Learning Strategies for Radar Clutter Classification

Pia Addabbo, *Senior Member, IEEE*, Sudan Han,
Danilo Orlando, *Senior Member, IEEE*, and Giuseppe Ricci

***Abstract*—In this paper, we address the problem of classifying clutter returns into statistically homogeneous subsets. The classification procedures are devised assuming latent variables, which represent the classes to which each range bin belongs, and three different models for the structure of the clutter covariance matrix. Then, the expectation-maximization algorithm is exploited in conjunction with cyclic estimation procedures to come up with suitable estimates of the unknown parameters. Finally, the classification is performed by maximizing the posterior probability that a range bin belongs to a specific class. The performance analysis of the proposed classifiers is conducted over synthetic data as well as real recorded data and highlights that they represent a viable means to cluster clutter returns with respect to their range.**

***Index Terms*—Clutter, Expectation-Maximization, Heterogeneous Environment, Interference Classification, Radar.**

I. Introduction

In the past ten years, improvements in digital architectures and miniaturization technologies have wielded a significant impact in the evolution of radar systems which, consequently, are being equipped with more and more reliable and sophisticated functions [1], [2]. This increase in computational resources has led the radar community to devise detection/estimation algorithms capable of facing with challenging scenarios and, more importantly, of capitalizing on specific *a priori* knowledge about either the system or the environment or both. In this context, a few examples related to the structural information about the interference covariance matrix are provided by [3]–[9], where, at the design stage, it is assumed that the system illuminates the surveillance area through a symmetrically spaced linear array of sensors. This assumption lends both the interference covariance matrix and the steering vector a special structure which yields interesting processing gains at the price of an additional computational load [10], [11].

Other approaches relying on a priori information exploit the possible symmetries in the interference spectral properties [5], [12], [13]. As a matter of fact, ground clutter returns collected by a monostatic steady radar experience a symmetric power spectral density centered around zero-Doppler frequency [14], [15]. Remarkably, such property allows to improve the estimation quality of the covariance matrix since, in this case, the real and imaginary parts of data are statistically independent and, hence, can be used to conceive enhanced estimation procedures as shown in [13]. Therefore, the above knowledge-based strategies represent an effective means to deal with situations where the amount of training data, used for the estimation of the interference covariance matrix, is limited (sample-starved condition) otherwise leading to low-quality estimates and, consequently, to a detection performance degradation. Besides the mentioned approaches, other widely used techniques to come up with suitable estimates of the interference covariance matrix consist in the regularization (or shrinkage) of the sample covariance matrix towards a given matrix [16]–[18].

However, in practice, it is not uncommon to meet situations where the presence of inhomogeneities and jammers makes the estimation of the interference properties an even more difficult task [19]–[25]. In fact, the homogeneity assumption for training or secondary data, namely data in spatial proximity to the cell under test to be used to estimate the covariance matrix of the disturbance, is very common in detector design [26]–[29, and references therein] and when it is no longer valid the performance degradation might become severe [30]. A more complete approach to the problem of generating homogeneous training sets would envisage an additional architectural layout capable of integrating and fusing information coming from potential heterogeneous sources to depict a clear picture of the clutter properties. These sources can be internal or external to the system and comprise mapping data, communication links, tracker feedback, or other inputs [31]–[34].

Now, note that environment maps might be useful to identify clutter edges and to cluster data into homogeneous subsets. In addition, for each subset, it is also possible to take advantage of a priori information to further increase the estimation quality of the interference covariance matrix as stated before. Thus, classifying (or, otherwise stated, clustering) clutter returns would represent a desirable feature for modern radar systems. Examples of clutter classifiers are provided by [35], [36], where the authors build up a neural network or process suitable features to distinguish between echoes from weather, birds, and aircrafts. Other classifiers are aimed at identifying the distribution for clutter data [37]–[39], the specific structure of the clutter covariance matrix [40], or the variability of clutter power over the range bins [41]. Supervised and unsupervised machine learning algorithms are proposed in [42]–[44], where an autoregressive model is assumed for the signal leading to a clutter covariance matrix with a Toeplitz/block-Toeplitz structure. In this context, a suitable metric is defined over the manifold of Hermitian Positive Definite Toeplitz/Block-Toeplitz matrices and incorporated into supervised or unsupervised clustering algorithms (k-nearest neighbors, k-means,

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

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

Danilo Orlando is with the Engineering Faculty of Università degli Studi "Niccolò Cusano", via Don Carlo Gnocchi 3, 00166 Roma, Italy. E-mail: danilo.orlando@unicusano.it.

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

mean-shift, agglomerative hierarchical clustering, etc.) that in turn are used to partition data into homogeneous subset based on an estimate of the interference covariance matrix of each range bin.

In this paper, we focus on the problem of partitioning training data into homogeneous subsets and we assume that only partial information about the environment is available at the radar receiver, namely that a given number of clutter boundaries is present. Then, we design a classification procedure capable of partitioning the secondary data set into subsets containing statistically homogeneous data. To this end, we jointly exploit the expectation-maximization (EM) algorithm [45] and the latent variable model [46]. The latter tool allows us to introduce hidden random variables which represent the classes, namely, uniform clutter regions, to which each range cell belongs. Thus, at the end of the procedure, the clustering is accomplished by estimating the *a posteriori* probability that a range bin belongs to a specific class. More importantly, at the design stage, we consider the following three models for the covariance matrix of the disturbance

- the disturbance of each class is characterized by its own Hermitian covariance matrix;
- different classes share a common (general Hermitian) structure of the covariance matrix, but they have different scaling factors representative of different power values (clutter-dominated environment); it is worth noticing that such model has been widely adopted to face with situations where the Gaussian assumption is not valid (namely, in the case of low grazing angles or high-resolution radars) and the interference contribution is better described by the so-called compound-Gaussian model resulting from the product of a correlated Gaussian process multiplied by a non-negative slowly-varying component [47]–[50];
- noise returns consist of a thermal noise component (whose power is independent of the class) plus a clutter component; the (general Hermitian) clutter covariance matrices are possibly different and their ranks are either known or unknown."

The preliminary performance analysis shows the effectiveness of the proposed methods in clustering data.

The remainder of the paper is organized as follows. The next section contains the problem formulation, whereas Section III is devoted to the design of the classification architectures. Illustrative examples and discussion about the classification performance are provided in Section IV. Finally, in Section V, we draw the conclusions and lay down possible future research lines. Derivations are confined to the Appendices.

A. Notation

In the sequel, vectors and matrices are denoted by boldface lower-case and upper-case letters, respectively. The (i,j)th entry of a matrix $\boldsymbol{A}$ is indicated by $\boldsymbol{A}(i,j)$. Symbols $\det(\cdot)$, $\mathrm{Tr}\,(\cdot)$, $(\cdot)^T$, and $(\cdot)^\dagger$ denote the determinant, trace, transpose, and conjugate transpose, respectively. As to numerical sets, $\mathbb{N}$ is the set of natural numbers, $\mathbb{R}$ 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 $M=1$). $\boldsymbol{I}$ and $\boldsymbol{0}$ stand for the identity matrix and the null vector or matrix of proper size. Given $a_1,\ldots,a_N\in\mathbb{C}$, $\mathrm{diag}\,(a_1,\ldots,a_N)\in\mathbb{C}^{N\times N}$ indicates the diagonal matrix whose ith diagonal element is a_i. The acronym pdf and pmf stand for probability density function and probability mass function, respectively, whereas the conditional pdf of a random variable x given another random variable y is denoted by $f(x|y)$. 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 $\boldsymbol{M}$ while given a matrix $\boldsymbol{X}=[\boldsymbol{x}_1\cdots\boldsymbol{x}_M]\in\mathbb{C}^{N\times M}$, writing $\boldsymbol{X}\sim\mathcal{CN}_N(\boldsymbol{m},\boldsymbol{M},\boldsymbol{I})$ means that $\boldsymbol{x}_i\sim\mathcal{CN}_N(\boldsymbol{m},\boldsymbol{M})$, $i=1,\ldots,M$, and the $\boldsymbol{x}_i$s are statistically independent.

II. Problem Formulation and Preliminary Definitions

Consider a radar system equipped with $N\geq 2$ space, time, or space-time channels which illuminates the operating area consisting of K range bins. The signals backscattered by these range cells are suitably conditioned and sampled by the signal-processing unit to form N-dimensional complex vectors denoted by $\boldsymbol{z}_1,\ldots,\boldsymbol{z}_K$. Now, let us assume that the statistical properties of the returns may change over the range due, for instance, to the presence of clutter boundaries [51]. Moreover, the number of edges and, hence, of subsets formed by statistically homogeneous data, L say, is assumed known. For instance, a rough estimate of this parameter can be obtained from the knowledge of the terrain types that compose the region of interest. Otherwise stated, we assume that the set of vectors can be partitioned into L subsets of statistically homogeneous data; the lth subset is denoted by $\Omega_l=\{\boldsymbol{z}_{i_{l,1}},\ldots,\boldsymbol{z}_{i_{l,K_l}}\}$, where K_l, $l=1,\ldots,L$, denotes its cardinality. Thus, the elements of Ω_l share the same distributional parameters which are generally different from those associated to the elements of Ω_m, $m\neq l$. Specifically, we assume that $[\boldsymbol{z}_{i_{l,1}}\cdots\boldsymbol{z}_{i_{l,K_l}}]\sim\mathcal{CN}_N(\boldsymbol{0},\boldsymbol{M}_l,\boldsymbol{I})$, $l=1,\ldots,L$, where $\boldsymbol{M}_l$ is unknown.

Summarizing, we are interested in estimating the subsets Ω_l along with the associated unknown parameter $\boldsymbol{M}_l$, $l=1,\ldots,L$. To this end, in the next section we devise a classification procedure relying on the joint exploitation of the EM algorithm [45] and the latent variable model [46]. Moreover, besides the most general structure for the clutter covariance matrix, we consider two additional models which account for possible clutter power variations and thermal noise.

III. Classification Architecture Designs

Data classification task is accomplished by introducing K independent and identically distributed discrete random variables, c_ks say, which take on values in $\{1,\ldots,L\}$ with unknown pmf $P(c_k=l)=p_l$, $k=1,\ldots,K$, and[1] such that

[1]Recall that $\sum_{l=1}^{L}p_l=1$.

when $c_k = l$, then $\boldsymbol{z}_k \sim \mathcal{CN}_N(\boldsymbol{0}, \boldsymbol{M}_l)$. Under this assumption, it naturally follows that the pdf of $\boldsymbol{z}_k$ can be written as

$$f(\boldsymbol{z}_k;\boldsymbol{\theta}) = \sum_{l=1}^{L} p_l f(\boldsymbol{z}_k|c_k = l;\boldsymbol{M}_l) = E_{c_k}[f(\boldsymbol{z}_k|c_k;\boldsymbol{\theta}))],$$

where $E_{c_k}[\cdot]$ denotes the statistical expectation with respect to c_k, $\boldsymbol{\theta} = \left[\boldsymbol{p}^T, \boldsymbol{\sigma}^T\right]^T$, $\boldsymbol{p} = [p_1 \cdots p_L]^T$, $\boldsymbol{\sigma} = \left[\boldsymbol{\nu}^T(\boldsymbol{M}_1) \cdots \boldsymbol{\nu}^T(\boldsymbol{M}_L)\right]^T$, $\boldsymbol{\nu}(\cdot)$ a vector-valued function selecting the generally distinct entries of the matrix argument, and

$$f(\boldsymbol{z}_k|c_k = l;\boldsymbol{M}_l) = \frac{1}{\pi^N \det(\boldsymbol{M}_l)} \exp\{-\mathrm{Tr}\left[\boldsymbol{M}_l^{-1}\boldsymbol{z}_k\boldsymbol{z}_k^\dagger\right]\}.$$

Now, obtaining possible closed-form maximum likelihood estimates of the unknown parameters, namely $\boldsymbol{M}_1, \ldots, \boldsymbol{M}_L$ and $\boldsymbol{p}$, is not an easy task (at least to the best of authors' knowledge). For this reason, we resort to EM-based algorithms [45] that provide closed-form updates for the parameter estimates at each step and reach at least a local stationary point. To this end, let us write the joint log-likelihood of $\boldsymbol{Z} = [\boldsymbol{z}_1 \cdots \boldsymbol{z}_K]$ as follows

$$\begin{aligned}\mathcal{L}(\boldsymbol{Z};\boldsymbol{\theta}) &= \sum_{k=1}^{K} \log \sum_{c_k=1}^{L} f(z_k, c_k;\boldsymbol{\theta}) \\ &= \sum_{k=1}^{K} \log \sum_{l=1}^{L} p_l f(z_k|c_k = l;\boldsymbol{M}_l). \end{aligned} \tag{1}$$

As observed before, the EM algorithm is a recursive approach to the estimation of the parameter $\boldsymbol{\theta}$: its hth iteration is aimed at computing $\widehat{\boldsymbol{\theta}}^{(h)}$ starting from the estimate at the previous iteration, $\widehat{\boldsymbol{\theta}}^{(h-1)}$ say, to form a nondecreasing sequence of log-likelihood values, namely $\mathcal{L}(\boldsymbol{Z};\widehat{\boldsymbol{\theta}}^{(h)}) \geq \mathcal{L}(\boldsymbol{Z};\widehat{\boldsymbol{\theta}}^{(h-1)})$. Obviously, an initial estimate of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}^{(0)}$ say, is necessary to initialize the algorithm as well as a reasonable stopping criterion as, for instance, a maximum number of iterations, $h_{\max}$ say. The EM consists of two steps referred to as the E-step and the M-step, respectively. The E-step leads to the computation of the following quantity

$$\begin{aligned} q_k^{(h-1)}(l) &= p(c_k = l|\boldsymbol{z}_k;\widehat{\boldsymbol{\theta}}^{(h-1)}) \\ &= \frac{f(\boldsymbol{z}_k|c_k = l;\widehat{\boldsymbol{M}}_l^{(h-1)})\widehat{p}_l^{(h-1)}}{f(\boldsymbol{z}_k;\widehat{\boldsymbol{\theta}}^{(h-1)})} \\ &= \frac{f(\boldsymbol{z}_k|c_k = l;\widehat{\boldsymbol{M}}_l^{(h-1)})\widehat{p}_l^{(h-1)}}{\sum_{l'=1}^{L} f(\boldsymbol{z}_k|c_k = l';\widehat{\boldsymbol{M}}_{l'}^{(h-1)})\widehat{p}_{l'}^{(h-1)}}, \end{aligned} \tag{2}$$

whereas the M-step requires to solve the following problem

$$\begin{aligned} \widehat{\boldsymbol{\theta}}^{(h)} &= \arg\max_{\boldsymbol{\theta}} \sum_{k=1}^{K}\sum_{l=1}^{L} q_k^{(h-1)}(l) \log \frac{f(\boldsymbol{z}_k|c_k = l;\boldsymbol{M}_l)p_l}{q_k^{(h-1)}(l)} \\ \Rightarrow \widehat{\boldsymbol{\theta}}^{(h)} &= \arg\max_{\boldsymbol{\theta}} \left\{ \sum_{k=1}^{K}\sum_{l=1}^{L} q_k^{(h-1)}(l) \log f(\boldsymbol{z}_k|c_k = l;\boldsymbol{M}_l) \right. \\ &\quad \left. + \sum_{k=1}^{K}\sum_{l=1}^{L} q_k^{(h-1)}(l) \log p_l \right\}. \end{aligned} \tag{3}$$

Note that the maximization with respect to p_l, $l = 1, \ldots, L$, is independent of that over $\boldsymbol{M}_l$, $l = 1, \ldots, L$, and, hence, we can proceed by separately addressing these two problems. Starting from the optimization over $\boldsymbol{p}$, observe that it can be solved by using the method of Lagrange multipliers, to take into account the constraint $\sum_{l=1}^{L} p_l = 1$.

Thus, it is not difficult to show that

$$\widehat{p}_l^{(h)} = \frac{1}{K}\sum_{k=1}^{K} q_k^{(h-1)}(l). \tag{4}$$

Finally, in order to come up with the estimates of $\boldsymbol{M}_1, \ldots, \boldsymbol{M}_L$, we solve the following problem

$$\widehat{\boldsymbol{\sigma}}^{(h)} = \arg\max_{\boldsymbol{\sigma}} \sum_{k=1}^{K}\sum_{l=1}^{L} q_k^{(h-1)}(l) \log f(\boldsymbol{z}_k|c_k = l;\boldsymbol{M}_l), \tag{5}$$

where three different forms for the $\boldsymbol{M}_l$, $l = 1, \ldots, L$, are considered, namely

1) $\boldsymbol{M}_l$ is a positive definite Hermitian matrix;
2) $\boldsymbol{M}_l = \sigma_{c,l}^2\boldsymbol{M}$, where $\sigma_{c,l}^2 > 0$ represents the clutter power which might vary over the range profile when a clutter edge occurs, while $\boldsymbol{M}$ is the common structure shared by the interference of the K range bins;
3) $\boldsymbol{M}_l = \sigma_n^2\boldsymbol{I} + \boldsymbol{R}_l$, where $\sigma_n^2 > 0$ is the unknown thermal noise power and $\boldsymbol{R}_l \in \mathbb{C}^{N\times N}$ denotes the clutter contribution to the interference of the lth range bin whose rank, r_l say, is assumed for the moment known.

Then, the estimates of the unknown parameters for the above cases are provided by the following propositions.

Proposition 1. *Assume that $K \geq N$, then an approximation to the relative maximum point of*

$$g_1(\boldsymbol{M}_1, \ldots, \boldsymbol{M}_L) = \sum_{k=1}^{K}\sum_{l=1}^{L} q_k^{(h-1)}(l) \log f(\boldsymbol{z}_k|c_k = l;\boldsymbol{M}_l) \tag{6}$$

has the following expression

$$\widehat{\boldsymbol{M}}_l^{(h)} = \frac{\sum_{k=1}^{K} q_k^{(h-1)}(l)\boldsymbol{z}_k\boldsymbol{z}_k^\dagger}{\sum_{k=1}^{K} q_k^{(h-1)}(l)}, \quad l = 1, \ldots, L. \tag{7}$$

Proof. See Appendix A. □

Proposition 2. *Assume that $K \geq N$ and form 2 for $\boldsymbol{M}_l$, then, given the function*

$$g_2(\boldsymbol{\sigma}_c^2, \boldsymbol{M}) = \sum_{k=1}^{K}\sum_{l=1}^{L} q_k^{(h-1)}(l) \log f(\boldsymbol{z}_k|c_k = l;\sigma_{c,l}^2\boldsymbol{M}) \tag{8}$$

where $\boldsymbol{\sigma}_c^2 = [\sigma_{c,1}^2 \cdots \sigma_{c,L}^2]^T$, an approximation to the relative maximum point can be achieved by means of the following cyclic procedure with respect to the iteration index t, $t = 1, \ldots, t_{\max}$, (with $t_{\max}$ a proper design parameter)

$$(\widehat{\sigma}_{c,l}^2)^{(1),(h)} = \frac{\sum_{k=1}^{K} q_k^{(h-1)}(l)\boldsymbol{z}_k^\dagger(\widehat{\boldsymbol{M}}^{(t_{\max}),(h-1)})^{-1}\boldsymbol{z}_k}{N\sum_{k=1}^{K} q_k^{(h-1)}(l)}, \tag{9}$$

$$\widehat{\boldsymbol{M}}^{(t),(h)} = \frac{1}{K}\sum_{k=1}^{K}\sum_{l=1}^{L} q_k^{(h-1)}(l)\frac{\boldsymbol{z}_k\boldsymbol{z}_k^{\dagger}}{(\widehat{\sigma}_{c,l}^2)^{(t),(h)}}, \quad (10)$$

$t = 1, \ldots, t_{\max}$, *and*

$$(\widehat{\sigma}_{c,l}^2)^{(t),(h)} = \frac{\sum_{k=1}^{K} q_k^{(h-1)}(l)\boldsymbol{z}_k^{\dagger}(\widehat{\boldsymbol{M}}^{(t-1),(h)})^{-1}\boldsymbol{z}_k}{N\sum_{k=1}^{K} q_k^{(h-1)}(l)}, \quad (11)$$

$t = 2, \ldots, t_{\max}$, $l = 1, \ldots, L$.

Proof. See Appendix B. □

Proposition 3. *Assume form* 3 *for* $\boldsymbol{M}_l$, $l = 1, \ldots, L$, *and that* $r_l < N$, $l = 1, \ldots, L$, *is known, then an approximation to the relative maximum point of the function*

$$g_3(\sigma_n^2, \boldsymbol{R}_1, \ldots, \boldsymbol{R}_L) = \sum_{k=1}^{K}\sum_{l=1}^{L} q_k^{(h-1)}(l)\log f(\boldsymbol{z}_k|c_k = l; \sigma_n^2\boldsymbol{I} + \boldsymbol{R}_l), \quad (12)$$

can be obtained as follows

$$\widehat{\sigma}_n^{2(h)} = \left\{\sum_{l=1}^{L}\sum_{n=r_l+1}^{N}\gamma_{l,n}^{(h-1)}\right\} \Big/ \left\{\sum_{l=1}^{L}\sum_{k=1}^{K} q_k^{(h-1)}(l)(N - r_l)\right\}, \quad (13)$$

$$\widehat{\boldsymbol{R}}_l^{(h)}(r_l) = \widehat{\boldsymbol{U}}_l^{(h)}\widehat{\boldsymbol{\Lambda}}_l^{(h)}(r_l)(\widehat{\boldsymbol{U}}_l^{(h)})^{\dagger}, \quad (14)$$

where $\widehat{\boldsymbol{U}}_l^{(h)}$ *is the unitary matrix whose columns are the eigenvectors corresponding to the eigenvalues* $\gamma_{l,1}^{(h-1)} \geq \gamma_{l,2}^{(h-1)} \geq \ldots \geq \gamma_{l,N}^{(h-1)}$ *of the matrix* $\boldsymbol{S}_l^{(h-1)} = \sum_{k=1}^{K} q_k^{(h-1)}(l)\boldsymbol{z}_k\boldsymbol{z}_k^{\dagger}$ *and*

$$\widehat{\boldsymbol{\Lambda}}_l^{(h)} = diag\left(\max\left\{\frac{\gamma_{l,1}^{(h-1)}}{\sum_{k=1}^{K} q_k^{(h-1)}(l)} - \widehat{\sigma}_n^{2(h)}, 0\right\}, \ldots, \max\left\{\frac{\gamma_{l,r_l}^{(h-1)}}{\sum_{k=1}^{K} q_k^{(h-1)}(l)} - \widehat{\sigma}_n^{2(h)}, 0\right\}, 0, \ldots, 0\right). \quad (15)$$

Proof. See Appendix C. □

Note that the last proposition supposes that r_l, $l = 1, \ldots, L$, is known. However, it is clear that such assumption does not exhibit a practical value; however, the results provided by Proposition 3 can suitably be exploited in conjunction with an estimator of $\boldsymbol{r} = [r_1, \ldots, r_L]^T$. To this end, we follow the lead of [52] and exploit model-order selection (MOS) rules [53], as the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the generalized information criterion (GIC), to build up the following estimator for[2] $\boldsymbol{r}$

$$\widehat{\boldsymbol{r}} = \arg\min_{\boldsymbol{r}}\left\{2\sum_{l=1}^{L}\sum_{m=1}^{r_l}\log\left(\frac{\gamma_{l,m}^{(h)}}{\sum_{k=1}^{K} q_k^{(h)}(l)}\right)\sum_{k=1}^{K} q_k^{(h)}(l) + 2\sum_{l=1}^{L}(N - r_l)\log\left[(\widehat{\sigma}_n^2)^{(h)}\right]\sum_{k=1}^{K} q_k^{(h)}(l) + 2\sum_{l=1}^{L} r_l\sum_{k=1}^{K} q_k^{(h)}(l) + \frac{2}{(\widehat{\sigma}_n^2)^{(h)}}\sum_{l=1}^{L}\sum_{m=r_l+1}^{N}\gamma_{l,m}^{(h)} + \xi(\boldsymbol{r})\right\}, \quad (16)$$

where $\xi(\boldsymbol{r})$ is a penalty term related to the number of unknown parameters and has the following expression $\xi(\boldsymbol{r}) = \left(1 + L + \sum_{l=1}^{L} r_l(2N - r_l)\right)k_p$ with [53]

$$k_p = \begin{cases} 2, & \text{AIC}, \\ 1 + a,\ a \geq 1 & \text{GIC}, \\ \log(2KN), & \text{BIC}. \end{cases} \quad (17)$$

Once the unknown quantities have been estimated, data classification can be accomplished by exploiting the following rule

$$\forall k = 1, \ldots, K : \boldsymbol{z}_k \sim \mathcal{CN}_N(\boldsymbol{0}, \widehat{\boldsymbol{M}}_{\hat{l}_k}) \quad (18)$$

where

$$\widehat{l}_k = \arg\max_{l=1,\ldots,L} q_k^{(h_{\max})}(l). \quad (19)$$

IV. Illustrative Examples and Discussion

In this section, the performance of the three proposed classification architectures are assessed drawing upon synthetic data as well as real recorded data. Specifically, in the next section, the analysis is conducted by means of standard Monte Carlo counting techniques, while in the last section, the procedures are applied to the Phase One data [14].

A. Simulated Data

In the following, data are generated resorting to independent Monte Carlo trials and using two different models for the structure of the clutter covariance matrix. In the first case, we suppose the prevalence of the clutter contribution assuming an exponential shaped clutter PSD, whereas, in the second case, we do not neglect the thermal noise contribution and model the clutter samples as the summation of the echoes from patches at distinct angles. The numerical examples assume $N = 16$, whereas two configurations for K and L are considered, namely, $K = 96$, $L = 3$ and $K = 32$, $L = 2$. Moreover, the presented analysis consists of a first qualitative part, where the classification outcomes of single Monte Carlo trial are shown, and a second quantitative part, where the root mean square classification error (RMSCE) is evaluated over 1000 independent Monte Carlo runs. The classification error is defined as the number of range bins whose class is not correctly identified.

1) Prevalence of the clutter contribution: The examples considered here are aimed at investigating the behavior of Proposition 1 and 2 when

$$\boldsymbol{M}_l = \sigma_{c,l}^2\boldsymbol{M}_c, \quad (20)$$

where $\sigma_{c,l}^2$ is the clutter power of the lth class, and $\boldsymbol{M}_c$ is the common clutter structure, such that $\boldsymbol{M}_c(i,j) = \rho^{|i-j|}$ with $\rho = 0.9$. It is important to observe that for the considered model, the classification procedure relying on Proposition 3 cannot be applied due to the fact that $r_l = N$, $l = 1, \ldots, L$.

[2]Notice that we are neglecting some constants that do not depend on r_l and, hence, do not enter the decision process.

As for the initialization of p_ls, we choose equiprobable priors, namely, $p_l = 1/L$, whereas the initial value of $\boldsymbol{M}_c$ is set by generating a random Hermitian structure as $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}^\dagger/\mathrm{Tr}\left(\boldsymbol{X}\boldsymbol{X}^\dagger\right)$, where $\boldsymbol{X}$ is a $N \times K$ matrix whose columns are complex Gaussian random vectors with zero mean and identity covariance matrix. Finally, the L clutter power levels are initialized as follows:

1) for each range bin, compute $g(k) = \frac{1}{N}\boldsymbol{z}_k^\dagger \boldsymbol{S}^{-1}\boldsymbol{z}_k$, $k = 1, \ldots, K$;
2) sort the above quantities in ascending order, $\tilde{g}(1) \leq \tilde{g}(2) \leq \ldots \leq \tilde{g}(K)$;
3) the mean values of the K/L subsets of the ordered powers is used to set the initial value of the clutter power levels, namely,

$$\widehat{\sigma}_{c,l}^2 = \frac{L}{K} \sum_{i=(l-1)\frac{K}{L}+1}^{l\frac{K}{L}} \tilde{g}(i), \quad l = 1, \ldots L. \tag{21}$$

As preliminary step, we analyze the requirements of the proposed procedures in terms of number of EM iterations. To this end, in Figure 1, we plot the average values of $\Delta\mathcal{L}(h) = |\mathcal{L}(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}^{(h)}) - \mathcal{L}(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}^{(h-1)})/\mathcal{L}(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}^{(h)})|$ over 1000 Monte Carlo trials versus h for Propositions 1 and 2. Specifically, the figure assumes $K_1 = 24$, $K_2 = 24$, $K_3 = 48$, $\sigma_{c,1}^2 = 20$ dB, $\sigma_{c,2}^2 = 30$ dB, $\sigma_{c,3}^2 = 40$ dB, and $t_{max} = 10$, where t_{max} is the iteration number for the alternating maximization procedure in Proposition 2. It turns out that, for the considered parameters,

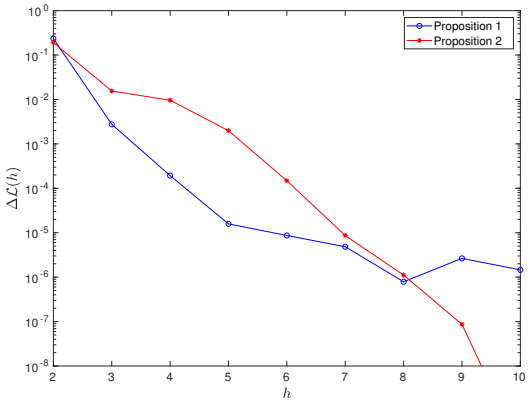


Fig. 1. Log-likelihood variation versus the iteration number of the EM procedure assuming the structure given by (20).

10 iterations are sufficient to achieve a variation lower than 10^{-4}. Similar results are obtained also for other parameter setting but for brevity are not shown here. They point out that 10 iterations are generally sufficient for convergence. Therefore, in the next numerical examples, we set $h_{max} = 10$. As for t_{max}, we have also analyzed its effect on the joint log-likelihood and the results show that $t_{max} = 10$ is a proper choice.

Now, we evaluate the effect of the clutter power levels on the classification performance. To this end, we assume $K_1 = 32$, $K_2 = 32$, $K_3 = 32$, and consider the following three cases for the clutter power levels: (1) [20,25,30] dB; (2) [20,30,40] dB; (3) [20,35,50] dB. Figure 2 shows a snapshot (to wit, a Monte Carlo outcome) for the three cases, where the estimated clutter classes are represented by "x" red stems, whereas the true classes by the "o" blue stems. The results highlight that for the considered parameters and from a qualitative point of view, the classification architecture based on Proposition 2 can achieve better performance than that based on Proposition 1. A more quantitative analysis can be obtained by resorting to

TABLE I
RMSCE FOR COVARIANCE MODEL (20) AND DIFFERENT CLUTTER POWERS ASSUMING $K = 96$

	case (1)	case (2)	case (3)
Proposition 1	19.85	2.87	0.29
Proposition 2	3.10	0.06	0

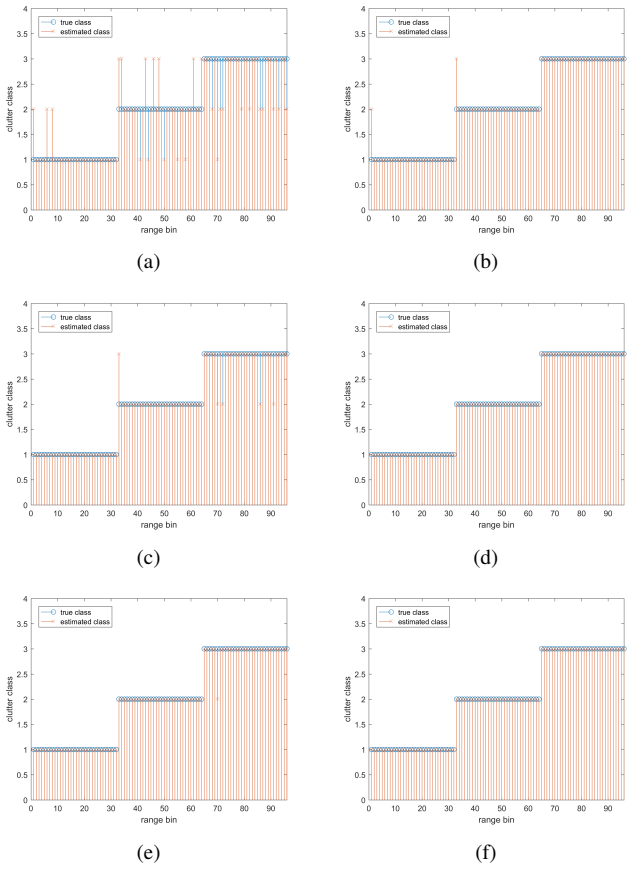


Fig. 2. Classification snapshots for different clutter power levels: (a) case (1) using Proposition 1; (b) case (1) using Proposition 2; (c) case (2) using Proposition 1; (d) case (2) using Proposition 2; (e) case (3) using Proposition 1; (f) case (3) using Proposition 2.

the RMSCE, whose values for the considered scenarios are reported in Table I. These values confirm the superiority of the algorithm based on Proposition 2 with respect to that relying on Proposition 1, indicating that a priori information about the structure of the covariance matrix can lead to a reduction of the unknown parameters and, hence, to better classification performance. In fact, the simulated covariance matrix, given by (20), adheres to the model of Proposition 2

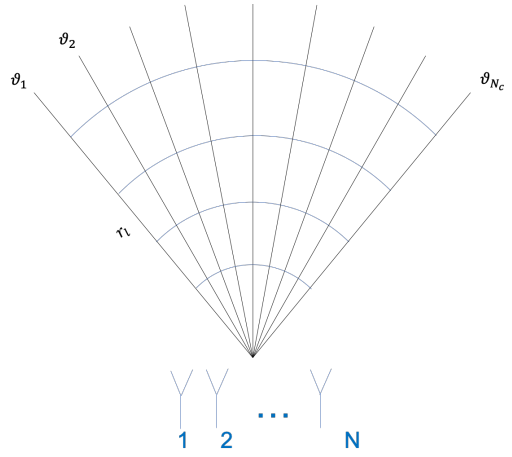


Fig. 3. Angular sector under surveillance.

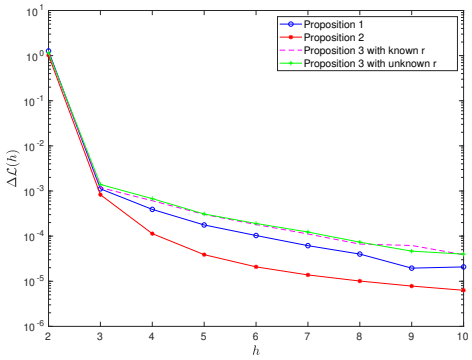


Fig. 4. Log-likelihood variation versus the iteration number of the EM procedure assuming the structure given by (22).

with the consequence that the algorithm based on Proposition 1 estimates more parameters than required. In addition, as expected, the larger the power separation between different clutter classes, the lower the error values.

Finally, we evaluate the effect of different configurations for the K_ls on the classification performance in terms of the RMSCE assuming $\sigma^2_{c,l} = 20 + 10(l-1)$ dB, $l = 1, 2, 3$. The classification results are shown in Table II. The superiority of the classification architecture based on Proposition 2 is further validated. Moreover, it is worth noticing that the more challenging case for both Propositions is $K_1 = 18$, $K_2 = 18$, and $K_3 = 60$. This behavior can be explained by the fact that the classification procedures seems to merge classes with similar powers and small cardinality.

In the last examples, we assume that $K = 2N = 32$ which is the classical length of the window containing secondary data used to estimate the interference covariance matrix before target detection [26], [28], [54, and references therein]. Such value is chosen to obtain a good compromise between two opposing objectives, namely limiting the detection loss due to adaptivity and preserving the homogeneous assumption within the secondary data window. Therefore, we assume that in this case $L = 2$. For brevity, we summarize the performance in terms of the RMSCE in Tables III and IV which are analogous to the previous tables. Specifically, in Table III, $K_1 = K_2 = 16$ and the cases are the same as in the previous Table I but for the third clutter region that is no longer present, whereas in Table IV, two configurations for K_1 and K_2 are considered with $\sigma^2_{c,l} = 20 + 10(l-1)$ dB, $l = 1, 2$. The inspection of these results confirm the superiority of the classifier based upon Proposition 2 over that relying on Proposition 1.

2) Distributed clutter plus thermal noise: In this subsection, we assume another clutter covariance matrix model[3]. As shown in Figure 3, we consider a uniformly-spaced linear array of N identical and isotropic sensors with inter-element distance equal to $\lambda/2$, where λ is the wavelength corresponding to the radar carrier frequency. We only consider the spatial processing for simplicity and model the clutter samples as the summation of individual patch returns at distinct angles [55], leading to the following covariance structure

$$\boldsymbol{M}_l = \sigma^2_{c,l} \sum_{\theta_i \in \Theta_l} \boldsymbol{v}(\theta_i)\boldsymbol{v}(\theta_i)^\dagger + \sigma^2_n \boldsymbol{I}, \tag{22}$$

where

- $\Theta_l = \{\theta^l_1, \theta^l_2, \ldots, \theta^l_{N^l_c}\}$ (for simplicity, we suppose that the number of the angular sectors is the same for each class, namely, $N^l_c = N_c$ for all l);
- $\boldsymbol{v}(\theta_i)$ is the spatial steering vector whose expression is given by $\boldsymbol{v}(\theta_i) = \frac{1}{\sqrt{N}}\left[1, e^{j\pi \sin\theta_i}, \ldots, e^{j\pi(N-1)\sin\theta_i}\right]^T \in \mathbb{C}^{N\times 1}$.

In the following, we set $N_c = 5$, the beam pointing direction to $0°$, and an angular sector within the first null beamwidth of $14°$, namely, $\Theta_l = \{-5.6°, -2.8°, 0°, 2.8°, 5.6°\}$.

The initialization method is the same as that in the Subsection IV-A1. Moreover, as to Proposition 3, we consider two situations, i.e., the clutter rank $\boldsymbol{r}$ is known and $\boldsymbol{r}$ is unknown. In the latter case, the GIC rule with $a = 2$ is exploited to estimate $\boldsymbol{r}$ using (16).

Figure 4 shows the average values of $\Delta\mathcal{L}(h)$, defined as in the previous subsection, versus h for $K_1 = 32$, $K_2 = 32$, $K_3 = 32$, $\sigma^2_{c,1} = 20$ dB, $\sigma^2_{c,2} = 30$ dB and $\sigma^2_{c,3} = 40$ dB. The curves confirm that a few iterations are sufficient for convergence.

A qualitative analysis is provided by Figures 5-7, where Monte Carlo outcomes for the same clutter power configurations as in Section IV-A1 are shown (assuming $K_1 = K_2 = K_3 = 32$). The curves highlight that, for the considered parameters, the classification methods based on Propositions 1 and 3 share almost the same performance. As to Proposition 2, it exhibits the worst performance. The behavior of the classifiers is explained observing that they are fed with data generated according to (22) that is the basis of Proposition 3 and in agreement with Proposition 1, but not with Proposition 2.

The RMSCE for different clutter power levels are shown in Table V. The obtained values confirm that the classification performances for Propositions 1 and 3 are very close to each other and are generally much better than that based

[3]In the case of time/space-time processing, a low-rank clutter covariance matrix can also arise when the clutter power spectral density exhibits a small spread with respect to the pulse repetition frequency.

TABLE II
RMSCE FOR DIFFERENT VALUES OF (K_1, K_2, K_3) AND COVARIANCE MODEL (20) ASSUMING $K = 96$

	(20,30,46)	(30,46,20)	(46,20,30)	(24,24,48)	(24,48,24)	(48,24,24)	(18,18,60)	(18,60,18)	(60,18,18)
Prop. 1	21.14	2.52	7.25	18.63	4.67	8.04	39.18	8.50	28.33
Prop. 2	0.13	0.09	0.09	0.10	0.09	0.08	18.99	0.09	1.04

TABLE III
RMSCE FOR COVARIANCE MODEL (20) AND DIFFERENT CLUTTER POWERS ASSUMING $K = 32$

	case (1)	case (2)	case (3)
Proposition 1	6.36	1.54	0.30
Proposition 2	2.29	0.96	0.14

TABLE IV
RMSCE FOR DIFFERENT VALUES OF (K_1, K_2) AND COVARIANCE MODEL (20) ASSUMING $K = 32$

	(12,20)	(20,12)
Proposition 1	2.23	2.61
Proposition 2	0.67	2.29

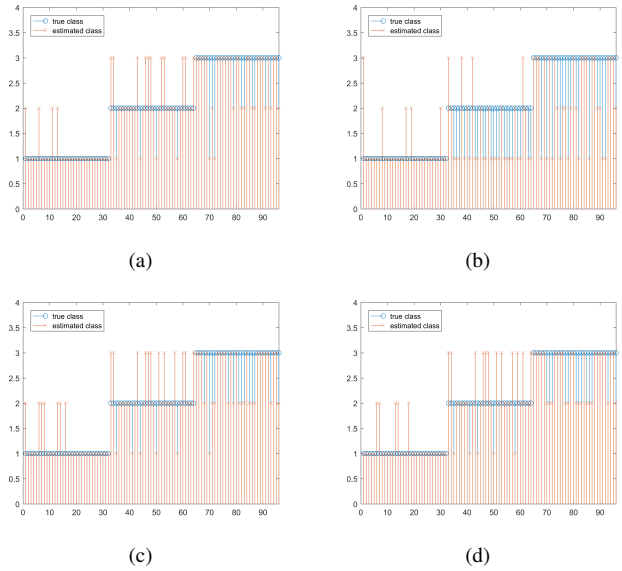
on Proposition 2. Table VI contains the RMSCE values for different K_l configurations. Inspection of the table indicates that the classification algorithm based on Proposition 3 can guarantee better performance than that based on Proposition 1. In addition, the knowledge of $\boldsymbol{r}$ does not significantly affects the resulting performance. Again, Proposition 2 continues to return the highest error values. Finally, we show the RMSCE for $K = 2N = 32$ and $L = 2$ in Tables VII and VIII, where the parameter setting is exactly the same as in Tables III and IV, respectively, but for the covariance matrix given by (22). The tables indicate that for this parameter setting the performance of the classifiers based upon Propositions 1 and 3 are similar and outperform the selection scheme based upon Proposition 2.

B. Real data

In this section, we assess the performance analysis on real L-band land clutter data, recorded in 1985 using the MIT Lincoln Laboratory Phase One radar at the Katahdin Hill site, MIT Lincoln Laboratory. We consider datasets contained in the files $H067037.2$ and $H067038.3$, which are composed of

TABLE V
RMSCE FOR DIFFERENT CLUTTER POWERS AND COVARIANCE MODEL (22) ASSUMING $K = 96$

	case (1)	case (2)	case (3)
Proposition 1	26.29	5.87	1.03
Proposition 2	43.89	33.08	29.21
Proposition 3, $\boldsymbol{r}$ known	29.16	5.64	0.87
Proposition 3, $\boldsymbol{r}$ unknown	29.94	5.81	0.97



(a) (b) (c) (d)

Fig. 5. Classification snapshots for $\sigma^2_{c,1} = 20$ dB, $\sigma^2_{c,2} = 25$ dB and $\sigma^2_{c,3} = 30$ dB: (a) Proposition 1; (b) Proposition 2; (c) Proposition 3 with known $\boldsymbol{r}$; (d) Proposition 3 with unknown $\boldsymbol{r}$.

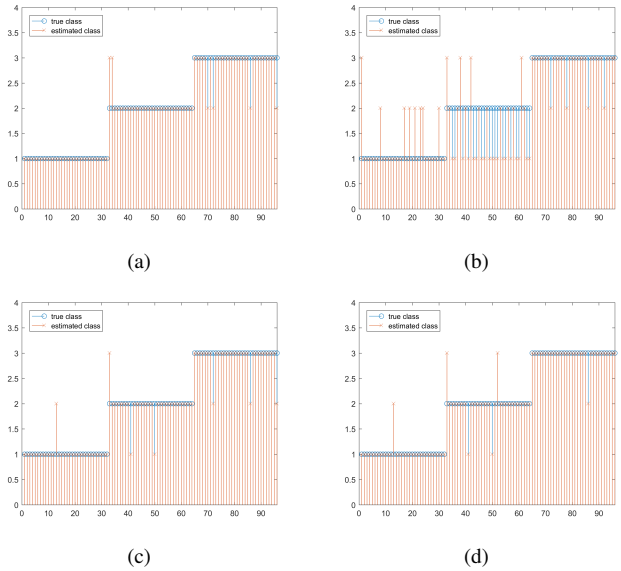
30720 temporal returns from 76 range cells with VV and HH-polarization, respectively. More details about this dataset can be found in [56]–[58] and references therein.

The 3-D clutter intensity field, from the Phase One file $H067037.2$, is plotted in Figure 8. It is evident the presence of two regions with different power levels (region 1 from cell 1 to cell 48 and region 2 from cell 49 to cell 76). This behavior, already observed in [58], is due to the fact that data were measured from range cells containing agricultural fields in contrast to windblown vegetation. Other five major terrain categories, distributed within the two major regions, are also evident, as indicated in figure. The 3-D normalized intensity plot relative to the $H067038.3$ data file is reported in Figure 9). Here, three major areas with different power levels can be identified.

In Figures 10 and 11, we plot the normalized power spectral density (PSD) for the VV and HH polarizations, respectively. To this end, we select 5 range bins each of them representative of the different terrains indicated in Figure 8 and apply Welch method over 8 partially overlapped segments of length 4096 pulses that are weighted using Blackman window. The figures show that although the general shape of the PSD is shared among the considered range bins, each curve decreases with slightly different rates leading to different decorrelation time intervals and, hence, to slightly different structures of the clutter covariance matrices. As shown below, this behavior justifies in part the observed classification outcomes.

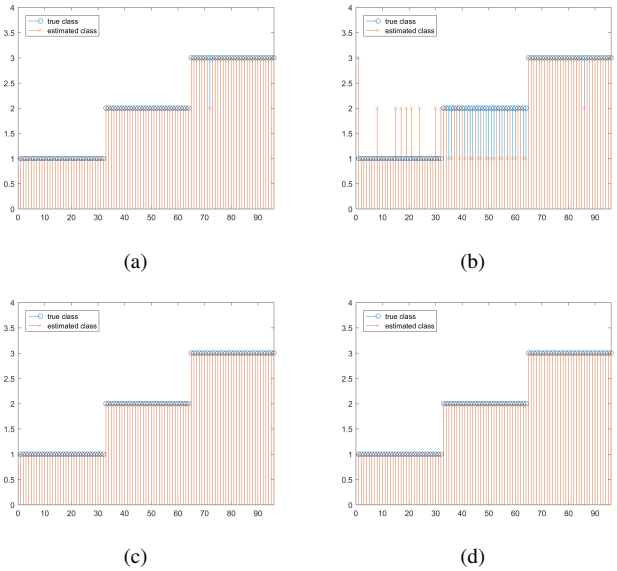
TABLE VI
RMSCE FOR DIFFERENT VALUES OF (K_1, K_2, K_3) AND COAVARIANCE MODEL (22) ASSUMING $K = 96$

	(20,30,46)	(30,46,20)	(46,20,30)	(24,24,48)	(24,48,24)	(48,24,24)	(18,18,60)	(18,60,18)	(60,18,18)
Prop. 1	20.48	8.32	10.29	18.37	8.29	15.70	37.36	13.12	38.32
Prop. 2	37.66	29.19	24.99	35.84	27.90	25.97	44.59	22.11	35.72
Prop. 3, $\boldsymbol{r}$ known	19.44	6.74	7.99	17.23	7.54	11.76	35.66	11.24	37.73
Prop. 3, $\boldsymbol{r}$ unknown	18.55	6.83	7.29	16.29	7.58	11.00	35.28	10.84	38.28



(a) (b) (c) (d)

Fig. 6. Classification snapshots for $\sigma^2_{c,1} = 20$ dB, $\sigma^2_{c,2} = 30$ dB and $\sigma^2_{c,3} = 40$ dB: (a) Proposition 1; (b) Proposition 2; (c) Proposition 3 with known $\boldsymbol{r}$; (d) Proposition 3 with unknown $\boldsymbol{r}$.



(a) (b) (c) (d)

Fig. 7. Classification snapshots for $\sigma^2_{c,1} = 20$ dB, $\sigma^2_{c,2} = 35$ dB and $\sigma^2_{c,3} = 50$ dB: (a) Proposition 1; (b) Proposition 2; (c) Proposition 3 with known $\boldsymbol{r}$; (d) Proposition 3 with unknown $\boldsymbol{r}$.

TABLE VII
RMSCE FOR COVARIANCE MODEL (22) AND DIFFERENT CLUTTER POWERS ASSUMING $K = 32$

	case (1)	case (2)	case (3)
Proposition 1	6.77	1.90	0.49
Proposition 2	15.37	12.25	8.96
Proposition 3, $\boldsymbol{r}$ known	7.39	2.06	0.47
Proposition 3, $\boldsymbol{r}$ unknown	8.02	2.39	0.68

TABLE VIII
RMSCE FOR DIFFERENT VALUES OF (K_1, K_2) AND COVARIANCE MODEL (22) ASSUMING $K = 32$

	(12,20)	(20,12)
Proposition 1	2.54	2.99
Proposition 2	8.94	11.51
Proposition 3, $\boldsymbol{r}$ known	2.59	3.17
Proposition 3, $\boldsymbol{r}$ unknown	3.00	3.23

Now, these data are fed to the proposed algorithms and the used parameters are: $N = 8$, $K = 75$, $L = 3$ or 5, and a maximum number iterations of 10 (for both EM and alternating procedure). Classification results, relative to the $H067037.2$ dataset, are reported in Figures 12 and 13, for a number of classes $L = 3$ and $L = 5$, respectively. Notice that data are characterized by small temporal variations of the power (variations in time on a given range cell, or on few cells) and of the covariance structure (as stated above) due to the inherent characteristics of the observed scene. However, since displaying the covariance structure variations over the range bins is not an easy task and it would lead to figures that are difficult to read, we compare the estimated classes with the power levels only, averaged over 100 temporal samples near the selected temporal N samples. The inspection of the figure points out that estimated classes almost follow the power profile for both $L = 3$ and $L = 5$ with the classification algorithm based upon Proposition 2 that exhibits a slightly lower discrimination capability with respect to the other schemes. As a matter of fact, the design assumptions behind Proposition 2 are not exactly met by the considered datasets since the clutter covariance structure does not keep constant as corroborated by the PSD analysis. Finally, notice that assuming five classes allows to distinguish between all the

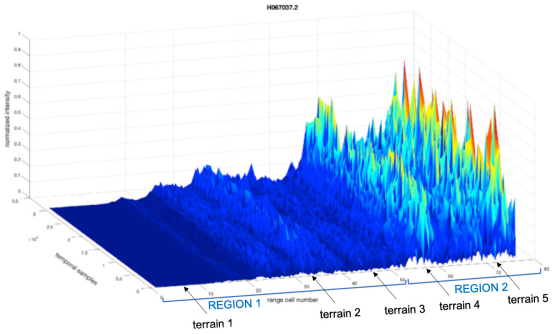


Fig. 8. 3-D normalized intensity field of clutter returns (H067037.2 dataset).

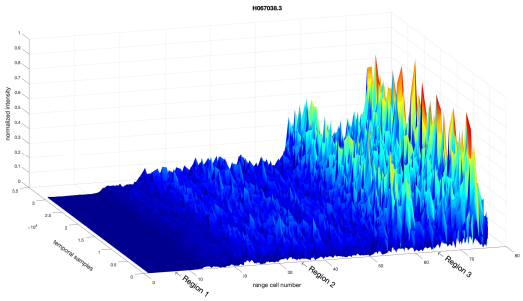


Fig. 9. 3-D normalized intensity field of clutter returns (H067038.3 dataset).

five terrains indicated in Figure 8. Other results obtained over different data blocks and not reported here for brevity confirm the behavior observed in the considered numerical examples.

The classification results for dataset $H067038.3$ are shown in Figures 14 and 15, respectively, and confirm what observed in the previous figures.

V. Conclusions

This paper has proposed several algorithms to classify clutter radar echoes with the goal of partitioning the possibly heterogeneous training dataset into homogeneous subsets, which, then, can be used for estimation/detection purposes. The algorithms have been designed using the EM algorithm in conjunction with the latent variable model. More precisely, considering three different structures for the clutter covariance

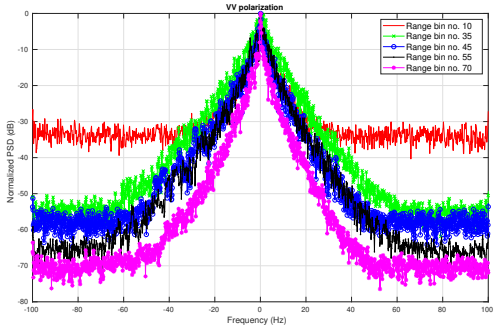


Fig. 10. Normalized power spectral densities of clutter for 5 range bins belonging to H067037.2 dataset (VV polarization).

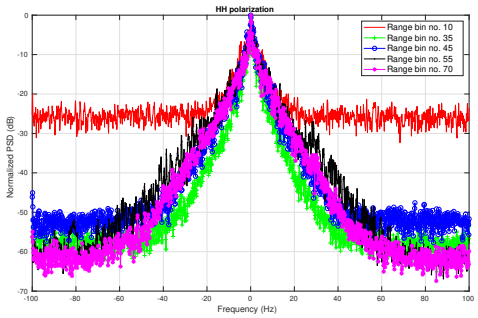


Fig. 11. Normalized power spectral densities of clutter for 5 range bins belonging to H067038.3 dataset (HH polarization).

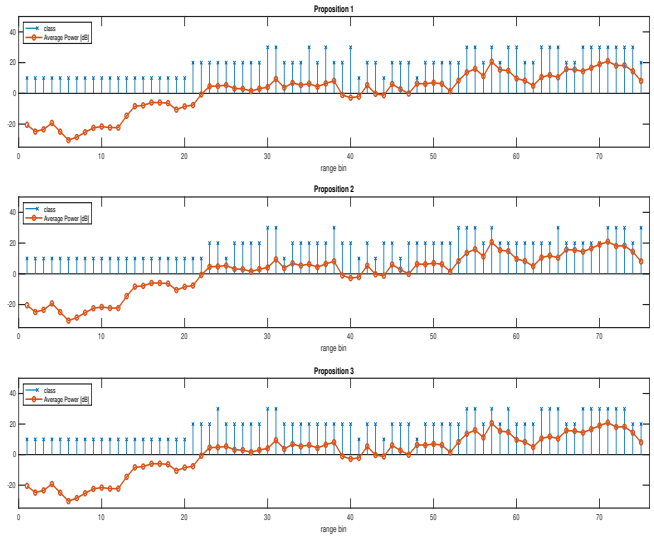


Fig. 12. Average power and estimated classes for $L = 3$ (H067037.2 dataset).

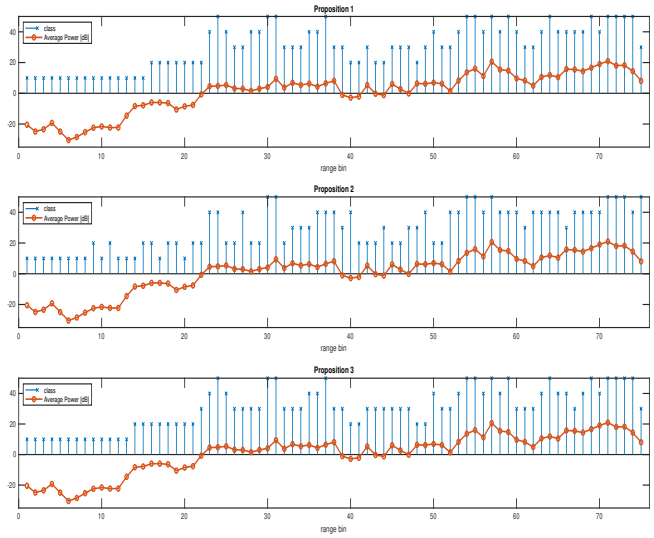


Fig. 13. Average power and estimated classes for $L = 5$ (H067037.2 dataset).

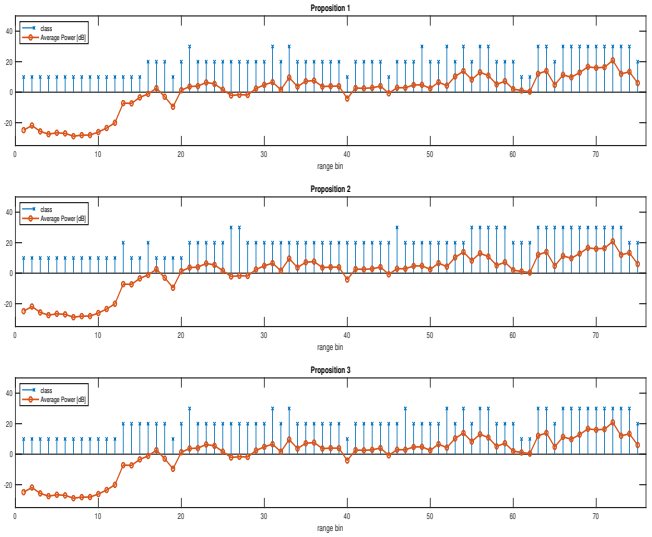


Fig. 14. Average power and estimated classes for $L = 3$ (H067038.3 dataset).

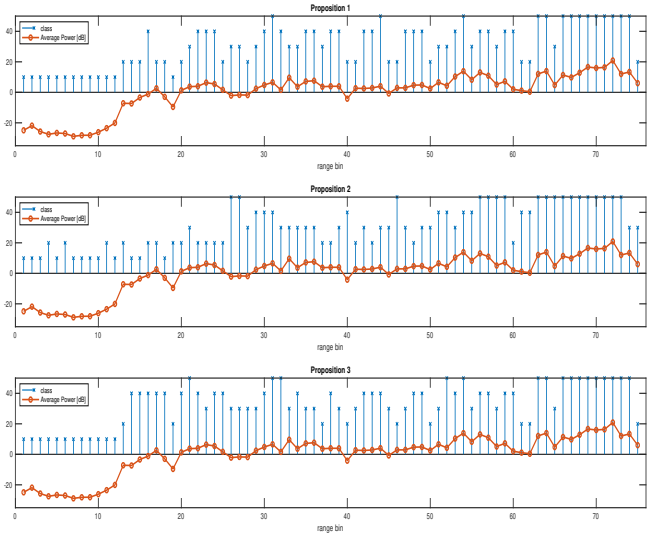


Fig. 15. Average power and estimated classes for $L = 5$ (H067038.3 dataset).

matrix (from the most general case of a Hermitian structure to the specific one where thermal noise contribution is accounted for) three different classification architectures have been introduced. Performance analysis for both simulated and real data has clearly shown the capability of the proposed approach to solve the problem of clutter data clustering. More importantly, these schemes can be used as preliminary stage of a detection architecture, where the detection stage exploits the information provided by the classifier to process homogeneous data.

Future research tracks include the design of clustering algorithms capable of handling the presence of outliers, which can be discarded once identified, and without any a priori assumption about the total number of homogeneous subsets. Another issue is related to further structures for the clutter covariance matrix that can improve the estimation quality and, hence, detection performance of those receivers relying on such estimates. Finally, the design of architectures for the joint detection and classification of multiple clutter edges represent a further important extension of this work. All the above topics represent the current research activity.

Appendix A
Proof of Proposition 1

Let us consider the following problem

$$\widehat{\boldsymbol{\sigma}}^{(h)} = \arg\max_{\boldsymbol{\sigma}} g_1(\boldsymbol{M}_1, \ldots, \boldsymbol{M}_L), \tag{23}$$

which is tantamount to solving

$$\widehat{\boldsymbol{M}}_l^{(h)} = \arg\max_{\boldsymbol{M}_l} \underbrace{\sum_{k=1}^{K} q_k^{(h-1)}(l) \left[-\log\det(\boldsymbol{M}_l) - \boldsymbol{z}_k^\dagger \boldsymbol{M}_l^{-1} \boldsymbol{z}_k \right]}_{d(\boldsymbol{M}_l)} \tag{24}$$

for each $l = 1, \ldots, L$. To this end, we set to zero the first derivative of $d(\boldsymbol{M}_l)$ with respect to $\boldsymbol{M}_l$ [59], namely

$$\frac{\partial}{\partial \boldsymbol{M}_l}[d(\boldsymbol{M}_l)] = -(\boldsymbol{M}_l^T)^{-1} \sum_{k=1}^{K} q_k^{(h-1)}(l) + (\boldsymbol{M}_l^T)^{-1} \left[\sum_{k=1}^{K} q_k^{(h-1)}(l) \boldsymbol{z}_k \boldsymbol{z}_k^\dagger \right]^T (\boldsymbol{M}_l^T)^{-1} = \boldsymbol{0}. \tag{25}$$

The solution of the above equation is given by

$$\widehat{\boldsymbol{M}}_l^{(h)} = \left[\sum_{k=1}^{K} q_k^{(h-1)}(l) \boldsymbol{z}_k \boldsymbol{z}_k^\dagger \right] / \left[\sum_{k=1}^{K} q_k^{(h-1)}(l) \right], \tag{26}$$

and the proof is complete.

Appendix B
Proof of Proposition 2

In order to come up with the estimates of the $\sigma_{c,l}^2$s and $\boldsymbol{M}$, we set to zero the first derivatives of $g_2(\boldsymbol{\sigma}_c^2, \boldsymbol{M})$ (with respect to the $\sigma_{c,l}^2$s and $\boldsymbol{M}$), namely

$$\forall l = 1, \ldots, L: \quad \frac{\partial g_2(\boldsymbol{\sigma}_c^2, \boldsymbol{M})}{\partial \sigma_{c,l}^2} = -\sum_{k=1}^{K} q_k^{(h-1)}(l) \times \left(\frac{N}{\sigma_{c,l}^2} - \frac{1}{\sigma_{c,l}^4} \boldsymbol{z}_k^\dagger \boldsymbol{M}^{-1} \boldsymbol{z}_k \right) = 0 \tag{27}$$

and

$$\frac{\partial g_2(\boldsymbol{\sigma}_c^2, \boldsymbol{M})}{\partial \boldsymbol{M}} = -\sum_{k=1}^{K} \sum_{l=1}^{L} q_k^{(h-1)}(l) \times \left(\boldsymbol{M}^{-1} - \boldsymbol{M}^{-1} \frac{1}{\sigma_{c,l}^2} \boldsymbol{z}_k \boldsymbol{z}_k^\dagger \boldsymbol{M}^{-1} \right)^T = \boldsymbol{0}. \tag{28}$$

The equations can be re-written as

$$\sigma_{c,l}^2 = \frac{\sum_{k=1}^{K} q_k^{(h-1)}(l) \boldsymbol{z}_k^\dagger \boldsymbol{M}^{-1} \boldsymbol{z}_k}{N \sum_{k=1}^{K} q_k^{(h-1)}(l)}, \quad l = 1, \ldots, L, \tag{29}$$

and

$$\boldsymbol{M} = \frac{1}{K} \sum_{k=1}^{K} \sum_{l=1}^{L} q_k^{(h-1)}(l) \frac{\boldsymbol{z}_k \boldsymbol{z}_k^\dagger}{\sigma_{c,l}^2}, \tag{30}$$

respectively, where we have used the fact that $\sum_{k=1}^{K}\sum_{l=1}^{L} q_k^{(h-1)}(l) = K$. Since the equation system formed by (29) and (30) does not admit a closed-form solution, we propose to resort to alternating maximization; based on the $(\widehat{\sigma}_{c,l}^2)^{(h-1)}$s and $\widehat{\boldsymbol{M}}^{(h-1)}$ we first compute the $(\widehat{\sigma}_{c,l}^2)^{(1),(h)}$s by plugging $\widehat{\boldsymbol{M}}^{(h-1)}$ into eqs. (29); then, we compute $\widehat{\boldsymbol{M}}^{(1),(h)}$ by plugging the $(\widehat{\sigma}_{c,l}^2)^{(1),(h)}$s into eq. (30). This procedure can be iterated obtaining, after t_{max} iterations, $(\widehat{\sigma}_{c,l}^2)^{(h)} = (\widehat{\sigma}_{c,l}^2)^{(t_{max}),(h)}$, $l = 1,\ldots,L$, and $\widehat{\boldsymbol{M}}^{(h)} = \widehat{\boldsymbol{M}}^{(t_{max}),(h)}$. To conclude the proof we observe that both EM and alternating maximization lead to a non decreasing sequence of likelihood values [48].

APPENDIX C
PROOF OF PROPOSITION 3

First we re-write (12) as follows

$$g_3(\sigma_n^2,\boldsymbol{R}_1,\ldots,\boldsymbol{R}_L) = \sum_{k=1}^{K}\sum_{l=1}^{L} q_k^{(h-1)}(l)\left[-\log\det(\sigma_n^2\boldsymbol{I} + \boldsymbol{R}_l) - N\log\pi - \boldsymbol{z}_k^\dagger(\sigma_n^2+\boldsymbol{R}_l)^{-1}\boldsymbol{z}_k\right]$$

and also as

$$g_3'(\sigma_n^2,\boldsymbol{R}_1,\ldots,\boldsymbol{R}_L) = \sum_{k=1}^{K}\sum_{l=1}^{L} q_k^{(h-1)}(l)\left\{-\log\det(\sigma_n^2\boldsymbol{I} + \boldsymbol{R}_l) - \mathrm{Tr}\left[(\sigma_n^2+\boldsymbol{R}_l)^{-1}\boldsymbol{S}_k\right]\right\} \quad (31)$$

where $\boldsymbol{S}_k = \boldsymbol{z}_k\boldsymbol{z}_k^\dagger$. Now, let us consider the eigendecomposition of $\boldsymbol{R}_l$, namely $\boldsymbol{R}_l = \boldsymbol{U}_l\boldsymbol{\Lambda}_l\boldsymbol{U}_l^\dagger$, where $\boldsymbol{U}_l \in \mathbb{C}^{N\times N}$ is a unitary matrix whose columns are the eigenvectors of $\boldsymbol{R}_l$ while Λ_l is the corresponding diagonal matrix of the eigenvalues of $\boldsymbol{R}_l$; $\boldsymbol{\Lambda}_l$ can be represented as $\boldsymbol{\Lambda}_l = \mathrm{diag}\,(\lambda_{l,1},\ldots,\lambda_{l,r_l},0,\ldots,0) \in \mathbb{R}^{N\times N}$ with $\lambda_{l,1} \geq \ldots \geq \lambda_{l,r_l} > 0$. It follows that the objective function becomes

$$g_3'(\sigma_n^2,\boldsymbol{R}_1,\ldots,\boldsymbol{R}_L) = \sum_{l=1}^{L}\sum_{k=1}^{K} q_k^{(h-1)}(l)\left\{-\log\det(\sigma_n^2\boldsymbol{I} + \boldsymbol{\Lambda}_l) - \mathrm{Tr}\left[\boldsymbol{U}_l(\sigma_n^2\boldsymbol{I}+\boldsymbol{\Lambda}_l)^{-1}\boldsymbol{U}_l^\dagger\boldsymbol{S}_k\right]\right\}$$

$$= \sum_{l=1}^{L}\left\{-\left(\sum_{k=1}^{K} q_k^{(h-1)}(l)\right)\log\left[(\sigma_n^2)^{N-r_l}\prod_{m=1}^{r_l}(\sigma_n^2+\lambda_{l,m})\right] - \mathrm{Tr}\left[\boldsymbol{U}_l(\sigma_n^2\boldsymbol{I}+\boldsymbol{\Lambda}_l)^{-1}\boldsymbol{U}_l^\dagger\boldsymbol{S}_l^{(h-1)}\right]\right\},$$

where $\boldsymbol{S}_l^{(h-1)} = \sum_{k=1}^{K} q_k^{(h-1)}(l)\boldsymbol{S}_k$. Replacing $\boldsymbol{S}_l^{(h-1)}$ by its eigendecomposition, we also come up with

$$\sum_{l=1}^{L}\left\{-\left(\sum_{k=1}^{K} q_k^{(h-1)}(l)\right)\log\left[(\sigma_n^2)^{N-r_l}\prod_{m=1}^{r_l}(\sigma_n^2+\lambda_{l,m})\right] - \mathrm{Tr}\left[\boldsymbol{U}_l(\sigma_n^2\boldsymbol{I}+\boldsymbol{\Lambda}_l)^{-1}\boldsymbol{U}_l^\dagger\boldsymbol{O}_l^{(h-1)}\boldsymbol{\Gamma}_l^{(h-1)}(\boldsymbol{O}_l^{(h-1)})^\dagger\right]\right\}$$

where $\boldsymbol{\Gamma}_l^{(h-1)} = \mathrm{diag}\,(\gamma_{l,1}^{(h-1)},\ldots,\gamma_{l,N}^{(h-1)})$ with $\gamma_{l,1}^{(h-1)} \geq \ldots \geq \gamma_{l,N}^{(h-1)}$ being the eigenvalues of $\boldsymbol{S}_l^{(h-1)}$ and $\boldsymbol{O}_l^{(h-1)}$ the unitary matrix of the corresponding eigenvectors. As a consequence, the objective function (31) can also be recast as

$$g_3''(\sigma_n^2,\boldsymbol{V}_l,\boldsymbol{\Lambda}_l,l=1,\ldots,L) = \sum_{l=1}^{L}\left\{-q^{(h-1)}(l) \times\log\left[(\sigma_n^2)^{N-r_l}\prod_{m=1}^{r_l}(\sigma_n^2+\lambda_{l,m})\right] - \mathrm{Tr}\left[\boldsymbol{V}_l(\sigma_n^2\boldsymbol{I}+\boldsymbol{\Lambda}_l)^{-1}\boldsymbol{V}_l^\dagger\boldsymbol{\Gamma}_l^{(h-1)}\right]\right\}$$

where $q^{(h-1)}(l) = \sum_{k=1}^{K} q_k^{(h-1)}(l)$ and $\boldsymbol{V}_l = (\boldsymbol{O}_l^{(h-1)})^\dagger\boldsymbol{U}_l$. Exploiting *Theorem 1* of [60], it is possible to show that $\forall l = 1,\ldots,L$

$$\arg\max_{\boldsymbol{V}_l} -\mathrm{Tr}\left[\boldsymbol{V}_l(\sigma_n^2\boldsymbol{I}+\boldsymbol{\Lambda}_l)^{-1}\boldsymbol{V}_l^\dagger\boldsymbol{\Gamma}_l^{(h-1)}\right] = \boldsymbol{I},$$

which implies that $\boldsymbol{U}_l^{(h)} = \boldsymbol{O}_l^{(h-1)}$. Then, we obtain

$$g_3'''(\sigma_n^2,\boldsymbol{\Lambda}_l,l=1,\ldots,L) = \max_{\substack{\boldsymbol{V}_l \\ l=1,\ldots,L}} g_3''(\sigma_n^2,\boldsymbol{V}_l,\boldsymbol{\Lambda}_l,l=1,\ldots,L) \quad (32)$$

$$= \sum_{l=1}^{L}\left\{-q^{(h-1)}(l)(N-r_l)\log\sigma_n^2 - q^{(h-1)}(l) \times\sum_{m=1}^{r_l}\log(\sigma_n^2+\lambda_{l,m}) - \sum_{m=1}^{r_l}\frac{\gamma_{l,m}^{(h-1)}}{\sigma_n^2+\lambda_{l,m}} - \sum_{m=r_l+1}^{N}\frac{\gamma_{l,m}^{(h-1)}}{\sigma_n^2}\right\}.$$

As the next step towards the final result, we set to zero the first derivative of the above objective function with respect to $\lambda_{l,m}$, $m = 1,\ldots,r_l$, namely

$$\frac{\partial}{\partial\lambda_{l,m}}\left[-q^{(h-1)}(l)\log(\sigma_n^2+\lambda_{l,m}) - \frac{\gamma_{l,m}^{(h-1)}}{\sigma_n^2+\lambda_{l,m}}\right] = 0$$

$$\Rightarrow -q^{(h-1)}(l)\frac{1}{(\sigma_n^2+\lambda_{l,m})} + \frac{\gamma_{l,m}^{(h-1)}}{(\sigma_n^2+\lambda_{l,m})^2} = 0$$

$$\Rightarrow \widehat{\lambda}_{l,m} = \begin{cases} \frac{\gamma_{l,m}^{(h-1)}}{q^{(h-1)}(l)} - \sigma_n^2, & \sigma_n^2 < \frac{\gamma_{l,m}^{(h-1)}}{q^{(h-1)}(l)}, \\ 0, & \text{otherwise.} \end{cases} \quad (33)$$

After replacing $\lambda_{l,m}$ with $\widehat{\lambda}_{l,m}$ in (32) it still remains to maximize with respect to σ_n^2. However, the corresponding partially-compressed likelihood is a complicated expression due to the "resulting" value of the rank which can be less than or equal to r_l, $l = 1,\ldots,L$, according to σ_n^2; thus, we approximate the partially-compressed likelihood with the expression obtained letting $\sigma_n^2 < \frac{\gamma_{l,m}^{(h-1)}}{q^{(h-1)}(l)}$, $l = 1,\ldots,L$, $m = 1,\ldots,r_l$. Thus, the last optimization is

$$\max_{\sigma_n^2}\sum_{l=1}^{L}\left\{-q^{(h-1)}(l)(N-r_l)\log\sigma_n^2 - q^{(h-1)}(l)\right.$$

$$\times \sum_{m=1}^{r_l} \log\left(\frac{\gamma_{l,m}^{(h-1)}}{q^{(h-1)}(l)}\right) - r_l q^{(h-1)}(l) - \sum_{m=r_l+1}^{N} \frac{\gamma_{l,m}^{(h-1)}}{\sigma_n^2}\Bigg\},$$

which can be solved by finding the zeros of the following function

$$\frac{\partial}{\partial \sigma_n^2}\left[\sum_{l=1}^{L}\left\{ - q^{(h-1)}(l)(N-r_l)\log\sigma_n^2 - \sum_{m=r_l+1}^{N}\frac{\gamma_{l,m}^{(h-1)}}{\sigma_n^2}\right]\right.$$
$$= -\frac{1}{\sigma_n^2}\sum_{l=1}^{L} q^{(h-1)}(l)(N-r_l) + \frac{1}{(\sigma_n^2)^2}\sum_{l=1}^{L}\sum_{m=r_l+1}^{N}\gamma_{l,m}^{(h-1)}.$$

The result is

$$\widehat{\sigma}_n^{2(h)} = \frac{\sum_{l=1}^{L}\sum_{m=r_l+1}^{N}\gamma_{l,m}^{(h-1)}}{\sum_{l=1}^{L} q^{(h-1)}(l)(N-r_l)}. \tag{34}$$

Finally, the estimate of $\lambda_{l,m}$, $l = 1, \ldots, L$, $m = 1, \ldots, r_l$, is given by

$$\widehat{\lambda}_{l,m}^{(h)} = \begin{cases} \frac{\gamma_{l,m}^{(h-1)}}{q^{(h-1)}(l)} - \widehat{\sigma}_n^{2(h)}, & \widehat{\sigma}_n^{2(h)} < \frac{\gamma_{l,m}^{(h-1)}}{q^{(h-1)}(l)}, \\ 0, & \text{otherwise}, \end{cases} \tag{35}$$

and the proof is complete.

REFERENCES

[1] W. L. Melvin and J. A. Scheer, *Principles of Modern Radar: Advanced Techniques*, S. Publishing, Ed., Edison, NJ, 2013.

[2] M. A. Richards, W. L. Melvin, J. A. Scheer, and W. A. Holm, *Principles of Modern Radar: Radar Applications, Volume 3*, ser. Electromagnetics and Radar. Institution of Engineering and Technology, 2013.

[3] J. Liu, W. Liu, H. Liu, B. Chen, X. G. Xia, and F. Dai, "Average SINR Calculation of a Persymmetric Sample Matrix Inversion Beamformer," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 2135–2145, April 2016.

[4] J. Liu, S. Sun, and W. Liu, "One-step persymmetric GLRT for subspace signals," *IEEE Transaction on Signal Processing*, vol. 14, no. 67, pp. 3639–3648, July 15 2019.

[5] G. Foglia, C. Hao, G. Giunta, and D. Orlando, "Knowledge-aided adaptive detection in partially homogeneous clutter: Joint exploitation of persymmetry and symmetric spectrum," *Digital Signal Processing*, vol. 67, no. Supplement C, pp. 131 – 138, 2017.

[6] L. Cai and H. Wang, "A Persymmetric Multiband GLR Algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 3, pp. 806–816, 1992.

[7] P. Wang, Z. Sahinoglu, M. Pun, and H. Li, "Persymmetric Parametric Adaptive Matched Filter for Multichannel Adaptive Signal Detection," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 3322–3328, 2012.

[8] C. Hao, S. Gazor, G. Foglia, B. Liu, and C. Hou, "Persymmetric adaptive detection and range estimation of a small target," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 4, pp. 2590–2604, 2015.

[9] G. Pailloux, P. Forster, J. P. Ovarlez, and F. Pascal, "Persymmetric Adaptive Radar Detectors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 4, pp. 2376–2390, 2011.

[10] R. Nitzberg, "Application of Maximum Likelihood Estimation of Persymmetric Covariance Matrices to Adaptive Processing," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 16, no. 1, pp. 124–127, 1980.

[11] H. L. Van Trees, *Optimum Array Processing (Detection, Estimation, and Modulation Theory, Part IV)*. John Wiley & Sons, 2002.

[12] G. Foglia, C. Hao, A. Farina, G. Giunta, D. Orlando, and C. Hou, "Adaptive Detection of Point-Like Targets in Partially Homogeneous Clutter With Symmetric Spectrum," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 4, pp. 2110–2119, 2017.

[13] A. De Maio, D. Orlando, C. Hao, and G. Foglia, "Adaptive Detection of Point-like Targets in Spectrally Symmetric Interference," *IEEE Transactions on Signal Processing*, vol. 64, no. 12, pp. 3207–3220, 2016.

[14] J. B. Billingsley, *Low-angle radar land clutter - Measurements and empirical models*. Norwich, NY: William Andrew Publishing, 2002.

[15] J. B. Billingsley, A. Farina, F. Gini, M. S. Greco, and L. Verrazzani, "Statistical Analyses of Measured Radar Ground Clutter Data," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 2, pp. 579–593, 1999.

[16] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4097–4107, Sep. 2011.

[17] E. Ollila and D. E. Tyler, "Regularized M-Estimators of Scatter Matrix," *IEEE Transactions on Signal Processing*, vol. 62, no. 22, pp. 6059–6070, Nov 2014.

[18] M. Steiner and K. Gerlach, "Fast converging adaptive processor or a structured covariance matrix," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 4, pp. 1115–1126, Oct 2000.

[19] M. C. Wicks, W. L. Melvin, and P. Chen, "An efficient architecture for nonhomogeneity detection in space-time adaptive processing airborne early warning radar," in *Radar 97 (Conf. Publ. No. 449)*, Oct 1997, pp. 295–299.

[20] R. S. Adve, T. B. Hale, and M. C. Wicks, "Transform domain localized processing using measured steering vectors and non-homogeneity detection," in *Proceedings of the 1999 IEEE Radar Conference. Radar into the Next Millennium (Cat. No.99CH36249)*, April 1999, pp. 285–290.

[21] B. Himed, Y. Salama, and J. H. Michels, "Improved detection of close proximity targets using two-step nhd," in *Record of the IEEE 2000 International Radar Conference [Cat. No. 00CH37037]*, May 2000, pp. 781–786.

[22] M. Rangaswamy, B. Himed, and J. H. Michels, "Performance analysis of the nonhomogeneity detector for stap applications," in *Proceedings of the 2001 IEEE Radar Conference (Cat. No.01CH37200)*, May 2001, pp. 193–197.

[23] Y. Sun, A. Breloy, P. Babu, D. P. Palomar, F. Pascal, and G. Ginolhac, "Low-complexity algorithms for low rank clutter parameters estimation in radar systems," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 1986–1998, 2016.

[24] B. Mériaux, C. Ren, M. N. El Korso, A. Breloy, and P. Forster, "Robust estimation of structured scatter matrices in (mis)matched models," *Signal Processing*, vol. 165, pp. 163 – 174, 2019.

[25] B. Mériaux, A. Breloy, C. Ren, M. N. El Korso, and P. Forster, "Modified sparse subspace clustering for radar detection in non-stationary clutter," in *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2019, pp. 669–673.

[26] E. J. Kelly, "An adaptive detection algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, no. 2, pp. 115–127, 1986.

[27] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 1, pp. 208–216, 1992.

[28] F. Bandiera, D. Orlando, and G. Ricci, *Advanced Radar Detection Schemes Under Mismatched Signal Models*. San Rafael, US: Synthesis Lectures on Signal Processing No. 8, Morgan & Claypool Publishers, 2009.

[29] E. Conte, A. De Maio, and G. Ricci, "GLRT-based adaptive detection algorithms for range-spread targets," *IEEE Transactions on Signal Processing*, vol. 49, no. 7, pp. 1336–1348, July 2001.

[30] W. L. Melvin, "Space-time Adaptive Radar Performance in Heterogeneous Clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 2, pp. 621–633, 2000.

[31] W. L. Melvin, M. Wicks, P. Antonik, Y. Salama, Ping Li, and H. Schuman, "Knowledge-based space-time adaptive processing for airborne early warning radar," *IEEE Aerospace and Electronic Systems Magazine*, vol. 13, no. 4, pp. 37–42, 1998.

[32] G. T. Capraro, A. Farina, H. Griffiths, and M. C. Wicks, "Knowledge-based radar signal and data processing: a tutorial review," *IEEE Signal Processing Magazine*, vol. 23, no. 1, pp. 18–29, 2006.

[33] M. C. Wicks, M. Rangaswamy, R. Adve, and T. B. Hale, "Space-time adaptive processing: a knowledge-based perspective for airborne radar," *IEEE Signal Processing Magazine*, vol. 23, no. 1, pp. 51–65, 2006.

[34] A. Benavoli, L. Chisci, A. Farina, S. Immediata, L. Timmoneri, and G. Zappa, "Knowledge-based system for multi-target tracking in a littoral environment," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 3, pp. 1100–1119, 2006.

[35] S. Haykin and C. Deng, "Classification of radar clutter using neural networks," *IEEE Transactions on Neural Networks*, vol. 2, no. 6, pp. 589–600, 1991.

[36] S. Haykin, W. Stehwien, C. Deng, P. Weber, and R. Mann, "Classification of radar clutter in an air traffic control environment," *Proceedings of the IEEE*, vol. 79, no. 6, pp. 742–772, 1991.

[37] V. Anastassopoulos and G. A. Lampropoulos, "High resolution radar clutter classification," in *Proceedings International Radar Conference*, 1995, pp. 662–667.

[38] M. A. Darzikolaei, A. Ebrahimzade, and E. Gholami, "Classification of radar clutters with Artificial Neural Network," in *2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, 2015, pp. 577–581.

[39] P. Formont, F. Pascal, G. Vasile, J. Ovarlez, and L. Ferro-Famil, "Statistical Classification for Heterogeneous Polarimetric SAR Images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 567–576, 2011.

[40] V. Carotenuto, A. De Maio, D. Orlando, and P. Stoica, "Model Order Selection Rules for Covariance Structure Classification in Radar," *IEEE Transactions on Signal Processing*, vol. 65, no. 20, pp. 5305–5317, 2017.

[41] J. Liu, F. Biondi, D. Orlando, and A. Farina, "Training Data Classification Algorithms for Radar Applications," *IEEE Signal Processing Letters*, vol. 26, no. 10, pp. 1446–1450, 2019.

[42] Y. Cabanes and J. B. F. Barbaresco, M. Arnaudon, "Matrix extension for pathological radar clutter machine learning," *hal-02875440 version 1*. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02875440/

[43] ——, "Toeplitz hermitian positive definite matrix machine learning based on fisher metric lecture notes on computer sciences," pp. 261–270, 2019.

[44] ——, "Unsupervised machine learning for pathological radar clutter clustering: the p-mean-shift algorithm," *French Mod Conference CE & SAR AI for Defense*. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02875430/

[45] 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.

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

[47] E. Conte, M. Lops, and G. Ricci, "Asymptotically optimum radar detection in compound-gaussian clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 31, no. 2, pp. 617–625, 1995.

[48] E. Conte, A. De Maio, and G. Ricci, "Recursive Estimation of the Covariance Matrix of a Compound-Gaussian Process and Its Application to Adaptive CFAR Detection," *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 1908–1915, 2002.

[49] E. Conte, A. De Maio, and G. Ricci, "Covariance matrix estimation for adaptive CFAR detection in compound-Gaussian clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 2, pp. 415–426, 2002.

[50] F. Pascal, Y. Chitour, J. Ovarlez, P. Forster, and P. Larzabal, "Covariance Structure Maximum-Likelihood Estimates in Compound Gaussian Noise: Existence and Algorithm Analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 34–48, 2008.

[51] M. A. Richards, J. A. Scheer, and W. A. Holm, *Principles of Modern Radar: Basic Principles*. Raleigh, NC: Scitech Publishing, 2010.

[52] L. Yan, P. Addabbo, C. Hao, D. Orlando, and A. Farina, "New ECCM Techniques Against Noise-like and/or Coherent Interferers," *IEEE Transactions on Aerospace and Electronic Systems*, 2019.

[53] 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.

[54] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid Convergence Rate in Adaptive Arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-10, no. 6, pp. 853–863, 1974.

[55] W. L. Melvin, Ed., *Principles of Modern Radar: Advanced techniques*, ser. Radar, Sonar and Navigation. Institution of Engineering and Technology, 2012. [Online]. Available: https://digital-library.theiet.org/content/books/ra/sbra020e

[56] J. B. Billingsley, A. Farina, F. Gini, M. V. Greco, and L. Verrazzani, "Statistical analyses of measured radar ground clutter data," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 2, pp. 579–593, 1999.

[57] M. Greco, F. Gini, A. Farina, and J. B. Billingsley, "Validation of windblown radar ground clutter spectral shape," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 2, pp. 538–548, 2001.

[58] E. Conte, A. De Maio, and A. Farina, "Statistical tests for higher order analysis of radar clutter: their application to L-band measured data," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 1, pp. 205–218, 2005.

[59] A. Hjørungnes, *Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications*. Cambridge University Press, 2011.

[60] L. Mirsky, "On the trace of matrix products," *Mathematische Nachrichten*, vol. 20, pp. 171–174, 1959.