

# AIS Based Shipping Routes Using the Dijkstra Algorithm

P. Silveira, A. P. Teixeira & C. Guedes Soares

*Centre for Marine Technology and Ocean Engineering (CENTEC), Instituto Superior Técnico, Universidade de Lisboa, Portugal*

**ABSTRACT:** This paper proposes an approach for identifying and characterizing shipping routes using information contained in Automatic Identification System messages broadcasted by ships and recorded by the coastal Vessel Traffic Service centre. The approach consists of using historical Automatic Identification System data to build a graph, where nodes are cells of a grid covering the geographical area being studied and the weights of directional edges are inversely related to ship movements between cells. Based on this graph, the Dijkstra algorithm is used to identify a potential safe route, assumed to be the most used route by ships between two locations. A second graph is created simultaneously, with the same nodes and edges, but with edge weights equal to the average speed of transitions between cells, thus allowing the determination of the average speed profile for any possible path within the graph. The proposed approach is applied to two scenarios: an approach to the port of Lisbon and the entry through the fairway to a RO-RO terminal in the port of Setubal in Portugal.

## 1 INTRODUCTION

For many years navigators have been relying on charts, pilot books (sailing directions) and other nautical publications as sources of information for safe voyage planning. Predefined shipping routes have been used since 1898, when they were used, due to safety reasons, by passenger ships across the Atlantic (IMO, 2017). The growth in the number of ships brought the need to organize maritime traffic, especially in congested waters where head-on encounters were likely to occur. Traffic Separation Schemes (TSSs) were adopted, first in the English Channel and progressively all over the world. Other routing measures have been adopted to organize traffic and to aid the navigator when planning a voyage, such as roundabouts, two-way routes, recommended routes, areas to be avoided, deep water routes and precautionary areas. However, in many

navigable waters around the world, the decisions regarding the ship routes are not constrained by these routing measures, and depend largely on the navigator's (and ultimately the Master's) choices. Moreover, when entering or leaving ports or during other critical maritime operations, a local pilot is often required. Local pilots are expert ship-handlers with extensive knowledge of the characteristics of particular waterways, acting as the ship master's advisors. In many cases, pilots have the conn (directly controlling the ship's manoeuvrability) during some or all stages of the operation. In this context, the definition of reliable safe paths in areas where traffic flow is not regulated and/or where navigation errors can quickly lead to accidents will reduce the uncertainty of each operation, thus increasing the overall safety level and the efficiency of the maritime operations. These paths can be identified and properly characterized by analysing historical

Automatic Identification System (AIS) data that have been collected over the last years, since AIS became mandatory to most commercial vessels.

AIS data have been used by several authors for maritime traffic characterization (e.g. Silveira et al., 2012, Rong et al., 2018), for collision risk assessment (e.g. Silveira et al., 2013, 2015; Rong et al., 2016) and for developing simulation models of ship navigation in congested waterways (e.g. Rong et al., 2015a,b).

Several algorithms have been proposed for characterising maritime traffic routes from historical AIS data. In particular, Pallotta et al., (2013) proposed a methodology called Traffic Route Extraction and Anomaly Detection (TREAD), which automatically learns a statistical model for maritime traffic from AIS data in an unsupervised way. Discontinuous events are clustered to form waypoint objects. The linking of these waypoint objects leads to the characterization of route objects. Anomalies can be detected by comparing the found routes with real-time traffic. This research builds on the work of Vespe, et al., (2012), which developed the framework aiming at automatically learning AIS maritime traffic patterns using an unsupervised approach that works in real-time. Etienne et al., (2010) proposed a process to extract trajectories from data of moving objects and to identify unusual behaviour. The process starts by storing positions in a spatio-temporal database, after which a zone graph is set up and a cluster of trajectories of objects following the same itinerary extracted from the database. Then, a statistical analysis is performed to compare patterns in order to qualify the behaviour of a mobile object. The process was applied using AIS data in the Brest area.

The present paper explores the use of the well-known Dijkstra's shortest path algorithm (Dijkstra, 1959) for establishing safe paths of ships based on historical AIS data. The Dijkstra algorithm, as well as several improved versions of the original method, have been widely used in different contexts such as for ship route planning, logistics management, and many others network optimization problems that can be formulated as shortest path problem (e.g. Joo et al., 2012; Takashima et al., 2009; Mannarini et al., 2013; Neumann, 2016).

The approach proposed in this paper consists of using AIS data to build two graphs. In the first one the nodes of the graph are cells of a grid covering the geographical area being studied and the weights of directional edges are inversely related to ship movements between cells. The second graph is created with the same nodes and edges, but with edge weights equal to the average speed of transitions between cells. Based on these graphs, the Dijkstra algorithm is used to identify the most used route by ships between two locations and the average speed profile for any possible path within the graph.

The proposed approach is applied to two scenarios: the approach to the port of Lisbon from the southbound traffic lanes of the "Off Cape Roca Traffic Separation Scheme" and the entry of the fairway to a RO-RO terminal in the port of Setubal in Portugal.

## 2 SAFE PATHS FROM AIS TRAFFIC DATA

This section details which information can be obtained from AIS data and describes the proposed method in detail. Some implementation aspects, such as grid resolution, ship speed and data quantity are discussed.

### 2.1 AIS data

AIS allows automatic exchange of information between stations (ships and coastal), using VHF radio waves. There are 27 message types defined in the International Telecommunication Union (ITU) recommendation M.1371-5 (ITU, 2014), and two classes of shipboard equipment: class A (used mainly by commercial vessels) and class B (used mainly by fishing vessels and pleasure craft). The reporting intervals of class A equipment vary between 2 seconds and 10 seconds (depending on the ship's speed and rate of turn) if the ship is not moored or at anchor. If the ship status is moored or at anchor, the reporting interval is 3 minutes, unless the speed is greater than 3 knots, which sets the re-reporting interval to 10 seconds. The message types used in this study were position reports (message type 1, 2 and 3) and static and voyage related data (message type 5) transmitted by class A equipment. Information contained in position reports includes date/time, Maritime Mobile Service Identity (MMSI) number, navigation state, rate of turn, speed over ground, position accuracy, latitude, longitude, course over ground and heading. Static and voyage related data messages include information on MMSI number, International Maritime Organization (IMO) number, ship's name, destination, Estimated Time of Arrival (ETA), callsign, type of ship, length, breadth and draught. AIS records used in this work were collected by the Portuguese coastal VTS, between 05/05/2012 and 21/06/2013.

### 2.2 Method description

The proposed method to define safe paths based on AIS records starts by decoding message types 1, 2, 3 and 5, transmitted by class A ship equipment. From these data, decoded messages originated from the geographical area under study are selected for further use. Maximum and minimum latitudes and longitudes define the geographical area. The next step consists of defining a grid with square cells, with cell side length defined according to the required resolution.

The number of grid rows and columns is adjusted so that the grid totally includes the area under study. Figure 1 shows an example of a grid on a traffic density map, covering one of the geographical areas selected for application of the described method: the approaches to the port of Lisbon.

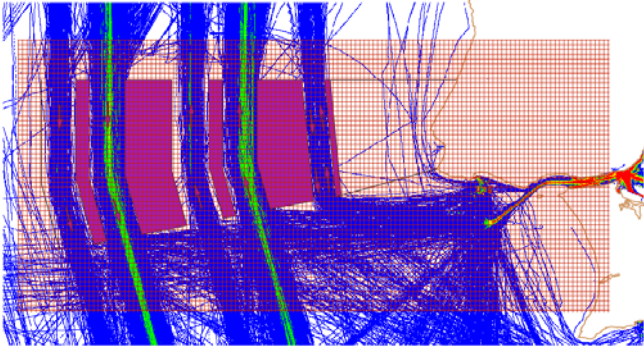


Figure 1. Example of traffic density map with grid, TSS and approaches to the port of Lisbon

Once the grid is defined, a weighted directed edge graph that represents the ship traffic is created. This graph may be created using all previously selected AIS messages or applying filters for ship type and/or length. Each grid cell is potentially a graph node. AIS messages are processed in the order they were received and for each position report (AIS message types 1, 2 and 3) the ship position is matched to a grid cell. When a ship changes cell, the following process takes place:

- 1 if the origin or destination cells are not yet graph nodes, these nodes are added to the graph;
- 2 if there is no edge connecting the origin and destination nodes, an edge is created with weight equal to one, otherwise, the edge weight is increased by one.

When all AIS messages are processed the graph creation process ends. At this stage, the weight of each edge of the graph is the number of times a transition between the origin and destination cells occurred.

The Dijkstra (1959) algorithm finds the shortest path between nodes in a directed graph, where weights represent distances between nodes (or vertices). The graph weights must be positive. The pseudocode (Cormen et al., 2009) for this well-known algorithm, which takes as arguments a directed graph and a source node or vertex, is:

*Inputs: A weighted, directed graph  $G = (V, E)$ , with all edge weights nonnegative; a source vertex  $s$ .*

```

for each vertex  $v \in G.V$ 
   $v.d = \infty$ 
   $v.\pi = NIL$ 
 $s.d = 0$ 

 $S = \emptyset$ 
 $Q = G.V$ 
while  $Q \neq \emptyset$ 
   $u = \text{EXTRACT-MIN}(Q)$ 
   $S = S \cup \{u\}$ 
  for each vertex  $v \in G.Adj[u]$ 
    if  $v.d > u.d + w(u, v)$ 
       $v.d = u.d + w(u, v)$ 
       $v.\pi = u$ 

```

*EXTRACT-MIN(Q) removes and returns the element of Q with the smallest weight*

Since in this case the intention is to find the “most used” route, the edge weights must be inverted and all weights must remain positive in order to apply the algorithm. This transformation is performed in the following way:

$$w_{ij} = w_{max} - w_{ij} + 1 \quad (1)$$

where  $w_{ij}$  is the weight of the directed edge between node  $i$  and node  $j$  and  $w_{max}$  is the maximum edge weight in the graph. Once the edge weights are calculated, the Dijkstra algorithm is used to find the most used trajectory between an origin node and any other node in the graph. However, it may be more interesting to find the best route between a group of origin cells (origin area) and a group of destination cells (destination area), because one cell may represent a very small geographical area, depending on the grid resolution.

Regarding the grouping of destination cells and since the Dijkstra algorithm solution includes the minimal cost path from an origin cell to any other cell in the graph, it is only necessary to choose the best solution from the set of destination cells.

As for the grouping of origin cells, an extra step is added to the proposed algorithm, which consists of joining all origin cells that are connected through edges to non-origin cells, updating the edge weights as cells are joined, until only one origin cell remains.

The path with the lowest cost is found by applying the Dijkstra algorithm on this new graph.

### 2.3 Path speed profiles

To determine the speed profile of any possible path within the previously mentioned graph, a second graph is created simultaneously. Nodes and edges are added as described before, the difference being in the weight of the edges, which in this second graph corresponds to the average Speed Over Ground (SOG) of ships when changing cells. Once a safe path is determined using the first graph, the speed profile for that path can be found by retrieving the path’s edge weights on the second graph.

### 2.4 Implementation aspects: Grid resolution, ship speed and data quantity

It should be noted that the choice of the grid cell side length not only influences the resolution of the path obtained, but may also bias the results towards the paths chosen by slower ships.

As stated before, class A AIS transmits position reports every 2 to 10 seconds, depending on ship speed and rate of turn. If the grid cell dimensions are small enough for faster ships to skip one or more cells in transitions between cells, the paths taken by slower ships end up contributing to more edge weights, therefore biasing the results. To avoid this, cell side in nautical miles (*nm*) should be chosen according to the following criterion:

$$cell\ side(nm) \geq \frac{V_f(kts) \times ri_{max}(s)}{3600(s)} \quad (2)$$

where  $V_f$  is the estimated speed of faster ship and  $ri_{max}$  is the maximum reporting interval.

A problem related to the choice of grid cell side length may occur when the cell dimensions are too small considering the amount of available AIS data, which, to the limit, may result in all edge weights in the graph being equal to one, before transformation.

In this case, assuming ships have equal speeds, the algorithm would choose the longer path, for which the sum of inverted edge weights is smaller.

### 3 APPLICATION TO PORT APPROACHES

The process previously described to define the safe paths is applied to a port approach scenario: the approach to the port of Lisbon from the southbound traffic lanes of the “Off Cape Roca TSS”. Four graphs are created, corresponding to four different periods of the year: May 2012, August 2012, November 2012 and February 2013. The parameters for creating the graph are:

- Grid cell side: 0.01 nautical miles;
- AIS ship types: 70-79 (cargo ships);
- Ship length: greater or equal to 100 m.

Figures 2 to 5 show the traffic density for the four time periods in the relevant geographical area. Figure 6 shows the paths obtained for the four time periods, using all AIS messages received from ships with selected types and length, as well as the grid, origin and destination areas.

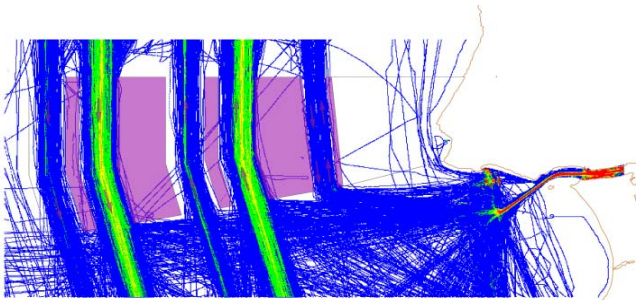


Figure 2. Traffic density, May 2012

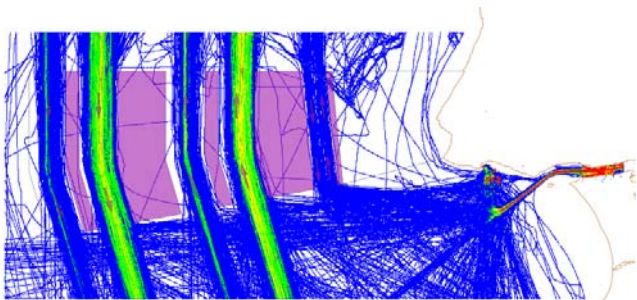


Figure 3. Traffic density, August 2012

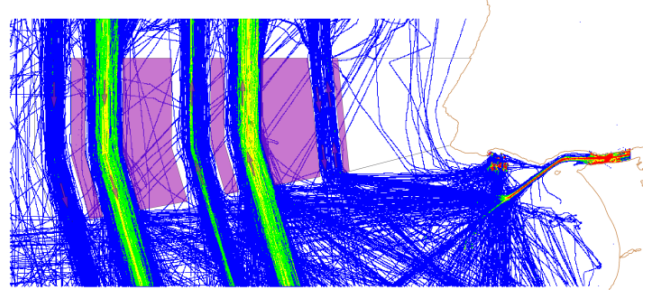


Figure 4. Traffic density, November 2012

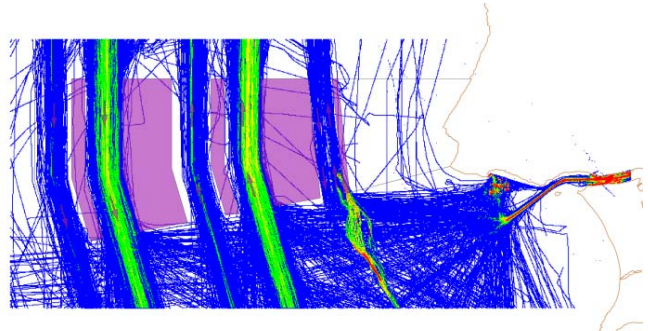


Figure 5. Traffic density, February 2013

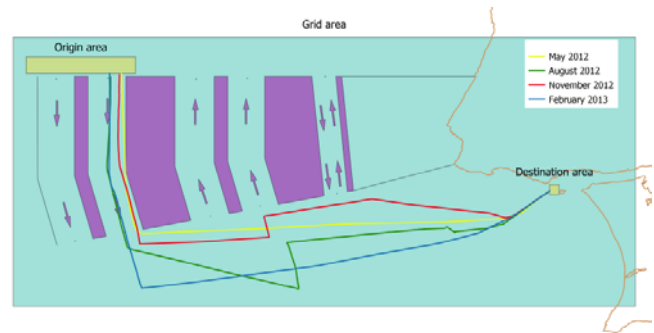


Figure 6. Paths obtained from all AIS messages, selected ship types and length

Although the May and February paths seem to be good candidates to a safe path between the origin and destination areas, the August and November paths are clearly suffering from the influence of the dense northbound traffic headed to the TSS. To remove this influence, a new graph is created following the same process, but this time using only AIS messages received from ships that travelled between the origin and destination areas (for the same ship types and length). Due to the existence of an anchorage in Cascais, it is possible for some ships to cross the origin area, proceed to the anchorage, where they may stay for any period of time, and then proceed inside the port of Lisbon, crossing the destination area. For that reason, all voyages taking more than 700 minutes (equivalent to a minimum average speed of about 4.7 knots) were excluded. Since the amount of selected AIS messages was much smaller, grid cell side was increased to 0.05 (see section 2.4). Figures 7 to 10 show the traffic density of the selected voyages from origin to destination area, for the four time periods. The graph parameters are:

- Grid cell side: 0.05 nautical miles;
- AIS ship types: 70-79 (cargo ships);
- Ship length:  $\geq 100$  m.

Figure 11 shows the paths obtained for the four time periods, using only voyage selected AIS messages received from ships with selected types and length, as well as the grid, origin and destination areas.

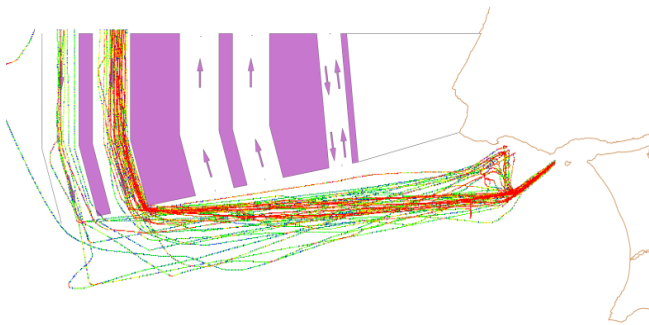


Figure 7. Traffic density, May 2012, selected voyages

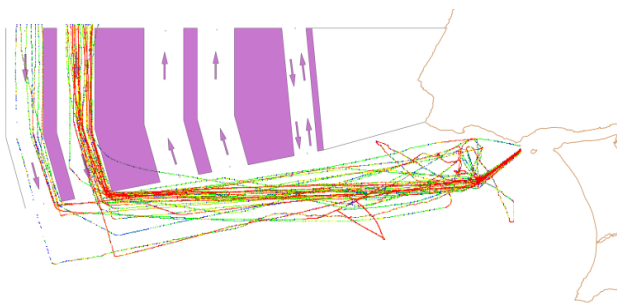


Figure 8. Traffic density, August 2012, selected voyages

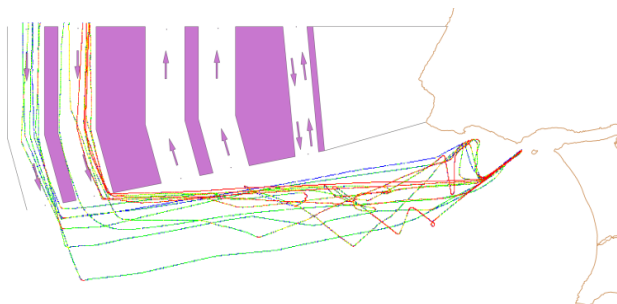


Figure 9. Traffic density, November 2012, selected voyages

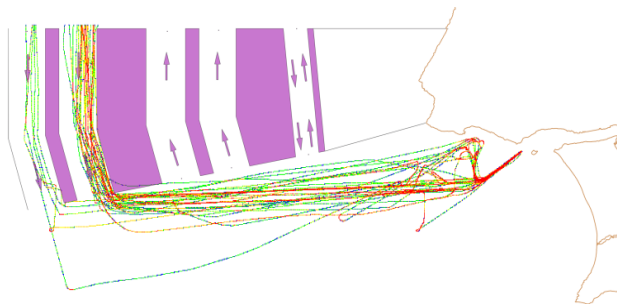


Figure 10. Traffic density, February 2013, selected voyages

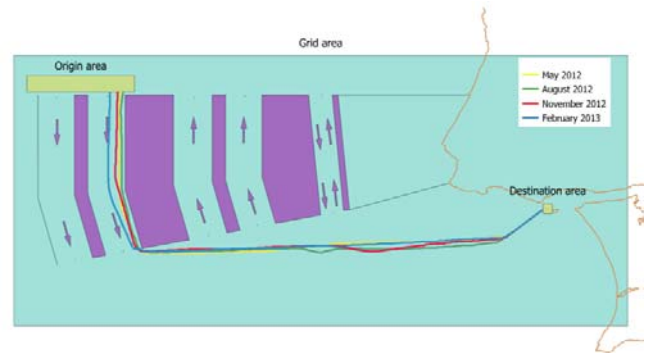


Figure 11. Paths obtained from selected AIS messages, received from ships bound to Lisbon

#### 4 APPLICATION TO IN-PORT ROUTES

The proposed approach is now applied to define a safe path in the port of Setubal, from the fairway entry to the RO-RO terminal. The entire available dataset is used, from May 2012 to June 2013. Only messages from ships that travelled between the origin and destination areas in under 300 minutes (equivalent to a minimum average speed of about 0.9 knots) are considered, once again to exclude paths taken by ships that anchor inside the port of Setubal before proceeding to the terminal. The adopted graph parameters are:

- Grid cell side: 0.01 nautical miles;
- AIS ship type: 70-79 (cargo ships);
- Ship length: greater or equal to 100 meters.

Figure 12 shows the traffic density as well as the origin and destination areas. Figure 13 shows the path obtained for these parameters.

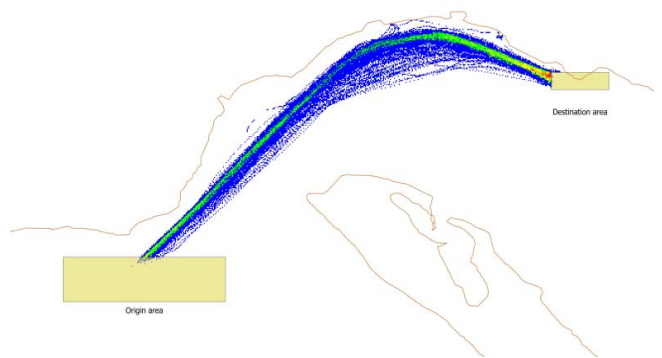


Figure 12. Port of Setubal: traffic density, origin and destination areas

As mentioned before, the proposed method allows the determination of the speed profile for the path. Figure 14 shows the speed profile for the path in Figure 13 and the smoothed profile using a Savitzky-Golay filter (Savitzky & Golay, 1964) with order = 1 and frame length = 51.

Path points 20 and 40 are shown in Figure 13 for reference. It is interesting to notice the slight increase in speed around path point 20, right before the turn is executed.

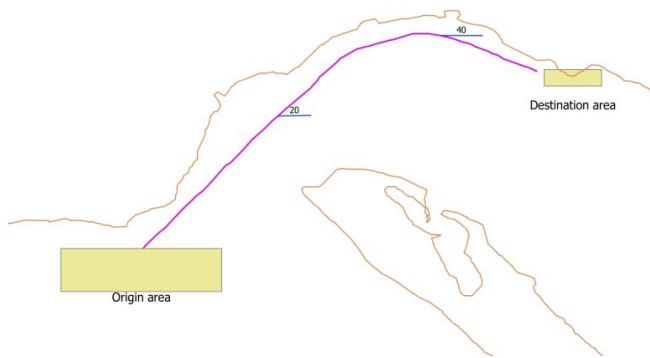


Figure 13. Port of Setúbal: path, origin and destination areas

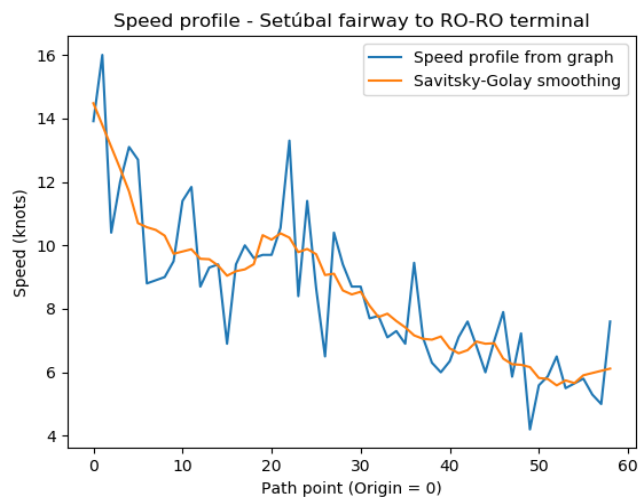


Figure 14. Speed profile and smoothed profile, Setúbal fairway to RO-RO terminal

## 5 CONCLUSIONS

The present paper has contributed with an approach for identifying typical routes of ships from historical AIS data. In general, after being certified as reliable by authorities and/or experts, these routes derived from AIS data have the potential of becoming the reference for ship's crews, pilots and Vessel Traffic Services (VTS), both in the planning stage and during the monitoring stage of a voyage. These paths may also be used as a reference in accident investigations and may become an enabler for future autonomous vessel operations.

The proposed method proved to be very effective in identifying the most used path between selected origin and destination areas, based on historical AIS data, given particular ship types and dimensions. The path obtained may be considered safe because it takes into account the predominant movements of vessels with the chosen types and dimensions during a certain time period.

The quality of the results is dependent on the choice of the grid cell size and on the amount of AIS messages available. These two factors are related with each other, to the point that the method can return the path of one particular vessel if the grid cell size is too small for the amount of AIS messages available, thus

distorting the overall objective. Therefore, these parameters must be carefully chosen. The method is also useful to obtain the speed profile of a particular path.

When implementing the method, some unexpected difficulties arose due to errors in AIS data. Possibly because of the fact that the AIS messages were received by several coastal stations and then combined in one file, some messages appeared to be out of sync. In some cases, when plotting consecutive positions of the same ship, it was possible to observe the ship's position alternating back and forth, thus introducing false transitions between grid cells when building the graph. To eliminate this error it was necessary to filter out the out of sync messages. This potential problem must be taken into account when applying the method to data collected by multiple stations.

## ACKNOWLEDGEMENTS

This work has been co-funded by the European Regional Development Fund (Fundo Europeu de Desenvolvimento Regional (FEDER) and by the Portuguese Foundation for Science and Technology (Fundação para a Ciência e a Tecnologia - FCT) under project "Integrated System for Traffic Monitoring and Maritime Risk Assessment (MoniRisk)", n.º 028746.

The authors would like to thank the Portuguese coastal Vessel Traffic Service (VTS) Control Centre (CCTMC) for providing the AIS data used in this study.

## REFERENCES

- Cormen, T., Leiserson, C., Rivest, R., & Clifford, S. (2009). *Introduction to Algorithms* (3rd ed.). The MIT Press.
- Dijkstra, E. W. (1959). A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, 1(1), 269–271. <http://doi.org/10.1007/BF01386390>
- Etienne, L., Devogele, T., & Bouju, A. (2010). Spatio-Temporal Trajectory Analysis of Mobile Objects Following the Same Itinerary. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(II), 86–91. <http://doi.org/10.1007/s11263-012-0594-8>
- IMO. (20178). *IMO - Ships' routing* (2017 Edition). International Maritime Organization.
- ITU. (2014). Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile frequency band, M.1371-5. International Telecommunication Union.
- Joo, S.Y., Cho, T.J., Cha, J.M., Yang, J.H., Kwon, Y.K. (2012). An economic ship routing system based on a minimal dynamic-cost path search algorithm. *KIPS Trans. Comput. Commun. Syst.* 1 (2), 79–86.
- Mannarini G., Coppini G., Oddo P., Pinardi N. (2013). A Prototype of Ship Routing Decision Support System for an Operational Oceanographic Service. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, Vol. 7, No. 1, doi:10.12716/1001.07.01.06, pp. 53-59.
- Neumann T. (2016). Vessels Route Planning Problem with Uncertain Data. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, Vol. 10, No. 3, pp. 459-464, doi:10.12716/1001.10.03.11.
- Pallotta, G., Vespe, M., & Bryan, K. (2013). Vessel pattern knowledge discovery from AIS data: A framework for

- anomaly detection and route prediction. *Entropy*, 15(6), 2218–2245. <http://doi.org/10.3390/e15062218>.
- Rong, H., Teixeira, A.P. and Guedes Soares, C., (2015a). Simulation and analysis of maritime traffic in the Tagus River Estuary using AIS data, *Maritime Technology and Engineering*, Guedes Soares, C. & Santos T.A. (Eds.), Taylor & Francis Group, London, UK, pp. 185-194.
- Rong, H., Teixeira, A.P. and Guedes Soares, C. (2015b). Evaluation of near-collisions in the Tagus River Estuary using a marine traffic simulation model, *Scientific Journals of the Maritime University of Szczecin*, 43 (115), pp. 68–78, (ISSN 1733-8670).
- Rong, H., Teixeira, A.P. and Guedes Soares, C. (2016). Assessment and characterization of near ship collision scenarios off the coast of Portugal. *Maritime Technology and Engineering 3*. Guedes Soares & Santos (Eds.), Taylor & Francis Group, London, 871–878.
- Rong, H., Teixeira, A.P. and Guedes Soares, C. (2018). A model for predicting ship destination routes based on AIS data, *Maritime Transportation and Harvesting of Sea Resources*, Guedes Soares & Teixeira (Eds.), Taylor & Francis Group, London, pp. 257-264.
- Savitzky, A., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8), 1627–1639.
- Silveira, P. Teixeira, A.P. and Guedes Soares, C. (2012). Analysis of maritime traffic off the coast of Portugal, *Maritime Engineering and Technology*, C. Guedes Soares, Y. Garbatov S. Sutulo T. A. Santos, (Eds.), pp. 35-41, Taylor & Francis Group.
- Silveira, P., Teixeira, A., & Guedes Soares, C. (2013). Use of AIS Data to Characterise Marine Traffic Patterns and Ship Collision Risk off the Coast of Portugal. *The Journal of Navigation*, 66, 879–898. <http://doi.org/10.1017/S0373463313000519>.
- Silveira, P., Teixeira, A.P. and Guedes Soares, C. (2015). Assessment of ship collision estimation methods using AIS data, *Maritime Technology and Engineering*, Guedes Soares, C. & Santos T.A. (Eds.), Taylor & Francis Group, London, UK, pp. 195-204.
- Takashima K., Mezaoui B., Shoji R. (2009). On the Fuel Saving Operation for Coastal Merchant Ships using Weather Routing. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, Vol. 3, No. 4, pp. 401-406.
- Vespe, M., Visentini, I., Bryan, K., & Braca, P. (2012). Unsupervised learning of maritime traffic patterns for anomaly detection. *9th IET Data Fusion & Target Tracking Conference (DF&TT 2012): Algorithms & Applications*, 14–14. <http://doi.org/10.1049/cp.2012.0414>.