Detecting Anomalous Deviations
from Standard Maritime Routes
Using the Ornstein-Uhlenbeck Process

Enrica d'Afflisio, Paolo Braca, *Senior Member, IEEE*, Leonardo M. Millefiori, *Member, IEEE* and
Peter Willett, *Fellow, IEEE*

Abstract—A novel anomaly detection procedure based on the Ornstein-Uhlenbeck (OU) mean-reverting stochastic process is presented. The considered anomaly is a vessel that deviates from a planned route, changing its nominal velocity v_0. In order to hide this behavior, the vessel switches off its Automatic Identification System (AIS) device for a time T, and then tries to revert to the previous nominal velocity v_0. The decision that has to be taken is either declaring that a deviation happened or not, relying only upon two consecutive AIS contacts. Furthermore, the extension to the scenario in which multiple contacts (e.g. radar) are available during the time period T is also considered.

A proper statistical hypothesis testing procedure that builds on the changes in the OU process long-term velocity parameter v_0 of the vessel is the core of the proposed approach and enables the solution of the anomaly detection problem. Closed analytical forms are provided for the detection and false alarm probabilities of the hypothesis test.

Index Terms—maritime surveillance, long term prediction of vessel motion, Ornstein-Uhlenbeck process, maritime anomaly detection, statistical hypothesis test, target tracking, automatic identification system, radar, real-world data.

I. INTRODUCTION

MARITIME traffic monitoring mostly relies on data collected by heterogeneous sensor systems, including, e.g. the self-reporting Automatic Identification System (AIS) [1]–[3], coastal radars [4]–[9], space-borne sensors devices such as Synthetic Aperture Radar (SAR) [10], video and infrared cameras [11]. *Stealth activities* [12], in which the perpetrators aim to remain hidden and undetected by law-enforcement bodies throughout the whole duration of the activity, are among the main issues to deal with. Such activities are basically drug smuggling, human trafficking, illegal, unreported or unregulated fishing, illegal immigration, marine pollution and waste dumping. Ships involved in these activities tend to follow set patterns depending on the illicit activity in which they are engaged: deviation from standard routes includes unexpected AIS activity, unexpected port arrival, close
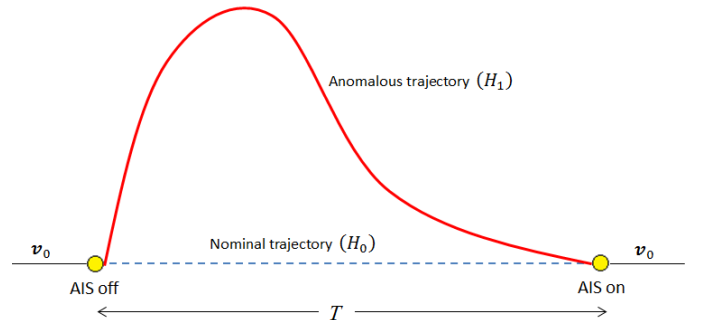
Fig. 1. A vessel turns off the AIS transponder and follows a different trajectory from the nominal one, changing its velocity.

approach, and zone entry [13]. The analysis of real-world AIS data shows that vessels in open seas maneuver very seldom, as it is advantageous for them to travel by the most economical route towards their destination, which is often either the shortest one or the one at constant bearing. Along their route, vessels also seek to optimize fuel consumption, tending to maintain a nearly constant speed. Therefore, the set patterns previously mentioned can be interpreted as deviations from such expected conduct and can therefore be associated to an *anomalous behavior*. As already pointed out in previous relevant works [13]–[19], an anomaly detection strategy is therefore essential.

Ships equipped with AIS transponders broadcast their location, course, speed and other details, such as their destination and ship identifier, at regular intervals. However, as suggested in [18], AIS reports can be counterfeit or vessels could simply turn off their AIS transmitters in order to hinder surveillance systems and operators from detecting illicit activities.

In this work, we study the anomaly detection problem depicted in Fig. 1, where a certain vessel deviates from its planned route, changing the nominal velocity v_0. The vessel can hide the deviation by switching off its AIS transponder for a time period T, and after the deviation it would then try to revert back to the planned route and to the original nominal velocity v_0. The decision that needs to be taken is whether a deviation happened or not, relying upon the available contacts (AIS, radar, etc.).

During the period of silence the vessel might have been loitering or drifting for an unspecified (and unknown) amount of time, perhaps to encounter other ships; all actions commonly

classified as anomalous [20]. As also documented in [12], a practical example could be a mother ship transporting a drug shipment from its place of origin to the waters off the country of the ultimate destination, where the drug is then transferred to a second vessel waiting in a pre-established location, which eventually brings it ashore.

Fig. 2. The track of cargo vessel reveals rendezvous with four fishing vessels in the Pacific.

A real-world example of anomalous behavior is provided in Fig. 2, where the about five-month track of a cargo vessel is shown. The vessel navigates with a nominal speed of about 5 m/s in the waters of the Pacific Ocean [21]. Nearby the Galápagos Exclusive Economic Zone (EEZ), the vessel shuts the engines down and starts drifting, with an apparent deviation from its route. The reason of this deviation is to rendezvous with four tuna longliners at about 1700 miles away from Galápagos. Each fishing vessel spends about 12 hours moving along with the vessel at a distance of about 30 m, which indicates the boats were likely tied up. This behavior suggests a substantial transfer of cargo was possible [21].

Unlike other works [13]–[19], here the anomaly detection problem is addressed relying on a hypothesis testing procedure that builds on the changes of the *Ornstein-Uhlenbeck* (OU) process long-run mean velocity parameter, and such a strategy will be tested against a trajectory of a real transshipment incident [21].

A ship motion model based on the OU process has been shown to be more realistic than other conventional kinematic models for the behavior of the real-world commercial maritime

traffic [3], [10], [22], [23]. In this framework, the OU model turns out to be a valuable tool when vessel information is not available, providing an accurate estimation of a ship's position and velocity, even after several hours.

The detection strategy proposed here is investigated assuming to have available multiple contacts before and after the possible anomaly, with possibly no contacts available during the anomaly itself. The case where only two contacts are available, for example the last contact before the AIS device shutdown and the first after the AIS device reactivation, is of particular interest for real-world applications. The use of multiple heterogeneous contacts associated with the vessel along its trajectory is then considered. This situation could be represented by a scenario where multiple radar contacts are available in addition to AIS contacts. Even though radar and AIS are by now standard ship equipment, the reality is that information from the two systems is seldom used jointly, apart from being manually combined, especially on a local scale by individual ships or port authorities. On a larger scale, at a command and control center level, the opportunity to have complementary observations of a target is possible thanks to long range coastal surveillance radar networks and satellite networks, which provide non-cooperative target measurements and represent the reference application for our investigations.

It is worth mentioning that, different from the AIS data, which contain vessel labeling information, other (especially non-collaborative) sensors (e.g. radar) suffer from the measurement-origin uncertainty [10]. In this work we assume that the association of contacts to the vessel of interest is solved in a preliminary stage, see e.g. [10], [24]. The possible association error, relevant when several multiple targets are close to each other (uncommon scenario in open sea), is neglected and left to future investigation.

The use of multiple contacts compared to the case of only two contacts can lead to a sensible improvement of detection performance. However, counterintuitively, there are scenarios in which the detection performance is degraded. Mathematical conditions and physical insights related to such scenarios are provided.

The paper is organized as follows. The problem is formulated in Section II, while the proposed solution is presented in Section III. Section IV describes the dynamic model employed for the vessel introducing the OU process. Section V is devoted to the development of the detection strategy, experimental results (analyses of both synthetic and real scenarios) are reported in Section VI, and, finally, conclusive remarks are provided in Section VII.

II. PROBLEM FORMULATION

Let us consider a vessel of interest, represented by a point in a two-dimensional space, which is following its planned route. The position and velocity of the vessel are expressed in Cartesian coordinates, resulting from the projection of the geographic coordinates reported by the on-board AIS transponder. The OU stochastic process is used to model the velocity of the vessel with a *long-run* mean parameter v_0, representing the nominal velocity of the ship [3]. In other
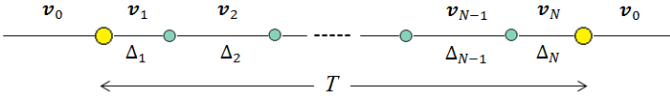
Fig. 3. Sequence of long-run mean velocities and time intervals characterizing the N-section path under hypothesis H_1.

words, the velocity of the vessel is a modified Wiener process so that there is a tendency of the process to move back towards the long-run mean value, with attraction proportional to the current deviation from that long-run mean. The velocity $\boldsymbol{v}_0$ is therefore a deterministic parameter of the OU stochastic process.

Let us now suppose that AIS data is unavailable for a time T after a given instant, due to a lack of communications from the ship (because of limited sensor coverage, interference, etc.) or an intentional shutdown of the AIS transponder. In this context, two hypotheses can be envisioned: the first one, denoted by H_0, that the vessel navigates according to the nominal condition (along the planned trajectory with a long-run mean velocity $\boldsymbol{v}_0$), and the alternative one, denoted by H_1, that the vessel moves away from the nominal condition once the AIS transponder has been shut down. At the end of the time interval T the AIS device is switched back on and the vessel keeps on moving under the nominal condition, as shown in Fig. 1.

We assume that, during the time it *went dark*, the vessel had been moving according to a sequence of OU processes with unknown long-run mean velocities or, equivalently, a single OU process with long-run mean velocity that is a piecewise-constant function of time. In Fig. 3 it is shown the sequence of long-run mean velocities, which identify an N-section path, that we represent with the $2N$-dimensional vector $\boldsymbol{v}_{1:N} = \left[\boldsymbol{v}_1^{\mathrm{T}} \dots \boldsymbol{v}_N^{\mathrm{T}}\right]^{\mathrm{T}}$. Considering the set of time instants $t_1 < \dots < t_n < \dots < t_N$, the period taken to cover the n-th section corresponds to the difference $\Delta_n = t_n - t_{n-1}$ and the sequence of these time intervals is denoted with $\mathcal{D}_N = \{\Delta_n\}_{n=1}^N$.

In other words, under H_1, the vessel velocity is modeled as a *piecewise OU model*, i.e., an OU process with a long-run mean velocity parameter that is a piecewise-constant function of time. The time period T, during which the AIS is disabled, can be expressed as the sum of all the different time intervals Δ_n, so that $T = \sum_{n=1}^N \Delta_n$.

The considered problem amounts to determining, in the absence of AIS data and without any other information during the time interval T, whether the vessel has been following the planned trajectory at the nominal velocity $\boldsymbol{v}_0$ or not, by means of a *composite hypothesis testing formulation* designed to identify changes in the velocity parameter. In the first instance, the problem is studied just relying only upon two contacts available at the instants t_0 and $t_N = T$, respectively. Then, the problem is extended to the case where multiple contacts are available, each one located in a generic point along the N-section path.

III. STATISTICAL HYPOTHESIS TESTING

Let us assume there are $K+1$ measurements available from the target, taken at $K+1$ arbitrary instants of time. Let us also assume that the target measurements can be combined into another vector of K elements, $\boldsymbol{y}$, by suitably taking differences between consecutive measurements. This new vector conveniently encapsulates all the available information about the target trajectory. The specific dependence of $\boldsymbol{y}$ from the target measurements varies depending on the measurement model and hypothesized motion model. Since it is not essential to the considerations that will follow in this section, its definition is delayed to Section IV.

The anomaly detection strategy investigated throughout this paper and fully developed in Section V is the *Generalized Likelihood Ratio Test* (GLRT). By denoting with $\Lambda(\mathbf{y})$ the test function, the GLRT can be expressed as

$$\Lambda(\mathbf{y}) = \arg\max_{\boldsymbol{\theta}} \left\{\ln\left[p_{\boldsymbol{\theta}}(\mathbf{y})\right]\right\} - \ln\left[p_{\boldsymbol{\theta}_0}(\mathbf{y})\right] \underset{H_0}{\overset{H_1}{\gtrless}} \tau, \quad (1)$$

where $p_{\boldsymbol{\theta}_0}$ is the probability distribution under the simple null hypothesis H_0 by which there have been no changes in the vessel velocity during the AIS transponder shutdown; conversely, $p_{\boldsymbol{\theta}}$ is the probability distribution, depending on the unknown parameter $\boldsymbol{\theta}$, under the composite alternative hypothesis H_1 by which the vessel has changed its velocity during the AIS transponder shutdown. With a minor abuse of notation, that we hope the reader will forgive, the threshold will be always identified by τ from now on.

It will be shown that the anomaly detection problem previously described can be traced back to the following Gaussian composite hypothesis testing problem:

$$\begin{cases} H_0 : \ \mathbf{y} \sim \mathcal{N}\left(\boldsymbol{\mu}_0, \mathbf{C}_{\mathbf{y}}\right) \\ H_1 : \ \mathbf{y} \sim \mathcal{N}\left(\boldsymbol{\mu}, \mathbf{C}_{\mathbf{y}}\right) \end{cases} \quad (2)$$

where the mean vectors are given by

$$\boldsymbol{\mu}_0 = \mathbf{H}\left(\mathcal{D}_N\right)\left(\mathbf{1}_N \otimes \boldsymbol{v}_0\right), \qquad \boldsymbol{\mu} = \mathbf{H}\left(\mathcal{D}_N\right)\boldsymbol{v}_{1:N}, \quad (3)$$

and $\mathbf{H}\left(\mathcal{D}_N\right)$ is the model matrix, whose expression will be fully provided in Section IV. The dependence of $\mathbf{H}$ from $\mathcal{D}_N$ will not be reported from now on for ease of notation. $\mathbf{1}_N$ is the vector of N ones, and $\otimes$ represents the Kronecker product.

Noting that $\boldsymbol{v}_{1:N}$ is always unknown, we will make a distinction between the case where the number of sections N and the sequence of the time intervals $\mathcal{D}_N$ are known quantities, and the case where N and $\mathcal{D}_N$ are instead unknown. Given $\boldsymbol{v}_0$, it will be shown that $\boldsymbol{\mu}_0$ is always known even if $\mathbf{H}\left(\mathcal{D}_N\right)$ is unknown. Then the mean term $\boldsymbol{\mu}$ in (3) can be expressed as

$$\boldsymbol{\mu} \triangleq \begin{cases} \mathbf{H}\,\boldsymbol{\theta} & \text{if } N, \mathcal{D}_N \text{ known} \\ \boldsymbol{\theta} & \text{if } N, \mathcal{D}_N \text{ unknown} \end{cases} \quad (4)$$

Under the condition of known parameters, $\mathbf{H}\left(\mathcal{D}_N\right)$ is a known matrix and $\boldsymbol{\theta} = \boldsymbol{v}_{1:N}$ is the unknown parameter vector of size $2N \times 1$. Under the alternative condition, it is preferable to directly estimate $\boldsymbol{\mu}$ (where $\boldsymbol{\theta} = \boldsymbol{\mu}$ represents a global unknown parameter) instead of jointly estimating N, $\mathcal{D}_N$ and

$\boldsymbol{v}_{1:N}$ which would require a numerical solution of the GLRT with no closed-form performance expression available.

The test performance is defined in terms of false alarm probability P_{FA}, i.e., the probability that the test statistic exceeds the threshold under H_0 and detection probability P_{D}, i.e., the probability that the test statistic exceeds the threshold under H_1. By letting the threshold τ vary, the P_{FA} and P_{D} values define a curve in the $(P_{\text{FA}}, P_{\text{D}})$ plane named the *Receiver Operating Characteristic* (ROC). In particular it will be seen that the GLRT for the hypothesis testing problem at hand can be easily traced back to the *GLRT for Gaussian linear model* [25], where the test statistics under the two hypotheses H_0 and H_1 are characterized, respectively, by a central and a non-central Chi-squared distributions, both with d degrees of freedom. In such a way, the detection performance is described by

$$P_{\text{FA}} = Q_{\chi_d^2}(\tau), \tag{5}$$

$$P_{\text{D}} = Q_{\chi_d'^2(\lambda)}(\tau), \tag{6}$$

where $Q_{\chi_d^2}$ and $Q_{\chi_d'^2(\lambda)}$ are the right tail probabilities of the central and non-central Chi-squared distributions, respectively.

IV. STATISTICAL REPRESENTATION OF DATA BASED ON THE ORNSTEIN-UHLENBECK PROCESS

The OU model [3], [26]–[28], validated against a real-world commercial maritime traffic dataset [3], enables a more accurate representation of the target state in the long-term when ships are not maneuvering. The OU process is distinguished from the conventional nearly-constant velocity (NCV) model mainly by a feedback loop, which ensures that the velocity of the target does not diverge with time, but is instead bounded around a desired value, i.e., the velocity of the process tends to drift over time towards its long-term mean.

It is worth mentioning that the NCV model could also be adopted in anomaly detection [14]. It is however less appropriate in two ways, equally important. First, OU appears to model reality better than NCV, and second, the elevated uncertainty growth under the NCV versus OU model means that anomalous excursions stand out less.

Let us indicate the four-dimensional target state at time $t \in \mathbb{R}_0^+$ with

$$\mathbf{s}(t) = \left[\mathbf{x}(t) \ \dot{\mathbf{x}}(t) \right]^{\text{T}}, \tag{7}$$

where $\mathbf{x}(t)$ and $\dot{\mathbf{x}}(t)$ denote the target position and velocity, respectively, in a two-dimensional Cartesian reference system

$$\mathbf{x}(t) \triangleq [x(t) \ y(t)] \ , \ \dot{\mathbf{x}}(t) \triangleq [\dot{x}(t) \ \dot{y}(t)]. \tag{8}$$

The target dynamics, in general, are modeled by a set of linear stochastic differential equations (SDEs) [29], and in [3] it is shown how the movement of real non-maneuvering vessels in the open sea can be represented by a mean-reverting stochastic process. Specifically, the velocity of the target is an OU process, and its position is an Integrated OU (IOU) process. Under this assumption, the SDE for the target motion model has the following form

$$\mathrm{d}\mathbf{s}(t) = \mathbf{A} \, \mathbf{s}(t) \, \mathrm{d}t + \mathbf{G} \, \boldsymbol{v} \, \mathrm{d}t + \mathbf{B} \, \mathrm{d}\boldsymbol{\omega}(t), \tag{9}$$

where $\boldsymbol{v} = [v_x \ v_y]^{\text{T}}$ is the long-run process mean, and $\boldsymbol{\omega}(t)$ is a standard bi-dimensional Wiener process. The matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{G}$ are defined as:

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{0} & -\boldsymbol{\Theta} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\Sigma} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\Theta} \end{bmatrix}, \tag{10}$$

where $\mathbf{0}$ is a 2×2 null matrix, $\boldsymbol{\Sigma}$ is a 2×2 matrix defining the noise process and $\boldsymbol{\Theta}$ is a 2×2 matrix quantifying the mean-reversion effect, meaning the rate at which the target will tend back to the desired velocity after a perturbation; its diagonal terms refer to the x and y components, while the off-diagonals quantify the coupling effect.

Unless otherwise stated, we will use hereafter subscripted indexes to denote time dependency, i.e., $\mathbf{x}_n = \mathbf{x}(t_n)$, $\dot{\mathbf{x}}_n = \dot{\mathbf{x}}(t_n)$ and $\mathbf{s}_n = \mathbf{s}(t_n)$ by definition. We also assume that $\boldsymbol{\Theta}$ has positive and distinct eigenvalues, so that an affine transformation can be found that projects the matrix $\boldsymbol{\Theta}$ onto another space, i.e., $\boldsymbol{\Theta} = \mathbf{R}\boldsymbol{\Gamma}\mathbf{R}^{-1}$, where $\mathbf{R}$ is the matrix whose columns contain the eigenvectors of $\boldsymbol{\Theta}$ and $\boldsymbol{\Gamma}$ is a diagonal matrix whose elements are the corresponding eigenvalues. For the sake of simplicity and without loss of generality we assume that $\mathbf{R} = \mathbf{I}$, so that $\boldsymbol{\Theta} = \boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma})$, with $\boldsymbol{\gamma} = [\gamma_x, \gamma_y]^{\text{T}}$. In short, as described in [3], there are three parameters for each coordinate: the long-run mean velocity $\boldsymbol{v}$, the reversion rate $\boldsymbol{\gamma}$ and the process noise $\boldsymbol{\sigma}$.

Clearly, (9) is only suitable to represent a non-maneuvering target, i.e., whose long-run mean velocity does not change in time. However the model can be easily extended to the case of waypoint navigation [23], that is relevant to our application, being the navigation mode of substantially all the commercial maritime traffic. Under this case, we can assume that the long-run mean velocity of the target is a piecewise-constant function of the time that takes values from a sequence $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N$. Under these assumptions, the target state at time t_i, given the target state at the previous $i-1$ times, can be written in matrix form, as

$$\mathbf{s}_i = \boldsymbol{\Phi}(t_i - t_{i-1}, \boldsymbol{\gamma})\mathbf{s}_{i-1} + \boldsymbol{\Psi}(t_i - t_{i-1}, \boldsymbol{\gamma})\boldsymbol{v}_i + \boldsymbol{\omega}_i, \tag{11}$$

where $\boldsymbol{v}_i$ is the long-run mean velocity in the time interval $\Delta_i = [t_{i-1}, t_i]$ and $\boldsymbol{\omega}_i = \boldsymbol{\omega}(\Delta_i)$ is a zero-mean Gaussian random variable with covariance reported in Appendix A. The state transition matrix and the control input function, $\boldsymbol{\Phi}(\Delta_i, \boldsymbol{\gamma})$ and $\boldsymbol{\Psi}(\Delta_i, \boldsymbol{\gamma})$, respectively, are defined as

$$\boldsymbol{\Phi}(\Delta_i, \boldsymbol{\gamma}) = \begin{bmatrix} \mathbf{I} & \left(\mathbf{I} - e^{-\boldsymbol{\Gamma}\Delta_i}\right) \boldsymbol{\Gamma}^{-1} \\ \mathbf{0} & e^{-\boldsymbol{\Gamma}\Delta_i} \end{bmatrix}, \tag{12}$$

$$\boldsymbol{\Psi}(\Delta_i, \boldsymbol{\gamma}) = \begin{bmatrix} \Delta_i\mathbf{I} - \left(\mathbf{I} - e^{-\boldsymbol{\Gamma}\Delta_i}\right) \boldsymbol{\Gamma}^{-1} \\ \mathbf{I} - e^{-\boldsymbol{\Gamma}\Delta_i} \end{bmatrix}. \tag{13}$$

The target state at time t_N, given the target states at the previous $N-1$ times, can be expressed recursively as (omitting

$$\mathbf{y}_k = \mathbf{z}_k - \mathbf{\Phi}(T_k)\mathbf{z}_0$$

$$= \mathbf{\Phi}(\Delta_{\lceil p_k \rceil}\delta_k)\underbrace{\left\{ \sum_{n=1}^{\lfloor p_k \rfloor - 1}\left[\prod_{i=n+1}^{\lfloor p_k \rfloor}\mathbf{\Phi}(\Delta_i)\right]\mathbf{\Psi}(\Delta_n)\boldsymbol{v}_n + \mathbf{\Psi}(\Delta_{\lfloor p_k \rfloor})\boldsymbol{v}_{\lfloor p_k \rfloor}\right\} + \mathbf{\Psi}(\Delta_{\lceil p_k \rceil}\delta_k)\boldsymbol{v}_{\lceil p_k \rceil} + \boldsymbol{\omega}(T_k) + \mathbf{n}_k - \mathbf{\Phi}(T_k)\mathbf{n}_0}_{\triangleq \boldsymbol{\mu}_k}$$

$$= \boldsymbol{\mu}_k + \boldsymbol{\omega}(T_k) + \mathbf{n}_k - \mathbf{\Phi}(T_k)\mathbf{n}_0 \sim \mathcal{N}\left(\boldsymbol{\mu}_k, \mathbf{C}_{\mathbf{y}_k}\right). \tag{14}$$

$\boldsymbol{\gamma}$ for clarity)

$$\mathbf{s}_N = \mathbf{\Phi}(\Delta_N)\mathbf{s}_{N-1} + \mathbf{\Psi}(\Delta_N)\boldsymbol{v}_N + \boldsymbol{\omega}_N$$
$$= \mathbf{\Phi}(\Delta_N)\left[\mathbf{\Phi}(\Delta_{N-1})\mathbf{s}_{N-2} + \mathbf{\Psi}(\Delta_{N-1})\boldsymbol{v}_{N-1} + \boldsymbol{\omega}_{N-1}\right]$$
$$+ \mathbf{\Psi}(\Delta_N)\boldsymbol{v}_N + \boldsymbol{\omega}_N$$
$$= \mathbf{\Phi}(T)\mathbf{s}_0 + \mathbf{\Psi}(\Delta_N)\boldsymbol{v}_N + \boldsymbol{\omega}_N$$
$$+ \sum_{n=1}^{N-1}\left[\prod_{i=n+1}^{N}\mathbf{\Phi}(\Delta_i)\right]\left[\mathbf{\Psi}(\Delta_n)\boldsymbol{v}_n + \boldsymbol{\omega}_n\right], \tag{15}$$

where we exploited the property of the state transition matrix by which $\mathbf{\Phi}(\Delta_1)\mathbf{\Phi}(\Delta_2)...\mathbf{\Phi}(\Delta_n) = \mathbf{\Phi}(\Delta_1 + \Delta_2 + ... + \Delta_n)$, $\forall n = 1, ..., N$, that can be derived by inspection from (12).

A. Two contacts available ($K = 1$)

Let us start by considering the case in which two contacts are available. Specifically, we denote with $\mathbf{z}$ and $\mathbf{z}_0$ the two available measurements, respectively, at time T and time t_0

$$\mathbf{z} = \mathbf{s}(T) + \mathbf{n}, \quad \mathbf{z}_0 = \mathbf{s}_0 + \mathbf{n}_0,$$

where $\mathbf{n}$ and $\mathbf{n}_0$ are independent zero-mean Gaussian noises with covariance matrices $\mathbf{C}_\mathbf{n}$ and $\mathbf{C}_{\mathbf{n}_0}$, respectively. Clearly, the measurement noise is independent of the OU process noise. Even if the distribution of $\mathbf{z}_0$ does not affect the hypothesis test, meaning that $\mathbf{s}_0$ has the same distribution under both hypotheses, such information is important because $\mathbf{s}_0$ represents the starting point of the kinematic terminating in $\mathbf{s}(T)$. Since $\mathbf{z}_0$ is the maximum likelihood estimate of $\mathbf{s}_0$, then it is possible to substitute $\mathbf{z}_0$ in the generalized likelihood of $\mathbf{z}$. Given the linearity of the previous equations and the fact that $\mathbf{z}$ is Gaussian, we can use the following vector of data to avoid the dependence on $\mathbf{s}_0$ in $\mathbf{z}$

$$\mathbf{y} = \mathbf{z} - \mathbf{\Phi}(T)\mathbf{z}_0$$
$$= \mathbf{\Psi}(\Delta_N)\boldsymbol{v}_N + \boldsymbol{\omega}(\Delta_N)$$
$$+ \sum_{n=1}^{N-1}\left[\prod_{i=n+1}^{N}\mathbf{\Phi}(\Delta_i)\right]\left[\mathbf{\Psi}(\Delta_n)\boldsymbol{v}_n + \boldsymbol{\omega}(\Delta_n)\right]$$
$$+ \mathbf{n} - \mathbf{\Phi}(T)\mathbf{n}_0. \tag{16}$$

The terms $\boldsymbol{\omega}_n = \boldsymbol{\omega}(\Delta_n)$ are independent zero-mean Gaussian random variables with covariance $\mathbf{C}(\Delta_n)$. It is shown in Appendix A that

$$\boldsymbol{\omega}(T) \triangleq \sum_{n=1}^{N-1}\left[\prod_{i=n+1}^{N}\mathbf{\Phi}(\Delta_i)\right]\boldsymbol{\omega}(\Delta_n) + \boldsymbol{\omega}(\Delta_N) \sim \mathcal{N}\left(\mathbf{0}, \mathbf{C}(T)\right)$$
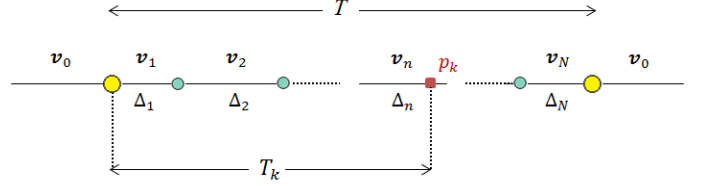$$\tag{17}$$

Fig. 4. The parameter p_k identifies the time location of the k-th contact along the N-section path covered by the vessel.

so that $\mathbf{y}$ can be recast as follows

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\omega}(T) + \mathbf{n} - \mathbf{\Phi}(T)\mathbf{n}_0$$
$$\sim \mathcal{N}\left(\boldsymbol{\mu}, \mathbf{C}(T) + \mathbf{C}_\mathbf{n} + \mathbf{\Phi}(T)\mathbf{C}_{\mathbf{n}_0}\mathbf{\Phi}(T)^{\mathrm{T}}\right), \tag{18}$$

where the expression of the mean term $\boldsymbol{\mu}$ can be easily traced back to the matrix format (3) as

$$\boldsymbol{\mu} = \sum_{n=1}^{N-1}\left[\prod_{i=n+1}^{N}\mathbf{\Phi}(\Delta_i)\right]\mathbf{\Psi}(\Delta_n)\boldsymbol{v}_n + \mathbf{\Psi}(\Delta_N)\boldsymbol{v}_N$$
$$= \mathbf{H}\,\boldsymbol{v}_{1:N}, \tag{19}$$

where in this case $\mathbf{H}$ is a single $4 \times 2N$ matrix incorporating the state transition matrices and the control input functions, dependent on the number of sections N and the time interval sequence $\mathcal{D}_N$:

$$\mathbf{H} \triangleq \underbrace{\left[\mathbf{\Phi}\left(\sum_{i=2}^{N}\Delta_i\right)\mathbf{\Psi}(\Delta_1), \ldots, \mathbf{\Phi}(\Delta_N)\mathbf{\Psi}(\Delta_{N-1}), \mathbf{\Psi}(\Delta_N)\right]}_{4 \times 2N}.$$
$$\tag{20}$$

Drawing a distinction between the two cases in which parameters N and $\mathcal{D}_N$ are known or not, $\boldsymbol{\mu}$ can be expressed as in (4). In the case of known parameters, such a matrix is known and $\boldsymbol{\theta} = \boldsymbol{v}_{1:N}$. Under the null hypothesis H_0, by defining the vector $\boldsymbol{\theta}_0 \triangleq \mathbf{1}_N \otimes \boldsymbol{v}_0$, it is possible to show that $\boldsymbol{\mu}_0 = \mathbf{\Psi}(T)\boldsymbol{v}_0 = \mathbf{H}\boldsymbol{\theta}_0$ by exploiting the following equality

$$\sum_{n=1}^{N-1}\left[\prod_{i=n+1}^{N}\mathbf{\Phi}(\Delta_i)\right]\mathbf{\Psi}(\Delta_n) + \mathbf{\Psi}(\Delta_N) = \mathbf{\Psi}(T). \tag{21}$$

In the case of N and $\mathcal{D}_N$ unknown, $\boldsymbol{\theta} = \boldsymbol{\mu}$ given in (19) and $\boldsymbol{\theta}_0 = \boldsymbol{\mu}_0 = \mathbf{\Psi}(T)\boldsymbol{v}_0$.

$$\mathbf{P}_k = \underbrace{\left[\mathbf{\Phi}(\Delta_{\lceil p_k \rceil}\delta_k)\mathbf{\Phi}\left(\sum_{i=2}^{\lfloor p_k \rfloor}\Delta_i\right)\mathbf{\Psi}(\Delta_1), \ldots, \mathbf{\Phi}(\Delta_{\lceil p_k \rceil}\delta_k)\mathbf{\Phi}\left(\sum_{i=\lceil p_k \rceil - 1}^{\lfloor p_k \rfloor}\Delta_i\right)\mathbf{\Psi}(\Delta_{\lceil p_k \rceil - 2}), \mathbf{\Phi}(\Delta_{\lceil p_k \rceil}\beta_k)\mathbf{\Psi}(\Delta_{\lceil p_k \rceil - 1}), \mathbf{\Psi}(\Delta_{\lceil p_k \rceil}\beta_k),}_{4 \times 2\lceil p_k \rceil} \underbrace{\begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}}_{4 \times 2(N - \lceil p_k \rceil)} \right]$$
(22)

B. Multiple contacts available

Let us now assume that, in addition to the contacts in t_0 and T, a number of contacts is available at any time during the period between t_0 and T. We hence assume that we have a stack of $K + 1$ measurements, as $[\mathbf{z}_0, \ldots, \mathbf{z}_k, \ldots, \mathbf{z}_K]^\mathrm{T}$, where the k-th measurement is given by $\mathbf{z}_k = \mathbf{s}(T_k) + \mathbf{n}_k$, and the measurement noise terms $\mathbf{n}_k$ are assumed to be independent and identically distributed according to a zero-mean Gaussian with covariance $\mathbf{C}_{\mathbf{n}_k}$. Accordingly, it is possible to consider the vector $\mathbf{y}$ of size $4K$, defined as $\mathbf{y} = [\mathbf{y}_1, \ldots, \mathbf{y}_k, \ldots, \mathbf{y}_K]^\mathrm{T}$.

The k-th measurement is available at time $T_k = \frac{p_k}{N}T$, where $p_k \in [0, N]$ is by definition a fraction of the interval $[0, N]$ representing the time location of the contact with respect to the N piecewise OU velocities, as shown in Fig. 4, where p_k is located at some point along the n-th section of the path. The k-th vector, $\mathbf{y}_k$, is defined in (14), and its components are analyzed hereafter.

Similarly to (17) we can compute the OU process noise $\boldsymbol{\omega}(T_k)$ at time T_k, which is given by

$$\boldsymbol{\omega}(T_k) \triangleq \mathbf{\Phi}(\Delta_{\lceil p_k \rceil}\delta_k)\sum_{n=1}^{\lfloor p_k \rfloor - 1}\left[\prod_{i=n+1}^{\lfloor p_k \rfloor}\mathbf{\Phi}(\Delta_i)\right]\boldsymbol{\omega}(\Delta_n)$$
$$+ \mathbf{\Phi}(\Delta_{\lceil p_k \rceil}\delta_k)\boldsymbol{\omega}(\Delta_{\lfloor p_k \rfloor}) + \boldsymbol{\omega}(\Delta_{\lceil p_k \rceil}\delta_k)$$
$$\sim \mathcal{N}\left(\mathbf{0}, \mathbf{C}(T_k)\right),$$
(23)

where $\delta_k = p_k - \lfloor p_k \rfloor$, with $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ denoting respectively the ceiling and the floor functions.

Consequently, the vector $\mathbf{y}_k$ is Gaussian with mean $\boldsymbol{\mu}_k$, defined in (14) and derived in the same way as the mean term (19) in the case of only two contacts, and covariance matrix given by

$$\mathbf{C}_{\mathbf{y}_k} \triangleq \mathbf{C}(T_k) + \mathbf{C}_{\mathbf{n}_k} + \mathbf{\Phi}(T_k)\mathbf{C}_{\mathbf{n}_0}\mathbf{\Phi}(T_k)^\mathrm{T},$$

assuming the independence of the noise terms in (14).

At this point we can proceed as done in the previous section computing the relation between the hypotheses and the vector of data. The mean term $\boldsymbol{\mu}_k$ in (14) under hypothesis H_1 can be traced back to the matrix format (4) as

$$\boldsymbol{\mu}_k = \mathbf{P}_k \, \boldsymbol{v}_{1:N},$$
(24)

where $\mathbf{P}_k$ is the k-th sub-matrix constituting $\mathbf{H}$, which includes the state transition matrices and the control input functions related to the k-th radar contact, whose expression is specified in (22), where

$$\beta_k = \begin{cases} 1 & \text{if } \delta_k = 0 \\ \delta_k & \text{if } \delta_k \neq 0 \end{cases}.$$

Specifically, the null matrix appearing in (22) cancels the long-run mean velocities contributing later than the time T_k of the k-th contact. Notice that, for $p_k = N$, $\mathbf{P}_k$ is just the matrix (20) found in the previous case where only two contacts are available.

The vector of data $\mathbf{y}$ is therefore characterized as follows

$$\boldsymbol{\mu} = E[\mathbf{y} \mid H_1] = \mathbf{H}\,\boldsymbol{v}_{1:N} = \begin{bmatrix} \mathbf{P}_1 \\ \vdots \\ \mathbf{P}_K \end{bmatrix}\boldsymbol{v}_{1:N},$$
(25)

$$\mathbf{C}_{\mathbf{y}} \triangleq \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1K} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \cdots & \mathbf{C}_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_{K1} & \mathbf{C}_{K2} & \cdots & \mathbf{C}_{KK} \end{bmatrix}$$
(26)

where $\forall i, j = 1, \ldots, K$ (see Appendix C)

$$\mathbf{C}_{ij} = \begin{cases} \mathbf{C}(T_i) + \mathbf{C}_{\mathbf{n}_i} + \mathbf{\Phi}(T_i)\mathbf{C}_{\mathbf{n}_0}\mathbf{\Phi}(T_i)^\mathrm{T} & \text{if } i = j \\ \mathbf{C}(T_i)\mathbf{\Phi}(T_j - T_i)^\mathrm{T} + \mathbf{\Phi}(T_i)\mathbf{C}_{\mathbf{n}_0}\mathbf{\Phi}(T_j)^\mathrm{T} & \text{if } i < j \end{cases}$$

and $\mathbf{C}_{ji} = \mathbf{C}_{ij}^\mathrm{T}$.

Under the null hypothesis H_0 the mean term is given by

$$\boldsymbol{\mu}_0 = E[\mathbf{y} \mid H_0] = \mathbf{\Psi}_{\mathbf{y}}\boldsymbol{v}_0,$$
(27)

being $\mathbf{\Psi}_{\mathbf{y}} \triangleq [\mathbf{\Psi}(T_1), \mathbf{\Psi}(T_2), \ldots, \mathbf{\Psi}(T_K)]^\mathrm{T}$ the control input matrix related to the total information available from the K contacts. Exploiting equality (21) we can recast $\boldsymbol{\mu}_0$ in terms of $\mathbf{H}$, as follows

$$\boldsymbol{\mu}_0 = \mathbf{H}\left(\mathbf{1}_N \otimes \boldsymbol{v}_0\right).$$
(28)

Exploiting equations (25)-(28) $\boldsymbol{\mu}$ can be expressed as in (4), with $\boldsymbol{\theta} = \boldsymbol{v}_{1:N}$ and $\boldsymbol{\theta}_0 = \mathbf{1}_N \otimes \boldsymbol{v}_0$ when the $\mathbf{H}$ is known (or equivalently when N and $\mathcal{D}_N$ are known), otherwise $\boldsymbol{\theta} = \boldsymbol{\mu}$ and $\boldsymbol{\theta}_0 = \boldsymbol{\mu}_0$.

V. DETECTION STRATEGY PERFORMANCE

In this section we develop the detection strategy for the case where multiple contacts are available, making a distinction between the case where N and $\mathcal{D}_N$ are unknown and the one where such parameters are known.

In particular, the detection strategy for the hypothesis testing problem at hand is based on the GLRT approach that can be easily traced back to the GLRT for Gaussian linear model [25], as anticipated in Section III. Finally the issue arising from this strategy will be highlighted and investigated.

A. Scenario 1: N and $\mathcal{D}_N$ unknown parameters

In this case we have $\boldsymbol{\mu} = \boldsymbol{\theta} = \mathbf{H}\,\boldsymbol{v}_{1:N}$ under H_1 and $\boldsymbol{\mu}_0 = \boldsymbol{\theta}_0 = \mathbf{H}\,(\mathbf{1}_N \otimes \boldsymbol{v}_0)$ under H_0. The GLRT is derived from the GLRT (1), introduced in Section III, as follows

$$\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)^{\mathrm{T}} \mathbf{C}_{\mathbf{y}}^{-1} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \underset{H_0}{\overset{H_1}{\gtrless}} \tau, \tag{29}$$

where $\widehat{\boldsymbol{\theta}}$ represents the ML estimate for the parameter $\boldsymbol{\theta}$. In Appendix B it is shown that $\widehat{\boldsymbol{\theta}} = \mathbf{y}$.

As anticipated in Section III, the test statistics under the two hypotheses H_0 and H_1 are characterized, respectively, by central and non-central Chi-squared distributions, both with $d = 4K$ degrees of freedom, corresponding to the size of the known parameter $\boldsymbol{\theta}_0$. The detection performance is therefore given by (5) and (6) with the following non-centrality parameter

$$\lambda = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^{\mathrm{T}} \mathbf{C}_{\mathbf{y}}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0), \tag{30}$$
$$= (\boldsymbol{v}_{1:N} - \mathbf{1}_N \otimes \boldsymbol{v}_0)^{\mathrm{T}} \mathbf{H}^{\mathrm{T}} \mathbf{C}_{\mathbf{y}}^{-1} \mathbf{H} \, (\boldsymbol{v}_{1:N} - \mathbf{1}_N \otimes \boldsymbol{v}_0).$$

B. Scenario 2: N and $\mathcal{D}_N$ known parameters

In this case $\boldsymbol{\theta} = \boldsymbol{v}_{1:N}$ under H_1 and $\boldsymbol{\theta}_0 = \mathbf{1}_N \otimes \boldsymbol{v}_0$ under H_0. Similarly to what seen in the previous case, the GLRT has the following form:

$$\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)^{\mathrm{T}} \mathbf{H}^{\mathrm{T}} \mathbf{C}_{\mathbf{y}}^{-1} \mathbf{H} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \underset{H_0}{\overset{H_1}{\gtrless}} \tau, \tag{31}$$

where $\widehat{\boldsymbol{\theta}}$ is the ML estimate of the parameter $\boldsymbol{\theta}$, given by (see Appendix B)

$$\widehat{\boldsymbol{\theta}} = \left(\mathbf{H}^{\mathrm{T}} \mathbf{C}_{\mathbf{y}}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^{\mathrm{T}} \mathbf{C}_{\mathbf{y}}^{-1} \mathbf{y}, \tag{32}$$

when $\mathbf{H}$ is a full rank matrix. Therefore it is possible to achieve a direct estimate of the vector incorporating the velocities assumed by the vessel under the hypothesis H_1 instead of an estimate of the global parameter as shown in the previous case[1]. The detection performance is given by (5) and (6), but in the current case the number of degrees of freedom for both central and non-central Chi-squared distributions is $d = 2N$, corresponding to the size of the known parameter $\boldsymbol{\theta}_0$, and the non-centrality parameter is given by (30).

On the other hand, when $\mathbf{H}$ is ill-conditioned, a problem of matrix inversion arises in the ML estimate expression (32). This is identified as a *rank deficiency* problem that can be approached by using the *Singular Value Decomposition* (SVD) [31] of matrix $\mathbf{H}$, given by

$$\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}}, \tag{33}$$

where $\mathbf{U}$ is a $4K \times 4K$ unitary matrix, $\mathbf{S}$ is a $4K \times 2N$ rectangular diagonal matrix with non-negative real numbers on the diagonal, corresponding to the singular values of $\mathbf{H}$, and $\mathbf{V}$ is a $2N \times 2N$ unitary matrix.

The central idea is to replace $\mathbf{H}$ by its rank reduced version, and this is referred to as truncating the SVD [30]. It is shown that $p \overset{\triangle}{=} \mathrm{rank}(\mathbf{H}) = \mathrm{rank}(\mathbf{S})$ and it is equal to the number

---

[1]When $\mathbf{H}$ holds a rank equal to the number of its rows, we fall back into the case where $\mathbf{H}$ is unknown, so that $\mathbf{H}\widehat{\boldsymbol{\theta}} = \mathbf{y}$, as shown in [30].

---

of non-zero singular values of $\mathbf{H}$. Therefore p provides the effective size of the matrices involved, meaning that $\mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}} = \widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widehat{\mathbf{V}}^{\mathrm{T}}$, where $\widetilde{\mathbf{U}}$, of size $4K \times p$, and $\widetilde{\mathbf{V}}^{\mathrm{T}}$, of size $p \times 2N$, are respectively the matrices of the left-singular vectors and of the right-singular vectors of $\mathbf{H}$, while $\widetilde{\mathbf{S}}$ is a $p \times p$ diagonal matrix. In such a way, by using the reduced rank version of $\mathbf{H}$, it is possible to get

$$\mathbf{H}\boldsymbol{\theta} = \widetilde{\mathbf{U}}\widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^{\mathrm{T}}\boldsymbol{\theta} = \widetilde{\mathbf{U}}\widetilde{\boldsymbol{\theta}}, \tag{34}$$

where $\widetilde{\boldsymbol{\theta}} = \widetilde{\mathbf{S}}\widetilde{\mathbf{V}}^{\mathrm{T}}\boldsymbol{\theta}$ is a p-dimensional unknown vector resulting from a process of rotation and scaling of the vector $\boldsymbol{\theta}$. In the same way, under hypothesis H_0, we obtain $\mathbf{H}\boldsymbol{\theta}_0 = \widetilde{\mathbf{U}}\widetilde{\boldsymbol{\theta}}_0$.

By applying the SVD (34) to the GLRT (31) and considering

$$\mathbf{C}_{\mathbf{y}}^{-1} = \widetilde{\mathbf{C}}^{\mathrm{T}}\widetilde{\mathbf{C}},$$

with $\widetilde{\mathbf{C}}$ lower triangular matrix with positive elements on the diagonal, derived from the *Cholesky Decomposition* [31], the reformulation the GLRT is achieved as

$$\left(\widehat{\widetilde{\boldsymbol{\theta}}} - \widetilde{\boldsymbol{\theta}}_0\right)^{\mathrm{T}} \widetilde{\mathbf{U}}^{\mathrm{T}}\widetilde{\mathbf{C}}^{\mathrm{T}}\widetilde{\mathbf{C}}\widetilde{\mathbf{U}} \left(\widehat{\widetilde{\boldsymbol{\theta}}} - \widetilde{\boldsymbol{\theta}}_0\right)$$
$$= \left(\widehat{\widetilde{\boldsymbol{\theta}}} - \widetilde{\boldsymbol{\theta}}_0\right)^{\mathrm{T}} \widetilde{\mathbf{H}}^{\mathrm{T}}\widetilde{\mathbf{H}} \left(\widehat{\widetilde{\boldsymbol{\theta}}} - \widetilde{\boldsymbol{\theta}}_0\right) \underset{H_0}{\overset{H_1}{\gtrless}} \tau, \tag{35}$$

where $\widetilde{\mathbf{H}} = \widetilde{\mathbf{C}}\widetilde{\mathbf{U}}$ and $\widehat{\widetilde{\boldsymbol{\theta}}}$ is the ML estimate of parameter $\widetilde{\boldsymbol{\theta}}$, given by

$$\widehat{\widetilde{\boldsymbol{\theta}}} = \left(\widetilde{\mathbf{H}}^{\mathrm{T}}\widetilde{\mathbf{H}}\right)^{-1} \widetilde{\mathbf{H}}^{\mathrm{T}}\widetilde{\mathbf{C}}\mathbf{y}, \tag{36}$$

where the problem of matrix inversion does not arise since $\widetilde{\mathbf{H}}^{\mathrm{T}}\widetilde{\mathbf{H}}$ is a full-rank matrix $p \times p$ by construction, and therefore invertible. In this case the number of degrees of freedom is $d = p$ corresponding to the size of the known parameter $\widetilde{\boldsymbol{\theta}}_0$, while the non-centrality parameter is always given by (30).

C. More on the detection performance

Assuming that we know the matrix $\mathbf{H}$ (see scenario 2) we expect that the detector (35) outperforms the detector (29) designed under the assumption that $\mathbf{H}$ is unknown (see scenario 1). Indeed, given that the parameter λ is equal for both the scenarios, the number of degrees of freedom d makes the difference. In scenario 1, $d = 4K$, while, for scenario 2, d depends on the structure of $\mathbf{H}$ (or equivalently of contact time locations): $d = p = \mathrm{rank}(\mathbf{H}) \leq \min(4K, 2N)$. Then, scenario 2 is never worse than scenario 1 when $p < 4K$ and equivalently when $p = 4K$.

In both scenarios by increasing K we would like to obtain an improvement of performance. Unfortunately, this is not guaranteed. However, when K is large enough and the structure of $\mathbf{H}$ allows λ to increase with K, the scenario 2 has performance improving with K, since d is bounded by $2N$, implying that $P_{\mathrm{D}} \longrightarrow 1$ for any fixed P_{FA}.

Conversely, in scenario 1, when λ increases with K, the performance is not guaranteed to improve as $d = 4K$ increases as well with K. In the following we give the scaling law of λ_K to obtain improving performance of the detection strategy in scenario 1.

Scaling law of λ_K

Let us indicate $d_K = 4K$. The decision statistic (29) has a Chi-squared distribution, that is equivalent to the sum of squares of d_K independent Normal random variables x_i^2 with finite mean under H_1 (zero mean under H_0) and unit variance

$$\mathcal{Q}_K = \sum_{i=1}^{d_K} x_i^2$$

$$\begin{cases} H_0: & x_i \sim \mathcal{N}(0,1), & \mathcal{Q}_K \sim \chi^2_{d_K} \\ H_1: & x_i \sim \mathcal{N}(\xi_i,1), & \mathcal{Q}_K \sim \chi'^2_{d_K}(\lambda_K), \end{cases}$$

where $\lambda_K = \sum_{i=1}^{d_K} \xi_i^2$. Moreover, we have

$$E[\mathcal{Q}_K] = \begin{cases} d_K & H_0 \\ \lambda_K + d_K & H_1 \end{cases}$$

$$Var[\mathcal{Q}_K] = \begin{cases} 2d_K & H_0 \\ 2(d_K + 2\lambda_K) & H_1 \end{cases}$$

Exploiting the *central limit theorem* [32] and normalizing the decision statistic Q_K under H_0, we obtain the normalized decision statistic

$$\widetilde{\mathcal{Q}}_K = \frac{\mathcal{Q}_K - d_K}{\sqrt{2d_K}} \underset{H_0}{\overset{H_1}{\gtrless}} \tau. \tag{37}$$

Such a distribution converges to Gaussian for large K:

$$\begin{cases} H_0: & \widetilde{\mathcal{Q}}_K \longrightarrow \mathcal{N}(0,1) \\ H_1: & \widetilde{\mathcal{Q}}_K \longrightarrow \mathcal{N}\left(\widetilde{\lambda}_\infty, \widetilde{\sigma}_\infty^2\right). \end{cases}$$

Under hypothesis H_1, $\widetilde{\mathcal{Q}}_K$ can be written as

$$\underbrace{\frac{\mathcal{Q}_K - \lambda_K - d_K}{\sqrt{2(d_K + 2\lambda_K)}}}_{\widetilde{\mathcal{Q}}_K^{(H_1)} \longrightarrow \mathcal{N}(0,1)} \underbrace{\sqrt{\frac{d_K + 2\lambda_K}{d_K}}}_{\widetilde{\sigma}_K \longrightarrow \widetilde{\sigma}_\infty} + \underbrace{\frac{\lambda_K}{\sqrt{2d_K}}}_{\widetilde{\lambda}_K \longrightarrow \widetilde{\lambda}_\infty}, \tag{38}$$

where the first limit is a convergence in distribution to a normal random variable while the others are limit of deterministic sequences. Summarizing we have

$$\widetilde{\lambda}_\infty = \lim_{K \to \infty} \widetilde{\lambda}_K = \lim_{K \to \infty} \frac{\lambda_K}{\sqrt{2d_K}},$$

$$\widetilde{\sigma}_\infty^2 = \lim_{K \to \infty} \widetilde{\sigma}_K^2 = 1 + \lim_{K \to \infty} 2\frac{\lambda_K}{d_K}.$$

Exploiting the previous convergence properties, we can analyze the asymptotic detection performance of the decision statistic $\tilde{Q}_K$, specifically we have

$$P_D^\infty = \lim_{K \to \infty} P_{D,K} = \lim_{K \to \infty} P\left(\widetilde{\mathcal{Q}}_K > \tau | H_1\right) \tag{39}$$

$$= Q\left(\frac{Q^{-1}(P_{FA}^\infty) - \widetilde{\lambda}_\infty}{\widetilde{\sigma}_\infty}\right), \tag{40}$$

where P_{FA}^∞ is the asymptotic false alarm probability, obtained exploiting the convergence in (37) and $\tau = Q^{-1}(P_{FA}^\infty)$.

Let us recall that $d_K = \mathcal{O}(K)$ and assume $\widetilde{\lambda}_K = \mathcal{O}(K^{\frac{n}{2}})$. We distinguish three different cases depending on the parameter $n \geq 0$. If $0 \leq n < 1$, the test cannot distinguish H_0 from H_1 because $\widetilde{\lambda}_\infty = 0$ and $\widetilde{\sigma}_\infty = 1$, consequently from (40)

$P_D^\infty = P_{FA}^\infty$. If $n = 1$ then $\widetilde{\lambda}_\infty < \infty$, $\widetilde{\sigma}_\infty = 1$ and the detection probability converges to a value less than 1 provided by (40). If $n > 1$ and $\tau < \infty$ then under H_1 dividing the decision statistic (38) by $\widetilde{\lambda}_K$ we obtain

$$\frac{\widetilde{\mathcal{Q}}_K}{\widetilde{\lambda}_K} = \underbrace{\widetilde{\mathcal{Q}}_K^{(H_1)}}_{\longrightarrow \mathcal{N}(0,1)} \underbrace{\widetilde{\sigma}_K / \widetilde{\lambda}_K}_{\longrightarrow 0} + 1,$$

implying that $\widetilde{\mathcal{Q}}_K / \widetilde{\lambda}_K$ converges to one in probability under H_1. Then, following (39) and given that $\tau / \widetilde{\lambda}_K \longrightarrow 0$, the detection probability converges to one. Summarizing for any $P_{FA}^\infty < 1$ we have:

$$\begin{cases} P_D^\infty & = P_{FA}^\infty & \text{if } 0 \leq n < 1, \\ P_D^\infty & = Q\left(Q^{-1}(P_{FA}^\infty) - \widetilde{\lambda}_\infty\right) & \text{if } n = 1, \\ P_D^\infty & = 1 & \text{if } n > 1. \end{cases}$$

Remark

Note that the sequence ξ_i, for $i = 1, \ldots, d_K$, where in general $\xi_i \neq \xi_j$ for $i \neq j$, represents the heterogeneity of the contacts. Each ξ_i is a measure of the information contained in a component of a single measurement (position or velocity along one of the Cartesian axes). Clearly, if $\xi_i \ll 1$ this component is adding mostly noise to the decision statistic degrading the detection performance. This concept will be clear in Section VI.

VI. Experimental results

A. Analysis of a synthetic scenario

Performance of tests (29) and (35) will be examined for a situation of concern where multiple contacts are considered. We set a configuration for the analyses of a synthetic scenario, where a vessel is navigating under nominal conditions along a straight route with velocity $\boldsymbol{v}_0 = [8 \; 0]^T$ m/s and at some point it turns its AIS device off for a time period $T = 12$ h. Under hypothesis H_1 the vessel is supposed to follow an N-section path and the time intervals are assumed all equal, so that $\Delta_n = T/N \; \forall n = 1, ..., N$. The reversion rate of the underlying OU dynamic model is set as $\gamma_x = \gamma_y = 0.9 \cdot 10^{-2}$ and $\boldsymbol{\Sigma\Sigma}^T = \sigma^2 \mathbf{I}$, with noise level $\sigma^2 = 10^{-2}$. The noise covariance matrix is set as $\mathbf{C}_{\boldsymbol{n}_k} = \text{diag}(50^2, 50^2, 1, 1) \; \forall k = 1, ..., K$, while $\mathbf{C}_{\boldsymbol{n}_0}$ is assumed to be negligible.

A comparison between the tests (29) and (35) derived in the previous sections is highlighted here, where we distinguish two specific case studies depending on the time location of multiple contacts. Specifically we assume to observe two AIS contacts (the last one before, and the first one just after the AIS device shutdown) and two radar contacts located along the path covered by the vessel, which under hypothesis H_1 consists of $N = 4$ sections characterized by the following sequence of long-run mean velocities $|\boldsymbol{v}_1| = 7.9$, $|\boldsymbol{v}_2| = 8.6$, $|\boldsymbol{v}_3| = 7.8$, $|\boldsymbol{v}_4| = 7.8$. The two radar contacts are denoted with k_1 and k_2 and the corresponding time locations with p_{k_1} and p_{k_2}. This means that the contacts are observed at $T_{1,2} = \frac{p_{k_{1,2}}}{N} T$.

Performance curves are provided in terms of missed detection probability, $1 - P_D$, versus false alarm probability, P_{FA},
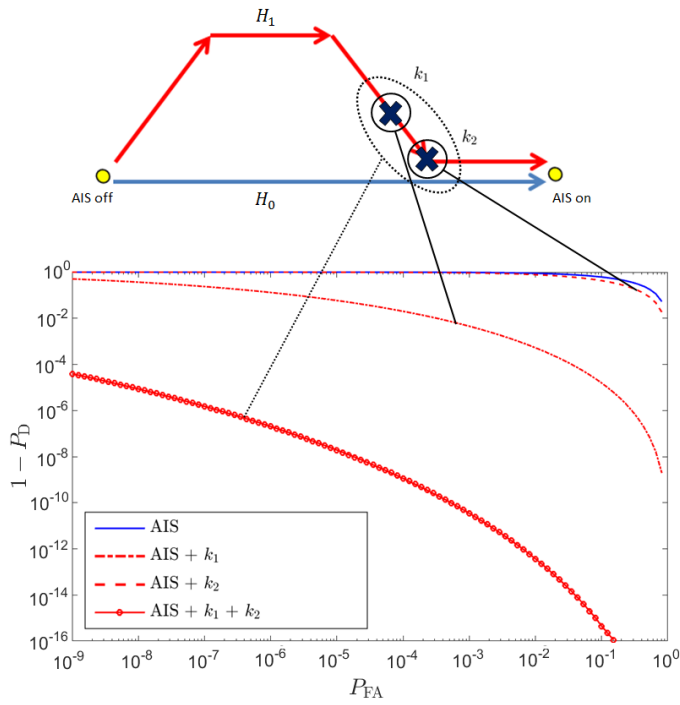
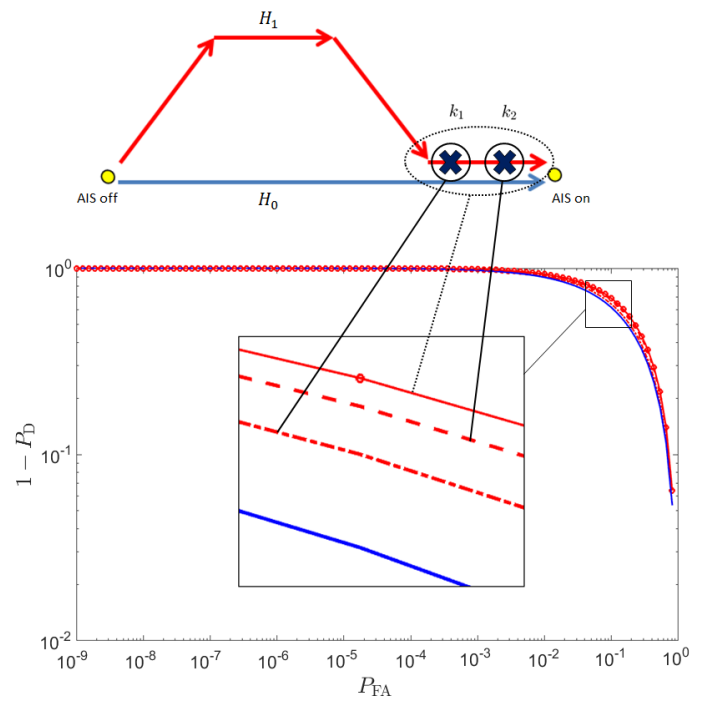Fig. 5. **Case study 1.(a)**: GLRT (29) performance for case study (a).

Fig. 6. **Case study 1.(b)**: GLRT (29) performance for case study (b).
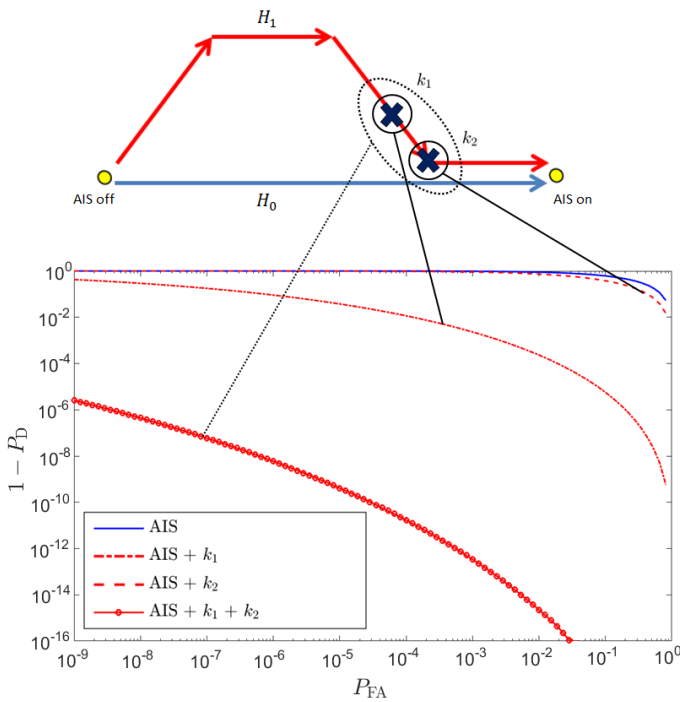

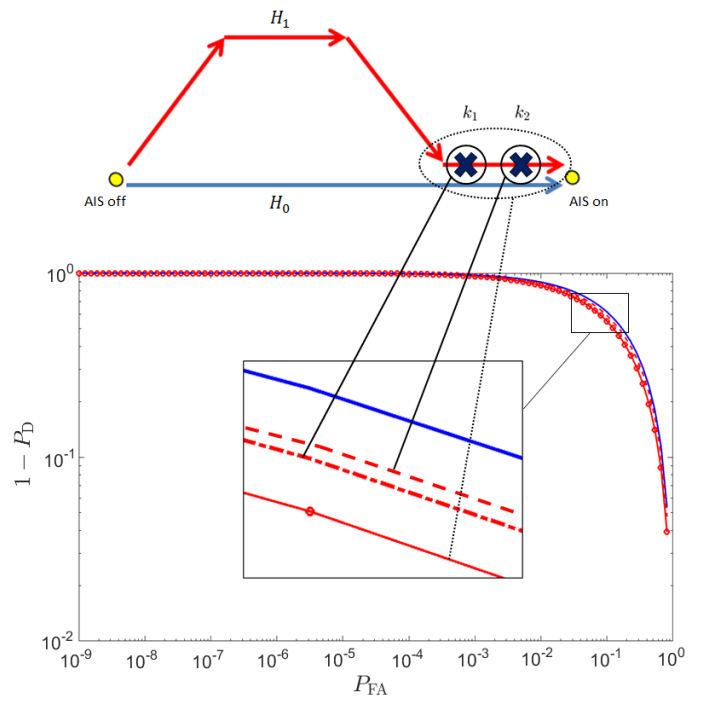Fig. 7. **Case study 2.(a)**: GLRT (35) performance for case study (a).

Fig. 8. **Case study 2.(b)**: GLRT (35) performance for case study (b).
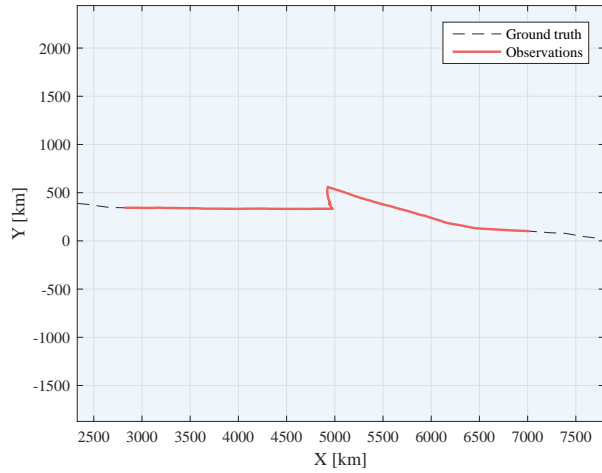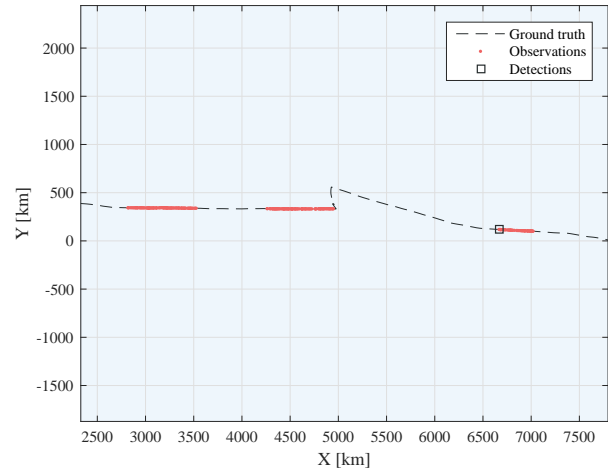
Fig. 9. Complete AIS track.

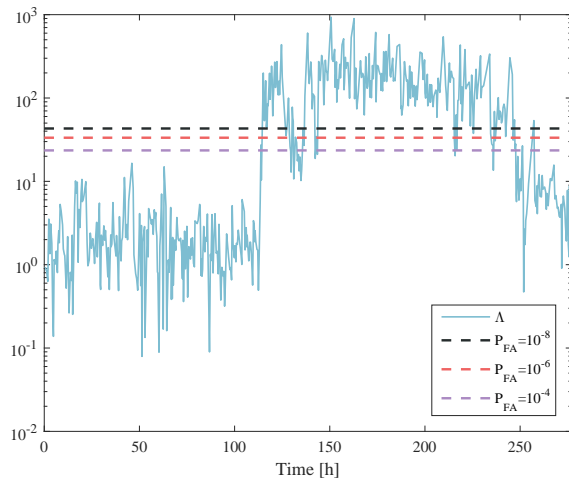Fig. 10. Simulated gaps in the AIS track.

Fig. 11. Test statistic (29) related to the complete AIS track.
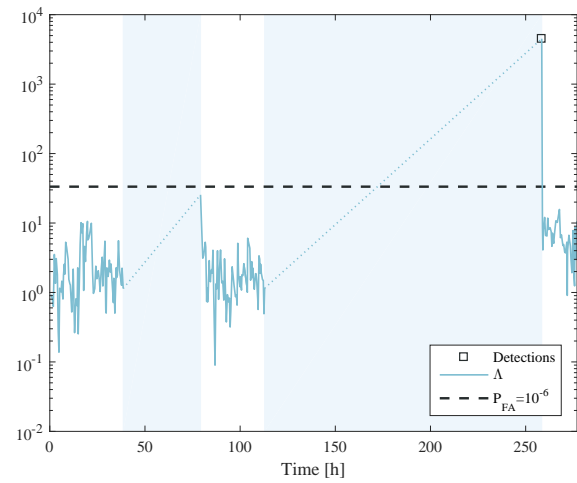

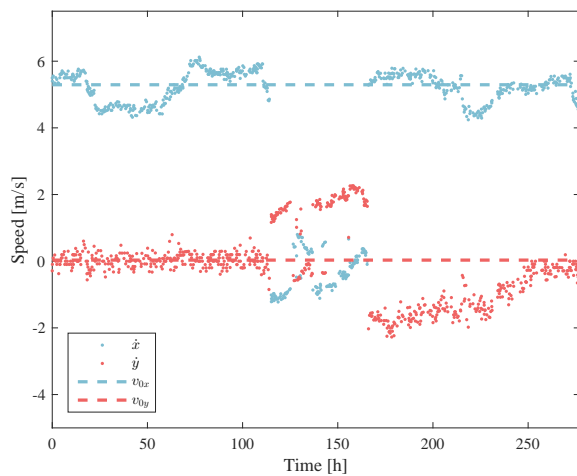Fig. 12. Test statistic (29) related to the incomplete AIS track.

Fig. 13. Velocity components related to the complete AIS track.
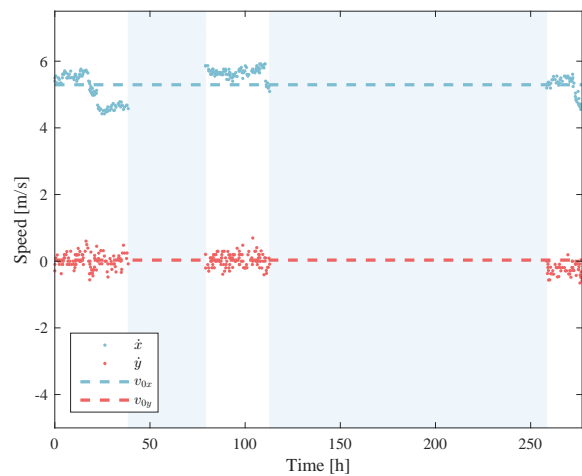

Fig. 14. Velocity components related to the incomplete AIS track.

TABLE I
NON-CENTRALITY PARAMETER AND DEGREES OF FREEDOM FOR RADAR
CONTACTS LOCATED IN $p_{k_1} = 2.7, p_{k_2} = 3$

		AIS	AIS + k_1	AIS + k_2	AIS + k_1 + k_2
	λ	3.3072	50.9995	7.4007	134.4314
d	1.(a)	4	8	8	12
	2.(a)	4	6	6	6

TABLE II
NON-CENTRALITY PARAMETER AND DEGREES OF FREEDOM FOR RADAR
CONTACTS LOCATED IN $p_{k_1} = 3.3, p_{k_2} = 3.7$

		AIS	AIS + k_1	AIS + k_2	AIS + k_1 + k_2
	λ	3.3072	4.1040	3.6086	4.1040
d	1.(b)	4	8	8	12
	2.(b)	4	4	4	4

(see Figs. 5-8). Specifically, for both case studies, while the solid line is related to the detection procedure performed by using the only AIS contribution, the dot-dashed and dashed lines represent the test performance achieved by using AIS data and the measurement provided by the radar contact k_1 and k_2, respectively. Furthermore performance related to the combined use of both radar contacts and AIS is denoted with an o-marker line.

Case study (a): Performance of GLRT (29) and (35) for the case study (a) is depicted in Fig. 5 and Fig. 7 respectively. The two radar contacts are located in $p_{k_1} = 2.7$ and in $p_{k_2} = 3$. It is easy to verify that, in both cases, the improvement obtained by using k_2 is not significant since it is located where the vessel is very close to the expected position as if the deviation never happened, while the use of k_1, located where the anomalous behavior and the nominal condition are significantly distant, provides a remarkable improvement. Finally the combined use of both radar contacts improves the performance with respect to the use of a single radar contact. We observe that in both cases performance improves with increasing number of radar contacts, however as discussed in Section V, given that λ is equal for both detectors, GLRT (35) exhibits better performance because d is smaller comparing to the case of GLRT (29), see also the discussion in Section V-C. In Table I are reported the values of λ and d for cases 1.(a) and 2.(a).

Case study (b): Performance of GLRT (29) and (35) for the case study (b) is depicted in Fig. 6 and Fig. 8 respectively, where the two radar contacts are now located at $p_{k_1} = 3.3$ and $p_{k_2} = 3.7$, that is, along a part of the path where the anomalous behavior is very close to the nominal condition, and therefore the two hypotheses are difficult to distinguish. In this case, compared to the case of AIS information only, the performance of the GLRT (29) gets worse with increasing number of radar contacts. This phenomenon is explained in Section V-C, and basically it is the case in which adding new contacts in location where H_1 is close to H_0 has the effect of adding mostly noise to the decision statistic. On the other hand, GLRT (35) shows increasingly better performance, even though the improvement is quite small because of the radar contacts contain limited information for discriminating the two hypotheses. In Table II are reported the values of λ and d for cases 1.(b) and 2.(b).

B. Analysis of real-world vessel traffic data

The strategy proposed in this work has been applied to the real-world AIS track depicted in Fig. 2.

Let us focus on the observation window confined to the area of concern where an actual anomalous behavior occurs.

The related AIS track, shown in Fig. 9, does indeed reveal a deviation from the normal route during a time frame of about 5 days, and Fig. 11 displays the test statistic (29) which exceeds the threshold (plotted for different values of the false alarm probability: $P_{\text{FA}} \in \{10^{-4}, 10^{-6}, 10^{-8}\}$) corresponding to the deviation from the nominal condition. In particular, the deviation from the nominal velocity is highlighted in Fig. 13 where the velocity components show an apparent change in that specific time frame.

The OU parameters are estimated in the path section immediately preceding the one where the deviation actually happens ($\boldsymbol{v}_0 = [5.29, \quad 0.03], \boldsymbol{\gamma} = [2.30 \times 10^{-4}, \quad 4.19 \times 10^{-3}]$, $\boldsymbol{\sigma} = [1.13 \times 10^{-2}, \quad 2.23 \times 10^{-2}]$), and the detection strategy is tested considering simulated gaps in AIS data, as shown in Fig. 10, with the corresponding velocity gaps shown in Fig. 14. The first gap occurs in a section of the trajectory where there is no deviation from the nominal conditions, while the second one occurs where the deviation actually happens. From the application of the detector (29) with $P_{\text{FA}} = 10^{-6}$, the deviation can be properly detected while no detection is correctly declared in the first gap, as shown in Fig. 12.

VII. CONCLUSION

In this paper the maritime anomaly detection problem has been studied assuming an OU mean-reverting stochastic motion model for the vessel dynamics. The aim was to reveal a possible deviation of the vessel under consideration from its nominal conditions, during an AIS device disablement, relying on a hypothesis testing procedure based on the generalized likelihood ratio decision statistic that builds on the changes in the OU process long-term velocity parameter.

A detailed description of the proposed detection strategy, built by exploiting multiple contacts has been provided, presenting both synthetic and real world analyses.

As confirmed by the numerical analysis, the joint use of radar and AIS information compared to the case of only AIS can lead to a remarkable improvement of detection performance, while it has been shown that under certain conditions detection performance not only does not improve, but actually deteriorates. Such conditions depend on the timing of radar contacts with respect to the differences between nominal and anomalous trajectories.

Moreover, a case of a real anomalous trajectory has been processed by exploiting the proposed detection strategy allowing to assess its performance.

For the sake of clarity the radar model does not consider false alarms and association error with other vessels, leading to a closed form expression for the detector and detection per-

$$\boldsymbol{\Omega}_2(\Delta_n) = \begin{bmatrix} f(\Delta_n\gamma_x) & h(\Delta_n,\boldsymbol{\gamma}) & k(\Delta_n\gamma_x) & \frac{g\left(\gamma_y\frac{\Delta_n}{2}\right)}{\gamma_y} - \frac{g\left((\gamma_x+\gamma_y)\frac{\Delta_n}{2}\right)}{\gamma_x+\gamma_y} \\ h(\Delta_n,\boldsymbol{\gamma}) & f(\Delta_n\gamma_y) & \frac{g\left(\gamma_x\frac{\Delta_n}{2}\right)}{\gamma_x} - \frac{g\left((\gamma_x+\gamma_y)\frac{\Delta_n}{2}\right)}{\gamma_x+\gamma_y} & k(\Delta_n\gamma_y) \\ k(\Delta_n\gamma_x) & \frac{g\left(\gamma_x\frac{\Delta_n}{2}\right)}{\gamma_x} - \frac{g\left((\gamma_x+\gamma_y)\frac{\Delta_n}{2}\right)}{\gamma_x+\gamma_y} & g(\Delta_n\gamma_x) & g\left((\gamma_x+\gamma_y)\frac{\Delta_n}{2}\right) \\ \frac{g\left(\gamma_y\frac{\Delta_n}{2}\right)}{\gamma_y} - \frac{g\left((\gamma_x+\gamma_y)\frac{\Delta_n}{2}\right)}{\gamma_x+\gamma_y} & k(\Delta_n\gamma_y) & g\left((\gamma_x+\gamma_y)\frac{\Delta_n}{2}\right) & g(\Delta_n\gamma_y) \end{bmatrix}$$
$$(41)$$

formance expressed as central chi-squared under the nominal condition and non-central chi-squared if anomaly occurs.

APPENDIX A
OU PROCESS

A. OU process covariance matrix

As given in [3], the OU process covariance matrix is

$$\mathbf{C}(\Delta_n) = \widetilde{\mathbf{R}}\boldsymbol{\Omega}\widetilde{\mathbf{R}}^{\mathrm{T}} \tag{42}$$

where $\boldsymbol{\Omega} = \boldsymbol{\Omega}_1 \circ \boldsymbol{\Omega}_2(\Delta_n)$, with the $\circ$ operator denoting the Hadamard product, and $\widetilde{\mathbf{R}} \triangleq [\mathbf{R}\ \mathbf{0}; \mathbf{0}\ \mathbf{R}]$. Matrix $\boldsymbol{\Omega}_1$ has the following form

$$\boldsymbol{\Omega}_1 = \begin{bmatrix} \frac{\sigma_x^2}{\gamma_x^3} & \frac{\sigma_{xy}}{\gamma_x\gamma_y} & \frac{\sigma_x^2}{2\gamma_x^2} & \frac{2\sigma_{xy}}{\gamma_x} \\ \frac{\sigma_{xy}}{\gamma_x\gamma_y} & \frac{\sigma_y^2}{\gamma_y^3} & \frac{2\sigma_{xy}}{\gamma_y} & \frac{\sigma_y^2}{2\gamma_y^2} \\ \frac{\sigma_x^2}{2\gamma_x^2} & \frac{2\sigma_{xy}}{\gamma_y} & \frac{\sigma_x^2}{\gamma_x} & \frac{2\sigma_{xy}}{\gamma_x+\gamma_y} \\ \frac{2\sigma_{xy}}{\gamma_x} & \frac{\sigma_y^2}{2\gamma_y^2} & \frac{2\sigma_{xy}}{\gamma_x+\gamma_y} & \frac{\sigma_y^2}{\gamma_y} \end{bmatrix}, \tag{43}$$

while $\boldsymbol{\Omega}_2(\Delta_n)$ is defined in (41) by using the following functions:

$$f(t) \triangleq \frac{1}{2}\left(2t + 4e^{-t} - e^{-2t} - 3\right),$$

$$g(t) \triangleq \frac{1}{2}\left(1 - e^{-2t}\right),$$

which are the prediction position and velocity error normalized variance, respectively, and

$$h(t,\boldsymbol{\gamma}) \triangleq t - \frac{1-e^{-t\gamma_x}}{\gamma_x} - \frac{1-e^{-t\gamma_y}}{\gamma_y} + \frac{1-e^{-t(\gamma_x+\gamma_y)}}{\gamma_x+\gamma_y},$$

$$k(t) \triangleq e^{-2t}\left(1 - e^t\right)^2.$$

B. Proof of $\boldsymbol{\omega}(T) \sim \mathcal{N}\left(\mathbf{0}, \mathbf{C}(T)\right)$

It can be easily shown returning the synthetic characterization of $\boldsymbol{\omega}(T)$ (17) as follows

$$E[\boldsymbol{\omega}(T)] = \sum_{n=1}^{N-1}\left[\prod_{i=n+1}^{N}\boldsymbol{\Phi}(\Delta_i)\right]E[\boldsymbol{\omega}(\Delta_n)] + E[\boldsymbol{\omega}(\Delta_N)] = 0.$$

$$Cov[\boldsymbol{\omega}(T)] = E\left[\boldsymbol{\omega}(T)\boldsymbol{\omega}(T)^{\mathrm{T}}\right]$$

$$\overset{i)}{=} \sum_{n=1}^{N-1}\left[\prod_{i=n+1}^{N}\boldsymbol{\Phi}(\Delta_i)\right]\mathbf{C}(\Delta_n)\left[\prod_{i=n+1}^{N}\boldsymbol{\Phi}(\Delta_i)^{\mathrm{T}}\right] + \mathbf{C}(\Delta_N)$$

$$\overset{ii)}{=} \mathbf{C}\left(\sum_{n=1}^{N}\Delta_n\right) = \mathbf{C}(T).$$

Since the $\boldsymbol{\omega}(\Delta_n)$ are independent zero-mean Gaussian random variables, the terms involving the mean of the mixed products, $E\left[\boldsymbol{\omega}(\Delta_n)\boldsymbol{\omega}(\Delta_m)\right]$, are zero, and equality $i)$ is valid.

The equality $ii)$ is proved by making the following considerations. For $N = 2$, by using the definitions of $\mathbf{C}(\Delta_n)$ in (42) and $\boldsymbol{\Phi}(\Delta_n)$ in (12), we can show that

$$\boldsymbol{\Phi}(\Delta_2)\mathbf{C}(\Delta_1)\boldsymbol{\Phi}(\Delta_2)^{\mathrm{T}} + \mathbf{C}(\Delta_2) = \mathbf{C}(\Delta_1 + \Delta_2).$$

It can be easily shown that for $N = 3$, exploiting the previous expression, we can get

$$\boldsymbol{\Phi}(\Delta_2 + \Delta_3)\mathbf{C}(\Delta_1)\boldsymbol{\Phi}(\Delta_2 + \Delta_3)^{\mathrm{T}} + \boldsymbol{\Phi}(\Delta_3)\mathbf{C}(\Delta_2)\boldsymbol{\Phi}(\Delta_3)^{\mathrm{T}} + \mathbf{C}(\Delta_3)$$
$$= \boldsymbol{\Phi}(\Delta_3)\left[\boldsymbol{\Phi}(\Delta_2)\mathbf{C}(\Delta_1)\boldsymbol{\Phi}(\Delta_2)^{\mathrm{T}} + \mathbf{C}(\Delta_2)\right]\boldsymbol{\Phi}(\Delta_3)^{\mathrm{T}} + \mathbf{C}(\Delta_3)$$
$$= \boldsymbol{\Phi}(\Delta_3)\mathbf{C}(\Delta_1 + \Delta_2)\boldsymbol{\Phi}(\Delta_3)^{\mathrm{T}} + \mathbf{C}(\Delta_3)$$
$$= \mathbf{C}(\Delta_1 + \Delta_2 + \Delta_3).$$

This procedure is valid for each N by induction.

APPENDIX B
MAXIMUM LIKELIHOOD ESTIMATION

Incorporating the term independent of the unknown parameter appearing in the GLRT (1) in the threshold τ, under the condition of scenario 1, in which N and $\mathcal{D}_N$ are unknown, the maximum likelihood estimate of $\boldsymbol{\mu} = \boldsymbol{\theta}$ is given by

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\left\{\ln\left[p_{\boldsymbol{\theta}}(\mathbf{y})\right]\right\}$$
$$= \arg\min_{\boldsymbol{\theta}}\left\{(\mathbf{y} - \boldsymbol{\theta})^{\mathrm{T}}\mathbf{C}_{\mathbf{y}}^{-1}(\mathbf{y} - \boldsymbol{\theta})\right\},$$

whose exact expression can be found by setting to zero the derivative with respect to $\boldsymbol{\theta}$:

$$\frac{\partial}{\partial\boldsymbol{\theta}}\left\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{C}_{\mathbf{y}}^{-1}\boldsymbol{\theta} - 2\boldsymbol{\theta}^{\mathrm{T}}\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{y}\right\}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = 0,$$

from which it follows that the ML estimation of the unknown parameter corresponds to data, $\widehat{\boldsymbol{\theta}} = \mathbf{y}$.

In the alternative case where N and $\mathcal{D}_N$ are known, the maximum likelihood estimation of the vector $\boldsymbol{\theta} = \boldsymbol{v}_{1:N}$ is given by

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}}\left\{(\mathbf{y} - \mathbf{H}\boldsymbol{\theta})^{\mathrm{T}}\mathbf{C}_{\mathbf{y}}^{-1}(\mathbf{y} - \mathbf{H}\boldsymbol{\theta})\right\},$$

whose exact expression can be found by setting to zero the derivative with respect to $\boldsymbol{\theta}$:

$$\frac{\partial}{\partial\boldsymbol{\theta}}\left\{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{H}\boldsymbol{\theta} - 2\boldsymbol{\theta}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{y}\right\}\bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = 0,$$

from which

$$\widehat{\boldsymbol{\theta}} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{y}.$$

In the same way it is shown that the ML estimate of the unknown parameter $\widetilde{\boldsymbol{\theta}}$, achieved by applying the rank reduction method, is given by

$$\widehat{\widetilde{\boldsymbol{\theta}}} = \left(\widetilde{\mathbf{H}}^{\mathrm{T}}\widetilde{\mathbf{H}}\right)^{-1}\widetilde{\mathbf{H}}^{\mathrm{T}}\widetilde{\mathbf{C}}\mathbf{y}.$$

APPENDIX C
OFF-DIAGONAL TERMS IN $\mathbf{C_y}$ CALCULATION

The off-diagonal terms in the covariance matrix $\mathbf{C_y}$ (26) can be derived as follows. Let us suppose for simplicity that p_i and p_j, denoting the time location of the i-th and the j-th contacts, respectively, are integer quantities. The result is still valid even if p_i and p_j are not integer.

$$
\begin{aligned}
\mathbf{C}_{ij} &= E\left[(\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_j - \boldsymbol{\mu}_j)^{\mathrm{T}}\right]\\
&= E\left[(\boldsymbol{\omega}(T_i) + \mathbf{n}_i + \boldsymbol{\Phi}(T_i)\mathbf{n}_0)(\boldsymbol{\omega}(T_j) + \mathbf{n}_j + \boldsymbol{\Phi}(T_j)\mathbf{n}_0)^{\mathrm{T}}\right]\\
&= E\left[\boldsymbol{\omega}(T_i)\boldsymbol{\omega}(T_j)^{\mathrm{T}}\right] + \boldsymbol{\Phi}(T_i)\mathbf{C}_{\mathbf{n}_0}\boldsymbol{\Phi}(T_j)^{\mathrm{T}}\\
&\overset{i)}{=} \sum_{n=1}^{p_i}\left[\prod_{l=n+1}^{p_i}\boldsymbol{\Phi}(\Delta_l)\right]\mathbf{C}(\Delta_n)\left[\prod_{l'=n+1}^{p_j}\boldsymbol{\Phi}(\Delta_{l'})^{\mathrm{T}}\right]\\
&\quad + \boldsymbol{\Phi}(T_i)\mathbf{C}_{\mathbf{n}_0}\boldsymbol{\Phi}(T_j)^{\mathrm{T}}\\
&= \mathbf{C}(T_i)\boldsymbol{\Phi}(T_j - T_i)^{\mathrm{T}} + \boldsymbol{\Phi}(T_i)\mathbf{C}_{\mathbf{n}_0}\boldsymbol{\Phi}(T_j)^{\mathrm{T}},
\end{aligned}
$$

$\forall i,j = 1,...,K$. In $i)$ it has been considered that the random variables $\boldsymbol{\omega}(\Delta_n)$ involved in the expressions of $\boldsymbol{\omega}(T_i)$ and $\boldsymbol{\omega}(T_j)$ given in (23), are all independent so that $E[\boldsymbol{\omega}(\Delta_n)\boldsymbol{\omega}(\Delta_m)^{\mathrm{T}}] = 0$ if $n \neq m$. In the same way it is shown that $\mathbf{C}_{ji} = \mathbf{C}_{ij}^{\mathrm{T}}$, $\forall i,j = 1,2,\ldots,K$ with $T_i < T_j$.

ACKNOWLEDGMENT

REFERENCES

[1] "International Convention of Safety of Life at Sea (SOLAS)," International Maritime Organization (IMO).

[2] *Technical characteristics for an automatic identification system using time-division multiple access in the VHF maritime mobile band*, ITU Recommendation M.1371, Rev. 5, Feb. 2014.

[3] L. M. Millefiori, P. Braca, K. Bryan, and P. Willett, "Modeling vessel kinematics using a stochastic mean-reverting process for long-term prediction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 5, pp. 2313–2330, October 2016.

[4] S. Maresca, P. Braca, J. Horstmann, and R. Grasso, "Maritime surveillance using multiple high-frequency surface-wave radars." *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 5056–5071, August 2014.

[5] P. Braca, S. Maresca, R. Grasso, K. Bryan, and J. Horstmann, "Maritime surveillance with multiple over-the-horizon HFSW radars: An overview of recent experimentation." *IEEE Aerospace and Electronic Systems Magazine*, vol. 30, no. 12, pp. 4 – 18, December 2015.

[6] K. Granström, A. Natale, P. Braca, G. Ludeno, and F. Serafino, "Gamma gaussian inverse wishart probability hypothesis density for extended target tracking using X-band marine radar data." *IEEE Geoscience and Remote Sensing Society*, vol. 53, no. 12, pp. 6617 – 6631, July 2015.

[7] G. Papa, P. Braca, S. Horn, S. Marano, V. Matta, and P. Willett, "Multisensor adaptive bayesian tracking under time-varying target detection probability." *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 5, pp. 2193 – 2209, October 2016.

[8] F. Meyer, T. Kropfreiter, J. Williams, R. Lau, F. Hlawatsch, P. Braca, and M. Z. Win, "Message passing algorithms for scalable multitarget tracking." *Proceedings of IEEE.*, vol. 106, no. 2, pp. 221 – 259, February 2018.

[9] F. Amato, M. Fiorini, S. Gallone, and G. Golino, "Fully solid state radar for vessel traffic services," in *11-th International Radar Sympsium*, June 2010, pp. 1–5.

[10] G. Vivone, L. M. Millefiori, P. Braca, and P. Willett, "Performance assessment of vessel dynamic models for long-term prediction using heterogeneous data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 11, pp. 6533–6546, November 2017.

[11] S. Giompapa, F. Gini, A. Farina, A. Graziano, R. Croci, and R. Distefano, "Maritime border control multisensor system," *IEEE Aerospace and Electronic Systems Magazine*, vol. 24, no. 8, pp. 9–15, Aug 2009.

[12] M. Jakob, O. Vaňek, S. Urban, P. Benda, and M. Pěchouček, "Adversarial modeling and reasoning in the maritime domain," Agent Technology Center, Department of Cybernetics, FEE Czech Technical University in Prague, Tech. Rep., December 2009.

[13] R. O. Lane, D. A. Nevell, S. D. Hayward, and T. W. Beaney, "Maritime anomaly detection and threat assessment," in *2010 13th International Conference on Information Fusion*, July 2010, pp. 1–8.

[14] B. Ristic, B. L. Scala, M. Morelande, and N. Gordon, "Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction," in *2008 11th International Conference on Information Fusion*, June 2008, pp. 1–7.

[15] M. Guerriero, S. Coraluppi, and C. Carthel, "Analysis of AIS intermittency and vessel characterization using a Hidden Markov Model," NURC, Tech. Rep. NURC-FR-2010-002, NATO UNCLASSIFIED, January 2010.

[16] K. Kowalska and L. Peel, "Maritime anomaly detection using Gaussian process active learning," in *2012 15th International Conference on Information Fusion*, July 2012, pp. 1164–1171.

[17] M. Vespe, I. Visentini, K. Bryan, and P. Braca, "Unsupervised learning of maritime traffic patterns for anomaly detection," in *9th IET Data Fusion Target Tracking Conference (DF TT 2012): Algorithms Applications*, May 2012, pp. 1–5.

[18] F. Katsilieris, P. Braca, and S. Coraluppi, "Detection of malicious AIS position spoofing by exploiting radar information," in *Proceedings of the 16th International Conference on Information Fusion*, July 2013, pp. 1196–1203.

[19] G. Pallotta, M. Vespe, and K. Bryan, "Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction," *Entropy*, vol. 15, no. 6, 2013.

[20] J. van Laere and M. Nilsson, "Evaluation of a workshop to capture knowledge from subject matter experts in maritime surveillance," in *2009 12th International Conference on Information Fusion*, July 2009, pp. 171–178.

[21] J. J. Alava, M. J. Barragán-Paladines, J. Denkinger, L. Muñoz-Abril, P. Jiménez, F. Paladines, and et al., "Massive Chinese fleet jeopardizes threatened shark species around the Galápagos marine reserve and waters off Ecuador: Implications for national and international fisheries policy," *International Journal of Fisheries Sci Res.*, 2017.

[22] L. M. Millefiori, P. Braca, and P. Willett, "Consistent estimation of randomly sampled Ornstein-Uhlenbeck process long-run mean for long-term target state prediction." *IEEE Signal Processing Letters*, vol. 23, no. 11, pp. 1562 – 1566, November 2016.

[23] P. Coscia, P. Braca, L. M. Millefiori, F. Palmieri, and P. Willett, "Maritime traffic representation based on sea-lanes graph construction criteria using multiple Ornstein-Uhlenbeck processes," *IEEE Transactions on Aerospace and Electronic Systems*, to be published, 2018.

[24] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.

[25] B. Porat and B. Friedlander, "Performance analysis of a class of transient detection algorithms-a unified framework," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2536–2546, Oct 1992.

[26] G. E. Uhlenbeck and L. S. Ornstein, "On the theory of the brownian motion," *Phys. Rev.*, vol. 36, pp. 823–841, Sep 1930.

[27] S. Coraluppi and C. Carthel, "Stability and stationarity in target kinematic modeling," in *2012 IEEE Aerospace Conference*, March 2012, pp. 1–8.

[28] S. Coraluppi, C. Carthel, P. Braca, and L. Millefiori, "The mixed Ornstein-Uhlenbeck process and context exploitation in multi-target

tracking," in *2016 19th International Conference on Information Fusion (FUSION)*, July 2016, pp. 217–224.

[29] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. Part I. Dynamic models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1333–1364, Oct 2003.

[30] A. Thorpe and L. Scharf, "Data adaptive rank-shaping methods for solving least squares problems," *IEEE Transactions on Signal Processing*, vol. 43, no. 7, pp. 1591–1601, July 1995.

[31] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. The Johns Hopkins University Press, 1996.

[32] E. Lehmann, *Elements of Large-Sample Theory*. New York: Springer Science & Business Media, 2004.

Enrica d'Afflisio received the M.Sc. degree (*summa cum laude*) in telecommunication engineering from the University of Naples Federico II, Naples, Italy, in 2016.

Since 2016, she has been a Visiting Researcher at the NATO Science and Technology Organization Center for Maritime Research and Experimentation, La Spezia, where she joined the Research Department as a Scientific Assistant in 2018. Her research interests include statistical signal processing, target tracking and data fusion.
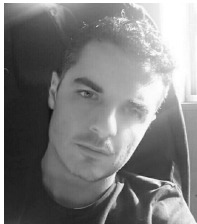
Paolo Braca (M'14–SM'17) received the Laurea degree (*summa cum laude*) in electronic engineering and the Ph.D. degree (highest rank) in information engineering from the University of Salerno, Salerno, Italy, in 2006 and 2010, respectively.

In 2009, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT, USA. From 2010 to 2011, he was a Post-Doctoral Associate with the University of Salerno. In 2011, he joined the NATO Science and Technology Organization Center for Maritime Research and Experimentation, La Spezia, as a Scientist with the Research Department. He has co-authored of over 100 publications in international scientific journals and conference proceedings. His research interests include statistical signal processing with an emphasis on detection and estimation theory, wireless sensor network, multiagent algorithms, target tracking and data fusion, adaptation and learning over graphs, and distributed radar (sonar) processing.

Dr. Braca was awarded the National Scientific Qualification to function as an Associate and Full Professor in Italian universities by a qualification committee of professors respectively in 2017 and 2018. He was a recipient of the Best Student Paper Award (first runner-up) at FUSION Conference in 2009. He received the NATO STO Scientific Achievement Award 2017 from the NATO Chief Scientist. He serves as an Associate Editor of the IEEE T-SP, the IEEE T-AES, the ISIF JAIF, and the EURASIP JASP. In 2017, he was the Lead Guest Editor of the Special Issue on "Sonar Multi-Sensor Applications and Techniques" in IET RSN. He served as an Associate Editor of the IEEE SPM (E-Newsletter) from 2014 to 2016.

Leonardo M. Millefiori (S'12–M'14) received the B.Sc. degree in aerospace information engineering and the M.Sc. degree (*summa cum laude*) in communication engineering with a focus on radar systems and remote sensing from the Sapienza University of Rome, Rome, Italy, in 2010 and 2013, respectively.

In 2013, he was a Visiting Researcher at the NATO Science and Technology Organization Center for Maritime Research and Experimentation, La Spezia, where he joined the Research Department under the Maritime Security Program as a Scientist in 2014. His research interests include target motion modeling, statistical signal processing, target tracking, and data fusion and radar systems.

Peter Willett (F'03) received the B.A.Sc. degree in engineering science from the University of Toronto, Toronto, ON, Canada, in 1982, and the Ph.D. degree from Princeton University, Princeton, NJ, USA, in 1986.

Since 1986, he has been a Faculty Member at the University of Connecticut, Storrs, CT, USA, where he has been a Professor since 1998. He is Chief Editor of the IEEE AES MAGAZINE, and was Editor-in-Chief of the IEEE SP LETTERS (2014-2016) and the IEEE TRANSACTIONS ON AES (2006-2011). He has held numerous positions in IEEE SPS snd AESS. He was the General Co-Chair of the IEEE/ISIF Fusion Conference in Florence in 2006, in Cologne in 2008, and in Chicago in 2011, and the IEEE SPS SAM in Sheffield in 2018.