Innovative Solutions based on the EM-algorithm for Covariance Structure Detection and Classification in Polarimetric SAR Images

Pia Addabbo, *Senior Member, IEEE*, Filippo Biondi, *Member, IEEE*, Carmine Clemente, *Senior Member, IEEE*,
Sudan Han, Danilo Orlando, *Senior Member, IEEE*, and Giuseppe Ricci, *Senior Member, IEEE*

***Abstract*—This paper addresses the challenge of identifying the polarimetric covariance matrix (PCM) structures associated with a polarimetric SAR image. Interestingly, such information can be used, for instance, to improve the scene interpretation or to enhance the performance of (possibly PCM-based) segmentation algorithms as well as other kinds of methods. To this end, a general framework to solve a multiple hypothesis test is introduced with the aim to detect and classify contextual spatial variations in polarimetric SAR images. Specifically, under the null hypothesis, only one unknown structure is assumed for data belonging to a 2-dimensional spatial sliding window, whereas under each alternative hypothesis, data are partitioned into subsets sharing different PCM structures. The problem of partition estimation is solved by resorting to hidden random variables representative of covariance structure classes and the expectation maximization (EM) algorithm. The effectiveness of the proposed detection strategies is demonstrated on both simulated and real polarimetric SAR data also in comparison with existing classification algorithms.**

***Index Terms*—Adaptive Radar Detection, Model Order Selection, Multiple Hypothesis Testing, Expectation Maximization, Polarimetric Radar, Radar, Synthetic Aperture Radar.**

I. INTRODUCTION

In the last 20 years, the benefits of information extraction from synthetic aperture radar (SAR) [1]–[3] and, in particular, polarimetric SAR images have been widely demonstrated in a range of applications including environmental monitoring [4]–[6], security [7], [8] and urban area monitoring [9], [10]. Thanks to the increasing number of use cases for this specific type of sensor, more and more current and future remote sensing missions use polarimetric SAR, despite their increased costs. A key aspect of polarimetric SAR is the capability to extract information about the scattering mechanisms of the scene of interest, thus allowing for a more advanced characterization of the scene as in [11] where a segmentation algorithm is devised exploiting reflectivity edges and the polarimetric covariance matrix (PCM) to identify the objects contained in the region of interest. Remarkably, the incoherent polarimetric scattering phenomenon of a medium can be completely described by using the covariance matrix [12]. Generally speaking, symmetric properties arise in the encountered medium, which are, in principle, detectable through the structure of the related covariance matrix. However, there exist different structures for the PCM due to the scattering symmetries of (spatially) distributed targets. As a matter of fact, in geophysical remote sensing, the encountered media usually possess symmetry properties [12]. For instance, water waves on the ocean surface have reflection symmetry with respect to the vertical plane parallel to the wind direction resulting in zero scattering coefficients of the PCM when the azimuthal direction coincides with the wind direction; agricultural plants usually are grown along rows that exhibit reflection symmetry, whereas azimuthal symmetry is often observed in forests. These properties give rise to specific PCM structures whose knowledge exhibits a practical value in various applications. For instance, PCM structure information can be used to improve the scene interpretation or to enhance the performance of the PCM-based segmentation methods in order to classify different cover types [12]. As a matter of fact, the a priori knowledge about the PCM structure can increase the PCM estimation quality by forcing the identified structures through suitable data transformations. In addition, this information can also be used in conjunction with other techniques such as multi-model based detection techniques exploiting PCM structure information in order to tune the background model. Another example is the use of the identified PCM-structures as features (in combination with other features) to be included in a scene classifier or even a target recognition framework. Otherwise stated, PCM-structure-based classification is complementary to the other segmentation techniques.

In the context of PCM structure detection and identification, the first approach is provided in [13] where the classification task is performed by processing data belonging to a window centered on the pixel under test. Such a window moves over the entire image in order to classify all its pixels. More importantly, the main design assumption is that polarimetric data within the sliding window are characterized by the same

---

The work of Dr. Clemente was supported by Physical Sciences Research Council under Grant EP/K014307/1.

Pia Addabbo is with Università degli studi "Giustino Fortunato", Benevento, Italy. E-mail: `p.addabbo@unifortunato.eu`.

Fillippo Biondi is with Italian Ministry of Defence. Email: `biopippo@gmail.com`.

Carmine Clemente is with the University of Strathclyde, Department of Electronic and Electrical Engineering, 204 George Street, G1 1XW, Glasgow, Scotland. E-mail: `carmine.clemente@strath.ac.uk`

S. Han is with the National Innovation Institute of Defense Technology, Beijing, China E-mail: `xiaoxiaosu0626@163.com`.

D. Orlando is with Università degli Studi "Niccolò Cusano", 00166 Roma, Italy. E-mail: `danilo.orlando@unicusano.it`.

G. Ricci is with the Dipartimento di Ingegneria dell'Innovazione, Università del Salento, Via Monteroni, 73100 Lecce, Italy. E-mail: `giuseppe.ricci@unisalento.it`.

PCM structure. However, in scenarios of practical interest, this assumption can be often broken and, consequently, the window contains data with different PCM structures. In these situations, methods devised in [13] would detect the dominant PCM structure (or symmetry) only, thus missing the details related to the other PCM structures present in the window that can be suitably used to improve the interpretation of the scene. Therefore, the extraction of the entire information about the PCM structures contained in a set of contiguous pixels represents a compelling task that has not been investigated in the literature.

With the above remarks in mind, in this paper we go beyond the formulation of [13], [14] and assume (at design stage) that vectors belonging to the window under test are not all representative of the same predominant scattering symmetry and, hence, PCM structure (see Section 3.3.4 of [15]). Otherwise stated, we assume that the window can be partitioned into subsets whose elements are characterized by the same PCM (and, hence, PCM structure), whereas elements of different subsets exhibit different PCM structures (and, hence, PCMs). As briefly mentioned before, we consider the structures arising from specific predominant scattering symmetries (see also [13]). Specifically, in addition to the more general case of Hermitian structure which is not representative of any predominant symmetry related to spatially distributed scatterers, we consider three PCM structures that are associated with the scattering mechanisms due to [15]: a distributed target having reflection symmetry in the plane normal to the line-of-sight, a distributed target having rotation symmetry around the line-of-sight, and a distributed target exhibiting both reflection and rotation symmetries (see the next section for further details). This kind of classification problem is new and, hence, investigated here for the first time (at least to the best of authors' knowledge). From a more formal point of view, we are interested in discriminating between the hypothesis where all data vectors under test share the same PCM structure and the situations that include different PCM structures. This problem can be naturally formulated in terms of a multiple hypothesis test comprising one null hypothesis and several alternative hypotheses whose number depends on the combinations of PCM structures that can characterize data window. As a consequence, each alternative hypothesis identifies a certain class of partitions (i.e., considering a specific combination of the available PCM structures). In order to solve such a difficult problem, we propose an approximation of the so-called penalized generalized log-likelihood ratio test (LLRT) devised in [16] where in place of the maximum likelihood estimates of the unknown parameters, we use suitable estimates obtained by means of new estimation procedures based upon the joint exploitation of the expectation maximization algorithm (EM-algorithm) [17] and hidden random variables representative of the PCM structure classes. Even though the above idea has been widely adopted in different fields (see, for instance, [18]–[20]), the specific problem at hand is new and cannot be solved using existing results. Thus, we conceive suitable and innovative modifications also dictated by the fact that the maximum likelihood approach (MLA) would lead to very time demanding estimation procedures. As a matter of fact, it requires a maximization over all the possible window partitions and for each combination of the available PCM structures. For this reason, we design here an alternative approach grounded on the equivalence between partitioning and labeling. Finally, notice that the classification task is accomplished by the penalized log-likelihood ratio test where the log-likelihood ratio is adjusted through suitable penalty terms borrowed from the model order selection (MOS) rules [21]. Interestingly, the classification results about the scattering symmetries could be suitably exploited to drive conventional [12] and AI-based classifiers. In fact, the comparisons between the proposed procedures coupled with the conventional H/α classifier [15], [22], the plain H/α method, and the classifier implemented in the software developed by the European Space Agency (ESA) called "Polarimetric SAR data Processing and Education Toolbox" [23], have shown promising results.

The remainder of this paper is organized as follows. The next section formally introduces the multiple hypothesis test defining the measurement models as well as the unknown parameters. Section III describes the estimation procedures along with the design of the detection architectures. Illustrative examples based upon both simulated and real-recorded data are confined to Section IV, whereas concluding remarks and possible future research lines are contained in Section V.

A. Notation

In the sequel, vectors and matrices are denoted by boldface lower-case and upper-case letters, respectively. The symbols $\det(\cdot)$, $\mathrm{Tr}\,(\cdot)$, $(\cdot)^T$, and $(\cdot)^\dagger$ denote the determinant, trace, transpose, and conjugate transpose, respectively. As to the numerical sets, $\mathbb{R}$ is the set of real numbers, $\mathbb{R}^{N\times M}$ is the Euclidean space of $(N \times M)$-dimensional real matrices (or vectors if $M = 1$), $\mathbb{C}$ is the set of complex numbers, and $\mathbb{C}^{N\times M}$ is the Euclidean space of $(N \times M)$-dimensional complex matrices (or vectors if $M = 1$). If A and B are two sets, $A \setminus B$ is the set containing the elements of A that do not belong to B; the empty set is denoted by $\emptyset$. The modulus of $x \in \mathbb{C}$ is denoted by $|x|$, whereas symbol $\propto$ means proportional to. Symbol $\Re e\left\{z\right\}$ indicates the real part of the complex number z. The acronyms PDF and IID mean probability density function and independent and identically distributed, respectively. $\boldsymbol{I}$ and $\boldsymbol{0}$ stand for the identity matrix and the null vector/matrix of proper size, respectively. Finally, we write $\boldsymbol{x} \sim \mathcal{CN}_N(\boldsymbol{m}, \boldsymbol{M})$ if $\boldsymbol{x}$ is a complex circular N-dimensional normal vector with mean $\boldsymbol{m}$ and positive definite covariance matrix $\boldsymbol{M}$.

II. SENSOR MODEL AND PROBLEM FORMULATION

A multipolarization SAR sensor generates an image (datacube) where each pixel is represented by a vector whose entries are the complex returns corresponding to the different polarimetric channels. Here, we assume that the medium is reciprocal allowing to deal with the three polarimetric channels HH, HV, and VV [12]. Let us denote by L and M the numbers of pixels along the vertical and horizontal dimensions of the polarimetric image, respectively, and assume that the sensor provides a single-look complex datacube of size $L\times M\times 3$ (see

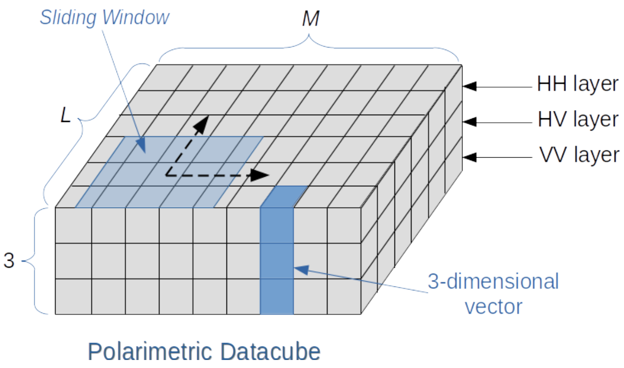


Figure 1: Schematic representation of a polarimetric SAR image as a datacube and sliding window used to obtain data under test.

Figure 1). Now, the set of vectors under test is selected using a sliding window that moves over the image and, assuming non-textured models only, contains $K = L \times M$ statistically independent random vectors $\boldsymbol{z}_k \in \mathbb{C}^{3\times 1}$, $k = 1, \ldots, K$, such that $\boldsymbol{z}_k \sim \mathcal{CN}_3(\boldsymbol{0}, \boldsymbol{M}_k)$, with[1] $\boldsymbol{M}_k \in \mathbb{C}^{3\times 3}$, $k = 1, \ldots, K$, the positive definite unknown PCM.

Moreover, since the coherency matrix can be expressed as the sum of certain target components that depend on spatial distributions of the scatterers leading to specific structures for the PCM, in order to simplify the ensuing developments, we consider four different types of predominant scattering symmetries [15]. It is clear that such an assumption is a simplification of the general scattering problem, but it allows for an analytical treatment of the classification problem at hand. As a matter of fact, there exists an uncountable number of possible scatterers' configurations and, hence, of target decompositions. Therefore, given the kth vector, we assume that the PCM structure takes on one of the following forms [12], [15]:

- in the presence of a reciprocal medium without any predominant scattering symmetry, we have that
$$\boldsymbol{M}_k = \begin{bmatrix} c_{hhhh} & c_{hhhv} & c_{hhvv} \\ c^*_{hhhv} & c_{hvhv} & c_{hvvv} \\ c^*_{hhvv} & c^*_{hvvv} & c_{vvvv} \end{bmatrix} = \boldsymbol{C}_1; \tag{1}$$
- in the presence of a predominant reflection symmetry with respect to a plane normal to the line-of-sight, the structure becomes
$$\boldsymbol{M}_k = \begin{bmatrix} c_{hhhh} & 0 & c_{hhvv} \\ 0 & c_{hvhv} & 0 \\ c^*_{hhvv} & 0 & c_{vvvv} \end{bmatrix} = \boldsymbol{C}_2; \tag{2}$$
- when a distributed target has rotation symmetry around the line-of-sight, we can write
$$\boldsymbol{M}_k = \begin{bmatrix} c_{hhhh} & c_{hhhv} & c_{hhvv} \\ -c_{hhhv} & c_{hvhv} & c_{hhhv} \\ c_{hhvv} & -c_{hhhv} & c_{hhhh} \end{bmatrix} = \boldsymbol{C}_3, \tag{3}$$
where $\mathfrak{Re}\{c_{hhhv}\} = 0$, $c_{hhvv} \in \mathbb{R}$ and $c_{hvhv} = (c_{hhhh} - c_{hhvv})/2$;
- in the presence of both reflection symmetry in some special plane and rotation symmetry (i.e., azimuth symmetry), the PCM is given by
$$\boldsymbol{M}_k = \begin{bmatrix} c_{hhhh} & 0 & c_{hhvv} \\ 0 & c_{hvhv} & 0 \\ c_{hhvv} & 0 & c_{hhhh} \end{bmatrix} = \boldsymbol{C}_4, \tag{4}$$
where $c_{hhvv} \in \mathbb{R}$ and $c_{hvhv} = (c_{hhhh} - c_{hhvv})/2$.

It turns out that within the sliding window containing the vectors under test, several situations may occur according to the involved structures. Specifically, the PCM can remain unaltered within the sliding window or at least two different forms appear within the window.

To be more precise, we are interested in distinguishing the case $\boldsymbol{M}_1 = \ldots = \boldsymbol{M}_K \in \mathcal{C} = \{\boldsymbol{C}_1, \ldots, \boldsymbol{C}_4\}$ from different configurations where the pixels are characterized by at least two PCMs. This problem can be formulated in terms of a multiple hypothesis test consisting of one null hypothesis and several alternative hypotheses, namely as

$$\begin{cases} H_0: \ \boldsymbol{z}_k \sim \mathcal{CN}_3(\boldsymbol{0}, \boldsymbol{C}_{i_0}), i_0 \in \{1, \ldots, 4\}, \\ H_{1,1}: \begin{cases} \boldsymbol{z}_k \sim \mathcal{CN}_3(\boldsymbol{0}, \boldsymbol{C}_{i_0}), k \in \Omega_1 \subset \Omega, \\ \boldsymbol{z}_k \sim \mathcal{CN}_3(\boldsymbol{0}, \boldsymbol{C}_{i_1}), k \in \Omega_2 = \Omega \setminus \Omega_1, \\ i_0 < i_1, i_0, i_1 \in \{1, \ldots, 4\}, \end{cases} \\ H_{1,2}: \begin{cases} \boldsymbol{z}_k \sim \mathcal{CN}_3(\boldsymbol{0}, \boldsymbol{C}_{i_0}), k \in \Omega_1 \subset \Omega, \\ \boldsymbol{z}_k \sim \mathcal{CN}_3(\boldsymbol{0}, \boldsymbol{C}_{i_1}), k \in \Omega_2 \subset \Omega \setminus \Omega_1, \\ \boldsymbol{z}_k \sim \mathcal{CN}_3(\boldsymbol{0}, \boldsymbol{C}_{i_2}), k \in \Omega_3 = \Omega \setminus \{\Omega_1 \cup \Omega_2\}, \\ i_0 < i_1 < i_2, i_0, i_1, i_2 \in \{1, \ldots, 4\}, \end{cases} \\ H_{1,3}: \begin{cases} \boldsymbol{z}_k \sim \mathcal{CN}_3(\boldsymbol{0}, \boldsymbol{C}_1), k \in \Omega_1 \subset \Omega, \\ \boldsymbol{z}_k \sim \mathcal{CN}_3(\boldsymbol{0}, \boldsymbol{C}_2), k \in \Omega_2 \subset \Omega \setminus \Omega_1, \\ \boldsymbol{z}_k \sim \mathcal{CN}_3(\boldsymbol{0}, \boldsymbol{C}_3), k \in \Omega_3 \subset \Omega \setminus \{\Omega_1 \cup \Omega_2\}, \\ \boldsymbol{z}_k \sim \mathcal{CN}_3(\boldsymbol{0}, \boldsymbol{C}_4), k \in \Omega_4 = \Omega \setminus \{\Omega_1 \cup \Omega_2 \cup \Omega_3\}, \end{cases} \end{cases} \tag{5}$$

where $\Omega = \{1, \ldots, K\}$ and the Ω_ls are unknown (except for Ω_{i+1} under $H_{1,i}$ that can be computed as $\Omega \setminus \cup_{j=1}^{i} \Omega_j$). Recalling that $\boldsymbol{z}_k$s are statistical independent, the joint PDF of $\boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K]$ under H_0 is given by

$$p_0(\boldsymbol{Z}; \boldsymbol{C}_{i_0}) = \frac{\exp\left\{-\mathrm{Tr}\left[\boldsymbol{C}_{i_0}^{-1} \boldsymbol{Z}\boldsymbol{Z}^\dagger\right]\right\}}{\pi^{3K}\left[\det\left(\boldsymbol{C}_{i_0}\right)\right]^K}, \tag{6}$$

whereas that under $H_{1,m}$, $m = 1, \ldots, 3$, can be written as

$$p_{1,m}(\boldsymbol{Z}; \boldsymbol{C}_{i_0}, \ldots, \boldsymbol{C}_{i_m}) = \prod_{l=0}^{m} \prod_{k \in \Omega_{l+1}} \frac{\exp\left\{-\mathrm{Tr}\left[\boldsymbol{C}_{i_l}^{-1} \boldsymbol{z}_k \boldsymbol{z}_k^\dagger\right]\right\}}{\pi^3 \left[\det(\boldsymbol{C}_{i_l})\right]} \tag{7}$$

with the constraints

$$\bigcup_{l=0}^{m} \Omega_{l+1} = \Omega \quad \text{and} \quad \Omega_i \cap \Omega_j = \emptyset, \ i \neq j. \tag{8}$$

For future reference, it is also useful to define the sets

$$\mathcal{A}_m = \{i_0, \ldots, i_m\} \subseteq \{1, \ldots, 4\}, \ m = 0, \ldots, 3, \tag{9}$$

and[2] denote by $\boldsymbol{\theta}_0(\mathcal{A}_0)$ and $\boldsymbol{\theta}_1(\mathcal{A}_m)$ the unknown parameters under H_0, given $\mathcal{A}_0$, and under $H_{1,m}$, given $\mathcal{A}_m$, respectively.

---

[1] Notice that this data model arises from simplified assumptions allowing for analytical tractability and is not always valid for real SAR images.

[2] Notice that $\mathcal{A}_3 = \{1, 2, 3, 4\}$.

III. DETECTION ARCHITECTURE DESIGNS

In this section, we provide some important remarks that are preparatory to the subsequent derivations and motivate the design choices. As specified below, the adopted decision rules rely on the LLRT where the unknown parameters are replaced by suitable estimates. However, the implementation of such a strategy for the problem at hand requires to circumvent two main drawbacks.

First of all, under $H_{1,m}$, the partition $\{\Omega_1, \ldots, \Omega_{m+1}\}$ of the pixels of the sliding window is not known. As a consequence, application of the MLA to obtain the parameter estimates would be a formidable task: we should consider all the combinations of $m+1$ PCM structures over the available options, namely $\binom{4}{m+1}$, and for each of them the different partitions of Ω into $m+1$ subsets. Therefore, in what follows, we propose two alternative solutions that abstain from the computation of all the possible partitions of Ω. These alternatives rely on the fact that, from an operating point of view, partitioning Ω is tantamount to labeling its elements. Therefore, we can follow the lead of [18] and introduce K IID hidden discrete random variables that are representative of the labels associated with the $\boldsymbol{z}_k$s under $H_{1,m}$, and $\mathcal{A}_m = \{i_0, \ldots, i_m\} \subseteq \{1, \ldots, 4\}$. In fact, such random variables take on values in $\mathcal{A}_m$. Then, we apply the EM-algorithm [17] to estimate the unknown parameters within the window under test. The herein proposed estimation procedures differ from each other in the way such hidden random variables are defined and used to build up the LLRT, a point better specified at the end of this section.

The second drawback of implementing a plain LLRT is originated by the fact that the elements of $\mathcal{C}$ are characterized by different numbers of unknowns. Thus, not only a balanced comparison of the hypotheses, but also of the different $\mathcal{A}_m$s, given m (i.e., given the hypothesis), requires introducing adequate penalty factors. To be more quantitative, we observe that the number n_i of unknown parameters associated with $\boldsymbol{C}_i$, $i = 1, \ldots, 4$, is given by

$$n_i = \begin{cases} 9 & \text{if } i = 1, \\ 5 & \text{if } i = 2, \\ 3 & \text{if } i = 3, \\ 2 & \text{if } i = 4. \end{cases} \tag{10}$$

Accordingly, the number of unknowns associated with $\mathcal{A}_m = \{i_0, \ldots, i_m\}$ can be computed as $u(\mathcal{A}_m) = \sum_{j=0}^{m} n_{i_j}$.

With the above remarks in mind, we devise decision schemes for problem (5) exploiting a penalized LLRT [24]. As a first step towards the introduction of such a penalized LLRT, we denote by $\widehat{\boldsymbol{\theta}}_0(\mathcal{A}_0)$ ($\widehat{\boldsymbol{\theta}}_1(\mathcal{A}_m)$) the estimate of the unknown parameters related to H_0 and $\mathcal{A}_0$ ($H_{1,m}$ and $\mathcal{A}_m$). Similarly, $\widehat{\boldsymbol{\theta}}_1(\mathcal{A}_{\widehat{m}})$ is the estimate of the unknown parameters associated with $H_{1,\widehat{m}}$ and $\mathcal{A}_{\widehat{m}}$. For the moment, we leave aside the description of the estimation procedures, which will be the object of the next subsections, and introduce the general structure of the penalized LLRT

$$\max_{m\in\{1,2,3\}} \max_{\mathcal{A}_m} \left[\log g_1\left(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}_1(\mathcal{A}_m)\right) - h_1(\mathcal{A}_m)\right]$$
$$- \max_{\mathcal{A}_0} \left[\log g_0\left(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}_0(\mathcal{A}_0)\right) - h_0(\mathcal{A}_0)\right] \underset{H_0}{\overset{H_{1,\widehat{m}}}{\gtrless}} \eta, \tag{11}$$

where

$$\widehat{m} = \arg\max_{m\in\{1,2,3\}} \left\{ \max_{\mathcal{A}_m} \left[\log g_1\left(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}_1(\mathcal{A}_m)\right) - h_1(\mathcal{A}_m)\right]\right\},$$

$g_0(\boldsymbol{Z}; \boldsymbol{\theta}_0(\mathcal{A}_0)) = p_0(\boldsymbol{Z}; \boldsymbol{C}_{i_0})$, $g_1(\boldsymbol{Z}; \boldsymbol{\theta}_1(\mathcal{A}_m))$ denotes the PDF of the observables under $H_{1,m}$ and $\mathcal{A}_m$, that will be specified in the subsequent sections, $h_1(\mathcal{A}_m)$, $m = 1, 2, 3$, is a penalty term accounting for the number of unknown parameters related to $H_{1,m}$ and $\mathcal{A}_m$, $h_0(\mathcal{A}_0)$ is a penalty term accounting for the number of unknown parameters related to H_0 and $\mathcal{A}_0$, and η is the detection threshold[3] to be set according to the probability of false alarm (P_{fa}). The penalty terms can be written as

$$h_1(\mathcal{A}_m) = \gamma\left(u(\mathcal{A}_m) + m + 1\right), \tag{12}$$
$$h_0(\mathcal{A}_0) = \gamma u(\mathcal{A}_0), \tag{13}$$

where we recall that $u(\mathcal{A}_m)$ is the number of unknown real-valued parameters associated with $\mathcal{A}_m$, $m+1$ is the number of unknowns arising from the probability mass function (PMF) of the hidden discrete random variables (such random variables take on values in $\mathcal{A}_m$), and γ is a factor borrowed from the MOS rules [21] as the Akaike information criterion (AIC), the generalized information criterion (GIC), and the Bayesian information criterion (BIC), i.e.,

$$\gamma = \begin{cases} 1, & \text{for AIC-based detector (AIC-D)}, \\ \log(6K)/2, & \text{for BIC-based detector (BIC-D)}, \\ (1+\rho)/2, \ \rho > 1, & \text{for GIC-based detector (GIC-D)}. \end{cases} \tag{14}$$

As already observed, under $H_{1,m}$ and $\mathcal{A}_m$, $u(\mathcal{A}_m)$ is obtained by partitioning the data set into $m+1$ subsets, associating with them specific structures, and summing the respective number of unknown parameters. The cardinality of each subset along with the coordinates of the vectors within it are also unknowns, but they are independent of $\mathcal{A}_m$ and, hence, irrelevant to the decision process.

It still remain to show how to estimate $\boldsymbol{\theta}_0(\mathcal{A}_0)$ and $\boldsymbol{\theta}_1(\mathcal{A}_m)$. As previously anticipated, we will follow the lead of [18] and introduce K independent and identically distributed hidden discrete random variables that "specify the characterization" of the $\boldsymbol{z}_k$s. Then, we apply the EM-algorithm [17] to estimate the unknown parameters. The herein proposed estimation procedures differ from each other in the way such hidden random variables are defined and used to build up the LLRT under $H_{1,m}$.

---

[3] Hereafter, we denote by η the generic detection threshold.

The first procedure assumes that under $H_{1,m}$ the hidden random variables, $c_{k,m}$ say, have alphabet $\mathcal{A}_m = \{i_0, \ldots, i_m\} \subseteq \{1 \ldots, 4\}$ with PMF

$$\begin{cases} P(c_{k,m} = l) = P_{l,m}, \ l \in \mathcal{A}_m, \\ \displaystyle\sum_{l \in \mathcal{A}_m} P_{l,m} = 1, \end{cases} \tag{15}$$

and that when $c_{k,m} = l$, $l \in \mathcal{A}_m$, then $\boldsymbol{M}_k = \boldsymbol{C}_l$. Therefore, we can write the PDF of $\boldsymbol{z}_k$ under $H_{1,m}$ exploiting the following mixture model [18]

$$f_{1,m}(\boldsymbol{z}_k; \boldsymbol{\theta}_1(\mathcal{A}_m)) = \sum_{l \in \mathcal{A}_m} P_{l,m} f(\boldsymbol{z}_k; \boldsymbol{C}_l), \tag{16}$$

where $f(\boldsymbol{z}_k; \boldsymbol{C}_l)$ is the PDF of $\boldsymbol{z}_k \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{C}_l)$. The above PDF will be used in place of the original PDF to form the LLRT. Notice that $\boldsymbol{\theta}_1(\mathcal{A}_m)$ depends on the specific choice for the alphabet of the hidden random variables. As a matter of fact, for each alternative hypothesis, each of the $\binom{4}{m+1}$ combinations of the PCM structures identifies an alphabet configuration. Thus, we come up with 6, 4, and 1 different alphabet configurations under $H_{1,1}$, $H_{1,2}$, and $H_{1,3}$, respectively. Nevertheless, as we will show in the next subsections, these configurations can be handled without a dramatic increase of the computational requirements.

The second approach does not account for the hypotheses $H_{1,1}$, $H_{1,2}$, and $H_{1,3}$ to set the number of classes but simply considers all classes. As consequence, the hidden random variables, c_k say, share the same alphabet $\mathcal{A} = \{1, 2, 3, 4\}$ and PMF $P(c_k = l) = P_l$, $l \in \mathcal{A}$. The LLRT under $H_{1,m}$ is formed by selecting the $m+1$ most probable PCM structures and modifying (16) according to the selected structures.

In the next subsections, we describe in the detail these procedures that are based upon the EM-algorithm and exploit the transformations introduced by *Lemma 3.1* of [13]. As a matter of fact, such transformations allow us to recast the PCM in the objective function as a block-diagonal matrix leading to different maximization problems according to the specific structure.

For the reader ease, before describing the estimation procedures, we recall here that $\boldsymbol{M}_k$ is the generic unknown PCM of $\boldsymbol{z}_k$; $\boldsymbol{C}_1, \ldots, \boldsymbol{C}_4$ are the structures that the PCM of $\boldsymbol{z}_k$ can take on; $\mathcal{A}_m$ contains the integers indexing a combination of m structures under $H_{1,m}$; $\Omega = \{1, \ldots, K\}$ is the set of integers indexing the data vectors within the window under test; Ω_i, $i = 1, \ldots, 4$, is a set of integers indexing a subset of data vectors that share the same PCM structure; $P_{l,m}$ is the probability that the hidden random variable $c_{k,m}$, associated with $\boldsymbol{z}_k$ under $H_{1,m}$ and for a specific $\mathcal{A}_m$, takes on the value l, with $l \in \mathcal{A}_m$; similarly, P_l is the prior of the hidden random variable c_k associated with $\boldsymbol{z}_k$ under $H_{1,3}$.

A. First EM-based Estimation Strategy

Let us assume that under $H_{1,m}$, $m = 1, 2, 3$, equation (16) holds true and focus on problem (5). Now, given a configuration for $\mathcal{A}_m$, the log-likelihood of $\boldsymbol{Z}$ is given by

$$\log g_1\left(\boldsymbol{Z}; \boldsymbol{\theta}_1\left(\mathcal{A}_m\right)\right) = \sum_{k=1}^{K} \log \sum_{l \in \mathcal{A}_m} P_{l,m} f(\boldsymbol{z}_k; \boldsymbol{C}_l). \tag{17}$$

The application of the EM-algorithm consists of the E-step that leads to [18], [25]

$$q_k^{(h-1)}(l,m) = \frac{f(\boldsymbol{z}_k; \widehat{\boldsymbol{C}}_l^{(h-1)}) \widehat{P}_{l,m}^{(h-1)}}{\displaystyle\sum_{n \in \mathcal{A}_m} f(\boldsymbol{z}_k; \widehat{\boldsymbol{C}}_n^{(h-1)}) \widehat{P}_{n,m}^{(h-1)}}, \quad l \in \mathcal{A}_m, \tag{18}$$

where $\widehat{P}_{l,m}^{(h-1)}$, $l \in \mathcal{A}_m$, and $\widehat{\boldsymbol{C}}_n^{(h-1)}$, $n \in \mathcal{A}_m$, are the available estimates at the $(h-1)$th step, and of the M-step requiring to solve the following joint optimization problem[4]

$$\max_{\mathbf{p}_m} \max_{\substack{\mathbf{C}_l \\ l \in \mathcal{A}_m}} \left\{ \sum_{k=1}^{K} \sum_{l \in \mathcal{A}_m} q_k^{(h-1)}(l,m) [-\log\det(\boldsymbol{C}_l) - \operatorname{Tr}(\boldsymbol{C}_l^{-1} \boldsymbol{z}_k \boldsymbol{z}_k^{\dagger})] - \sum_{k=1}^{K} \sum_{l \in \mathcal{A}_m} q_k^{(h-1)}(l,m) \log P_{l,m} \right\}, \tag{19}$$

where[5] $\boldsymbol{p}_m = [P_{i_0,m}, \ldots, P_{i_m,m}]^T \in \mathbb{R}^{(m+1)\times 1}$.

It is not difficult to show that the maximization with respect to $\boldsymbol{p}_m$, accomplished under the constraint

$$\sum_{l \in \mathcal{A}_m} P_{l,m} = 1, \tag{20}$$

returns the following stationary points

$$\widehat{P}_{l,m}^{(h)} = \frac{1}{K} \sum_{k=1}^{K} q_k^{(h-1)}(l,m), \quad l \in \mathcal{A}_m. \tag{21}$$

On the other hand, the maximization with respect to $\boldsymbol{C}_l$ is accomplished by finding the maximizer for each PCM structure and, then, plugging those indexed by $\mathcal{A}_m$ into (19). Therefore, given $\bar{l} \in \mathcal{A}_m$, we solve

$$\max_{\boldsymbol{C}_{\bar{l}}} \sum_{k=1}^{K} q_k^{(h-1)}(\bar{l},m) \left[-\log\det(\boldsymbol{C}_{\bar{l}}) - \operatorname{Tr}\left(\boldsymbol{C}_{\bar{l}}^{-1} \boldsymbol{z}_k \boldsymbol{z}_k^{\dagger} \right) \right]. \tag{22}$$

Let us start with the case $\bar{l} = 1$ and, as shown in Appendix A, the update rule for the estimate of $\boldsymbol{C}_1$ is given by

$$\widehat{\boldsymbol{C}}_1^{(h)} = \frac{\displaystyle\sum_{k=1}^{K} q_k^{(h-1)}(1,m) \boldsymbol{z}_k \boldsymbol{z}_k^{\dagger}}{\displaystyle\sum_{k=1}^{K} q_k^{(h-1)}(1,m)}. \tag{23}$$

Now, assume that $\bar{l} = 2$. Then, in Appendix B, we prove that the estimate of $\boldsymbol{C}_2$ can be updated according to

$$\widehat{\boldsymbol{C}}_2^{(h)} = \boldsymbol{U}^{\dagger} \begin{bmatrix} \widehat{\boldsymbol{A}}^{(h)} & \boldsymbol{0} \\ \boldsymbol{0} & \widehat{d}^{(h)} \end{bmatrix} \boldsymbol{U}, \tag{24}$$

---

[4] Equation (19) can be obtained recalling that $\boldsymbol{z}_k$s are modeled in terms of a convex combination of complex normal distributions. For brevity, we have omitted some derivation details of the EM-algorithm and refer the interested reader to [18], [25] for further information.

[5] Notice that the entries of $\boldsymbol{p}_m$ are nonnegative.

where $\boldsymbol{U}$ is the unitary matrix defined by (6) of [13, *Lemma 3.1*] and

$$\begin{cases} \widehat{d}^{(h)} = \dfrac{\sum_{k=1}^{K} q_k^{(h-1)}(2,m)|z_{k,2}|^2}{\sum_{k=1}^{K} q_k^{(h-1)}(2,m)}, \\ \widehat{\boldsymbol{A}}^{(h)} = \dfrac{\sum_{k=1}^{K} q_k^{(h-1)}(2,m)\boldsymbol{z}_{k,1}\boldsymbol{z}_{k,1}^{\dagger}}{\sum_{k=1}^{K} q_k^{(h-1)}(2,m)}, \end{cases} \tag{25}$$

where $\boldsymbol{U}\boldsymbol{z}_k = [\boldsymbol{z}_{k,1}^T \ z_{k,2}]^T$ with $\boldsymbol{z}_{k,1} \in \mathbb{C}^{2\times 1}$ and $z_{k,2} \in \mathbb{C}$. The next case is $\bar{l} = 3$. Appendix C contains the derivations leading to the following estimate of $\boldsymbol{C}_3$

$$\widehat{\boldsymbol{C}}_3^{(h)} = \boldsymbol{T}^{\dagger}\boldsymbol{E}^{-1}\boldsymbol{V}^{\dagger}\begin{bmatrix}\widehat{a}^{(h)} & \boldsymbol{0} \\ \boldsymbol{0} & \widehat{\boldsymbol{B}}^{(h)}\end{bmatrix}\boldsymbol{V}\boldsymbol{E}^{-1}\boldsymbol{T}, \tag{26}$$

where the transformations $\boldsymbol{E}$, $\boldsymbol{T}$, and $\boldsymbol{V}$ are used in (7) of [13, *Lemma 3.1*] while

$$\begin{cases} \widehat{a}^{(h)} = \dfrac{\sum_{k=1}^{K} q_k^{(h-1)}(3,m)|x_{k,1}|^2}{\sum_{k=1}^{K} q_k^{(h-1)}(3,m)}, \\ \widehat{\boldsymbol{B}}^{(h)} = \dfrac{1}{2}\dfrac{\sum_{k=1}^{K} q_k^{(h-1)}(3,m)\left(\boldsymbol{x}_{k,2}\boldsymbol{x}_{k,2}^{\dagger} + \boldsymbol{J}\boldsymbol{x}_{k,2}\boldsymbol{x}_{k,2}^{\dagger}\boldsymbol{J}\right)}{\sum_{k=1}^{K} q_k^{(h-1)}(3,m)}, \end{cases} \tag{27}$$

with $\boldsymbol{V}\boldsymbol{E}\boldsymbol{T}\boldsymbol{z}_k = [x_{k,1} \ \boldsymbol{x}_{k,2}^T]^T$, $x_{k,1} \in \mathbb{C}$, $\boldsymbol{x}_{k,2} \in \mathbb{C}^{2\times 1}$, and $\boldsymbol{J} \in \mathbb{R}^{2\times 2}$ a permutation matrix (see Appendix C).

The derivations related to the final case, i.e., $\bar{l} = 4$, are confined to Appendix D. Specifically, it is shown that the update rule for the estimate of $\boldsymbol{C}_4$ is (see also (8) of [13, *Lemma 3.1*])

$$\widehat{\boldsymbol{C}}_4^{(h)} = \boldsymbol{T}^{\dagger}\boldsymbol{E}^{-1}\begin{bmatrix}\widehat{b}^{(h)} & 0 & 0 \\ 0 & \widehat{c}^{(h)} & 0 \\ 0 & 0 & \widehat{c}^{(h)}\end{bmatrix}\boldsymbol{E}^{-1}\boldsymbol{T}, \tag{28}$$

where

$$\begin{cases} \widehat{b}^{(h)} = \dfrac{\sum_{k=1}^{K} q_k^{(h-1)}(4,m)|y_{k,1}|^2}{\sum_{k=1}^{K} q_k^{(h-1)}(4,m)}, \\ \widehat{c}^{(h)} = \dfrac{1}{2}\dfrac{\sum_{k=1}^{K} q_k^{(h-1)}(4,m)\boldsymbol{y}_{k,2}^{\dagger}\boldsymbol{y}_{k,2}}{\sum_{k=1}^{K} q_k^{(h-1)}(4,m)}. \end{cases} \tag{29}$$

with $\boldsymbol{E}\boldsymbol{T}\boldsymbol{z}_k = [y_{k,1} \ \boldsymbol{y}_{k,2}^T]^T$, $y_{k,1} \in \mathbb{C}$, and $\boldsymbol{y}_{k,2} \in \mathbb{C}^{2\times 1}$.

The actual implementation of the EM-algorithm, necessary to obtain an estimate of $\boldsymbol{\theta}_1(\mathcal{A}_m)$, needs to specify the convergence criterion that can be used to terminate the iterations. In what follows, for each $\mathcal{A}_m$, $m = 1,2,3$, we adopt the following criterion

$$\Delta\mathcal{L}_m(h) = \left|\left[\mathcal{L}_m\left(\widehat{\boldsymbol{\theta}}_1^{(h)}(\mathcal{A}_m);\boldsymbol{Z}\right)\right.\right. \\ \left.\left. -\mathcal{L}_m\left(\widehat{\boldsymbol{\theta}}_1^{(h-1)}(\mathcal{A}_m);\boldsymbol{Z}\right)\right]/\mathcal{L}_m\left(\widehat{\boldsymbol{\theta}}_1^{(h-1)}(\mathcal{A}_m);\boldsymbol{Z}\right)\right| < \epsilon_m, \tag{30}$$

where $\mathcal{L}_m(\widehat{\boldsymbol{\theta}}_1^{(h)}(\mathcal{A}_m);\boldsymbol{Z}) = \log g_1(\boldsymbol{Z};\widehat{\boldsymbol{\theta}}_1^{(h)}(\mathcal{A}_m))$ (see (17)) and $\epsilon_m > 0$ is set accounting for the requirements in terms of system reactivity. The above estimation procedure is summarized in Algorithm 1.

The decision statistic of test (11) also requires to estimate the unknown parameters under H_0. The MLE of $\boldsymbol{C}_i$ is given by *Proposition 3.2* of [13].

B. Second EM-based Estimation Strategy

The second procedure builds up the term associated with $H_{1,m}$ of the left-hand side of (11) by considering the estimates obtained through the first procedure under $H_{1,3}$ only. Specifically, it relies on the following steps

1) assume that there exists at least one pixel for each considered PCM structure within the window;
2) estimate the unknown parameters through the procedure based upon the EM-algorithm under $H_{1,3}$;
3) exploit the estimates of the class priors at the previous step to select the most plausible subset of PCM structures associated with the window.

As shown below, the third item allows us to build up an estimate of $\mathcal{A}_m$ based upon the most plausible structures. Thus, starting from the first step, let us define $\mathcal{A}_3 = \{1,2,3,4\}$ and, given m, select the $m+1$ structures corresponding to the indices of the $m+1$ highest entries of the final estimate of $\boldsymbol{p}_3$ that is denoted by $\widehat{\boldsymbol{p}}_3 = \left[\widehat{P}_{1,3}, \widehat{P}_{2,3}, \widehat{P}_{3,3}, \widehat{P}_{4,3}\right]^T$.

To be more formal, let us sort the $\widehat{P}_{l,3}$s in descending order, namely

$$\widehat{P}_{l_0,3} \geq \widehat{P}_{l_1,3} \geq \widehat{P}_{l_2,3} \geq \widehat{P}_{l_3,3}, \tag{31}$$

and form the following subsets $\tilde{\mathcal{A}}_m = \{l_0,\ldots,l_m\}$, $m = 1,2,3$, along with the estimate $\tilde{\boldsymbol{\theta}}_1(\tilde{\mathcal{A}}_m)$ that can be drawn from $\widehat{\boldsymbol{\theta}}_1(\mathcal{A}_3)$ by picking the components corresponding to the indices $l_0,\ldots,l_m$. Then, decision rule (11) becomes

$$\max_{m\in\{1,2,3\}}\left[\log g_1\left(\boldsymbol{Z};\tilde{\boldsymbol{\theta}}_1\left(\tilde{\mathcal{A}}_m\right)\right) - h_1\left(\tilde{\mathcal{A}}_m\right)\right] \\ -\max_{\mathcal{A}_0}\left[\log g_0\left(\boldsymbol{Z};\widehat{\boldsymbol{\theta}}_0\left(\mathcal{A}_0\right)\right) - h_0\left(\mathcal{A}_0\right)\right] \underset{H_0}{\overset{H_{1,\widehat{m}}}{\gtrless}} \eta, \tag{32}$$

where

$$\log g_1\left(\boldsymbol{Z};\tilde{\boldsymbol{\theta}}_1\left(\tilde{\mathcal{A}}_m\right)\right) = \sum_{k=1}^{K}\log\sum_{l\in\tilde{\mathcal{A}}_m}\widehat{P}_{l,m}f(\boldsymbol{z}_k;\widehat{\boldsymbol{C}}_l). \tag{33}$$

The entire procedure is summarized in Algorithm 2.

The estimation of the unknown parameters under H_0 is the same as that of previous subsection.

C. Architecture Summary and Classification Rules

According to the specific penalty term in (11), that depends on (14), and the procedure pursued to come up with the parameter estimates, we can form the following architectures (see also Figure 2):

- AIC-D coupled with procedure 1 (AIC-D-P1) or procedure 2 (AIC-D-P2);
- BIC-D coupled with procedure 1 (BIC-D-P1) or procedure 2 (BIC-D-P2);
- GIC-D coupled with procedure 1 (GIC-D-P1) or procedure 2 (GIC-D-P2).

Finally, regardless of the estimation procedure applied to data under test, once the unknown parameters have been estimated, data classification is accomplished according to the following association procedure

- if test (11) returns H_0, then $\boldsymbol{z}_1,\ldots,\boldsymbol{z}_K$ are characterized by the same PCM structure whose estimate is obtained by selecting $\boldsymbol{C}_{\hat{i}_0}$ where
$$\widehat{\mathcal{A}}_0 = \left\{\hat{i}_0\right\} = \arg\max_{\mathcal{A}_0} \left[\log g_0\left(\boldsymbol{Z};\widehat{\boldsymbol{\theta}}_0\left(\mathcal{A}_0\right)\right) - h_0\left(\mathcal{A}_0\right)\right]; \tag{34}$$
- on the other hand, when (11) selects $H_{1,\widehat{m}}$, then we know that the number of structures within the window is $\widehat{m}$ (that is the final estimate of m) and the structure forms are indexed by $\widehat{\mathcal{A}}_{\widehat{m}}$ (that is the final estimate of $\mathcal{A}$). In this case, the association procedure is based upon the following maximum a posteriori rule
$$\boldsymbol{z}_k \longrightarrow \boldsymbol{C}_{\hat{i}} \quad \text{if } \hat{i} = \arg\max_{l\in\widehat{\mathcal{A}}_{\widehat{m}}} q_k^{(h_{\max})}(l,\widehat{m}), \tag{35}$$
where $h_{\max}$ is the maximum number of EM iterations used to come up with the final parameter estimates.

It is worth noticing that such an association rule exploits the estimates obtained by processing all the vectors within the window under test.

IV. NUMERICAL EXAMPLES

In this section, the behaviors of the proposed architectures are assessed using synthetic data as well as real polarimetric SAR data. Specifically, the first subsection contains the performance results over simulated data obtained by means of standard Monte Carlo (MC) counting techniques, whereas in Subsections IV-B and IV-C, the proposed procedures are tested using real polarimetric SAR data. Subsection IV-B contains an illustrative example that provides evidence of the practical value as well as the advantages of the PCM structure information. In fact, it shows how such an information can be exploited to enhance the segmentation performance of the H/α method [15], [22]. Specifically, we force the PCM structures identified by means of the proposed approaches on the PCMs processed by the H/α method and compare the results with the plain H/α method and the classifier implemented in the software developed by the European Space Agency (ESA) called "Polarimetric SAR data Processing and Education Toolbox" [23]. Notice that other PCM-based segmentation methods or frameworks can be considered such as multi-model based detection, scene classification, or even target recognition. In Subsection IV-C another interesting application of the proposed algorithms is described. More precisely, we show that the PCM structure can be used to identify vehicles that are parking in or departing from a car park.

Algorithm 1 First EM-based estimation procedure

```
Require: $\mathcal{A}_m$, $\widehat{\boldsymbol{C}}_l^{(0)}$, $\widehat{P}_{l,m}^{(0)}$, $\epsilon_m$, $l \in \mathcal{A}_m$, $m = 1,2,3$
Ensure: $\widehat{\mathcal{A}}_{\widehat{m}}$ and the related estimates $\widehat{\boldsymbol{C}}_l$, $\widehat{P}_{l,\widehat{m}}$, $l \in \widehat{\mathcal{A}}_{\widehat{m}}$
 1: For $m = 1,2,3$ do
 2:    For each $\mathcal{A}_m$ do
 3:       Compute $h_1(\mathcal{A}_m)$ (eq. (12))
 4:       Set $h = 1$
 5:       Compute $q_k^{(h-1)}(l,m)$, $l \in \mathcal{A}_m$ (eq. (18))
 6:       For each $l \in \mathcal{A}_m$ do
 7:          Compute $\widehat{\boldsymbol{C}}_l^{(h)}$ using (23), (24), (26), or (28)
 8:          Compute $\widehat{P}_{l,m}^{(h)}$ using (21)
 9:       End For
10:       If $\Delta\mathcal{L}_m(h) \geq \epsilon_m$ (eq. (30)) set $h = h+1$ and
          go to line 5 else go to line 11
11:    End For
12:    Select the $\mathcal{A}_m$ and the $\widehat{\boldsymbol{C}}_l^{(h)}$s, $l \in \mathcal{A}_m$, that maximize
       $y_1(\mathcal{A}_m, \widehat{\boldsymbol{\theta}}_1(\mathcal{A}_m)) = \log g_1(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}_1(\mathcal{A}_m)) - h_1(\mathcal{A}_m)$,
       where $\log g_1(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}_1(\mathcal{A}_m))$ is defined by (17)
13: End For
14: Select the $m$ that maximizes $y_1(\widehat{\mathcal{A}}_m, \widehat{\boldsymbol{\theta}}_1(\widehat{\mathcal{A}}_m))$, where $\widehat{\mathcal{A}}_m$
    is the $\mathcal{A}_m$ selected at line 12
15: Return: $\widehat{\mathcal{A}}_{\widehat{m}}$ and $\widehat{\boldsymbol{\theta}}_1(\widehat{\mathcal{A}}_{\widehat{m}})$ where $\widehat{m}$ has been selected at
    line 14
```

Algorithm 2 Second EM-based estimation procedure

```
Require: $\mathcal{A}_3$, $\widehat{\boldsymbol{C}}_l^{(0)}$, $\widehat{P}_{l,3}^{(0)}$, $\epsilon_3$, $l \in \mathcal{A}_3$
Ensure: $\tilde{\mathcal{A}}_{\widehat{m}}$ and the related estimates $\widehat{\boldsymbol{C}}_l$, $\widehat{P}_{l,\widehat{m}}$, $l \in \tilde{\mathcal{A}}_{\widehat{m}}$
 1: Set $h = 1$
 2: Compute $q_k^{(h-1)}(l,3)$, $l \in \mathcal{A}_3$ (eq. (18))
 3: For each $l \in \mathcal{A}_3 = \{1,2,3,4\}$ do
 4:    Compute $\widehat{\boldsymbol{C}}_l^{(h)}$ using (23), (24), (26), or (28)
 5:    Compute $\widehat{P}_{l,3}^{(h)}$ using (21)
 6: End For
 7: If $\Delta\mathcal{L}_3(h) \geq \epsilon_3$ (eq. (30)) set $h = h+1$ and
    go to line 2 else go to line 8
 8: Sort $\widehat{P}_{l,3}^{(h)}$, $l \in \mathcal{A}_3$, in decreasing order to form $\widehat{P}_{l_0,3} \geq$
    $\widehat{P}_{l_1,3} \geq \widehat{P}_{l_2,3} \geq \widehat{P}_{l_3,3}$
 9: For $m = 1,2,3$ do
10:    Form $\tilde{\mathcal{A}}_m = \{l_0,\ldots,l_m\}$
11:    Select $\tilde{\boldsymbol{\theta}}_1(\tilde{\mathcal{A}}_m)$ from $\widehat{\boldsymbol{\theta}}_1(\mathcal{A}_3)$
12:    Compute $y_1(\tilde{\mathcal{A}}_m, \tilde{\boldsymbol{\theta}}_1(\tilde{\mathcal{A}}_m)) = \log g_1(\boldsymbol{Z}; \tilde{\boldsymbol{\theta}}_1(\tilde{\mathcal{A}}_m))$
       $-h_1(\tilde{\mathcal{A}}_m)$, where $\log g_1(\boldsymbol{Z}; \tilde{\boldsymbol{\theta}}_1(\tilde{\mathcal{A}}_m))$ and $h_1(\tilde{\mathcal{A}}_m)$ are
       defined by (33) and (12), respectively
13: End For
14: Select $m$ that maximizes $y_1(\tilde{\boldsymbol{\theta}}_1(\tilde{\mathcal{A}}_m), \tilde{\mathcal{A}}_m)$
15: Return: $\tilde{\mathcal{A}}_{\widehat{m}}$ and $\tilde{\boldsymbol{\theta}}_1(\tilde{\mathcal{A}}_{\widehat{m}})$ where $\widehat{m}$ has been selected at
    line 14
```

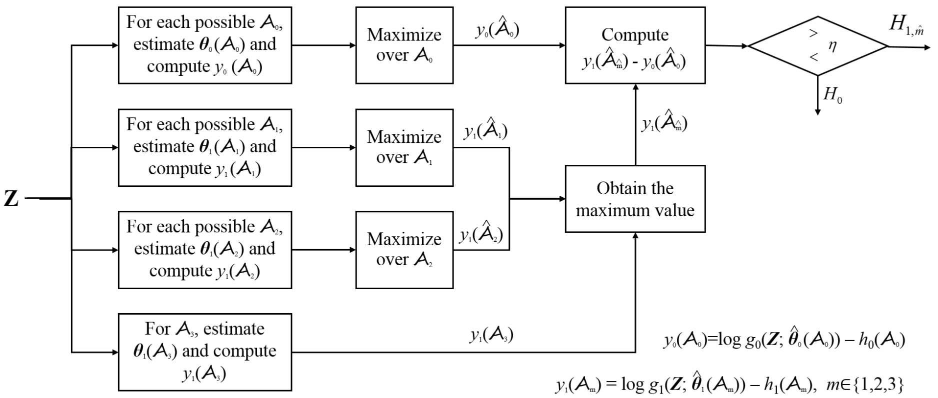


Figure 2: Block diagram for the computation of decision rule (11) (using symbols of Algorithm 1).

From the point of view of classification, natural competitors of the proposed techniques are represented by the solutions proposed in [13] that exploit the same classification features used here, namely, the PCM structures corresponding to specific scattering symmetries (see (1)-(4)). These competitors are devised assuming that all the pixels within the window under classification share the same PCM structure. Finally, we also consider a *more conventional* competitor based upon the well-known K-nearest neighbors (KNN) algorithm trained with the actual values of the PCMs [25].

A. Simulated Data

The simulated data obey the multivariate circular complex Gaussian distribution with zero mean and nominal covariance matrices related to four scenarios: no symmetry, reflection, rotation, and azimuth symmetries. Specifically, they are given by

$$\boldsymbol{C}_1 = \begin{bmatrix} 1 & 0.2+0.3j & 0.5-0.3j \\ 0.2-0.3j & 0.25 & -0.2-0.2j \\ 0.5+0.3j & -0.2+0.2j & 0.8 \end{bmatrix}, \quad (36)$$

$$\boldsymbol{C}_2 = \begin{bmatrix} 1 & 0 & 0.5-0.3j \\ 0 & 0.25 & 0 \\ 0.5+0.3j & 0 & 0.4 \end{bmatrix}, \quad (37)$$

$$\boldsymbol{C}_3 = \begin{bmatrix} 1 & 0.3j & 0.2 \\ -0.3j & 0.4 & 0.3j \\ 0.2 & -0.3j & 1 \end{bmatrix}, \quad (38)$$

$$\boldsymbol{C}_4 = \begin{bmatrix} 1 & 0 & 0.5 \\ 0 & 0.25 & 0 \\ 0.5 & 0 & 1 \end{bmatrix}, \quad (39)$$

respectively. In the numerical examples below, the number of data (K) ranges from 60 to 240, and data are partitioned into adjacent subsets characterized by different PCM structures. The parameter ρ (of GIC-based architectures) is set to 3 for the competitor [13], 1.3 for GIC-D-P1, and 11 for GIC-D-P2 (these values are selected in order to guarantee a good compromise between underestimation and overestimation of the model order). Finally, we consider $P_{fa} = 10^{-2}$ and the related detection thresholds are estimated as follows

1) compute the detection threshold under H_0 and for each PCM structure;
2) the final threshold (namely, η in (11)) is set by selecting the maximum of the thresholds obtained at the previous step.

The above procedure guarantees that the actual P_{fa} is less than or equal to the nominal P_{fa}.

As a preliminary analysis, we focus on the requirements of the proposed procedures in terms of the EM iterations. To this end, in Figure 3, we plot the log-likelihood variations, i.e., $\Delta\mathcal{L}_m(h)$, $m = 1, 2, 3$, as a function of h, averaged over 1000 MC trials. It turns out that, for all the analyzed cases, a number of 10 iterations (this value will be used in the subsequent analysis) is sufficient to ensure log-likelihood variations less than 10^{-4}, namely, $\epsilon_m < 10^{-4}$.

In Figures 4-7, we investigate the instantaneous behavior of the proposed architectures by showing the classification outcomes of a single Monte Carlo trial using a window of size 9×20. These figures are obtained by generating data as follows

- under H_0, all data share $\boldsymbol{C}_1$;
- under $H_{1,1}$, data are split in two equal parts, where the PCM of the first and second halves are $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$, respectively;
- under $H_{1,2}$, data are partitioned into three subsets with the same cardinality and characterized by $\boldsymbol{C}_1$, $\boldsymbol{C}_2$, and $\boldsymbol{C}_3$;
- under $H_{1,3}$, four equal subsets are generated and, clearly, all the PCMs are used.

In these figures, the estimated structure is mapped to its structure index, namely, i ($\in \{1, \ldots, 4\}$) means that $\boldsymbol{C}_i$ has been selected. The ground truth is reported at the beginning of each subfigure. As for the KNN-based competitor, the training stage comprises 500 vectors for each PCM class, assuming that the PCMs are perfectly known, whereas, from an operating point of view, the number of neighbors is 50.

From the figures' inspection, it is evident the advantage (at least from a qualitative point of view) of the proposed architectures. As a matter of fact, they overcome the KNN-based competitor under each hypothesis, whereas, competitor [13] has worse performance than the new architectures when $H_{1,m}, m = 1, \ldots, 3$, is in force. Moreover, as expected, it clearly turns out that $H_{1,3}$ represents the most challenging

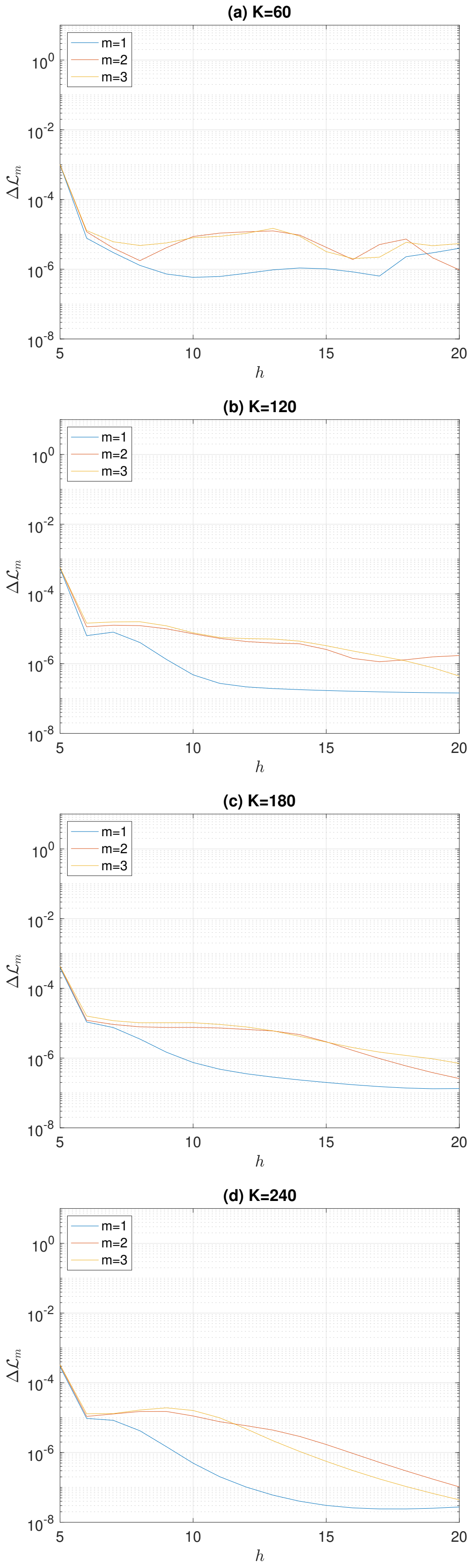
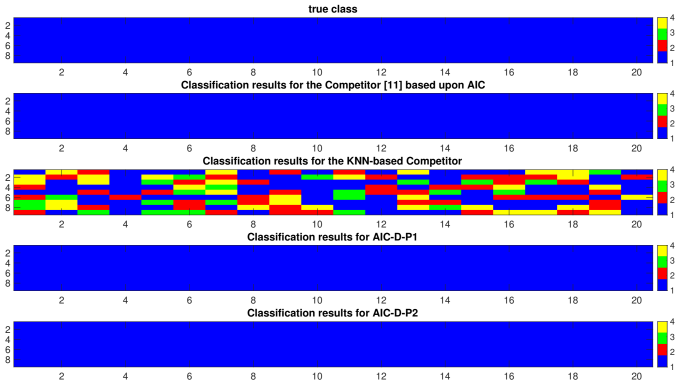
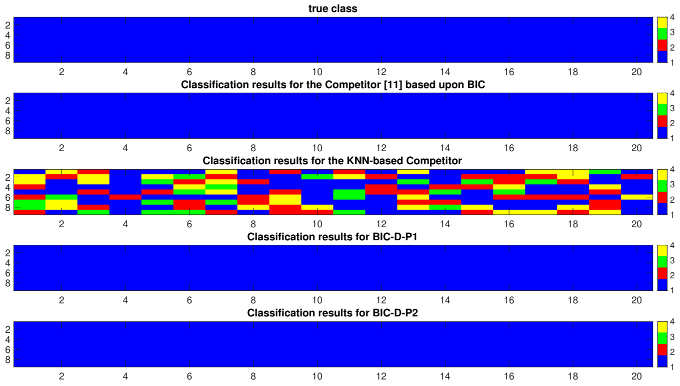


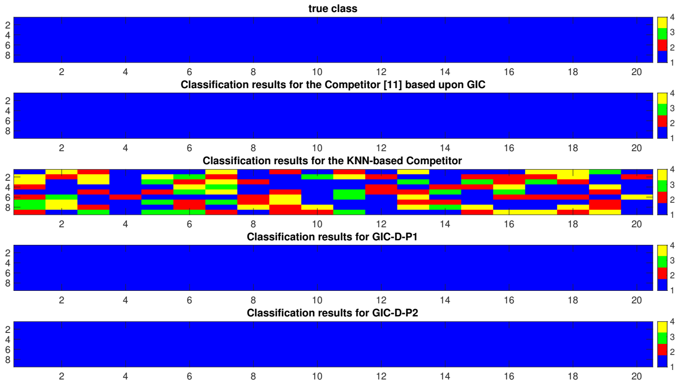
Figure 3: Log-likelihood variations versus the iteration number h of the EM-algorithm for different values of K.



(a) AIC-based



(b) BIC-based



(c) GIC-based

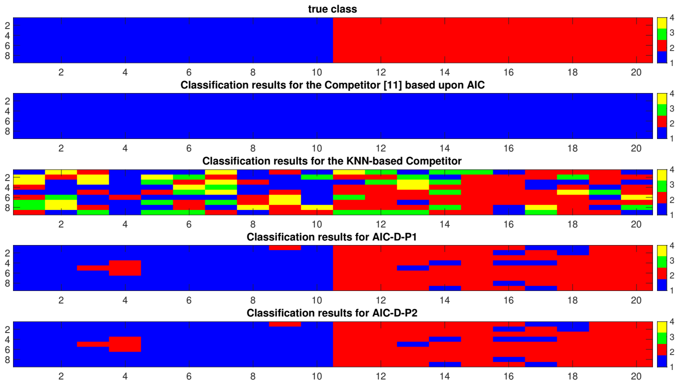
Figure 4: Classification results for a single MC trial under H_0.

scenario with the major difficulties in correctly classifying the azimuth symmetry (yellow pixels present in the last partition of data set). As a matter of fact, AIC-D-P1, AIC-D-P2, BIC-D-P2, GIC-D-P1, and GIC-D-P2 are capable of only partially classifying such pixels as characterized by azimuth symmetry.

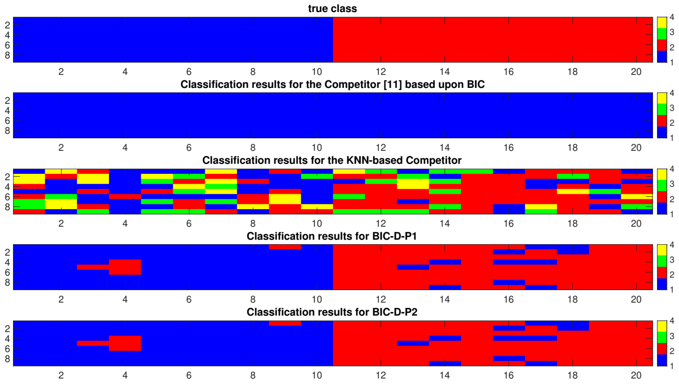
In Figures 8 and 9, we investigate the behavior of the proposed architectures[6] from a more quantitative standpoint. In fact, these figures contain the histograms of correct classification over 1000 independent MC trials assuming $K = 120$ and $K = 180$, respectively. Such histograms are representative of the probability of correct classification (P_c) defined as the probability of declaring H_0 or $H_{1,m}$, $m = 1, \dots, 3$, under H_0

---

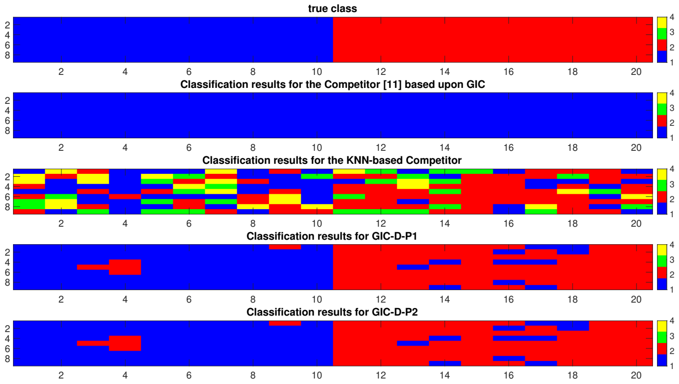
[6] We do not include the competitors in this analysis since they experience poor performances.



(a) AIC-based

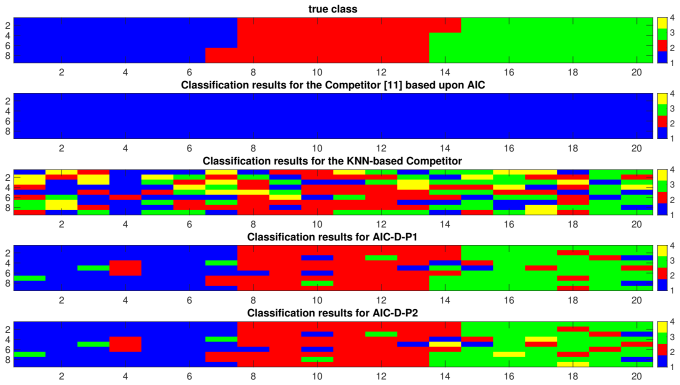


(b) BIC-based

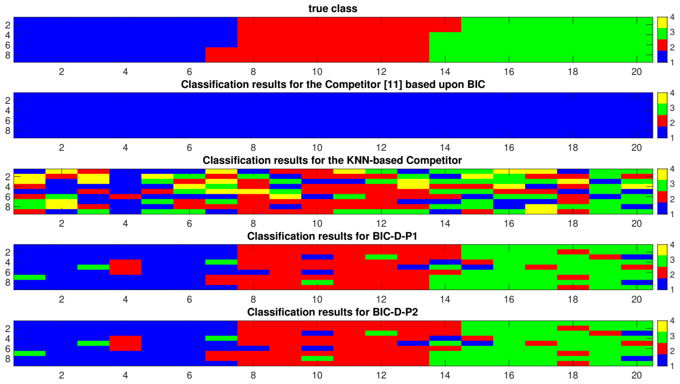


(c) GIC-based

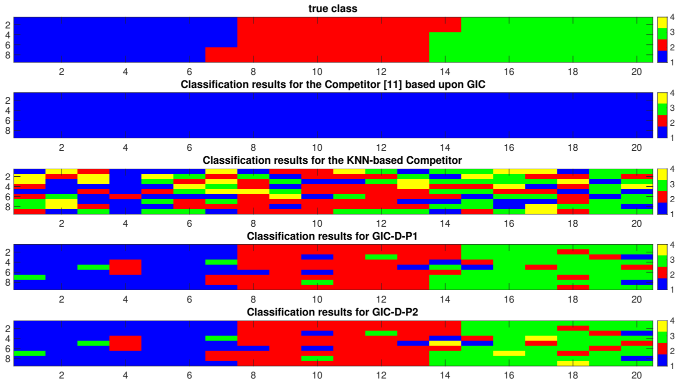
Figure 5: Classification results for a single MC trial under $H_{1,1}$.



(a) AIC-based



(b) BIC-based



(c) GIC-based

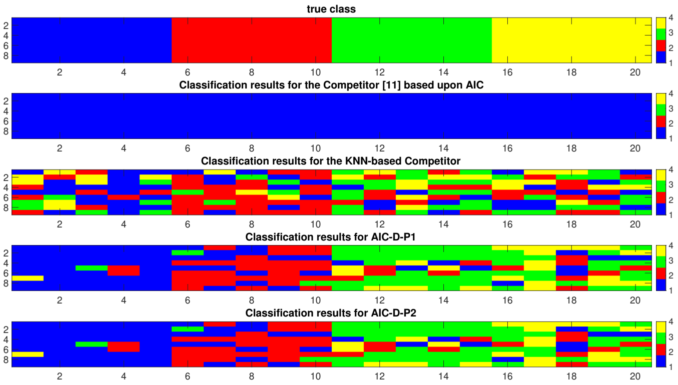
Figure 6: Classification results for a single MC trial under $H_{1,2}$.

Algorithm 1	Algorithm 2	Competitor [11]	KNN Competitor
448e-3	72.8e-3	0.719e-3	347.1e-3

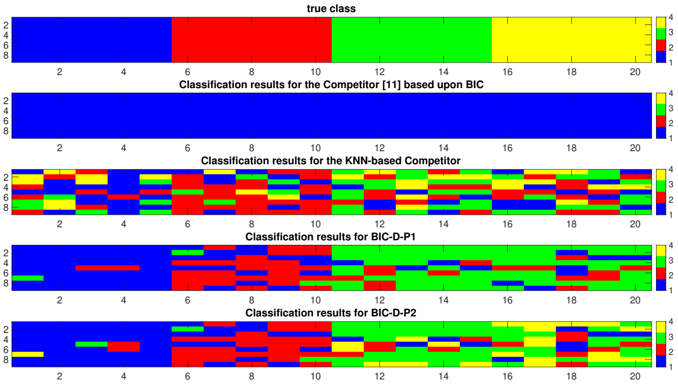
Table I: Computation times in [sec] for a single MC trial.

or $H_{1,m}$, respectively. As expected, under H_0, all the proposed architectures return P_c values very close to 100%. Under $H_{1,1}$, all the considered architectures can provide percentages of correct classification close to 100% except for AIC-D-P2 whose P_c values are around 0.70. Almost similar behaviors can be observed under $H_{1,2}$ with the difference that architectures based upon the second EM-based procedure have lower classification capabilities with respect to the results under $H_{1,1}$. Under this hypothesis, the performance of AIC-D-P2 is very poor due to a strong overestimation inclination. Such inclination is also experienced by BIC-D-P2 for $K = 180$ since the resulting P_c is about 0.56. Under $H_{1,3}$, which represents the most challenging case, we notice that for $K = 120$ the classification values are below 0.75 for all the considered architectures with BIC-D-P1 returning the worst performance. When K increases to 180, the situation is clearly better than for $K = 120$ even though the classification performance of BIC-D-P1 is less than 0.5. The other architectures ensure P_c values greater than 92%.

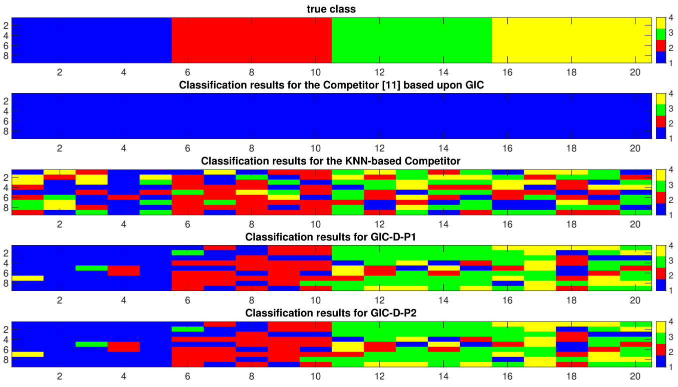
The curves reported in Figure 10 pertain to the probability of PCM variation detection (P_d) and the normalized root mean square classification error (RMSCE) values both as functions of K; notice that the P_d is defined as the probability of



(a) AIC-based



(b) BIC-based



(c) GIC-based

Figure 7: Classification results for a single MC trial under $H_{1,3}$.

rejecting H_0 under $H_{1,m}$, whereas the RMSCE is the root mean square number of misclassified vectors divided[7] by K. Data are generated under the most challenging hypothesis $H_{1,3}$ and, again, the performance parameters are estimated over 1000 MC independent trials. From Subfigure 10(a), it turns out that the curves associated with the considered architectures are close to each other when $K > 120$ with a maximum difference of about 0.1. This difference becomes negligible as K increases. As a matter of fact, AIC-D-P1, AIC-D-P2, and GIC-D-P1 are capable of achieving $P_d = 1$ at $K = 240$, whereas BIC-D-P1, BIC-D-P2, and GIC-D-P2 return $P_d = 0.998$, $P_d = 0.994$, and $P_d = 0.978$, respectively, at $K = 240$. In Subfigure 10(b), we plot the normalized RMSCE versus K. The figure points out that the error curves for AIC-D-P1, AIC-D-P2, and GIC-D-P1 are almost overlapped outperforming the other classifiers at least for $K < 240$. The worst performance is returned by BIC-D-P1 as expected from the analysis of the classification histograms.

Summarizing, the above analysis indicates that AIC-D-P1 and GIC-D-P1 can guarantee an excellent compromise between detection performance and classification results under each hypothesis for $K > 120$. In addition, notice that if we consider subsets of hypotheses, other architectures can provide reliable classification and detection performance starting from $K > 120$.

Finally, in Table I, the computation times, expressed in seconds and evaluated for a single run of the procedures, are reported for both the Algorithm 1 and 2 and compared with the competitors. As expected, it turns out that the competitors [13] have the smallest computation time. In addition, the second EM-based estimation strategy (Algorithm 2) outperforms the first one (Algorithm 1). As for the KNN-based competitor, it sightly outperforms Algorithm 1 but it requires more computation time than Algorithm 2.

B. *Real Recorded Data Set from EMISAR Sensor*

In this subsection, we consider the fully polarimetric SAR data acquired by the EMISAR airborne sensor[8] in the L-band (1.25 GHz). The set is formed by 1750 rows and 1000 columns. The scene under investigation, shown in Figure 11, is over the Foulum Area, Denmark. Precisely, in the figure we show on the left the Pauli decomposition of the fully-polarimetric SAR image while on the right we have the same image acquired through an optical sensor. It turns out that the scene contains a mixed urban, vegetation, as well as water scene. Therefore, it is representative of different scattering mechanisms that allow us to suitably verify the classification capabilities of the proposed algorithms in a real-world manifold scenario. The rectangular boxes in the figure highlight the two urban areas of Tjele and Orum. In the ensuing numerical examples, we do not include the classification results provided by the KNN-based algorithm since it requires a training stage that should be fed by data adhering as much as possible to the ground truth. However, in this case, the actual values for the PCMs are not available and, given the classification results of Subsection IV-A, the KNN-based competitor would suffer from a further performance degradation due to the mismatched operating conditions. Nevertheless, we show the advantages of the proposed approach when coupled with conventional classifiers such as the H/α method in comparison with the plain H/α itself and that implemented in the ESA classification software. The classifier of this software exploits a matrix distance based upon the Wishart distribution and initializes the image classes by means of the H/α algorithm. Then, these initial classes are refined by resorting to the k-mean algorithm [23].

---

[7] This normalization is necessary for comparison purposes.

[8] Data can be downloaded at: https://earth.esa.int/web/polsarpro/data-sources/sampledatasets.

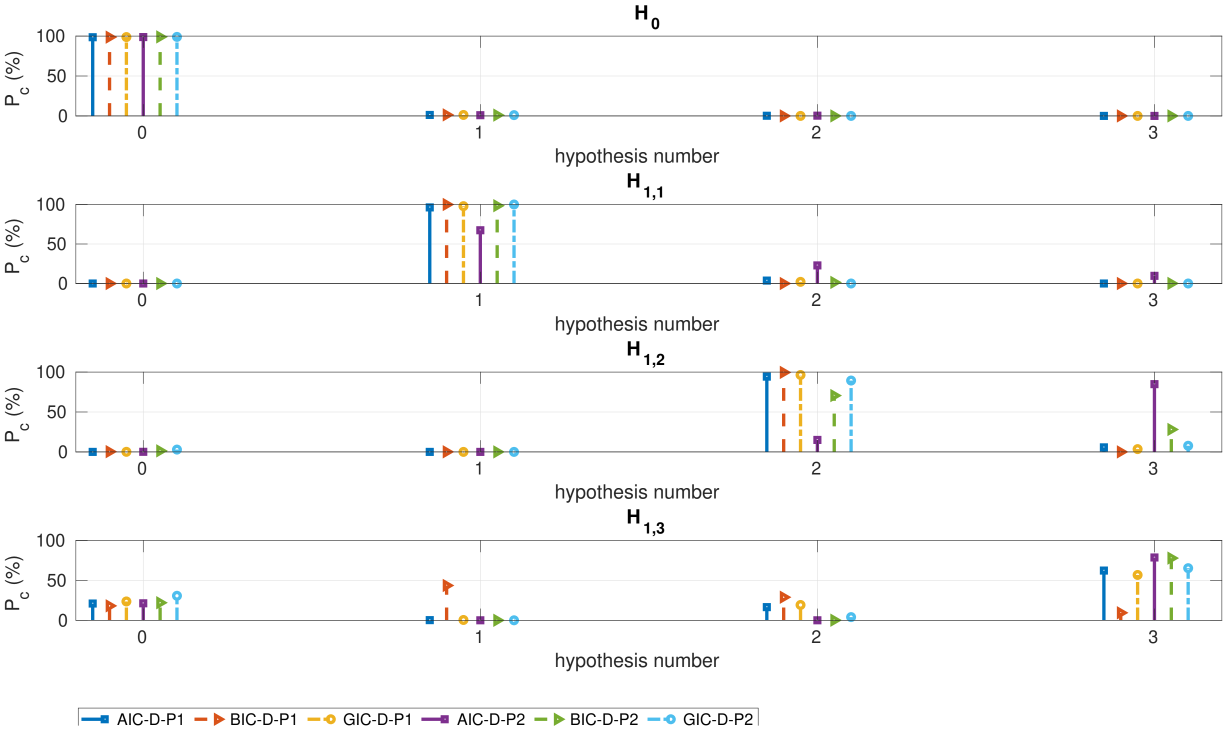


Figure 8: P_c (%) versus H_0 and $H_{1,m}$, $m = 1, 2, 3$, assuming that $K = 120$.

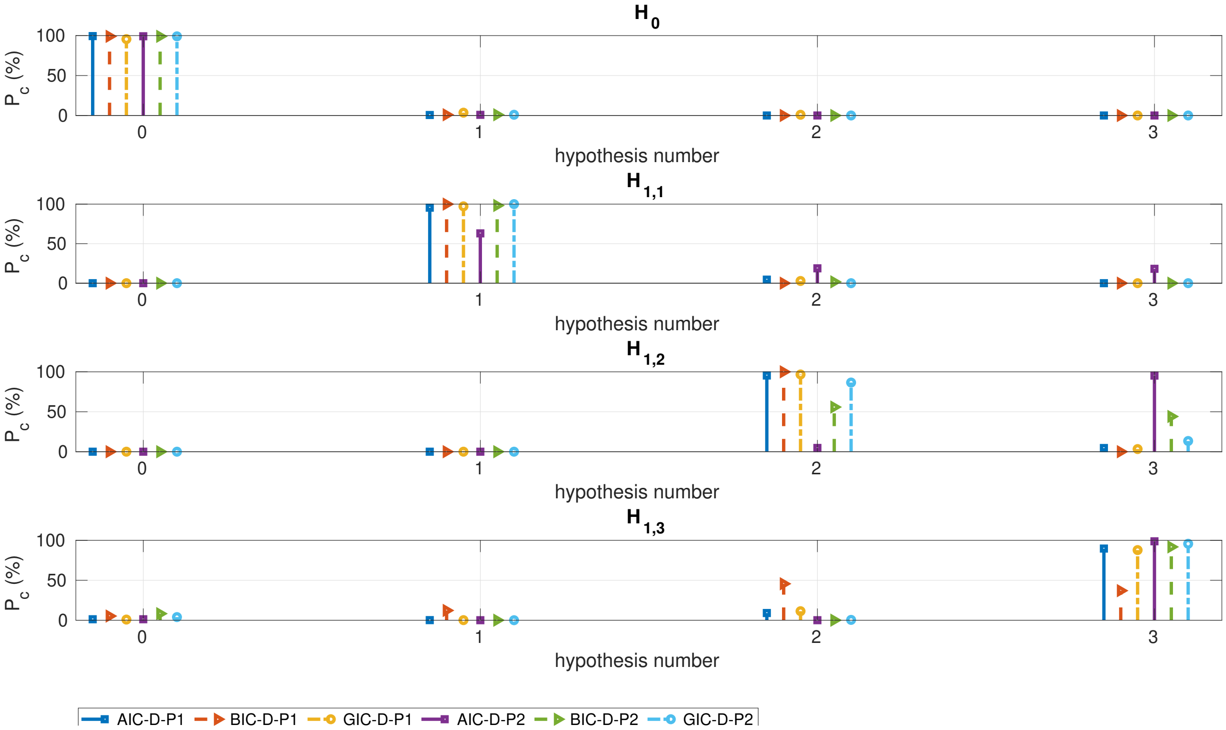
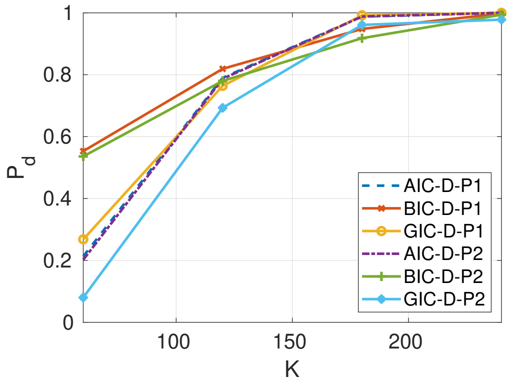
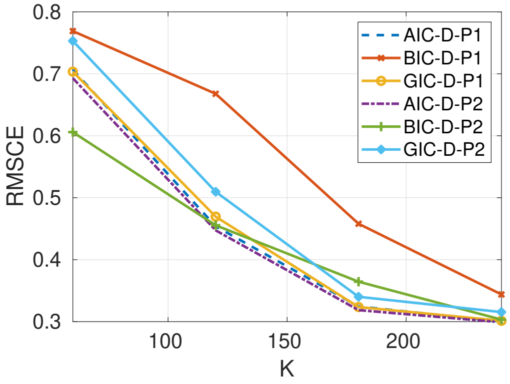


Figure 9: P_c (%) versus H_0 and $H_{1,m}$, $m = 1, 2, 3$, assuming that $K = 180$.



(a) P_d



(b) RMSCE

Figure 10: P_d and RMSCE versus K assuming that $H_{1,3}$ is true.

In Figure 12, we compare the classification results of the proposed architectures with those of the natural competitors [13]. Thus, such a figure contains maps of structures (identified by the considered algorithms) that can be used as described at the beginning of this section. A window of size 11×11 pixels is used[9] and the threshold is set to the value obtained with the synthetic simulations. The figure clearly sheds light on the fact that the proposed architectures are capable of providing enhanced details and finer resolutions with respect to the competitor due to the inherent best classification capabilities. In all the considered cases, the absence of symmetry (blue pixels) is revealed over the water. Red pixels indicating a detected reflection symmetry, in place of crops and bare fields, are predominant for the BIC-like and GIC-D-P2 architectures. Yellow pixels (azimuth symmetry) are classified in the presence of forest areas. Rotation symmetries (green pixels) are very few for the classification maps obtained by the competitor, whereas, they are more present in the results obtained with the proposed architectures and appear in the regions containing buildings (for example in the two highlighted urban areas) and roads (that are more clearly visible for the proposed architectures with respect to the competitor).

In order to show the practical value of the structure information and, hence, of the proposed methods, in what follows we present an illustrative example of how PCM structure information can be used in conjunction with a classical segmentation method. More precisely, we focus on the H/α classifier, which returns nine classes as defined in [15], [22], and compare its classification results when data are preprocessed using the PCM structures identified by the proposed algorithms with those obtained without any side information. The processing scheme exploiting the PCM structure information is referred to in the following as EM H/α and, for conciseness, only the AIC-D-P1 detector is considered in this analysis (the other architectures provide similar results).

The confusion matrix between the two mentioned classifiers is reported in Table II, in which the H/α classes are assumed as references. The results obtained are in agreement with those in [13], with the larger difference between the EM H/α and the H/α, visible in the migration of areas classified as medium entropy vegetation (class 5) to high entropy vegetation (class 2). The row corresponding to class 3 contains zero entries due to the fact that the original H/α classifier does not label any pixel as belonging to that class. Moreover, it turns out that structures corresponding to buildings would result more visible in the image produced by the EM H/α classifier (i.e., pixels classified as 1 that are representative of double-bounce mechanisms). This is evident in the two highlighted urban areas of Figure 13 that contain the classes returned by the H/α method when it is aided by PCM structure information and when it does not use any side information. The figure highlights that for the EM H/α classifier, the Tjele and the Orum areas contain more details with respect to the classic H/α classifier that returns more homogeneous regions. In fact, exploiting the PCM structure identification obtained through the proposed EM-based algorithm, pixels of these areas are labeled as class-1, which corresponds to echoes from buildings.

Finally, in Figure 14, the classification results returned by the ESA software based upon Wishart distribution, the H/α classifier, and the k-mean algorithm are shown [23]. From the comparison with Figure 13, it turns out that the ESA software and the EM H/α classifier provide fine details on the urban areas even though the ESA software is more inclined to label pixels as belonging to the class 3 which is the same class assigned to the lake visible on the left bottom part of the figure. This behavior is more evident in the rural fields where the green color associated to the class 3 is dominant.

C. Analysis on Real X-band Data Set with Ground Truth

In order to demonstrate the capability of the proposed approach to classify different symmetries and relate them to a verified truth on the ground, in this subsection an analysis performed using real X-band data is presented. The used data set is the Coherent Change Detection Challenge data set acquired by the Air Force Research Laboratory [26], [27]. The overall data set provides 10 complex images for each of the three available polarizations (HH, VV, and HV), acquired the same day. The original images size is of 4501×4501 pixels and the images have been coherently aligned to a single reference (per polarization) with the help of Digital

---

[9] The window moves over the entire image without data overlapping between consecutive positions.

EM H/α

		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
H/α **classes**	**1**	99.27	0.73	0	0	0	0	0	0	0
	2	3.39	96.61	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0
	4	9.17	11.57	0	79.26	0	0	0	0	0
	5	0	17.02	0.51	2.93	76.85	2.69	0	0	0
	6	0	0	0	0	1.09	98.91	0	0	0
	7	8.25	5.71	0	7.77	0.51	0	77.77	0	0
	8	0	12.02	0.05	3.47	5.73	0.05	1.71	76.96	0
	9	0	0	0	0	0.01	3.09	0	0	96.89

Table II: Confusion matrix (values expressed in percentage).

Elevation Map information [26]. In this analysis two passes have been considered, the acquisition named "FP0121" is used as reference pass, while the acquisition "FP0124" is used as a test pass and a ground truth is derived as in [28]. In particular, an area over a car park is selected, this region is represented by a sub-image of 1000×1000 pixels and is composed of several parking lots which are occupied by numerous parked (i.e., stationary) vehicles. This scenario is particularly useful as it provides the opportunity to assess the capability of the proposed approach to pick changes in the PCM structure between the reference and test image, namely when a vehicle has been parked or has left the car park between the two acquisitions. Figure 15 shows the reference and test sub-images for the HH mode.

The proposed approach[10] is applied to both acquisitions and the change between classification output is analysed. In particular, as it is expected that on the tarmac no predominant symmetry would be returned, while in the presence of a vehicle the dominant symmetry would be the reflection one, the transitions between these two detected symmetries are considered. So when on the expected change there is a transition from no dominant to reflection symmetry it can indicate an "arrival" of a vehicle in the scene between acquisitions, while a transition from reflection to no dominant symmetry indicates a "departed" car. The obtained result together with the ground truth is shown in Figure 16.

By visually comparing the obtained symmetry transition map with the ground truth it is possible to observe that in most cases the transition between the two symmetries is a good indicator of a departure or of an arrival of a vehicle in the scene in between acquisitions, and on the total of 34 changes between the two images, only in 9 cases a clear decision could not be made by confidently relying on the symmetry transition map only.

V. CONCLUSIONS

In this paper, we have addressed the problem of detecting and classifying PCM structure variations within a data window moving over a polarimetric SAR image. Unlike the existing structure classification procedures that assume a specific PCM structure for all vectors belonging to the sliding window, in this case, data might exhibit different unknown PCM structures. More importantly, the partition of the entire data set according to the respective PCM structures is unknown and must be estimated. This problem naturally leads to a multiple hypothesis test with one null hypothesis and multiple alternative hypotheses. In order to avoid a significant computational load, we have devised a design framework, grounded on hidden random variables, which assign a PCM structure label to data vectors, and the EM-algorithm tailored to the considered PCM structures. Interestingly, PCM structure information can be exploited in conjunction with other (possibly PCM-based) segmentation techniques or frameworks such as multi-model based detection, scene classification, or even target recognition.

The performance analysis, conducted on simulated also in comparison with suitable competitors, has highlighted that AIC-D-P1 and GIC-D-P1 are capable of providing an excellent compromise between detection and classification performance under all the considered hypotheses and for $K > 120$. In addition, if we restrict the set of hypotheses of interest, other architectures can guarantee good classification/detection performance at least for values of K greater than 120. The analysis over real-recorded data, in comparison with the plain H/α method and the ESA software, has shown promising results.

Future research tracks might encompass the extension of such architectures to the heterogeneous environment where the reflectivity coefficient within the window under investigation is not spatially stationary.

APPENDIX A
DERIVATION OF (23) (UPDATE RULE FOR $\boldsymbol{C}_1$)

When $\bar{l} = 1$, then the objective function in (22) can be recast as

$$\sum_{k=1}^{K} q_k^{(h-1)}(1,m)\left[-\log\det(\boldsymbol{C}_1) - \mathrm{Tr}\left(\boldsymbol{C}_1^{-1}\boldsymbol{z}_k\boldsymbol{z}_k^{\dagger}\right)\right]$$
$$= q^{(h-1)}(1,m)\left\{\log\det(\boldsymbol{C}_1^{-1}) - \mathrm{Tr}\left[\boldsymbol{C}_1^{-1}\boldsymbol{S}_q(1,m)^{(h-1)}\right]\right\}, \qquad (40)$$

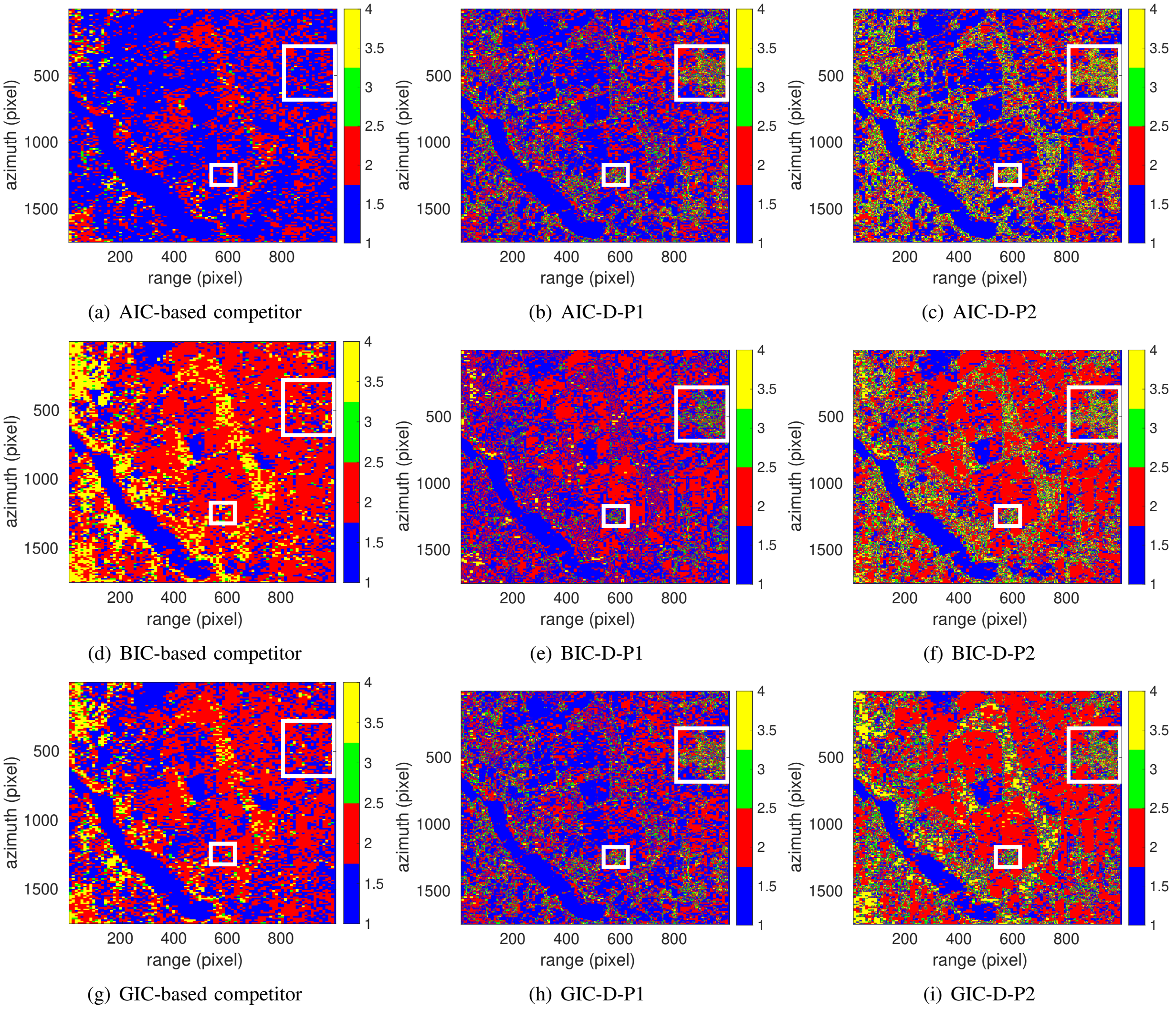
where

$$q^{(h-1)}(1,m) = \sum_{k=1}^{K} q_k^{(h-1)}(1,m),$$
$$\boldsymbol{S}_q(1,m)^{(h-1)} = \sum_{k=1}^{K} q_k^{(h-1)}(1,m)\boldsymbol{z}_k\boldsymbol{z}_k^{\dagger}/q^{(h-1)}(1,m).$$

---

[10] For conciseness here, the use of AIC-D-P1 only is discussed.

Figure 11: Observed scene. Left subplot: Pauli decomposition of the fully-polarimetric SAR image. Right subplot: optical image drawn from Google Earth ©.



(a) AIC-based competitor (b) AIC-D-P1 (c) AIC-D-P2

(d) BIC-based competitor (e) BIC-D-P1 (f) BIC-D-P2

(g) GIC-based competitor (h) GIC-D-P1 (i) GIC-D-P2

Figure 12: Classification maps with real SAR data: urban area of Tjele (small rectangle) and Orum (great rectangle).

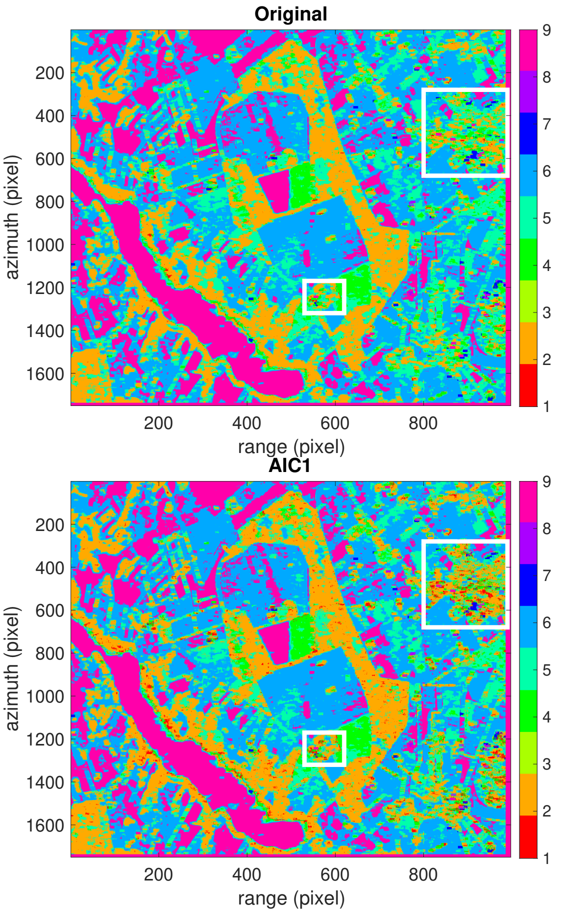


Figure 13: H/α classification. On the top: H/α classifier. On the bottom: EM H/α classifier.

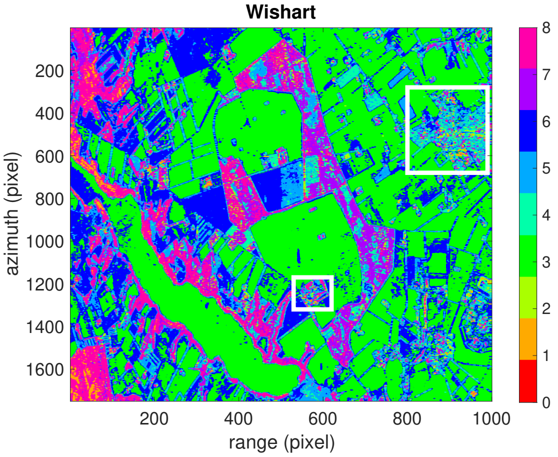


Figure 14: Wishart H/α classifier (ESA software).

It follows that

$$\max_{\boldsymbol{C}_1} \log\det(\boldsymbol{C}_1^{-1}) - \mathrm{Tr}\left[\boldsymbol{C}_1^{-1}\boldsymbol{S}_q(1,m)^{(h-1)}\right] \tag{41}$$

is tantamount to maximize

$$\log\det\left(\boldsymbol{C}_1^{-1}\boldsymbol{S}_q(1,m)^{(h-1)}\right) - \log\det\left(\boldsymbol{S}_q(1,m)^{(h-1)}\right) - \mathrm{Tr}\left[\boldsymbol{C}_1^{-1}\boldsymbol{S}_q(1,m)^{(h-1)}\right]. \tag{42}$$

The maximizer can be obtained resorting to the following inequality [29]

$$\log\det(\boldsymbol{A}) \le \mathrm{Tr}\,[\boldsymbol{A}] - N, \tag{43}$$

where $\boldsymbol{A}$ is any N-dimensional matrix with nonnegative eigenvalues, and, hence, we obtain

$$\begin{aligned}\widehat{\boldsymbol{C}}_1^{(h)} &= \arg\max_{\boldsymbol{C}_1} \sum_{k=1}^{K} q_k^{(h-1)}(1,m) \\ &\quad\times\left[-\log\det(\boldsymbol{C}_1) - \mathrm{Tr}\left(\boldsymbol{C}_1^{-1}\boldsymbol{z}_k\boldsymbol{z}_k^{\dagger}\right)\right] \\ &= \boldsymbol{S}_q(1,m)^{(h-1)}.\end{aligned}$$

The last equality concludes the proof.

APPENDIX B
DERIVATION OF (24) (UPDATE RULE FOR $\boldsymbol{C}_2$)

As first step, let us notice that

$$\boldsymbol{U}\boldsymbol{C}_2\boldsymbol{U}^{\dagger} = \begin{bmatrix}\boldsymbol{A} & \boldsymbol{0}\\ \boldsymbol{0} & d\end{bmatrix}, \tag{44}$$

where $\boldsymbol{A} \in \mathbb{C}^{2\times 2}$ is positive definite and $d > 0$. It follows that problem (22) can be recast as

$$\max_{\boldsymbol{A}}\max_{d>0} \sum_{k=1}^{K} q_k^{(h-1)}(2,m)\left[-\log\det(\boldsymbol{A}) - \log d - \boldsymbol{z}_{k,1}^{\dagger}\boldsymbol{A}^{-1}\boldsymbol{z}_{k,1} - |z_{k,2}|^2 d^{-1}\right] \tag{45}$$

where $\boldsymbol{U}\boldsymbol{z}_k = [\boldsymbol{z}_{k,1}^T \; z_{k,2}]^T$ with $\boldsymbol{z}_{k,1} \in \mathbb{C}^{2\times 1}$ and $z_{k,2} \in \mathbb{C}$.

In order to maximize (45) with respect to d, it is not difficult to show that

$$\lim_{\substack{d\to 0\\ d\to+\infty}} \left[-\sum_{k=1}^{K} q_k^{(h-1)}(2,m)(\log d + |z_{k,2}|^2 d^{-1})\right] = -\infty. \tag{46}$$

Thus, the stationary points over $d > 0$ can be found by setting to zero the first derivative with respect to d of the argument of (45), to obtain

$$-\sum_{k=1}^{K} \frac{q_k^{(h-1)}(2,m)}{d} + \frac{1}{d^2}\sum_{k=1}^{K} q_k^{(h-1)}(2,m)|z_{k,2}|^2 = 0. \tag{47}$$

Thus, the update of the estimate of d is given by the first row of (25). As for $\boldsymbol{A}$, let us consider

$$\max_{\boldsymbol{A}} \sum_{k=1}^{K} q_k^{(h-1)}(2,m)\left[-\log\det(\boldsymbol{A}) - \boldsymbol{z}_{k,1}^{\dagger}\boldsymbol{A}^{-1}\boldsymbol{z}_{k,1}\right], \tag{48}$$

which can be recast as

$$\begin{aligned}&\max_{\boldsymbol{A}} q^{(h-1)}(2,m)\log\det(\boldsymbol{A}^{-1}) - \text{Tr}\left[\boldsymbol{A}^{-1}\boldsymbol{S}(2,m)^{(h-1)}\right],\\ &\Rightarrow \max_{\boldsymbol{A}} \log\det\left[\boldsymbol{A}^{-1}\boldsymbol{S}(2,m)^{(h-1)}/q^{(h-1)}(2,m)\right]\\ &\quad - \text{Tr}\left[\boldsymbol{A}^{-1}\boldsymbol{S}(2,m)^{(h-1)}/q^{(h-1)}(2,m)\right], \qquad (49)\end{aligned}$$

where

$$\begin{aligned}q^{(h-1)}(2,m) &= \sum_{k=1}^{K} q_k^{(h-1)}(2,m),\\ \boldsymbol{S}(2,m)^{(h-1)} &= \sum_{k=1}^{K} q_k^{(h-1)}(2,m)\boldsymbol{z}_{k,1}\boldsymbol{z}_{k,1}^{\dagger}.\end{aligned}$$

Exploiting (43), it turns out that

$$\widehat{\boldsymbol{A}}^{(h)} = \boldsymbol{S}(2,m)^{(h-1)}/q^{(h-1)}(2,m). \qquad (50)$$

Thus, (24) naturally follows and the proof is complete.

APPENDIX C
DERIVATION OF (26) (UPDATE RULE FOR $\boldsymbol{C}_3$)

Matrix $\boldsymbol{C}_3$ can be suitably manipulated by applying the transformations $\boldsymbol{E}$, $\boldsymbol{T}$, and $\boldsymbol{V}$ (see [13, *Lemma 3.1*]) as follows

$$\boldsymbol{V}\boldsymbol{E}\boldsymbol{T}\boldsymbol{C}_3\boldsymbol{T}^{\dagger}\boldsymbol{E}\boldsymbol{V}^{\dagger} = \begin{bmatrix} a & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{B}\end{bmatrix}, \qquad (51)$$

where $a > 0$ and $\boldsymbol{B} \in \mathbb{R}^{2\times 2}$ is centrosymmetric.[11] Then, the objective function can be accordingly expressed as

$$\begin{aligned}&\max_{a}\max_{\boldsymbol{B}} \sum_{k=1}^{K} q_k^{(h-1)}(3,m)\left[-\log a - \log\det(\boldsymbol{B})\right.\\ &\qquad\qquad \left.-\frac{|x_{k,1}|^2}{a} - \boldsymbol{x}_{k,2}^{\dagger}\boldsymbol{B}^{-1}\boldsymbol{x}_{k,2}\right], \qquad (52)\end{aligned}$$

where $\boldsymbol{V}\boldsymbol{E}\boldsymbol{T}\boldsymbol{z}_k = [x_{k,1}\ \boldsymbol{x}_{k,2}^T]^T$ with $x_{k,1} \in \mathbb{C}$ and $\boldsymbol{x}_{k,2} \in \mathbb{C}^{2\times 1}$. Since $\boldsymbol{B}$ is centrosymmetric, the following equality

$$\boldsymbol{B}^{-1} = \frac{(\boldsymbol{B}^{-1} + \boldsymbol{J}\boldsymbol{B}^{-1}\boldsymbol{J})}{2} \qquad (53)$$

holds. As a result, (52) can be written as

$$\begin{aligned}&\max_{a}\max_{\boldsymbol{B}} \sum_{k=1}^{K} q_k^{(h-1)}(3,m)\left[-\log a - \log\det(\boldsymbol{B})\right.\\ &\left.-\frac{|x_{k,1}|^2}{a} - \frac{1}{2}\text{Tr}\left[\boldsymbol{B}^{-1}\left(\boldsymbol{x}_{k,2}\boldsymbol{x}_{k,2}^{\dagger} + \boldsymbol{J}\boldsymbol{x}_{k,2}\boldsymbol{x}_{k,2}^{\dagger}\boldsymbol{J}\right)\right]\right]. \qquad (54)\end{aligned}$$

Exploiting the same line of reasoning as for the estimation of d and $\boldsymbol{A}$ in Appendix B, it is possible to show that the estimate of a is given by the first row of (27), whereas, using (43), the estimate of $\boldsymbol{B}$ can be written as the second row of (27). Gathering the above results, (26) follows and we conclude the proof.

APPENDIX D
DERIVATION OF (28) (UPDATE RULE FOR $\boldsymbol{C}_4$)

Exploiting equation (8) of [13, *Lemma 3.1*], $\boldsymbol{C}_4$ can be transformed as follows

$$\mathbf{E}\mathbf{T}\boldsymbol{C}_4\mathbf{T}^{\dagger}\mathbf{E} = \begin{bmatrix} b & 0 & 0\\ 0 & c & 0\\ 0 & 0 & c\end{bmatrix} \in \mathbb{R}^{3\times 3}, \qquad (55)$$

where $b > 0$ and $c > 0$. As a consequence, the optimization problem to be solved becomes

$$\begin{aligned}&\max_{b>0}\max_{c>0} \sum_{k=1}^{K} q_k^{(h-1)}(4,m)\left[-\log b - 2\log c\right.\\ &\qquad\qquad \left.-\frac{|y_{k,1}|^2}{b} - \frac{1}{c}\boldsymbol{y}_{k,2}^{\dagger}\boldsymbol{y}_{k,2}\right], \qquad (56)\end{aligned}$$

where $\boldsymbol{E}\boldsymbol{T}\boldsymbol{z}_k = [y_{k,1}\ \boldsymbol{y}_{k,2}^T]^T$ with $y_{k,1} \in \mathbb{C}$ and $\boldsymbol{y}_{k,2} \in \mathbb{C}^{2\times 1}$. Now, observe that

$$\lim_{\substack{b\to 0\\ b\to+\infty}} \left\{-\sum_{k=1}^{K} q_k^{(h-1)}(4,m)\left[\log b + \frac{|y_{k,1}|^2}{b}\right]\right\} = -\infty, \qquad (57)$$

$$\lim_{\substack{c\to 0\\ c\to+\infty}} \left\{-\sum_{k=1}^{K} q_k^{(h-1)}(4,m)\left[2\log c + \frac{1}{c}\boldsymbol{y}_{k,2}^{\dagger}\boldsymbol{y}_{k,2}\right]\right\} = -\infty, \qquad (58)$$

and, hence, setting to zero the first derivatives of the above functions with respect to b and c, respectively, it is not difficult to obtain (29) and the final estimate of $\boldsymbol{C}_4$ defined by (28).

REFERENCES

[1] F. Biondi, P. Addabbo, C. Clemente, and D. Orlando, "Measurements of Surface River Doppler Velocities With Along-Track InSAR Using a Single Antenna," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 987–997, 2020.

[2] F. Biondi, A. Tarpanelli, P. Addabbo, C. Clemente, and D. Orlando, "Water Level measurement using COSMO-SkyMed Synthetic Aperture Radar," in *2020 IEEE 7th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, 2020, pp. 148–153.

[3] J. Zheng, T. Su, L. Zhang, W. Zhu, and Q. H. Liu, "ISAR Imaging of Targets With Complex Motion Based on the Chirp Rate–Quadratic Chirp Rate Distribution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 7276–7289, 2014.

[4] F. Bandiera, A. Masciullo, and G. Ricci, "A Bayesian Approach to Oil Slicks Edge Detection Based on SAR Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2901–2909, 2014.

[5] A. Salberg, O. Rudjord, and A. H. S. Solberg, "Oil Spill Detection in Hybrid-Polarimetric SAR Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 10, pp. 6521–6533, 2014.

[6] S. Mermoz, S. Allain-Bailhache, M. Bernier, E. Pottier, J. J. Van Der Sanden, and K. Chokmani, "Retrieval of River Ice Thickness From C-Band PolSAR Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 6, pp. 3052–3062, 2014.

[7] L. M. Novak, S. D. Halversen, G. Owirka, and M. Hiett, "Effects of polarization and resolution on SAR ATR," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 1, pp. 102–116, 1997.

[8] D. Gaglione, C. Clemente, L. Pallotta, I. Proudler, A. De Maio, and J. J. Soraghan, "Krogager decomposition and Pseudo-Zernike moments for polarimetric distributed ATR," in *2014 Sensor Signal Processing for Defence (SSPD)*, 2014, pp. 1–5.

[9] F. Garestier, P. Dubois-Fernandez, X. Dupuis, P. Paillou, and I. Hajnsek, "PolInSAR analysis of X-band data over vegetated and urban areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 2, pp. 356–364, 2006.

---

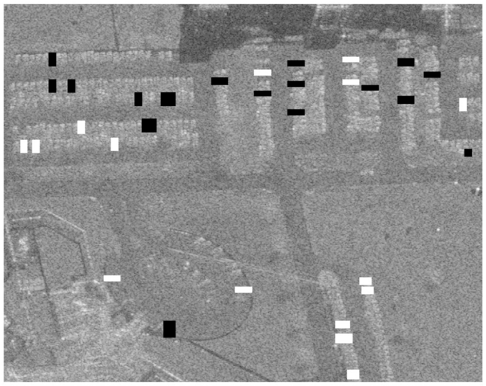
[11] $\boldsymbol{B}$ is such that $\boldsymbol{B} = \boldsymbol{J}\boldsymbol{B}\boldsymbol{J}$, where

$$\boldsymbol{J} = \begin{bmatrix} 0 & 1\\ 1 & 0\end{bmatrix}.$$

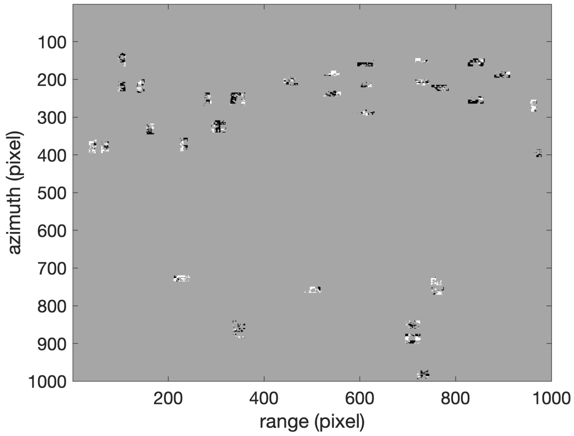
(a) Reference image.

(b) Test image.

Figure 15: Reference and test images gathered in HH mode.



(a) Ground Truth



(b) Symmetry Transition Map

Figure 16: Ground truth and map of the detected PCM symmetries, on the ground truth white boxes represent vehicle departures, while black boxes represent vehicle arrivals, on the map of detect PCM symmetries, black is used to represent transitions from no dominant symmetry to reflection symmetry, while white color is used to highlight a transition from reflection symmetry to no dominant symmetry.

[10] A. Lönnqvist, Y. Rauste, M. Molinier, and T. Häme, "Polarimetric SAR Data in Land Cover Mapping in Boreal Zone," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3652–3662, 2010.

[11] D. Xiang, Y. Ban, W. Wang, and Y. Su, "Adaptive Superpixel Generation for Polarimetric SAR Images With Local Iterative Clustering and SIRV Model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3115–3131, 2017.

[12] S. V. Nghiem, S. H. Yueh, R. Kwok, and F. K. Li, "Symmetry properties in polarimetric remote sensing," *Radio Science*, vol. 27, no. 5, pp. 693–711, 1992.

[13] L. Pallotta, C. Clemente, A. De Maio, and J. J. Soraghan, "Detecting Covariance Symmetries in Polarimetric SAR Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 1, pp. 80–95, 2017.

[14] L. Pallotta, A. De Maio, and D. Orlando, "A Robust Framework for Covariance Classification in Heterogeneous Polarimetric SAR Images and Its Application to L-Band Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 104–119, 2019.

[15] J. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*, ser. Optical Science and Engineering. CRC Press, 2017.

[16] P. Addabbo, S. Han, F. Biondi, G. Giunta, and D. Orlando, "Adaptive Radar Detection in the Presence of Multiple Alternative Hypotheses Using Kullback-Leibler Information Criterion-Part I: Detector Designs," *IEEE Transactions on Signal Processing*, vol. 69, pp. 3730–3741, 2021.

[17] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society (Series B - Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[18] P. Addabbo, S. Han, D. Orlando, and G. Ricci, "Learning Strategies for Radar Clutter Classification," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1070–1082, 2021.

[19] A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Classification With a Non-Gaussian Model for PolSAR Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 10, pp. 2999–3009, 2008.

[20] Y. Zhou, A. Rangarajan, and P. D. Gader, "A Gaussian Mixture Model Representation of Endmember Variability in Hyperspectral Unmixing," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2242–2256, 2018.

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

[22] S. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 1, pp. 68–78, 1997.

[23] L. Ferro-Famil, E. Pottier, and J.-S. Lee, "Unsupervised classification of multifrequency and fully polarimetric SAR images based on the H/A/Alpha-Wishart classifier," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 11, pp. 2332–2342, 2001.

[24] H. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*, ser. Detection, Estimation, and Modulation Theory. Wiley, 2004.

[25] K. Murphy, *Machine Learning: A Probabilistic Perspective*, ser. Adaptive Computation and Machine Learning series. MIT Press, 2012.

[26] "U.S. Air Force Sensor Data Management System, Coherent Change Detection Challenge Problem." [Online]. Available: https://www.sdms.afrl.af.mil/index.php?collection=ccd_challenge

[27] S. Scarborough, L. Gorham, M. J. Minardi, U. K. Majumdar, M. G.

Judge, L. Moore, L. Novak, S. Jaroszewski, L. Spoldi, and A. Pieramico, "Water Level measurement using COSMO-SkyMed Synthetic Aperture Radar, 76990U," in *SPIE 7699, Algorithms for Synthetic Aperture Radar Imagery XVII*, 2010.

[28] V. Carotenuto, A. De Maio, C. Clemente, J. J. Soraghan, and G. Alfano, "Forcing scale invariance in multipolarization SAR change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 36–50, January 2016. [Online]. Available: https://doi.org/10.1109/TGRS.2015.2449332

[29] H. Lütkepohl, *Handbook of Matrices*, ser. Handbook of Matrices. Wiley, 1996.