Received April 2, 2019, accepted May 15, 2019, date of publication May 20, 2019, date of current version June 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2917950

An Approximate Regularized ML Approach to Censor Outliers in Gaussian Radar Data

SUDAN HAN[1], (Student Member, IEEE), LUCA PALLOTTA[2], (Senior Member, IEEE), VINCENZO CAROTENUTO[2], (Senior Member, IEEE), ANTONIO DE MAIO[3], (Fellow, IEEE), AND XIAOTAO HUANG[1], (Member, IEEE)

[1]College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China
[2]CNIT udr Università degli Studi di Napoli 'FedericoII', I-80125 Napoli, Italy
[3]Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione, Università degli Studi di Napoli 'Federico II', I-80125 Napoli, Italy

Corresponding author: Sudan Han (xiaoxiaosu0626@163.com)

ABSTRACT This paper considers the problem of censoring outliers from the secondary dataset in a radar scenario where the sample support is limited. To this end, the generalized regularized likelihood function (GRLF) criterion is used and the corresponding regularized maximum likelihood (RML) estimate of the outlier subset is derived. Since the exact RML estimate involves the solution of a combinatorial optimization problem, a reduced complexity but approximate RML (ARML) procedure is also designed. As to the selection of the regularization parameter, both the expected likelihood (EL) principle and the cross-validation (CV) technique are exploited. At the analysis stage, the performance of the RML and ARML procedure is evaluated based on simulated data in comparison with some previously proposed methods. The results highlight that the RML and ARML algorithm achieves, in general, a satisfactory performance level whereas the previously proposed techniques often experience some performance degradation when the volume of training data is dramatically limited.

INDEX TERMS Outlier removal, regularized ML, limited training data, expected likelihood, cross-validation.

I. INTRODUCTION

Adaptive radar processing techniques are based on the accurate estimate of the interference covariance matrix usually performed the sample covariance matrix (SCM) [1]–[8]. However, SCM presents a serious disadvantage requiring a sample support large enough (usually more than twice the system degrees of freedom (DOFs)) in order for the estimate to achieve convergence. Besides, the underlying independent and identically distributed (iid) assumption for training data is commonly infringed in practical scenarios due to clutter power variations, clutter discretes, undesired outlier signals of different types, and so on [9] (heterogeneous environment). A possible way to overcome the deleterious effects of heterogeneous environments relies on the use of structured covariance matrix estimators which needs less training data than the SCM to achieve convergence [10]–[17]. Besides, an alternative approach consists in the selection of homogeneous training data before the SCM computation [18]–[22, and references therein].[1] As an example, in [26], a de-emphasis weighting approach is devised to suppress the effect of outliers during the process of SCM computation. Moreover, in [27] forward methods are utilized for joint CFAR testing to discern among the absence or presence of outliers in noise data. In [28], data selection is performed exploiting the generalized inner product (GIP) metric in conjunction with a covariance estimate based on diagonal loading. The authors of [29] have interpreted the training data selection as a parameter estimation problem and have exploited a sparsity-based approach to solve it. Finally, in [30] training data for covariance estimation are selected as those having a spectrum close to that of the cell under test. Other possible criteria that could inspire secondary data selection could be borrowed from the Machine Learning domain. A first approach, referred to as ''transfer learning'' [31], consists in applying some previously learned knowledge from a specific source to a different context. In addition the ''adversarial learning'' philosophy can be

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaolong Chen.

[1]As a matter of fact, censoring outliers is also a hot topic in statistical field, as for instance in [23]–[25].

directly utilized to classify deviant behaviors as possible outliers to excise from data, as extensively done in the field of Internet of things (IoT) [32].

In [33], [34], the authors considered the maximum likelihood (ML) estimate of the outlier subset according to the generalized likelihood function (GLF) criterion and devised an approximate ML (AML) procedure to improve the computational efficiency. The ML (AML) procedure can often completely eliminate the outlier masking effect and usually can outperform the GIP and reiterative censored GIP (RCGIP). However, the ML (AML) algorithm imposes a strict requirement on the number of selected homogeneous training data (sample support). Specifically, it is necessary that the sample support is at least equal to the system DOFs in order to ensure the non-singularity of the estimated covariance matrix. Unfortunately, in the real world, the number of homogeneous secondary data is often limited, which may lead to a performance degradation or failure of the ML (AML) method. To circumvent this drawback, in this paper the generalized regularized likelihood function (GRLF) criterion is introduced to estimate the outlier subset. In particular, a suitable regularization function necessary to ensure well conditioned estimates is used. Moreover, like the AML technique, an approximate regularized ML (ARML) procedure is also devised to reduce the computational burden.

With reference to the selection of the regularization parameter, two distinct techniques are employed: the expected likelihood (EL) [35], [36] and the cross-validation (CV) [37], [38]. The EL procedure seeks for the specified regularized covariance matrix which provides the likelihood ratio (LR) value consistent with what is expected from the true covariance matrix. The CV approach starts with partitioning the secondary dataset into two subsets. The former is used as training to derive the outlier subset estimate and the corresponding regularized ML (RML) interference covariance estimate whereas the latter plays the role of a validation set to assess the risk associated with the obtained covariance value.

Summarizing, the main contributions of this paper are:

- the introduction of a novel technique to estimate the outlier subset based on the generalized regularized likelihood function (GRLF) criterion in order to circumvent the lack of a conspicuous amount of secondary data;
- the exploitation of two techniques for the selection of the regularization parameter for the specific problem, i.e., EL and CV;
- the evidence of interesting performance gains of the new procedures over some counterparts available in the open literature in terms of capability of correctly excise outliers.

The remainder of the paper is organized as follows. Section II is devoted to the design of the regularized outlier censoring scheme while Section III focuses on the choice of the regularization parameter. In Section IV, the property of the proposed algorithm is investigated whereas in Section V the performance assessment is conducted. Finally, concluding remarks and future research are discussed in Section VI.

Notation: In the sequel, vectors and matrices are denoted by boldface lower-case and upper-case letters, respectively. Symbols $\det(\cdot)$, $\mathrm{Tr}(\cdot)$, $(\cdot)^\dagger$, $(\cdot)^{-1}$ denote the determinant, trace, conjugate transpose, and inverse, respectively. Given a vector $\boldsymbol{a}$, $\mathbf{diag}(\boldsymbol{a})$ indicates the diagonal matrix whose ith diagonal element is the ith entry of $\boldsymbol{a}$. Given a matrix $\boldsymbol{R}$ of size $N \times N$, $\lambda_i(\boldsymbol{R})$, $i = 1, \cdots, N$, denotes the eigenvalues of $\boldsymbol{R}$. For a finite set A, $|A|$ stands for its cardinality. Finally, the symbol $\mathbb{E}[\cdot]$ denotes the statistical expectation.

II. DESIGN OF THE REGULARIZED OUTLIER CENSORING SCHEME

The aim of this section is to design an outlier censoring scheme resorting to the GRLF criterion. To this end, assume the same model as [33]: a radar system collects data from N channels (spatial and/or temporal) and the returns from K range cells are sampled. Denote by $\boldsymbol{x}_i$, $i = 1, \cdots, K$, the N-dimensional vector of the samples from the ith range cell and suppose that

$$\begin{cases} \boldsymbol{x}_i = \boldsymbol{c}_i, & \forall i \in \Omega - \Omega_0 \\ \boldsymbol{x}_i = \boldsymbol{c}_i + \boldsymbol{p}_i, & \forall i \in \Omega_0 \end{cases} \tag{1}$$

where $\boldsymbol{c}_1, \cdots, \boldsymbol{c}_K$, are independent, circular, zero-mean, complex Gaussian random vectors sharing the same covariance matrix $\boldsymbol{R}$ (assumed to be positive definite and modeled as an unknown deterministic parameter), representing the homogeneous interference components; $\Omega = \{1, \cdots, K\}$ is a set of size K which contains the indices of all the considered range cells; $\Omega_0 = \{i_1, \cdots, i_M\}$ is a subset of Ω with distinct elements and of size M ($M < K$), which contains the indices of the non-homogeneous range cells (see Fig. 1, the range cells in red represent the non-homogeneous samples), and $\boldsymbol{p}_i$s are non-homogeneous interference components, representing outliers, and modeled at design stage as deterministic and unknown vectors,[2] which are introduced by varying terrain, clutter discretes, moving targets and so forth.

The goal of the outlier censoring procedure is to estimate the subset Ω_0 and to excise from the data $\boldsymbol{x}_i$, $i = 1, \cdots, K$, the vectors whose indices belong to the estimated outlier subset. Toward this goal, exploiting the statistical independence of $\boldsymbol{c}_1, \cdots, \boldsymbol{c}_K$, in conjunction with the circular complex Gaussian property, the regularized negative log-likelihood function (NLLF) of $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_K$ can be expressed as

$$\begin{aligned} g(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_K) = {} & \ln\left[\det(\boldsymbol{R})\right] \\ & + \frac{1}{K} \sum_{i \in \Omega_0} \left[(\boldsymbol{x}_i - \boldsymbol{p}_i)^\dagger \boldsymbol{R}^{-1} (\boldsymbol{x}_i - \boldsymbol{p}_i) \right] \\ & + \frac{1}{K} \sum_{i \in \Omega - \Omega_0} \boldsymbol{x}_i^\dagger \boldsymbol{R}^{-1} \boldsymbol{x}_i + \alpha \mathrm{Tr}(\boldsymbol{R}^{-1}) \end{aligned} \tag{2}$$

[2]Notice that, the adopted model does not preclude to simulate the outliers at the analysis stage either as deterministic or as random vectors with specific covariance properties different from those of the other homogeneous data. By doing so a variety of covariance non-homogeneities can be accounted for.

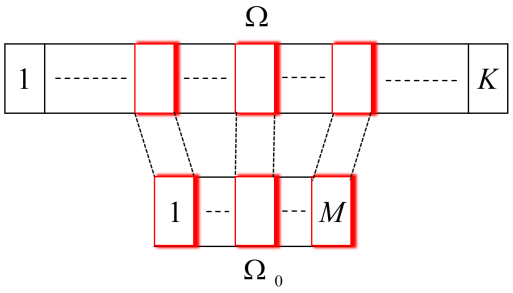

FIGURE 1. Diagram showing Ω and Ω_0.

where $\alpha > 0$ is a regularization (or penalty) parameter and $\mathrm{Tr}(\boldsymbol{R}^{-1})$ is the chosen regularization function. Since $\mathrm{Tr}(\boldsymbol{R}^{-1}) = \sum_{i=1}^{N} \frac{1}{\lambda_i(\boldsymbol{R})}$, the adopted regularization function restricts $\frac{1}{\lambda_i(\boldsymbol{R})}$ from becoming unbounded and thus shrinks the estimate toward a well-conditioned covariance matrix.

According to the GRLF criterion, the regularized maximum likelihood (RML) estimate of Ω_0 is the solution to the following problem

$$\widehat{\Omega}_0(\alpha) = \arg\min_{\Omega_0}[\min_{\boldsymbol{R}} \min_{\boldsymbol{p}_{i_1},\cdots,\boldsymbol{p}_{i_M}} g(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_K)] \tag{3}$$

where $\arg\min_{\Omega_0}(\cdot)$ denotes a set among the $\binom{K}{M} = \frac{K!}{M!(K-M)!}$ subsets of Ω with distinct elements and of size M which optimizes the argument.

Minimizing (2) over $\boldsymbol{p}_i$, $\forall i \in \Omega_0$, yields

$$\min_{\boldsymbol{p}_{i_1},\cdots,\boldsymbol{p}_{i_M}} g(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_K) = \ln[\det(\boldsymbol{R})] + \frac{1}{K}\sum_{i\in\Omega-\Omega_0} \boldsymbol{x}_i^{\dagger}\boldsymbol{R}^{-1}\boldsymbol{x}_i + \alpha\mathrm{Tr}(\boldsymbol{R}^{-1}), \tag{4}$$

observing that the minimum of each non-negative quadratic form is achieved for $\boldsymbol{p}_i = \boldsymbol{x}_i$.

Optimizing (4) over $\boldsymbol{R}$ yields

$$\min_{\boldsymbol{R}} \min_{\boldsymbol{p}_{i_1},\cdots,\boldsymbol{p}_{i_M}} g(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_K) = \ln\left[\det(\widetilde{\boldsymbol{R}})\right] + N \tag{5}$$

where $\widetilde{\boldsymbol{R}} = \frac{1}{K}\sum_{i\in\Omega-\Omega_0} \boldsymbol{x}_i\boldsymbol{x}_i^{\dagger} + \alpha\boldsymbol{I}$.

As a result, the solution to (3) is tantamount to solving the following problem

$$\widehat{\Omega}_0(\beta) = \arg\min_{\Omega_0}\left[\det(\widehat{\boldsymbol{R}})\right] \tag{6}$$

where $\widehat{\boldsymbol{R}} = \frac{1}{K-M}\sum_{i\in\Omega-\Omega_0} \boldsymbol{x}_i\boldsymbol{x}_i^{\dagger} + \beta\boldsymbol{I}$ is the regularized sample covariance matrix (RSCM) with regularization parameter $\beta = \frac{K}{K-M}\alpha > 0$. As to the choice of M, we may exploit some a-priori information to obtain an upper-bound (see [33] for more details on this issue).

Problem (6) is a combinatorial optimization problem which requires screening among all the possible $\binom{K}{M}$ subsets of size M and selecting the one leading to the lowest regularized sample covariance determinant. The computational burden connected with the exhaustive search could be prohibitive, especially when K is large. It is thus of interest developing approximate procedures characterized by a more affordable computational load and at the same time ensuring good quality solutions. In this respect, inspired by [23], [33], [34], we give the following theorem, which represents the theoretical foundation for the development of an effective regularized outlier censoring scheme.

Theorem 1. Consider a dataset $\boldsymbol{X} = \{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_K\}$ containing K random complex vectors of size N. Let $H_1 \subset \{1,\cdots,K\}$ with $|H_1| = h$ $(1 \le h \le K)$ and evaluate the RSCM $\boldsymbol{S}_1 = (1/h)\sum_{i\in H_1} \boldsymbol{x}_i\boldsymbol{x}_i^{\dagger} + \beta\boldsymbol{I}$ $(\beta > 0)$. Then compute the GIP values based on $\boldsymbol{S}_1$ for all the data

$$\boldsymbol{d}_1(i) = \boldsymbol{x}_i^{\dagger}\boldsymbol{S}_1^{-1}\boldsymbol{x}_i, \quad \text{for} \quad i = 1,\cdots,K \tag{7}$$

Now take H_2 such that

$$\{\boldsymbol{d}_1(i);\, i \in H_2\} = \{(\boldsymbol{d}_1)_{1:K},\cdots,(\boldsymbol{d}_1)_{h:K}\} \tag{8}$$

where $(\boldsymbol{d}_1)_{1:K} \le (\boldsymbol{d}_1)_{2:K} \le \cdots \le (\boldsymbol{d}_1)_{K:K}$ are the ordered GIP values, and compute the corresponding RSCM $\boldsymbol{S}_2 = (1/h)\sum_{i\in H_2} \boldsymbol{x}_i\boldsymbol{x}_i^{\dagger} + \beta\boldsymbol{I}$. Then

$$\det(\boldsymbol{S}_2) \le \det(\boldsymbol{S}_1) \tag{9}$$

with equality if and only if $\boldsymbol{S}_2 = \boldsymbol{S}_1$.

The proof is an extention of those given in [23] and [33].

The procedure in Theorem 1 is referred to as concentration-step (C-step) since it is possible to obtain a more concentrated RSCM $\boldsymbol{S}_2$ (i.e. sharing a lower determinant than the initial RSCM $\boldsymbol{S}_1$). Exploiting this C-step iteratively, a sequence of secondary datasets (with cardinality h) characterized by a non-increasing regularized sample covariance determinant can be obtained. More precisely, given the cardinality-h subset H_{old} and the corresponding RSCM $\boldsymbol{S}_{old}$, the generic iteration of the procedure involves three steps:

1. compute the GIP values using $\boldsymbol{S}_{old}$ for all the data $\boldsymbol{d}_{old}(i) = \boldsymbol{x}_i^{\dagger}\boldsymbol{S}_{old}^{-1}\boldsymbol{x}_i$, $i = 1,\cdots K$;
2. sort them in ascending order and select the indices corresponding to the lowest h GIP values to construct the new cardinality-h subset H_{new};
3. compute the new RSCM $\boldsymbol{S}_{new} = (1/h)\sum_{i\in H_{new}} \boldsymbol{x}_i\boldsymbol{x}_i^{\dagger} + \beta\boldsymbol{I}$.

Since there are only finitely many cardinality-h subsets, the iterative algorithm must converge, and the stopping criterion can be set as $\det(\boldsymbol{S}_m) = \det(\boldsymbol{S}_{m-1})$ or a maximum number of C-steps N_{cstep} is reached.[3] It is finally worth pointing out that there is no guarantee that the algorithm converges to the global optimum of the minimal regularized

[3]Simulation examples have highlighted that a quite small number of C-steps is necessary to achieve $\det(\boldsymbol{S}_m) = \det(\boldsymbol{S}_{m-1})$ and in all the developed analysis we set $N_{cstep} = 5$.

sample covariance determinant problem. However, if we take more distinct initial subsets, apply C-steps on each subset until convergence and select among the output solutions that leading to the lowest regularized sample covariance determinant, the probability of falling in the global optimum solution will increase. This is the basic idea behind the proposed ARML method. As to the selection of initial subsets, we could choose randomly distinct cardinality-h subsets. Summarizing, the pseudocode of the ARML method is summarized in Algorithm 1.

Algorithm 1 : Pseudocode of the ARML Method

1: Construct $N_{initial}$ distinct cardinality-h subsets and for each of them carry out C-steps until the stopping criterion is reached, obtaining $N_{initial}$ candidate solutions;
2: Report the subset with the lowest regularized sample covariance determinant chosen among the $N_{initial}$ candidates.

The ARML procedure can allow for a quite significant reduction in computational complexity as compared with the corresponding RML counterpart. Specifically, for a given regularization parameter, the overall computational complexity of the ARML procedure is approximately $O(\max(K,N)N_{initial}N_{cstep}N^2)$ floating point operations per seconds (FLOPs) whereas that of the exact RML procedure is $O\left(\max(K,N)\binom{K}{M}N^2\right)$ FLOPs, where $O(n)$ indicates that an operation requires a number of FLOPs proportional to n.

III. CHOICE OF THE REGULARIZATION PARAMETER

In order to benefit from the RML (ARML) method, it is vital to choose an appropriate regularization parameter β. A reasonable choice for β could be the value which minimizes the outlier error selection probability. However, since there is no a-priori knowledge about the exact outlier subset, it is not possible in general to exploit the aforementioned criterion. In order to circumvent this drawback and come up with practically implementable outlier selectors, in the following the use of the EL and CV techniques [35]–[38] is suggested to select the regularization parameter β.

A. EL SELECTION TECHNIQUE

The EL method aims to find a parametric (regularized) covariance matrix which generates a specific likelihood ratio (LR) value consistent with what is expected from the true covariance matrix [35], [36]. The basic principle behind this technique relies on the invariance property of the LR, namely, for the true covariance matrix, the probability density function (pdf) of the LR does not depend on this true matrix. Precisely, consider T iid N-dimensional circular, zero-mean, complex Gaussian observations sharing the same positive definite covariance matrix $\boldsymbol{R}$, $\boldsymbol{Z}_T = \{z_1, \cdots, z_T\}$, the LR of the regularized covariance matrix $\boldsymbol{R}(\beta)$ is given by[4]

$$LR(\boldsymbol{Z}_T, \boldsymbol{R}(\beta)) = \frac{\prod_{j=1}^{\min(T,N)} \lambda_j\left(\boldsymbol{R}(\beta)^{-1}\widehat{\boldsymbol{R}}_T\right)\exp(\min(T,N))}{\exp\left[\sum_{j=1}^{\min(T,N)} \lambda_j\left(\boldsymbol{R}(\beta)^{-1}\widehat{\boldsymbol{R}}_T\right)\right]} \tag{10}$$

where $\widehat{\boldsymbol{R}}_T = (1/T)\boldsymbol{Z}_T\boldsymbol{Z}_T^\dagger$ is the SCM and $\lambda_j\left(\boldsymbol{R}(\beta)^{-1}\widehat{\boldsymbol{R}}_T\right)$ denotes the jth largest eigenvalue of the matrix $\boldsymbol{R}(\beta)^{-1}\widehat{\boldsymbol{R}}_T$. As proved in [35], [36], when the regularized covariance matrix $\boldsymbol{R}(\beta)$ coincides with the true covariance $\boldsymbol{R}$, the pdf of the LR ($LR(\boldsymbol{Z}_T, \boldsymbol{R})$) does not depend on $\boldsymbol{R}$ and are specified by the parameters N and T only (scenario free situation).

In the above formulation, it is required that the observations $\boldsymbol{Z}_T$ should be outlier-free and iid. To this end, we first select T $(1 \leq T \leq K)$ samples from the secondary dataset $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_K\}$ using the RML (ARML) procedure with an initial regularization parameter β_0 and denote by $\Phi_0 = \{i_1, \cdots, i_T\}$ the set of indices for the selected samples. The regularized covariance matrix is then given by

$$\boldsymbol{R}(\beta) = \frac{1}{T}\sum_{i\in\Phi_0} \boldsymbol{x}_i\boldsymbol{x}_i^\dagger + \beta\boldsymbol{I}. \tag{11}$$

According to the EL principle, the optimal regularization parameter is the solution to the following problem

$$\widehat{\beta} = \arg\left\{LR\left(\boldsymbol{x}_{i_1}, \cdots, \boldsymbol{x}_{i_T}, \boldsymbol{R}(\beta)\right) = LR_0\right\}, \tag{12}$$

where LR_0 is selected by referring to the scenario-free pdf. For example, the mean or median value of that pdf could be chosen. More precisely, LR_0 is a-priori computed as the value corresponding to the mean, median, or other percentiles of the scenario-free pdf. For more details see also references [35], [36].

To possibly improve the result, we could substitute the estimated $\widehat{\beta}$ for the initial β_0 and repeat the above procedure iteratively. The motivation behind the iteration is that we might choose more qualified data with the estimated $\widehat{\beta}$, namely, the probability that the selected T samples do not contain outliers might increase. In addition, the iteration procedure can be stopped when $\widehat{\beta}_m = \widehat{\beta}_{m-1}$ or when the number of iterations reach a pre-set value N_{iter}. Simulation results have highlighted that the performance of the RML (ARML) procedure based on the EL technique, which is evaluated in terms of correctly censoring all the outliers, is not very sensitive to the parameter β_0 and approximately achieves satisfactory performance when $K - M \leq T \leq K$ (based on this observation in the sequel, we set $T = K$).

B. CV SELECTION TECHNIQUE

CV technique relies on partitioning the original secondary dataset into two subsets, one used as training set to estimate the unknown parameters and the other used as validation set to establish the risk associated with the training choice.

[4]Note that this expression can be applied for both the cases $T \geq N$ [35] and $T < N$ [36].

To implement the CV procedure, it is necessary specifying the selection of the risk function and the choice of the partitioning strategy. The likelihood of the validation data (evaluated using the training-based interference covariance estimate) could be selected as the risk function. Moreover, just for computational reasons, the negative logarithmic of the mentioned likelihood can be equivalently chosen.

Like the EL method, we require that the validation set is outlier-free. To this end, we first select L_1 $(1 < L_1 \leq K - M)$ samples from the secondary dataset using the RML (ARML) method with a fixed regularization parameter β_1 and denote by $\Phi_1 = \{i_1, \cdots, i_{L_1}\}$ the set of indices for the corresponding vectors, which are indeed chosen as representative. As to the choice of β_1, we will discuss it later in subsection III-B.3.

1) CHOICE OF THE PARTITIONING STRATEGY AND CV PROCEDURE

We adopt the Q_1-fold cross-validation strategy [38]. Specifically, we divide Φ_1 into Q_1 $(1 < Q_1 \leq L_1)$ nonoverlapping groups such that $\Phi_1 = \bigcup_{q=1}^{Q_1} \Phi_{1_q}$ with N_{1_q} denoting the number of indices in the qth group (N_{1_q}s, $q = 1, \cdots, Q_1$, are often selected to be approximately equal. For the qth group, $\boldsymbol{X}_{1_q} = [\boldsymbol{x}_i, i \in \Phi_{1_q}]$ is used as validation set whereas all the remaining data $\boldsymbol{X} - \boldsymbol{X}_{1_q} = [\boldsymbol{x}_i, i \in \Omega - \Phi_{1_q}]$ are exploited as training set to estimate the interference covariance matrix, whose expression is given by

$$\widehat{\boldsymbol{R}}_{1_q}(\beta) = \frac{1}{K - N_{1_q} - M} \sum_{i \in \Omega - \Phi_{1_q} - \widehat{\Omega}_{1_q}(\beta)} \boldsymbol{x}_i \boldsymbol{x}_i^{\dagger} + \beta \boldsymbol{I}, \quad (13)$$

where $\widehat{\Omega}_{1_q}(\beta)$ is the estimated outlier subset of size M based on the training set $\boldsymbol{X} - \boldsymbol{X}_{1_q}$ for a specific regularization parameter β. As a consequence, the risk function for the qth group can be written as

$$f_{1_q}(\beta) = N_{1_q} \ln\left[\det\left(\widehat{\boldsymbol{R}}_{1_q}(\beta)\right)\right] + \mathrm{Tr}\left[\widehat{\boldsymbol{R}}_{1_q}^{-1}(\beta) \boldsymbol{X}_{1_q} \boldsymbol{X}_{1_q}^{\dagger}\right]. \quad (14)$$

The cross-validated risk function is defined as the average over all the Q_1 risks

$$g_1(\beta) = \frac{1}{Q_1} \sum_{q=1}^{Q_1} f_{1_q}(\beta) \ (\beta > 0), \quad (15)$$

and the optimal regularization parameter β is the solution to the following optimization problem

$$\widehat{\beta} = \arg \min_{\beta > 0} g_1(\beta). \quad (16)$$

Since we cannot claim neither the existence of the minimum nor we can obtain a closed form expression for $\widehat{\beta}$, we define a grid of values $\beta_p > 0, p = 1, \cdots, P_1$, evaluate (15) in correspondence of each β_p and select the β_p leading to the lowest value as optimizer. Otherwise stated, for a specific β_p, the interference covariance matrix estimate for the qth group in (13) is given by

$$\widehat{\boldsymbol{R}}_{1_q}(\beta_p) = \frac{1}{K - N_{1_q} - M} \sum_{i \in \Omega - \Phi_{1_q} - \widehat{\Omega}_{1_q}(\beta_p)} \boldsymbol{x}_i \boldsymbol{x}_i^{\dagger} + \beta_p \boldsymbol{I}, \quad (17)$$

where $\widehat{\Omega}_{1_q}(\beta_p)$ is the estimated outlier subset for β_p and for the qth group.

Finally, the optimal β is approximated by the value on the grid corresponding to the lowest cross-validated risk function,

$$\widehat{\beta} = \beta_{\widehat{p}}, \ \text{with} \ \widehat{p} = \arg \min_{p \in [1, \cdots, P_1]} g_1(\beta_p). \quad (18)$$

The result can be possibly improved replacing β_1 with the estimated $\widehat{\beta}$ and triggering an iterative procedure. The same stopping criterion as for the EL technique can be used. In conclusion, the pseudocode of the CV method is summarized in Algorithm 2.

Algorithm 2 : Pseudocode of the CV Method

1: Select L_1 $(1 < L_1 \leq K - M)$ samples from the secondary dataset using the RML (ARML) method with a fixed regularization parameter β_1 and denote by $\Phi_1 = \{i_1, \cdots, i_{L_1}\}$ the set of indices for the selected vectors.
2: Divide Φ_1 into Q_1 $(1 < Q_1 \leq L_1)$ nonoverlapping groups such that $\Phi_1 = \bigcup_{q=1}^{Q_1} \Phi_{1_q}$.
3: For each group, evaluate the risk function (14) for each β_p, $p = 1, \cdots, P_1$.
4: Average all the Q_1 risk functions to obtain the cross-validated risk function (15) for β_p, $p = 1, \cdots, P_1$.
5: Obtain the estimated regularization parameter $\widehat{\beta}$ selecting the specific β_p corresponding to the lowest cross-validated risk function.
6: Replace β_1 with $\widehat{\beta}$ and repeat steps 1-5 until the stopping criterion is satisfied.
7: Report the final $\widehat{\beta}$ as the selected regularization parameter.

2) REDUCED COMPLEXITY CV PROCEDURE

The main drawback of the above CV approach is the heavy computational burden because the RML (ARML) procedure should be applied $Q_1 \times P_1$ times to estimate the outlier subset for each group and β_p value. In order to circumvent this drawback, we can preliminarily estimate an outlier subset and use it for all the groups of data and for all the values of β. By doing so only one RML (ARML) procedure is necessary to estimate the outlier subset and this leads to a significant reduction of the overall complexity.

Formalizing, let us denote by $\widehat{\Omega}_2$ [5] an outlier subset of size M $(M < K)$ obtained exploiting the RML (ARML) procedure with a fixed regularization parameter β_2. [6] In addition, we select L_2 $(1 < L_2 \leq K - M)$ samples from the remaining data to form the outlier-free validation set also based on the RML (ARML) procedure with the same regularization parameter β_2 and denote by $\Phi_2 = \{i_1, \cdots, i_{L_2}\}$ the set of indices for the selected validation set. After that, like the CV method, we partition Φ_2 into Q_2 $(1 < Q_2 \leq L_2)$

[5] It is worth noting that when the outlier power is low or when the value of M is smaller than the true number of the outliers, the remaining data (i.e., data whose indices belong to $\Omega - \hat{\Omega}_2$) might still contain some outliers.

[6] The discussions on the selection of β_2 is provided in subsection III-B.3.

nonoverlapping groups such that $\Phi_2 = \bigcup_{q=1}^{Q_2} \Phi_{2_q}$, with N_{2_q} denoting the number of indices in the qth group. For the qth group, $\boldsymbol{X}_{2_q} = [\boldsymbol{x}_i, i \in \Phi_{2_q}]$ is used as validation set and the remaining data $\boldsymbol{X} - \boldsymbol{X}_{2_q} = [\boldsymbol{x}_i, i \in \Omega - \Phi_{2_q}]$ are exploited for training. The distinction between the modified and original CV method relies in the fact that the same $\hat{\Omega}_2$ is used for all the groups. As a consequence, the interference covariance matrix estimate for the qth group can be written as

$$\widehat{\boldsymbol{R}}_{2_q}(\beta) = \frac{1}{K - N_{2_q} - M} \sum_{i \in \Omega - \widehat{\Omega}_2 - \Phi_{2_q}} \boldsymbol{x}_i \boldsymbol{x}_i^{\dagger} + \beta \boldsymbol{I}. \quad (19)$$

The corresponding risk function for the qth group and the cross-validated risk function are given by

$$f_{2_q}(\beta) = N_{2_q} \ln\left[\det\left(\widehat{\boldsymbol{R}}_{2_q}(\beta)\right)\right] + \mathrm{Tr}\left[\widehat{\boldsymbol{R}}_{2_q}^{-1}(\beta) \boldsymbol{X}_{2_q} \boldsymbol{X}_{2_q}^{\dagger}\right], \quad (20)$$

$$g_2(\beta) = \frac{1}{Q_2} \sum_{q=1}^{Q_2} f_{2_q}(\beta), \quad (21)$$

respectively, and the optimal β is the solution to the optimization problem

$$\widehat{\beta} = \arg\min_{\beta > 0} g_2(\beta). \quad (22)$$

In order to get some insights toward the solution of (22), let us denote by

$$\boldsymbol{S}_{2_q} = \frac{1}{K - N_{2_q} - M} \sum_{i \in \Omega - \widehat{\Omega}_2 - \Phi_{2_q}} \boldsymbol{x}_i \boldsymbol{x}_i^{\dagger} \quad (23)$$

the SCM for the qth group and $\boldsymbol{S}_{2_q} = \boldsymbol{U}_{2_q} \boldsymbol{\Lambda}_{2_q} \boldsymbol{U}_{2_q}^{\dagger}$ its eigenvalue decomposition, where $\boldsymbol{U}_{2_q} = [\boldsymbol{u}_{2_q,1}, \cdots, \boldsymbol{u}_{2_q,N}]$ is a unitary matrix of the eigenvectors, $\boldsymbol{\Lambda}_{2_q} = \mathbf{diag}\,(\boldsymbol{r}_{2_q})$ is the diagonal matrix of the corresponding eigenvalues $\boldsymbol{r}_{2_q} = [r_{2_q,1}, \cdots, r_{2_q,N}]$. Thus, the objective function of (22) can be recast as

$$\begin{aligned} g_3(\beta) &= \frac{1}{Q_2} \sum_{q=1}^{Q_2} N_{2_q} \ln\left[\det(\boldsymbol{S}_{2_q} + \beta \boldsymbol{I})\right] \\ &\quad + \mathrm{Tr}\left[(\boldsymbol{S}_{2_q} + \beta \boldsymbol{I})^{-1} \boldsymbol{X}_{2_q} \boldsymbol{X}_{2_q}^{\dagger}\right] \\ &= \frac{1}{Q_2} \sum_{q=1}^{Q_2} N_{2_q} \left\{ \sum_{i=1}^{N} \left[\ln(r_{2_q,i} + \beta) + \left(\frac{c_{2_q,i}}{r_{2_q,i} + \beta} \right) \right] \right\}, \end{aligned} \quad (24)$$

where $c_{2_q,i} = \frac{1}{N_{2_q}} \boldsymbol{u}_{2_q,i}^{\dagger} \boldsymbol{X}_{2_q} \boldsymbol{X}_{2_q}^{\dagger} \boldsymbol{u}_{2_q,i}$, $q = 1, \cdots, Q_2$, $i = 1, \cdots, N$.

Now, let us focus on $g_3(\beta)$ and observe that it is continuous and differentiable on (0,+∞). Its derivative is given by

$$\begin{aligned} g_4(\beta) &= \frac{dg_3(\beta)}{d\beta} \\ &= \frac{1}{Q_2} \sum_{q=1}^{Q_2} N_{2_q} \sum_{i=1}^{N} \left[\frac{1}{r_{2q,i} + \beta} - \frac{c_{2q,i}}{(r_{2q,i} + \beta)^2} \right]. \end{aligned} \quad (25)$$

It is clear that $\lim_{\beta \to +\infty} g_3(\beta) \to +\infty$ and, as a consequence, if the minimum of (22) exists it is achieved for a finite value of β to be searched among the real positive stationary points, namely it is a real positive root of $g_4(\beta) = 0$.

To find the roots of $g_4(\beta) = 0$, it is necessary to observe that the equation (25) can be recast as a polynomial equation of degree $2NQ_2 - 1$. Typically, there are no closed-form expressions for the roots of the above equation unless $2NQ_2 \leq 5$ (Abel-Ruffini theorem [39]). When $2NQ_2 > 5$ (which is the practical situation), roots can be only computed using numerical algorithms. Nevertheless, it is always possible to provide an upper-bound on the real positive roots of the equation (which can be helpful in restricting the search interval for a numerical procedure) resorting to the following theorem [40]:

Cauchy's theorem: Denote by $p(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_k x^{(n-k)} + \cdots + a_{n-1} x + a_n$, $(a_0 > 0)$ a polynomial of degree $n > 0$. If N_{neg} is the number of negative coefficients and is greater than 0, then an upper-bound on the values of the real positive roots of $p(x)$ is given by

$$u_b = \max_{1 \leq k \leq n : a_k < 0} \sqrt[k]{-\frac{N_{neg} a_k}{a_0}} \quad (26)$$

If $N_{neg} = 0$, there are no real positive roots.

According to the above theorem, it is not only possible to determine whether or not the real positive roots of equation (25) exist, but one can also estimate an upper-bound on them as long as they exist. Once the upper-bound is available, one could also get an approximate solution to problem (22) resorting to the discretization technique. Specifically, one can sample uniformly the interval (0,u_b] to obtain a grid of P_2 points and select that minimizing the function g_3. In addition, when $N_{neg} = 0$, the smallest β ($\beta > 0$) value on the search grid would lead to the minimum of g_3 on the search grid. Under this situation, $\widehat{\beta} = 0.01$ is set.

To improve the result, the estimated $\hat{\beta}$ can also be used to replace β_2 thus triggering an iterative procedure. Moreover, the same stopping criterion as for the original CV method can be exploited. Summarizing, the pseudocode of the modified CV technique is reported in Algorithm 3.

3) PERFORMANCE ISSUES

Some considerations are now provided based on extensive numerical experiments.

- The performance of the CV method is not very sensitive to the value of the initial regularization parameter β_1, thus it could be selected randomly and the step 6 in Algorithm 2 is no longer required (in the subsequent simulations, we set $\beta_1 = 1$ and $N_{iter} = 0$);
- As to the selection of L_1, the best performance is normally ensured when L_1 is equal to or a bit smaller than $K - M$. Furthermore, simulations highlight that when step 1 in Algorithm 2 is omitted, namely, all the secondary data (including the data contaminated by outliers) are exploited as the candidate validation data,

Algorithm 3 : Pseudocode of the Modified CV Method

1: Estimate an outlier subset $\widehat{\Omega}_2$ of size M using the RML (ARML) procedure with a fixed regularization parameter β_2.
2: Select L_2 $(1 < L_2 \leq K - M)$ samples from the data whose indices belong to $\Omega - \widehat{\Omega}_2$ exploiting the RML (ARML) procedure with β_2 and denote by $\Phi_2 = \{i_1, \cdots, i_{L_2}\}$ the set of indices for the selected vectors.
3: Partition Φ_2 into Q_2 $(1 < Q_2 \leq L_2)$ nonoverlapping groups such that $\Phi_2 = \bigcup_{q=1}^{Q_2} \Phi_{2_q}$.
4: For each group, estimate the SCM and perform the eigenvalue decomposition.
5: Construct the polynomial equation corresponding to equation (25), then expand it and count the number of negative coefficients N_{neg} in the polynomial.
6: If $N_{neg} > 0$, then estimate an upper-bound on the real positive roots of equation (25) resorting to Cauchy's theorem and evaluate $\widehat{\beta}$ exploiting the finite grid search method. Otherwise, $\widehat{\beta} = 0.01$.
7: Replace β_2 with $\widehat{\beta}$ and repeat steps 1-6 until the stopping criterion is satisfied.
8: Report the final $\widehat{\beta}$ as the selected regularization parameter.

the behavior degrades slightly (in the sequel, we set $L_1 = K - M$);

- In principle, there is no effective method to determine an upper-bound or a relatively narrow interval of the optimal β for the CV method and, hence, to obtain the approximate global minimum of problem (16), it is necessary to search among all the possible values of β in the range (0,+∞), which is a prohibitive task. Thus, if there is no a-priori knowledge about the range of the possible β values, the CV method is hardly to implement. Nevertheless, simulations show that a proper range of β for the CV method is (0,10] for the considered experiments, which is adopted in the sequel;
- Simulations indicate that $P_1 = 11$ and $Q_1 = 4$ represent a good compromise between the performance and computational complexity for the reported experiments. Precisely, when $P_1 > 11$ and $Q_1 > 4$, the performance almost keeps the same level;
- Like the CV method, the behavior of the modified CV method is not too much sensitive to the value of the initial regularization parameter β_2. As a result, we also set $\beta_2 = 1$ and $N_{iter} = 0$;
- Simulations highlight that the best performance of the modified CV method is often achieved when $L_2 = K - M$, as a consequence, this value is adopted in the sequel.
- Like the CV method, P_2 and Q_2 should be selected properly in order to realize a reasonable tradeoff between performance and computational complexity. According to simulation results, $P_2 = 41$ and $Q_2 = 4$ are good choices.

IV. PROPERTY OF THE RML (ARML) PROCEDURE

An interesting property of the above RML (ARML) procedure is the unitary invariance (i.e. invariance with respect to unitary transformations of the data). Precisely, let $\boldsymbol{Y} = \boldsymbol{BX}$, where $\boldsymbol{Y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_K]$ is the transformed data matrix and $\boldsymbol{B}$ is an arbitrary $N \times N$ unitary matrix. If $\widehat{\Omega}_0$ is the solution to (6), then it also solves:

$$\widehat{\Omega}_0 = \arg\min_{\Omega_0}[\det(\boldsymbol{R}_Y)], \tag{27}$$

where $\boldsymbol{R}_Y = \dfrac{1}{K-M}\sum_{i\in\Omega-\Omega_0}\boldsymbol{y}_i\boldsymbol{y}_i^\dagger + \beta\boldsymbol{I}$.

In fact, for a specific Ω_0, denote by $\boldsymbol{X}_0 = \{\boldsymbol{x}_i | \boldsymbol{x}_i \in \Omega - \Omega_0\}$, $\boldsymbol{R}_X = \dfrac{1}{K-M}\boldsymbol{X}_0\boldsymbol{X}_0^\dagger + \beta\boldsymbol{I}$, $\boldsymbol{Y}_0 = \boldsymbol{BX}_0 = \{\boldsymbol{y}_i | \boldsymbol{y}_i \in \Omega - \Omega_0\}$, and $\boldsymbol{R}_Y = \dfrac{1}{K-M}\boldsymbol{Y}_0\boldsymbol{Y}_0^\dagger + \beta\boldsymbol{I}$, then

$$\begin{aligned}\boldsymbol{R}_Y &= \frac{1}{K-M}\boldsymbol{Y}_0\boldsymbol{Y}_0^\dagger + \beta\boldsymbol{I} = \boldsymbol{B}\boldsymbol{R}_X\boldsymbol{B}^\dagger,\\ \det(\boldsymbol{R}_Y) &= \det(\boldsymbol{B}\boldsymbol{R}_X\boldsymbol{B}^\dagger) = \det(\boldsymbol{R}_X).\end{aligned} \tag{28}$$

Therefore, for a given regularization parameter β, an unitary transformation of the data gives rise to the same RML estimate of the outlier subset.

As to the ARML procedure, let Ω_1 be the initial subset of size $(K - M)$, $\boldsymbol{X}_1 = \{\boldsymbol{x}_i | \boldsymbol{x}_i \in \Omega_1\}$ and $\boldsymbol{Y}_1 = \{\boldsymbol{y}_i | \boldsymbol{y}_i \in \Omega_1\}$ the corresponding transformed data, then the initial RSCMs are given by

$$\begin{aligned}\boldsymbol{R}_{X,1} &= \frac{1}{K-M}\boldsymbol{X}_1\boldsymbol{X}_1^\dagger + \beta\boldsymbol{I},\\ \boldsymbol{R}_{Y,1} &= \frac{1}{K-M}\boldsymbol{Y}_1\boldsymbol{Y}_1^\dagger + \beta\boldsymbol{I},\end{aligned} \tag{29}$$

and the corresponding GIP values exhibit the following property

$$\boldsymbol{y}_i^\dagger\boldsymbol{R}_{Y,1}^{-1}\boldsymbol{y}_i = \boldsymbol{x}_i^\dagger\boldsymbol{R}_{X,1}^{-1}\boldsymbol{x}_i, \quad i = 1, \cdots, K. \tag{30}$$

Combined with (28), it follows that given the regularization parameter, the ARML is unitary invariant.

As to the selection of the regularization parameter, it can be proved that using the unitary invariance of the eigenvalues, the same regularization parameter will be selected after the unitary transformation of the data for both the EL and the CV techniques.

V. RESULTS ON SIMULATED DATA

In this section, we evaluate the performance of the RML and ARML methods, also in comparison with the GIP [18], [21], RCGIP [22], [41], AML [33], [34], and exact ML [33], [34] algorithms. For notational simplicity, we refer to the RML procedure exploiting the EL and unmodified/modified CV techniques as EL-RML and U-RML/M-RML, respectively. All the outlier subset estimation procedures involved in the U-RML and M-RML algorithms are implemented using the RML technique. The EL-ARML and M-ARML are defined in a similar way, where the ARML procedure is applied for all the steps associated with the outlier subset estimation.

We consider a Doppler processing and exploit the following interference covariance matrix model [19], [20]

$$\boldsymbol{R} = \boldsymbol{R}_c + \sigma_n^2 \boldsymbol{I} \tag{31}$$

where $\boldsymbol{R}_c$ and $\boldsymbol{I}$ account for the clutter and thermal noise component, respectively, and σ_n^2 is the thermal noise power, assumed in the following, without loss of generality, equal to 0 dB. Besides, the clutter covariance matrix is assumed exponentially-shaped [19], [20],

$$\boldsymbol{R}_c(i,j) = \sigma_c^2 \rho^{|i-j|} e^{j2\pi f_{d_c}(i-j)}, \quad i,j = 1, \cdots, N \tag{32}$$

where σ_c^2 is the clutter power, ρ is the one-lag correlation coefficient of the clutter, and f_{d_c} is the normalized clutter Doppler frequency (in the subsequent simulations, $\sigma_c^2 = 20$ dB, $\rho = 0.95$, and $f_{d_c} = 0$).

Moreover, M outliers are randomly injected into distinct vectors of the secondary dataset. The temporal steering vectors of these outliers are given by

$$\boldsymbol{p}_i = \alpha_i [1, e^{j2\pi f_{d_{o,i}}}, \cdots, e^{j2\pi(N-1)f_{d_{o,i}}}]^T, \quad i = 1, \cdots, M \tag{33}$$

where $f_{d_{o,i}}$ is the normalized Doppler frequency of the ith outlier and $\alpha_1, \cdots, \alpha_M$ are independent, zero-mean, complex Gaussian random variables, representing the complex amplitudes. In the sequel, the outliers are assumed to share the same power, namely, σ_o^2. As to the Doppler frequencies, two situations are considered. The former assumes $f_{d_{o,i}} = 0.15$, $i = 1, \cdots, M$, namely, equal Doppler outliers. In fact, this situation might arise when considering a scenario containing multiple outlying targets moving at the same radial velocity or considering an outlying target which occupies more than one range cell, i.e., the range-distributed target. The latter assumes that the normalized Doppler frequencies of the outliers are modeled as statistically independent random variables uniformly distributed within the interval [0.1,0.2] (random Doppler outliers). Finally, denote by M_o ($M_o < K$) the number of data vectors to be removed from the K snapshots (thus $h = K - M_o$).

The performance is assessed using as figure of merit the probability of correctly removing all the outliers, namely, $\Omega_0 \subseteq \widehat{\Omega}_0$, denoted in the following by P_r. Due to the lack of a closed-form expression for P_r, the analysis is conducted resorting to standard Monte Carlo (MC) simulations based on 1000 independent trials. Fig. 2 shows P_r versus the number of initial subsets $N_{initial}$ for $N = 8$, $M = 3$, $M_o = 3$, and equal Doppler outliers under different choices for K and σ_o^2. Specifically, four cases are considered, whose results are displayed in subplots (a)-(d), respectively: (a) $K = 12$ and $\sigma_o^2 = 15$ dB, which corresponds to a limited sample support scenario, namely, the number of homogeneous training data is less than twice the system DOFs; (b) $K = 12$ and $\sigma_o^2 = 30$ dB; (c) $K = 20$ and $\sigma_o^2 = 15$ dB, where the number of homogeneous training data exceeds twice the system DOFs; (d) $K = 20$ and $\sigma_o^2 = 30$ dB. The plots highlight that the EL-RML, U-RML and M-RML methods almost achieve

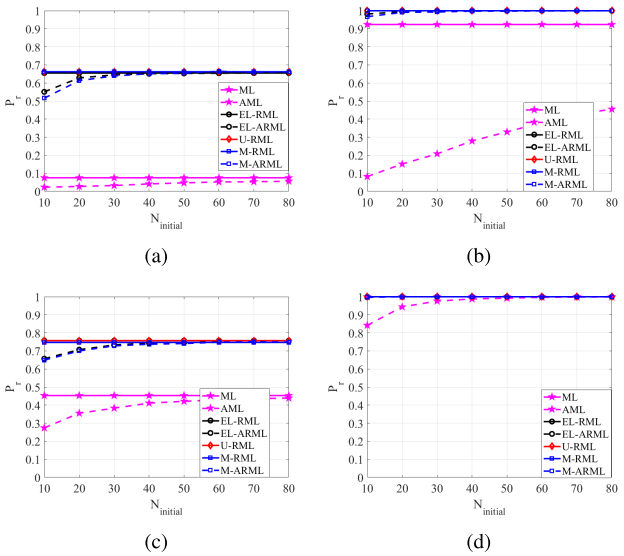

FIGURE 2. P_r versus $N_{initial}$ of the exact ML (pentagram-marked solid curve), AML (pentagram-marked dashed curve), EL-RML (circle-marked solid curve), EL-ARML (circle-marked dashed curve), U-RML (diamond-marked curve), M-RML (square-marked solid curve), and M-ARML (square-marked dashed curve) for $N = 8$, $M = 3$, $M_o = 3$, and equal Doppler outliers. (a) $K = 12$ and $\sigma_o^2 = 15$ dB. (b) $K = 12$ and $\sigma_o^2 = 30$ dB. (c) $K = 20$ and $\sigma_o^2 = 15$ dB. (d) $K = 20$ and $\sigma_o^2 = 30$ dB.

the same performance level and generally outperform the corresponding ML counterpart. Moreover, like the AML procedure, the performance of the ARML (EL-ARML and M-ARML) methods improves as $N_{initial}$ increases. However, for the selected parameters, the value of $N_{initial}$ required for the ARML to achieve the limit performance is much less than that for the AML, especially when the sample size is small and the outlier power is large.

In Fig. 3, P_r versus $N_{initial}$ is plotted for $N = 8$, $M = 6$, $M_o = 6$, and equal Doppler outliers under different choices for K and σ_o^2, namely, a relatively outlier dense scenario. We also consider four cases: (a) $K = 20$ and $\sigma_o^2 = 15$ dB; (b) $K = 20$ and $\sigma_o^2 = 30$ dB; (c) $K = 24$ and $\sigma_o^2 = 15$ dB; (d) $K = 24$ and $\sigma_o^2 = 30$ dB. The curves show that a larger number of $N_{initial}$ than in Fig. 2 is required for the ARML methods to achieve a satisfactory performance level, especially when the outlier power is relatively low. Specifically, for $\sigma_o^2 = 15$ dB, 80 initial subsets are not adequate for the ARML methods to ensure the best performance when $M = 6$ whereas 40 initial subsets are sufficient when $M = 3$. Usually, the EL-ARML method slightly outperforms the M-ARML method for the same value of $N_{initial}$.

Fig. 4 shows P_r versus $N_{initial}$ for $K = 20$, $N = 8$, $M = 6$, $M_o = 8$, equal Doppler outliers and some different values for σ_o^2. The plots highlight that the requirement for $N_{initial}$ to achieve a good performance level can be reduced increasing M_o. Nevertheless, even if P_r improves as M_o increases, the reduction on the number of selected training data might impose a deleterious impact on the estimation accuracy of the interference covariance matrix. In addition, the plots of

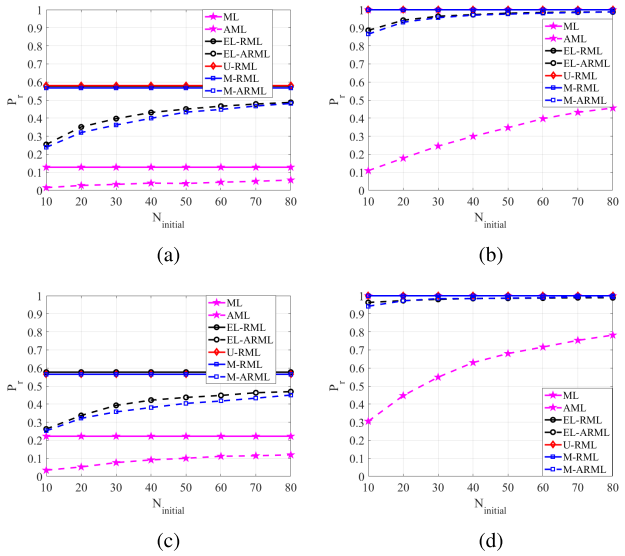

FIGURE 3. P_r versus $N_{initial}$ of the exact ML (pentagram-marked solid curve), AML (pentagram-marked dashed curve), EL-RML (circle-marked solid curve), EL-ARML (circle-marked dashed curve), U-RML (diamond-marked curve), M-RML (square-marked solid curve), and M-ARML (square-marked dashed curve) for $N = 8$, $M = 6$, $M_o = 6$, and equal Doppler outliers. (a) $K = 20$ and $\sigma_o^2 = 15$ dB. (b) $K = 20$ and $\sigma_o^2 = 30$ dB. (c) $K = 24$ and $\sigma_o^2 = 15$ dB. (d) $K = 24$ and $\sigma_o^2 = 30$ dB.

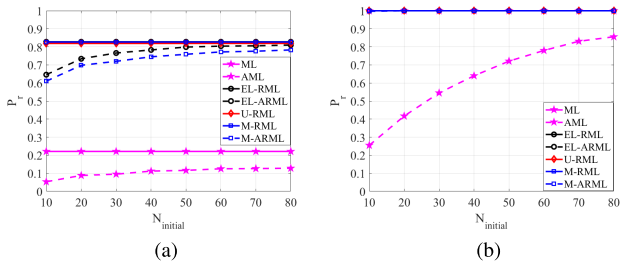

FIGURE 4. P_r versus $N_{initial}$ of the exact ML (pentagram-marked solid curve), AML (pentagram-marked dashed curve), EL-RML (circle-marked solid curve), EL-ARML (circle-marked dashed curve), U-RML (diamond-marked curve), M-RML (square-marked solid curve), and M-ARML (square-marked dashed curve) for $K = 20$, $N = 8$, $M = 6$, $M_o = 8$, and equal Doppler outliers. (a) $\sigma_o^2 = 15$ dB. (b) $\sigma_o^2 = 30$ dB.

P_r versus $N_{initial}$ exhibit a similar behavior when the Doppler frequencies of the outliers are different: these results are not reported here for brevity.

Fig. 5 shows P_r versus the outlier power σ_o^2 for $N = 8$, $M = 3$, $M_o = 3$, and $N_{initial} = 40$ for different values of K and f_{d_o}. Specifically, four cases are considered, namely, (a) $K = 12$ and equal Doppler outliers; (b) $K = 12$ and random Doppler outliers; (c) $K = 20$ and equal Doppler outliers; (d) $K = 20$ and random Doppler outliers. The performance of the EL-RML, EL-ARML, U-RML, M-RML and M-ARML is compared with the GIP, RCGIP, ML and AML methods. The plots highlight that P_r of all the algorithms improves as σ_o^2 increases. Moreover, the EL-RML, EL-ARML, U-RML, M-RML, and M-ARML almost share the same behavior and outperform the other techniques, especially when the sample size is small. Specifically, for $\sigma_o^2 = 20$ dB, P_rs of the RML

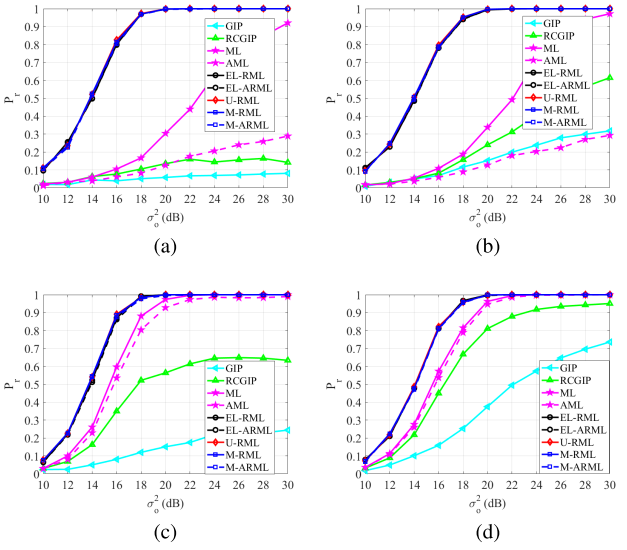

FIGURE 5. P_r versus σ_o^2 of the GIP (left-triangle-marked curve), RCGIP (upper-triangle-marked curve), exact ML (pentagram-marked solid curve), AML (pentagram-marked dashed curve), EL-RML (circle-marked solid curve), EL-ARML (circle-marked dashed curve), U-RML (diamond-marked curve), M-RML (square-marked solid curve), and M-ARML (square-marked dashed curve) for $N = 8$, $M = 3$, $M_o = 3$, and $N_{initial} = 40$. (a) $K = 12$ and equal Doppler outliers. (b) $K = 12$ and random Doppler outliers. (c) $K = 20$ and equal Doppler outliers. (d) $K = 20$ and random Doppler outliers.

and ARML methods are approximately equal to 1 for all the studied cases whereas those of the exact ML are about 0.31, 0.33, 0.96 and 0.97, respectively, for cases (a)-(d). In addition, the behavior of the EL-RML, EL-ARML, U-RML, M-RML, M-ARML, ML and AML algorithms is robust with respect to the Doppler frequencies of the outliers.

In Fig. 6, P_r versus σ_o^2 is plotted for $K = 20$, $N = 8$, $M = 6$, and $N_{initial} = 40$ for different values of M_o and f_{d_o}. Four specific cases are considered: (a) $M_o = 6$ and equal Doppler outliers; (b) $M_o = 6$ and random Doppler outliers; (c) $M_o = 8$ and equal Doppler outliers; (d) $M_o = 8$ and random Doppler outliers. The performance of the EL-RML, U-RML, M-RML and ML is not evaluated due to the heavy computational burden connected with their implementation. The ARML methods (EL-ARML and M-ARML) still ensure the best performance for all the considered cases. Specifically, when σ_o^2 reaches 20 dB, the ARML can reach a P_r value higher than 0.9. Nevertheless, P_rs of the AML are about 0.16, 0.31, 0.42 and 0.6, respectively, for cases (a)-(d).

Fig. 7 shows P_r versus the number of available secondary data K for $N = 8$, $M = 3$, $M_o = 3$, and $N_{initial} = 40$ for some values of σ_o^2 and outlier Doppler frequencies. More specifically, the following four cases are considered: (a) $\sigma_o^2 = 15$ dB and equal Doppler outliers; (b) $\sigma_o^2 = 15$ dB and random Doppler outliers; (c) $\sigma_o^2 = 30$ dB and equal Doppler outliers; (d) $\sigma_o^2 = 30$ dB and random Doppler outliers. The plots highlight that the RML and ARML methods ensure the best performance. Moreover, for the considered parameter values, 12 secondary data seem sufficient for the performance

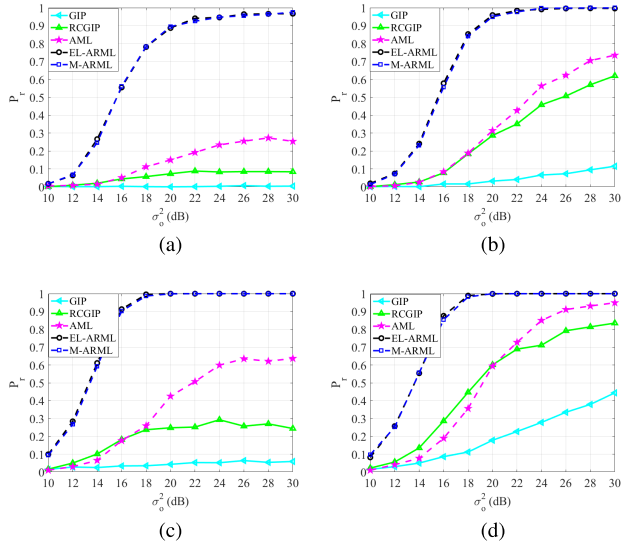

FIGURE 6. P_r versus σ_o^2 of the GIP (left-triangle-marked curve), RCGIP (upper-triangle-marked curve), AML (pentagram-marked dashed curve), EL-ARML (circle-marked dashed curve), and M-ARML (square-marked dashed curve) for $K = 20$, $N = 8$, $M = 6$, and $N_{initial} = 40$. (a) $M_o = 6$ and equal Doppler outliers. (b) $M_o = 6$ and random Doppler outliers. (c) $M_o = 8$ and equal Doppler outliers. (d) $M_o = 8$ and random Doppler outliers.

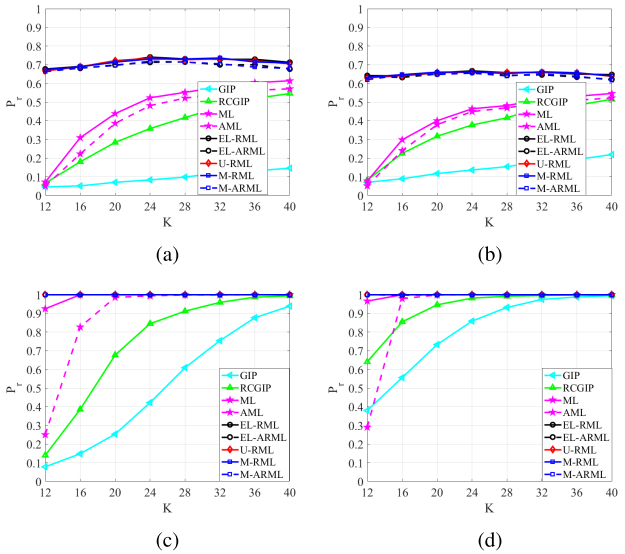

FIGURE 7. P_r versus K of the GIP (left-triangle-marked curve), RCGIP (upper-triangle-marked curve), exact ML (pentagram-marked solid curve), AML (pentagram-marked dashed curve), EL-RML (circle-marked solid curve), EL-ARML (circle-marked dashed curve), U-RML (diamond-marked curve), M-RML (square-marked solid curve), and M-ARML (square-marked dashed curve) for $N = 8$, $M = 3$, $M_o = 3$, and $N_{initial} = 40$. (a) $\sigma_o^2 = 15$ dB and equal Doppler outliers. (b) $\sigma_o^2 = 15$ dB and random Doppler outliers. (c) $\sigma_o^2 = 30$ dB and equal Doppler outliers. (d) $\sigma_o^2 = 30$ dB and random Doppler outliers.

of the regularized algorithms to achieve convergence.[7] By contrast, the GIP, RCGIP, ML and AML normally requires a larger K to reach convergence, especially when the outlier power is relatively low. Precisely, when $\sigma_o^2 = 15$ dB, the RCGIP, ML and AML requires about 36 secondary data whereas the GIP needs more than 40 vectors. Moreover, when $\sigma_o^2 = 30$ dB, the required number of secondary data for the GIP, RCGIP, ML and AML is around 40, 36, 20 and 16, (36, 24, 16 and 16), respectively, for the case of equal (random) Doppler outliers.

In Fig. 8, P_r versus K is plotted for $N = 8$, $M = 6$, $N_{initial} = 40$, and equal Doppler outliers for different values of M_o and σ_o^2. The following four specific cases are considered: (a) $M_o = 6$ and $\sigma_o^2 = 15$ dB; (b) $M_o = 6$ and $\sigma_o^2 = 30$ dB; (c) $M_o = 8$ and $\sigma_o^2 = 15$ dB; (d) $M_o = 8$ and $\sigma_o^2 = 30$ dB. The performance of the EL-RML, U-RML, M-RML and ML is not evaluated due to the heavy computational burden connected with their implementation. The ARML methods still require the smallest number of secondary data to achieve convergence. Precisely, the ARML methods only require about 20 secondary data whereas the other methods need at least 30 training vectors. Moreover, when $M_o = M$, the necessary number of secondary data for the GIP, RCGIP and AML to achieve convergence in the outlier dense scenario would increase dramatically as

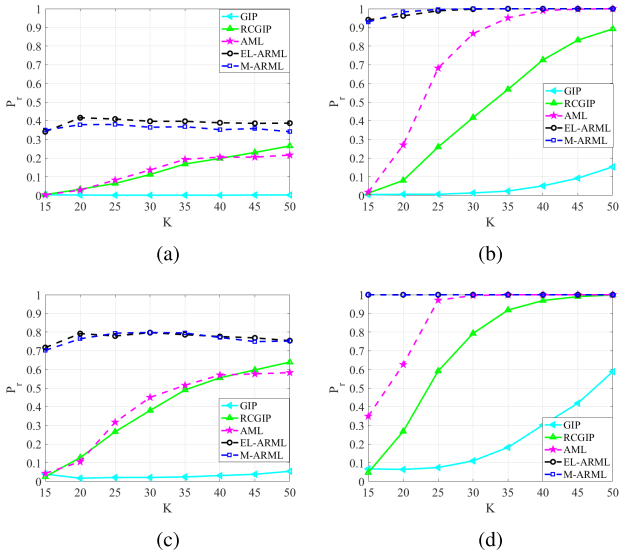

FIGURE 8. P_r versus K of the GIP (left-triangle-marked curve), RCGIP (upper-triangle-marked curve), AML (pentagram-marked dashed curve), EL-ARML (circle-marked dashed curve), and M-ARML (square-marked dashed curve) for $N = 8$, $M = 6$, $N_{initial} = 40$, and equal Doppler outliers. (a) $M_o = 6$ and $\sigma_o^2 = 15$ dB. (b) $M_o = 6$ and $\sigma_o^2 = 30$ dB. (c) $M_o = 8$ and $\sigma_o^2 = 15$ dB. (d) $M_o = 8$ and $\sigma_o^2 = 30$ dB.

compared with the outlier sparse situation, especially when the outlier power is high.

Summarizing, the conducted analysis suggests to recommend the EL-ARML as a viable candidate for a practical

[7] It is worth noting that a larger number of secondary data might lead to a slight performance degradation instead of an improvement when the outlier power is relatively low. This can be explained observing that the increase in the secondary data number might determine a corresponding increase in the probability that some homogeneous data possess a larger GIP value than the data containing outliers.

implementation thanks to the satisfactory performance and the acceptable computational complexity.

VI. CONCLUSION

In this paper the problem of censoring data vectors contaminated by outliers from secondary datasets with small cardinality has been considered. Specifically, the RML estimate of the outlier subset has been derived resorting to the GRLF criterion and an ARML procedure has been proposed to boost the computational efficiency. As to the selection of the regularization parameter, both the EL and CV criteria have been exploited. In particular, two different CV schemes have been devised: the former estimates the outlier subset independently for each partition, whereas the latter exploits the same outlier subset estimate for all the partitions. It has also been proved that the RML (ARML) procedure exhibits the unitary invariance property, i.e., invariance with respect to unitary transformations of the data.

The performance of the proposed algorithms has been analyzed in the presence of simulated data. The illustrative examples have shown that the EL-ARML method represents a very good compromise between performance and complexity. It is thus recommended as a candidate for a practical implementation in scenarios characterized by a limited number of secondary data. Possible future research avenues might concern the application of the EL-ARML censoring procedure as a pre-processing step to adaptive radar detectors.

REFERENCES

[1] F. Bandiera, D. Orlando, and G. Ricci, *Advanced Radar Detection Schemes Under Mismatched Signal Models* (Synthesis Lectures on Signal Processing). San Rafael, CA, USA: Morgan & Claypool, 2009.

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

[3] P. Stoica and K. C. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 1132–1143, Jul. 1990.

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

[5] W. L. Melvin, "A STAP overview," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 19, no. 1, pp. 19–35, Jan. 2004.

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

[7] A. D. Maio, "Rao test for adaptive detection in Gaussian interference with unknown covariance matrix," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3577–3584, Jul. 2007.

[8] P. Stoica, J. Li, X. Zhu, and J. R. Guerci, "On using *a priori* knowledge in space-time adaptive processing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2598–2602, Jun. 2008.

[9] W. L. Melvin, "Space-time adaptive radar performance in heterogeneous clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 621–633, Apr. 2000.

[10] M. Steiner and K. Gerlach, "Fast converging adaptive processor or a structured covariance matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1115–1126, Oct. 2000.

[11] A. Aubry, A. D. Maio, and L. Pallotta, "A geometric approach to covariance matrix estimation and its applications to radar problems," *IEEE Trans. Signal Process.*, vol. 66, no. 4, pp. 907–922, Feb. 2018.

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

[13] A. Aubry, A. D. Maio, L. Pallotta, and A. Farina, "Radar detection of distributed targets in homogeneous interference whose inverse covariance structure is defined via unitary invariant functions," *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 4949–4961, Oct. 2013.

[14] C. Hao, D. Orlando, G. Foglia, and G. Giunta, "Knowledge-based adaptive detection: Joint exploitation of clutter and system symmetry properties," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1489–1493, Oct. 2016.

[15] S. Yan, D. Massaro, D. Orlando, C. Hao, and A. Farina, "Adaptive detection and range estimation of point-like targets with symmetric spectrum," *IEEE Signal Process. Lett.*, vol. 24, no. 11, pp. 1744–1748, Nov. 2017.

[16] J. Liu, W. Liu, Y. Gao, S. Zhou, and X.-G. Xia, "Persymmetric adaptive detection of subspace signals: Algorithms and performance analysis," *IEEE Trans. Signal Process.*, vol. 66, no. 23, pp. 6124–6136, Dec. 2018.

[17] S. Han, C. Fan, and X. Huang, "A novel STAP based on spectrum-aided reduced-dimension clutter sparse recovery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 2, pp. 213–217, Feb. 2017.

[18] W. L. Melvin, M. C. Wicks, and R. D. Brown, "Assessment of multichannel airborne radar measurements for analysis and design of space-time processing architectures and algorithms," in *Proc. IEEE Radar Nat. Conf.*, May 1996, pp. 130–135.

[19] A. Aubry, A. D. Maio, L. Pallotta, and A. Farina, "Median matrices and their application to radar training data selection," *IET Radar, Sonar Navigat.*, vol. 8, no. 4, pp. 265–274, Apr. 2014.

[20] A. Aubry, A. D. Maio, L. Pallotta, and A. Farina, "Covariance matrix estimation via geometric barycenters and its application to radar training data selection," *IET Radar, Sonar Navigat.*, vol. 7, no. 6, pp. 600–614, Jul. 2013.

[21] P. Chen, W. L. Melvin, and M. C. Wicks, "Screening among multivariate normal data," *J. Multivariate Anal.*, vol. 69, no. 1, pp. 10–29, 1999.

[22] K. Gerlach, "Outlier resistant adaptive matched filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 3, pp. 885–901, Jul. 2002.

[23] P. J. Rousseeuw and K. Van Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, Aug. 1999.

[24] D. M. Rocke and D. L. Woodruff, "Identification of outliers in multivariate data," *J. Amer. Stat. Assoc.*, vol. 91, no. 435, pp. 1047–1061, 1996.

[25] A. S. Hadi, "Identifying multiple outliers in multivariate data," *J. Roy. Stat. Soc., B (Methodol.)*, vol. 54, no. 3, pp. 761–771, 1992.

[26] M. McDonald and B. Bhashyam, "Outlier suppression in adaptive filtering through de-emphasis weighting," *IET Radar, Sonar Navigat.*, vol. 1, no. 1, pp. 38–49, Feb. 2007.

[27] J. J. Lehtomaki, J. Vartiainen, M. Juntti, and H. Saarnisaari, "CFAR outlier detection with forward methods," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4702–4706, Sep. 2007.

[28] B. Tang, J. Tang, and Y. Peng, "Detection of heterogeneous samples based on loaded generalized inner product method," *Digit. Signal Process.*, vol. 22, no. 4, pp. 605–613, 2012.

[29] B. Dai, T. Wang, T. Bai, J. Wu, and Z. Bao, "Training data selection method for adaptive beamforming," *Electron. Lett.*, vol. 50, no. 17, pp. 1242–1244, Aug. 2014.

[30] W. Yifeng, W. Tong, W. Jianxin, and D. Jia, "Robust training samples selection algorithm based on spectral similarity for space—Time adaptive processing in heterogeneous interference environments," *IET Radar, Sonar Navigat.*, vol. 9, no. 7, pp. 778–782, 2015.

[31] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[32] D. Li, D. Chen, J. Goh, and S.-K. Ng, "Anomaly detection with generative adversarial networks for multivariate time series," Sep. 2018, *arXiv:1809.04758*. [Online]. Available: https://arxiv.org/abs/1809.04758

[33] S. Han, A. D. Maio, V. Carotenuto, L. Pallotta, and X. Huang, "Censoring outliers in radar data: An approximate ML approach and its analysis," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 2, pp. 534–546, Apr. 2019.

[34] S. Han, X. Huang, A. D. Maio, L. Pallotta, and S. Iommelli, "A two-stage detector for operation in outlier-dense scenarios," in *Proc. Sensor Signal Process. Defence Conf. (SSPD)*, Dec. 2017, pp. 1–5.

[35] Y. I. Abramovich, N. K. Spencer, and A. Y. Gorokhov, "Modified GLRT and AMF framework for adaptive detectors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 3, pp. 1017–1051, Jul. 2007.

[36] Y. I. Abramovich and B. A. Johnson, "GLRT-based detection-estimation for undersampled training conditions," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3600–3612, Aug. 2008.

[37] D. I. Warton, "Penalized normal likelihood and ridge regularization of correlation and covariance matrices," *J. Amer. Stat. Assoc.*, vol. 103, no. 481, pp. 340–349, 2008.

[38] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statist. Surv.*, vol. 4, pp. 40–79, Jul. 2010. doi: 10.1214/09-SS054.
[39] R. G. Ayoub, "Paolo Ruffini's contributions to the quintic," *Arch. History Exact Sci.*, vol. 23, no. 3, pp. 253–277, Sep. 1980.
[40] A. G. Akritas, "Linear and quadratic complexity bounds on the values of the positive roots of polynomials," *J. Universal Comput. Sci.*, vol. 15, no. 3, pp. 523–537, 2009.
[41] K. Gerlach, S. D. Blunt, and M. L. Picciolo, "Robust adaptive matched filtering using the FRACTA algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 3, pp. 929–945, Jul. 2004.

SUDAN HAN received the B.S. degree in information engineering and the M.S. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2012 and 2015, respectively, where she is currently pursuing the Ph.D. degree. From 2016 to 2017, she was with the University of Naples Federico II, Naples, Italy, involved in the training data selection. Her main research interests include clutter suppression and target detection.

LUCA PALLOTTA (S'12–M'15–SM'18) received the Laurea Specialistica degree *(cum laude)* in telecommunication engineering from the University of Sannio, Benevento, Italy, in 2009, and the Ph.D. degree in electronic and telecommunication engineering from the University of Naples Federico II, Naples, Italy, in 2014. His research interest includes the field of statistical signal processing, with emphasis on radar signal processing and radar targets classification. He won the Student Paper Competition at the IEEE Radar Conference, in 2013.

VINCENZO CAROTENUTO (S'12–M'16–SM'19) received the M.Sc. degree in telecommunication engineering and the Ph.D. degree in electronic and telecommunication engineering from the University of Naples Federico II, Naples, Italy, in 2010 and 2015, respectively. His research interest includes the field of statistical signal processing, with an emphasis on radar signal processing. He was a co-recipient of the Best Radar Paper Award from the IEEE Metrology for Aerospace, in 2018.

ANTONIO DE MAIO (S'01–A'02–M'03–SM'07–F'13) was born in Sorrento, Italy, in 1974. He received the Dr.Eng. degree (Hons.) and the Ph.D. degree in information engineering from the University of Naples Federico II, Naples, Italy, in 1998 and 2002, respectively.

In 2004, he was a Visiting Researcher with the U.S. Air Force Research Laboratory, Rome, NY. In 2007, he was a Visiting Researcher with the Chinese University of Hong Kong, Hong Kong. He is currently a Professor with the Università degli Studi di Napoli 'Federico II'. His research interests include the field of statistical signal processing, with emphasis on radar detection, optimization theory applied to radar signal processing, and multiple-access communications.

Dr. De Maio was a recipient of the 2010 IEEE Fred Nathanson Memorial Award as the young (less than 40 years of age) AESS Radar Engineer 2010 whose performance is particularly noteworthy as evidenced by contributions to the radar art over a period of several years, with the following citation for robust CFAR detection, knowledge-based radar signal processing, and waveform design and diversity. He was a co-recipient of the 2013 Best Paper Award (entitled to B. Carlton) of the IEEE Transactions on Aerospace and Electronic Systems with the contribution for Knowledge-Aided (Potentially Cognitive) Transmit Signal and Receive Filter Design in Signal-Dependent Clutter.

XIAOTAO HUANG (M'02) received the B.S. and Ph.D. degrees in information and communication engineering from the National University of Defense Technology, Changsha, China, in 1990 and 1999, respectively. He is currently a Professor with the National University of Defense Technology. His fields of interest include radar theory, signal processing, and radio frequency signal suppression.