EM-based Algorithm for Unsupervised Clustering of Measurements from a Radar Sensor Network

Linjie Yan, *Member, IEEE*, Pia Addabbo, *Senior Member, IEEE*, Nicomino Fiscante, Carmine Clemente, *Senior Member, IEEE*, Chengpeng Hao, *Senior Member, IEEE*, Gaetano Giunta, *Senior Member, IEEE*, and Danilo Orlando, *Senior Member, IEEE*

Abstract—This paper deals with the problem of clustering data returned by a radar sensor network that monitors a region where multiple moving targets are present. The network is formed by nodes with limited functionalities that transmit the estimates of target positions (after a detection) to a fusion center without any association between measurements and targets. To solve the problem at hand, we resort to model-based learning algorithms and instead of applying the plain maximum likelihood approach, due to the related computational requirements, we exploit the latent variable model coupled with the expectation-maximization algorithm. The devised estimation procedure returns posterior probabilities that are used to cluster the huge amount of data collected by the fusion center. Remarkably, we also consider challenging scenarios with an unknown number of targets and estimate it by means of the model order selection rules. The clustering performance of the proposed strategy is compared to that of conventional data-driven methods over synthetic data. The numerical examples point out that the herein proposed solutions can provide reliable clustering performance overcoming the considered competitors.

Index Terms—Batch algorithms, expectation-maximization, measurement clustering, multiple moving targets, radar, sensor network, unsupervised learning.

I. INTRODUCTION

IN the recent years, the increase of computational resources has heavily promoted the development and the implementation of sophisticated signal processing techniques in real radar systems. More importantly, with the advent of the big data era, statistical signal processing algorithms have been incorporated into a wider class called "Machine Learning" that has become more and more popular over the years [1], [2] and also comprises deep learning techniques. Two main design

Linjie Yan and Chengpeng Hao are with the Institute of Acoustics, Chinese Academy of Sciences, Beijing, 100190, China. E-mail: yanlinjie16@163.com; haochengp@mail.ioa.ac.cn.

Pia Addabbo is with Università degli Studi "Giustino Fortunato", viale Raffale Delcogliano, 12, 82100 Benevento, Italy. E-mail: p.addabbo@unifortunato.eu

Nicomino Fiscante and Gaetano Giunta are with Industrial, Electronic and Mechanical Engineering Department, University of Roma Tre, 00146 Rome, Italy. E-mail: nicomino.fiscante@virgilio.it; gaetano.giunta@uniroma3.it.

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

Danilo Orlando is with Dipartimento di Ingegneria dell'Informazione, Università di Pisa, Via Caruso 16, 56122 Pisa, Italy (e-mail: danilo.orlando@unipi.it).

approaches can be identified within this wide class: data-driven and model-based oriented designs. The former operates through a learning stage that relies on training data, namely a set of input-output pairs used to learn the algorithm from the available data, whereas the latter relies on a model grounded on the first physics principles. In both cases, the exploitation of high performance boards becomes unavoidable to fulfill the tight (time) requirements of real radar systems. The methods devised in this paper are framed in the model-based class.

In parallel with the technological advancements, the operating scenarios have become more and more challenging as well as the corresponding estimation/optimization problems. A related example is represented by scenarios where swarms of (possibly noncooperative) targets are moving in the region of interest that is monitored by a network of passive/active radars [3]–[13]. As a matter of fact, the advantages arising from the use of a sensor network are well-known and we mention here the spatial diversity and the energy integration. For this reason, in modern radar systems, the cooperation between systems distributed over the region under surveillance becomes a key factor to improve the reliability of the entire surveillance system. For instance, with focus on ground radars, the spatial diversity of the sensor network can be exploited to face deception/saturation jamming techniques such as range gate stealing [14], [15], that generate false targets to make the radar lose range track on the target.

However, radar networks require special attention in handling a huge amount of data provided by each node, especially in the presence of multiple moving targets [16]–[18]. In this case, the detection can be either centralized or decentralized. In centralized detection, sending raw observations from radar sensors to the fusion center (FC), where the final decision process takes place, imposes a large communication burden. On the other hand, in decentralized detection, radar sensors take their local decisions on the presence of prospective targets and, then, transmit the results of such decisions (compressed data) to the FC. In this case, the amount of transmitted data is lower than in the previous configuration with a consequent energy saving but at the price of performance loss. As a matter of fact, the overall system does not take advantage of diversity. The design of an optimized decentralized detector network, which entails the design of both optimal local detectors and fusion rules at FC, does not represent an easy task [19] and some strategies to solve this difficult problem have been proposed in the past [20]–[22]. Moreover, in radar networks, the tracking of multiple targets represents a very challenging

problem due to the limited data computational capacity of the FC as well as of the transmit energy of each radar [22]–[25]. Generally speaking, this implies the necessity of using sophisticated (centralized or decentralized) fusion methods [26]–[29]. Anyway, the implementation of centralized fusion methods is computationally expensive [26]. Nevertheless, a preliminary stage that preprocesses and clusters the received observations into homogeneous measurement sets, might reduce the computational load at the FC [30]–[35]. As a matter of fact, although the data volume handled by the FC does not decrease after clustering, a preliminary data clustering, according to some criterion, can provide an indication of the number of objects that move within the region of interest along with the associated measurements. As a consequence, the tracking algorithms used to form target trajectories can be fed by these clusters and simplified (from a computational point of view) since they are not required to solve a data association problem.

Clustering algorithms have been also applied for the design of CFAR and/or selective/robust detectors [36], [37]. Particularly, in [36], received data are transformed to generate specific features based upon the maximal invariant statistic [38] for that problem. Then, such features are clustered in a two-dimensional plane to come up with a detector that is invariant to the disturbance parameters and, hence, can guarantee the constant false alarm rate property. Another design methodology, based on the previous approach, is proposed in [37], where sub-optimal strategies with low complexity have been developed. The expectation-maximization (EM) algorithm is also used for clustering data that can be modeled as a Gaussian Mixture [1]. In particular, in the context of radar systems, the EM algorithm is used to partition clutter returns based upon their spectral properties [39]. To this end, fictitious hidden random variables are introduced to represent the clutter type of each range bin. Interestingly, the classification algorithms have been devised by accounting for different models of the clutter covariance matrix. Whereas, in [40], joint detection and classification architectures have been proposed by extending the work of [39] to the case where an unknown number of multiple point-like targets are present in the region of interest. In [41], the EM algorithm is still used for the target detection in heterogeneous clutter scenarios.

In this paper, we focus on the design of a suitable preliminary stage that processes data in order to form clusters from measurements collected by a radar sensor network that senses a crowded region. As stated above, the aim of this preliminary stage consists in providing a preliminary solution of the data association problem that can be used to simplify the ensuing data processing stages responsible for track formation. Actually, this issue is getting more and more critical with the growing number of scenarios in which multiple targets need to be monitored in crowded spaces [42]. As a matter of fact, conventional radar scenarios are getting more densely populated with the arrival of new classes of targets such as unmanned/autonomous vehicles (either in air, land, or sea). The same remark also holds for less conventional but equally challenging radar scenarios such as monitoring space targets. Therefore, we focus on a radar sensor network whose

monostatic nodes illuminate the same region of interest where an unknown number of targets are moving. Each node does not perform any association between the measurements related to a detection and the detected targets, and sends to the FC the position estimates corresponding to each detection within a common observation time window. At the design stage, we assume that measurements are collected over a time interval such that the target trajectories can be approximated as straight lines (with a not necessarily constant velocity). Assuming a specific distribution for the measurement noise, we exploit the likelihood function of data to solve the clustering problem. In this respect, we do not resort to the maximum likelihood principle, because it requires the evaluation of the joint likelihood function for each partition of the entire measurement set and targets' number. It is clear that such a task is unacceptable from a computational standpoint and, more importantly, the maximum likelihood approach would return estimates corresponding to the maximum allowed model order since the likelihood function monotonically increases with the number of unknown parameters [43], [44]. Therefore, to circumvent the above limitations, we introduce fictitious and unobserved discrete random variables that represent target labels associated with the measurements gathered by the FC. Then, exploiting the joint distribution of measurements and the hidden labels, we develop an estimation procedure grounded on the EM algorithm [45] that allows us to obtain a nondecreasing sequence of likelihood values as well as closed-form expressions for the updates of the estimates. The clusters are formed by applying the maximum a posteriori rule, whereas an estimate of the number of targets is returned through the Model Order Selection (MOS) rules [43], [44], [46]–[48].

The performance analysis is carried out over synthetic data and starts from the case where the number of targets is known to proceed with the case where the latter is unknown. As terms of comparison, we consider two conventional data-driven algorithms. The numerical examples show the superiority of the proposed approach over the considered competitors in classifying the collected measurements.

The remainder of this paper is organized as follows. In the next section, we describe the surveillance system and provide a formal statement of the problem. In Section III, we devise the EM-based estimation procedure when the number of targets is known, whereas in Section IV we extend such a procedure to the case of an unknown number of targets. The numerical examples are contained in Section V and, finally, concluding remarks along with the description of possible future research lines are confined to Section VI.

A. Notation

In the sequel, vectors are denoted by boldface lower-case. As to numerical sets, $\mathbb{N}$ is the set of natural numbers, $\mathbb{R}$ is the set of real numbers, and $\mathbb{R}^{N \times M}$ is the Euclidean space of $(N \times M)$-dimensional real matrices (or vectors if $M = 1$). The Cartesian product of two sets A and B is denoted by $A \times B$. The acronyms 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 $f(x|y)$. The probability of an event $\mathcal{A}$ is defined as $P\{\mathcal{A}\}$, whereas the conditional probability of $\mathcal{A}$ given another event $\mathcal{B}$ is $P\{\mathcal{A}|\mathcal{B}\}$. Symbol $\lfloor \cdot \rfloor$ represents the highest integer lower than the argument while $|x|$ is the absolute value of $x \in \mathbb{R}$. Finally, we write $x \sim \mathcal{N}(m, \sigma^2)$ if x is a Gaussian random variable with mean m and variance $\sigma^2 > 0$.

II. SIGNAL MODEL AND PROBLEM FORMULATION

Let us consider a sensor network of $K \in \mathbb{N}$ monostatic radars deployed to illuminate the same region of interest. The capabilities of such systems are limited to target detection and rough estimation of its range and azimuth. Assuming that the region of interest is populated by an unknown number, L say, of multiple point-like moving targets, the measurements obtained by each sensor are transmitted to a fusion center that converts the received polar coordinates into Cartesian coordinates. Notice that each radar system works in an asynchronous way with respect to the other systems. At a given time instant t and for the lth target, the fusion center receives $0 \leq K_l \leq K$ measurements from the nodes that have detected the same target and we denote the corresponding coordinates by $(x_{l,k,t}, y_{l,k,t}) \in \mathbb{R} \times \mathbb{R}$, $l = 1, \ldots, L$ and $k \in \Omega_l \subseteq \{1, \ldots, K\}$, where Ω_l is the set of systems that have detected the lth target and measured the related parameters. The fusion center collects measurements for a preassigned time interval without knowing which target is associated with the received measurements. The first operation performed by the fusion center consists in clustering data under the assumption that targets' motion over short time intervals can be approximated as a straight line. The motivation behind this assumption arises from the fact that, for some platforms such as unmanned/autonomous vehicles, their routes are planned according to a specific mission or constrained by spatial obstacles that force the platform to follow a given path. For suitable time intervals, such trajectories can be approximated by straight lines. Moreover, since the shortest path between two spatial points is a straight line (neglecting the space curvature), the latter is used to reduce fuel/energy consumption [49, and references therein]. The system geometry and the considered scenario are depicted in Fig. 1. Thus, in the absence of disturbance and at a given time instant t, we consider the following model for the measurements

$$y_{l,k,t} = a_l x_{l,k,t} + b_l, \tag{1}$$

where $a_l \in \mathbb{R}$ and $b_l \in \mathbb{R}$ are unknown coefficients associated with the motion of the lth target and they are assumed stationary over the considered time interval. In what follows, we neglect the temporal dimension represented by t since the proposed approach works in asynchronous way with respect to the time of arrival of the measurements. Therefore, measurements are indexed by a generic integer n. It is important to observe that, in principle, the above clustering problem could be solved by using the maximum likelihood approach. However, the solution would experience a combinatorial computational complexity due to the fact
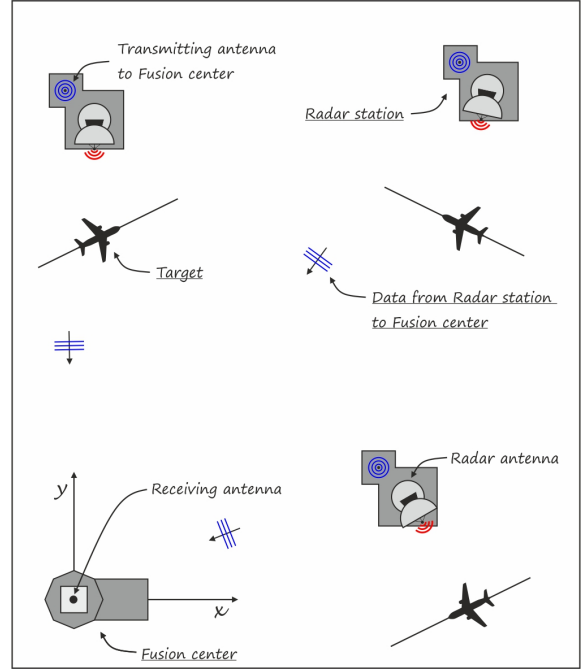

Fig. 1: System geometry and coordinates.

that the likelihood function is computed over all the possible associations and number of targets. In addition, the maximum likelihood approach would always select the maximum number of targets since the likelihood function is monotonically increasing with the number of parameters [44]. To overcome the above drawbacks, we resort to the following approaches:

- the Latent Variable Model [1], [39]–[41], [50] that allows us to reduce the computational complexity related to each possible association;
- the MOS rules [43], [44], [46]–[48] that mitigate the overfitting inclination of the maximum likelihood approach for the estimation of the target number.

Starting from the first item, let us assume for the moment that the number of targets is known and denote by N the entire number of measurements, y_n say, collected by the fusion center. Then, we reformulate the problem at hand by introducing N independent and identically distributed hidden discrete random variables, c_n say, taking on values in $\mathcal{A} = \{1, \ldots, L\}$, such that for the generic nth measurement we have that

$$y_n | c_n = l \sim \mathcal{N}(a_l x_n + b_l, \sigma_l^2), \quad n = 1, \ldots, N, \tag{2}$$

where the measurement noise related to the lth target is modeled as a Gaussian random variable with unknown variance $\sigma_l^2 > 0$. Notice that random variables c_n represent target identifiers associated with each measurement and, as described below, can be used to suitably cluster measurements. Assuming that the measurements are statistically independent and that the PMF of c_n is denoted by $\pi_l = P\{c_n = l\}$, $l \in \mathcal{A}$,

the joint PDF of $y_1, \ldots, y_N$ can be written as:

$$f(y_1, \ldots, y_N; \boldsymbol{\theta}, \pi_1, \ldots, \pi_L)$$
$$= \prod_{n=1}^{N} f(y_n; \boldsymbol{\theta}, \pi_1, \ldots, \pi_L)$$
$$= \prod_{n=1}^{N} \sum_{l \in \mathcal{A}} f(y_n, c_n = l; \boldsymbol{\theta}, \pi_1, \ldots, \pi_L) \quad (3)$$
$$= \prod_{n=1}^{N} \sum_{l \in \mathcal{A}} f(y_n | c_n = l; \boldsymbol{\theta}) \pi_l,$$

where $\boldsymbol{\theta} = [a_1, \ldots, a_L, b_1, \ldots, b_L, \sigma_1^2, \ldots, \sigma_L^2]^T \in \mathbb{R}^{3L \times 1}$.

Finally, we exploit the above PDF to obtain the estimates of the unknown parameters that will be used to build up a maximum a posteriori rule allowing for measurement association and, hence, target clustering. Specifically, the unknown parameters are estimated by solving

$$\max_{\substack{\boldsymbol{\theta} \\ \pi_1, \ldots, \pi_L}} \prod_{n=1}^{N} \sum_{l \in \mathcal{A}} f(y_n | c_n = l; \boldsymbol{\theta}) \pi_l \quad (4)$$

or equivalently

$$\max_{\substack{\boldsymbol{\theta} \\ \pi_1, \ldots, \pi_L}} \sum_{n=1}^{N} \log \left[\sum_{l \in \mathcal{A}} f(y_n | c_n = l; \boldsymbol{\theta}) \pi_l \right]. \quad (5)$$

Denoting by $\widehat{\boldsymbol{\theta}}$ and $\widehat{\pi}_l$, $l \in \mathcal{A}$, the estimates of $\boldsymbol{\theta}$ and π_l, $l \in \mathcal{A}$, respectively, we associate measurement y_n with the target identifier $\widehat{l}$ exhibiting the highest a posteriori probability, namely

$$\widehat{l} = \arg\max_{l \in \mathcal{A}} P\left\{c_n = l | y_n; \widehat{\boldsymbol{\theta}}, \widehat{\pi}_l\right\}. \quad (6)$$

In the next section, we design an estimation procedure to solve problem (5) and, as a byproduct, (6), grounded on the EM-algorithm. Then, in Section IV, we address the case where L is unknown.

III. ESTIMATION PROCEDURE FOR KNOWN NUMBER OF TARGETS

From a mathematical point of view, the plain maximization in (5) is a difficult task at least to the best of the authors' knowledge. For this reason, we resort to the EM-algorithm that is an iterative procedure with closed-form updates for the estimates of interest and provides at least a local maximum [1], [45], [51]. The EM-algorithm repeats two steps called E-step and M-step until a stopping criterion is not satisfied. The former consists in updating the a posteriori probability of the event $c_n = l$ given the nth measurement y_n whereas in the latter step, the log-likelihood function is maximized to obtain updated parameter estimates.

Thus, let us start form the E-step and denote by $\widehat{\boldsymbol{\theta}}^{(h-1)}$ and $\widehat{\pi}_l^{(h-1)}$, $l \in \mathcal{A}$, the estimates of $\boldsymbol{\theta}$ and π_l, $l \in \mathcal{A}$, at the $(h-1)$th iteration, respectively. The E-step leads to the computation of

$$p_n^{(h-1)}(l) = P\left\{c_n = l | y_n; \widehat{\boldsymbol{\theta}}^{(h-1)}, \widehat{\pi}_l^{(h-1)}\right\}$$
$$= \frac{f\left(y_n | c_n = l; \widehat{\boldsymbol{\theta}}^{(h-1)}\right) \widehat{\pi}_l^{(h-1)}}{\sum_{\bar{l} \in \mathcal{A}} f\left(y_n | c_n = \bar{l}; \widehat{\boldsymbol{\theta}}^{(h-1)}\right) \widehat{\pi}_{\bar{l}}^{(h-1)}} \quad (7)$$

for $l \in \mathcal{A}$ and $n = 1, \ldots, N$. As for the M-step, after applying the Jensen inequality to the argument of (5), we come up with the following optimization problem

$$\max_{\substack{\boldsymbol{\theta} \\ \pi_l, l \in \mathcal{A}}} \sum_{n=1}^{N} \sum_{l \in \mathcal{A}} p_n^{(h-1)}(l) \log \left(\frac{f(y_n | c_n = l; \boldsymbol{\theta}) \pi_l}{p_n^{(h-1)}(l)} \right), \quad (8)$$

where

$$f(y_n | c_n = l; \boldsymbol{\theta}) = \frac{\exp\left[-\frac{1}{2\sigma_l^2}(y_n - a_l x_n - b_l)^2\right]}{\sqrt{2\pi}\sigma_l}. \quad (9)$$

Problem (8) is tantamount to

$$\max_{\substack{\boldsymbol{\theta} \\ \pi_l, l \in \mathcal{A}}} \sum_{n=1}^{N} \sum_{l \in \mathcal{A}} \Big[p_n^{(h-1)}(l) \log[f(y_n | c_n = l; \boldsymbol{\theta})]$$
$$+ p_n^{(h-1)}(l) \log(\pi_l) \Big]. \quad (10)$$

Thus, the maximization over π_l, $l \in \mathcal{A}$, can be accomplished by solving

$$\begin{cases} \max_{\pi_l, l \in \mathcal{A}} \sum_{n=1}^{N} \sum_{l \in \mathcal{A}} p_n^{(h-1)}(l) \log(\pi_l), \\ \text{subject to} \quad \sum_{l \in \mathcal{A}} \pi_l = 1. \end{cases} \quad (11)$$

According to the method of Lagrange multipliers, we set to zero the first derivative with respect to the unknowns of the Langrangian function whose expression is

$$\sum_{n=1}^{N} \sum_{l \in \mathcal{A}} p_n^{(h-1)}(l) \log(\pi_l) - \lambda \left(\sum_{l \in \mathcal{A}} \pi_l - 1 \right), \quad (12)$$

where λ is the Lagrange multiplier. Proceeding in this way, we obtain that

$$\sum_{n=1}^{N} p_n^{(h-1)}(l) \frac{1}{\pi_l} - \lambda = 0$$
$$\implies \pi_l = \frac{1}{\lambda} \sum_{n=1}^{N} p_n^{(h-1)}(l), \quad l \in \mathcal{A}. \quad (13)$$

Considering the constraint leads to

$$\frac{1}{\lambda} \sum_{l \in \mathcal{A}} \sum_{n=1}^{N} p_n^{(h-1)}(l) = 1$$
$$\implies \lambda = \sum_{n=1}^{N} \sum_{l \in \mathcal{A}} p_n^{(h-1)}(l) = N \quad (14)$$

and, hence,

$$\pi_l^{(h)} = \frac{1}{N} \sum_{n=1}^{N} p_n^{(h-1)}(l), \quad l \in \mathcal{A}. \qquad (15)$$

It still remains to maximize the objective function with respect to the other parameters. Thus, neglecting the irrelevant constants, the optimization problem to be solved can be formulated as

$$\max_{\boldsymbol{\theta}} g(\boldsymbol{\theta}), \qquad (16)$$

where

$$g(\boldsymbol{\theta}) = \sum_{n=1}^{N} \sum_{l \in \mathcal{A}} p_n^{(h-1)}(l)$$
$$\times \left[-\frac{1}{2} \log(2\pi\sigma_l^2) - \frac{1}{2\sigma_l^2} (y_n - a_l x_n - b_l)^2 \right]. \qquad (17)$$

Focusing on σ_l^2, $l \in \mathcal{A}$, we firstly notice that $\forall l \in \mathcal{A}$

$$\lim_{\sigma_l^2 \to +\infty} g(\boldsymbol{\theta}) = -\infty \quad \text{and} \quad \lim_{\sigma_l^2 \to 0} g(\boldsymbol{\theta}) = -\infty. \qquad (18)$$

Thus, the maximum over σ_l^2 can be found by setting to zero the first derivative of $g(\boldsymbol{\theta})$ with respect to σ_l^2, namely

$$\sum_{n=1}^{N} p_n^{(h-1)}(l) \left[-\frac{1}{2\sigma_l^2} + \frac{1}{2(\sigma_l^2)^2} (y_n - a_l x_n - b_l)^2 \right] = 0 \qquad (19)$$

and solving with respect to σ_l^2 we obtain

$$\tilde{\sigma}_l^2 = \frac{\sum\limits_{n=1}^{N} (y_n - a_l x_n - b_l)^2 p_n^{(h-1)}(l)}{\sum\limits_{n=1}^{N} p_n^{(h-1)}(l)}, \quad l \in \mathcal{A} \qquad (20)$$

Moreover, when $\sigma_l^2 < (\tilde{\sigma}_l^2)^{(h)}$, the derivative is positive (increasing function), whereas for $\sigma_l^2 > (\tilde{\sigma}_l^2)^{(h)}$, the derivative is negative (decreasing function). Replacing (20) in (17) and neglecting the terms that do not enter the optimization problem, the latter is equivalent to

$$\min_{\substack{a_l, b_l \\ l \in \mathcal{A}}} \sum_{l \in \mathcal{A}} g(a_l, b_l), \qquad (21)$$

where

$$g(a_l, b_l) = \left(\sum_{n=1}^{N} p_n^{(h-1)}(l) \right)$$
$$\times \log \left[\sum_{n=1}^{N} (y_n - a_l x_n - b_l)^2 p_n^{(h-1)}(l) \right]. \qquad (22)$$

Let us study the behavior of $g(a_l, b_l)$ at the endpoints of its domain. To this end, it is not difficult to show that

$$\lim_{\substack{|a_l| \to +\infty \\ |b_l| \to +\infty}} g(a_l, b_l) = +\infty, \quad l \in \mathcal{A}. \qquad (23)$$

Therefore, we set to zero the first derivative over a_l of $g(a_l, b_l)$ to obtain

$$\left(\sum_{n=1}^{N} p_n^{(h-1)}(l) \right) \frac{1}{\sum\limits_{n=1}^{N} (y_n - a_l x_n - b_l)^2 p_n^{(h-1)}(l)}$$
$$\times \left[\sum_{n=1}^{N} p_n^{(h-1)}(l) 2(y_n - a_l x_n - b_l)(-x_n) \right] = 0 \qquad (24)$$

and, hence,

$$\tilde{a}_l = \frac{\sum\limits_{n=1}^{N} p_n^{(h-1)}(l)(y_n - b_l)x_n}{\sum\limits_{n=1}^{N} p_n^{(h-1)}(l)x_n^2}, \quad l \in \mathcal{A}. \qquad (25)$$

Replacing (25) into (22) and neglecting the irrelevant constants, we can consider

$$\min_{b_l} \log g_1(b_l), \qquad (26)$$

where

$$g_1(b_l) = \sum_{n=1}^{N} \left\{ \frac{p_n^{(h-1)}(l)}{\left(B_l^{(h-1)} \right)^2} \left[y_n B_l^{(h-1)} \right. \right.$$
$$- x_n \sum_{\bar{n}=1}^{N} p_{\bar{n}}^{(h-1)}(l) y_{\bar{n}} x_{\bar{n}}$$
$$\left. \left. + b_l \left(x_n A_l^{(h-1)} - B_l^{(h-1)} \right) \right]^2 \right\}, \qquad (27)$$

with

$$A_l^{(h-1)} = \sum_{\bar{n}=1}^{N} p_{\bar{n}}^{(h-1)}(l)x_{\bar{n}},$$
$$B_l^{(h-1)} = \sum_{\bar{n}=1}^{N} p_{\bar{n}}^{(h-1)}(l)x_{\bar{n}}^2. \qquad (28)$$

Now, setting to zero the first derivative of $\log g_1(b_l)$ with respect to b_l leads to

$$\frac{1}{g_1(b_l)} \sum_{n=1}^{N} \frac{2 p_n^{(h-1)}(l)}{B_l^{(h-1)}} \left[y_n B_l^{(h-1)} \right.$$
$$\left. - x_n \sum_{\bar{n}=1}^{N} p_{\bar{n}}^{(h-1)}(l) y_{\bar{n}} x_{\bar{n}} + b_l \left(x_n A_l^{(h-1)} - B_l^{(h-1)} \right) \right]$$
$$\times \left(\frac{x_n A_l^{(h-1)} - B_l^{(h-1)}}{B_l^{(h-1)}} \right) = 0, \qquad (29)$$

and, as a consequence, the estimate update for b_l is given by

$$
\widehat{b}_l^{(h)} = \frac{1}{\sum_{n=1}^{N} p_n^{(h-1)}(l) \left(x_n A_l^{(h-1)} - B_l^{(h-1)}\right)^2}
$$

$$
\times \left\{ \sum_{n=1}^{N} \left[p_n^{(h-1)}(l) x_n \left(x_n A_l^{(h-1)} - B_l^{(h-1)}\right) \right. \right.
$$

$$
\times \sum_{\bar{n}=1}^{N} p_{\bar{n}}^{(h-1)}(l) y_{\bar{n}} x_{\bar{n}} \right] - \sum_{n=1}^{N} \left[p_n^{(h-1)}(l) y_n \right.
$$

$$
\left. \left. \times \left(x_n A_l^{(h-1)} - B_l^{(h-1)}\right) \sum_{\bar{n}=1}^{N} p_{\bar{n}}^{(h-1)}(l) x_{\bar{n}}^2 \right] \right\}. \quad (30)
$$

Finally, the updates for the estimates of a_l and σ_l^2 can be written as

$$
\widehat{a}_l^{(h)} = \frac{\sum_{n=1}^{N} p_n^{(h-1)}(l) \left(y_n - b_l^{(h)}\right) x_n}{\sum_{n=1}^{N} p_n^{(h-1)}(l) x_n^2}, \quad l \in \mathcal{A}. \quad (31)
$$

and

$$
(\widehat{\sigma}_l^2)^{(h)} = \frac{\sum_{n=1}^{N} \left(y_n - a_l^{(h)} x_n - b_l^{(h)}\right)^2 p_n^{(h-1)}(l)}{\sum_{n=1}^{N} p_n^{(h-1)}(l)}, \quad l \in \mathcal{A}, \quad (32)
$$

respectively.

Summarizing, the E-step given by (7) and the above updates obtained from the M-step are repeated until a stopping criterion is not satisfied. Specifically, the iterations end when

$$
\Delta \mathcal{L}(h) = \left| \left[\mathcal{L}\left(y_1, \ldots, y_N; \widehat{\boldsymbol{\theta}}^{(h)}, \widehat{\pi}_1^{(h)}, \ldots, \widehat{\pi}_L^{(h)}\right) \right. \right.
$$

$$
\left. \left. - \mathcal{L}\left(y_1, \ldots, y_N; \widehat{\boldsymbol{\theta}}^{(h-1)}, \widehat{\pi}_l^{(h-1)}, \ldots, \widehat{\pi}_L^{(h-1)}\right) \right] \right|
$$

$$
\left/ \left| \mathcal{L}\left(y_1, \ldots, y_N; \widehat{\boldsymbol{\theta}}^{(h)}, \widehat{\pi}_1^{(h)}, \ldots, \widehat{\pi}_L^{(h)}\right) \right| < \epsilon, \quad (33) \right.
$$

where $\epsilon > 0$ and

$$
\mathcal{L}\left(y_1, \ldots, y_N; \widehat{\boldsymbol{\theta}}^{(h)}, \widehat{\pi}_1^{(h)}, \ldots, \widehat{\pi}_L^{(h)}\right)
$$

$$
= \sum_{n=1}^{N} \log \left[\sum_{l \in \mathcal{A}} f\left(y_n | c_n = l; \widehat{\boldsymbol{\theta}}^{(h)}\right) \pi_l^{(h)} \right], \quad (34)
$$

or after a maximum number of iterations denoted by h_{max}.

In the next section, we show how to incorporate the estimates obtained through the EM-algorithm into the MOS rules to determine the number of targets L.

IV. ESTIMATION PROCEDURE FOR UNKNOWN NUMBER OF TARGETS

The estimation of the number of targets relies on the MOS rules since the hypotheses corresponding to scenarios with different numbers of targets are nested. As stated in Section I, the maximum likelihood approach experiences an overestimation of the parameter space size [44], [52] and, hence, a penalty term is required to balance the corresponding growth of the likelihood function. In what follows, we consider the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the Generalized Information Criterion (GIC) [44], whose general structure is

$$
\widehat{L} = \underset{L \in \{1, \ldots, L_{\max}\}}{\arg \min} \left\{ -2\mathcal{L}\left(y_1, \ldots, y_N; \widehat{\boldsymbol{\theta}}^{(h_L)}, \right. \right.
$$

$$
\left. \left. \widehat{\pi}_1^{(h_L)}, \ldots, \widehat{\pi}_L^{(h_L)}\right) + p(L) \right\}, \quad (35)
$$

where h_L is the number of iterations used by the EM-based estimation procedure introduced in Section III assuming L targets and $p(L)$ is a penalty term defined as follows

$$
p(L) = \begin{cases} 2n_p(L), & \text{AIC}, \\ (1+\rho)n_p(L), \ \rho \geq 1, & \text{GIC}, \\ n_p(L) \log(N), & \text{BIC}, \end{cases} \quad (36)
$$

with $n_p(L) = 4L$ being the number of unknown parameters in the presence of L targets.

Once $\widehat{L}$ is computed through (35), we can define $\widehat{\mathcal{A}} = \{1, \ldots, \widehat{L}\}$ and the association rule (6) becomes

$$
\widehat{l} = \underset{l \in \widehat{\mathcal{A}}}{\arg \max} \, P\left\{c_n = l | y_n; \widehat{\boldsymbol{\theta}}^{(h_{\widehat{L}})}, \widehat{\pi}_l^{(h_{\widehat{L}})}\right\}, \quad (37)
$$

where $\widehat{\boldsymbol{\theta}}^{(h_{\widehat{L}})}$ and $\widehat{\pi}_l^{(h_{\widehat{L}})}$, $l \in \widehat{\mathcal{A}}$, are the estimates corresponding to the case $L = \widehat{L}$.

V. NUMERICAL EXAMPLES AND DISCUSSION

In this section, we provide illustrative examples aimed at assessing the classification/clustering performance of the proposed approach by resorting to Monte Carlo (MC) counting techniques. Specifically, we start the analysis by assuming that L is known and consider two operating scenarios with $L = 5$ and $L = 10$ targets. Then, we focus on the estimation capabilities for the number of targets by setting $L_{\max} = 10$ and the actual value of L equal to 3 and 7.[1]

In each scenario, we directly generate the trajectories of multiple targets in Cartesian coordinates with the number of measurements for the lth target, N_l say, $l \in \mathcal{A}$,[2] being a uniformly distributed random variable in $[60, 90]$. In addition, trajectory intersections are included in the considered scenarios making the classification task more challenging.

For comparison purposes, we compare the proposed approach with two classical machine learning clustering techniques, namely the K-Nearest-Neighbor (KNN) and K-means [1], [53], [54]. The former is a supervised method, whereas the latter is unsupervised. In the ensuing examples, we assume that the ratio between the quantity of training and validation

---

[1]A guideline for the choice of $L_{\max}$ should account for both computational load and target number underestimation. Specifically, $L_{\max}$ should be large enough to account for all the targets moving in the region of interest and, at the same time, to guarantee a reasonable computation time.

[2]Notice that the constraint $\sum_{l=1}^{L} N_l = N$ holds.

Algorithm 1: Initialization of a_l and b_l.

Input: L, y_n, x_n, $n = 1, \ldots, N$
Output: a_l, b_l, $l \in \mathcal{A}$
Initialization: set $N'_1 = N$, $y_{n'_1} = y_n$, $x_{n'_1} = x_n$,
$n'_1 = n = 1, \ldots, N$
for $l = 1, \ldots, L$ **do**

  1. compute a_l and b_l of the linear regression by

$$a_l = \left(\sum_{n'_l=1}^{N'_l} (x_{n'_l} - \bar{x})(y_{n'_l} - \bar{y}) \right) / \sum_{n'_l=1}^{N'_l} (x_{n'_l} - \bar{x})^2$$

  and $b_l = \bar{y} - a_l \bar{x}$ using N'_l measurements with $\bar{x}$
  and $\bar{y}$ the mean value of $x_{n'_l}$ and $y_{n'_l}$, respectively;

  2. compute the standard deviation of $(x_{n'_l}, y_{n'_l})$
  with respect to the fitted line, namely,
  $d_l(n'_l) = |y_{n'_l} - a_l x_{n'_l} - b_l| / \sqrt{a_l^2 + b_l^2}$,
  $n'_l = 1, \ldots, N'_l$;

  3. set $l = l + 1$;

  4. select the $\lfloor N/L \rfloor$ measurements with the
  minimum $d_l(n'_l)$ and discard them, we have
  $N'_l = N'_{l-1} - \lfloor N/L \rfloor$;

  5. go to step 1 using the remaining measurements
  until the loop ends.

end

---

data for the KNN is $4/6$ and the number of nearest neighbors is 50. Moreover, we set the GIC parameter[3] $\rho = 2$.

As for the initialization of the EM-based procedure, it is important to observe that it plays a role of primary importance. As a matter of fact, an unreliable initialization might significantly impair the estimation performance of iterative procedures and, more importantly, it should tightly account for the design assumptions of the iterative estimation procedures. For this reason, we derive an *ad hoc* initialization strategy, summarized in Algorithm 1, that finds initial labels for the unknown coefficients associated with the motion of each target exploiting the a priori information about trajectory (i.e., its approximation to a straight line over the observation interval). Finally, the initial values for the PMF of c_n are $\pi_l = 1/L$, $l \in \mathcal{A}$. Possible initial values the noise variances are

$$\sigma_l^2 = \sum_{n'_l=1}^{N'_l} (y_{n'_l} - a_l x_{n'_l} - b_l)^2 / N'_l, \ l \in \mathcal{A}, \qquad (38)$$

where N'_l denotes the number of measurements used to compute the initial trajectory coefficients of the lth target in Algorithm 1.

In each scenario, if not explicitly specified, the measurements are affected by an uncertainty with variance $\sigma_l^2 = 50$, $l \in \mathcal{A}$ that lead to a significant "mix" of the targets' measurements. Finally, we anticipate here that, in order to achieve a satisfactory compromise between convergence and computational load, ϵ is set to 10^{-5} and h_{max} is set to 150 and 250 for $L = 5$ and $L = 10$, respectively, as corroborated by the subsequent convergence analysis.

---

[3]Notice that typical values for ρ are within the interval $[1, 5]$ [44] and $\rho = 2$, for this specific application, represents a reasonable compromise to limit the overestimation of the number of targets L [40], [55].

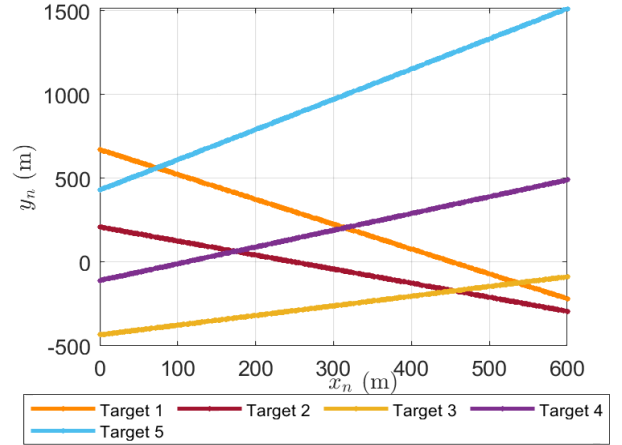
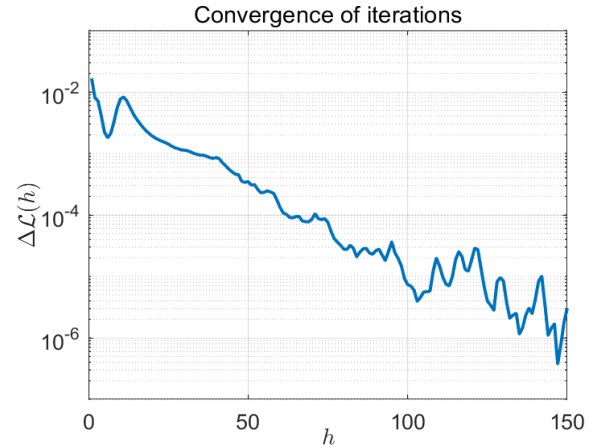Fig. 2: Clean target trajectories without measurement noise (first scenario).

Fig. 3: $\Delta\mathcal{L}(h)$ versus h for the EM-based procedure (first scenario).

A. First Operating Scenario for Known Number of Targets $L = 5$

In this scenario, we consider five targets whose measurements are generated as follows

- **Target 1**: $y_{n_1} | c_{n_1} = 1 \sim \mathcal{N}(-1.5 x_{n_1} + 671, \ \sigma_1^2)$, $n_1 = 1, \ldots, N_1$;
- **Target 2**: $y_{n_2} | c_{n_2} = 2 \sim \mathcal{N}(-0.8 x_{n_2} + 310, \ \sigma_2^2)$, $n_2 = 1, \ldots, N_2$;
- **Target 3**: $y_{n_3} | c_{n_3} = 3 \sim \mathcal{N}(0.6 x_{n_3} - 434, \ \sigma_3^2)$, $n_3 = 1, \ldots, N_3$;
- **Target 4**: $y_{n_4} | c_{n_4} = 4 \sim \mathcal{N}(x_{n_4} - 110, \ \sigma_4^2)$, $n_4 = 1, \ldots, N_4$;
- **Target 5**: $y_{n_5} | c_{n_5} = 5 \sim \mathcal{N}(1.8 x_{n_5} + 430, \ \sigma_5^2)$, $n_5 = 1, \ldots, N_5$;

where x_{n_l}, $l \in \mathcal{A}$, are generated by randomly selecting integers from 1 to N. The above target trajectories without measurement noise are shown in Fig. 2.

In Fig. 3, we plot the curves of $\Delta\mathcal{L}(h)$ averaged over 1000 independent trials to select a suitable value for h_{max}. It can be seen that the relative variation of the log-likelihood function
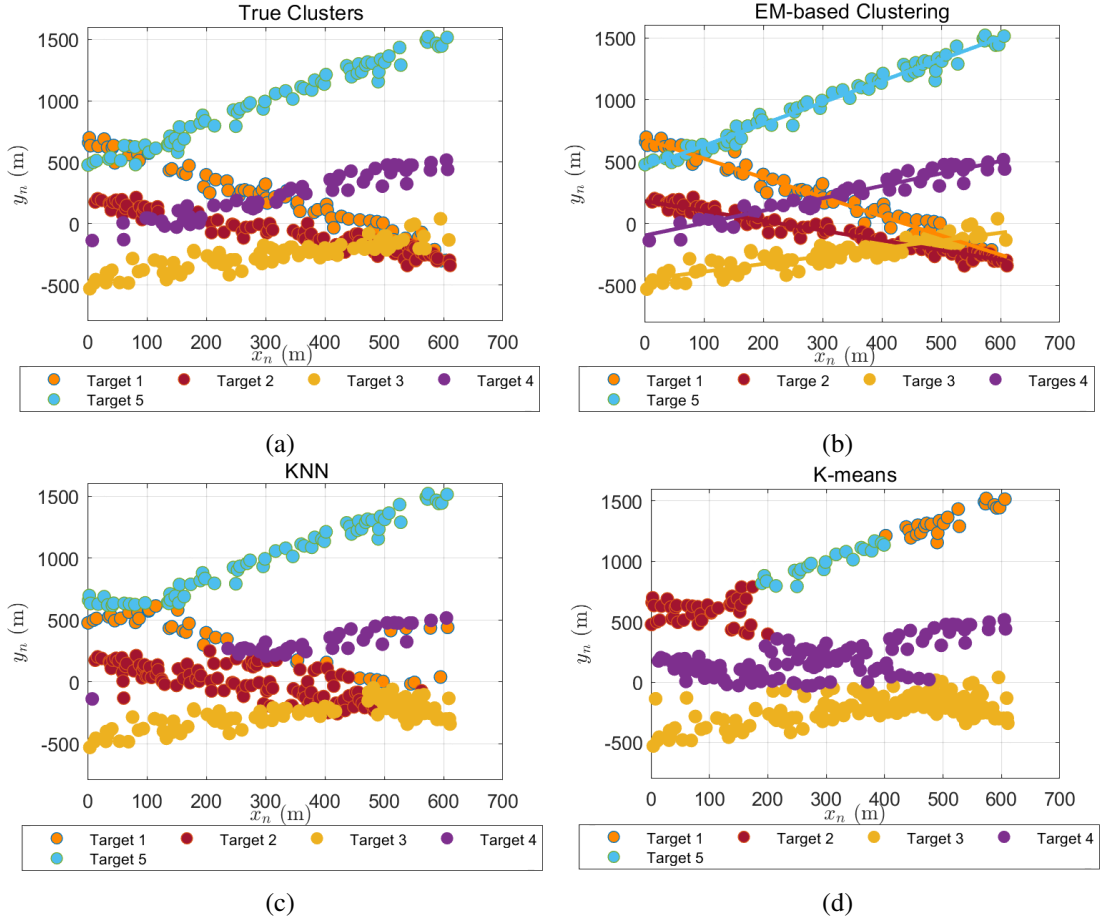
Fig. 4: Cartesian coordinates diagrams for each target over a single MC trial (first scenario): (a) true measurement association; (b) classification (scatter points) results and trajectory fitting (straight lines) for the proposed architecture; (c) classification results for the KNN; (d) classification results for the K-means.

is lower than 10^{-5} after 150 iterations and, hence, for this scenario, we set $h_{max} = 150$.

The classification capabilities of the proposed architecture in comparison with the two considered competitors are investigated in Figs. 4-6. Specifically, the true clusters and the classification results over a single MC trial are shown in Fig. 4. The inspection of this figure clearly points out the superiority of the proposed method in measurement labeling over the counterparts. As a matter of fact, from a qualitative point of view, both KNN and K-means experience evident misclassification errors due to the association of measurements to wrong targets. As a consequence, the trajectory of a given target appears divided into segments corresponding to different targets. This kind of segmentation is more evident in Subfigure 4(d) that reports the classification results of K-means. Moreover, Subfigure 4(b) contains the target trajectories, obtained through the estimates of a_l and b_l, $l \in \mathcal{A}$, that[4] perfectly fit with the measurements (at least for the considered parameter setting).

Fig. 5 shows the mean classification consistency (%), which is defined as the ratio between the number of correct classi-

[4]Observe that analogous curves are not reported in the other subfigures since the KNN and K-means cannot provide such estimates.
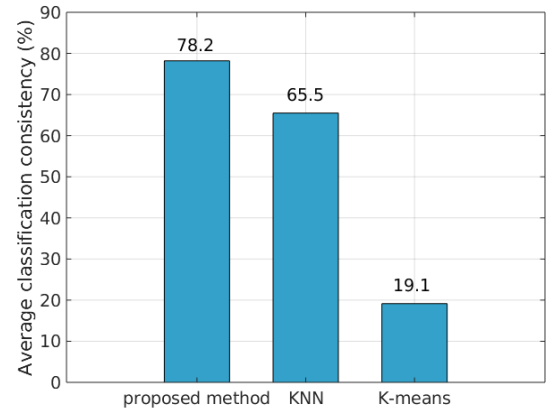

Fig. 5: Average classification consistency (%) over 1000 independent trials for the considered classifiers (first scenario).

fications over the true target categories averaged over 1000 independent trials. It turns out that the proposed EM-based classifier achieves a classification gain of approximately 12.6% and 59% with respect to the KNN and K-means, respectively. Fig. 6 shows the mean classification error (%) averaged over
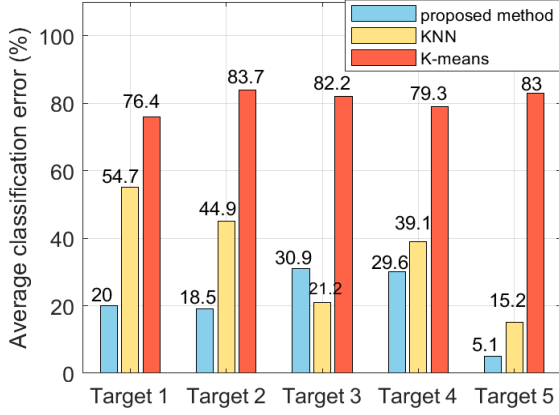
Fig. 6: Average classification error (%) of each target over 1000 independent trials for the considered classifiers (first scenario).

1000 MC trials, namely the ratio between the number of misclassified measurements for a given class (target) and the true quantity for that category. Again, the advantage of the proposed method over the considered competitors is quite evident.

Finally, to assess the estimation accuracy for the estimates of a_l and b_l at the n_{mc}th MC trial, which are generally denoted by $\widehat{a}_l(n_{mc})$ and $\widehat{b}_l(n_{mc})$, $l \in \mathcal{A}$, respectively, in Table I we evaluate the Percentage Root Mean Square Error (PRMSE) relative to the true values that is defined as

$$
\begin{cases}
\text{PRMSE}_{a_l} = \sqrt{\sum_{n_{mc}=1}^{N_{mc}} \dfrac{\min\limits_{l' \in \mathcal{A}} \left(a_l - \widehat{a}_{l'}\right)^2}{N_{mc}}} \times \dfrac{100}{|a_l|}, \\[3em]
\text{PRMSE}_{b_l} = \sqrt{\sum_{n_{mc}=1}^{N_{mc}} \dfrac{\min\limits_{l' \in \mathcal{A}} \left(b_l - \widehat{b}_{l'}\right)^2}{N_{mc}}} \times \dfrac{100}{|b_l|}
\end{cases}
\tag{39}
$$

with $N_{mc} = 1000$ being the number of MC trials. The table highlights that for targets 3 and 4 the estimate of a_l gives rise to errors greater than 20% due to the fact that the trajectories of these targets are characterized by intersections with other lines whose angular coefficient is considerably different. The same remark also holds for what concerns the errors related to the estimate of b_l. Otherwise stated, even though the percentage of correct classification is high, when an error occurs, its value can be high due to line intersections. Nevertheless, such errors can be mitigated by filtering the estimates over several consecutive processed batches of measurements.

Before assessing the performance when the number of targets grows, we investigate here the behavior of the considered classifiers when the measurement noise variance increases to $\sigma_l^2 = 80$, $l = 1, \ldots, L$. The analysis is analogous to that for $\sigma_l^2 = 50$. Specifically, Figs. 7-10 and Table II show that although the performance of the EM-based approach is slightly worse (as expected) with respect to the case $\sigma_l^2 = 50$, $l \in \mathcal{A}$, the proposed scheme still maintains more reliable capabilities

TABLE I: PRMSE values (%) for a_l and b_l , $l = 1, \ldots, L$, over 1000 independent trials (first scenario).

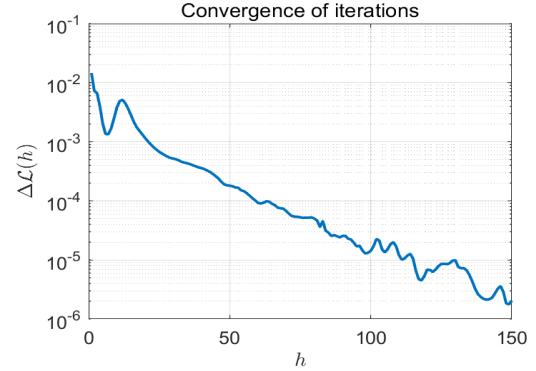|	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$
a_l	5.4	17.5	58.5	36.8	1.9
b_l	4.7	25.8	13.3	45.2	3.2

Fig. 7: $\Delta\mathcal{L}(h)$ versus h for the EM-based procedure (first scenario with $\sigma_l^2 = 80$, $l \in \mathcal{A}$).

in terms of measurements classification then the considered competitors.

B. Second Operating Scenario for Known Number of Targets $L = 10$

Now, we consider a more challenging scenario where $L = 10$ targets are present with different trajectories and intersections. As for the previous case, the KNN and K-means classifiers are taken into account as natural competitors. Target measurements are generated as follows

- **Target 1**: $y_{n_1}|c_{n_1} = 1 \sim \mathcal{N}(-4x_{n_1} - 4897, \ \sigma_1^2), \ n_1 = 1, \ldots, N_1$;
- **Target 2**: $y_{n_2}|c_{n_2} = 2 \sim \mathcal{N}(14.3x_{n_2} - 6230, \ \sigma_2^2), \ n_2 = 1, \ldots, N_2$;
- **Target 3**: $y_{n_3}|c_{n_3} = 3 \sim \mathcal{N}(0.1x_{n_3} - 2936, \ \sigma_3^2), \ n_3 = 1, \ldots, N_3$;
- **Target 4**: $y_{n_4}|c_{n_4} = 4 \sim \mathcal{N}(-1.9x_{n_4} - 1774, \ \sigma_4^2), \ n_4 = 1, \ldots, N_4$;
- **Target 5**: $y_{n_5}|c_{n_5} = 5 \sim \mathcal{N}(1.1x_{n_5} + 330, \ \sigma_5^2), \ n_5 = 1, \ldots, N_5$;
- **Target 6**: $y_{n_6}|c_{n_6} = 6 \sim \mathcal{N}(-0.7x_{n_6} + 1997, \ \sigma_6^2), \ n_6 = 1, \ldots, N_6$;
- **Target 7**: $y_{n_7}|c_{n_7} = 7 \sim \mathcal{N}(0.4x_{n_7} + 3245, \ \sigma_7^2), \ n_7 = 1, \ldots, N_7$;
- **Target 8**: $y_{n_8}|c_{n_8} = 8 \sim \mathcal{N}(0.5x_{n_8} + 4588, \ \sigma_8^2), \ n_8 = 1, \ldots, N_8$;
- **Target 9**: $y_{n_9}|c_{n_9} = 9 \sim \mathcal{N}(2.6x_{n_9} + 5846, \ \sigma_9^2), \ n_9 = 1, \ldots, N_9$;

TABLE II: PRMSE values (%) for a_l and b_l , $l \in \mathcal{A}$, over 1000 independent trials (first scenario with $\sigma_l^2 = 80$, $l \in \mathcal{A}$).

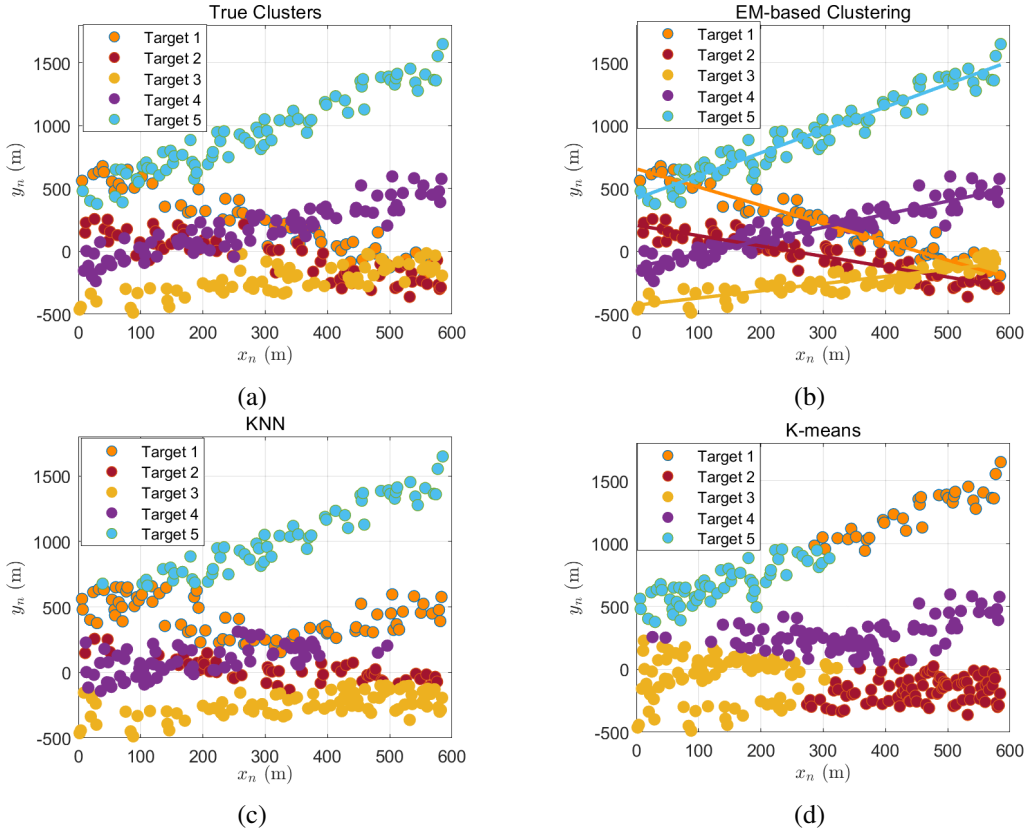|	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$
a_l	8.7	26.3	61.2	30.3	3.4
b_l	6.7	34.9	14.3	58.0	5.8

Fig. 8: Cartesian coordinates diagrams for each target over a single MC trial (first scenario with $\sigma_l^2 = 80$, $l \in \mathcal{A}$): (a) true measurement association; (b) classification (scatter points) results and trajectory fitting (straight lines) for the proposed architecture; (c) classification results for the KNN; (d) classification results for the K-means.
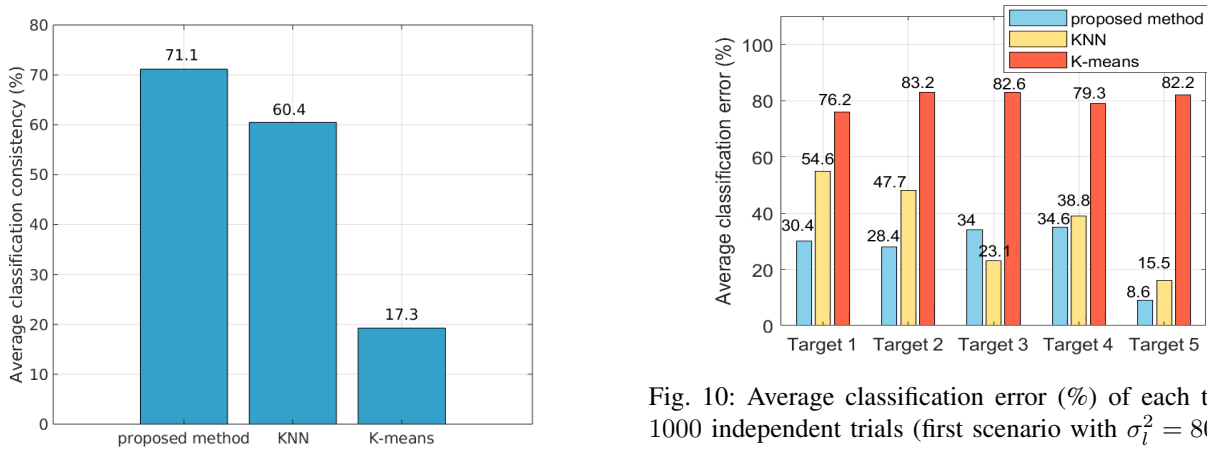

Fig. 9: Average classification consistency (%) over 1000 independent trials for the considered classifiers (first scenario with $\sigma_l^2 = 80$, $l \in \mathcal{A}$).

Fig. 10: Average classification error (%) of each target over 1000 independent trials (first scenario with $\sigma_l^2 = 80$, $l \in \mathcal{A}$).

• **Target 10**: $y_{n_{10}}|c_{n_{10}} = 10 \sim \mathcal{N}(-2.4x_{n_{10}} + 6706, \ \sigma_{10}^2)$, $n_{10} = 1, \ldots, N_{10}$.

The number of measurements for each target is generated as in the previous case and the clean trajectories are shown in Fig. 11. The classification performances of the proposed architecture are investigated assuming $h_{max} = 250$. Such a value is selected from Fig. 12 where the relative variation of the log-likelihood is below 10^{-5} for $h = 250$.

In Figs. 13-15, we provide a qualitative and quantitative assessment of the classification performance for the three algorithms. As observed for the case $L = 5$, the EM-based classifier is less inclined to partition the measurement set corresponding to a given target into subsets associated to other targets. This behavior is evident in Fig. 13 where the K-means experiences the worst performance corroborating what

TABLE III: PRMSE values (%) for a_l and b_l , $l \in \mathcal{A}$, over 1000 independent trials (second scenario).

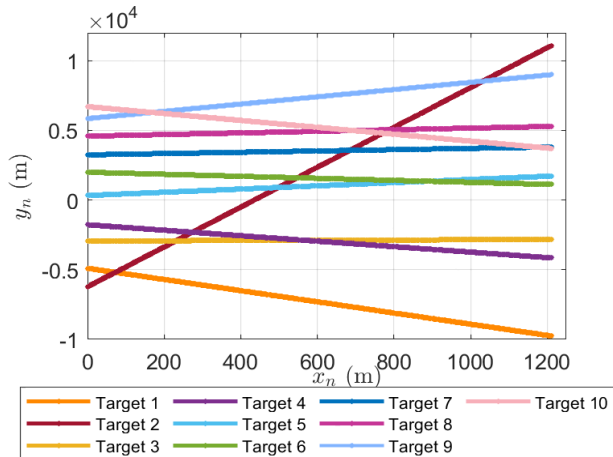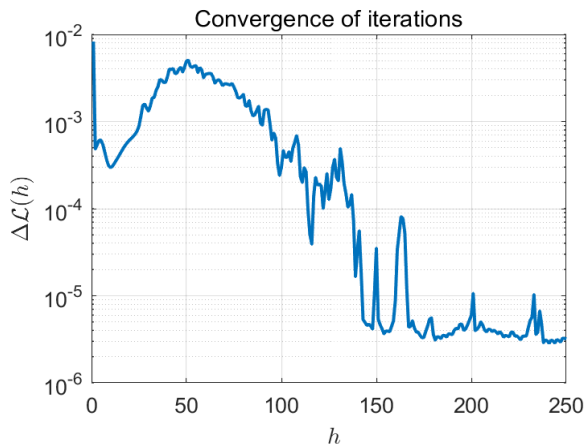|	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$	$l=6$	$l=7$	$l=8$	$l=9$	$l=10$
a_l	5.1	1.7	39.4	53.9	23.3	16.7	23.5	21.8	11.5	53.6
b_l	0.9	0.2	17.5	28.9	79.0	11.6	10.4	10.4	1.0	5.7

Fig. 11: Clean target trajectories without measurement noise (second scenario).

Fig. 12: $\Delta\mathcal{L}$ versus h for the EM-based procedure (second scenario).

indicated by Fig. 5. Figs. 14-15 point out the EM-based method provides an overall performance that is superior with respect to that of the considered competitors, even though for some targets the KNN (that is a supervised method) can return lower classification errors with respect ot the EM-based classifier (that is an unsupervised method).

Finally, in Table III, we report the PRMSE for the estimates of a_l and b_l, $l = 1, \ldots, 10$. In this case, the errors related to b_l are small except for target 5 whose trajectory intersects that of target 6 maintaining a low separation. As for the errors related to a_l, the highest values are returned for targets 3, 4, and 10. In fact, the lines corresponding to these targets share an intersection with the line associated to target 2 whose angular coefficient is significantly different from the other.

TABLE IV: PRMSE values (%) for a_l and b_l , $l \in \mathcal{A}$, using AIC, BIC, and GIC over 1000 independent trials ($L = 3$).

		$l=1$	$l=2$	$l=3$
AIC-based	a_l	4.2	30.1	8.6
	b_l	1.9	3.8	12.0
BIC-based	a_l	3.8	27.1	7.5
	b_l	1.7	3.2	10.1
GIC-based	a_l	3.9	27.4	7.8
	b_l	1.7	3.3	10.5

C. Operating Scenario where Number of Targets is Unknown

In this section, the classification and estimation performance is assessed when the number of targets is unknown under the constraint $L_{\max} = 10$. The scenario considered in what follows comprises $L = 3$ targets whose trajectories are

- **Target 1**: $y_{n_1}|c_{n_1} = 1 \sim \mathcal{N}(-1.8807x_{n_1} + 771, \ \sigma_1^2)$, $n_1 = 1, \ldots, N_1$;
- **Target 2**: $y_{n_2}|c_{n_2} = 2 \sim \mathcal{N}(-0.2679x_{n_2} + 410, \ \sigma_2^2)$, $n_2 = 1, \ldots, N_2$;
- **Target 3**: $y_{n_3}|c_{n_3} = 3 \sim \mathcal{N}(x_{n_3} - 129, \ \sigma_3^2)$, $n_3 = 1, \ldots, N_3$,

More specifically, true clusters with measurement noise variance $\sigma_l^2 = 50$ and the clean trajectories are shown in Fig. 16. Before presenting the estimation results, the convergence rate of the EM procedure over 1000 independent trials using AIC, BIC, and GIC is depicted in Fig. 17. The results confirm that $h_{max} = 50$ returns a relative variation of $\Delta\mathcal{L}$ lower than 10^{-4}. The estimation performance related to a_l, b_l, and L is investigated using the metrics defined in (39) and

$$\text{RMSE}_L = \sqrt{\sum_{n_{mc}=1}^{N_{mc}} \left(\widehat{L}(n_{mc}) - L\right)^2 / N_{mc}} \quad (40)$$

with $N_{mc} = 1000$. Table IV contains the PRMSE values for what concerns the estimation of the trajectory parameters and points out that the BIC-based clustering architectures provide better results than the classifiers based upon the AIC and GIC. In Fig. 18, the RMSE values associated with the estimation of L confirm the superiority of the BIC-based classifier that returns an error lower than 0.5 at least for the considered parameters.

VI. CONCLUSIONS AND FUTURE WORKS

In this work, we have proposed a solution for clustering data generated by the nodes of a radar network where each node has limited processing capabilities. In fact, we have assumed that the fusion center collects (2-dimensional) position measurements without any side information that can be used to create clusters associated with the targets in the region of interest.
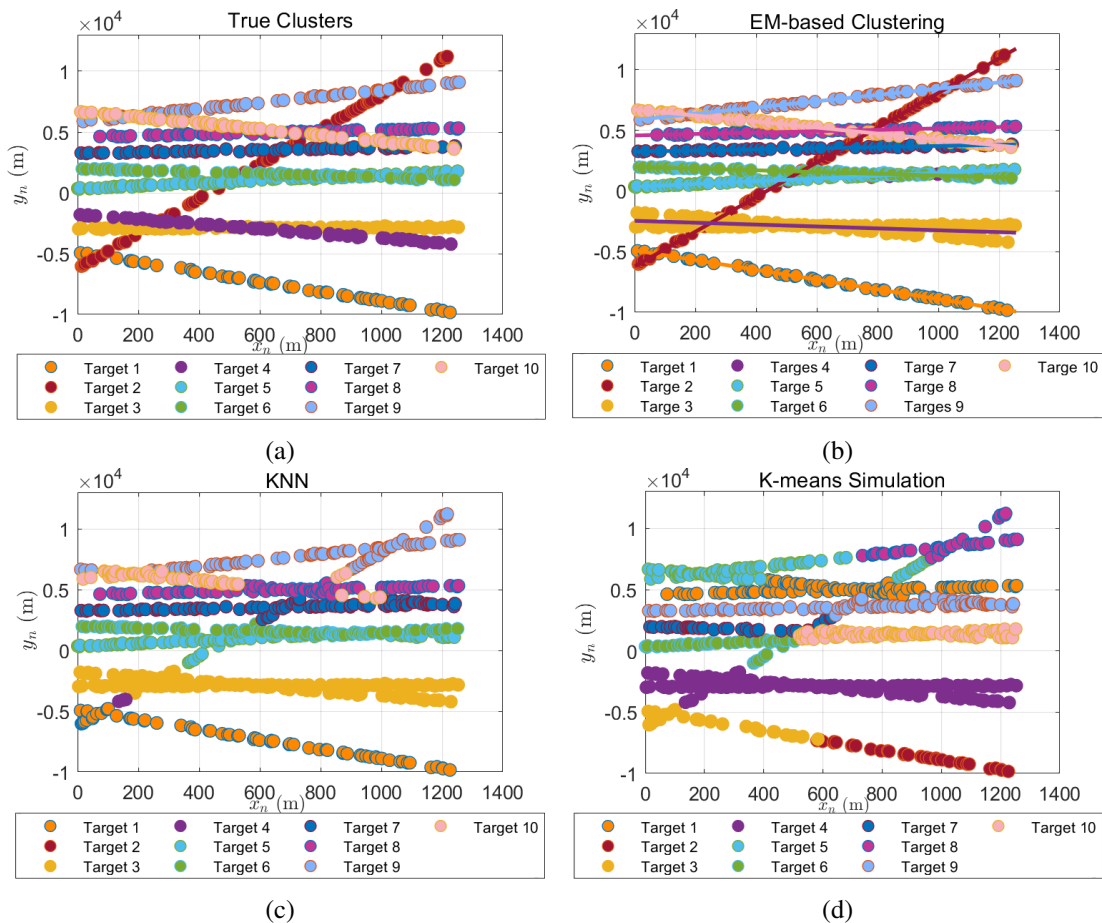
Fig. 13: Scatter diagrams for each target over one trial (second scenario): (a) true measurement association; (b) classification (scatter points) results and trajectory fitting (straight lines) for the proposed architecture; (c) classification results for the KNN; (d) classification results for the K-means.
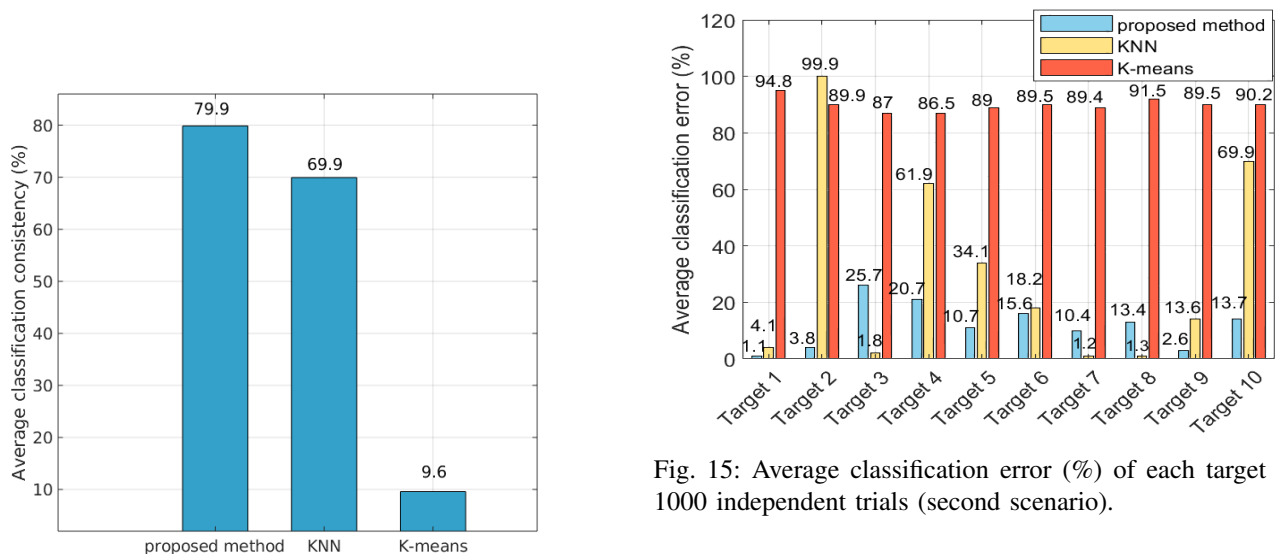

Fig. 14: Average classification consistency (%) over 1000 independent trials (second scenario).

Fig. 15: Average classification error (%) of each target over 1000 independent trials (second scenario).

To this end, we have used fictitious unobservable random variables that represent the label of each measurement. Then, we have estimated the posterior probability that a label takes on a specific value given the corresponding measurement by resorting to the EM-algorithm. The clustering is performed by
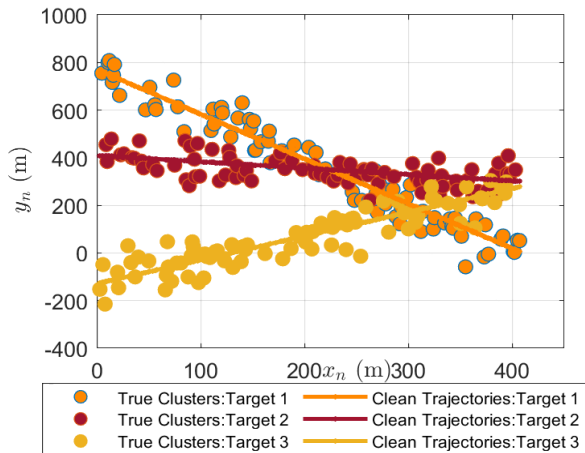
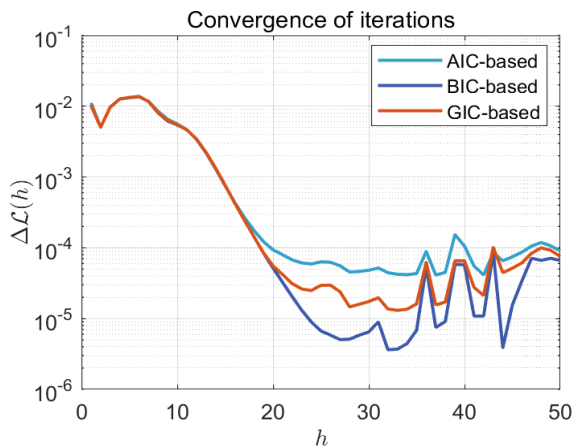Fig. 16: True clusters and clean trajectories ($L = 3$).

Fig. 17: $\Delta\mathcal{L}$ versus h of the EM procedure for the classifiers based on AIC, BIC, and GIC ($L = 3$).
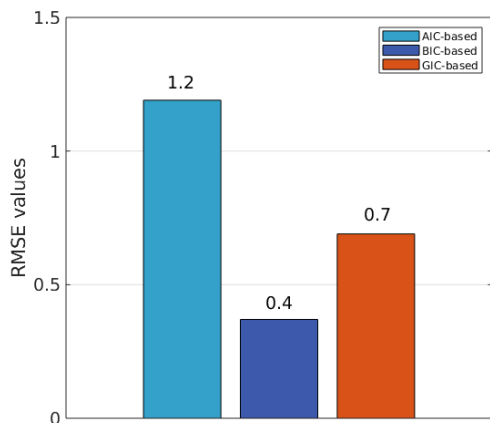

Fig. 18: RMSE values for L for the classifiers based upon the AIC, BIC, and GIC over 1000 independent trials ($L = 3$).

selecting the label that returns the highest posterior probability. This method is clearly less time demanding than the plain maximum likelihood approach whose computational load depends on the total number of data partitions as well as the number of targets. The performance assessment has been conducted by using synthetic data and in comparison with well-known data-driven clustering algorithms such as the KNN and K-means. Different challenging scenarios have been considered and for each of them, the proposed algorithm is capable of outperforming the considered competitors for what concerns the clustering and estimation capabilities.

Future research tracks might encompass the extension of the proposed approach to the case where the measurements contain information related to a third-dimension or the target Doppler frequency. The validation of the proposed approach with real recorded data from a radar network is part of the current research activity.

REFERENCES

[1] K. Murphy, *Machine Learning: A Probabilistic Perspective*, ser. Adaptive Computation and Machine Learning series. MIT Press, 2012.
[2] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*. Elsevier Science, 2015.
[3] V. S. Chernyak, *Fundamentals of Multisite Radar Systems*. Routledge, 1998.
[4] G. R. F. Ehlers, D. Orlando, "Batch tracking algorithm for multistatic sonars," *IET Radar, Sonar & Navigation*, vol. 6, pp. 746–752(6), 2012.
[5] L. Yan, L. Pallotta, G. Giunta, and D. Orlando, "Robust Target Localization for Multistatic Passive Radar Networks," *IEEE Sensors Letters*, vol. 7, no. 7, pp. 1–4, 2023.
[6] A. Filip and D. Shutin, "Ambiguity function analysis for ofdm-based ldacs passive multistatic radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 3, pp. 1323–1340, 2018.
[7] F. Santi, F. Pieralice, and D. Pastina, "Joint Detection and Localization of Vessels at Sea With a GNSS-Based Multistatic Radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5894–5913, 2019.
[8] J. Liu, H. Li, and B. Himed, "Two Target Detection Algorithms for Passive Multistatic Radar," *IEEE Transactions on Signal Processing*, vol. 62, no. 22, pp. 5930–5939, 2014.
[9] B. Sobhani, E. Paolini, A. Giorgetti, M. Mazzotti, and M. Chiani, "Target Tracking for UWB Multistatic Radar Sensor Networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 1, pp. 125–136, 2014.
[10] M. Boutkhil, A. Driouach, and A. Khamlichi, "Detecting and localizing moving targets using multistatic radar system," *Procedia Manufacturing*, vol. 22, pp. 455–462, 2018, 11th International Conference Interdisciplinarity in Engineering, INTER-ENG 2017, 5-6 October 2017, Tirgu Mures, Romania. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2351978918303640
[11] J. Ma and X. Rui, "Compressive detection of multiple targets in passive bistatic radar," *IET Radar, Sonar & Navigation*, vol. 17, no. 3, pp. 537–544, 2023. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/rsn2.12358
[12] F. Colone, D. W. O'Hagan, P. Lombardo, and C. J. Baker, "A Multistage Processing Algorithm for Disturbance Removal and Target Detection in Passive Bistatic Radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 2, pp. 698–722, 2009.
[13] C. Lu, J. Zheng, T. Yang, R. Chang, and H. Liu, "DOA Estimation Based on Coherent Integration-Sparse Bayesian Learning With Time-Variant Gain-Phase Errors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 6, pp. 7951–7962, 2023.
[14] D. Adamy, *EW101: A First Course in Electronic Warfare*, A. House, Ed., Norwood, MA, 2001.
[15] G. Stimson, H. Griffiths, C. Baker, D. Adamy, and D. Adamy, *Stimson's Introduction to Airborne Radar*, ser. Radar, Sonar and Navigation. SciTech Publishing, 2014.
[16] S. H. Javadi and A. Farina, "Radar networks: A review of features and challenges," *Information Fusion*, vol. 61, pp. 48–55, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253519307535

[17] S. H. Javadi, "Decision fusion: Sparse network vs. dense network," in *2016 24th Iranian Conference on Electrical Engineering (ICEE)*, 2016, pp. 1821–1824.

[18] E. Conte, E. D'Addio, A. Farina, and M. Longo, "Multistatic radar detection: synthesis and comparison of optimum and suboptimum receivers," *IEE Proceedings F (Communications, Radar and Signal Processing)*, vol. 130, pp. 484–494(10), October 1983. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/ip-f-1.1983.0078

[19] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-17, no. 4, p. 501 – 510, 1981.

[20] J. Fang and H. Li, "Hyperplane-based vector quantization for distributed estimation in wireless sensor networks," *IEEE Transactions on Information Theory*, vol. 55, no. 12, p. 5682 – 5699, 2009.

[21] R. Niu, P. K. Varshney, and Q. Cheng, "Distributed detection in a large wireless sensor network," *Information Fusion*, vol. 7, no. 4, pp. 380–394, 2006, special Issue on the Seventh International Conference on Information Fusion-Part I. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253505000710

[22] S. H. Javadi and A. Peiravi, "Fusion of weighted decisions in wireless sensor networks," *IET Wireless Sensor Systems*, vol. 5, no. 2, pp. 97–105, 2015. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-wss.2013.0116

[23] J. Yan, W. Pu, S. Zhou, H. Liu, and Z. Bao, "Collaborative detection and power allocation framework for target tracking in multiple radar system," *Information Fusion*, vol. 55, pp. 173–183, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253518301210

[24] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," *IEEE Control Systems Magazine*, vol. 29, no. 6, pp. 82–100, 2009.

[25] Y. Bar-Shalom and X.-R. Li, *Multitarget-multisensor tracking: principles and techniques*. YBs Storrs, CT, 1995, vol. 19.

[26] S. Nannuru, S. Blouin, M. Coates, and M. Rabbat, "Multisensor cphd filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 4, pp. 1834–1854, 2016.

[27] X. Yang, W.-A. Zhang, L. Yu, and K. Xing, "Multi-rate distributed fusion estimation for sensor network-based target tracking," *IEEE Sensors Journal*, vol. 16, no. 5, pp. 1233–1242, 2015.

[28] K.-C. Chang, R. K. Saha, and Y. Bar-Shalom, "On optimal track-to-track fusion," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 4, pp. 1271–1276, 1997.

[29] G. Battistelli, L. Chisci, C. Fantacci, A. Farina, and A. Graziano, "Consensus cphd filter for distributed multitarget tracking," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 508–520, 2013.

[30] T. Li, J. Prieto, H. Fan, and J. M. Corchado, "A robust multi-sensor phd filter based on multi-sensor measurement clustering," *IEEE Communications Letters*, vol. 22, no. 10, pp. 2064–2067, 2018.

[31] J. Hu, L. Zuo, P. K. Varshney, and Y. Gao, "Resource Allocation for Distributed Multitarget Tracking in Radar Networks With Missing Data," *IEEE Transactions on Signal Processing*, vol. 72, pp. 718–734, 2024.

[32] H. Sun, M. Li, L. Zuo, and P. Zhang, "Resource Allocation for Multitarget Tracking and Data Reduction in Radar Network With Sensor Location Uncertainty," *IEEE Transactions on Signal Processing*, vol. 69, pp. 4843–4858, 2021.

[33] L. Zuo, J. Hu, H. Sun, and Y. Gao, "Resource allocation for target tracking in multiple radar architectures over lossy networks," *Signal Processing*, vol. 208, p. 108973, 2023.

[34] Y. Yuan, W. Yi, and W. Choi, "Dynamic Sensor Scheduling for Target Tracking in Wireless Sensor Networks With Cost Minimization Objective," *IEEE Internet of Things Journal*, vol. 9, no. 21, pp. 20 957–20 974, 2022.

[35] M. Xie, W. Yi, T. Kirubarajan, and L. Kong, "Joint Node Selection and Power Allocation Strategy for Multitarget Tracking in Decentralized Radar Networks," *IEEE Transactions on Signal Processing*, vol. 66, no. 3, pp. 729–743, 2018.

[36] A. Coluccia, A. Fascista, and G. Ricci, "CFAR Feature Plane: A Novel Framework for the Analysis and Design of Radar Detectors," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3903–3916, 2020.

[37] ——, "Design of Customized Adaptive Radar Detectors in the CFAR Feature Plane," *IEEE Transactions on Signal Processing*, vol. 70, pp. 5133–5147, 2022.

[38] L. Scharf and C. Demeure, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, ser. Addison-Wesley series in electrical and computer engineering. Addison-Wesley Publishing Company, 1991.

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

[40] L. Yan, S. Han, C. Hao, D. Orlando, and G. Ricci, "Innovative Cognitive Approaches for Joint Radar Clutter Classification and Multiple Target Detection in Heterogeneous Environments," *IEEE Transactions on Signal Processing*, vol. 71, pp. 1010–1022, 2023.

[41] A. Coluccia, A. Fascista, D. Orlando, and G. Ricci, "Adaptive Radar Detection in Heterogeneous Clutter Plus Thermal Noise via the Expectation-Maximization Algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 1–13, 2023.

[42] J. Zheng, T. Yang, H. Liu, T. Su, and L. Wan, "Accurate Detection and Localization of Unmanned Aerial Vehicle Swarms-Enabled Mobile Edge Computing System," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 5059–5067, 2021.

[43] S. M. Kay, "The multifamily likelihood ratio test for multiple signal model detection," *IEEE Signal Processing Letters*, vol. 12, no. 5, pp. 369–371, 2005.

[44] 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.

[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] P. Stoica and P. Babu, "On the proper forms of bic for model order selection," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4956–4961, September 2012.

[47] ——, "On the exponentially embedded family (eef) rule for model order selection," *IEEE Signal Processing Letters*, vol. 19, no. 9, pp. 551–554, September 2012.

[48] S. Kay, "Exponentially embedded families-new approaches to model order estimation," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 41, no. 1, pp. 333–345, January 2005.

[49] Y. Yang, X. Xiong, and Y. Yan, "UAV Formation Trajectory Planning Algorithms: A Review," *Drones*, vol. 7, no. 1, 2023.

[50] S. Han, P. Addabbo, F. Biondi, C. Clemente, D. Orlando, and G. Ricci, "Innovative Solutions Based on the EM-Algorithm for Covariance Structure Detection and Classification in Polarimetric SAR Images," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 1, pp. 209–227, 2023.

[51] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 1st ed. Springer, 2007.

[52] S. M. Kay, "The Multifamily Likelihood Ratio Test for Multiple Signal Model Detection," *IEEE Signal Processing Letters*, vol. 12, no. 5, pp. 369–371, May 2005.

[53] Y. Wang, Z. Pan, and Y. Pan, "A training data set cleaning method by classification ability ranking for the k-nearest neighbor classifier," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 5, pp. 1544–1556, 2020.

[54] B. Balusamy, N. Abirami R, S. Kadry, and A. H. Gandomi, *Cluster Analysis*, 2021, pp. 259–292.

[55] L. Yan, P. Addabbo, C. Hao, D. Orlando, and A. Farina, "New ECCM Techniques Against Noiselike and/or Coherent Interferers," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 2, pp. 1172–1188, 2020.

Linjie Yan received the B.E. degree in communication engineering from the Shandong University of Science and Technology, in 2016, and the Ph.D. degree in signal and information processing from the Institute of Acoustics, Chinese Academy of Sciences, Beijing, China, in 2021. She is currently a Postdoctoral Researcher with the Institute of Acoustics, Chinese Academy of Sciences, Beijing. Her research interests include statistical signal processing, with more emphasis on adaptive sonar and radar signal processing.
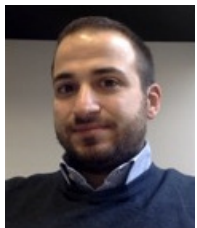
Pia Addabbo (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunication engineering, and the Ph.D. degree in information engineering from the Università degli Studi del Sannio, Benevento, Italy, in 2005, 2008, and 2012, respectively. She is currently an Associate Professor with the Giustino Fortunato University, Benevento, Italy. She is the coauthor of scientific publications in international journals and conferences. Her research interests include statistical signal processing applied to radar target recognition, global navigation satellite system reflectometry, and hyperspectral unmixing. She is currently an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and IEEE ACCESS.

Chengpeng Hao (Senior Member, IEEE) received the B.S. and M.S. degrees in electronic engineering from Beijing Broadcasting Institute, Beijing, China, in 1998 and 2001, respectively, and the Ph.D. degree in signal and information processing from the Institute of Acoustics, Chinese Academy of Sciences, Beijing, China, in 2004. He is currently a Professor with the Autonomous Underwater Vehicles Laboratory of Information Technology, Chinese Academy of Sciences. He has held a visiting position with the Electrical and Computer Engineering Department, Queens University, Kingston, ON, Canada, from July 2013 to July 2014. He authored or coauthored more than 100 journal and conference papers. His research interests are in the fields of statistical signal processing with more emphasis on adaptive sonar and radar signal processing. Dr. Hao is currently an Associate Editor for several international journals, including IEEE ACCESS , the Signal, Image and Video Processing (Springer), and the Open Electrical and Electronic Engineering Journal. He was a Guest Editor for EURASIP Journal on Advances in Signal Processing for the special issue entitled Advanced Techniques for Radar Signal Processing.

Nicomino Fiscante was born in Benevento, Italy, in 1982. He received the M.Sc. degree (cum laude) in telecommunication engineering from the University of Sannio, Benevento, Italy, in 2007, and the Ph.D. degree in applied electronics from the University of Roma TRE, Italy, in 2023. From January to June 2007, he was a Trainee with the Antenna Group, Netherlands Organization for Applied Scientific Research (TNO Defense, Security, and Safety), The Hague, Netherlands. From 2008 to 2012, he was with Mediterranean Agency for Remote Sensing and Environmental Control (MARSec), Benevento, Italy. From 2012 to 2014, he was with Advanced Research Technology Spa in Passignanosul Trasimeno, Italy. Since 2014, he has been a Project Manager with Geoslab (now GeneGIS GI), where he has gained extensive work experience both nationally and internationally through collaborations with major companies, research centers, universities, and institutions. His research interests include remote sensing, GIS, data processing, TLC and EM systems, mobile mapping systems, and renewable resources.

Gaetano Giunta (Senior Member, IEEE) received the degree in electronic engineering from the University of Pisa, Pisa, Italy, in 1985, and the Ph.D. degree in information and communication engineering from the University of Rome La Sapienza, Rome, Italy, in 1990. Since 1989, he has been a Research Fellow with the Signal Processing Laboratory (LTS), EPFL, Lausanne, Switzerland. In 1992, he was an Assistant Professor with the INFO-COM Department, University of Rome La Sapienza. From 2001 to 2005, he was with the University Roma Tre, Rome, as an Associate Professor. Since 2005, he has been a Full Professor of telecommunications with the University Roma Tre, where he currently teaches the courses "Digital Signal Processing," "Signal Processing for Biomedical Engineering," and "Mobile Communications." His research interests include signal processing for mobile communications, image communications, and security. Dr. Giunta has been a Member of the IEEE Societies of Communications, Signal Processing, and Vehicular Technology. He has served as a Reviewer for several IEEE Transactions, IET (formerly IEE) proceedings, and EURASIP journals, and a TPC Member for several international conferences and symposia in the abovementioned fields.

Carmine Clemente (Senior Member, IEEE) received the Laurea (B.Sc.) (cum laude) and Laurea Specialistica (M.Sc.) (cum laude) degrees in telecommunications engineering from the Università degli Studi del Sannio, Benevento, Italy, in 2006 and 2009, respectively, and the Ph.D. degree in electronic and electrical engineering from the University of Strathclyde, Glasgow, U.K., in 2012. He is currently a Senior Lecturer with the Department of Electronic and Electrical Engineering, University of Strathclyde, working on advanced radar signal processing algorithms, micro-Doppler analysis, SAR image processing, and advanced radar applications. He directs the Sensor Signal Processing and Security Laboratories, University of Strathclyde. Research interests of his group include micro-Doppler signature analysis and extraction, waveform design, passive and multistatic radars, automatic target recognition, electronic surveillance, space situation awareness, ballistic missile defense, automotive radars, applications of radars in Industry 4.0, and Agritech and statistical signal processing. Dr. Clemente is the Chair of the Advanced Signal Processing Focus Group of the Electromagnetic Sensing Interest Group-the UK Radar Society, and an Associate Editor for the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS.

Danilo Orlando (Senior Member, IEEE) was born in Gagliano del Capo, Italy, in 1978. Hereceived the Dr. Eng. degree(with honors) in computer engineering and the Ph.D. degree (with maximum score) in information engineering from the University of Salento (formerly University of Lecce), Lecce, Italy, in 2004 and 2008, respectively. From 2007 to 2010, he was with the University of Cassino, Cassino, Italy, engaged in a research project on algorithms for track-before-detect of multiple targets in uncertain scenarios. From September to November 2009, he was a Visiting Scientist with the NATO Undersea Research Centre, La Spezia, Italy. From 2011 to 2015, he was with Elettronica S.p.A. engaged as System Analyst in the field of electronic warfare. From 2015 to 2024, he joined Università degli Studi "Niccolo Cusano," Rome, Italy, as an Associate Professor. From 2024, he is an Associate Professor with the Department of Information Engineering, Università degli Studi di Pisa, Pisa, Italy. In 2007, he has held visiting positions with the Department of Avionics and Systems, ENSICA (now Institut Supérieur de l'Aéronautique et de l'Espace (ISAE)), Toulouse, France, and from 2017 to 2019, with the Chinese Academy of Science, Beijing, China. He has authored or coauthored more than 190 scientific publications in international journals, conferences, and books. His main research interests include statistical signal processing with more emphasis on adaptive detection and tracking of multiple targets in multisensor scenarios. Dr. Orlando was Senior Area Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING and is currently an Associate Editor for IEEE OPEN JOURNAL OF SIGNAL PROCESSING, IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS, EURASIP Journal on Advances in Signal Processing, and Remote Sensing (MDPI).