Maritime Traffic Networks: From Historical Positioning Data to Unsupervised Maritime Traffic Monitoring

Virginia Fernandez Arguedas, Giuliana Pallotta, and Michele Vespe

***Abstract*— The large maritime traffic volume and its implications in economy, environment, safety, and security require an unsupervised system to monitor maritime traffic. In this paper, a method is proposed to automatically produce synthetic maritime traffic representations from historical self-reporting positioning data, more specifically from automatic identification system data. The method builds a two-layer network that represents the maritime traffic in the monitored area, where the external layer presents the network's basic structure and the inner layer provides precision and granularity to the representation. The method is tested in a specific scenario with high traffic density, the Baltic Sea. Experimental results reveal a decrease of over 99% storage data with a negligible precision drop. Finally, the novel method presents a light and structured representation of the maritime traffic, which sets the foundations to real-time automatic maritime traffic monitoring, anomaly detection, and situation prediction.**

***Index Terms*— Maritime traffic representation, maritime surveillance, anomaly detection, traffic monitoring, AIS.**

I. Introduction

OCEANS cover a 71 % of the Earth surface, their great extension hinders their monitoring and surveillance, implying the supervision of coastal and open sea areas. Such enormous playground is also scenario of numerous activities which have a high impact on security, safety, economy and environment. Maritime surveillance represents a challenging research field due to the broadness of the coverage area and the variety of monitoring activities, e.g. irregular migration, piracy, fisheries control or traffic monitoring.

The main target in maritime surveillance is to enable the automatic monitoring, analysis and understanding of activities at sea, typically named Martime Situational Awareness (MSA). In 2009, European Commission defined Maritime Situational Awareness as "the effective understanding of activity associated with the maritime domain that could impact the security, safety, economy, or environment of the European Union and its Member States." [3]. For the maritime surveillance enhancement, diverse technologies have been deployed, such as vessel positioning systems or meteorological and oceanographic sensors. As a result, a large amount of heterogeneous data sources feed the maritime surveillance systems. The fusion of such data sources enables the creation of a real-time picture of the situation at sea and enlarges the available knowledge.

Working towards MSA, a maritime knowledge acquisition, discovery, representation and understanding system is required to provide a universal framework capable to (1) acquire, aggregate and represent heterogeneous data; (2) analyse data either from individual sensors or from combined and heterogeneous sources of information; (3) contextualise results, offering a complete view of the situation at sea, and so, offering new possibilities to tackle challenging scenarios such as irregular migration, piracy, smuggling or illegal fishing. The aim is to automatise the data acquisition, aggregation and knowledge discovery process, offering the users (e.g. operational authorities) the possibility to monitor large and remote marine areas with minimum or null human supervision.

In this paper, we propose to analyse, model and represent big maritime traffic data to enhance the MSA, to generate an efficient representation, capable to sustain real-time maritime traffic monitoring and early anomaly detection at sea. We propose a method that analyses historical self-reporting positioning data to create a maritime traffic representation based on networks. The proposed method is built up from the system presented in [14]. Two novelties are presented. First, the distinction and separation between *semantic routes* and *routes*, separating maritime traffic sharing the same origin and destination but presenting different behaviours in order to improve the representation precision. Second, a new method to detect vessel behavioural changes is proposed. The new method, based on Douglas and Peucker [10], simplifies the maritime traffic network, increasing the representation precision and the data storage compression. Both novelties tackle a unique challenge, to automatically generate data-driven representation of real navigation routes. The data-driven algorithm ensures the adaptation to the navigation peculiarities of the scenario, marine characteristics, bathimetry of the area and areas with special navigation requirements (i.e. marine protected areas). The consideration and adaptation to real navigation characteristics will enable the creation of a realistic and adaptive maritime traffic network capable to support and assist real-

Manuscript received June 23, 2016; revised January 16, 2017 and March 21, 2017; accepted April 14, 2017. Date of publication May 25, 2017; date of current version March 7, 2018. The Associate Editor for this paper was N. Geroliminis. *(Corresponding author: Virginia Fernandez Arguedas.)*

V. Fernandez Arguedas was with European Commission, Joint Research Centre, 21027 Ispra, Italy. She is now with the Deimos–Elecnor Group, 8760 Madrid, Spain (e-mail: virginiafernandeza@gmail.com).

G. Pallotta was with the NATO Centre for Maritime Research and Experimentation, 19126 La Spezia, Italy. She is now with the Lawrence Livermore National Laboratory, Livermore, CA 94550 USA (e-mail: pallotta2@llnl.gov).

M. Vespe is with European Commission, Joint Research Centre, 21027 Ispra, Italy (e-mail: michele.vespe@ec.europa.eu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2017.2699635

time monitoring of the activity at sea, early anomaly detection and navigation aid in difficult meteorological situations.

The remainder of the paper is organised as follows. Section II presents the scenario under analysis. Section III reviews the literature on maritime traffic representation. In Section IV, the proposed Maritime Traffic Knowledge Discovery and Representation System is detailed. Sections V to VII will introduce each stage of the process and the intermediate results. Section VIII presents and analyses the experimental results. In Section IX, the paper ends drawing some conclusions, remaining challenges and future lines of exploitation.

II. Maritime Traffic Analysis in the Baltic Sea

The Baltic Sea is one of the most densely trafficked sea regions in the world [25]. Its total sea area covers $404.354 km^2$ and it is confined by nine countries populated by around 90 million people. Since the mid-90s, maritime transportation has suffered an enormous increment [23] and a forecast predicts a huge growth in the incoming years [20]. Such increment is mainly related to three factors: (i) the expansion and construction of oil terminals on the shores of the Gulf of Finland, (ii) the regional economic growth [25] and (iii) the increment of cruise ships in the area [5]. The increasing ship-presence enlarges the risk of accidents, threatening the Baltic Sea ecosystem and its population safety, security and economy.

In the attempt to increase the security, maritime traffic in the Baltic must be monitored. Currently, there is a great variety of sensors that enable maritime traffic monitoring [4]. Amongst them, self-reporting positioning sensors enable constant access and update to vessel positioning data. The main ship self-reporting systems currently used are Automatic Identification System (AIS), Long Range Identification and Tracking (LRIT), and Vessel Monitoring System (VMS). These systems were designed for different purposes and, hence, have different technical specifications, related to the coverage, latency or the controlling agency. The technical specifications of such systems are regulated at international level and the relevant tracking capabilities are characterised by refresh rate, spatial coverage, data latency between transmission and availability of the information, quantity and types of compliant vessels [4]. Among the existing self-reporting positioning systems, AIS demonstrated to be a reliable source of information for maritime traffic monitoring purposes in terms of spatial coverage, vessels coverage and message transmission frequency. Two types of AIS systems are currently deployed, terrestrial- and satellite-AIS. Their main differences reside in their spatial coverage and probability of detection. Terrestrial-AIS offers good tracking capabilities in the sensors' line of sight areas. However, satellite-AIS offers virtually global spatial coverage to detriment of the probability of detection in busy areas due to message collision. Moreover, satellite-AIS depends on the number of satellites in the constellation and data latency is affected on the visibility of ground station. Finally, terrestrial and satellite-AIS complementarity improves the spatial coverage, despite arising issues related to data persistence.

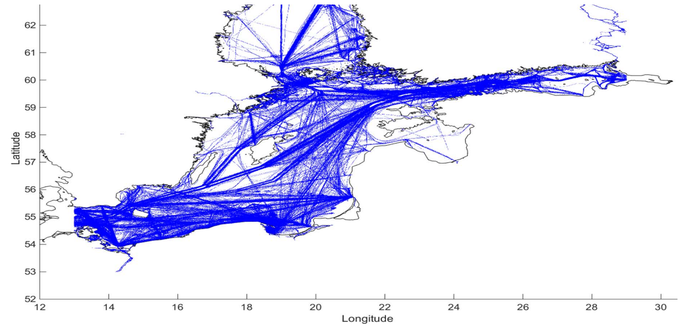

Fig. 1. Two months of AIS messages within the monitored area.

The complexity of the maritime traffic in the Baltic Sea, with more than 200 commercial ports and managing more than 800 million tonnes of cargo and 91 million passengers [25], represents a complex and challenging scenario for the analysis of maritime traffic. In this paper, the area under analysis covers latitude from 53 to 63 degrees and longitude from 13 to 29 degrees. The study includes the largest shipping lanes connecting cities such as Stockholm or Tallinn; it reveals the impact of gas and oil extraction and shipping in the Baltic Sea or it locates the marine protected areas via data analysis. The processed historical self-reporting data comprises two months of terrestrial-AIS messages, from 15/06/2014 to 15/08/2014. Figure 1 depicts all AIS messages in the Baltic Sea for time of analysis. Such data constitutes the input to our analysis.

III. Literature Review

In the literature, several methods have tackled maritime traffic representation from self-reporting data. Such techniques can be classified into two main categories, named *spatial-grid techniques* and *spatio-temporal techniques*. On the one hand, spatial grid techniques typically partition the monitored area in a spatial grid whose cells are characterised by the motion properties of the crossing vessels [6], [17]. Spatial grid techniques are a fast approach to detect and visualise low-level maritime intelligence layers. However, such techniques present two main drawbacks. First, they suffer from computational burden when increasing scale of the monitored area, hindering maritime traffic analysis on global scale. Second, the aggregation and manipulation makes the detection of structured anomalies (e.g. track birth/deaths or start/stop sailing) difficult to be implemented. On the other hand, several techniques have been proposed to detect and model motion patterns and vessel behaviours inherent in the maritime traffic raw data based on *spatio-temporal data mining* techniques. Su and Chang [24] presented a method to monitor fishing vessels, clustering vessels based on the spatial information provided in the VMS data. An extension of this work was presented in [16]. Chang [8] investigated methods to extract vessel traffic patterns that could support high-risk spot identification analysing coastal AIS data. Their proposed method extracted routes using density-based clustering and generalised tracks detecting turns using angle thresholding. Later, a method based on an adapted version of the Shared Nearest Neighbour algorithm was presented to cluster vessels moving towards the

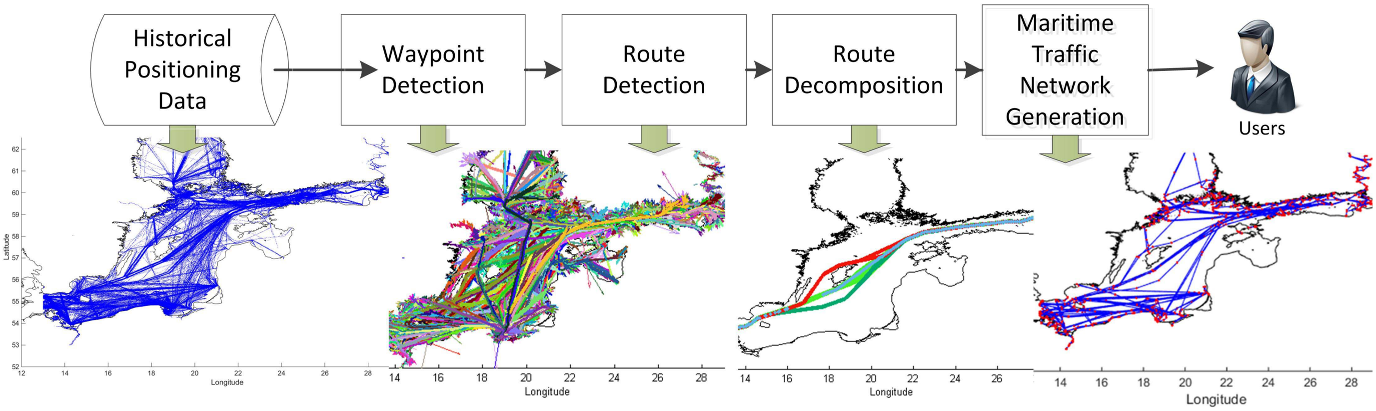

Fig. 2. Maritime Traffic Knowledge Discovery and Representation System.

same direction. The clustering process analysed the vessels position and direction to identify maritime traffic routes [22]. In [21], a vectorial approach was presented, where trajectories were envisaged as a set of lines connecting nodes. The authors proposed a method to automatically detect trajectories and cluster them together into routes based on their spatio-temporal information. In [14], this method was used to detect routes and waypoints as initial structures of the maritime traffic, then, the authors proposed a method to synthesise and combine the routes to build networks capable to represent maritime traffic. The preliminary results based on the analysis of the maritime traffic in the Dover Strait revealed a precise representation based on a light structure. Chen *et al.* [9] proposed a method to extract principal fairways in western Taiwan Strait based on the route extraction method described in [21]. The main advantage of such approaches resides on the capacity to perform global maritime traffic analysis, detect route-based anomalies and perform vessel route prediction. However, these techniques were not conceived to detect anomalies related to vessels interactions, which still are an open challenge. Finally, IALA [1] created the "IALA Waterway Risk Assessment Program" (IWRAP) [2] whose main objective was to provide users with tools to quantify the risks of a ship transiting in specific geographical areas. IWRAP provides a supervised approach to assists analysts on the risks in terms of groundings and collisions from manually selected routes, where AIS data is analysed and information is provided to analysts that evaluate the risks.

In this paper, we present an unsupervised approach to generate maritime traffic representation from historical self-reporting positioning data based on spatio-temporal data analysis techniques. The proposed representation, based on networks, provides a light, structured and precise representation. The interest in network analysis has had an interdisciplinary impact on diverse fields and the maritime domain is one of the most recent application areas. In the book [11], the concept of maritime networks and the connectivity of maritime nodes like ports and terminals is discussed. To our best knowledge, this is the most recent example in the literature where the analysis of maritime networks is extensively analysed. They represent the traffic flows between a predefined set of nodes in a graph-based fashion and the networks are derived as a topographical map of the shipping connections. In [7], the analysis of shipping networks is analysed via a network representation, where the main indicators of network connectivity are derived. The strategy presented in the present paper first derives the points of interest using a learning approach; then it estimates the average intermediate paths between them by using the actual in-between vessel tracks which connect the points. The advantage is the reduction of the storage requirements for the maritime routes, but without loosing the real geographical mapping of the connections between the nodes.

IV. Maritime Traffic Knowledge Discovery and Representation System

The proposed Maritime Traffic Knowledge Discovery and Representation System aims at the analysis of vessels historical self-reporting positioning data to extract and model navigation patterns in order to synthetically represent maritime traffic, offering a simple but precise representation. The proposed method detects vessel behavioural changes as the first step towards discovering the de facto routes followed by vessels. The approach is fully automatic and data-driven, building models over real navigation data.

The proposed approach bases its analysis in AIS self-reporting positioning data. The fusion of AIS data provides a real-time MSP, revealing vessels location at a current time. However, historical AIS positioning data requires a sophisticated knowledge discovery process based on pattern recognition and spatio-temporal data mining to reveal the underlying patterns of navigation in the Baltic Sea. Our goal is to automatically derive a maritime traffic network from real maritime data which synthetically represents the navigational patterns in the Baltic Sea. Three are the future application areas of our maritime traffic network. First, to offer operational authorities the means to automatically monitor maritime traffic at the Baltic Sea and to support early anomaly detection capabilities. Second, to offer ships up-to-date knowledge of navigational routes and current traffic, minimising the risk of collisions and groundings (one of the main risks in the Baltic). Third, to offer policy makers fundamental tools for the ex-post evaluation of policies impact on maritime traffic.

Figure 2 presents the Maritime Traffic Knowledge Discovery and Representation System. Our system consists of a sequential process of four stages. First, the waypoints (entries/

TABLE I
PARAMETERS ADOPTED IN THE ANALYSED SCENARIO

Parameter	Adopted Value
Search radius for Port detection	0.02 nm
Search radius for Entry/Exit gate detection	0.02 nm
Min Number of Neighbours for Port detection	3
Min Number of Neighbours for Entry/Exit gate detection	5

exits and ports) within the area of interest must be detected, revealing the first nodes of our maritime traffic network. Second, semantic route detection is proposed to cluster all vessels sharing same origin and destination. Third, the detected semantic routes are decomposed. Vessels travelling from the same origin and to the same destination often follow very diverse itineraries to perform the same percourse/semantic route. Different itineraries are separated into different routes to add precision to the maritime traffic modelling. Finally, the maritime traffic network is created based on the analysis of waypoints and routes, by extrapolating behaviour patterns inherent to the spatio-temporal information.

The Baltic Sea presents a great AIS coverage feeding maritime monitoring systems with large amount of data. The data is typically stored in databases, requiring of a pre-processing stage to eliminate redundant information. The pre-processing eliminates duplicated messages, appearing due to the existence of multiple AIS receptors storing the vessel's reporting messages; and it discards messages from the same vessel that are too close both in time and space to each other. The advantages of the pre-processing are the elimination of duplicated messages arriving with different delays due to their platform and the reduction of the data to process.

In the following Sections, each stage of the Maritime Traffic Knowledge Discovery and Representation System is detailed. The pseudocode of the system is presented in Algorithm 1.

V. WAYPOINTS DETECTION AND ROUTE EXTRACTION

Considering our area of interest, defined in Section II, the first stage to create the maritime traffic network is to determine the waypoints, including Entry, Exit and Port areas. Entry and Exit waypoints define the gates to the monitored area, so they depend on its geographical definition. On the other hand, Ports identify local ports, offshore platforms and stationary areas, hence, they are reference points, invariant with respect to the monitored area. The waypoint detection method is based on TREAD [21]. In this method, entry/exit gates are created and dynamically updated, when a vessel enters/leaves the area of interest, generating "birth"/"death" events (corresponding to the transition of the vessel status from "transmitting" to "lost" and vice versa). In contrast, Ports are detected by speed gating.

Maritime traffic changes over time implying the need to create evolutionary waypoints. Thus, Ports and Entry/Exit gates are created, expanded and merged progressively using an incremental Density-Based Spatial Clustering procedure (refer to [13] and [12]), where the clustering parameters are set based on specific traffic density, intensity and regularity in the area of interest (refer to Table I). Density-based clustering methods

Algorithm 1 Maritime Traffic Knowledge Discovery and Representation System

Require: AIS messages containing static and dynamic information
  $[waypoints] \Leftarrow Waypoint\_Detection(AIS\_messages)$
  $[semantic\_routes] \Leftarrow Route\_Detection(AIS\_messages)$
  for $semantic\_route \in [semantic\_routes]$ **do**
    $[routes] \Leftarrow Route\_Decomposition(semantic\_route)$
  end for
  for $route \in [routes]$ **do**
    To compute the distance between the two waypoints defining the route
    $distanceWP = Distance(waypoint_1, waypoint_2)$
    Travelled_distance is the distance percoursed by the route and indicated in its AIS data. If it is outside those limits, it is an anomaly, i.e. uncomplete routes or detours
    $condition1 \Leftarrow travelled\_distance > distanceWP$
    $condition2 \Leftarrow travelled\_distance < 3 * distanceWP$
    if condition1 & condition2 **then**
      $[\hat{R}] \Leftarrow Synthetic\_Route\_Computation(route)$
      $[\bar{R}] \Leftarrow Route\_Temporally\_Unfolding(\hat{R})$
      $[breakpoints] \Leftarrow Breakpoint\_Detection(\bar{R})$
      $segment \Leftarrow [breakpoints]$
      $[segments] \Leftarrow segment$
      A maritime lane is built by a concatenation of segments
      $maritime\_lane = [segments]$
      $[maritime\_lanes] \Leftarrow maritime\_lane$
    end if
  end for
  Start of the External Layer algorithm
  for $maritime\_lane \in [maritime\_lanes]$ **do**
    $[all\_segments] \Leftarrow maritime\_lane\{segments\}$
  end for
  $MTN \Leftarrow Initialisation([all\_segments])$
  Comparison between each segment in $[all\_segments]$ and all the segments stored in the Maritime Traffic Network
  for $segment \in [all\_segments]$ **do**
    $ans = Compare(segment, MTN\{segments\})$
    if ans == TRUE **then**
      $new\_seg = Merge(segment, MTN\{segment_i\})$
      $[MTN] \Leftarrow new\_seg$
      $[updated\_segments] \Leftarrow Update\_Route(new\_seg)$
    end if
  end for
  $MTN\_optimised \Leftarrow Initialisation([updated\_segments])$
  for $segment \in [updated\_segments]$ **do**
    $ans = Compare(segment, MTN\{segments\})$
    if ans == TRUE **then**
      $new\_seg = Merge(segment, MTN\{segment_i\})$
      $[MTN\_optimised] \Leftarrow new\_seg$
      $[optimised\_segments] \Leftarrow Update\_Route(new\_seg)$
    end if
  end for

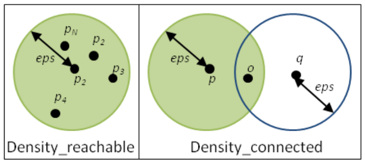

Fig. 3. Points belong to a cluster when they are *density-reachable* or *density-connected*. N points are *density-reachable*, when the number of points in a neighbourhood exceeds a minimum threshold. Two points p and q are *density-connected* when both are *density-reachable* from a third point o. The parameters used are described in Table I.

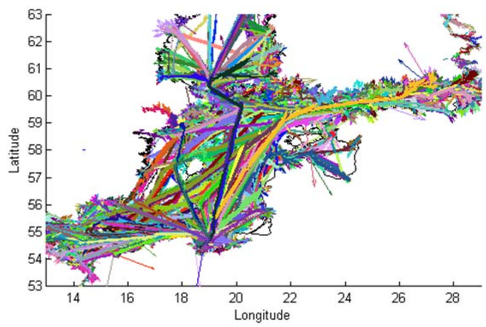

Fig. 4. All semantic routes detected in the monitored area.

have shown to be particularly convenient in the maritime applications since the number of clusters does not need to be pre-set and clusters of arbitrary shapes can be detected.

The semantic routes are derived thereafter, once we obtain some active waypoints and we start seeing vessel trips for one waypoint to another. A semantic route becomes active once a minimum number of trips is observed along it. Moreover, the management of the semantic routes is dynamic since they follow the evolution of waypoints. More specifically, semantic route objects get updated if Ports and Entry/Exit gates are created, expanded and merged. DBSCAN is used on the routes as well, as a post-processing tool to filter out outliers and noise points from the route detections. Given a specific point p, if the cardinality of the neighbourhood of a given radius *eps* is greater than a certain minimum threshold for the number of point in the neighbourhood, then such points are *density-reachable* from p and belong to the same cluster. Moreover, two points p and q are *density-connected* if there is a third point o such that p and q are density-reachable from o. Points that are *density-connected* to each other belong to the same cluster, and points that are *density-connected* to any point of the cluster are also part of the cluster. All the points in a cluster define a semantic route. At the end of the process, those points that are not *density-connected* to other points do not belong to any cluster and, hence, they are considered as noise. A graphic definition of the clustering process is presented in Figure 3. The route detection method is based on TREAD [21].

Finally, the detected waypoints will be used as the main nodes in the Maritime Traffic Network Generation (refer to Section VII). Within the area under analysis, $1,315$ waypoints have been detected, distinguishing between entries, exits and ports areas. Figure 4 presents the maritime traffic amongst such ports defined by $5,827$ semantic routes, which connect the existing waypoints and come determined by their directionality (a.k.a. a route from port X to port Y is different from a route from port Y to port X). Such distinction is relevant to avoid collisions and to respect existing traffic separation schemes.

Considering the large amount of semantic routes co-existing in the area of interest, the longest and most densely populated semantic routes have been selected to evaluate every stage of the system. The 10 selected semantic routes comprise 7 waypoints representing entry points, including 2 entries and 5 ports; whilst the subset presents 6 waypoints acting as exits, including 5 exit points and 1 port. At the end of the paper, in Section VIII, all the semantic routes and ports are analysed to provide the final maritime traffic network in the Baltic Sea.

At this point, the maritime traffic in the Baltic Sea is represented by waypoints and semantic routes, where (1) waypoints are elements providing information such as vessels entrying/exiting/stopping, vessels' kinematic information and geographical area belonging to the waypoint; whilst (2) semantic routes are complex elements including these attributes:

- Time (t): vector determining the timestamp when certain vessels entered and exited the area of interest.
- Route ID (R_{id}).
- Number of vessels associated to the route during the analysed period of time (V).
- MMSI list ($MMSI$): list of the MMSI numbers of the vessels associated to the route.[1]
- Ship type code (STC) list: each vessel in the MMSI list is categorised with a STC, ranging from 20 to 99.
- Way in/out ($EE_{i,j}$): associates the $i-th$ route with one of the Entry/Exit/Port areas detected in the monitored area.
- Points (P): each route contains a set of coordinates and attributes that define its spatio-temporal evolution. Thus, the $i-th$ route contains M attribute vectors describing the AIS messages, as $P_i = (p_{i,j})_{j=1}^{M}$. Each message, $p_{i,j}$, is defined with the following elements: latitude and longitude coordinates (x, y), Course Over Ground (COG), Speed Over Ground (SOG), MMSI number and timestamp (t) as derived by the AIS data standard [19]. A few other attributes associated to the route are also part of $p_{i,j}$, including STC and EE. Formally, the $j-th$ point associated to $i-th$ route is thus defined as:

$$p_{i,j} = (x_{i,j}, y_{i,j}, COG_{i,j}, SOG_{i,j}, MMSI_{i,j}, t_{i,j}, STC_{i,j}, EE_{i,j}). \quad (1)$$

VI. Route Decomposition

Real maritime traffic presents complex scenarios where semantic routes can present different itineraries. Figure 5 (Top) shows a real semantic route where vessels follow different itineraries, implying large geographical deviations. Thus, to accurately estimate vessels behaviour, a prior step must differentiate between different itineraries, decomposing each semantic route into different sub-routes.

The Route Decomposition process proposes an aggregation technique to cluster together similar vessel behaviours

[1] The Maritime Mobile Service Identity (MMSI) is used by AIS to identify vessels, assigning a unique number to each vessel.

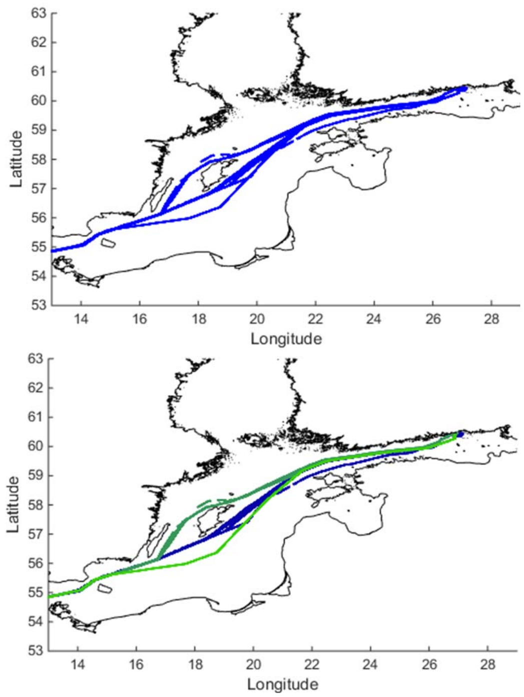

Fig. 5. Route Decomposition. (Top) Semantic route including all vessels following different itineraries. (Down) Semantic route decomposed into three itineraries (from now onwards referred as "routes").

present in a semantic route. The process analyses each vessel itinerary as an individual entity with a unique behaviour. Then, the distance between all vessels traversing the semantic route is compared. Hausdorff distance is proposed as metric to compute the distance between two sub-sets within the same metric space and enabling different set size comparison. Hausdorff distance or $H(A, B)$ "measures the degree of mismatch between two sets by measuring the distance of the point of A that is furthest from any point of B and viceversa" [18]. Moreover, the Hausdorff distance computes the maximum distance between points of different sets without establishing a specific pairing of points between sets. The Route Decomposition process computes the Hausdorff distance in an all-against-all system, calculating the maximum distance between all vessels traversing the semantic route. Finally, vessels are clustered based on their Hausdorff distances, associating vessels whose itineraries present a similar behaviour. The sequential algorithm analyses each individual vessels itinerary, compares it with the other vessels and associates the vessel trajectory with the closest itineraries to form a cluster or "route". The association process thresholds the computed Hausdorff distances, associating vessels whose itineraries distances is less than 0.6 km.[2]

Figure 5 (Down) reveals the three different itineraries detected within the semantic route presented in Figure 5 (Top). The existence of different de-facto trajectories that share the same origin and destination can be due to different reasons, from geographical (i.e. bathymetric characteristics of the area under analysis) to logistic reasons (i.e. fuel consumption optimisation). Considering the different vessel behaviour, distance and characteristics of all the itineraries within a semantic route, from this point onwards, the process will analyse each itinerary separately, named "route", rather than the semantic route, as such entity reveals similar vessel behavioural patterns.

[2]This threshold has been empirically determined and it is tailored to the scenario under analysis

VII. Maritime Traffic Network Generation

The Maritime Traffic Network Generation proposes an algorithm to analyse the decomposed routes, towards the discovery of vessels behavioural patterns in an attempt to synthetically and accurately represent the maritime traffic in the monitored area. The proposed algorithm consists in an unsupervised hierarchical approach extrapolating vessels behaviours from AIS historical data and using such patterns to build the maritime synthetic representation, offering a synthetic view of the activities in the area. The maritime traffic representation is envisaged as a directed-graph or a directed-network, composed of a set of nodes and tracklets. Despite a single layer network could represent maritime traffic, such simplistic approach would neglect real-world situations, including high traffic density areas, confined waterways or no-crossing points, amongst others. Hence, the proposed hierarchical maritime traffic network is built upon two layers: external and internal.

A. Internal Layer

In the internal layer, each route is individually analysed, detecting the changes in the vessels behaviour associated to restrictions in navigation in order to synthetically represent each route. Despite the oceans surface provides an enormous playground for maritime traffic, navigation is typically constrained to certain areas, delineated by economic, legal, logistic, security or safety reasons. These factors can be classified depending on their regulatory or optional character. Some regulatory limitations include maritime protected areas (where maritime traffic is forbidden or restricted), country-restricted maritime traffic, preserved fishing areas or areas regulated by traffic separation schemes. The limitations of optional character are related to economic, logistic or security reasons, including fuel-efficiency, risk and exposition to piracy attacks, exploitation of strategic ports positioning (efficient maritime transport, fuel refilling stations, etc.), amongst others. Whilst the regulatory limitations are clearly established in border definitions and traffic separation schemes, the optional limitations are reflected in *de facto* maritime routes.

The internal layer is composed by nodes named *breakpoints*, which reflect the vessels constant and stable changes of behaviour; and edges named *tracklets*, which represent the trajectory followed by the vessel. The internal layer is meant to add granularity to the graph, adding detail and accuracy to the representation, while maintaining its synthetic character.

A recursive algorithm is proposed to construct the internal layer, analysing each route individually, and based on two steps: route temporal unfolding and breakpoint detection.

1) ***Route Temporal Unfolding****:* All points associated to a route are originally organised in the AIS databases by their latitude/longitude, neglecting the temporal character inherent in the route. Thus, it is necessary to unfold the route on the temporal dimension. The developed algorithm is based on [15]. Initially, all route points are used to compute the route skeleton of the route, named synthetic route. The synthetic route is computed through a recursive algorithm based on two steps, (1) latitude/longitude computation and (2) position prediction.

The algorithm is initialised, assigning one of the waypoints delimiting the route as the initial point, p_{ini}. In (1), the algorithm searches for all points within a circular search cell centered in p_{ini} and with radius of 0.02 degrees. Once the neighbouring points are found, their mean is computed (p_{synth}) and assigned to the synthetic route, together with the estimated bearing. In (2), the next position (p_{next}) is predicted based on the mean computed in (1), p_{synth}, and the estimated bearing. Then p_{ini} is updated with p_{next} and the algorithm returns to step (1), continuing until reaching the other route waypoint. The algorithm constructs the synthetic route, formally defined as $\hat{R} = (p_{synth_m})_{m=1}^{X}$, where X is the number of points in the synthetic route and each p_{synth_m} represents the mean computed from the neighbouring points.

The synthetic route is then used as guideline to unfold the route on the temporal axis. Considering that the synthetic route reveals the route temporal evolution, an aggregation algorithm is proposed to unfold all the points of the route over the synthetic route. The algorithm establishes a search area in the route for each of the X points building the synthetic route, $\hat{R}$. All the points belonging to the $m-th$ searching area are aggregated to p_{synth_m}. Thus, each point of the synthetic route has associated L points, $p_{synth_m} \leftarrow (p_j, COG_j, t_j)_{j=1}^{L}$, where each point comes defined by a set of attributes including latitude and longitude coordinates, ($p_j = (x_j, y_j)$), Course Over Ground (COG_j) and timestamp (t_j). Then the L points associated to p_{synth_m} are organised by their timestamp, $(t_j)_{j=1}^{L}$. Finally, all the points associated to each p_{synth_m} are concatenated, resulting on the temporally unfolded route, $\bar{R}$:

$$\bar{R} = ((p_j, COG_j, t_j)_{j=1}^{L})_{m=1}^{X} \tag{2}$$

where $t_j < t_{j+1} \forall j \epsilon \{1, ..., L\}$ and where the sum of the points associated to each p_{synth_m} is equal to the total number of points of the route under analysis, M, hence $M = \sum_{m=1}^{X} L_m$.

*1) **Breakpoint Detection**:* Vessels perform constant changes in their heading in an attempt to maintain a final direction, however, vessels rarely present large COG variations. Observations on the vessels percourses and coordinates revealed that, vessels perform small COG variations to maintain a direction, addressing meteorological, oceanographic and other external factors; whilst vessels perform constant and stable COG variations when a different direction must be followed. Thus, a vessel route can be primarily identified by their initial and final port/waypoint and later based on the intermediate changes in its direction. The intermediate changes of direction, or *breakpoints (BP)*, determine the geographical points where the vessel changed its spatio-temporal behaviour. The proposed breakpoint detection process relies on the Douglas-Peucker algorithm [10]. The algorithm proposes to optimise a curve representation, given that a curve is formed by linear segments, by reducing the points or segments composing the curve, selecting only the most representative.

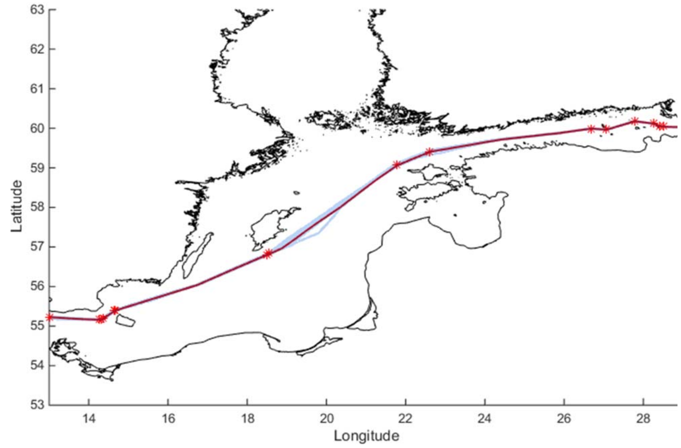

Fig. 6. Analysis of a real route and the resulting network representation. The network representation is composed of the external nodes (entry/exit/port), inner nodes (or breakpoints) and tracklets connecting them.

The algorithm has a recursive character, dividing the curve into segments and computing the most approximated line to the curve segment. Douglas-Peucker algorithm requires the setting of a parameter limiting the maximal distance allowed between the line and the vertex, named *tolerance*, in our implementation, tolerance is set to 0.04 km. Figure 6 shows the Breakpoint Detection process results on a single route.

B. External Layer

The external layer presents the visible elements of the maritime traffic network, including the previously detected routes and waypoints. Based on graph-theory, the waypoints are represented by nodes/vertices, while the routes are represented by edges/lines. The external layer depicts the main characteristics of the route, such origin, destination or directionality, width and precision, as well as the relationship among various routes or the total maritime traffic. A three stages algorithm is proposed to construct the external layer, providing with access to the maritime traffic network and its façade. First, *Maritime Traffic Network Initialisation*, all the segments from one semantic route are archived in the Maritime Traffic Network, shaped as a list of segments and storing uniquely the geographical coordinates delimiting each segment, e.g. $\{x_1, y_1, x_2, y_2\}$. Each route, then, is represented by the route's synthetic representation. Such representation is composed by a list of segments which sequentially ordered form the *maritime lane*. Second, *Maritime Traffic Network Construction*, each segment of a route is compared with the segments stored in the Maritime Traffic Network. If the distance between the segments is under a threshold, both segments are considered to be the same one (in different routes) and so to simplify the graph, both segments are merged. In our implementation, the maximum distance allowed between segments is 10 km. Then, the route and the Maritime Traffic Network are updated. Otherwise, the segment from the route is unique and so it is included in the Maritime Traffic Network. Third, *Maritime Traffic Network Update*, considering that in the previous step, similar segments were merged to optimise the traffic representation, the network should be updated. The second step is repeated to address the

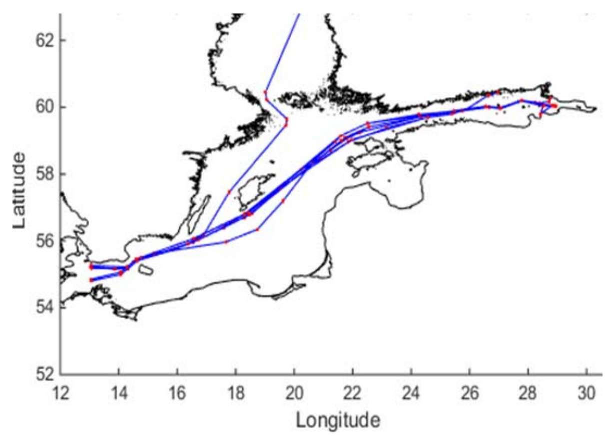

Fig. 7. Maritime Traffic Network automatically obtained analysing the reduced dataset of the Baltic Sea.

route content modification. The process ends when all routes and segments have been merged or added to the Maritime Traffic Network. The resulting Maritime Traffic Network is externally composed by waypoints and routes, while each route is detailed by the fields: (1) *Route ID*, (2) *Route parent* is the semantic route that originated the route, (3) *Segments* represent the tracklets that concatenated compose the route, each segment is represented by $[x_i, y_i; x_{i+1}, y_{i+1}]$, (4) average *COG* of the route, (5) *Route width* and (6) *Route precision*.

C. Performance Evaluation

In this Section, we evaluate the performance of the Maritime Traffic Network Generation over a reduced dataset presenting the longest and most active routes in the Baltic Sea. The dataset is composed of 10 semantic routes equivalent to 106, 281 geographical points.

In previous sections, the concrete analysis of a semantic route was presented. Figure 6 presents a route (blue), its network representation (red) and the detected breakpoints (red asterisks) obtained during the *Inner Layer* algorithm. While the semantic route includes 13, 459 geographic points, the computed Maritime Lane stores uniquely the external waypoints (entry/port and exit/port), the detected breakpoints and the tracklets/segments together with their corresponding attributes. Thus, 13, 459 geographic points are substituted by 2 waypoints and 17 segments, implying a decrease of the 99.75% of stored data to represent that specific route.

Figure 7 presents the obtained Maritime Traffic Network. The dataset, composed of 106, 281 geographical points, was clustered in 10 semantic routes and initially decomposed into 13 routes. The analysis of each route in the Inner Layer and the routes association in the External Layer resulted in a Maritime Traffic Network composed of 180 segments. Each segment is defined as two coordinate points, resulting in 360 points, a 0.34% of the previously required data storage for the original data. Our maritime traffic representation method decreases in a 99.66% the data required to represent the maritime traffic, offering a solution to palliate the exponential increase of storage for maritime traffic data. Nevertheless, the greatest advantage is to provide light, structured and fast access to maritime traffic data, supporting MSA capabilities such as real-time maritime monitoring and anomaly detection.

VIII. Experimental Results

The Baltic Sea is considered one of the most densely trafficked sea regions in the world. Additionally, its geographical position, its large extension, the presence of oil extraction platforms and marine protected areas and its delicate ecosystem expose it as a highly sophisticated scenario. Section II introduced the monitored area statistics and key factors affecting the maritime traffic, its density and distribution.

The Maritime Traffic Knowledge Discovery and Representation System proposed an in-built hierarchical algorithm to analyse self-reporting positioning data, detect vessels patterns and model vessels behaviours. The aim was to automatically create a network representation of the maritime traffic in the monitored area, to support automatic maritime traffic monitoring, situation prediction and anomaly detection. Our system was divided into four algorithms. The impact and results of each stage were partially presented along with the algorithms. In this Section, experimental results are detailed on the complete monitored area, in order to demonstrate the systems performance and to present its benefits.

In the Waypoint and Route Detection algorithms, 1, 315 waypoints were detected (1, 284 ports, 15 entries, 16 exits) and 5, 827 semantic routes (from which 1, 754 Active and 4, 073 Inactive). Figure 1 presents all points of the 5, 827 semantic routes, active and inactive, in the monitored area. Inactive routes depict semantic routes that are not currently in use, hence, revealing non-representative traffic. Our system only considers Active routes, as they represent the current navigation lanes, limiting the analysis to 1, 754 semantic routes. During the learning phase, all the points which do not comply with the routes, incrementally derived, are discarded, i.e. kept in a separate list of noise points and not contributing to build the traffic normalcy. Although this might seem a waste of data, the massive amount of data on one side, and the risks of contamination from noise in the data in the subsequent knowledge exploitation phase, make this choice reasonable. For a theoretical standpoint, the choice is justified by the classical two-phase approach adopted in the statistical process control practice, where during the learning phase only the points compatible with the normal control limits derived are kept in the system. All the points falling outside the normalcy bounds are discarded and the control limits are recursively computed. These limits are used in the action phase to perform the classification of newly seen points. In the same way, we incrementally build a system of routes discarding the noise points during the learning phase, and use that system of routes to perform route classification of new points. Following the same rationale, an additional requirement was set, semantic routes should be traversed by 10 or more different vessels, limiting our analysis to 886 semantic routes, depicted in Figure 8. The routes not fulfilling such minimum are discarded in the learning phase to build the maritime traffic network, being stored in a outliers route list considering their behaviour deviates from majority.

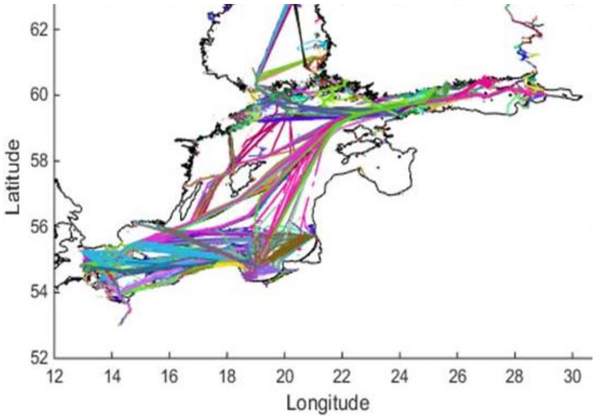

Fig. 8. Complete dataset under analysis. 886 active semantic routes were analysed to build the Maritime Traffic Network.

In order to envision the data storage and handling capabilities required to monitor the Baltic Sea, the whole active traffic must be considered and processed, implying the analysis of 1, 842, 928 geographical points defining the 886 selected semantic routes. Each route not only provides geographical information but also static information such as vessel list, distance travelled, etc. Such big data requires a large storage space and processing time to retrieve any information, hindering maritime traffic monitoring and detection of anomalies. Hence, our method proposed to synthetically represent all maritime traffic to alleviate maritime traffic monitoring systems.

Later, the selected semantic routes were analysed in order to model their inherent behaviour. To ensure the representation accuracy, the semantic routes were decomposed into *routes*. All the routes belonging to the same semantic route shared origin and destination but followed different percourses. As a result, the 886 semantic routes were decomposed in 1, 136 routes. The route decomposition lead to a numerous amount of routes. However, some presented partial itineraries or outliers. In order to maintain the representativity in the system, only routes with more than 250 geographical points were considered, such measure reduced the dataset to 454 routes.

The final stage of our system was the Maritime Traffic Network Generation algorithm, where all routes were analysed to model the maritime traffic and present a synthetic representation, which drastically reduced the amount of data representing the traffic, reducing the data storage and, most importantly, providing a structured and light representation capable to support real-time maritime traffic monitoring and anomaly detection at sea. The proposed Maritime Traffic Network Generation algorithm was divided into two algorithms. First, to model the external network layer, providing the network skeleton and giving a rough approximation of the maritime traffic. Second, to model the internal network layer, providing granularity and precision to the analysis, subdividing the routes into tracklets/segments by detecting the vessels behavioural changes. The resulting network was composed of 454 routes defined by a total of 2, 095 segments (equivalent to 4, 190 points). Therefore, the proposed maritime traffic representation decreases the data volume in a 99.77%.

TABLE II
INDIVIDUAL ROUTES PRECISION ANALYSIS

Route ID	Route Error (km)	Number of Points	Error per point (km)
1	42.77	11003	0.0039
2	141.88	5054	0.0281
3	21.70	6244	0.0035
4	75.01	5687	0.0132
5	80.31	4928	0.0163
6	233.38	7429	0.0314
7	104.49	6785	0.0154
8	8.05	14920	0.0005
9	137.55	2589	0.0531
10	41.83	2134	0.0196
11	102.62	6597	0.0156
12	3.89	5975	0.0007
13	439.36	11552	0.0380

Maximising the data compression consists on reducing the amount of stored data, which in our case translates into representing routes in a simpler way while maintaining the equilibrium between compression and goodness of fit. Two measures can be proposed to account the proposed approach performance: (i) Root Mean Square Error (RMSE) and (ii) complexity of the obtained maritime traffic network.

The RMSE intends to measure approximation provided by the maritime traffic network representation by evaluating the average point standard deviation between the computed representation and the real route points. RMSE is computed applying Equation 3. The distance between the route point $X_i^r = (x_i, y_i)$, and the $r-th$ segment representing the route can be represented as the distance between the point, X_i^r, and the closest set of points from the discretised segments, $\hat{X}_t^r$, where $\hat{X}_t^r = \{X_t^r\}$. The RMSE is then computed for all the points of each route, n_r, and combined for all detected routes within the system, L.

$$RMSE = \sqrt{\sum_{r=1}^{L} \frac{1}{n_r} \sum_{i=1}^{n_r} (X_i^r - \hat{X}_t^r)^2} \quad (3)$$

The complexity of the maritime traffic network is directly related to the number of tracklets/segments.

The 886 active routes comprise 1, 842, 928 points. Applying our method to model and represent maritime traffic over the whole Baltic Sea scenario resulted in a Maritime Traffic Network composed of 2, 095 tracklets. Figure 9 presents the obtained maritime traffic network which compared with the original data presents a RMSE of 19.35 km. The error is reversely represented by the 99.77% decrease in the required storage space to represent the maritime traffic.

To further evaluate the system performance, the RMSE of each route in the reduced dataset is presented in Table II. The results reveal an average error below 30 m per point, in the best cases below 1 m and in one case up to 380 m. Figure 10 shows the approximation of a single route to its network representation, revealing the network goodness of fit.

Figure 9 shows the automatically computed maritime traffic network representing the Baltic Sea maritime traffic extracted from AIS historical data. The resulting maritime traffic network is composed of 454 maritime lanes composed of 2, 095 tracklets/segments and presents an RMSE of 19.35 km. Such error can be neglected in an area as large as the Baltic Sea.

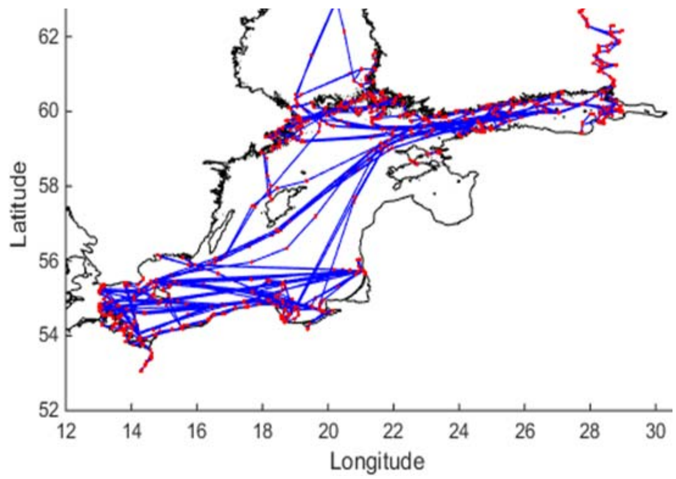

Fig. 9. Maritime Traffic Network obtained from the analysis of the complete Baltic Sea maritime traffic dataset.

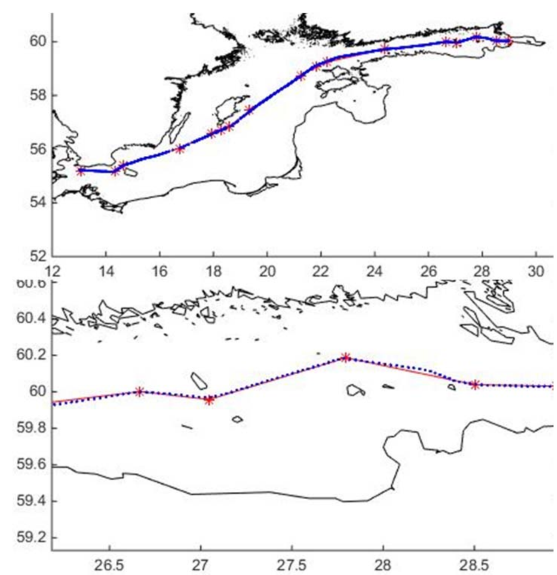

Fig. 10. Route precision. Qualitative comparison between an individual route and its computed network representation. (Top) Overall view of the network representation adaptation to the route. (Bottom) Close-up view of the network representation adaptation to the real maritime traffic data.

Figure 9 demonstrates that the Maritime Traffic Network provides a good trade-off between precision and goodness of fit. Moreover, a light and structured maritime traffic representation is provided, enabling visualisation capabilities as well as further supporting applications such as real-time maritime traffic monitoring, on-the-fly anomaly detection or situation prediction. To enlighten the use of the provided Maritime Traffic Networks for unsupervised traffic monitoring and anomaly detection, Figure 11 presents an example of how the provided network can monitor maritime traffic on-the-fly and detect abnormal behaviours. In Figure 11 a vessel, whose declared origin and destination implied that it would follow the de-facto route represented by the network representation (black), during the first part of its trajectory, the vessel follows such itinerary (blue), however, short after, the vessel deviates from the declared route (red). The provided Maritime Traffic

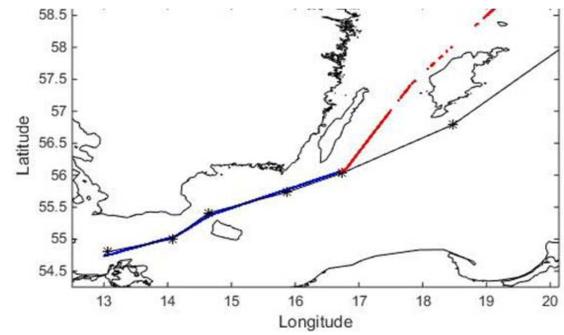

Fig. 11. Example of maritime traffic monitoring and anomaly detection based on the use of the provided Maritime Traffic Network.

Network was used to automatically detect deviations from normality as a first step towards the development of a fully unsupervised maritime traffic monitoring system.

IX. CONCLUSIONS & FUTURE WORK

The proposed approach analysed a representative portion of the Baltic Sea maritime traffic to detect de-facto routes and more specifically to model the traffic on this heavily-transitted area, resulting in a highly-compressed representation, capable to visualise the maritime traffic and to provide the necessary knowledge for real-time maritime traffic monitoring, anomaly detection and situation prediction, enhancing the MSA.

The resulting maritime traffic network provides synthetic representation of the Baltic Sea maritime traffic, reducing the storage space for each route in more than a 99%, while maintaining the representation accuracy. The proposed method tackles maritime traffic representation in the Baltic Sea and reduces $1,842,298$ geographical points to $2,095$ tracklets.

Nowadays, most maritime applications rely on systems built over end-user knowledge rather than on derived maritime traffic knowledge. Despite end-users provide incomparable knowledge, maritime traffic applications could benefit from more precise and numerical knowledge as the one derived from real maritime traffic data, e.g. the presented maritime traffic networks, to support unsupervised maritime traffic monitoring or anomaly detection activities. However, the knowledge and benefits provided by the Maritime Traffic Networks could be exponentially enlarged if combined with other heterogeneous sources of information. The fusion of maritime traffic knowledge with weather forecasts, ship characteristics, ocean currents or special cargo requirements could expand the maritime traffic network applications scope to new fields. Two examples would include: *grounding avoidance* and *weather routing*. In the one hand, maritime traffic networks, if combined with bathimetric maps and ship characteristics, could avoid groundings and tailor routes to vessels considering their specific characteristics. On the other hand, if maritime traffic networks were to be combined with continuously updated weather forecasts, vessels would be able to optimally recompute their routes in real-time, not only to avoid weather-related risk situations but also to control fuel consumption or to maintain vessel scheduling. Thus, future work should focus on the combination and exploitation of derived maritime traffic knowledge and heterogeneous sources of information to enhance the maritime situational picture and to provide additional tools and applications to the end-users.

ACKNOWLEDGMENT

The U.S. Navy and the Volpe Centre of the U.S. Department of Transport are acknowledged for the MSSIS data.

REFERENCES

[1] *IALA*. [Online]. Available: http://www.iala-aism.org/

[2] IALA. *IWRAP Tool*. [Online]. Available: http://www.iala-aism.org/wiki/iwrap/index.php/Main_Page

[3] "Towards the integration of maritime surveillance: A common information sharing environment for the eu maritime domain," European Commission, Brussels, Belgium, Tech. Rep. COM/2009/0538, 2009.

[4] A. Alessandrini *et al.*, "Data driven contextual knowledge from and for maritime situational awareness," in *Proc. 1st Int. Workshop Context-Awareness Geogr. Inf. Services*, 2014, pp. 39–49.

[5] C. Baltic, "News release: Sustained growth of the cruise industry in the baltic sea region," Tech. Rep., 2010.

[6] N. A. Bomberger, B. J. Rhodes, M. Seibert, and A. M. Waxman, "Associative learning of vessel motion patterns for maritime situation awareness," in *Proc. Int. Conf. Inf. Fusion*, Jul. 2006, pp. 1–8.

[7] S. Caschili, F. Medda, F. Parola, and C. Ferrari, "An analysis of shipping agreements: The cooperative container network," *Netw. Spatial Econ.*, vol. 14, no. 3, pp. 357–377, Dec. 2014.

[8] S.-J. Chang, G.-Y. Hsu, J.-A. Yang, K.-N. Chen, Y.-F. Chiu, and F.-T. Chang, "Vessel traffic analysis for maritime intelligent transportation system," in *Proc. IEEE Veh. Technol. Conf.*, May 2010, pp. 1–4.

[9] J. Chen, F. Lu, and G. Peng, "A quantitative approach for delineating principal fairways of ship passages through a strait," *Ocean Eng.*, vol. 103, pp. 188–197, Jul. 2015.

[10] D. H. Douglas and T. K. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," in *Classics in Cartography: Reflections on Influential Articles From Cartographica*, vol. 2. Univ. Toronto Press, 1973. [Online]. Available: http://www.utpjournals.press/doi/abs/10.3138/FM57-6770-U75U-7727

[11] C. Ducruet, Ed., *Maritime Networks: Spatial Structures and Time Dynamics*. Evanston, IL, USA: Routledge, 2015.

[12] M. Ester, H. Kriegel, J. Sander, M. Wimmer, and X. Xu, "Incremental clustering for mining in a data warehousing environment," in *Proc. Int. Conf. Very Large Data Bases*, 1998, pp. 323–333.

[13] M. Ester, H. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. Int. Conf. Knowl. Discovery Data Mining*, 1996, pp. 226–231.

[14] V. F. Arguedas, G. Pallotta, and M. Vespe, "Automatic generation of geographical networks for maritime traffic surveillance," in *Proc. Int. Conf. Inf. Fusion (FUSION)*, Jul. 2014, pp. 1–8.

[15] V. F. Arguedas, G. Pallotta, and M. Vespe, "Unsupervised maritime pattern analysis to enhance contextual awareness," in *Proc. CAGIS*, 2014, pp. 50–61.

[16] M. Fiorini and S.-J. Chang, "Vessel route analysis by *in situ* measurements a pre-requisite for maritime spatial planning," in *Proc. Int. Conf. ITS Telecommun.*, Nov. 2012, pp. 380–384.

[17] J. George, J. Crassidis, T. Singh, and A. Fosbury, "Anomaly detection using context-aided target tracking," *J. Adv. Inf. Fusion*, vol. 6, no. 1, pp. 39–56, 2011.

[18] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, Sep. 1993.

[19] *Technical Characteristics for an Automatic Identification System Using Time Division Multiple Access in the VHF Maritime Mobile Band*, document Rec. ITU-R M.1371-3, ITU Standard, 2014.

[20] M. Kadin, "Trends and scenarios exemplifying the future of baltic sea and skagerack—Ecological impacts of not taking action," Swedish Environmental Protection Agency, Stockholm, Sweden, Tech. Rep. 5875, 2008.

[21] 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, pp. 2218–2245, 2013.

[22] M. Y. Santos, J. P. Silva, J. Moura-Pires, and M. Wachowicz, "Automated traffic route identification through the shared nearest neighbour algorithm," in *Bridging the Geographic Information Sciences*, Berlin, Germany: Springer, 2012, pp. 231–248.

[23] A. Saurama and J. Särkijärvi, "Baltic maritime transport," *Baltic Transp. J.*, vol. 1, pp. 16–18, 2010.

[24] Y.-Y. Su and S.-J. Chang, "Spatial cluster detection for the fishing vessel monitoring systems," in *Proc. OCEANS*, Apr. 2008, pp. 1–4.

[25] *Future Trends in the Baltic Sea*, WWF, Morges, Switzerland, 2010.

Virginia Fernandez Arguedas received the B.S. and M.S. degrees in Telecommunication Engineering from Universidad Autonoma de Madrid in 2008 and the Ph.D. degree in Electronic Engineering from Queen Mary University of London (QMUL) in 2012. From 2008 to 2013, she was a Researcher with QMUL. From 2013 to 2016, she was a Scientific/Technical Project Officer with the European Commission, Joint Research Centre, where she was involved in maritime awareness through the analysis of satellite imagery and self-reporting positioning data. She is currently a Senior Image Processing Engineer with the Deimos–Elecnor Group.

Giuliana Pallotta received the Ph.D. degree in applied statistics in 2009. In 2011, she was a Post-Doctoral Research Associate with University of Washington. From 2012 to 2015, she was a Research Scientist of Maritime Big Data Analytics with the NATO Science and Technology Organization Centre for Maritime Research and Experimentation. She has been a Data Scientist with the Applied Statistics Group, Lawrence Livermore National Laboratory, since 2015. She was a recipient of a Fulbright Scholarship and of the 2013 IEEE JPLC Information Fusion Best Paper Award for her work with Michele Vespe on pattern recognition and anomaly detection.

Michele Vespe received the degree in Telecommunications Engineering from University of Florence in 2003 and the Ph.D. degree in signal processing from University College London, in 2006. From 2008 to 2011, he was a Post-Doctoral Researcher with the Joint Research Centre (JRC). From 2011 to 2013, he was a Senior Scientist with the NATO Centre for Maritime Research and Experimentation, where he was involved in maritime data mining and anomaly detection. He is currently a Scientific Officer with the European Commission, JRC, where he coordinates the activities of a research team in the fields of data analytics, data fusion, and knowledge discovery.