Traffic Knowledge Discovery from AIS Data

Giuliana Pallotta, Michele Vespe and Karna Bryan
NATO Science and Technology Organization (STO)
Centre for Maritime Research and Experimentation (CMRE)
La Spezia, Italy
Email: {pallotta, vespe, bryan}@cmre.nato.int

Abstract—Maritime Situational Awareness (i.e., an effective understanding of activities in and impacting the maritime environment) can be significantly improved by knowledge discovery of maritime traffic patterns. The recent build-up of terrestrial networks and satellite constellations of Automatic Identification System (AIS) receivers provides a rich source of cooperative vessel movement information. This vast amount of information can not be fully utilized by human operators and poses new storage and computational challenges. A compact representation of this rapidly increasing amount of information gives operational utility to data which would otherwise be ignored. This paper proposes an unsupervised and incremental learning approach to extract the historical traffic patterns from AIS data.

The presented methodology called Traffic Route Extraction for Anomaly Detection (TREAD) effectively processes raw AIS data to infer different levels of contextual information, spanning from the identification of ports and off-shore platforms to spatial and temporal distributions of traffic routes. Furthermore, the accurate understanding of the historical traffic enables the classification and prediction of vessel behaviours as well as the detection of low-likelihood behaviours, or anomalies. The ultimate goal is to provide operators with a configurable knowledge framework supporting day by day decision making and general awareness of vessel pattern of life activity. The methodology is demonstrated via a real-world case study, which can be used as a reference data set for further analysis.

I. INTRODUCTION

Maritime transportation is by volume the most utilized form of transportation supporting the global supply chain, making maritime safety and security an important concern. Maritime surveillance data are increasingly used to achieve a high level picture of situational awareness. Cooperative self-reporting vessel location systems, including the Automatic Identification System (AIS), provide a vast amount of near-real time information ([1], [2]), calling for an ever increasingly degree of automation in transforming data to decision support elements. The AIS system was originally conceived for collision avoidance, allowing vessels to broadcast information on their location (position information) depending on their degree of predictability. As the main use of the system is local and real time there are challenges to the use of AIS beyond this scope. For example, vessel originated messages do not contain timestamps but only a reference to the seconds of the UTC time. Despite those shortcomings, AIS still provides a wealth of valuable surveillance data and the messages can be effectively archived in databases or in centralized networks of receivers. Also, the system autonomously organizes in cells whose size depend on the traffic density. Thus a weakness of the system for surveillance data is that in congestion areas the transmission power is decreased.

To further use of the AIS system as a means of global tracking, AIS messages can now be received also from space [3] and a growing number of satellite based data providers have emerged. As opposed to terrestrial networks of AIS receivers whose performance is characterized by high persistence but limited coverage, satellite based systems can pick up messages from open sea, far away from the coastline. Nonetheless, space based receivers are mounted on Low Earth Orbit (LEO) satellites, therefore coverage is linked to the orbiting platform revisit time and, thus, not globally persistent, due to the limited coverage of each satellite.

This methodology is proposed to effectively utilize these data sources while still accommodating data intermittencies found both with terrestrial and space-based networks. A large amount of data is converted into decision support elements independently of the number, platform and performance of the receivers to extract maritime traffic patterns. The extracted maritime traffic patterns can be used for the following applications:

• Traffic knowledge for human operators, providing an up-to-date situation assessment information (e.g. level 2 processing in the Joint Directors of Laboratories (JDL) model [4]). The way routes and traffic change with time and season can help operators in enhancing the knowledge of vessel pattern of life activity and analysts in predicting vessel movement and the impact on traffic when the routing systems are modified.

• Low-likelihood and rule-based Anomaly Detection, detecting deviations from the “normality”. This paper enhances the detection of “low-likelihood” behaviours providing an analytic framework for vessel behaviours. Rule-based approach may also be used to generate alerts based on a set of rules [5], such as maximum speed allowed in a port, presence in areas restricted to navigation or inconsistencies between ship claimed and actual activity.

• Vessel models for modelling and simulation applications. The extracted knowledge provides information on traffic spatial distribution, daily patterns and travel times differentiated by vessel types. This enables realistic simulations of traffic, useful to, e.g., test and evaluate novel tracking techniques and/or emerging technologies [6] and [7].
- Knowledge-based tracking and classification. The distribution and characterisation of traffic represents geographical information that can be used for augmenting remote sensing tracking and classification performance [8]. Specifically, the knowledge of vessels patterns can be used for i) connecting tracks originated by the same target and broken by gaps in coverage or reduced observability and/or ii) providing a prior on the vessel type for classification purposes.

- White shipping density prediction, predicting the density of commercial traffic (white shipping) at a given time. This can be useful for decision support applications requiring knowledge of predicted vessel activity and shipping density. For example, in counter piracy applications it can be used to identify potential high risk areas associated to the joint predicted presence of white shipping density and Pirates Action Groups (PAG) (see, e.g., [9]). Backward and forward tracking of vessels can also be significantly improved by the learned maritime traffic when attempting to fuse AIS and space-based optical or Synthetic Aperture Radar (SAR) information (e.g. [10] and [11]).

- A Source of High level/low level data fusion, allowing AIS static and position information to provide a level of knowledge that can be integrated for threat assessment applications. In particular, a combination of events not necessarily anomalous can lead to more complex and structured alerts ([12]).

The remainder of the paper is organized as follows. Section II reviews related work in the field of traffic analysis. Section III gives an overview of the proposed approach. Section IV applies the methodology to a real data set and conclusions are reported in Section V.

II. RELATED WORK

Several methods have been proposed to derive motion patterns from a collection of trajectories as applied in video surveillance and image processing, where the traffic flows are constrained to stay in specific areas (see [13], [14], [15], [16]). The application of such techniques in maritime domain has gained a recent interest. Previous work has shown the limitations of grid-based methods, when applied in large scale areas or in areas with complex traffic (e.g. [10], [17], [18]). The grid-based methods are effective for small area surveillance applications but require heavy computational burden when increasing the scale of the monitored area and, additionally, they need a priori selection of the optimal cell size for the grid. In areas characterised by complex traffic like intersecting sea lanes, the resulting multi-modal distribution of behaviours would lead to complex algorithms, difficult to be exploited for route propagation and anomaly detection. To overcome these difficulties, some vector-based methods have been proposed. Within this framework, routes are thought of as a set of straight lanes connecting waypoints. As an example, in [19] and [20] a vector-based approach is combined with directional statistics. However these methods do not always perform effectively in unregulated areas. The approach adopted here is based on a preliminary clustering of waypoints, highlighting stationary areas as well as entry and exit points, given a bounding box. Trajectories are then identified between such waypoints. Different from other vector-based methods, the routes are formed by the sequences of flow vectors of the vessels whose trajectories linked such waypoints. As first discussed in [21], it is possible to consistently capture maritime patterns in a compact and accurate way. It is also feasible to extract temporal information like route travel time distributions and daily patterns, as well as to associate historical route patterns to vessels. These features enable the discovery of maritime traffic knowledge that can be used to implement higher level anomaly detection tools. Most available methods require a pre-processing of trajectories since commonly used similarity measures require equally spaced and properly aligned trajectories. In addition, most approaches are thought to work with complete trajectories. This can be difficult in areas where positional data are received only intermittently. Moreover this is not optimal for surveillance purposes where the detection of anomalies needs to be performed on-line. In [22] AIS training data are supposed to be already extracted to perform sequential anomaly detection. The methodology proposed in this work considers data points as single trajectory points. Thus, the traffic representation does not need complete trajectories and performs an event clustering rather than a trajectory clustering, minimizing the issues of unequal-length, incomplete tracks or tracks with gaps, commonplace due to the highly variable refresh rate of AIS messages. In the literature such an approach may be referred to as a point-based approach (see, e.g., [23], [24]), in contrast to trajectory-based approaches (see, e.g., [15]) and enables on-line anomaly detection.

III. TRAFFIC KNOWLEDGE DISCOVERY OVERVIEW

The proposed methodology called Traffic Route Extraction for Anomaly Detection (TREAD) automatically learns a synthetic representation of maritime traffic patterns from low-level AIS data in an unsupervised way. The functional architecture is summarized in Fig. 1.

The achieved knowledge is shaped in a compact form via Vessel Objects, Waypoints Objects and Route Objects, created and updated from the sequence of input AIS messages. Meaningful events are generated in the vessel state vector sequences, including events like a break in observation updates. The clustering of such events, initiated by different vessels objects (Vs), enables to form waypoint objects (WPs) which identify either stationary objects (POs), entry points (ENs) and exit points (EXs) within the selected bounding box. The linking of such waypoints ultimately leads to the detection and statistical characterisation of routes objects (Rs). The general assumption is that the feature values of the data points come from a stationary distribution of normal traffic, estimated using training data.
messages with different time rate. The result of the track
temporal, and attribute properties, is applied to synchronize
A complex strategy, based on a set of rules involving spatial,
stop events can themselves indicate an anomalous behaviour.
The way in which the tracks are split is crucial since those
periods near land indicates a point at which to split the track.
when the vessel is stopped or is not transmitting for long
MMSI number) is handled using velocity gating. The time
i.e. the elimination of duplicate data (i.e.,
field (which is present both in the static and kinematic fields)
receiving the MMSI number and contains both static and dynamic properties. While
the former are linked to the identification of the vessel (e.g.,
vessel type, call sign, name, IMO number, size), the latter
are related to state vector (e.g., position, Course Over Ground
(COG), Speed Over Ground (SOG)) and to historical and
and they are related to state vector (e.g., position, Course Over Ground
(COG), Speed Over Ground (SOG)) and to historical and
current route patterns. Changes of status of vessel objects
are events of interest such as “lost” when not observed for
a given time window, which is a multiple of the maximum
AIS message refresh rate in the area of interest. Additional
vessel status are “stationary”/“sailing”, and their transitions
identify other events of interest such as when the vessel stops
or starts sailing again from a stoppage. Such events create or
update waypoint objects $WPs$.

Waypoints Objects: Stationary Objects $POs$ (i.e., ports,
offshore platforms and stationary areas) are formed by vessels
having a speed lower than a given threshold. Specifically, sta-
tionary events are detected by speed gating. The last observed
displacement in the vessel position is used to empirically
derive an average speed since the field SOG in AIS messages
is unreliable to be used in detecting such stationary events.
Similarly, Entry/Exit Points ($ENs$ and $EXs$, respectively) are
created and dynamically updated, when a vessel enters/leaves
the selected area of interest, being associated to “birth”/“death”
events (corresponding to vessel status transition “transmit-
ting”/“lost” and vice versa). As in image processing and visual
surveillance (see, e.g., [16]), entry and exit points are related
to the monitored scene and may change depending on the
bounding box area, while stationary objects are fixed reference
points. Both Stationary Objects and Entry/Exit Points are
created, expanded and merged progressively using an incre-
mental DBSCAN (i.e., Density-Based Spatial Clustering of
Applications with Noise) procedure (see [25] and [26]). The
clustering parameters are set, based on the nature of the $WPs$
(i.e., whether they are $POs$ or $ENs/EXs$) and on the specific
scale of the selected area. More specifically, DBSCAN forms
clusters of elements, on the basis of the density of the points in
their neighbourhood. 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

C. AIS Data transformation into an Object-based structure

As soon as a new vessel enters the selected area of interest,
a detection occurs and the management of the Object-based
structure is initialized as follows.

Vessel Objects: The list of Vessel Objects $Vs$ is updated
according to the information content of each decoded AIS
message (or database record when performing historical data
analysis). Every Vessel Object is identified by the MMSI
number and contains both static and dynamic properties. While
the former are linked to the identification of the vessel (e.g.,
vessel type, call sign, name, IMO number, size), the latter
are related to state vector (e.g., position, Course Over Ground
(COG), Speed Over Ground (SOG)) and to historical and
current route patterns. Changes of status of vessel objects
are events of interest such as “lost” when not observed for
a given time window, which is a multiple of the maximum
AIS message refresh rate in the area of interest. Additional
vessel status are “stationary”/“sailing”, and their transitions
identify other events of interest such as when the vessel stops
or starts sailing again from a stoppage. Such events create or
update waypoint objects $WPs$.

Waypoints Objects: Stationary Objects $POs$ (i.e., ports,
offshore platforms and stationary areas) are formed by vessels
having a speed lower than a given threshold. Specifically, sta-
tionary events are detected by speed gating. The last observed
displacement in the vessel position is used to empirically
derive an average speed since the field SOG in AIS messages
is unreliable to be used in detecting such stationary events.
Similarly, Entry/Exit Points ($ENs$ and $EXs$, respectively) are
created and dynamically updated, when a vessel enters/leaves
the selected area of interest, being associated to “birth”/“death”
events (corresponding to vessel status transition “transmit-
ting”/“lost” and vice versa). As in image processing and visual
surveillance (see, e.g., [16]), entry and exit points are related
to the monitored scene and may change depending on the
bounding box area, while stationary objects are fixed reference
points. Both Stationary Objects and Entry/Exit Points are
created, expanded and merged progressively using an incre-
mental DBSCAN (i.e., Density-Based Spatial Clustering of
Applications with Noise) procedure (see [25] and [26]). The
clustering parameters are set, based on the nature of the $WPs$
(i.e., whether they are $POs$ or $ENs/EXs$) and on the specific
scale of the selected area. More specifically, DBSCAN forms
clusters of elements, on the basis of the density of the points in
their neighbourhood. 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

A. Target Database Selection

Input parameters can be selected such as the timeframe,
the utilized sensors (i.e., ground-based and/or satellite-AIS
receivers), the bounding box area (which can be arbitrarily
shaped) and the parameters of the clustering and filtering
procedures. The methodology can also work in a single-route
mode, retrieving the historical routes which connect two pre-
selected areas. This could be used to monitor the activity of a
specific port or the ongoing traffic streams between two points
of interest. Additionally, given a ship type of interest, a filtered
description of the traffic can be learned and specific ship-type
routes are derived.

B. AIS data pre-processing

The first step to create an Object-based structure is to pre-
process AIS data, before using them as training data for the
learning. The pre-processing involves a preliminary cleaning
phase before applying the track management (i.e., separating
the AIS data stream into vessel tracks). In this phase the MMSI
field (which is present both in the static and kinematic fields)
is used to assign each message to a distinguished track. Then
the elimination of duplicate data (i.e., different vessels sharing
MMSI number) is handled using velocity gating. The time
when the vessel is stopped or is not transmitting for long
periods near land indicates a point at which to split the track.
The way in which the tracks are split is crucial since those
stop events can themselves indicate an anomalous behaviour.
A complex strategy, based on a set of rules involving spatial,
temporal, and attribute properties, is applied to synchronize
messages with different time rate. The result of the track
management is the creation/update of the Vessel Objects with
the relating tracks stored.
part of the cluster. In this framework, those points that are.* not density-connected* to other points do not belong to any cluster and are considered as noise. Differently from centroid-based clustering, DBSCAN does not require the number of clusters a priori, while arbitrarily shaped clusters can be easily found as often observed within the maritime traffic context. For instance, centroid-based methods can fail in discriminating different ports whose centroids are close to each other, when they are located along the coast line. Moreover, DBSCAN introduces a way to classify noise points, which can be used to detect and filter outliers, as will be shown hereafter.

**Route Objects:** Route Objects $Rs$ are extracted by connecting the derived $WPs$. The Route Objects do not merely count the registered transiting vessels but synthesize both the static (e.g., related to the type of vessel and its identifier) and dynamic (state vector observations) features inherited by the vessels that created or updated them. $Rs$ are dynamically managed, also following the $WPs$ creation/expansion/merge operations. The Route Object is activated when a minimum number of detections (i.e., number of not necessarily unique vessels that transited along the route) has been reached. The set of $Rs$ forms historical atlases, which are a compact representation of the maritime traffic clusters over the considered area.

The marine traffic patterns in the area can be distinguished into three types of traffic concepts or sailing routes:

- **Local routes:** routes between stationary locations (ports and off-shore platforms) within the area of interest;
- **Destination routes:** routes from stationary locations (ports and off-shore platforms) inside the the area of interest to destinations outside of the region and vice versa;
- **Transit routes:** routes crossing the area of interest by entering and exiting the bounding box.

**Synthetic Route Extraction:** To enhance the operational monitoring over busy areas of interest, the obtained route density clusters can be compressed into a compact form called synthetic routes. Each synthetic route summarises the expected “central” behaviour along the route. This is obtained by sequentially applying the Algorithm *Synthetic Route Generator* whose pseudo-code is reported below. Generally speaking, a vessel track $V$ is a time series of $T$ observed state vectors $v_i$:

$$V = \{v_1, v_2, ..., v_T\}$$ (1)

where the $t$-th state vector observation $v_t$ is directly isolated from the broadcast AIS information. In this study it includes both position and velocity information as extracted by the vessel track properties:

$$v_t = [x_t, y_t, \dot{x}_t, \dot{y}_t]^T$$ (2)

where $x_t$ and $y_t$ are related to the vessel coordinates and the velocity components $[\dot{x}_t, \dot{y}_t]$ are derived by combining SOG and COG information.

The algorithm incrementally tracks the expected position $[x_P, y_P]$ of a vessel (of a given type $v$) along the $k$-th route $R^k_v$.

The synthetic route starts from the centroid of the waypoint $WP_1$, and is propagated according to the kinematic features distribution $S_P$ in the $\epsilon$-neighbourhood of the current position $[x_P, y_P]$ until the waypoint $WP_2$ polygon is reached. The radius $\epsilon$ of the neighbourhood is dynamically adapted, based on the specific route cluster spread. The time increment $step_t$ can be conveniently chosen, depending on the complexity of the route and it affects the ability of the interpolated route to follow the actual route (see [22]).

**Algorithm 1 Synthetic Route Generator**

Require: $R^k_v$, $WP_1$, $WP_2$, $step_t$, $\epsilon$

1: RouteLength $\leftarrow 0$
2: $[x_P(1), y_P(1)]$ $\leftarrow$ centroid($WP_1$)
3: while not(inpolygon($[x_P(end), y_P(end)]$, $WP_2$)) do
4:     find $\ell$ s.t. $\forall l : \|R^k_v(l)[x, y] - [x_P(end), y_P(end)]\| \leq \epsilon$
5:     $[\dot{x}_P(end), \dot{y}_P(end)]$ $\leftarrow$ median($S_P$)
6:     $[x_P(end + 1), y_P(end + 1)]$ $\leftarrow$ $[x_P(end), y_P(end)]$
7:     RouteLength $\leftarrow$ RouteLength + dist($[x_P(end), y_P(end)]$, $[x_P(end - 1), y_P(end - 1)]$)
8: end while
9: return $[x_P, y_P]$, RouteLength

**D. Traffic Knowledge discovery**

Knowledge Discovery (KD) is generally defined as the process of extracting unknown patterns from large volumes of raw data with innovative methods or tools [27]. In this study the analysis of the obtained traffic patterns (active and highly populated routes) leads to the knowledge discovery of the traffic over the area of interest. TREAD converts a large amount of AIS data over a time window. As typical of Information Theory, part of these data is valuable and is included into the historical traffic pattern model. Some other data do not contribute to the traffic information gain and are discarded. Thus, the learning capability of the proposed methodology can be linked to the notion of entropy, which can be used to evaluate the degree of disorder within a single route cluster and over the whole area of interest. The more the entropy, the more is the disorder in the system. One important consequence of such evaluation is the extent to which the traffic can be predicted on the basis of the historical patterns over the area. Entropy quantifies the information gain that the derived traffic patterns will provide for prediction [28]. In Information Theory entropy is widely employed to predict human mobility, ATM traffic streams and cellular networks traffic [29]. In geographical clustering studies the notion of entropy has been suggested in [30].

**Route Complexity Evaluation:** Entropy is a potential descriptor of motion patterns disorder [31]. The classical Shannon entropy of a discrete set of $n$ random variables $x_i$, with probabilities $P(x_i)$ can be expressed as follows:

$$H = - \sum_{i=1}^{n} P(x_i) \log_2 P(x_i)$$ (3)

where $\sum_{i=1}^{n} P(x_i) = 1$. This broad concept has recently found interesting applications in cluster quality assessment.
and traffic anomaly detection. As an example, in [32] an entropy approach is proposed to detect abnormal activities in video surveillance, based on the degree of randomness in both directions and displacements in video frames.

Similarly, the entropy of a route cluster can be assessed with respect to both position and course over ground (COG) of the vessels travelling along the route. In the future, additional features (i.e., speed over ground) can be included into the entropy computation. A within-route entropy can be defined, based on the similarity and homogeneity purposes for which AIS data clustering is being done. This gives an insight into the degree of uncertainty in the derived clusters and provides an additional post-processing metric to evaluate the learned system of routes. Thus, the entropy of a generic route cluster \( R_k \) can be expressed as:

\[
E_{R_k} = \sum_{j=1}^{J} P_j \log_2 \left( \frac{1}{P_j} \right)
\]  

(4)

where the index \( j \) is related to the basis of selected route features. Low values of entropy are an indication of vessel concentration in terms of positions and directions.

The synthetic route can be used as the reference “expected” path for the vessels travelling along the given route. The Euclidean distances of the point positions from the synthetic route are used to measure the route cluster spatial compactness. The root mean square deviation gives a measure of the extent (or dispersion) shown by the points with respect to the synthetic route. Similarly, the vessels’ dis-alignments are computed as the relative angular distances of the vessel COG values from the expected COG value, estimated based on the synthetic route neighbours, using the circular statistics principles (see, e.g., [33]). The circular median of such relative angular distances is used to express the degree of directional dis-alignment of the vessels in the route. As suggested in [32] the dispersion in positions and directions needs to be dimensionless and normalized in order to be combined into an entropy metric. The normalization is obtained by dividing the quantities by the maximum observed average dispersion in the area. Since each route cluster has a relative weight depending on its size (i.e., linked to the number of points forming the cluster and to the number of transits along that route), we can also derive an overall traffic entropy as a weighted average of the single route entropy values, as suggested in [34].

IV. CASE STUDY: CMRE TYRRHENIAN DATA SET

The NATO Science and Technology Organisation Centre for Maritime Research and Experimentation (STO-CMRE) owns an AIS receiver and collects data for research purposes. The navigation traffic data collected using this receiver are used to illustrate the knowledge discovery process as described in Section III. The data set used for the analysis is real world AIS data covering the time frame between January 1st to February 20th 2013 for a section of the Northern Tyrrhenian Sea framing La Spezia (see Fig. 2). The data initially contained eight fields: the vessels MMSI, the timestamp, the latitude and longitude of the vessel, the vessel speed, course and type.

<table>
<thead>
<tr>
<th>VESSEL TYPE DISTRIBUTION AS DERIVED FROM THE AIS DATA IN THE AREA AND PERIOD OF INTEREST.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cargo Vessels</td>
</tr>
<tr>
<td>Unclassified</td>
</tr>
<tr>
<td>Tankers</td>
</tr>
<tr>
<td>Passenger Vessels</td>
</tr>
<tr>
<td>Others</td>
</tr>
<tr>
<td>Pleasure Crafts</td>
</tr>
<tr>
<td>Tugs</td>
</tr>
<tr>
<td>Sailing</td>
</tr>
<tr>
<td>Towing</td>
</tr>
<tr>
<td>Dredging or underwater ops.</td>
</tr>
<tr>
<td>Total number of tracked vessels</td>
</tr>
</tbody>
</table>

The navigational status field was removed from the data since in most cases was unreliable (containing invalid information) or either redundant (containing information already covered more reliably by other fields such as SOG). Table I summarises the overall distribution of vessel type in the area. The selected bounding box covers an area of about 46 x 60 nautical miles in the Ligurian Sea, as shown in Fig. 2.

<table>
<thead>
<tr>
<th>SUMMARY OF OBJECTS DETECTED IN THE AREA.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object Type</td>
</tr>
<tr>
<td>Stationary areas</td>
</tr>
<tr>
<td>Entry Points</td>
</tr>
<tr>
<td>Exit Points</td>
</tr>
<tr>
<td>Routes</td>
</tr>
</tbody>
</table>

¹Only Objects with a critical mass of vessels (classified as "Active") are displayed, out of all the detected ones.
Table II reports the number of the Objects detected in the area. Each Object stores historical information for second-level analysis.

A. Historical Route Atlas

The data clustering of vessel patterns leads to the filtered color-coded route clusters summarised in Fig. 3a.

The resulting atlas of historical routes provides a codebook of traffic activity in the area of interest as shown in Fig. 4. Fig. 3b displays the corresponding synthetic routes in the area as discussed in Section III-C.

B. Route Object features

Each route incorporates different levels of information which enable spatio-temporal and attribute analyses. As an example, Fig. 5 gives the summary for a selected route. Some additional statistics are also produced such as the route length, the median speed along the route, the number of transits, their median duration and the list of vessels which travelled along the route.

C. Route Cluster evaluation

The route cluster evaluation described in Section III-D can be performed to measure the level of traffic disorder in the area. Fig 6 gives an example of such evaluation. As can be observed, the level of entropy associated to a route is in direct proportion with the relevant route trajectories.

V. RESULTS AND DISCUSSION

A novel framework has been presented for deriving motion patterns in maritime traffic in a fully unsupervised way using AIS data only. The methodology has demonstrated the ability to effectively recognize the main routes and agrees with the nautical charts.
A compact and configurable knowledge structure underlines the traffic patterns and can be displayed and studied for the specific analysis purposes. The application has shown a modelling technique which can be used to enhance route prediction and the detection of low-likelihood behaviours, discovering the needed prior knowledge of maritime domain. The example provided in this paper demonstrates the often unconstrained nature of maritime traffic and shows the challenges in providing automated techniques to identify low-likelihood behaviours based on “normal” traffic patterns. As future work, the authors propose development of an entropy measure which can provide a level of traffic order (disorder) in an area as a precursor to the development of completely automated routines.

Fig. 5. Single Route summary statistics: route cluster; synthetic route; transit duration distribution along the route; Vessel type distribution; COG distribution; SOG distribution

Fig. 6. An example of route entropy ranking in the area of interest

REFERENCES

[1] Safety of Life at Sea (SOLAS) convention Chapter V. Regulation 19.


