# Multi-class random matrix filtering for adaptive learning

Paolo Braca, *Senior Member, IEEE*, Augusto Aubry, *Senior Member, IEEE*
Leonardo M. Millefiori *Member, IEEE*, Antonio De Maio, *Fellow, IEEE*, Stefano Marano

***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 learning strategy to track both the covariance matrix of data and its structure (class). We assume that, given the class, the posterior distribution of the covariance is described through a mixture of inverse Wishart distributions, while the class evolves according to a Markov chain. Hence, we devise a novel and general filtering strategy, called multi-class inverse Wishart mixture filter, able to capitalize on previous observations so as to accurately track and estimate the covariance. Some case studies are provided to highlight the effectiveness of the proposed technique, which is shown to outperform alternative methods in terms of both covariance estimation accuracy and probability of correct model selection. Specifically, the proposed filter is compared with class-clairvoyant covariance estimators, e.g., the maximum likelihood and the knowledge-based recursive least square filter, and with the model order selection method based on the Bayesian information criterion.**

***Index Terms*—Random matrices, covariance matrix estimation, interference covariance matrix, model classification, Bayesian information criterion, multi-class inverse Wishart mixture filter, radar and sonar signal processing, adaptive signal processing.**

## I. Introduction

The estimation of the 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, target tracking, and spectral analysis.

In radar and sonar applications, to predict the interference covariance matrix (ICM), conventional adaptive strategies (such as sample matrix inversion (SMI) [1] and 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 system's degrees of freedom, i.e., the useful signal dimension [3]. The above requirements represent important limitations, since typical size of homogeneous data is usually quite limited and, more important, a poor training data selection can imply severe performance degradation [4].

A possible strategy to circumvent the lack of a sufficient number of homogeneous secondary data is to exploit some a-priori information about the scene illuminated by the radar and reduce the unknown parameters at the estimation stage by enforcing appropriate structural models on the ICM. In this respect, several approaches have been proposed, where the training data is modeled as independent and identically distributed (i.i.d.), zero-mean, circularly symmetric Gaussian vectors (homogeneous environment). In [5] the maximum likelihood (ML) covariance matrix estimator is derived by modeling the disturbance as the sum of a colored interference plus a white contribution; relying on the ML principle and suitable covariance structures, advanced estimators have been proposed in [6]–[12], both for homogeneous and heterogeneous environments. Finally, Bayesian ICM estimators have been developed to cope with training data scarcity [13]–[19], where the ICM prior is assumed to be a complex inverse Wishart distribution. Evidently, adaptive signal processing algorithms based on the aforementioned covariance estimators may suffer performance degradation in the presence of model mismatches, e.g., due to changes in the operative conditions arising from meteorological phenomena and terrain changes or the appearance and disappearance of interference. A first attempt to overcome this drawback has been pursued in [20], where an adaptive classification of the ICM structure is addressed by resorting to the theory of model order selection (MOS) [21]. By doing so, the actual ICM model can be adaptively predicted and mismatch loss avoided.

### A. Contribution and related work

A general filtering framework is proposed to track, at each time scan $k$, a hybrid state that is composed by a discrete random variable $C_k$, representing the model or class, and a positive definite random matrix $\boldsymbol{R}_k$ whose dimension depends on $C_k$. The posterior distribution of $\boldsymbol{R}_k$, conditioned on $C_k$, is modeled by an inverse Wishart mixture (IWM) distribution, while $C_k$ is modeled as a Markov chain. In particular, similarly to the Gaussian sum filter framework [22], where the posterior is approximated by a Gaussian mixture (GM) distribution, in the proposed approach the covariance posterior, conditioned on $C_k$, is approximated by an IWM distribution. This choice is not arbitrary, as the IWM and GM are natural representations of

P. Braca and L. M. Millefiori are with the North Atlantic Treaty Organization (NATO) Science and Technology Organization (STO) Centre for Maritime Research and Experimentation (CMRE), La Spezia, Italy. E-mail: paolo.braca@cmre.nato.int, leonardo.millefiori@cmre.nato.int. P. Braca and L. M. Millefiori were supported by NATO Allied Command Transformation (ACT) via the Data Knowledge Operational Effectiveness (DKOE) project.

A. Aubry and A. De Maio are with the University of Naples, "Federico II", DIETI, Napoli, Italy. Email: augusto.aubry@unina.it, ademaio@unina.it.

S. Marano is with the University of Salerno, DIEM, Fisciano, Italy. Email: marano@unisa.it.

A preliminary version of this work was presented at EUSIPCO 2018, Rome, Italy, September 2018.

posterior distributions for positive definite random matrices [23] and random vectors [22], respectively. The proposed hybrid tracking approach is named multi-class inverse Wishart mixture (MC-IWM) filter.

Hybrid state approaches are common in the tracking literature and used when the primary objective is to track not only (possibly multiple) targets, but also to estimate nuisance parameters, for instance those of the process noise. A notable example is the interactive multiple model (IMM) filter [24], where the process noise switches among discrete predefined classes. More recent developments are reported in [25]–[27], where the approach is applied also to other parameters, such as target detection probability and clutter rate. In [28], [29] the IMM approach has been combined with the measurement noise covariance estimation in the context of adaptive filtering for jump Markov systems. The main difference with the aforementioned works is that the MC-IWM filter tracks positive definite random matrices, i.e., covariance matrices, whose dimensions depend on the underlying models, which can also be nested, as usual in MOS problems [20], [21].

Random matrices are commonly used in tracking problems to model extended targets. Originally proposed by Koch [30], this approach adopts the IW distribution for the posterior distribution of the target extension (e.g., as measured by high resolution sensors), or to model a coordinated group of targets. Similarly, the random finite set (RFS) implementation of the extended target tracking (ETT), proposed in [31], [32], models the probability hypothesis density (PHD) as a mixture of gamma Gaussian IW distributions. Nowadays, the ETT is applied to a number of practical problems involving different sensor technologies, including camera, X-band radar, light detection and ranging (LIDAR) [23]. In this context, the MC-IWM filter can be regarded as a natural extension of random matrix based ETT filters to the case of targets whose measurements can be generated by different models, such as targets with peculiar symmetries or negligible dimensions. Although the proposed approach is general enough to be applicable also in this context, the present work is focused on adaptive environment classification and awareness, specifically for radar and sonar sensors; its application to ETT is left to future investigations.

In the landscape of adaptive environment classification techniques for radar application, the proposed approach has the following innovative peculiar features. First, the MC-IWM filter sequentially processes multiple time scan observations of the scene, whereas in [20] observations are processed one by one. Second, the proposed filter tracks both the ICM and its structure, so as to achieve a dynamic environment learning capability. To this end, as in [14], [18], we resort to Bayesian methods, except that a sequential approach on multiple time scans is herein adopted.

Summarizing, a unified environment model is proposed that includes several disturbance classes of practical interest, and a suitable probabilistic time transition from $(\boldsymbol{R}_k, C_k)$ to $(\boldsymbol{R}_{k+1}, C_{k+1})$. Given that the transition from $C_k$ to $C_{k+1}$ can imply a change of the matrix dimensions, a suitable moment matching approach is developed to address prediction and update steps. Analytical closed-form solutions are provided for some scenarios of practical interest. For arbitrary models, a tailored Monte Carlo-based solution for the prediction and the update steps is also presented. Other than the moment matching, the criterion of minimizing the Kullback-Leibler divergence can be also considered, see e.g. [33]–[36].

The paper is organized as follows. In Sec. II we formalize the problem, while in Sec. III the proposed filtering strategy is derived and the algorithm structure is discussed. Specific ICM models are provided in Sec. IV. Results of computer experiments are reported in Sec. V and concluding remarks are given in Sec. VI. Finally, three appendices are devoted to the mathematical details related to the moment matching and the derivation of the prediction and update steps. A preliminary version of the MC-IWM filter has been presented in [37].

## Notation

We adopt the notation of using boldface for vectors $\boldsymbol{a}$ (lower case), and matrices $\boldsymbol{A}$ (upper case). The transpose, the conjugate, and the conjugate transpose operators are denoted by the symbols $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^\dagger$, respectively. $\boldsymbol{I}$ and $\boldsymbol{0}$ denote respectively the identity matrix and the matrix with zero entries (their size is determined from the context). The Kronecker and Hadamard products are indicated as $\otimes$ and $\odot$, respectively. $\mathbb{C}^N$, $\mathbb{C}^{N,K}$, $\mathbb{S}^N_{++}$, $\mathbb{H}^N$, $\mathbb{H}^N_+$, and $\mathbb{H}^N_{++}$ are respectively the sets of $N$-dimensional vectors of complex numbers, of $N \times K$ matrices of complex numbers, of $N \times N$ positive definite symmetric matrices, of $N \times N$ Hermitian matrices, of $N \times N$ positive semidefinite Hermitian matrices, and of $N \times N$ positive definite Hermitian matrices. For any $\boldsymbol{R} \in \mathbb{C}^{N,N}$, $\mathbf{diag}_{K_1}(\boldsymbol{R}) \in \mathbb{C}^{K_1 N, K_1 N}$ is a block diagonal matrix with $K_1$ blocks given by $\boldsymbol{R}$. For any matrix $\boldsymbol{A} \in \mathbb{C}^{K_1 m, K_1 m}$, $\mathrm{D}[\boldsymbol{A}]_i^m$, $i = 0, \ldots, K_1 - 1$, denotes the square matrix of size $m$ obtaining extracting from $\boldsymbol{A}$ the entries $i_1, i_2 \in \{im+1, \ldots, (i+1)m\}$. The curled inequality symbol $\succeq$ (and its strict form $\succ$) is used to denote generalized matrix inequality: for any $\boldsymbol{A} \in \mathbb{H}^N$, $\boldsymbol{A} \succeq \boldsymbol{0}$ means that $\boldsymbol{A}$ is a positive semi-definite matrix ($\boldsymbol{A} \succ \boldsymbol{0}$ for positive definiteness). $|\boldsymbol{X}|$ denotes the determinant of $\boldsymbol{X} \in \mathbb{C}^{N,N}$. Besides, for any set $\mathcal{A}$, $|\mathcal{A}|$ represents the cardinality of $\mathcal{A}$. The real part of $\boldsymbol{A}$ is indicated with $\Re\{A\}$. Finally, $\mathbb{E}[\cdot]$ denotes statistical expectation.

## II. Problem Formulation

Let $\boldsymbol{Z}_k = [\boldsymbol{z}_{1,k}, \cdots, \boldsymbol{z}_{N,k}] \in \mathcal{Z}^{m,N}$, $k \geq 1$, be the measurements collected by the acquisition system at the $k$-th scan, whose entries can be complex $\mathcal{Z} = \mathbb{C}$ or real $\mathcal{Z} = \mathbb{R}$. The goal is to sequentially estimate the positive definite matrix $\boldsymbol{M}_k$ associated with the measurements $\boldsymbol{Z}_k$ given the data up to $k$, i.e., $\boldsymbol{Z}_{1:k} := \{\boldsymbol{Z}_1, \boldsymbol{Z}_2, \cdots, \boldsymbol{Z}_k\}$. In general, the measurement equation can be expressed as follows

$$\boldsymbol{Z}_k = \boldsymbol{h}(\boldsymbol{M}_k, \boldsymbol{Z}_k^w), \tag{1}$$

where $\boldsymbol{h}(\cdot)$ is a generic function, $\boldsymbol{M}_k$ is a positive definite random matrix of size $m$, $\boldsymbol{Z}_k^w$ is a random matrix whose entries are assumed i.i.d. The sequence $\{\boldsymbol{Z}_1^w, \boldsymbol{Z}_2^w, \cdots, \boldsymbol{Z}_k^w\}$ is assumed to be statistically independent from $\{\boldsymbol{M}_1, \boldsymbol{M}_2, \cdots, \boldsymbol{M}_k\}, \forall k \geq 1$.

In the case of radar (sonar) systems [20] $\boldsymbol{z}_{i,k}$, $i = 1, \ldots, N$, may represent the measurements from the $i$-th range bin in the surveillance area, and $m$ may denote the number of space, time or space-time channels. Then, focusing on the ICM estimation problem in the presence of a homogeneous clutter environment the previous equation is specialized as follows [20]

$$\boldsymbol{Z}_k = \boldsymbol{M}_k^{\frac{1}{2}} \boldsymbol{Z}_k^w, \tag{2}$$

where $\boldsymbol{M}_k$ is the ICM, and the entries of $\boldsymbol{Z}_k^w \in \mathbb{C}^{m,N}$ are zero-mean circularly symmetric complex Gaussian random variables with unit variance. This means that, conditioned on $\boldsymbol{M}_k$, the columns of $\boldsymbol{Z}_k$ are i.i.d. zero-mean circularly symmetric complex Gaussian random vectors with covariance matrix $\boldsymbol{M}_k$.

In the ETT application, see e.g. [30], $\boldsymbol{z}_{i,k}$ represents the $i$-th detection coordinate of an extended target at scan $k$, while $m$ denotes the coordinate dimension. Assuming a linear model, $\boldsymbol{z}_{i,k}$ is normally distributed around the target position with covariance $\boldsymbol{M}_k$, the measurement equation (1) is specialized as follows

$$\boldsymbol{Z}_k = \boldsymbol{1}_N^T \otimes \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{M}_k^{\frac{1}{2}} \boldsymbol{Z}_k^w, \tag{3}$$

where $\boldsymbol{1}_N$ is the unitary column vector of dimension $N$, $\boldsymbol{x}_k \in \mathbb{R}^n$ is the target kinematic state, $\boldsymbol{H}_k \in \mathbb{R}^{m,n}$ is the observation matrix, $\boldsymbol{M}_k$ contains the information about the target extension, and the entries of $\boldsymbol{Z}_k^w \in \mathbb{R}^{m,N}$ are zero-mean Gaussian random variables with unit variance. The measurement model in (3) can be framed immediately as in (2) assuming known the target kinematic $\boldsymbol{x}_k$. The extension taking into account both the target kinematic and the target size can be derived following the ETT framework, e.g. see [30]–[32], but is out of the scope of this work.

### A. Covariance matrix class

It is assumed the availability of prior knowledge about the possible structure of the covariance $\boldsymbol{M}_k$, which induces a partition $\{\mathcal{A}_1, \mathcal{A}_2, \ldots \mathcal{A}_{N_C}\}$ on the set of positive definite matrices. Specifically, at each time scan $k$, $\boldsymbol{M}_k$ belongs to the class $C_k \in \mathcal{C} := \{1, 2, \ldots, N_C\}$, such that $\boldsymbol{M}_k \in \mathcal{A}_{C_k}$, where $\mathcal{A}_{C_k}$ is the set of the matrices exhibiting a specific structure. In this context, $\mathcal{A}_C$ represents the range of a function $f_C(\cdot)$, whose domain is the set of positive definite matrices with a specific size (not necessarily equal to $m$) depending on the specific class $C$. Thus, the matrix $\boldsymbol{M}_k$ can be written as follows:

$$\boldsymbol{M}_k = f_{C_k}(\boldsymbol{R}_k), \qquad C_k \in \mathcal{C}, \tag{4}$$

where $\boldsymbol{R}_k$ is a positive definite matrix, whose size is a function of $C_k$. Otherwise stated, $\boldsymbol{M}_k$ belongs to a specific (but unknown) class of the positive definite matrices, parametrized via a (possibly lower dimensional) matrix $\boldsymbol{R}_k$. For instance, the class of white noise is described by $\boldsymbol{M} = f(R) = R\,\boldsymbol{I}$, where $R > 0$ is a one-dimensional variable and $\boldsymbol{I}$ 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 $k$, $\boldsymbol{Z}_{1:k} = \{\boldsymbol{Z}_1, \boldsymbol{Z}_2, \cdots, \boldsymbol{Z}_k\}$. The current hybrid state at step $k$ is defined as $\boldsymbol{X}_k := \{\boldsymbol{R}_k, C_k\}$ and uniquely determines the matrix $\boldsymbol{M}_k$. The estimation is based on the posterior distribution of $\boldsymbol{X}_k$ given the data observed up to time $k$

$$\begin{aligned}\mathcal{P}_{k|k}(\boldsymbol{X}_k) &:= \mathcal{P}(\boldsymbol{X}_k \,|\boldsymbol{Z}_{1:k}) \\ &= \mathcal{P}(\boldsymbol{R}_k \,|C_k, \boldsymbol{Z}_{1:k})\, \mathcal{P}(C_k \,|\boldsymbol{Z}_{1:k}) \\ &:= \mathcal{P}_{k|k}(\boldsymbol{R}_k \,|C_k)\, \mathcal{P}_{k|k}(C_k),\end{aligned} \tag{5}$$

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

### B. Hierarchical Markov assumption

In the following, it is supposed that the class $C_k$ evolves according to a Markov chain with finite sample space and transition matrix defined as $\pi_{ij}$, i.e.,

$$\mathcal{P}(C_{k+1} = i \,|C_k = j) = \pi_{ij}, \quad i, j \in \mathcal{C}. \tag{6}$$

In particular, we assume a hierarchical structure between the class and the matrix evolution as follows

$$\mathcal{P}(C_{k+1} \,|C_k, \boldsymbol{R}_k, \ldots, C_1, \boldsymbol{R}_1,) = \mathcal{P}(C_{k+1} \,|C_k). \tag{7}$$

Moreover, conditioned on $C_{k+1}, C_k$ and $\boldsymbol{R}_k$, we assume that $\boldsymbol{R}_{k+1}$ is statistically independent of the previous states, namely

$$\begin{aligned}\mathcal{P}(\boldsymbol{R}_{k+1} \,|C_{k+1}, C_k, \boldsymbol{R}_k, \ldots, C_1, \boldsymbol{R}_1) = \\ \mathcal{P}(\boldsymbol{R}_{k+1} \,|C_{k+1}, C_k, \boldsymbol{R}_k).\end{aligned} \tag{8}$$

Based on the aforementioned hierarchical evolution model, it follows that

$$\begin{aligned}\mathcal{P}(\boldsymbol{X}_{k+1} \mid \boldsymbol{X}_k, \ldots, \boldsymbol{X}_1) &= \mathcal{P}(C_{k+1} \mid \boldsymbol{X}_k, \ldots, \boldsymbol{X}_1) \\ &\quad \times \mathcal{P}(\boldsymbol{R}_{k+1} \mid C_{k+1}, \boldsymbol{X}_k, \ldots, \boldsymbol{X}_1) \\ &= \mathcal{P}(C_{k+1} \,|C_k) \\ &\quad \times \mathcal{P}(\boldsymbol{R}_{k+1} \,|C_{k+1}, C_k, \boldsymbol{R}_k) \\ &= \mathcal{P}(\boldsymbol{X}_{k+1} \mid \boldsymbol{X}_k);\end{aligned}$$

namely, the time evolution of $\boldsymbol{X}_k$ follows a Markov process. Fig. 1 illustrates a notional behavior of the considered hidden hierarchical Markov chain. In particular, the state $\boldsymbol{X}_k$ can only be observed indirectly (hidden model) via the available measurements $\boldsymbol{Z}_k$. Additionally, the state is described through two different random objects, i.e., $C_k$ and $\boldsymbol{R}_k$, that present a hierarchical relationship, and the structure of $\boldsymbol{R}_k$ depends on the actual value of $C_k$.

### C. Random matrix state transition

To fix ideas, we provide practical state transition rule valid for both the cases of ICM estimation and ETT. If $C_{k+1}$ and $C_k$ are the same, i.e., $C_{k+1} = C_k = i$, then it is reasonable to have a transition with conditional constant mean [30], i.e.,

$$\mathbb{E}[\boldsymbol{R}_{k+1} \,|C_k = i, C_{k+1} = i, \boldsymbol{R}_k] = \boldsymbol{R}_k.$$

Clearly, more sophisticated transitions can also be taken into account. Considering that we deal with positive definite

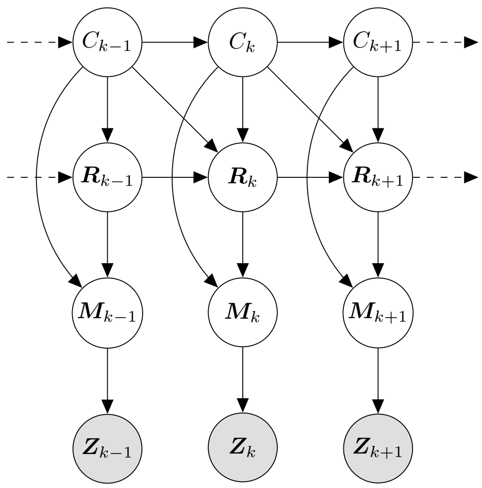



Fig. 1. Hierarchical Markov model of the hybrid state $\boldsymbol{X}_k = \{\boldsymbol{R}_k, C_k\}$.

matrices, we assume that the spread around the mean is ruled by a Wishart distribution[1], yielding [30]

$$\mathcal{P}\left(\boldsymbol{R}_{k+1}\,|C_k = i, C_{k+1} = i, \boldsymbol{R}_k\right) = \mathcal{W}\left(\boldsymbol{R}_{k+1}; \frac{\boldsymbol{R}_k}{\nu}, \nu\right), \tag{9}$$

where we have indicated, with some abuse of symbolism, by $\mathcal{W}(\boldsymbol{R}; \boldsymbol{M}, \nu)$, both the real and the complex Wishart distribution with scale matrix $\boldsymbol{M}$ and $\nu$ degrees of freedom. The choice between real and complex Wishart distribution depends on the field where $\boldsymbol{R}_k$, given $C_k$, is defined and will be easily determined from the context. If necessary, the transition of $\boldsymbol{R}_k$ can involve rotation and scaling transformation, see [35].

Finally, if $C_k \neq C_{k+1}$, a Wishart distribution, but with different and suitable parameters, can be used to model the transition from $k$ to $k+1$. Essentially, a suitable mapping of $\boldsymbol{M}_k = f_{C_k}(\boldsymbol{R}_k)$ onto $\mathcal{A}_{C_{k+1}}$ is considered as the mean matrix of the Wishart distribution; the details are reported in Appendix III and specific examples will be illustrated in the following sections.

## III. Multi-class inverse Wishart mixture filter

With the assumptions made in the previous section, the random matrix under each class can be inferred by approximating posterior distributions of the corresponding random matrix $\boldsymbol{R}_k$ with mixtures of IW components. This is similar to the approach taken in [22], where posterior distributions are approximated by mixtures of Gaussian components. Besides, the choice of IW components originates from the Gaussian nature of the data $\boldsymbol{Z}_k$ (conditioned on $\boldsymbol{M}_k$) and from the IW distribution being the conjugate prior for $\boldsymbol{M}_k$, e.g., see [16], [30]. In general, the moment matching approximation can be applied to preserve the IWM structure of the posterior.

The proposed filtering approach allows to sensibly reduce the complexity w.r.t. a brute force particle filtering strategy. Indeed, as $m$ increases, but even for moderately small values, the accurate representation of random objects in $\mathbb{C}^{m\times m}$ would require a prohibitively large number of particles, see also the discussion in [39].

In order to proceed, let us assume that the predicted posterior for each class at time $k$ given $\boldsymbol{Z}_{1:k-1}$ is a mixture of (complex or real, according to the field where $\boldsymbol{R}_k$, given $C_k$, is defined) inverse Wishart [16], [30]

$$\mathcal{P}_{k|k-1}\left(\boldsymbol{R}_k\,|C_k\right) = \sum_{n=1}^{N_W} w_{k|k-1}^{(n,C_k)}\,\mathcal{IW}(\boldsymbol{R}_k;\, \widehat{\boldsymbol{R}}_{k|k-1}^{(n,C_k)},\, \hat{\nu}_{k|k-1}^{(n,C_k)}), \tag{10}$$

where $\hat{\boldsymbol{R}}_{k|k-1}^{(n,C_k)}$ and $\hat{\nu}_{k|k-1}^{(n,C_k)}$ are the scale matrix and degrees of freedom associated with the $n$-th inverse Wishart component of the mixture characterizing the class $C_k$. The weight $w_{k|k-1}^{(n,C_k)}$ is the probability of the $n$-th component of the mixture at time $k$ for the class $C_k$, having observed the data up to time $k-1$:

$$w_{k|k-1}^{(n,C_k)} := \mathcal{P}(N_k = n\,|C_k,\, \boldsymbol{Z}_{1:k-1}) = \mathcal{P}_{k|k-1}(n\,|C_k), \tag{11}$$

where $N_k$ defines the auxiliary discrete random variable that models the switch among the different modes, from $n=1$ to $n=N_W$. Hereafter, the variable $n$ refers to a specific value of the random variable $N_k$, i.e., $N_k = n$. The quantities $\hat{\boldsymbol{R}}_{k|k-1}^{(n,C_k)}$, $\hat{\nu}_{k|k-1}^{(n,C_k)}$, and $w_{k|k-1}^{(n,C_k)}$ summarize the information acquired from the data up to time $k-1$, related to both $C_k$ and $\boldsymbol{R}_k$. When a new set of observables $\boldsymbol{Z}_k$ is gathered, the aforementioned quantities are updated.

### A. Matrix Update Step

In this subsection we compute $\mathcal{P}_{k|k}\left(\boldsymbol{R}_k\,|C_k\right)$, $C_k \in \mathcal{C}$, namely, the statistical characterization of $\boldsymbol{R}_k$, given $C_k$ and $\boldsymbol{Z}_{1:k}$. From the law of total probability it stems

$$\mathcal{P}_{k|k}(\boldsymbol{R}_k|C_k) = \sum_{n=1}^{N_W} w_{k|k}^{(n,C_k)}\,\mathcal{P}_{k|k}(\boldsymbol{R}_k|n, C_k), \tag{12}$$

where

$$w_{k|k}^{(n,C_k)} := \mathcal{P}(N_k = n|C_k, \boldsymbol{Z}_{1:k}),$$
$$\mathcal{P}_{k|k}(\boldsymbol{R}_k|n, C_k) := \mathcal{P}(\boldsymbol{R}_k|N_k = n, C_k, \boldsymbol{Z}_{1:k}).$$

The updated mixture weight is computed by Bayes' rule as

$$\begin{aligned} w_{k|k}^{(n,C_k)} &= \frac{\mathcal{P}(\boldsymbol{Z}_k|n, C_k, \boldsymbol{Z}_{1:k-1})w_{k|k-1}^{(n,C_k)}}{\sum_{n=1}^{N_W}\mathcal{P}(\boldsymbol{Z}_k|n, C_k, \boldsymbol{Z}_{1:k-1})w_{k|k-1}^{(n,C_k)}} \\ &= \frac{\alpha_k^{(n,C_k)} w_{k|k-1}^{(n,C_k)}}{\sum_{n=1}^{N_W}\alpha_k^{(n,C_k)} w_{k|k-1}^{(n,C_k)}}, \end{aligned} \tag{13}$$

where we have defined

$$\begin{aligned} \alpha_k^{(n,C_k)} &:= \mathcal{P}(\boldsymbol{Z}_k|n, C_k, \boldsymbol{Z}_{1:k-1}) \qquad (14) \\ &\overset{(a)}{=} \int \mathcal{P}(\boldsymbol{Z}_k|C_k, \boldsymbol{R}_k)\,\mathcal{P}_{k|k-1}(\boldsymbol{R}_k|n, C_k)\,\mathrm{d}\boldsymbol{R}_k \\ &= \int \mathcal{P}(\boldsymbol{Z}_k|C_k, \boldsymbol{R}_k) \\ &\quad \times \mathcal{IW}\left(\boldsymbol{R}_k;\, \widehat{\boldsymbol{R}}_{k|k-1}^{(n,C_k)},\, \hat{\nu}_{k|k-1}^{(n,C_k)}\right)\mathrm{d}\boldsymbol{R}_k, \end{aligned}$$

and in $(a)$ we have exploited the conditional independence of $\boldsymbol{Z}_k$ from $\boldsymbol{Z}_{1:k-1}$ and $N_k = n$, given the state $\boldsymbol{X}_k = (C_k, \boldsymbol{R}_k)$, i.e.,

$$\mathcal{P}(\boldsymbol{Z}_k|n, C_k, \boldsymbol{R}_k, \boldsymbol{Z}_{1:k-1}) = \mathcal{P}(\boldsymbol{Z}_k|C_k, \boldsymbol{R}_k). \tag{15}$$

[1]The use of the matrix-Gamma distribution [38] can be also considered.

We proceed analogously to compute the updated components of the $n$-th mixand:

$$\mathcal{P}_{k|k}(\boldsymbol{R}_k|n, C_k) = \frac{\mathcal{P}(\boldsymbol{Z}_k|C_k, \boldsymbol{R}_k)\mathcal{P}_{k|k-1}(\boldsymbol{R}_k|n, C_k)}{\int \mathcal{P}(\boldsymbol{Z}_k|C_k, \boldsymbol{R}_k)\mathcal{P}_{k|k-1}(\boldsymbol{R}_k|n, C_k)\,\mathrm{d}\boldsymbol{R}_k}. \tag{16}$$

Being the prior a complex (or real) IW, i.e.,

$$\mathcal{P}_{k|k-1}(\boldsymbol{R}_k|n, C_k) = \mathcal{IW}(\boldsymbol{R}_k;\, \widehat{\boldsymbol{R}}^{(n,C_k)}_{k|k-1},\, \hat{\nu}^{(n,C_k)}_{k|k-1})$$

and the data complex (or real) multivariate Gaussian distributed, eq. (16) is again (or can be approximated as) an IW (see Appendix I):

$$\mathcal{P}_{k|k}(\boldsymbol{R}_k|n, C_k) = \mathcal{IW}(\boldsymbol{R}_k;\, \widehat{\boldsymbol{R}}^{(n,C_k)}_{k|k},\, \hat{\nu}^{(n,C_k)}_{k|k}), \tag{17}$$

with the parameter $\hat{\boldsymbol{R}}^{(n,C_k)}_{k|k}$ and the degrees of freedom $\hat{\nu}^{(n,C_k)}_{k|k}$, that clearly depend on $\boldsymbol{Z}_k$, the previous IW parameters, as well as the class model. Formally:

$$\left(\widehat{\boldsymbol{R}}^{(n,C_k)}_{k|k},\, \hat{\nu}^{(n,C_k)}_{k|k}\right) = f_U\left(\boldsymbol{Z}_k, \widehat{\boldsymbol{R}}^{(n,C_k)}_{k|k-1},\, \hat{\nu}^{(n,C_k)}_{k|k-1}, C_k\right), \tag{18}$$

where the function $f_U(\cdot)$ can be obtained according to the guidelines in Appendix I; specific instances of this update equation are provided in Sec. IV. Summarizing, the posterior update of $\boldsymbol{R}_k$ is given by

$$\boxed{\begin{aligned} \mathcal{P}_{k|k}(\boldsymbol{R}_k|C_k) &= \sum_{n=1}^{N_W} w^{(n,C_k)}_{k|k} \mathcal{IW}\left(\boldsymbol{R}_k;\, \widehat{\boldsymbol{R}}^{(n,C_k)}_{k|k},\, \hat{\nu}^{(n,C_k)}_{k|k}\right) \\ w^{(n,C_k)}_{k|k} &= \frac{\alpha^{(n,C_k)}_k w^{(n,C_k)}_{k|k-1}}{\sum_{n=1}^{N_W} \alpha^{(n,C_k)}_k w^{(n,C_k)}_{k|k-1}} \end{aligned}}$$

where the the IW parameters are provided in (18) and $\alpha^{(n,C_k)}_k$ is given by (14).

### B. Class Update Step

Let us assume now that the class probability $\mathcal{P}_{k|k-1}(C_k)$, $C_k \in \mathcal{C}$, is available at time $k$, given the data up to time $k-1$. For ease of notation, let us define the class probability as $p_{k|j}(C_k) := \mathcal{P}_{k|j}(C_k)$. When the new observation $\boldsymbol{Z}_k$ is available, the class probability is updated as follows

$$p_{k|k}(C_k) = \frac{\mathcal{P}(\boldsymbol{Z}_k|C_k, \boldsymbol{Z}_{1:k-1})p_{k|k-1}(C_k)}{\sum_{c\in\mathcal{C}} \mathcal{P}(\boldsymbol{Z}_k|C_k = c, \boldsymbol{Z}_{1:k-1})p_{k|k-1}(c)}. \tag{19}$$

Using the law of total probability, the first term of the numerator in (19) can be expressed as

$$\begin{aligned} \mathcal{P}(\boldsymbol{Z}_k|C_k, \boldsymbol{Z}_{1:k-1}) &= \sum_{n=1}^{N_W} \mathcal{P}(\boldsymbol{Z}_k|n, C_k, \boldsymbol{Z}_{1:k-1}) w^{(n,C_k)}_{k|k-1}, \\ &\overset{(a)}{=} \sum_{n=1}^{N_W} \alpha^{(n,C_k)}_k w^{(n,C_k)}_{k|k-1}, \end{aligned} \tag{20}$$

where in $(a)$ we have used (14). Substituting (20) in (19) we finally obtain the update rule for the class probability

$$\boxed{p_{k|k}(C_k) = \frac{\left(\sum_{n=1}^{N_W} w^{(n,C_k)}_{k|k-1}\, \alpha^{(n,C_k)}_k\right) p_{k|k-1}(C_k)}{\sum_{c\in\mathcal{C}} \sum_{n=1}^{N_W} w^{(n,c)}_{k|k-1}\, \alpha^{(n,c)}_k p_{k|k-1}(c)}.}$$

The update steps are summarized in Algorithm 1.

In the following subsections we will describe the prediction procedure to establish the joint posterior distribution of the class and the related random matrix, namely $\boldsymbol{X}_{k+1} = \{\boldsymbol{R}_{k+1}, C_{k+1}\}$, at time $k+1$, given the observations up to time $k$:

$$\mathcal{P}_{k+1|k}(\boldsymbol{X}_{k+1}) = \mathcal{P}_{k+1|k}(\boldsymbol{R}_{k+1}|C_{k+1})\, p_{k+1|k}(C_{k+1}). \tag{21}$$

### C. Class Prediction Step

The predicted class probability is computed by means of the Markov chain assumption (6) as follows:

$$\begin{aligned} p_{k+1|k}(C_{k+1}) &= \sum_{C_k\in\mathcal{C}} \mathcal{P}(C_{k+1}|C_k, \boldsymbol{Z}_{1:k})\, \mathcal{P}(C_k|\boldsymbol{Z}_{1:k}) \\ &\overset{(a)}{=} \sum_{C_k\in\mathcal{C}} \mathcal{P}(C_{k+1}|C_k) p_{k|k}(C_k), \end{aligned} \tag{22}$$

where in (a) we have exploited the independence of $C_{k+1}$ from the data up to $k$ given $C_k$, which stems from the hierarchical Markov assumption discussed in Sec. II-B.

In summary, the predicted class probability is

$$\boxed{p_{k+1|k}(C_{k+1}) = \sum_{C_k\in\mathcal{C}} \mathcal{P}(C_{k+1}|C_k)\, p_{k|k}(C_k).}$$

### D. Matrix Prediction Step

Focusing now on the matrix prediction step, i.e., $\boldsymbol{R}_{k+1}$ given $\boldsymbol{Z}_{1:k}$ and $C_{k+1}$, the law of total probability on $C_k$ can be exploited as follows

$$\begin{aligned} \mathcal{P}_{k+1|k}(\boldsymbol{R}_{k+1}|C_{k+1}) = \sum_{C_k\in\mathcal{C}} &\mathcal{P}(\boldsymbol{R}_{k+1}|C_{k+1}, C_k, \boldsymbol{Z}_{1:k}) \\ &\times \mathcal{P}(C_k|C_{k+1}, \boldsymbol{Z}_{1:k}), \end{aligned} \tag{23}$$

where the last term in the previous summation can be recast applying Bayes' rule:

$$\begin{aligned} \mathcal{P}(C_k|C_{k+1}, \boldsymbol{Z}_{1:k}) &= \frac{\overbrace{\mathcal{P}(C_{k+1}|C_k, \boldsymbol{Z}_{1:k})}^{\mathcal{P}(C_{k+1}|C_k)}\, \overbrace{\mathcal{P}(C_k|\boldsymbol{Z}_{1:k})}^{\mathcal{P}_{k|k}(C_k)}}{\underbrace{\mathcal{P}(C_{k+1}|\boldsymbol{Z}_{1:k})}_{\mathcal{P}_{k+1|k}(C_{k+1})}} \\ &= \frac{\mathcal{P}(C_{k+1}|C_k)p_{k|k}(C_k)}{p_{k+1|k}(C_{k+1})}. \end{aligned} \tag{24}$$

Using again the total probability rule on $\boldsymbol{R}_k$, the first factor at the right-hand side of (23) can be expressed as

$$\begin{aligned} \mathcal{P}(\boldsymbol{R}_{k+1}|C_{k+1}, C_k, \boldsymbol{Z}_{1:k}) = \int &\overbrace{\mathcal{P}(\boldsymbol{R}_{k+1}|C_{k+1}, C_k, \boldsymbol{R}_k, \boldsymbol{Z}_{1:k})}^{\mathcal{P}(\boldsymbol{R}_{k+1}|C_{k+1}, C_k, \boldsymbol{R}_k)} \\ &\times \mathcal{P}(\boldsymbol{R}_k|C_{k+1}, C_k, \boldsymbol{Z}_{1:k})\mathrm{d}\boldsymbol{R}_k, \end{aligned} \tag{25}$$

where the dependence on $\boldsymbol{Z}_{1:k}$ is deleted in the first term of the integrand, being $\boldsymbol{R}_{k+1}$ independent of $\boldsymbol{Z}_{1:k}$ conditioned on $C_{k+1}, C_k$, and $\boldsymbol{R}_k$. The second term can be computed as

$$\begin{aligned}\mathcal{P}(\boldsymbol{R}_k|C_{k+1},C_k,\boldsymbol{Z}_{1:k}) &= \frac{\mathcal{P}(\boldsymbol{R}_k,C_{k+1},C_k|\boldsymbol{Z}_{1:k})}{\mathcal{P}(C_{k+1},C_k|\boldsymbol{Z}_{1:k})}\\ &\overset{(a)}{=} \frac{\mathcal{P}(\boldsymbol{R}_k,C_k|\boldsymbol{Z}_{1:k})\,\overbrace{\mathcal{P}(C_{k+1}|C_k,\boldsymbol{R}_k,\boldsymbol{Z}_{1:k})}^{\mathcal{P}(C_{k+1}|C_k)}}{\mathcal{P}(C_k|\boldsymbol{Z}_{1:k})\,\underbrace{\mathcal{P}(C_{k+1}|C_k,\boldsymbol{Z}_{1:k})}_{\mathcal{P}(C_{k+1}|C_k)}}\\ &= \mathcal{P}_{k|k}(\boldsymbol{R}_k|C_k), \end{aligned} \tag{26}$$

where in $(a)$ we have exploited the hierarchical Markov property by which $C_{k+1}$ conditioned on $C_k$ is independent of $\boldsymbol{R}_k$ and $\boldsymbol{Z}_{1:k}$. Substituting (26) in (25), we obtain:

$$\begin{aligned}&\mathcal{P}(\boldsymbol{R}_{k+1}|C_{k+1},C_k,\boldsymbol{Z}_{1:k})\\ &\quad= \int \mathcal{P}(\boldsymbol{R}_{k+1}|C_{k+1},C_k,\boldsymbol{R}_k)\,\mathcal{P}_{k|k}(\boldsymbol{R}_k|C_k)\mathrm{d}\boldsymbol{R}_k\\ &\quad= \sum_{n=1}^{N_W} w_{k|k}^{(n,C_k)} \int \mathcal{P}(\boldsymbol{R}_{k+1}|C_{k+1},C_k,\boldsymbol{R}_k)\\ &\qquad\times \mathcal{IW}(\boldsymbol{R}_k;\, \widehat{\boldsymbol{R}}_{k|k}^{(n,C_k)};\, \hat{\nu}(n,C_k))\,\mathrm{d}\boldsymbol{R}_k. \end{aligned}\tag{27}$$

Enforcing the solution of the integral in the last equality of (27) to be again an IW via moment matching (see details in Appendix II) and marginalizing (27) w.r.t. $C_k$, see eq. (23), we obtain the prediction matrix distribution

$$\begin{aligned}\mathcal{P}_{k+1|k}(\boldsymbol{R}_{k+1}|C_{k+1}) &= \sum_{n=1}^{N_W}\sum_{C_k\in\mathcal{C}} w_{k+1|k}^{(n,C_{k+1},C_k)}\\ &\times\mathcal{IW}(\boldsymbol{R}_{k+1};\, \widehat{\boldsymbol{R}}_{k+1|k}^{(n,C_{k+1},C_k)};\, \hat{\nu}(n,C_{k+1},C_k)),\end{aligned}\tag{28}$$

where

$$\begin{aligned}w_{k+1|k}^{(n,C_{k+1},C_k)} &= w_{k|k}^{(n,C_k)}\,\mathcal{P}(C_k|C_{k+1},\boldsymbol{Z}_{1:k})\\ &= w_{k|k}^{(n,C_k)}\,\frac{\mathcal{P}(C_{k+1}|C_k)\,p_{k|k}(C_k)}{p_{k+1|k}(C_{k+1})}.\end{aligned}\tag{29}$$

The IW parameters $\widehat{\boldsymbol{R}}_{k+1|k}^{(n,C_{k+1},C_k)}$ and $\hat{\nu}^{(n,C_{k+1},C_k)}$ of the prediction step follow a suitable moment matching approximation, reported in Appendix II and III, that generalizes the approach in [30] including the cases with $C_k \neq C_{k+1}$ that accounts for the different matrix structures of distinct classes. In particular, each of the $N_W$ modes associated to a given $C_k$ will lead to a new mixand component. Such update is expressed by the function

$$\begin{aligned}&\left(\widehat{\boldsymbol{R}}_{k+1|k}^{(n,C_{k+1},C_k)},\, \hat{\nu}_{k+1|k}^{(n,C_{k+1},C_k)}\right)\\ &\qquad= f_P\left(\widehat{\boldsymbol{R}}_{k|k}^{(n,C_k)},\, \hat{\nu}_{k|k}^{(n,C_k)}, C_k, C_{k+1}\right).\end{aligned}\tag{30}$$

The function $f_P(\cdot)$ can be obtained according to the guidelines provided in Appendix II and III.

**Remark:** In the prediction step, the number of components of the mixture in each class increases to $N_W \times N_C$. In order to avoid an increase of the computational complexity, following [22], at each time update a pruning criterion is adopted. For instance, only the first $N_W$ components, sorted according to their weights are retained. Furthermore IW components close to each other can be also merged, see e.g. [33]. The new weights are then normalized to one and the modes again indexed via the finite set $\{1,\ldots,N_W\}$ (having removed, without loss of generality, the explicit dependency on the previous class).

In summary, the prediction distribution of the matrix conditioned on the class is

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

where $w_{k+1|k}^{(n,C_{k+1})}$, $\widehat{\boldsymbol{R}}_{k+1|k}^{(n,C_{k+1})}$ and $\hat{\nu}_{k+1|k}^{(n,C_{k+1})}$ refer, respectively, to the update weights (their sum over $n$ is equal to one), the scale matrix, and the degrees of freedom associated to the $n$-th, $n = 1,\ldots,N_W$, selected mode for the class $C_{k+1}$.

## IV. ICM Models for Homogeneous Clutter

This section is devoted to the specialization of the MC-IWM filter to radar and sonar applications, for the adaptive classification and estimation of the ICM of homogeneous clutter, e.g., see [1], [20]. This scenario is of primary interest in radar, as it approximates real operating environments [2], [40]–[42]. Note that in the presence of clutter outliers, a selection procedure should be exploited to censor outliers and come up with a set of homogeneous data [20].

Here we consider a scenario with three classes. The first class is the white noise class, with $\boldsymbol{M}_k = R_k\,\boldsymbol{I}$ and $R_k > 0$; the second class describes persymmetric matrices (which model the radar ICM for symmetrically spaced temporal or spatial samples), i.e., $\boldsymbol{M}_k = \boldsymbol{U}^\dagger\boldsymbol{R}_k\boldsymbol{U}$ [41], with $\boldsymbol{R}_k \in \mathbb{S}_{++}^m$ and being $\boldsymbol{U}$ the specific unitary matrix defined in [41]. Finally, in the third class $\boldsymbol{M}_k = \boldsymbol{R}_k$ is a generic $m \times m$ ICM.

Assuming homogeneous clutter, the data are distributed according to (2) and then the likelihood is given by

$$\begin{aligned}&\mathcal{P}(\boldsymbol{Z}_k|\boldsymbol{R}_k,C_k)\\ &\quad= \begin{cases} (\pi R_k)^{-Nm}\, e^{-\mathrm{Tr}\left(\boldsymbol{Z}_k\boldsymbol{Z}_k^\dagger R_k^{-1}\right)} & C_k = 1,\\ |\pi\boldsymbol{R}_k|^{-N}\, e^{-\mathrm{Tr}\left(\Re\left\{\boldsymbol{U}\boldsymbol{Z}_k\boldsymbol{Z}_k^\dagger\boldsymbol{U}^\dagger\right\}\boldsymbol{R}_k^{-1}\right)} & C_k = 2,\\ |\pi\boldsymbol{R}_k|^{-N}\, e^{-\mathrm{Tr}\left(\boldsymbol{Z}_k\boldsymbol{Z}_k^\dagger\boldsymbol{R}_k^{-1}\right)} & C_k = 3.\end{cases}\end{aligned}\tag{31}$$

The variables $\alpha_k^{(n,C_k)}$ in Algorithm 1 are given as follows

$$\alpha_k^{(n,C_k)} = \int \mathcal{P}\left(\boldsymbol{Z}_k|C_k,\boldsymbol{R}\right)\mathcal{IW}\left(\boldsymbol{R};\, \widehat{\boldsymbol{R}}_{k|k-1}^{(n,C_k)},\, \hat{\nu}_{k|k-1}^{(n,C_k)}\right)\mathrm{d}\boldsymbol{R}. \tag{32}$$

Specifically, for $C_k = 1$, eq. (32) specializes to

$$\begin{aligned}\alpha_k^{(n,1)} &= 2^{Nm}\pi^{-Nm}\left|\widehat{R}_k^{(n,1)}\right|^{\frac{\hat{\nu}_{k|k-1}^{(n,1)}}{2}}\,\frac{\kappa_1(2Nm+\hat{\nu}_{k|k-1}^{(n,1)})}{\kappa_1(\hat{\nu}_{k|k-1}^{(n,1)})}\\ &\quad\times\left|2\mathrm{Tr}\left(\boldsymbol{Z}_k\boldsymbol{Z}_k^\dagger\right)+\widehat{R}_{k|k-1}^{(n,1)}\right|^{-\frac{2Nm+\hat{\nu}_{k|k-1}^{(n,1)}}{2}},\end{aligned}\tag{33}$$

for $C_k = 2$ to

$$\alpha_k^{(n,2)} = 2^{Nm}\pi^{-Nm}\left|\widehat{\boldsymbol{R}}_{k|k-1}^{(n,2)}\right|^{\frac{\hat{\nu}_{k|k-1}^{(n,2)}}{2}} \frac{\kappa_m(2N+\hat{\nu}_{k|k-1}^{(n,2)})}{\kappa_m(\hat{\nu}_{k|k-1}^{(n,2)})} \times \left|2\Re\left\{\boldsymbol{U}\boldsymbol{Z}_k\boldsymbol{Z}_k^\dagger\boldsymbol{U}^\dagger\right\}+\widehat{\boldsymbol{R}}_{k|k-1}^{(n,2)}\right|^{-\frac{2N+\hat{\nu}_{k|k-1}^{(n,2)}}{2}}, \quad (34)$$

and for $C_k = 3$ to

$$\alpha_k^{(n,3)} = \pi^{-Nm}\left|\widehat{\boldsymbol{R}}_{k|k-1}^{(n,3)}\right|^{\hat{\nu}_{k|k-1}^{(n,3)}} \frac{\tilde{\kappa}_m(N+\hat{\nu}_{k|k-1}^{(n,3)})}{\tilde{\kappa}_m(\hat{\nu}_{k|k-1}^{(n,3)})} \times \left|\boldsymbol{Z}_k\boldsymbol{Z}_k^\dagger+\widehat{\boldsymbol{R}}_{k|k-1}^{(n,3)}\right|^{-(N+\hat{\nu}_{k|k-1}^{(n,3)})}, \quad (35)$$

where $\kappa_p(\nu)$ and $\tilde{\kappa}_p(\nu)$ are defined as

$$\kappa_p(\nu) = \pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left[\frac{1}{2}(\nu+1-i)\right], \quad (36)$$

$$\tilde{\kappa}_p(\nu) = \pi^{\frac{p(p-1)}{2}}\prod_{i=1}^{p}\Gamma\left[(\nu+1-i)\right]. \quad (37)$$

Furthermore, for any mode $n = 1,\ldots,N_W$, being $f_{C_k}(\boldsymbol{R}_k) = \boldsymbol{A}_{C_k}\mathbf{diag}_{K_{C_k}}(\boldsymbol{R}_k)\,\boldsymbol{A}_{C_k}^\dagger$, with $K_C = m$ for $C = 1$, $K_C = 1$ for $C = 2,3$, $\boldsymbol{A}_{C_k} \in \mathbb{C}^{m,m}$ and $\left(\boldsymbol{A}_{C_k}\boldsymbol{A}_{C_k}^\dagger\right) \succ \mathbf{0}$, the IW distribution is the the conjugate prior for $\boldsymbol{R}_k$, see details in Appendix I. Hence, it follows that the function $f_U(\cdot)$ defined in (18) is given by

$$\widehat{\boldsymbol{R}}_{k|k}^{(n,C_k)} = \begin{cases} 2\,\mathrm{Tr}\left(\boldsymbol{Z}_k\boldsymbol{Z}_k^\dagger\right)+\widehat{R}_{k|k-1}^{(n,C_k)} & C_k = 1, \\ 2\Re\left\{\boldsymbol{U}\boldsymbol{Z}_k\boldsymbol{Z}_k^\dagger\boldsymbol{U}^\dagger\right\}+\widehat{\boldsymbol{R}}_{k|k-1}^{(n,C_k)} & C_k = 2, \\ \boldsymbol{Z}_k\boldsymbol{Z}_k^\dagger+\widehat{\boldsymbol{R}}_{k|k-1}^{(n,C_k)} & C_k = 3, \end{cases}$$

and

$$\hat{\nu}_{k|k}^{(n,C_k)} = \begin{cases} 2Nm+\hat{\nu}_{k|k-1}^{(n,C_k)} & C_k = 1, \\ 2N+\hat{\nu}_{k|k-1}^{(n,C_k)} & C_k = 2, \\ N+\hat{\nu}_{k|k-1}^{(n,C_k)} & C_k = 3. \end{cases}$$

Finally, the specialization of (28) to the three-class example is not reported here for brevity, but easily follows from the general treatment developed in Appendix II and III.

## V. Computer Experiments

This section is devoted to the performance analysis of the proposed MC-IWM filter on simulated data and its comparison with other methods. We assume homogeneous Gaussian disturbance generated according to (2) and several time-varying structures of the ICM $\boldsymbol{M}_k$; a case study with the appearance of an additional interference term is also included. In the computer simulations, the hybrid state $(\boldsymbol{R}_k, C_k)$ over time does not necessarily evolve according to the statistical model assumed in the previous sections (i.e., the state can be piecewise constant in time). This mismatch is intentional and proves the robustness of the proposed approach. In all the simulations, the degrees of freedom of the transition Wishart distribution (9) is set to $\nu = N\tau$, with $N\tau > m+4$, if $C_k = C_{k+1}$, and to $\nu = m+4$ otherwise, in order to allow a larger spread of any IW around its mean when the class changes from $k$ to $k+1$; the parameter $\tau$ can be seen as the analogous of a forgetting factor for the MC-IWM filter, and can be tuned depending on the specific application. Among all the available Bayesian estimators, we select the posterior mean, defined as:

$$\widehat{\boldsymbol{M}}_{k|k} = \mathbb{E}\left[\boldsymbol{M}_k|\boldsymbol{Z}_{1:k}\right] = \sum_{c=1}^{N_C}\sum_{n=1}^{N_W} p_{k|k}(c)\,w_{k|k}^{(n,c)}\,\widehat{\boldsymbol{M}}_{k|k}(n,c),$$

where $\widehat{\boldsymbol{M}}_{k|k}(n,c)$ is the expected mean conditioned to the $n$-th mode of the IW mixture of 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 $\widehat{\boldsymbol{R}}/(\hat{\nu}-p-1)$ if real, or $\widehat{\boldsymbol{R}}/(\hat{\nu}-p)$ if complex. For the classes defined in Sec. IV we have

$$\widehat{\boldsymbol{M}}_{k|k}(n,c) = \begin{cases} \left(\hat{\nu}_{k|k}^{(n,1)}-2\right)^{-1}\widehat{R}_{k|k}^{(n,1)}\,\boldsymbol{I}, & c = 1, \\ \left(\hat{\nu}_{k|k}^{(n,2)}-m-1\right)^{-1}\widehat{\boldsymbol{R}}_{k|k}^{(n,2)}, & c = 2, \\ \left(\hat{\nu}_{k|k}^{(n,3)}-m\right)^{-1}\widehat{\boldsymbol{R}}_{k|k}^{(n,3)}, & c = 3. \end{cases}$$

### A. Single class

In the first experiment we assume a homogeneous Gaussian clutter environment, with ICM

$$\boldsymbol{M}_k = \boldsymbol{M}_c + \sigma_a^2\boldsymbol{I}. \quad (38)$$

The ICM $\boldsymbol{M}_c$ is exponentially shaped, with entries given by

$$\boldsymbol{M}_c(k,l) = \sigma_c^2\rho^{|k-l|^p}e^{j2\pi(k-l)f_c}, \quad (39)$$

where $j = \sqrt{-1}$ is the imaginary unit, $\rho$ is the one-lag correlation coefficient of the clutter [43], and $\sigma_c$ and $f_c$ denote, respectively, the clutter power level and normalized Doppler frequency. In this first experiment, we simulate a homogeneous land environment, assuming $\rho = \rho^{(l)} = 0.999$, $p = p^{(l)} = 1$, $f_c = f_c^{(l)} = 0$ and clutter to noise ratio $\mathrm{CNR}^{(l)} = \sigma_c^2/\sigma_a^2 = 30\,\mathrm{dB}$. In Fig. 2 the Frobenius norm of the MC-IWM filter estimation error (suitably normalized w.r.t. its trace) of $\boldsymbol{M}_k$ is reported, for several values of $\tau$. As expected, the estimation error decays slower as $\tau$ increases. The error does not converge to zero: there exists a floor value to which the estimation error converges for large $k$. The larger is $\tau$, the smaller the floor level is, and later the convergence takes place. To the limit, when $\tau \to \infty$, the estimation error of the proposed MC-IWM filter converges to zero as $k \to \infty$ and to the performance of a ML filter that operates on all the data observed up to time step $k$, depicted in the same figure with a solid black line. Finally, the performance is compared with that of a knowledge-based recursive least square (KB-RLS) filter [44], also in its version that takes advantage of the persymmetric structure of the covariance matrix (KB-RLS-P). For fairness of comparison, it should be noted that the increased performance of the MC-IWM filter comes at a cost of a higher computational complexity.

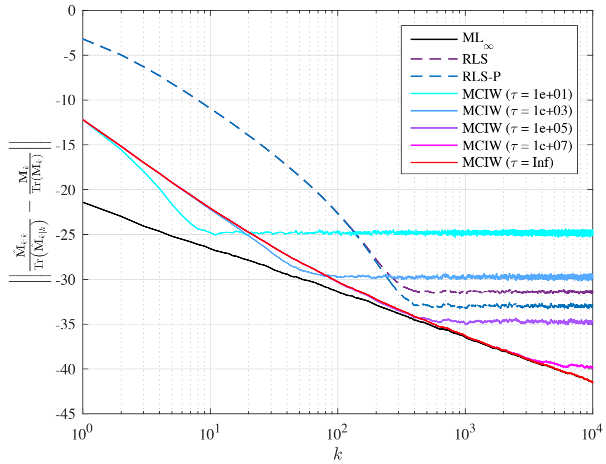



Fig. 2. Single-class analysis of the covariance matrix estimation error of the proposed MC-IWM filter compared with that of other approaches available in literature. The covariance matrix estimation error is suitably normalized w.r.t. its trace in order to be comparable with that of the RLS and RLS-P approaches. Values on the $y$ axis are in decibel units. Data is generated as described in Sec. V-A. Dashed lines to the RLS and RLS-P algorithms [44], with the forgetting factor $\lambda$ set to 0.99. Solid lines refer to the estimation error of the MC-IWM filter, for several values of the adaptation rate $\tau$. Finally, the solid black line refer to a ML estimator that operates on all the observed data. Curves are averaged over 100 runs and $m = 8$.

### B. Sea-land-sea clutter transition

A second experiment is devoted to the analysis of performance when the clutter environment is not homogeneous in space, being instead composed of three different regions; the first and last regions contain homogeneous sea clutter, while the second one contains homogeneous land clutter. Also, smooth transitions at the interfaces between regions with different clutter properties are simulated by gradually increasing the ratio of measurements from the regions before and after the transition. The land clutter ICM is simulated as in the previous experiment; the sea clutter ICM is also exponentially shaped and generated according to (38) and (39), but with $p = p^{(s)} = 2$, a lower one-lag correlation coefficient $\rho = \rho^{(s)} = 0.8$ and a non-null normalized Doppler frequency $f_c = f_c^{(s)} = 0.2$. The performance of the MC-IWM filter, configured as described in Sec. IV, is depicted in Fig. 3a and 3c, in terms of classification capability and ICM estimation error, respectively. Having both the sea and land clutter a persymmetric covariance matrix, the proposed filter correctly classifies the data as belonging to class 2, as it is reported in Fig. 3a. This is a consistent trend over $k$, with the exception of the transition regions, where the posterior probability for class 2 drops, for a couple of samples, below $0.5$. Conversely, the BIC seems unaffected by the transition, and correctly assigns a posterior probability of 1 to class 2. Fig. 3c, showing the norm of the ICM estimation error, suitably normalized w.r.t. its trace, offers a comparison of the proposed approach with other techniques available in literature. First, it must be noted the accuracy improvement over both the ML estimator and its clairvoyant version, which has knowledge of the true disturbance class and uses it to achieve a higher estimation accuracy. From the same figure, the estimation error can be compared also with that of the KB-RLS and KB-RLS-P [44], which initially perform better than the MC-IWM filter, but then prove to be less reactive to the change of environment than the proposed solution. Moreover, the KB-RLS and KB-RLS-P filters require prior knowledge on the position of the clutter edges, which is not required with the proposed approach. A priori information can also be employed in the proposed approach, for instance by allowing the greatest possible spread around the mean for the predictive distribution of the ICM. This is also reported in Fig. 3c by the curve labeled KB-MC-IWM (knowledge-based MC-IWM), which slightly improves the accuracy over clutter edges. Finally, it should be noted that the decay of the classification performance around the transitions does not affect the accuracy of the ICM estimation.

### C. Three classes

In the third experiment, a heterogeneous scenario composed of three different clutter regions is considered. In the first region, only white disturbance is present; the second region contains white disturbance, land and sea clutter; finally, the third region contains all before mentioned, plus a fixed interference term. Formally, we have

$$\boldsymbol{M}_k = \begin{cases} \sigma_a^2 \boldsymbol{I} & \text{data class } 1 \\ \sigma_a^2 \boldsymbol{I} + \boldsymbol{M}_c^{(l)} + \boldsymbol{M}_c^{(s)} & \text{data class } 2 \\ \sigma_a^2 \boldsymbol{I} + \boldsymbol{M}_c^{(l)} + \boldsymbol{M}_c^{(s)} + \boldsymbol{P} & \text{data class } 3, \end{cases} \tag{40}$$

where $\boldsymbol{P}$ is a fixed interference term with the following diagonal structure

$$\boldsymbol{P}(i,j) = \begin{cases} p_{ii} > 0 & i = j = 1, \ldots, n_{\text{int}} \\ 0 & i \neq j, i = n_{\text{int}} + 1, \ldots, m \end{cases} \tag{41}$$

with $p_{\max} \geq \boldsymbol{P}(i,i) \geq \boldsymbol{P}(i+1,i+1) \geq p_{\min}$; this further term may account, from a physical point of view, for the presence of additional interference due, e.g., to the appearance of a frequency modulated radar, or also the activation of a frequency-hopping jammer of shorter duration than the PRI and in the same operating band as the radar. The performance of the proposed MC-IWM filter is reported in Figures 3b and 3d, in terms of class detection probability and ICM estimation error, respectively. As it is apparent from Fig. 3b, the proposed filter is able to correctly detect all three classes; additionally, for class 3, it also outperforms the classification performance of the BIC. Running in a single snapshot configuration, the BIC is in fact able to correctly identify only classes 1 and 2, while class 3 is mistaken for class 2. A complementary perspective on the filter performance is provided in Fig. 3d, where the estimation accuracy of the proposed approach is compared with that of a generic ML estimator and a clairvoyant ML estimator, which has knowledge of the true ICM structure; the ML approaches are both outperformed by the proposed MC-IWM filter, with a noticeable improvement in accuracy.

### D. Classification performance vs BIC

Finally, one last experiment is devoted to a more extensive comparison of the classification performance of the proposed filter against that of the BIC. The simulation setup is similar to the previous experiment: the data is generated as in (40), but now the clutter regions are considered separately, i.e., once at a time, and the data class does not change over time in each run. In order to let vanish the effects of a possibly unmatched prior, the proposed filter has been let run sufficiently long (specifically, for 5 samples) before evaluating the probability of correct classification. The results of the analysis are reported in Table I, for $m = 8$ and several values of $N$. The analysis shows that the proposed MC-IWM filter is able to correctly identify all three classes; moreover, its performance is consistently superior w.r.t. the BIC, especially for low values of $\rho$ in the second class, and also in the third class, where the proposed approach can evidently achieve the same detection capability of the BIC with approximately half data (in each snapshot).

## VI. Conclusion

In this paper, we proposed a novel and general filtering strategy, able to process sequential observations for tracking random matrices, that are defined on multiple nested classes. The proposed filter, referred to as multi-class inverse Wishart mixture (MC-IWM), relies on a hybrid state composed by a discrete random variable representing the class and a positive definite random matrix. We focused on the filtering of the interference covariance matrix (ICM) from the secondary radar (sonar) data in homogeneous clutter environments. The performance assessment of the proposed method has been evaluated in terms of both classification, i.e., environment identification, and ICM estimation accuracy. The results have shown that the proposed approach may provide better performance against both single-scan techniques, such as the Bayesian Information Criterion (BIC), in terms of classification, as well as against sequential techniques, such as maximum likelihood estimation and knowledge-based recursive least square filtering, in terms of ICM estimation accuracy. Possible future research tracks include the analysis and the performance assessment of the MC-IWM filter with real data sets, collected by either radar or sonar systems, as well as the extension of the proposed approach to the case of compound-Gaussian clutter [45]. Finally, the proposed filter can also be applied in the context of extended target tracking to model structured shape extensions, including targets with length equal to the width, negligible length (or width), or composed by multiple ellipses.

## Appendix I
## Class Parameters Posterior Update

In this section, it is shown that (17) holds true whenever

$$\boldsymbol{M}_k = f_{C_k}(\boldsymbol{R}_k) = \boldsymbol{A}_{C_k} \mathbf{diag}_{K_{C_k}}(\boldsymbol{R}_k)\, \boldsymbol{A}_{C_k}^{\dagger}, \qquad (42)$$

being $\boldsymbol{A}_{C_k} \in \mathbb{C}^{m,m}$ a class-dependent matrix such that $\left(\boldsymbol{A}_{C_k}\boldsymbol{A}_{C_k}^{\dagger}\right) \succ \mathbf{0}$, i.e., it is full-rank. This model encompasses many classes of practical interest, as illustrated in Sec. IV.

**Input** : $\boldsymbol{Z}_k = [\boldsymbol{z}_{1,k}, \cdots, \boldsymbol{z}_{N,k}]$, $k = 1, \ldots, K$
$\left(\boldsymbol{R}_0^{(n,C_1)}, \nu_0^{(n,C_1)}\right)$, $n = 1, \ldots, N_W$, $C_1 \in \mathcal{C}$
**Output** : $\left(\widehat{\boldsymbol{R}}_{i|k}^{(n,C_i)}, \hat{\nu}_{i|k}^{(n,C_i)}\right)$, $w_{i|k}^{(n,C_i)}$, $p_{i|k}(C_i)$, for $C_i \in \mathcal{C}$,
$i = k, k+1$, $k = 1, \ldots, K$, $\mathcal{C} = \{1, \ldots, N_C\}$

**Initialization**
$k \longleftarrow 1, \quad p_{k|k-1}(c) \longleftarrow N_C^{-1}$
**for** $C_1 \in \mathcal{C}$ **do**
  **for** $n \in \{1, \ldots, N_W\}$ **do**
    $w_{k|k-1}^{(n,C_1)} \longleftarrow N_W^{-1}$
    $\left(\widehat{\boldsymbol{R}}_{k|k-1}^{(n,C_1)}, \hat{\nu}_{k|k-1}^{(n,C_1)}\right) \longleftarrow \left(\boldsymbol{R}_0^{(n,C_1)}, \nu_0^{(n,C_1)}\right)$
  **end**
**end**
**for** $k \in \{1, \ldots, K\}$ **do**
  **Update**
  Observe new data $\boldsymbol{Z}_k$
  **for** $C_k \in \{1, \ldots, N_C\}$ **do**
    **for** $n \in \{1, \ldots, N_W\}$ **do**
      $\alpha_k^{(n,C_k)} \longleftarrow \int \mathcal{P}\left(\boldsymbol{Z}_k | C_k, \boldsymbol{R}_k\right) \times \mathcal{IW}\left(\boldsymbol{R}; \widehat{\boldsymbol{R}}_{k|k-1}^{(n,C_k)}, \hat{\nu}_{k|k-1}^{(n,C_k)}\right) \mathrm{d}\boldsymbol{R}$
      $w_{k|k}^{(n,C_k)} \longleftarrow \dfrac{\alpha_k^{(n,C_k)} w_{k|k-1}^{(n,C_k)}}{\sum_{n=1}^{N_W} \alpha_k^{(n,C_k)} w_{k|k-1}^{(n,C_k)}}$
      $\left(\widehat{\boldsymbol{R}}_{k|k}^{(n,C_k)}, \hat{\nu}_{k|k}^{(n,C_k)}\right) \longleftarrow f_U\left(\boldsymbol{Z}_k, \widehat{\boldsymbol{R}}_{k|k-1}^{(n,C_k)}, \hat{\nu}_{k|k-1}^{(n,C_k)}, C_k\right)$
    **end**
    $p_{k|k}(C_k) \longleftarrow \dfrac{\left(\sum_{n'=1}^{N_W} w_{k|k-1}^{(n',C_k)} \alpha_k^{(n',C_k)}\right) p_{k|k-1}(C_k)}{\sum_{C_k'=1}^{N_C} \sum_{n'=1}^{N_W} w_{k|k-1}^{(n',C_k')} \alpha_k^{(n',C_k')} p_{k|k-1}(C_k')}$
  **end**
  **Prediction**
  **for** $C_{k+1} \in \{1, \ldots, N_C\}$ **do**
    $n' \longleftarrow 1$
    $p_{k+1|k}(C_{k+1}) \longleftarrow \sum_{C_k'} \pi_{C_{k+1} C_k'}\, p_{k|k}(C_k')$
    **for** $C_k \in \{1, \ldots, N_C\}$ **do**
      **for** $n \in \{1, \ldots, N_W\}$ **do**
        $w_{k+1|k}^{(n',C_{k+1})} = w_{k+1|k}^{(n,C_{k+1},C_k)} \longleftarrow w_{k|k}^{(n,C_k)} \dfrac{\pi_{C_{k+1}C_k} p_{k|k}(C_k)}{p_{k+1|k}(C_{k+1})}$
        $\left(\widehat{\boldsymbol{R}}_{k+1|k}^{(n',C_{k+1})}, \hat{\nu}_{k|k}^{(n',C_{k+1})}\right) = \left(\widehat{\boldsymbol{R}}_{k+1|k}^{(n,C_{k+1},C_k)}, \hat{\nu}_{k|k}^{(n,C_{k+1},C_k)}\right) \longleftarrow f_P\left(\widehat{\boldsymbol{R}}_{k|k}^{(n,C_k)}, \hat{\nu}_{k|k}^{(n,C_k)}, C_{k+1}, C_k\right)$
        $n' \longleftarrow n' + 1$
      **end**
    **end**
  **end**
  **Pruning**
  **for** $C_{k+1} \in \{1, \ldots, N_C\}$ **do**
    Sort the $N_W \times N_C$ mixture components according to the weights $\{w_{k+1|k}^{(n',C_{k+1})}\}_{n'=1}^{N_W\, N_C}$ retaining the first $N_W$ elements
    **for** $n \in \{1, \ldots, N_W\}$ **do**
      $w_{k+1|k}^{(n,C_{k+1})} \longleftarrow \dfrac{w_{k+1|k}^{(n,C_{k+1})}}{\sum_{n'=1}^{N_W} w_{k+1|k}^{(n',C_{k+1})}}$
    **end**
  **end**
**end**

**Algorithm 1:** Multi-class inverse Wishart mixture filter.

TABLE I
PROBABILITY OF CORRECT CLASSIFICATION

| | $\mathcal{P}(C_k=1)$ | | $\mathcal{P}(C_k=2)$ | | | | | | | | | | | | $\mathcal{P}(C_k=3)$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\rho=0.2$ | | $\rho=0.3$ | | $\rho=0.4$ | | $\rho=0.6$ | | $\rho=0.8$ | | $\rho=0.99$ | | | |
| $N$ | MC-IWM | BIC | MC-IWM | BIC | MC-IWM | BIC | MC-IWM | BIC | MC-IWM | BIC | MC-IWM | BIC | MC-IWM | BIC | MC-IWM | BIC |
| 10 | 1.000 | 1.000 | 0.010 | 0.000 | 0.340 | 0.000 | 0.915 | 0.000 | 1.000 | 0.000 | 0.995 | 0.277 | 1.000 | 1.000 | 1.000 | 0.166 |
| 12 | 0.990 | 1.000 | 0.010 | 0.000 | 0.420 | 0.000 | 0.965 | 0.000 | 1.000 | 0.001 | 0.995 | 0.476 | 1.000 | 1.000 | 1.000 | 0.412 |
| 16 | 1.000 | 1.000 | 0.025 | 0.000 | 0.655 | 0.000 | 0.995 | 0.000 | 1.000 | 0.003 | 1.000 | 0.872 | 1.000 | 1.000 | 1.000 | 0.926 |
| 20 | 1.000 | 1.000 | 0.065 | 0.000 | 0.905 | 0.000 | 0.990 | 0.000 | 1.000 | 0.013 | 1.000 | 0.987 | 1.000 | 1.000 | 1.000 | 0.998 |
| 24 | 1.000 | 1.000 | 0.135 | 0.000 | 0.970 | 0.000 | 1.000 | 0.000 | 1.000 | 0.089 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 28 | 1.000 | 1.000 | 0.235 | 0.000 | 0.980 | 0.000 | 1.000 | 0.000 | 1.000 | 0.240 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 32 | 1.000 | 1.000 | 0.385 | 0.000 | 0.990 | 0.000 | 1.000 | 0.000 | 1.000 | 0.499 | 1.000 | 1.000 | 0.995 | 1.000 | 1.000 | 1.000 |
| 64 | 1.000 | 1.000 | 0.995 | 0.000 | 1.000 | 0.000 | 1.000 | 0.026 | 1.000 | 1.000 | 1.000 | 1.000 | 0.985 | 1.000 | 1.000 | 1.000 |
| 128 | 1.000 | 1.000 | 1.000 | 0.000 | 1.000 | 0.020 | 1.000 | 0.988 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Probability of correct environment classification of the proposed MC-IWM filter versus the BIC approach. The MC-IWM performance is evaluated after 5 samples, while the BIC works on single snapshots. The thermal noise power level is set to $\sigma_a^2 = 0\,\mathrm{dB}$; the statistical properties of the sea clutter are $\mathrm{CNR}^{(s)} = 20\,\mathrm{dB}$, $\rho^{(s)} = 0.8$, $p^{(s)} = 2$ and $f_c^{(s)} = 0.2$; the statistical properties of the land clutter are $\mathrm{CNR}^{(l)} = 30\,\mathrm{dB}$, $\rho^{(l)} = 0.999$, $p^{(l)} = 1$, and $f_c^{(l)} = 0$. The fixed interference matrix $\boldsymbol{P}$ in class 3 is a diagonal matrix with diagonal components set to $[40\,\mathrm{dB}, 36\,\mathrm{dB}, 32\,\mathrm{dB}, 28\,\mathrm{dB}, 24\,\mathrm{dB}, 20\,\mathrm{dB}, 0, 0]$. The MC-IWM filter adaptation rate was set to $\tau = 10$ and the components of the transition matrix to $\pi_{ij} = 0.98$ if $i = j$ and to $\pi_{ij} = 0.01$ if, otherwise, $i \neq j$. Results are averaged over 200 runs and $m = 8$.

In order to proceed, let us focus on a specific class and, to simplify the notation, denote $c = C_k$, $\boldsymbol{R} = \boldsymbol{R}_k$, $\boldsymbol{Z} = \boldsymbol{Z}_k$, $K = K_{C_k} = m/m_c$ (with $m_c$ the size of $\boldsymbol{R}_k$), and $\boldsymbol{A}_{C_k} = \boldsymbol{A}$. We assume that $\boldsymbol{R} \sim \mathcal{IW}\left(\boldsymbol{R};\, \widehat{\boldsymbol{R}}, \widehat{\nu}\right)$. Then, if $\boldsymbol{R} \in \mathbb{S}_{++}^{m_c}$, the joint distribution between $\boldsymbol{R}$ and $\boldsymbol{Z}$ is proportional to

$$
\begin{aligned}
\mathcal{P}(\boldsymbol{R}, \boldsymbol{Z}) &\overset{(a)}{\propto} |\boldsymbol{R}|^{-\frac{m_c+\widehat{\nu}+1}{2}}\, e^{-\frac{1}{2}\mathrm{Tr}\left(\boldsymbol{R}^{-1}\widehat{\boldsymbol{R}}\right)} \\
&\quad \times \left|\boldsymbol{A}\, \mathbf{diag}_K\left(\boldsymbol{R}\right) \boldsymbol{A}^{\dagger}\right|^{-N} \\
&\quad \times e^{-\mathrm{Tr}\left(\boldsymbol{Z}\boldsymbol{Z}^{\dagger}\left(\boldsymbol{A}\, \mathbf{diag}_K(\boldsymbol{R})\boldsymbol{A}^{\dagger}\right)^{-1}\right)} \\
&\overset{(b)}{\propto} |\boldsymbol{R}|^{-\frac{m_c+\widehat{\nu}+1+2KN}{2}}\, e^{-\frac{1}{2}\mathrm{Tr}\left(\boldsymbol{R}^{-1}\widehat{\boldsymbol{R}}_1\right)}
\end{aligned}
$$

where in $(a)$ we have used the likelihood expression of $\boldsymbol{Z}$ and the IW distribution of $\boldsymbol{R}$ [46], and $(b)$ follows by algebraic manipulations, not reported for the sake of brevity. In the above expression we have defined:

$$
\widehat{\boldsymbol{R}}_1 := \widehat{\boldsymbol{R}} + 2\sum_{i=0}^{K-1} \mathrm{D}\left[\Re\left(\boldsymbol{A}^{-1}\boldsymbol{Z}\boldsymbol{Z}^{\dagger}\boldsymbol{A}^{\dagger^{-1}}\right)\right]_i^{m_c}.
$$

Thus, the posterior of $\boldsymbol{R}$ is a real IW with scale matrix $\widehat{\boldsymbol{R}}_1$ and $\widehat{\nu} + 2KN$ degrees of freedom. If instead $\boldsymbol{R} \in \mathbb{H}_{++}^{m_c}$, similarly to the previous case the joint distribution between $\boldsymbol{R}$ and $\boldsymbol{Z}$ is proportional to

$$
\begin{aligned}
\mathcal{P}(\boldsymbol{R}, \boldsymbol{Z}) &\propto |\boldsymbol{R}|^{-(m_c+\widehat{\nu})}\, e^{-\mathrm{Tr}\left(\boldsymbol{R}^{-1}\widehat{\boldsymbol{R}}\right)} \left|\boldsymbol{A}\mathbf{diag}_K\left(\boldsymbol{R}\right)\boldsymbol{A}^{\dagger}\right|^{-N} \\
&\quad \times e^{-\mathrm{Tr}\left(\boldsymbol{Z}\boldsymbol{Z}^{\dagger}\left(\boldsymbol{A}\mathbf{diag}_K(\boldsymbol{R})\boldsymbol{A}^{\dagger}\right)^{-1}\right)} \\
&\propto |\boldsymbol{R}|^{-(m_c+\widehat{\nu}+KN)}\, e^{-\mathrm{Tr}\left(\boldsymbol{R}^{-1}\widehat{\boldsymbol{R}}_1\right)},
\end{aligned}
$$

where again we have exploited the likelihood and IW expressions [15], [19], [47] with

$$
\widehat{\boldsymbol{R}}_1 = \widehat{\boldsymbol{R}} + \sum_{i=0}^{K-1} \mathrm{D}\left[\boldsymbol{A}^{-1}\boldsymbol{Z}\boldsymbol{Z}^{\dagger}\boldsymbol{A}^{\dagger^{-1}}\right]_i^{m_c}.
$$

Thus, the posterior of $\boldsymbol{R}$ is a complex IW with scale matrix $\widehat{\boldsymbol{R}}_1$ and degrees of freedom $\widehat{\nu} + KN$.

### *A. General update model*

If the classes cannot be described with the model in (42), and a more general model is required, then, in order to preserve the IW structure, the general update step (16) can be formulated in terms of both the moment matching approximation (see details in Appendix II) and Monte Carlo methods (MCM) [48]. Specifically, the idea is to compute the first and second moment of the random matrix $\boldsymbol{R}_k | \boldsymbol{Z}_{1:k}, n, C_k$ via MCM and then enforcing the update distribution to be IW applying the moment matching technique detailed in Appendix II. To this end, let $h(\boldsymbol{R}_k)$ be a scalar function of the matrix argument $\boldsymbol{R}_k$, for instance an entry of the matrix or its power, and denote by $\boldsymbol{R}_k^{(i)}$, $i = 1, \ldots, N_{\mathrm{MCM}}$, i.i.d. matrices drawn from $\mathcal{IW}(\boldsymbol{R}_k;\, \widehat{\boldsymbol{R}}_{k|k-1}^{(n,C_k)}, \widehat{\nu}_{k|k-1}^{(n,C_k)})$ made available from the prediction step given the data up to time $k-1$. Then, applying the MCM to the distribution given in the update equation (16), it follows that [48]

$$
\mathbb{E}\left[h\left(\boldsymbol{R}_k\right) | \boldsymbol{Z}_{1:k}, n, C_k\right] \approx \frac{\displaystyle\sum_{i=1}^{N_{\mathrm{MCM}}} h(\boldsymbol{R}_k^{(i)})\mathcal{P}(\boldsymbol{Z}_k | C_k, \boldsymbol{R}_k^{(i)})}{\displaystyle\sum_{i=1}^{N_{\mathrm{MCM}}} \mathcal{P}(\boldsymbol{Z}_k | C_k, \boldsymbol{R}_k^{(i)})},
$$

where, under weak assumptions, the previous equation converges almost surely when $N_{\mathrm{MCM}}$ diverges [48].

## APPENDIX II
## MOMENT MATCHING

Let $\widetilde{\boldsymbol{R}}$ and $\widetilde{\sigma}^2$ be the mean matrix and the total variance of a random matrix $\boldsymbol{R}$, i.e.,

$$
\widetilde{\sigma}^2 = \mathrm{Tr}\left(\mathbb{E}\left[\left(\boldsymbol{R} - \widetilde{\boldsymbol{R}}\right) \odot \left(\boldsymbol{R}^* - \widetilde{\boldsymbol{R}}^*\right)\right]\right). \tag{43}
$$

The aim is to approximate the distribution of $\boldsymbol{R}$ with an IW, whose parameters, $\boldsymbol{R}_{IW}$ and $\nu_{IW}$, fulfill the moment matching conditions. Conditioned on the class, the specific IW mixand, and the data up to time $k$, the random matrix $\boldsymbol{R}$ can refer either to $\boldsymbol{R}_k$ (update) or to $\boldsymbol{R}_{k+1}$ (prediction). The conditional moments of $\boldsymbol{R}_k$ can be computed[2] as in Appendix I-A while those of $\boldsymbol{R}_{k+1}$ can be evaluated as in Appendix III. The following subsection details the equation systems that ensure the moment matching.

### A. Relation between IW parameters and moments of $\boldsymbol{R}$

The relation between the IW parameters and the moments of $\boldsymbol{R}$ are based on the real and complex IW moments expressions [15], [19], [46], [47]:

$$\begin{cases} \widetilde{\boldsymbol{R}} = \frac{\boldsymbol{R}_{IW}}{\nu_{IW}-m_c-1}, & \\ \widetilde{\sigma}^2 = \sum_{i=1}^{m_c} \frac{2\boldsymbol{R}^2_{IW}(i,i)}{(\nu_{IW}-m_c-1)^2(\nu_{IW}-m_c-3)}, & \boldsymbol{R} \in \mathbb{S}^{m_c}_{++}, \\ \widetilde{\boldsymbol{R}} = \frac{\boldsymbol{R}_{IW}}{\nu_{IW}-m_c}, & \\ \widetilde{\sigma}^2 = \sum_{i=1}^{m_c} \frac{\boldsymbol{R}^2_{IW}(i,i)}{(\nu_{IW}-m_c)^2(\nu_{IW}-m_c-1)}, & \boldsymbol{R} \in \mathbb{H}^{m_c}_{++}. \end{cases}$$

Using the previous equations we can easily obtain the expression of the IW parameters for both real and complex matrices:

$$\begin{cases} \nu_{IW} = m_c + 3 + \frac{2}{\widetilde{\sigma}^2}\sum_{i=1}^{m_c}\widetilde{\boldsymbol{R}}^2(i,i), & \\ \boldsymbol{R}_{IW} = \widetilde{\boldsymbol{R}}\left(2 + \frac{2}{\widetilde{\sigma}^2}\sum_{i=1}^{m_c}\widetilde{\boldsymbol{R}}^2(i,i)\right), & \boldsymbol{R} \in \mathbb{S}^{m_c}_{++}, \\ \nu_{IW} = m_c + 1 + \frac{1}{\widetilde{\sigma}^2}\sum_{i=1}^{m_c}\widetilde{\boldsymbol{R}}^2(i,i), & \\ \boldsymbol{R}_{IW} = \widetilde{\boldsymbol{R}}\left(1 + \frac{1}{\widetilde{\sigma}^2}\sum_{i=1}^{m_c}\widetilde{\boldsymbol{R}}^2(i,i)\right), & \boldsymbol{R} \in \mathbb{H}^{m_c}_{++}. \end{cases}$$

## Appendix III
## Prediction Statistics Evaluation

In this subsection, guidelines for the evaluation of the prediction statistics are provided, as required for the moment-matching-based IW approximation of the integral in the last equality of (27). To this end, two covariance classes are considered, corresponding to the models

$$f_{C_k}(\boldsymbol{R}) = \boldsymbol{A}_j \mathbf{diag}_{K_j}(\boldsymbol{R})\boldsymbol{A}_j^\dagger, \quad C_k = j, \quad j \in \{1, 2\},$$

where if $C_k = 1$ then $\boldsymbol{R} \in \mathbb{S}^{m_1}_{++}$ and $m_1 K_1 = m$; otherwise if $C_k = 2$ then $\boldsymbol{R} \in \mathbb{H}^{m_2}_{++}$ and $m_2 K_2 = m$. In the following, both the mode $n$ and the observations $\boldsymbol{Z}_{1:k}$ are assumed fixed; also, all expectations are conditioned to $(C_{k+1}, C_k)$.

### A. Switch between $C_k = 1$ and $C_{k+1} = 1$

In this case, assuming $\nu > 0$, we have

$$\boldsymbol{R}_{k+1}|C_{k+1}, C_k, \boldsymbol{R}_k \sim \mathcal{W}\left(\boldsymbol{R}_{k+1}, \frac{\boldsymbol{R}_k}{\nu}, \nu\right),$$
$$\boldsymbol{R}_k|C_k \sim \mathcal{IW}\left(\boldsymbol{R}_k, \widehat{\boldsymbol{R}}_{IW}, \widehat{\nu}_{IW}\right),$$

where $\boldsymbol{R}_k$ is conditionally independent of $C_{k+1}$ given $C_k$. The conditional moments of $\boldsymbol{R}_{k+1}$ are given as follows

$$\widetilde{\boldsymbol{R}} = \mathbb{E}[\boldsymbol{R}_{k+1}] = \mathbb{E}[\mathbb{E}[\boldsymbol{R}_{k+1}|\boldsymbol{R}_k]] = \frac{\widehat{\boldsymbol{R}}_{IW}}{(\widehat{\nu}_{IW} - m_1 - 1)}, \quad (44)$$

$$\widetilde{\sigma}^2 = \sum_{i=1}^{m_1}\left(\mathbb{E}\left[\boldsymbol{R}^2_{k+1}(i,i)\right] - \mathbb{E}^2\left[\boldsymbol{R}_{k+1}(i,i)\right]\right). \quad (45)$$

[2]In the update step, the moment matching approximation is most useful when model (42) does not hold; otherwise, the exact forms of Appendix I can be used to compute the moments.

Hence, from the expressions of the moments reported in [46] it follows that

$$\mathbb{E}\left[\boldsymbol{R}^2_{k+1}(i,i)\right] = \left(1 + \frac{2}{\nu}\right)\left(\frac{2}{\widehat{\nu}_{IW} - m_1 - 3} + 1\right) \times \frac{\widehat{\boldsymbol{R}}^2_{IW}(i,i)}{(\widehat{\nu}_{IW} - m_1 - 1)^2}. \quad (46)$$

Thus, exploiting (44), (45) and (46), we obtain

$$\widetilde{\sigma}^2 = \left[\left(1 + \frac{2}{\nu}\right)\left(\frac{2}{\widehat{\nu}_{IW} - m_1 - 3} + 1\right) - 1\right] \times \left(\sum_{i=1}^{m_1}\widehat{\boldsymbol{R}}^2_{IW}(i,i)\right)\frac{1}{(\widehat{\nu}_{IW} - m_1 - 1)^2}. \quad (47)$$

### B. Switch between $C_k = 1$ and $C_{k+1} = 2$

We have:

$$\boldsymbol{R}_{k+1}|C_{k+1}, C_k, \boldsymbol{R}_k \sim \mathcal{W}\left(\boldsymbol{R}_{k+1}, \sum_{i=0}^{K_2-1}\frac{\mathrm{D}\left[g(\boldsymbol{R}_k)\right]_i^{m_2}}{\nu K_2}, \nu\right),$$
$$\boldsymbol{R}_k|C_k \sim \mathcal{IW}\left(\boldsymbol{R}_k, \widehat{\boldsymbol{R}}_{IW}, \widehat{\nu}_{IW}\right),$$

where $\boldsymbol{R}_k$ is conditionally independent of $C_{k+1}$ given $C_k$ and

$$g(\boldsymbol{R}_k) = \boldsymbol{A}_{21}\mathbf{diag}_{K_1}(\boldsymbol{R}_k)\boldsymbol{A}_{21}^\dagger, \quad \boldsymbol{A}_{21} := \boldsymbol{A}_2^{-1}\boldsymbol{A}_1,$$

Note that the matrix $\frac{1}{K_2}\sum_{i=0}^{K_2-1}\mathrm{D}\left[g(\boldsymbol{R}_k)\right]_i^{m_2}$, involved in the Wishart distribution of $\boldsymbol{R}_{k+1}$ considered before, is defined according to a sort of ML-based projection. Precisely,

$$\boldsymbol{A}_2\mathbf{diag}_{K_2}\left(\frac{1}{K_2}\sum_{i=0}^{K_2-1}\mathrm{D}\left[g(\boldsymbol{R}_k)\right]_i^{m_2}\right)\boldsymbol{A}_2^\dagger$$

represents the covariance matrix within the second class that maximizes the likelihood function given the actual covariance observed at the previous scan, i.e., $\boldsymbol{A}_1\mathbf{diag}_{K_1}(\boldsymbol{R}_k)\boldsymbol{A}_1^\dagger$. Let us now evaluate the statistics of $\boldsymbol{R}_{k+1}|C_{k+1}, C_k$. Based on [15], [19], [46], [47] and exploiting the linearity of the operators $D[\cdot]_i^{m_2}$ and $g(\cdot)$ we have

$$\widetilde{\boldsymbol{R}} = \mathbb{E}[\boldsymbol{R}_{k+1}] = \frac{\sum_{i=0}^{K_2-1}\mathrm{D}\left[g\left(\widehat{\boldsymbol{R}}_{IW}\right)\right]_i^{m_2}}{K_2(\widehat{\nu}_{IW} - m_1 - 1)}$$

$$\widetilde{\sigma}^2 = \sum_{i=1}^{m_2}\left(\mathbb{E}\left[\boldsymbol{R}^2_{k+1}(i,i)\right] - \mathbb{E}^2\left[\boldsymbol{R}_{k+1}(i,i)\right]\right),$$

where

$$\mathbb{E}\left[\boldsymbol{R}^2_{k+1}(i,i)\right] = \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{R}^2_{k+1}(i,i)|\boldsymbol{R}_k\right]\right].$$

According to [47], it follows that

$$\mathbb{E}\left[\boldsymbol{R}^2_{k+1}(i,i)|\boldsymbol{R}_k\right] = \left(\frac{1}{K_2}\sum_{h=0}^{K_2-1}\mathrm{D}[g(\boldsymbol{R}_k)]_h^{m_2}(i,i)\right)^2\left(1 + \frac{1}{\nu}\right).$$

$$\mathbb{E}\left[\boldsymbol{R}_k(i_1,i_2)\boldsymbol{R}_k(h_1,h_2)\right] = \frac{\widehat{\boldsymbol{R}}_{IW}(i_1,i_2)\widehat{\boldsymbol{R}}_{IW}(h_1,h_2)}{(\widehat{\nu}_{IW}-m_1-1)^2} + \frac{2\widehat{\boldsymbol{R}}_{IW}(i_1,i_2)\widehat{\boldsymbol{R}}_{IW}(h_1,h_2)}{(\widehat{\nu}_{IW}-m_1)(\widehat{\nu}_{IW}-m_1-1)^2(\widehat{\nu}_{IW}-m_1-3)} + \frac{\widehat{\boldsymbol{R}}_{IW}(i_1,h_1)\widehat{\boldsymbol{R}}_{IW}(i_2,h_2)+\widehat{\boldsymbol{R}}_{IW}(i_1,h_2)\widehat{\boldsymbol{R}}_{IW}(h_1,i_2)}{(\widehat{\nu}_{IW}-m_1)(\widehat{\nu}_{IW}-m_1-1)(\widehat{\nu}_{IW}-m_1-3)} \tag{48}$$

In order to proceed, let now assume $\boldsymbol{A}_{21}^{\dagger} = [\boldsymbol{a}_1,\ldots,\boldsymbol{a}_m]$ and observe that

$$\begin{aligned}\mathrm{D}\left[g(\boldsymbol{R}_k)\right]_h^{m_2}(l,l) &= \boldsymbol{a}_{l+hm_2}^{\dagger}\mathbf{diag}_{K_1}(\boldsymbol{R}_k)\,\boldsymbol{a}_{l+hm_2}\\ &= \mathrm{Tr}\left(\boldsymbol{R}_k\left(\sum_{e=0}^{K_1-1}\mathrm{D}\left[\boldsymbol{a}_{l+hm_2}\boldsymbol{a}_{l+hm_2}^{\dagger}\right]_e^{m_1}\right)\right).\end{aligned}$$

Hence,

$$\left(\frac{1}{K_2}\sum_{h=0}^{K_2-1}\mathrm{D}\left[g(\boldsymbol{R}_k)\right]_h^{m_2}(i,i)\right) = \mathrm{Tr}\left(\boldsymbol{R}_k\boldsymbol{B}^i\right),$$

where

$$\boldsymbol{B}^i = \frac{1}{K_2}\sum_{h=0}^{K_2-1}\sum_{e=0}^{K_1-1}\mathrm{D}\left[\boldsymbol{a}_{i+hm_2}\boldsymbol{a}_{i+hm_2}^{\dagger}\right]_e^{m_1}.$$

As a consequence,

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{R}_{k+1}^2(i,i)\right] &= \left(1+\frac{1}{\nu}\right)\mathbb{E}\left[\left(\mathrm{Tr}\left(\boldsymbol{R}_k\boldsymbol{B}^i\right)\right)^2\right]\\ &= \left(1+\frac{1}{\nu}\right)\mathrm{vec}\left(\boldsymbol{B}^{i^T}\right)^T\boldsymbol{E}_{\boldsymbol{R}_k}\mathrm{vec}\left(\boldsymbol{B}^{i^T}\right)\end{aligned}$$

where $\boldsymbol{E}_{\boldsymbol{R}_k} = \mathbb{E}\left[\mathrm{vec}(\boldsymbol{R}_k)\mathrm{vec}(\boldsymbol{R}_k)^T\right]$ can be computed via (48). Thus

$$\widetilde{\sigma}^2 = \left(1+\frac{1}{\nu}\right)\mathrm{Tr}\left(\boldsymbol{E}_{\boldsymbol{R}_k}\boldsymbol{B}\right) - \frac{\sum_{i=1}^{m_2}\left(\mathrm{Tr}\left(\widehat{\boldsymbol{R}}_{IW}\boldsymbol{B}^i\right)\right)^2}{(\widehat{\nu}_{IW}-m_1-1)^2}$$

with

$$\boldsymbol{B} = \sum_{i=1}^{m_2}\mathrm{vec}\left(\boldsymbol{B}^{i^T}\right)\mathrm{vec}\left(\boldsymbol{B}^{i^T}\right)^T$$

### C. Switch between $C_k = 2$ and $C_{k+1} = 1$

We have:

$$\begin{aligned}\boldsymbol{R}_{k+1}|C_{k+1},C_k,\boldsymbol{R}_k &\sim \mathcal{W}\left(\boldsymbol{R}_{k+1},\frac{1}{\nu K_1}\sum_{i=0}^{K_1-1}\mathrm{D}\left[g(\boldsymbol{R}_k)\right]_i^{m_1},\nu\right),\\ \boldsymbol{R}_k|C_k &\sim \mathcal{IW}\left(\widehat{\boldsymbol{R}}_{IW},\widehat{\nu}_{IW}\right),\end{aligned}$$

where $\boldsymbol{R}_k$ is conditionally independent of $C_{k+1}$ given $C_k$ and

$$g(\boldsymbol{R}_k) = \Re\left\{\boldsymbol{A}_{12}\mathbf{diag}_{K_2}(\boldsymbol{R}_k)\,\boldsymbol{A}_{12}^{\dagger}\right\},\qquad \boldsymbol{A}_{12} = \boldsymbol{A}_1^{-1}\boldsymbol{A}_2.$$

Also in this case, the scale matrix of the Wishart distribution is related to the ML-based projection discussed in Appendix III-B. Let us evaluate the statistics of $\boldsymbol{R}_{k+1}|C_{k+1},C_k$, leveraging the expressions of the moments of interest given in [15], [19], [46], [47] and the linearity of the operators $D[\cdot]_i^{m_1}$ and $g(\cdot)$:

$$\widetilde{\boldsymbol{R}} = \mathbb{E}\left[\boldsymbol{R}_{k+1}\right] = \frac{\sum_{i=0}^{K_1-1}\mathrm{D}\left[g\left(\widehat{\boldsymbol{R}}_{IW}\right)\right]_i^{m_1}}{K_1(\widehat{\nu}_{IW}-m_2)}.$$

Given that the computation of the total variance is similar to the previous case, we report below only the final expression:

$$\widetilde{\sigma}^2 = \left(1+\frac{2}{\nu}\right)\mathrm{Tr}\left(\tilde{\boldsymbol{E}}_{\boldsymbol{R}_k}\tilde{\boldsymbol{B}}\right) - \frac{\sum_{i=1}^{m_1}\mathrm{Tr}\left(\widehat{\boldsymbol{R}}_{IW}\tilde{\boldsymbol{B}}^i\right)^2}{(\widehat{\nu}_{IW}-m_2)^2},$$

with

$$\begin{aligned}\tilde{\boldsymbol{B}} &= \sum_{i=1}^{m_1}\mathrm{vec}(\tilde{\boldsymbol{B}}^{i^T})\mathrm{vec}(\tilde{\boldsymbol{B}}^{i^T})^T,\\ \tilde{\boldsymbol{B}}^i &= \frac{1}{K_1}\sum_{h=0}^{K_1-1}\sum_{e=0}^{K_2-1}\mathrm{D}\left[\tilde{\boldsymbol{a}}_{i+hm_1}\tilde{\boldsymbol{a}}_{i+hm_1}^{\dagger}\right]_e^{m_2},\\ \boldsymbol{A}_{12}^{\dagger} &= \left[\tilde{\boldsymbol{a}}_1,\ldots,\tilde{\boldsymbol{a}}_m\right],\tilde{\boldsymbol{E}}_{\boldsymbol{R}_k} = \mathbb{E}\left[\mathrm{vec}\left(\boldsymbol{R}_k\right)\mathrm{vec}\left(\boldsymbol{R}_k\right)^T\right].\end{aligned}$$

where $\tilde{\boldsymbol{E}}_{\boldsymbol{R}_k}$ can be evaluated according to (49).

### D. Switch between $C_k = 2$ and $C_{k+1} = 2$

We have:

$$\begin{aligned}\boldsymbol{R}_{k+1}|C_{k+1},C_k,\boldsymbol{R}_k &\sim \mathcal{W}\left(\boldsymbol{R}_{k+1},\frac{\boldsymbol{R}_k}{\nu},\nu\right),\\ \boldsymbol{R}_k|C_k &\sim \mathcal{IW}\left(\boldsymbol{R}_k,\widehat{\boldsymbol{R}}_{IW},\widehat{\nu}_{IW}\right),\end{aligned}$$

where $\boldsymbol{R}_k$ is conditionally independent of $C_{k+1}$ given $C_k$. The moments of $\boldsymbol{R}_{k+1}|C_{k+1},C_k$ can be computed as done for the previous cases. We report in the following the final expressions:

$$\begin{aligned}\widetilde{\boldsymbol{R}} &= \mathbb{E}\left[\boldsymbol{R}_{k+1}\right] = \frac{\widehat{\boldsymbol{R}}_{IW}}{\widehat{\nu}_{IW}-m_2},\\ \widetilde{\sigma}^2 &= \frac{\sum_{i=1}^{m_2}\widehat{\boldsymbol{R}}_{IW}^2(i,i))}{(\widehat{\nu}_{IW}-m_2)^2}\left(\frac{\left(1+\frac{1}{\nu}\right)(\widehat{\nu}_{IW}-m_2)}{\widehat{\nu}_{IW}-m_2-1}-1\right).\end{aligned}$$

### E. General prediction model

For an arbitrary class model, MCM can be possibly exploited to evaluate the moments of interest. Precisely, let $\boldsymbol{Y}^{(i)}$, $i = 1,\ldots,M_c$, be i.i.d. random matrices drawn from $\mathcal{IW}(\boldsymbol{Y};\,\widehat{\boldsymbol{R}}_{k|k}^{(n,C_k)};\,\hat{\nu}_{k|k}(n,C_k))$. Then, generating $M_c$ independent random matrices $\boldsymbol{R}_{k+1}^{(i)}$, $i = 1,\ldots,M_c$, according to $\mathcal{P}(\boldsymbol{R}_{k+1}|C_{k+1},C_k,\boldsymbol{Y}^{(i)})$, it follows that, $\boldsymbol{R}_{k+1}^{(i)}$, $i = 1,\ldots,M_c$ are i.i.d. and distributed as the integral in the last equality of (27). Hence, as done in the Appendix I-A the samples $\boldsymbol{R}_{k+1}^{i}$, $i = 1,\ldots,M_c$, can be exploited to estimate the desired moments to use in the moment matching procedure.

$$\mathbb{E}\left[\boldsymbol{R}_k(i_1,i_2)\boldsymbol{R}_k(h_1,h_2)\right] = \frac{\widehat{\boldsymbol{R}}_{IW}(i_1,i_2)\widehat{\boldsymbol{R}}_{IW}(h_1,h_2) + \frac{1}{\nu_{IW}-m_2}\widehat{\boldsymbol{R}}_{IW}(i_1,h_1)\widehat{\boldsymbol{R}}_{IW}(i_2,h_2)}{(\widehat{\nu}_{IW}-m_2)^2-1} \tag{49}$$

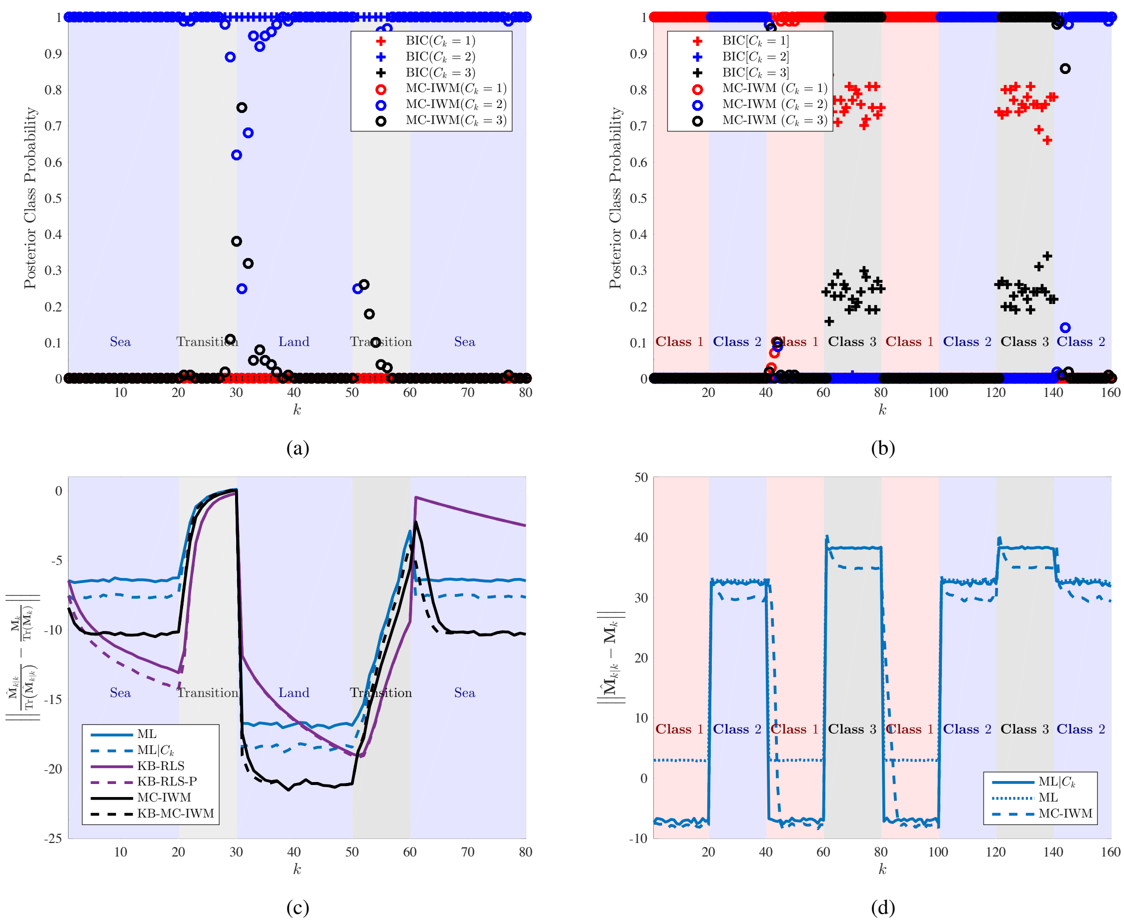



Fig. 3. Performance analysis and comparison with other approaches of the proposed MC-IWM filter in terms of classification capability (a)–(b) and ICM estimation error, in decibel units, (c)–(d). Panels (a) and (c) refer to the simulation study reported in Sec. V-B, while panels (b) and (d) to the experiment described in Sec. V-C. The background color represents the true environment class over time; in panels (a) and (b), for each time instant, a correct classification is represented by the highest marker's color being the same as the background; at each time scan $k$ the class with highest probability is selected and then decisions are averaged over all Monte Carlo runs. The white noise power level is set to $\sigma_a^2 = 0\,\text{dB}$; the statistical properties of the sea clutter are $\text{CNR}^{(s)} = 20\,\text{dB}$, $\rho^{(s)} = 0.8$, $p^{(s)} = 2$, and $f_c^{(s)} = 0.2$; the statistical properties of the land clutter are $\text{CNR}^{(l)} = 30\,\text{dB}$, $\rho^{(l)} = 0.999$, $p^{(l)} = 1$ and $f_c^{(l)} = 0$. The fixed interference matrix $\boldsymbol{P}$ in class 3 is a diagonal matrix with diagonal components set to $[40\,\text{dB}, 36.7\,\text{dB}, 33.3\,\text{dB}, 30\,\text{dB}, 26.7\,\text{dB}, 23.3\,\text{dB}, 20\,\text{dB}, 0]$. The MC-IWM filter adaptation rate $\tau$ is set to 5. The forgetting factor $\lambda$ of the KB-RLS and KB-RLS-P approach is set to 0.99 in the land and sea regions and linearly increases from 0.5 to 0.99 in the transition transition regions. Results are averaged over 100 runs; $m$ is set to 8 and $N$ to 16.

# References


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

[2] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 2, pp. 115–127, Mar. 1986.

[3] I. Reed, J. Mallett, and L. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, no. 6, pp. 853–863, Nov. 1974.

[4] B. Himed and W. L. Melvin, "Analyzing space-time adaptive processors using measured data," in *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers*, Nov. 1997, pp. 930–935.

[5] 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, Aug. 2000.

[6] 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, Jun. 2012.

[7] B. Kang, V. Monga, and M. Rangaswamy, "Rank-constrained maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 501–515, Jan. 2014.

[8] R. Nitzberg, "Application of maximum likelihood estimation of persymmetric covariance matrices to adaptive processing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-16, no. 1, pp. 124–127, Jan. 1980.

[9] H. Li, P. Stoica, and J. Li, "Computationally efficient maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1314–1323, May 1999.

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

[11] J. R. Roman, M. Rangaswamy, D. W. Davis, Q. Zhang, B. Himed, and J. H. Michels, "Parametric adaptive matched filter for airborne radar applications," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 677–692, Apr. 2000.

[12] I. Soloveychik, D. Trushin, and A. Wiesel, "Group symmetric robust covariance estimation," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 244–257, Jan. 2016.

[13] A. De Maio, A. Farina, and G. Foglia, "Adaptive radar detection: A Bayesian approach," in *2007 IEEE Radar Conference*, Apr. 2007, pp. 624–629.

[14] 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, May 2007.

[15] S. Bidon, O. Besson, and J. Tourneret, "A Bayesian approach to adaptive detection in nonhomogeneous environments," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 205–217, Jan. 2008.

[16] A. De 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, pp. 170–183, Jan. 2010.

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

[18] 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, Apr. 2010.

[19] S. Bidon, O. Besson, and J. Tourneret, "Knowledge-aided STAP in heterogeneous clutter using a hierarchical Bayesian algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 3, pp. 1863–1879, Jul. 2011.

[20] V. Carotenuto, A. De Maio, D. Orlando, and P. Stoica, "Model order selection rules for covariance structure classification in radar," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5305–5317, Oct. 2017.

[21] 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, Jul. 2004.

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

[23] K. Granström, M. Baum, and S. Reuter, "Extended object tracking: Introduction, overview, and applications," *Journal of Advances in Information Fusion*, vol. 12, no. 2, 2017.

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

[25] 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, Oct. 2016.

[26] G. Soldi and P. Braca, "Online estimation of unknown parameters in multisensor-multitarget tracking: a belief propagation approach," in *2018 21st International Conference on Information Fusion (FUSION)*, Jul. 2018, pp. 2151–2157.

[27] G. Soldi, F. Meyer, P. Braca, and F. Hlawatsch, "Self-tuning algorithms for multisensor-multitarget tracking using belief propagation," *IEEE Trans. Signal Process.*, (accepted).

[28] W. Li and Y. Jia, "Adaptive filtering for jump Markov systems with unknown noise covariance," *IET Control Theory Applications*, vol. 7, no. 13, pp. 1765–1772, 2013.

[29] ——, "Kullback–Leibler divergence for interacting multiple model estimation with random matrices," *IET Signal Processing*, vol. 10, no. 1, pp. 12–18, 2016.

[30] 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, Jul. 2008.

[31] C. Lundquist, K. Granström, and U. Orguner, "An extended target CPHD filter and a Gamma Gaussian inverse Wishart implementation," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 3, pp. 472–483, June 2013.

[32] 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, Dec. 2015.

[33] K. Granström and U. Orguner, "On the reduction of gaussian inverse wishart mixtures," in *2012 15th International Conference on Information Fusion*, July 2012, pp. 2162–2169.

[34] K. Granstrom and U. Orguner, "On spawning and combination of extended/group targets modeled with random matrices," *IEEE Transactions on Signal Processing*, vol. 61, no. 3, pp. 678–692, Feb 2013.

[35] K. Granström and U. Orguner, "New prediction for extended targets with random matrices," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 2, pp. 1577–1589, April 2014.

[36] K. Granström and J. Bramstång, "Bayesian smoothing for the extended object random matrix model," *IEEE Transactions on Signal Processing*, vol. 67, no. 14, pp. 3732–3742, July 2019.

[37] P. Braca, A. Aubry, L. M. Millefiori, A. D. Maio, and S. Marano, "Bayesian multi-class covariance matrix filtering for adaptive environment learning," in *26th European Signal Processing Conference (EUSIPCO)*, Sep. 2018, pp. 266–270.

[38] G. Alfano, A. De Maio, and A. M. Tulino, "A theoretical framework for LMS MIMO communication systems performance analysis," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5614–5630, Nov. 2010.

[39] J. C.-C. Chan and I. Jeliazkov, "MCMC estimation of restricted covariance matrices," *Journal of Computational and Graphical Statistics*, vol. 18, no. 2, pp. 457–480, Jan. 2009.

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

[41] G. Pailloux, P. Forster, J. P. Ovarlez, and F. Pascal, "Persymmetric adaptive radar detectors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 4, pp. 2376–2390, Oct. 2011.

[42] R. S. Raghavan, "Maximal invariants and performance of some invariant hypothesis tests for an adaptive detection problem," *IEEE Trans. Signal Process.*, vol. 61, no. 14, pp. 3607–3619, Jul. 2013.

[43] A. Farina, F. Gini, M. Greco, and P. Lee, "Improvement factor for real sea-clutter Doppler frequency spectra," *IEE Proc. Radar Sonar Navig.*, vol. 143, no. 5, pp. 341–344, 1996.

[44] A. De Maio, A. Farina, and G. Foglia, "Knowledge-based recursive least squares techniques for heterogeneous clutter suppression," *IET Radar, Sonar & Navigation*, vol. 1, no. 2, pp. 106–115, 2007.

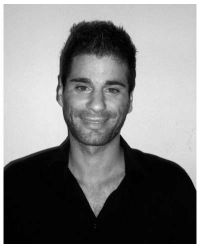
[45] F. Gini and M. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Processing*, vol. 82, no. 12, pp. 1847–1859, dec 2002.

[46] A. K. Gupta and D. K. Nagar, *Matrix variate distributions*, ser. Chapman & Hall/CRC monographs and surveys in pure and applied mathematics. Chapman & Hall, 2000.

[47] D. Maiwald and D. Kraus, "Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 147, no. 4, pp. 162–168, Aug. 2000.

[48] A. Doucet, N. Freitas, and N. Gordon, "An introduction to sequential Monte Carlo methods," in *Sequential Monte Carlo Methods in Practice*. Springer New York, 2001, pp. 3–14.

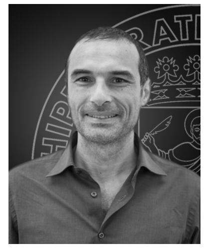

**Paolo Braca** (M'14, SM'17) received the Laurea degree (*summa cum laude*) in electronic engineering, and the Ph.D. degree (highest rank) in information engineering from the University of Salerno, Italy, in 2006 and 2010, respectively. In 2009, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT, USA. In 2010—2011, he was a Postdoctoral Associate with the University of Salerno, Italy. In 2011, Dr. Braca joined the NATO Science & Technology Organization Centre for Maritime Research and Experimentation (CMRE), and is currently a Senior Scientist with the Research Department. As project manager, he has led the Data Knowledge Operational Effectiveness (DKOE) project, and several other research projects funded by the EU Horizon 2020, the US Office of Naval Research, and the Defence Research and Development Canada. In 2017 and 2018, Dr. Braca obtained the National Scientific Qualification to function as Associate and Full Professor in Italian universities, respectively. In 2019, he was appointed as Adjunct Professor with the University of Cassino, Italy. He conducts research in the general area of statistical signal processing with emphasis on detection and estimation theory, wireless sensor network, multi-agent algorithms, target tracking and data fusion, adaptation and learning over graphs, radar (sonar) signal processing. He is co-author of over one hundred publications in international scientific journals and conference proceedings. Dr. Braca serves as Associate Editor of IEEE Transactions on Signal Processing, IEEE Transactions on Aerospace and Electronic Systems, ISIF Journal of Advances in Information Fusion, EURASIP Journal on Advances in Signal Processing, and IET Radar Sonar and Navigation. In 2017 he was Lead Guest Editor of the special issue "Sonar Multi-Sensor Applications and Techniques" in IET Radar Sonar and Navigation. He served as Associate Editor of IEEE Signal Processing Magazine (E-Newsletter) from 2014 to 2016. He is in the Technical Committee of the major international conferences in the field of signal processing and data fusion. He was the recipient of the Best Student Paper Award (first runner-up) at the 12$^{\text{th}}$ International Conference on Information Fusion in 2009. He received the NATO STO Scientific Achievement Award 2017 for his contribution to "Development and Demonstration of Networked Autonomous ASW" from the NATO Chief Scientist. He co-authored the paper which was awarded with the Best Paper Award (first runner-up) at the Sensor Signal Processing for Defence (SSPD) 2019.



**Augusto Aubry** (M'12, SM'16) received the Dr.Eng. degree (with honors) and the Ph.D. degree in information engineering, both from the University of Naples Federico II, Naples, Italy, in 2007 and 2011, respectively. From February to April 2012, he was a Visiting Researcher with the Hong Kong Baptist University, Hong Kong. He is currently under research agreement with the Department of Electrical and Information Technology Engineering, University of Naples Federico II. His research interests include statistical signal processing and optimization theory, with emphasis on MIMO communications and radar signal processing. Dr. Aubry is also the co-recipient of the 2013 best paper award (entitled to B. Carlton) of the IEEE Transactions on Aerospace and Electronic Systems with the contribution "Knowledge-Aided (Potentially Cognitive) Transmit Signal and Receive Filter Design in Signal-Dependent Clutter".

**Leonardo M. Millefiori** (S'12–M'14) received the B.Sc. degree in aerospace information engineering and the M.Sc. degree (*summa cum laude*) in communication engineering with focus on radar systems and remote sensing from the Sapienza University of Rome, Italy, in 2010 and 2013, respectively. He was a visiting researcher with the NATO Science and Technology Organization Center for Maritime Research and Experimentation, La Spezia, where he joined the Research Department as research scientist in 2014. His research interests include target motion modeling, statistical learning and signal processing, and target tracking and data fusion.

**Antonio De Maio** received the Dr.Eng. (Hons.) and Ph.D. degrees in information engineering from the University of Naples Federico II, Naples, Italy, in 1998 and 2002, respectively. From October to December 2004, he was a Visiting Researcher with the U.S. Air Force Research Laboratory, Rome, NY, USA. From November to December 2007, he was a Visiting Researcher with the Chinese University of Hong Kong, Hong Kong. He is currently a Professor with the University of Naples Federico II. His research interest lies in the field of statistical signal processing, with emphasis on radar detection, optimization theory applied to radar signal processing, and multiple-access communications. He is the recipient of the 2010 IEEE Fred Nathanson Memorial Award as the young (less than 40 years of age) AESS Radar Engineer 2010 whose performance is particularly noteworthy as evidenced by contributions to the radar art over a period of several years, with the following citation for "robust CFAR detection, knowledge-based radar signal processing, and waveform design and diversity". He is the corecipient of the 2013 best paper award (entitled to B. Carlton) of the IEEE Transactions On Aerospace And Electronic Systems with the contribution "Knowledge-Aided (Potentially Cognitive) Transmit Signal and Receive Filter Design in Signal-Dependent Clutter".



**Stefano Marano** received the Laurea degree in Electronic Engineering and the Ph.D. degree in Electronic Engineering and Computer Science, both from the University of Naples, Italy, in 1993 and 1997, respectively. Since 1999 he has been with the University of Salerno, Italy, where he is now is a Professor with DIEM. He has held visiting positions at Physics Department, University of Wales, College of Cardiff, and at the ECE Dept., University of California at San Diego, in 1996 and 2013, respectively. His areas of interest include statistical signal processing with emphasis on distributed inference, sensor networks, and information theory. Stefano Marano was awarded the IEEE Transactions on Antennas and Propagation 1999 Best Paper Award, and co-authored the paper winning the Best Student Paper Award (2nd place) at the 12th Conference on Information Fusion in 2009. He was in the Organizing Committee of FUSION 2006 and RADARCON 2008. Stefano Marano is Member (and subcommittee chair) of the EURASIP Special Area Team in Theoretical and Methodological Trends in Signal Processing, and Guest Editor of a Special Issue of EURASIP Journal on Advances in Signal Processing. He served as Associate Editor for the IEEE Transactions On Signal Processing (2010–2014) and as Associate Editor and Technical Editor for the IEEE Transactions On Aerospace And Electronic Systems (2009–2016). He is a Senior Member of the IEEE.