



**HAL**  
open science

## A geometry-based fuzzy approach for long-term association of vessels to maritime routes

Clément Iphar, Anne-Laure Joussetme

► **To cite this version:**

Clément Iphar, Anne-Laure Joussetme. A geometry-based fuzzy approach for long-term association of vessels to maritime routes. *Ocean Engineering*, 2023, 281, pp.114755. 10.1016/j.oceaneng.2023.114755 . hal-04106500

**HAL Id: hal-04106500**

**<https://hal.univ-brest.fr/hal-04106500>**

Submitted on 25 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# A geometry-based fuzzy approach for long-term association of vessels to maritime routes

Clément Iphar<sup>a,c</sup>, Anne-Laure Joussetme<sup>b,c</sup>

<sup>a</sup> *Université de Brest, UMR 6554 LETG  
Brest, France*

<sup>b</sup> *CS GROUP Research Lab  
La Garde, France*

<sup>c</sup> *NATO STO Centre for Maritime Research and Experimentation  
La Spezia, Italy*

---

## Abstract

Either for recreational or professional reasons, ships travel across the globe generating a network of maritime traffic with routes connecting key areas such as ports, off-shore facilities or fishing areas. Monitoring vessels' position relatively to maritime routes provides crucial information about their destination, and can help reducing the risk of collision. In this paper, we implement a fuzzy logic approach for associating vessels to maritime routes, suitable to variable surveillance contexts and very sparse data. Moreover, the framework is agnostic to the way maritime routes are provided, either reflecting patterns-of-life from statistical models extracted from real data or being hand-crafted by a user. Fuzzy membership functions enable expressing that vessels can belong more or less to route corridors, while they follow only one of the possible routes. The computation of membership scores relies on a precise distance computation involving geometrical properties of Earth, valid for very large route segments. The defuzzification step allows non-specific outputs. Several instantiations with aggregation operators of different semantics are compared on a real dataset of tracklets from the Automatic Identification System, with ground truth labels of routes. The performance is assessed in a quality space along with the three dimensions of *correctness*, *specificity* and *confidence*.

---

\*Corresponding author: C. Iphar, paper available online on May 17, 2023.

©CC-BY-NC-ND 4.0 ; doi: 10.1016/j.oceaneng.2023.114755

*Email addresses:* `clement.iphar@univ-brest.fr` (Clément Iphar),  
`anne-laure.joussetme@csgroup.eu` (Anne-Laure Joussetme)

*Keywords:* Maritime surveillance, Route segments, Fuzzy logic classifier,  
Performance metrics

---

## 1. Introduction

In the maritime domain, the ocean is facing an ever-increasing traffic and pressure from human activities. For instance, energy and goods transportation represents together up to 90% of the global maritime traffic [1] and defines a world-wide network of transportation and distribution connecting ports across the world. Other activities at sea such as fishing, sailing, cruising also generate a network of their own, with maritime routes linking key areas such as ports, off-shore facilities or fishing areas.

As the number of vessels entering and leaving ports increases, navigational issues arise. In particular, big vessels such as tankers or cargo vessels follow freely routes with loose temporal and spatial constraints, while other types such as fishing vessels or passengers follow paths dictated by their respective professional activities, often with a shorter time-scale (*e.g.*, fishing areas may change from one day to another one). In this respect, to prevent collisions or hazardous situations due to lousy spatial constraints at sea but also to anticipate future needs of autonomous shipping [2, 3], policies and reporting systems have been implemented [4].

Monitoring the maritime traffic involves knowing for instance the origin and destination of vessels, the route they follow or plan to follow, the estimated time of arrival (ETA) in ports, as well as possible changes in this respect [5]. In particular, knowing which maritime route a vessel is following can provide an estimation of its origin and destination, but also predict future long-term position or detect abnormal behaviour in transiting. Moreover, the position of vessels relatively to some areas is also an indicator to detect possible illegal activities or behaviours (*e.g.*, illegal fishing) or to prevent hazardous situations such as a vessel engaged in fishing in an area known for frequent traffic.

Before the inception of satellite technologies enabling a tracking of sea-going vessels and the large use of the Internet to communicate information worldwide, maritime traffic used to be largely unknown. The companies knew when their

30 vessels were in ports but in-between ports of call, the precise location of the  
31 vessel was not tracked on a high frequency rate. Port authorities knew incoming  
32 vessels with a short notice and coastal authorities could use coastal radars to  
33 be aware of local traffic. Similarly vessels, when at sea, were aware of their  
34 immediate surroundings and performed a paper chart-based navigation with,  
35 more recently, assistance from electronic sensors and charts.

36 Thanks to the recent development of satellites harvesting vessel signals on  
37 a large scale possibly combined to coastal systems such as radars, and the In-  
38 ternet as a worldwide platform for data sharing, vessels no longer disappear  
39 beyond the horizon line [6]. The Automatic Identification System (AIS) is a  
40 legally-enforced system put in place by the International Maritime Organisa-  
41 tion (IMO). Originally deployed on-board vessels to prevent collision risk, it has  
42 become a source of data on maritime navigation [7] widely used for extracting  
43 patterns of maritime traffic or understanding the maritime situation. Its high  
44 rate of transmission and vast network of receiving antennas allow in particular  
45 estimating the normal maritime traffic as well as identifying the main maritime  
46 areas such as ports, fishing areas, or anchorage areas. As maritime navigation  
47 aims at developing greener technologies [8], AIS data are exploited for a wide  
48 range of applications which make use of the modelling of the maritime traffic at  
49 sea (such as for ship collision avoidance [9]), and its economical, environmental  
50 and societal impacts [10]: exhaust of greenhouse gases from cruise and ferry  
51 operations [11] or commercial shipping [12], and impact on populations [13], as  
52 well as port activity [14] and maritime routing [15].

53 Today, optimisation in maritime routing is based on several criteria depend-  
54 ing on the particular objective targeted [16], while the most frequent ones are  
55 the travel time [17], the estimated time of arrival, carbon emissions [18], fuel  
56 consumption [19], various costs [20, 21], the bathymetric or ice risk [22, 23]  
57 or the sailing time in static sea state [24]. Either considered individually or  
58 combined, these criteria help providing optimised maritime routes.

59 The notion of maritime route is vague in itself. Although it indeed denotes  
60 the path to be followed by a vessel between two ports, its data-driven estimation  
61 is an object with fuzzy boundaries. The generation of maritime routes from a  
62 dataset (typically the AIS) is usually based on clustering algorithms [25, 26],

63 providing a synthetic representation of the routes followed by vessels over a past  
64 period, also enabling anomaly detection from those maritime routes [27]. In a  
65 similar fashion than highways link cities in a road network, data analytics draws  
66 a network of routes between the main maritime areas of interest [28], some of  
67 the ports being hubs, *i.e.* major nodes of the network [29]. The main difference  
68 with highways or railroads is that the maritime network is neither closed nor  
69 constrained to physical paths, but open (although subject to bathymetric, legal  
70 and geopolitical restrictions), making the maritime routes estimated from data  
71 fuzzy objects.

72 Moreover, the user's information needs about the route followed by a vessel  
73 depend on the current mission context, the geographical area, the type of vessel,  
74 or other subjective considerations. And such subjectivity inevitably appear in  
75 the basic task of associating a vessel to an established maritime route. We are  
76 thus looking at a solution which would (1) be robust to the lack of data, (2)  
77 flexible to contextual users' needs and (3) capture the vagueness related to the  
78 notion of maritime route. We propose in this paper we framework addressing  
79 these three requirements. It is based on a fuzzy logic approach integrated as a  
80 decision support tool.

81 In this work, we propose an association of vessel to route method based on  
82 fuzzy logic, which is agnostic to the way the maritime route is defined (data cloud  
83 of AIS contacts, manually drawn). The approach relies on a precise computation  
84 of spatial distance in position and course over ground which accounts for long  
85 range travels.

86 The paper is organised as follows. We first provide some background in  
87 Section 2, where we present basic geometric concepts pertaining to the definition  
88 of maritime trajectories with long distances, basics on fuzzy set theory and the  
89 AIS dataset used in this study. In Section 3, we further detail the geometry  
90 of maritime routes and provide detailed computation of the distance between  
91 a vessel and a maritime route along the two features of position and course  
92 over ground. We present in Section 4 a semantic framework for vessel to route  
93 association, where membership degrees of vessel to routes are computed through  
94 several aggregation operators. The approach is illustrated in Section 5 on a  
95 real AIS dataset, enriched with ground truth labeled tracklets, while results

96 are presented along the three quality dimensions of correctness, precision and  
97 confidence. Finally, conclusions and future work are discussed in Section 6.

## 98 2. Background

99 To optimise their route, ship captains can rely on dedicated pieces of software  
100 such as Seaware<sup>1</sup>, SPOS<sup>2</sup> (Ship Performance Optimization System) or BonVoy-  
101 age System<sup>3</sup>, amongst others. Those pieces of software take into consideration  
102 elements such as the weather forecast, winds and waves effects and other vessels  
103 of the fleets in order to save fuel, save time, assist captains and deck crew, and  
104 as a whole enhance the safety of travel. The maritime routes followed by the  
105 vessels are then drawn accordingly.

106 To compute the optimal route, these software thus use some definition of  
107 the object of “maritime route”. However, this object has not a clear definition  
108 and the interpretation of a maritime route varies across domains. While it  
109 could be defined by mariners through a sequence of waypoints for the prescribed  
110 trajectory between an origin and a destination, the data analytics field rather  
111 defines maritime routes through areas were vessels have frequently traveled in  
112 the past.

113 In this work, we will specifically address the study of the position of vessels  
114 relatively to maritime routes. This thus includes estimating “which route is  
115 followed by the vessel”. Similarly, such a vessel to route association would be  
116 efficient enough to discriminate between vessel located within a maritime route  
117 corridor and following it, so having as next port of call the destination port of  
118 the route, and a vessel physically located on the route extent, but for any reason  
119 not following it. For instance, this happens when a vessel crosses a maritime  
120 route with any angle, or when a vessel conducts another activity on the maritime  
121 route, such as fishing. This semantic discrimination is of paramount importance  
122 to properly assess maritime traffic and behaviours.

---

<sup>1</sup><http://www.amiwx.com/seawareenroute.html>

<sup>2</sup><https://www.dtn.com/weather/shipping/spos/>

<sup>3</sup><https://www.stormgeo.com/products/s-suite/s-planner/bonvoyage-system-bvs/>

## 123 2.1. Geometrical considerations for maritime routes

124 We first introduce some approximations relatively to the geometry of Earth,  
125 and then provide some background on computation of distances between points  
126 and trajectories.

### 127 2.1.1. Approximations

128 In order to fully define tracks and trajectories in  $\mathbb{R}^3$  in a mathematical represen-  
129 tation of the Earth such as the WGS84 ellipsoid that most positioning systems  
130 use, a series of parameters has to be determined. Each point  $A_i$ ,  $i = 1, \dots, n$ ,  
131 defining either a trajectory or a data point in tracks has actually three coor-  
132 dinates, a longitude  $\lambda_i$ , a latitude  $\varphi_i$  and a height  $h_i$ . The latitude and the  
133 height are directly linked to the ellipsoid itself defined by parameters  $a$  and  $b$ ,  
134 with  $a \geq b > 0$ , denoting the lengths of the semi-major axis and semi-minor  
135 axis respectively. The eccentricity of the generatrix ellipse denoted by  $e$  is then  
136 defined as  $e = \sqrt{1 - \frac{b^2}{a^2}}$ .

Our formalisation involves two approximations: first, we neglect the height parameter as the tidal effects on geodetic distances are negligible on distance computation, as vessels are located on the geoid (and not the ellipsoid); second, it has been shown [30] that the use of a sphere rather than an ellipsoid, for the same set of coordinates, gives variation in distance that do not exceed 0.5% of the said distance, the actual variation being in most cases much smaller. Consequently,  $h$  can be removed from the set of parameters while  $a$  and  $b$  can be replaced by  $\rho$ , the radius of the Earth. Since the local radius of the Earth changes with respect to the local latitude, a mean radius is chosen, as:

$$\rho = \frac{2 \times \rho_{eq} + \rho_{pol}}{3}$$

137 where  $\rho_{eq} = 6378137.0$  m is the equatorial radius and  $\rho_{pol} = 6356752.3$  m is the  
138 polar radius of the Earth<sup>4</sup>.

Discrete trajectories and tracks are thus uniquely defined by the sets of points  $A_i$ s together with  $\rho$  as a parameter. A pair a consecutive points  $(A, B)$

---

<sup>4</sup>For applications that are area-specific, it is possible to set the radius value to the best-fitting local Earth radius value, for marginal gain in precision of distance computation.

defines itself a segment that can be used as a dual representation:

$$T_\rho = \{A_1, \dots, A_n\} \equiv \{\overline{A_1 A_2}, \dots, \overline{A_{n-1} A_n}\}$$

139 The  $n - 1$  segments are represented by the smallest great circle arc between  
 140 pairs of consecutive points, on the sphere of radius  $\rho$ . This relies on the reason-  
 141 able assumption that two consecutive points are always linked by the smallest of  
 142 the great circle arcs (*i.e.*, no segment can be longer than an half-circle length).  
 143 With such a set of parameters and assumptions, trajectories are then uniquely  
 144 and minimally defined in space.

### 145 2.1.2. Distance from vessel to maritime trajectories

**The Haversine distance** When it comes to computing the distance between two points on the Earth, the Haversine distance is mostly used in the literature, although some instances of the use of Euclidean distances [31, 32] can be found. Given the point  $A$  of coordinates  $(\lambda_A, \varphi_A)$  and the point  $B$  of coordinates  $(\lambda_B, \varphi_B)$ , the Haversine distance on a sphere of radius  $\rho$  is denoted  $H(A, B)$  and defined as:

$$H(A, B) = 2\rho \arcsin \left( \left( \sin^2 \left( \frac{\varphi_B - \varphi_A}{2} \right) + \cos(\varphi_B) \cos(\varphi_A) \sin^2 \left( \frac{\lambda_B - \lambda_A}{2} \right) \right)^{\frac{1}{2}} \right) \quad (2.1)$$

146 The Haversine distance appears to have nice properties: it allows a better  
 147 consideration of the curvature of the Earth [33] in the round-Earth approxima-  
 148 tion than Euclidean distances, it is simple to compute, it allows reliable long-  
 149 distance computations without deviating that much from reality, than would  
 150 require the more complex and time-consuming ellipsoid distance.

151 The difference in distance between a great circle and a great ellipse arc  
 152 reaches a maximum of 0.5% [30]. Therefore, the consistent use of a great circle  
 153 has a negligible effect on the measure of the distance itself and will be used in  
 154 this work.

155 **Distances to and between trajectories** In the literature, distance compu-  
 156 tation between two trajectories mostly consider trajectories as a set of points,  
 157 either under the form of a cluster, or under the form of a series of points (that



158 we defined as a track, when the points originate from sensor data). This re-  
159 duces the distance between two trajectories to the distance between two sets of  
160 points, and the Hausdorff (or similar) distance [34] is thus often applied [35],  
161 sometimes modified [36] for a measure of dissimilarity [37], as a matching de-  
162 gree of trajectories between radar and satellite imagery [38], for the mining of  
163 clusters [39]. Other distances such as cost distance [40], or similarity measures  
164 Fréchet or discrete Fréchet [41] are also used.

165 Since tracks represent the movement of mobile objects, time considerations  
166 should be included when computing distances between several moving objects  
167 [42], with interpolation between points for approximating missing values.

168 We propose in this paper a geometrical approach that not only takes the  
169 points as elements of computation, but includes the lines that link those points.  
170 Computed as a centreline [43], defining the central behaviour of the vessels  
171 that follow the route of interest, the distance between the vessel and the route  
172 of interest is defined as the distance to the great arc circle that interpolates  
173 intermediate waypoints along the route.

174 When the closest point of segment lies in the segment between two vertices  
175 of the route prototype, it is necessary to compute the value of the local bearing  
176 (equivalent to the local course over ground of a vessel that would be located  
177 in this point and heading to the next vertex). Indeed, the local bearing of a  
178 segment evolves as the mobile travels along it, from its initial vertex until the  
179 final vertex. This is exemplified by Figure 1 which shows a (long) segment  
180 from Brest, France (48.4 N, 4.5 W) to New York City, USA (40.6 N, 74.0 W).  
181 Although the final point is South of the original point, the vessel takes an initial  
182 bearing of 288 (West-NorthWest), which goes slightly towards the North, and  
183 the northernmost point of the trajectory occurs at circa 30% of the total distance  
184 of the direct route (geodesic line on the sphere).

185 More particularly, our approach allows the computation over long distances,  
186 avoiding the inaccuracies of planimetric approaches, and is resilient to data  
187 blackout areas (in which vessels can be very close to the route but far away from  
188 any data points). Actually, the synthetic track can be reduced up to two points  
189 (the origin and the destination), and such route prototype can be handled in  
190 the same framework as a route prototype computed from data points, regardless

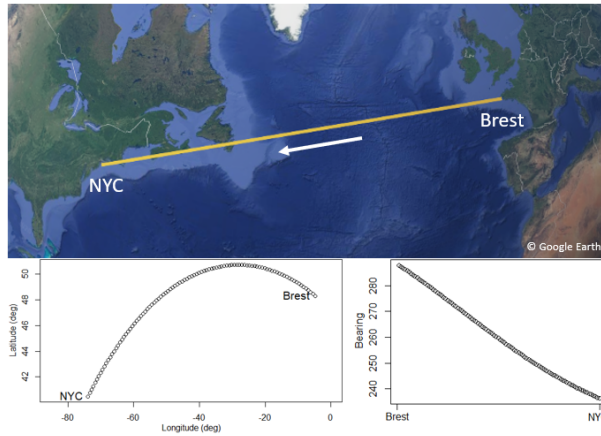


Figure 1: Illustration of the variation of the bearing along one segment. Top: geodesic line linking Brest and New York City (great circle arc). Bottom left: representation of the segment on a traditional planimetric map. Bottom right: evolution of the value of the local bearing of a mobile along the segment from Brest (start, left) to New York City (arrival, right). Angles are in degrees from the North, therefore 270 is full West.

191 of the number and spatial distribution thereof. Details of the computation on  
 192 those route prototypes are presented hereafter in Section 3.

## 193 2.2. Fuzzy logic

194 Fuzzy set theory was developed by Lotfi Zadeh to reason with linguistic vari-  
 195 ables, describing vague information such as human language descriptions (“*small*,  
 196 *large*, *quick*, *young*”), and concepts which cannot be defined by an interval with  
 197 strict limits, such as “*the vessel is small*” or “*the vessel has a quite low speed*”  
 198 [44, 45]. Fuzzy set theory provides a mathematical setting for reasoning with  
 199 subjective concepts themselves represented by membership functions [46].

200 Let  $\mathcal{X}$  be a set of possible values, named the universe of discourse. A fuzzy  
 201 set  $\mu_A$  is a set with imprecise (not well defined) boundaries, *i.e.*, *fuzzy boundaries*  
 202 and extends the concept of classical (crisp) set. In classical set theory, a subset  
 203  $A \subseteq \mathcal{X}$  is said to be crisp and is represented by a characteristic function  $\mu_A$   
 204 such that  $\mu_A : \mathcal{X} \rightarrow \{0; 1\}$  with  $\mu_A(x) = 1$  if  $x \in A$  and  $\mu_A(x) = 0$  if  $x \notin A$ . A  
 205 fuzzy set is defined by a membership function  $\mu_A(x)$  which is a generalisation  
 206 of the characteristic function and can take its values in the interval  $[0, 1]$ , when  
 207 normalised. A normalised fuzzy set  $\mu_A$  over  $\mathcal{X}$  can be represented by a set of

208 ordered pairs:

$$\mu_A = \{(x, \mu_A(x)) \mid x \in \mathcal{X}\} \quad (2.2)$$

209 with  $\mu_A(x) \in [0; 1]$  being the degree of membership of  $x$  to the fuzzy set  $\mu_A$ .  
 210 The complement of a fuzzy set is defined classically by  $\mu_{\bar{A}}(x) = 1 - \mu_A(x)$ ,  $\forall x \in$   
 211  $\mathcal{X}$ .

212 Fuzzy sets are combined using  $t$ -norms (*i.e.*, triangular norms) and  $t$ -conorms  
 213 operators, acting as conjunctive and disjunctive operators respectively [46, 47].  
 214 In Section 4.1, we will compare classical  $t$ -norms and  $t$ -conorms, as the minimum  
 215 (resp. maximum), the product, and Łukasiewicz that we will denote by  $t_M$  (resp.  
 216  $s_M$ ),  $t_P$  and  $s_P$ ,  $t_L$  and  $s_L$ , respectively.

217 Compared to probabilities which define degrees of belief regarding the oc-  
 218 currence (or truth) of an event, being itself either true or false, fuzzy sets define  
 219 degrees of truth for events which are thus allowed to be more or less true. Fuzzy  
 220 sets capture thus the notion of *graduality* [48] and will be used in the following  
 221 to define fuzzy concepts related to the association of vessels to maritime routes  
 222 such as “*vessel close to a maritime corridor*” or “*vessel travelling in the direc-*  
 223 *tion of the maritime corridor*” (Section 2.2). Fuzzy concepts allow to express  
 224 that vessels can be more or less close to a route, or travel more or less in the  
 225 direction of the route. This semantics contrasts with a probabilistic approach  
 226 which expresses that vessels are probably on the route or are probably travelling  
 227 in the direction of the route. As such, it does not rely on prior statistics.

Mamdani inference rule allows to combine two fuzzy concepts to define a new fuzzy one as:

$$\text{if } x_1 \text{ is } \mu_A \text{ and } x_2 \text{ is } \mu_B \text{ then } y \text{ is } \mu_C \quad (2.3)$$

where  $\mu_A$ ,  $\mu_B$  and  $\mu_C$  are three fuzzy sets defined on different universal sets.  
 $\mu_A$  and  $\mu_B$  are the antecedent fuzzy sets while  $\mu_C$  is the consequent fuzzy set.  
 Such kind of rules can be used to define fuzzy rule-based classifiers, where each  
 class  $k$  is associated to one or more fuzzy rules ( $\Gamma_k$ ) of the type:

$$(\Gamma_k) : \text{if } x_1 \text{ is } \mu_A \text{ and } x_2 \text{ is } \mu_B \text{ then } y \text{ is of class } C_k \text{ with confidence } c_k \quad (2.4)$$

Several approaches exist to derive the confidence (or certainty) factors. A single  
 winner rule decision would lead to select the class which has the maximum

weighted firing strength:

$$C_k = \arg \max_k (t(\mu_A(x_1), \mu_B(x_2))w_k) \quad (2.5)$$

228 where  $t$  is a  $t$ -norm, and  $w_k$  is the weight assigned to rule  $(\Gamma_k)$ .

229 In the following, we will propose another method for selecting classes which  
230 allows more expressiveness and less drastic decision. While fuzzy logic has  
231 been used in maritime applications as discussed in Section 1, to the best of our  
232 knowledge, it has never been used to estimate the route followed by a vessel.

### 233 3. Geometry of maritime routes

234 The global distance from a vessel to a maritime route can be defined as an  
235 aggregation of individual distances along different features. Although many  
236 features can be considered to characterise the vessel position relatively to mar-  
237 itime routes [49], we will focus in this work on the only features of position,  
238 with the two coordinates of latitude and longitude, and the course over ground.

239 We detail below the individual contribution of the whereabouts on the one  
240 hand (Section 3.1) and the course over ground on the other hand (Section 3.2).  
241 These distance values will further serve as a basis for the global computation of  
242 the membership of a vessel to a route, in Section 4.

$$R_i = \{A_1, \dots, A_n\}$$

243 where  $A_j(\lambda_j, \varphi_j)$  is a point of the synthetic route. Points are either synthetic  
244 or real waypoints. Two consecutive points  $(A_i, A_{i+1})$  define a segment.

#### 245 3.1. Contribution of the whereabouts

246 The problem of the computation of the distance between a point and a trajectory  
247 is equivalent to the problem of the computation of the distance between a point  
248 and a segment of this trajectory. For maritime trajectories, a segment is the  
249 great circle arc linking two consecutive synthetic waypoints, as explained in the  
250 description of the route prototype in Section 2.1. The distance of a vessel to a  
251 route is then defined as the smallest distance to any single segment of the route  
252 prototype.

253 For the computation of the distance of a point  $X$  of coordinates  $(\lambda_X, \varphi_X)$   
 254 to the smallest great circle arc linking  $A$   $(\lambda_A, \varphi_A)$  and  $B$   $(\lambda_B, \varphi_B)$ , we consider  
 255 the vectors

$$\vec{A} = \begin{bmatrix} \cos(\varphi_A) \cdot \cos(\lambda_A) \\ \cos(\varphi_A) \cdot \sin(\lambda_A) \\ \sin(\varphi_A) \end{bmatrix}, \vec{B} = \begin{bmatrix} \cos(\varphi_B) \cdot \cos(\lambda_B) \\ \cos(\varphi_B) \cdot \sin(\lambda_B) \\ \sin(\varphi_B) \end{bmatrix}, \vec{X} = \begin{bmatrix} \cos(\varphi_X) \cdot \cos(\lambda_X) \\ \cos(\varphi_X) \cdot \sin(\lambda_X) \\ \sin(\varphi_X) \end{bmatrix} \quad (3.1)$$

256 We have a twofold situation for the point  $X$ : either its closest point on the  
 257 great circle generated by  $\overline{AB}$  falls within the smallest arc between  $A$  and  $B$ .  
 258 This case will be called an *inside* case as the projection of  $X$  on the great circle  
 259 lies between  $A$  and  $B$ ; it is displayed in blue in Figure 2(a). Alternatively, the  
 260 closest point to  $X$  falls within the largest arc between  $A$  and  $B$ . This case will  
 be called an *outside* case and is displayed in red in Figure 2(a).

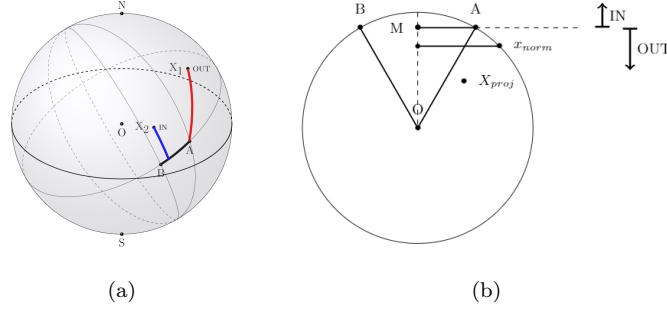


Figure 2: The position of the vessel relative to the trajectory segment drives different computational cases. (a) Distance from a point to the smallest great circle arc. The great circle ( $AB$ ) is divided into two arcs: the smallest one (in bold) is the arc of interest in our study, while the largest one is the arc going around the sphere. (b)  $X_{proj}$  is the projection of the point of interest  $X$  on the great circle plane, itself further projected on the great circle, to decide if the closest point to  $X$  falls inside or outside  $\overline{AB}$

261

To decide to which case (*inside* or *outside*) belongs the point of interest, we consider the plane passing through  $A$ ,  $B$  and  $O$  (*i.e.*, the great circle plane), on which we project the point  $X$ . The projection,

$$\vec{X}_{proj} = \vec{X} - (\vec{n} \cdot \vec{X}) \cdot \vec{n} \quad (3.2)$$

262 is necessarily inside the circle of centre  $O$  and of radius  $\rho$ , because of the proper-  
 263 ties of spheres and great circles. The normalisation enables to bring the point on

264 the great circle as  $\vec{x}_{norm} = \frac{\vec{X}_{proj}}{\|\vec{X}_{proj}\|}$ . We introduce  $\vec{M} = \frac{\vec{A} + \vec{B}}{2}$ . The comparison  
 265 of the scalar (or dot) products of  $(\vec{x}_{norm} \cdot \vec{M})$  and  $(\vec{A} \cdot \vec{M})$  enables to decide.  
 266 Indeed, the only case where  $(\vec{A} \cdot \vec{x}_{norm})$  can be greater than  $(\vec{A} \cdot \vec{M})$  is when  
 267 the normalised projection of  $X$  is between  $A$  and  $B$  on the great circle. This  
 268 situation is explained in Figure 2(b), showing the great circle plan.

In an *inside* case, the distance to the segment is equal to the distance to the closest point of the arc. We first compute a normal vector to the great circle plan with a vector product as  $\vec{n} = \vec{A} \times \vec{B}$  and we measure the minimum distance to the point  $X$  using the properties of the scalar product to compute the central angle and deduct the distance:

$$D_{inside}(X, \overline{AB}) = \rho \cdot \arcsin(\vec{n} \cdot \vec{X}) \quad (3.3)$$

In an *outside* case, the distance of the point to the segment  $\overline{AB}$  is equal to the smallest distance to any of the vertices of the segment (see Figure 2(a)). We define the *orthogonal area* of a segment the area covering all the points for which the closest point to the segment is not one of the two vertices of the segment (*i.e.*, the area for which the projection lies within the smallest great circle arc). The distance will then be defined using the Haversine distance (see Eq. (2.1)):

$$D_{outside}(X, \overline{AB}) = \min\left(H(A, X), H(B, X)\right) \quad (3.4)$$

The distance of the vessel to the entire trajectory  $T$  (*i.e.*, the maritime route prototype) is then given by:

$$D(X, T) = \min_{k \in S} D_k(X, k) \quad (3.5)$$

269 where  $S$  is the set of segments that build the trajectory  $T$  and  $D_k$  is either  
 270  $D_{inside}$  or  $D_{outside}$  depending on the position of  $X$  relatively to the segment  $k$   
 271 as detailed above.

272 Now, we will compute the contribution of the course over ground to the  
 273 global membership score, as it depends on the positional feature values.

## 274 3.2. Contribution of the Course Over Ground

275 **Principles** The course over ground provides directional information which is  
 276 precious in the association of a tracklet to a given route, enabling to distinguish

277 between the vessels sailing alongside the route and those sailing either in the  
 278 opposite direction or perpendicularly to that route, while still being on the  
 279 spatial extent of the route (*i.e.*, the route area).

280 As the AIS message provides the *instantaneous* course over ground of the  
 281 vessel, we want to compute the value of the *local* course over ground of the  
 282 trajectory. In our case, the trajectory is a synthetic trajectory composed of  
 283 consecutive waypoints linked by great circle arcs. The trajectory is therefore  
 284 divided into segments, as shown in Figure 3.

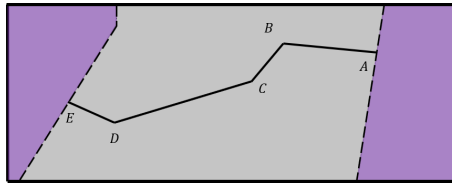


Figure 3: Representation of a synthetic maritime route under the form of a 4-segment route prototype with respect to its environment. In purple, the “outside area”, which is defined as the area “before” the first segment and “after” the last segment, and in grey, the “inside area”.

285 Out of the segments, two of them are of particular interest for the compu-  
 286 tation of the course over ground:

- 287 • The closest segment to the vessel position, while the distance to the seg-  
 288 ments have been computed following the formulas presented in Section  
 289 3.1;
- 290 • The closest segment to the point of interest amongst the (at most two)  
 291 adjacent segments to the first one (which will be called by convention the  
 292 second closest, albeit other non-adjacent segments to the closest could be  
 293 closer than it).

294 In Figure 5, three segments are presented, linking four vertices, named  $A$ ,  $B$ ,  
 295  $C$  and  $D$ , ordered chronologically, and the segment of interest is  $\overline{BC}$ . Depending  
 296 on the situation, the second closest segment will then be either  $\overline{AB}$  or  $\overline{CD}$ .

297 Those segments are shown in Figure 4 in three different scenarios, in which  
 298 the closest segment is coloured in green while the second segment is coloured in  
 299 orange. The bottom-right corner of Figure 4 shows a peculiar situation in which  
 300 the point  $x_3$  is not in the orthogonal area of both segments of interest. In this

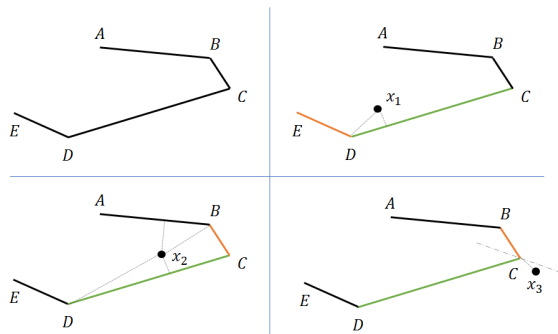


Figure 4: Closest (green) and second closest (orange) segments of a route prototype with respect to a point of interest. Top-left: a 4-segment prototype, from  $A$  to  $E$ . Top-right: classical case. Bottom-left:  $\overline{BC}$  is the second segment, because adjacent to  $\overline{CD}$  (the closest), although  $\overline{AB}$  is closer. Bottom-right:  $\overline{CD}$  and  $\overline{BC}$  are at the same distance

301 case, its distance to both segments  $\overline{BC}$  and  $\overline{CD}$  is identical (equals the distance  
 302 between  $C$  and  $x_3$ ), and the said “closest” and “second closest” segments are  
 303 determined by the side of the bisector straight line of the angle  $B\hat{C}D$  (figured  
 304 as a dotted line) on which the point of interest lies.

305 Figure 5 also shows the orthogonal areas for each of the three segments. The  
 306 position of the point of interest with respect to those areas will be of paramount  
 307 importance for the computation of the local course over ground.

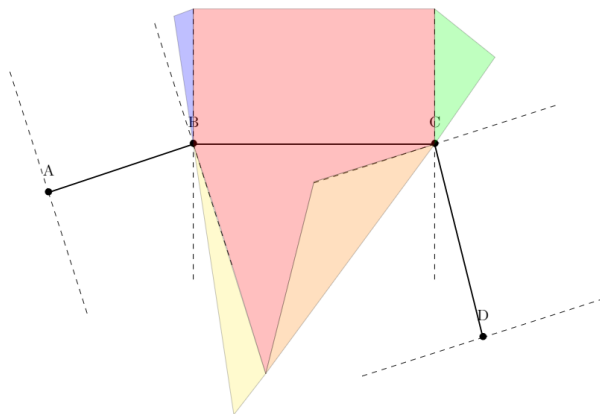


Figure 5: Different areas taken into consideration for the COG computation

308 We distinguish between five cases of point location, which share three com-  
 309 putational cases:

- 310 • The point is in the orthogonal area of the closest segment and not in the



311 orthogonal area of the second closest (red area) [Computational case 1]

312 • The point is in both orthogonal areas of the closest and the second closest  
313 segments, and the second closest segment is before the closest (yellow area)  
314 [Computational case 2]

315 • The point is in both orthogonal areas of the closest and the second closest  
316 segments, and the second closest segment is after the closest (orange area)  
317 [Computational case 2]

318 • The point is not in the orthogonal area of the closest segment, and the  
319 second closest segment is before (blue area) [Computational case 3]

320 • The point is not in the orthogonal area of the closest segment, and the  
321 second closest segment is after (green area) [Computational case 3]

322 On top of those cases that apply if the point is in the grey area of Figure  
323 3, another case applies if the point is in the purple area of Figure 3, which is  
324 presented in the computational case 0.

325 **Computational case 0** This case, represented by the purple area of the  
326 Figure 3, has two subcases: when the point is before the first segment or when  
327 the point is after the last segment.

328 In the first case, the retained value for the local course over ground of the  
329 prototype is the initial bearing of the first segment. In the second case, the  
330 retained value for the local course over ground is the final bearing of the last  
331 segment.

332 **Computational case 1** The computational case number 1, represented by  
333 the red area of the Figure 5 is pretty straightforward. The sole closest seg-  
334 ment is used, and the heading of the segment at the closest point to  $X$  is  
335 considered as being the local course over ground value of the trajectory. In-  
336 deed, the heading generally evolves along a great circle arc (*cf.* Section 2.1.2),  
337 therefore, given  $B$  the initial point of the segment,  $C$  the final point of the  
338 segment and  $K$  the closest point of the  $\overline{BC}$  segment to the point  $X$ , we define  
339  $\vec{K} = \rho \cdot \frac{\vec{X} - (\vec{X} \cdot (\vec{B} \times \vec{C})) \cdot (\vec{B} \times \vec{C})}{\|\vec{X} - (\vec{X} \cdot (\vec{B} \times \vec{C})) \cdot (\vec{B} \times \vec{C})\|}$ . Given that  $\arctan 2(y, x) = \arg(x + iy)$ , where  
340  $i^2 = -1$ , we compute the latitude, longitude and heading of the trajectory at

341 the point  $K$ , reduced to the computation of the initial heading  $G_{K \rightarrow C}$  of the  
 342  $\overline{KC}$  segment, as:

$$\lambda_k = \arctan 2(K_y, K_x) \quad (3.6)$$

$$\varphi_K = \arctan 2\left(K_z, \sqrt{K_x^2 + K_y^2}\right) \quad (3.7)$$

$$\begin{aligned} G_X = G_{K \rightarrow C} &= \arctan 2(\cos \varphi_C \cdot \sin(\lambda_C - \lambda_K), \\ &\cos \varphi_K \cdot \sin \varphi_C - \sin \varphi_K \cdot \cos \varphi_C \cdot \cos(\lambda_C - \lambda_K)) \end{aligned} \quad (3.8)$$

343 **Computational case 2** The computational case number 2, represented by  
 344 the yellow and orange areas in the Figure 5 is less obvious in its computation.  
 345 Indeed, not only the closest segment is relevant to the determination of the  
 346 local heading, but also the second closest segment, which can be the one before  
 347 (yellow area of Figure 5) or the one after (orange area of Figure 5).

348 As in this case, both projections are located on the two great circles along  
 349 the two considered segments, it is possible to compute a heading for both, using  
 350 the same formulas as in computational case 1. Let  $K_1$  be the closest point to  $X$   
 351 in the closest segment  $\overline{BC}$ , and let  $K_2$  be the closest point to  $X$  in the second  
 352 closest segment, so either  $\overline{AB}$  or  $\overline{CD}$ . Let us denote the headings in  $K_1$  and  $K_2$   
 353 as  $G_1 = G_{K_1 \rightarrow C}$ , and either  $G_2 = G_{K_2 \rightarrow B}$  or  $G_2 = G_{K_2 \rightarrow D}$ , respectively.

Because the difference in distance between  $\overline{KK_1}$  and  $\overline{KK_2}$  can be high, a  
 weighted mean of the heading values is used, giving more weight to the heading  
 of the closest segment. We thus define  $p = \lfloor \frac{\overline{KK_1}}{\overline{KK_2}} \rfloor$  the integer representing the  
 difference between the values. If  $\overline{KK_2}$  does not exceed twice the value of  $\overline{KK_1}$ ,  
 a simple circular mean value is computed, as  $G_X = \Delta_c(G_1, G_2)$ . However, if  
 $\overline{KK_2} > 2 \cdot \overline{KK_1}$ , the circular mean is computed recursively as:

$$G_X = G(p) = \Delta_c(G_1, G(p-1)), \text{ with } G(0) = G_2 \quad (3.9)$$

354 **Computational case 3** The computational case number 3, represented by  
 355 the blue and green areas in Figure 5, needs also two separate heading computa-  
 356 tions. Indeed, as being located outside the orthogonal areas of both the closest  
 357 and the second closest segment, computational cases 1 and 2 do not apply.

358 Whether the second closest segment is located before (blue case) or after  
359 (green case) the closest one does not change the computation method. As in  
360 this case  $X$  is in-between two segments, we call those two segments the *former*  
361 *segment* (which is the second closest in the blue case and the closest in the green  
362 case) and the *next segment* (which is the closest in the blue case and the second  
363 closest in the green case) respectively. Two headings are computed: the initial  
364 heading of the next segment and the final heading of the former segment<sup>5</sup>.

365 Therefore, considering the final heading of the former segment as  $G_1$  and  
366 the initial heading of the next segment as  $G_2$ , we compute, according to the  
367 right-hand side of the equation (3.8) and the schema presented in Figure 5:

$$368 \quad G_1 = \begin{cases} G_{B \rightarrow A} + \pi & , \text{ if blue case} \\ G_{C \rightarrow B} + \pi & , \text{ if green case} \end{cases} \quad G_2 = \begin{cases} G_{B \rightarrow C} & , \text{ if blue case} \\ G_{C \rightarrow D} & , \text{ if green case} \end{cases}$$

369 Then the estimated local heading of point  $X$  is computed as the circular  
370 mean of the two angles. Note that no weight is allocated as the distances to the  
371 closest and the second closest segment are in fact, by definition, equal, but the  
372 point is allocated to the segment in its quadrant.

$$G_X = \Delta_c(G_1, G_2) \quad (3.10)$$

373 At this point, and whatever the computational case in which the point of  
374 interest falls, a course over ground  $G_X$  has been computed, its value depicting  
375 the course over ground of the maritime route prototype at its closest point with  
376 the point of interest.

377 In this section, we computed the elements that allow us, from our consid-  
378 eration of the geometry of maritime routes, to get the information needed to  
379 compute the contribution of various features in the vessel to route association  
380 process. As we focus on the whereabouts and the course over ground, their  
381 contributions were presented in this Section, and will be used in the following  
382 Section for membership score computation, aggregation and route association.

---

<sup>5</sup>The final heading of a segment is the initial heading of the same segment considered backwards, to which is added  $\pi$  rad.

## 383 4. Semantic framework for vessel to maritime route 384 association

385 Whereas the semantics of maritime route can be various [50], it appears that  
386 maritime routes have two different meanings, one related to the “itinerary” a  
387 vessel is planning to follow, one related to their synthetic representation which  
388 itself defines a surface over the sea (*i.e.*, the route area or corridor). In this  
389 respect, we can consider the synthetic maritime route as a fuzzy set, to which  
390 vessels belong more or less. This allows to distinguish between the events “being  
391 in a route corridor” (*e.g.*, a vessel fishing on the area where other vessels generally  
392 traveled while transiting between two ports) and “traveling in the direction of  
393 a route” (*e.g.*, a vessel transiting between two ports with a trajectory more or  
394 less parallel to a given maritime route). These two events are mainly defined  
395 and discriminated by the two features of position (latitude and longitude) and  
396 course over ground.

397 In this section, we present the computation of the membership scores to a  
398 route  $R$  of a vessel represented by its local features  $\phi = (\lambda, \varphi, \theta)$  collected as  
399 data points. The local features are estimated based on the computation details  
400 provided in Section 3. Generally, a series of  $n \in \mathbb{N}^*$  consecutive data points  
401 is used, averaged through an arithmetic mean for the positional values of the  
402 whereabouts, and a circular angular mean for the course of ground. In the  
403 experimentation reported in Section 5,  $n$  was set to 5. The general framework  
404 is displayed in Figure 6 which elements will be detailed in the next sections.

### 405 4.1. A fuzzy logic approach to route association

406 For a given route  $R$ , the combination of the distance of the vessel to the route  
407 and its direction relatively to the route will together provide information if the  
408 vessel follows that route. We define the event “The vessel follows route  $R$ ” as the  
409 conjunction the two fuzzy events “The vessel is on route  $R$ ” and “The vessel is  
410 in the direction of route  $R$ ”. We will first define the fuzzy membership functions  
411 corresponding to the two sets of fuzzy events along the respective features of  
412 whereabouts and course over ground relatively to  $R$  (Section 4.1.1). We will

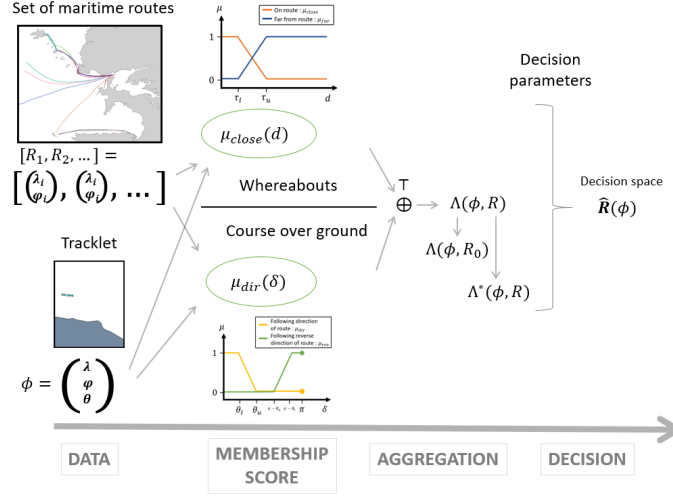


Figure 6: General framework for vessel-to-route association

413 then define in Section 4.1.2 the fuzzy rules allowing to infer if the vessel is  
 414 following route  $R$  or if it is “off-route” (*i.e.*, not following any route). Finally, in  
 415 Section 4.2 we will define two decision schemes which provide the set of routes  
 416 possibly followed by the vessel.

#### 417 4.1.1. Fuzzy membership functions

418 The translation of distances along the whereabouts and course over ground  
 419 respectively, into membership scores is done through the definition of fuzzy  
 420 membership functions which express fuzzy events such as “Vessel close to a  
 421 route” or “Vessel in the same direction of the route”. The fuzzy sets thus defined  
 422 allow expressing the fuzziness of the maritime route, allowing vessels to be “more  
 423 or less on a route”, to be “more or less in the same direction than the route”, or  
 424 to “follow more or less a route”.

**Score from the whereabouts** From the geodesic distance  $d$  between the synthetic route and the vessel track we derive a membership score (hereafter referred as the whereabouts score) ranging from 0 to 1, where 0 translates into a “the vessel is extremely far from the route” and 1 score translates into “the vessel is on the route” (or maximally close to the route), from the exclusive point of view of the whereabouts. A score between 0 and 1 translates into a

partial membership. While many decreasing functions can adequately represent the membership degree of a vessel to a route according to its distance  $d$  to the route prototype, we will use here a trapezoidal function  $\mu_{\text{close}}^{(w)}(d) : \mathbb{R}^+ \rightarrow [0, 1]$  defined as:

$$\mu_{\text{close}}^{(w)}(d) = \begin{cases} 1 & \text{if } d \leq \tau_l \\ 0 & \text{if } d \geq \tau_u \\ \frac{d - \tau_u}{\tau_l - \tau_u} & \text{otherwise} \end{cases} \quad (4.1)$$

425 where  $\tau_l$  is threshold below which the membership score is maximal,  $\tau_u$  the  
426 threshold above which the membership score is minimal.

427 **Score from the course over ground** Similarly, from the angular difference  
428  $\delta$  between the course over ground of the point and the local course over ground  
429 of the route, we define a membership function where 1 translates into “the vessel  
430 travels in a parallel direction to the route” and 0 translates into “the vessel travels  
431 in the opposite direction to the route”. A score between 0 and 1 translates into  
432 a partial agreement between the two directions.

We distinguish between two angles: the angle with respect to the local North reported by the vessel, denoted by  $\hat{C}_X$  and the computed angle with respect to the local North of the route, projected on the point  $X$ , denoted by  $G_X$ . Since  $\hat{C}_X \in [0, 2\pi[$  and  $G_X \in [0, 2\pi[$ , let us note their angular difference  $\hat{D}_0 \in ]-\pi, \pi]$  mod( $2\pi$ ) as  $\hat{D}_0 = \hat{C}_X - G_X$ . Since we are not interested in the sign of the circular angular difference, in the rest of this paper we will use  $\hat{D} \in [0, \pi]$  defined as  $\hat{D} = |\hat{D}_0|$ . For a data contact  $\phi = (\lambda, \varphi, \theta)$ , and a point  $X = (\lambda, \varphi)$ :

$$\delta(\phi, R) = |\theta - G_X|$$

Using also a trapezoidal function, the fuzzy membership function is defined as  $\mu_{\text{dir}}^{(c)}(\delta) : [0, \pi] \rightarrow [0, 1]$ :

$$\mu_{\text{dir}}^{(c)}(\delta) = \begin{cases} 1 & \text{if } \delta < \theta_l \\ 0 & \text{if } \delta > \theta_u \\ \frac{\delta - \theta_u}{\theta_l - \theta_u} & \text{otherwise} \end{cases} \quad (4.2)$$

433 where  $\theta_u \geq \theta_l$ , such that  $\theta_l$  is the threshold below which the membership score  
434 is maximal and  $\theta_u$  the threshold above which the membership score is minimal.

435 **Remark 1.** For each route  $R_i$ , we define a set of membership functions charac-  
436 terising the travel of the vessel relatively to that specific route. The parameters  
437  $\tau_l, \tau_u, \theta_l$  and  $\theta_u$  have to be chosen so they represent the width of the route for  
438 instance, or some subjective tolerance regarding both the distance and angle of  
439 the vessel relatively to that route.

#### 440 4.1.2. Inference and aggregation

441 A vessel is following a route if its position is close enough to the spatial extent  
442 of the route **and** if its direction is similar to the route. The inference is thus  
443 performed through the series of fuzzy rules valid for each of the route  $R_i$ , of the  
444 form:

$$\begin{aligned}
 & \text{If} \quad \text{Vessel is on the area of route } R_i \\
 (\Gamma_i) : \quad & \text{and} \quad \text{Vessel is in the direction of route } R_i \qquad (4.3) \\
 & \text{then} \quad \text{Vessel is following route } R_i
 \end{aligned}$$

Thus, we will combine conjunctively the scores on the individual features of position  $\mu_{\text{close}}^{(w)}$  and course over ground  $\mu_{\text{dir}}^{(c)}$  previously described. Let us denote by  $\Lambda(\phi, R_i)$  the global (or aggregated) membership score of the vessel, represented by its data contact or tracklet  $\phi$  with respect to the maritime route  $R_i$ , itself defined as the route prototype. We have:

$$\Lambda(\phi, R_i) = t\left(\mu_{\text{close}}^{(w)}(d), \mu_{\text{dir}}^{(c)}(\delta)\right) \qquad (4.4)$$

445 where  $t$  is a  $t$ -norm as introduced in Section 2.2. Some comparative behaviour  
446 of these three operators will be presented later in Section 5.

447 The inference being performed relatively to each route, it allows then to  
448 identify the set of routes likely or possibly followed by the vessel, and to detect  
449 vessels likely off-route.

#### 450 4.1.3. Vessels following a route or off-route

A vessel is said to follow a route if it follows at least one of the relevant routes  $R_i, i = 1, \dots, m$ . Hence, this is expressed by the disjunction of the previous fuzzy events and defined as:

$$\Lambda(\phi, R) = s(\Lambda(\phi, R_1), \dots, \Lambda(\phi, R_m)) \qquad (4.5)$$

where  $s(\cdot, \cdot)$  denotes a  $t$ -conorm. If we use the bounded sum (or Łukasiewicz)  $s_L$  as a  $t$ -conorm, we get:

$$\Lambda(\phi, R) = \min \left( \sum_{i=1}^m \Lambda(\phi, R_i), 1 \right) \quad (4.6)$$

or with the maximum  $t$ -conorm  $s_M$ :

$$\Lambda(\phi, R) = \max_{i \in \llbracket 1, m \rrbracket} \Lambda(\phi, R_i) \quad (4.7)$$

The complement event “Vessel following no route” (or “off-route”) is thus defined using the classical negation operator:

$$\Lambda(\phi, R_0) = 1 - \Lambda(\phi, R) \quad (4.8)$$

451  $R_0$  is used as a convention to denote “no route”.

In case the sum of all scores is greater than 1, a normalisation process can be performed in order to get all the normalised scores summing up to 1. In this respect,  $\forall i \in \llbracket 1, m \rrbracket$ , we define

$$\Lambda^*(\phi, R_i) = \frac{\Lambda(\phi, R_i)}{\sum_{j=1}^m (\Lambda(\phi, R_j))} \quad (4.9)$$

452 where  $\Lambda^*(\phi, R_i)$  denotes the normalised value of  $\Lambda(\phi, R_i)$ .

## 453 4.2. Route association

Let us denote by  $\mathcal{R} = \{R_1, \dots, R_m\}$  the set of possible routes and by  $\Omega_R$  the set of labels to be assigned by the fuzzy classifier to the vessel. Depending on the classification problem, we will define:

$$\Omega_R^{(2)} = \{R_0, R\} \text{ and } \Omega_R^{(m)} = \{R_0, R_1, \dots, R_m\}$$

454 We present below two decision methods for route association.

455 The decision methods allow to assign a set of possible classes to the vessel,  
 456 and thus enable more or less specific results. Let us denote by  $\hat{\mathbf{R}}(\phi) \subseteq \Omega_R$   
 457 the corresponding set of routes assigned by the classifier, and  $k$  the number of  
 458 classes associated to  $\phi$  so that  $|\hat{\mathbf{R}}(\phi)| = k$ .



459 **4.2.1. Top-k**

For a number of possible routes  $k$  fixed, the  $k$  routes corresponding to the highest membership scores are retrieved. Let denote by  $\Psi_{\Omega_R}$  the set of classes in decreasing order according to their score and by  $\psi_{R_i}$  the rank of the class  $R_i$  in  $\Psi_{\Omega_R}$ , then  $\forall i, j \in \llbracket 0, m \rrbracket^2, i \neq j$ ,

$$\Lambda(\phi, R_i) \geq \Lambda(\phi, R_j) \implies \psi_{R_i} \leq \psi_{R_j}. \quad (4.10)$$

Then, the set of retrieved routes is:

$$\hat{\mathbf{R}}(\phi) = \{R_i \in \Omega_R \mid \psi_{R_i} \leq k\} \quad (4.11)$$

460 Although in the general case,  $|\hat{\mathbf{R}}(\phi)| = k$ , this may not be true as in particular:

- 461 • in case some scores tie, several routes will have the same rank and thus
- 462  $|\hat{\mathbf{R}}(\phi)| \geq k$ ;
- 463 • routes with 0 scores will be excluded, leading to  $|\hat{\mathbf{R}}(\phi)| \leq k$ .

464 The quality characterisation of the classifier will still allow to capture this  
465 through the specificity measure as described in Section 5.1.

466 The advantage of the Top-k approach is to be able to set the desired number  
467 of output classes, whatever the scores of the classes in the association process.  
468 The drawback though is that in case a single class has a very high score, other  
469 irrelevant classes will still be assigned to  $\phi$ .

470 **4.2.2. Threshold**

In the threshold decision method, we fix instead a threshold value  $\varepsilon_i \in [0; 1]$ , for each  $i = 0, \dots, m$ , and are only kept the classes having a score superior or equal to this threshold. Then, the set of routes possibly followed by the vessel is:

$$\hat{\mathbf{R}}(\phi) = \{R_i \in \Omega_R \mid \Lambda(\phi, R_i) \geq \varepsilon_i\} \quad (4.12)$$

471 The advantage of the threshold method is to be able discard classes with  
472 scores too low to be relevant. Moreover, the thresholds can be set individually  
473 to the different routes and thus somehow capture their geometry. The drawback

474 however is to define properly the thresholds. Although expert knowledge elici-  
475 tation methods can be used, we will rather set in this work arbitrary thresholds  
476 based on our own knowledge of the area.

477 **Remark 2.** It can result in no class being selected, so that  $\hat{\mathbf{R}}(\phi) = \emptyset$  is the  
478 empty set. In this case, a Top-1 method will be applied. The confidence will be  
479 quite low (see Section 5.1).

480 **Remark 3.** Additionally, the two decision methods Top-k and Threshold can  
481 both be refined to consider a set of weights  $\xi_i$  for each route and allow further  
482 consideration independently of the membership score. These weights could be  
483 used for instance to give more importance to some routes than to others, based  
484 on a specific operational request.

## 485 5. Illustration on real data

486 We will illustrate our approach on real AIS data by considering two classification  
487 problems:

- 488 • on the one hand, the association of tracklets to the 18 classes constituted  
489 by the 17 routes and the “off-route” class, so that  $\Omega_R^{(18)} = \{R_0, R_1, \dots, R_{17}\}$   
490 is the set of possible labels;
- 491 • and on the other hand the association of tracklets to the two “off-route”  
492 and “on-route” classes, so that  $\Omega_R^{(2)} = \{R_0, R\}$  is the set of possible labels.

493 We name those problems the “18-class problem” and “2-class problem”, respec-  
494 tively. For the 18-class problem, we use the dataset denoted by  $I^*$  of 400 on-route  
495 tracklets with 17 labels, while for the 2-class problem we use the whole dataset  
496 of 800 tracklets, with only two labels, denoted by  $I$ .

### 497 5.1. Association quality dimensions

498 We will characterise the performances of different instantiations of the vessel to  
499 route association classifier along the three dimensions of *correctness*, *specificity*  
500 and *confidence* as described below. The different instantiations of the classifier  
501 are defined by a set of parameters  $\chi$  referring in particular to the definition of  
502 membership functions, aggregation and the decision methods.

503 For each tracklet  $\phi_j$  of the dataset  $I$ , the set of possible route labels output  
504 by the classifier under the computational set of parameters  $\chi$  is denoted by  
505  $\hat{\mathbf{R}}_\chi(\phi_j) \subseteq \Omega_R$  while the ground truth is denoted by  $R^*(\phi_j) \in \Omega_R$ .

### 506 5.1.1. Correctness

The correctness characterises the ability of the classifier to output the correct route followed by the vessel. It is defined here as the frequency of correct associations over the whole testing dataset. An association is deemed correct if one of the output classes is the ground truth class  $R^*(\phi_j)$ . For a given dataset  $D \in \{I; I^*\}$  and a classifier with parameters  $\chi$ , the correctness measure is defined as:

$$\Gamma(\chi, D) = \frac{1}{|D|} \sum_{j \in D} \Delta(\hat{\mathbf{R}}_\chi(\phi_j)) \quad \text{where} \quad \Delta(A) = \begin{cases} 1 & \text{if } A \ni R^*(\phi_j) \\ 0 & \text{else} \end{cases} \quad (5.1)$$

In the two-class problem, such as for the  $I$  dataset, the correctness measure reduces to the classical accuracy measure:

$$\Gamma(\chi, I) = \frac{TP + TN}{TP + FP + FN + TN} \quad (5.2)$$

507 where  $TP$ ,  $TN$ ,  $FP$  and  $FN$  are elements of the confusion matrix (see Table  
508 1). The correctness measure is thus maximal (and equal to 1) if all the samples

		Real	
		On-route	Off-route
Estimated	On-route	True Positive (TP)	False Positive (FP)
	Off-route	False Negative (FN)	True Negative (TN)

Table 1: Confusion matrix

508  
509  $\phi_j$  of the dataset are assigned a set of classes which contains the true class.

### 510 5.1.2. Specificity

The specificity is related to the number of output classes and is defined relatively to the normalised Hartley measure, averaged over all samples of the dataset:

$$S(\chi, D) = \frac{1}{|D|} \sum_{j \in D} H_1(\hat{\mathbf{R}}_\chi(\phi_j)) \quad \text{where} \quad H_1(A) = 1 - \frac{\log_2(|A|)}{\log_2(|\Omega_R|)} \quad (5.3)$$

511 where  $A \subseteq \Omega_R$ , and  $\Omega_R$  is the universal set.  $H_1(A)$  is maximum and equals to 1  
512 if and only if the classifier outputs a single class, while it is minimum and equals  
513 0 if the classifier outputs all classes of  $\Omega_R$ . In case we have only two classes,  
514 *i.e.*  $|\Omega_R| = 2$  the specificity will always be maximum as a single class will be  
515 output, and this measure is thus irrelevant in this case.

### 516 5.1.3. Confidence

The confidence is defined as the degree of trust that the correct class is within  
the set of output classes. We define it relatively to the scores of output classes,  
as:

$$C(\chi, D) = \frac{1}{|D|} \sum_{j \in D} \gamma(\hat{\mathbf{R}}_\chi(\phi_j)) \quad \text{where} \quad \gamma(A) = \frac{\sum_{r \in A} \Lambda(\phi_j, r)}{\sum_{r \in \Omega_R} \Lambda(\phi_j, r)} \quad (5.4)$$

517 where  $\Lambda(\phi_j, r)$  is the aggregated membership score for route  $r$  assigned to track-  
518 let  $\phi_j$ .

519 Thus, the local scores (for each route) are aggregated with a disjunctive  
520 operator, meaning that one of them is true. The confidence value is linked to  
521 the number of possible routes (*i.e.*, the specificity), as the bigger the set, the  
522 higher the confidence. The confidence value is also linked to the values of each  
523 individual score for the selected routes. It is 1 if all the routes are selected.

## 524 5.2. Vessels travelling on maritime routes

525 We provide herein the results of our fuzzy classifier applied to a dataset of real  
526 data extracted from the AIS, with ground truth route labels (available at [51]),  
527 as detailed in [52]. We first analyse the capability of the classifier to discriminate  
528 between the 17 routes according to different parameters such as the  $t$ -norm or  
529 defuzzification (*i.e.*, decision) method. We then highlight the robustness of the  
530 classifier to discriminate between vessels travelling on overlapping twin-routes  
531 (reverse origin and destination). We finally show that our approach is agnostic  
532 to the way the routes are constructed by appending the set of synthetic routes  
533 from TREAD by two hand-crafted routes for which no data were available.

534 **5.2.1. Route association: 18-class problem**

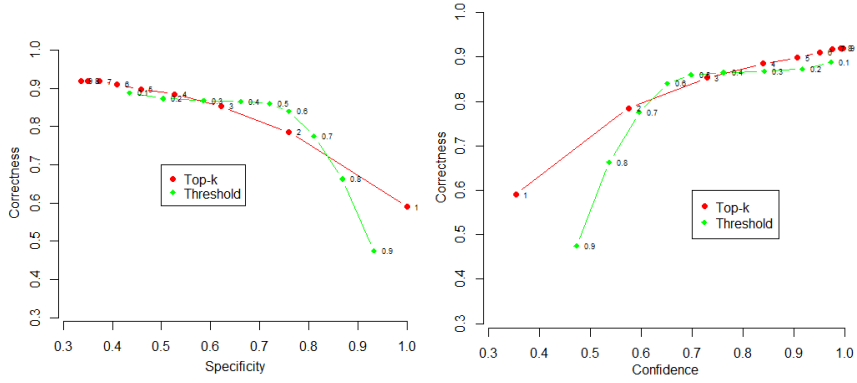
535 We use here the dataset of the 400 “on-route” tracklets (denoted dataset  $I^*$ )  
536 with 17 possible labels corresponding to the 17 routes of  $\mathcal{R}$ . The output of the  
537 classifier will be provided over  $\Omega_R = \{R_0, R_1, \dots, R_{17}\}$ , where a rejection class  
538  $R_0$  is added for “off-route vessels” understood as “vessel not travelling on the 17  
539 routes”.

540 **Comparison of decision methods** We first set the  $t$ -norm as the product,  
541  $t_P$  and observe the comparative results for the two decision methods “Top-k”  
542 and “Threshold” and corresponding parameter values. Figure 7 displays the  
543 results of the association. The three dimensions of correctness, specificity and  
544 confidence are presented pairwise in three distinct charts. Together with dots  
545 of the curves are displayed specific parameter values:  $k$ , for the Top-k method  
546 (red curve) and  $\varepsilon$ , for the Threshold method (green curve). The same threshold  
547 value is used for all the routes.

548 We observe a natural decrease of the correctness as the specificity increases  
549 (Fig. 7(a)), a natural increase of correctness as the confidence increases (Fig.  
550 7(b)) as well as a natural decrease of specificity as the confidence increases  
551 (Fig. 7(c)). An optimum seems to be reached for a threshold around  $\varepsilon = 0.6$   
552 providing a correctness just below 0.9, for a specificity of 0.8 (corresponding to  
553 2 routes output). When a single route is output though (maximum specificity),  
554 the correctness drops significantly. This can be explained by the pairs of twin  
555 routes in our dataset, very close to each other. Further analysis is performed in  
556 Section 5.2.2 below.

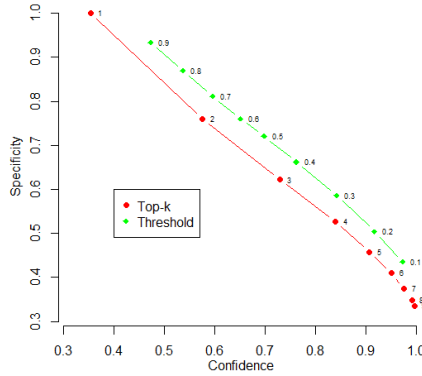
557 The Threshold approach outperforms the Top-k approach within a range  
558 of  $\varepsilon$  values between 0.3 and 0.7. This is confirmed by Figure 7(c), which also  
559 shows that in terms of specificity, a threshold value of 0.6 is worth a Top-2, and  
560 we can approximate a value of 0.35 being worth a Top-3. This means that, on  
561 average across the 400 tracklets, a Threshold approach with a 0.6 value outputs  
562 2 classes. Figure 7(b) shows that there is no clear improvement in correctness  
563 when the threshold value ranges from 0.2 to 0.6.

564 **Remark 4.** We observe a maximum correctness at 0.91, which means that in  
565 about 9% of the cases, some routes truly assigned to vessels are assigned null



(a) Correctness versus specificity

(b) Correctness versus confidence



(c) Specificity versus confidence

Figure 7: Association quality for different decision methods and a product  $t$ -norm as aggregation operator. Parameter values ( $k$  or  $\varepsilon$ ) are displayed on the dots

566 scores by the classifier. This is due to a rough estimation of the membership  
 567 functions which does not capture the actual spatial extent of the routes. Indeed,  
 568 some routes have a low width, while other have a wider width. In the later case,  
 569 a too low  $\tau_l$  leads inevitably to a null membership score.

570 **Remark 5.** If we interpret the relationship between confidence and correctness  
 571 as under- versus over-confidence of the classifier, the median in the graph of Fig.  
 572 7(b) would display an exact confidence. We can thus read these results as the  
 573 classifier being under-confident (or cautious) for threshold values between circa  
 574 0.3 and 0.88, and over-confident for the over values.

575 **Comparison of aggregation methods** We set now the decision method to  
 576 the Threshold method with varying parameter from 0.1 to 0.9, with 0.1 steps.  
 577 Figure 8 displays comparative results using three different  $t$ -norms introduced  
 in Section 2.2 for the aggregation along the two features. Figure 8(a) shows

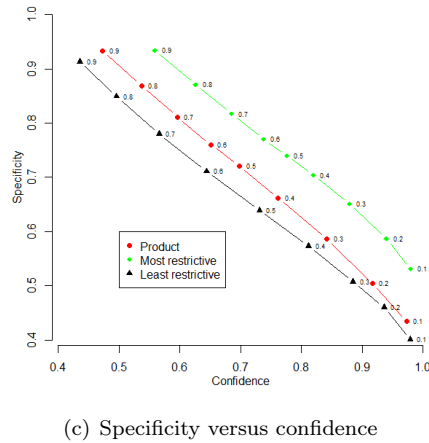
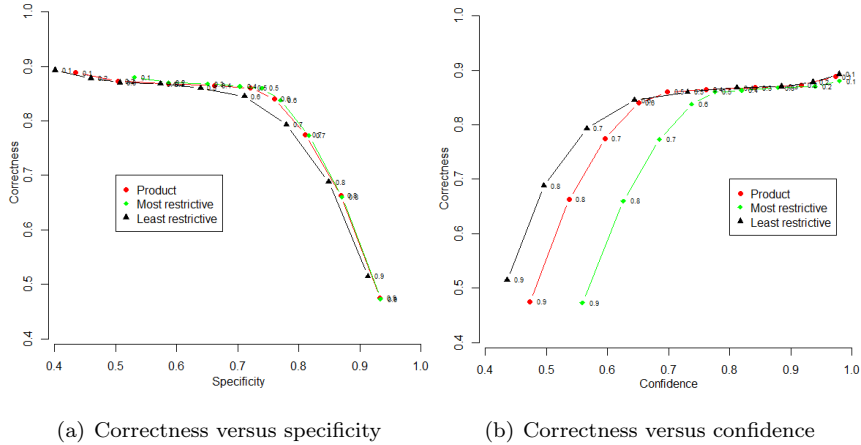


Figure 8: Association quality for different feature aggregation operators, using the Threshold method for decision. Threshold values  $\epsilon$  are displayed on each dot of the graphs

578  
 579 that the optimum at parameter values between 0.6 and 0.7 is consistent across  
 580 the aggregation methods used, while the aggregation method does not impact  
 581 much the results in terms of specificity-correctness. Figure 8(b) shows that  
 582 correctness reaches a plateau for values of the threshold between 0.2 and 0.6,  
 583 although the confidence is higher for the stronger  $t$ -norm. Figure 8(c) shows  
 584 that for both specificity and confidence, the weakest the  $t$ -norm, the better the  
 585 outcome, which is a result that is expected by the very nature of those two

586 quality measures. Best results are thus obtained with  $t_L$ .

### 587 5.2.2. Association to reverse route

588 We consider now the case of “twin-routes”, being pairs of routes having the  
589 same origin and destination ports, but with opposite directions. Six pairs (so  
590 12 routes) fall within this category in our dataset of routes. The twin-routes of  
591 our dataset are  $\{R_{01}, R_{02}\}$ ,  $\{R_{03}, R_{04}\}$ ,  $\{R_{05}, R_{06}\}$ ,  $\{R_{07}, R_{08}\}$ ,  $\{R_{10}, R_{11}\}$  and  
592  $\{R_{12}, R_{13}\}$ .

593 We would like to test the robustness of our classifier to correctly associate  
594 vessels travelling within the corridor of the opposite route. To such an aim, we  
595 modified the dataset reversing the COG of all tracklets on one route of the pair  
596 together with the label. We thus obtain sets of tracklets on the extent of routes,  
597 in the opposite direction. We refer below to this dataset as “reverse”, while the  
598 previous one is referred to as “original”.

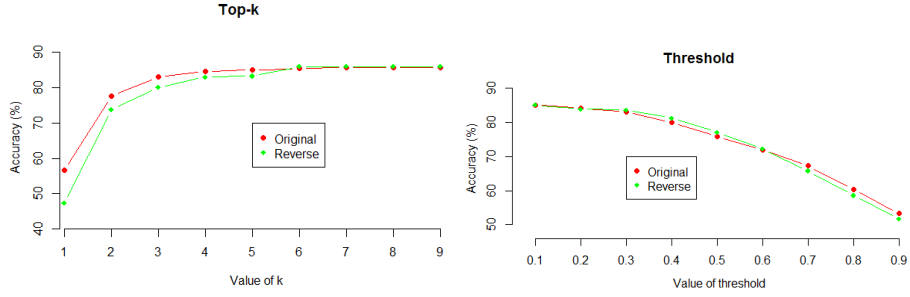
599 The association results are presented in Figure 9, where the correctness is  
600 displayed for varying values of decision parameters for both decision methods.  
601 Top-k in Fig. 9(a) and Threshold in Fig. 9(b).

602 The results show that on both datasets said “original” and “reverse” the  
603 association only shows significant discrepancies ( $> 4\%$ ) occur for the Top-1  
604 technique. Only for this parameter value, the correctness is low in the Top-k  
605 approach ( $< 70\%$ ). In general, the rest of the Top-k approach gives a slight  
606 advantage to the “original” association, although the significance of the  $k > 1$   
607 cases must be further assessed and is possibly route-related. The Threshold  
608 approach though, show very similar associating results regardless the threshold  
609 selected. That means that vessels travelling on the spatial extent of a route  
610 which overlaps with the opposite direction, will still be assigned to the correct  
611 route. These results show the robustness of our approach to correct association  
612 to twin-routes with overlapping spatial extents, as they are best discriminated  
613 by their direction.

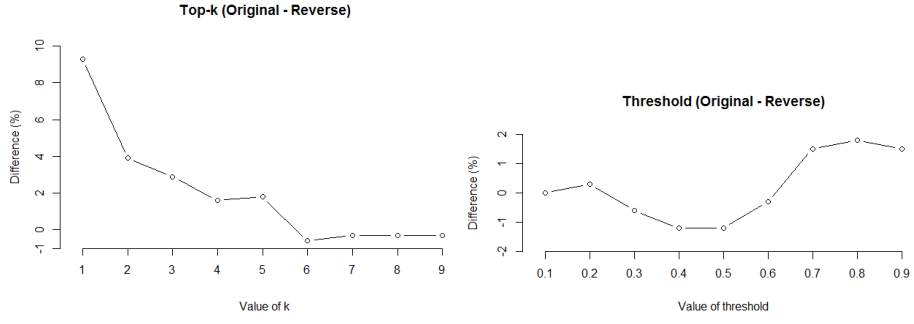
### 614 5.2.3. Hand-crafted maritime route: Brest-Douarnenez example

615 One original feature of the proposed approach is to be agnostic to the represen-  
616 tation of the synthetic route. Here we consider an hand-crafted route, which is





(a) Correctness for both original and reverse directions according to the value of  $k$  in the Top-k approach (b) Correctness of accuracy for both original and reverse direction according to the threshold value



(c) Difference in correctness between original and reverse direction in the Top-k approach (d) Difference in correctness between original and reverse direction in the Threshold approach

Figure 9: Comparative results for Top-k and Threshold approaches in terms of correctness for association with both original maritime route and with their counterpart

617 not the result of some AIS data processing, but rather provided by an operator  
 618 aware of that route.

619 We selected the ports of Brest and Douarnenez, being both already in the set  
 620 of ports of the dataset. We hand-picked 5 virtual waypoints and generated two  
 621 routes, one for each direction. Figure 10 shows the associated route prototypes  
 622 generated by the 5 virtual waypoints, that will serve as basis for computation.

623 We thus append the set of original routes  $\mathcal{R}$  with these two hand-crafted  
 624 routes, so that now  $|\mathcal{R}'| = 19$ . We ran the vessel-to-route association classifier  
 625 over the 400 off-route tracklets, *i.e.* over  $I \setminus I^*$  dataset, using a Top-1 decision  
 626 method. All tracklets assigned to one of the two routes as best score were first  
 627 isolated and their true (*i.e.*, labeled) origin and destination retrieved from the

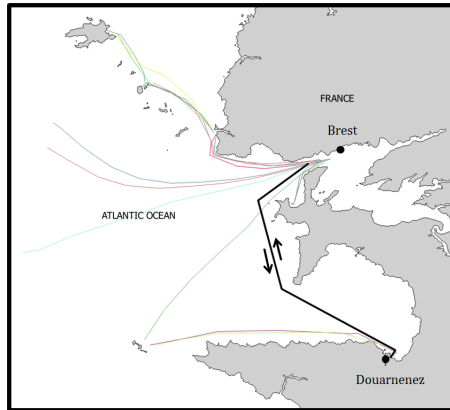


Figure 10: In black, the bidirectional hand-crafted route between Brest and Douarnenez together with the 17 routes of the dataset

628 original labeled AIS dataset. Table 2 shows all the selected tracklets, together  
 629 with their score, true and estimated origins and destinations.

630 This table shows that although the number of vessels navigating between  
 631 Brest and Douarnenez is not large enough to generate a maritime route via  
 632 the TREAD software (and be reflected in our dataset), some vessels still travel  
 633 through this route (for instance tracklets 475 and 650, cf. Table 2).

634 Table 3 shows the confusion matrix for the origins (on the left-hand side) and  
 635 the destinations (on the right-hand side) of the tracklets for which one of the two  
 636 hand-crafted routes were output by the classifier. It is worth noticing that out of  
 637 the 17 tracklets which had Douarnenez as either origin or destination), 14 were  
 638 actually estimated correctly (82%). Instead, Brest was correctly estimated for  
 639 only 4 tracklets. This can be explained by the relative isolation of Douarnenez  
 640 compared to the other maritime routes, while Brest is the nodal point of the  
 641 local network. Also, out of the 17 tracklets, only 2 tracklets were actually corre-  
 642 sponding to Douarnenez to Brest journeys while none of the actual routes that  
 643 the vessels followed were one of the other 17 routes of the dataset. An interest-  
 644 ing result is that one of these two tracklets was the only one to be assigned a  
 645 confidence score of 1, proving the utility of including hand-crafted routes and  
 646 the ability of the classifier to process them together with data-extracted routes.

#	Confidence	Estimated Origin	Estimated Destination	True Origin	True Destination
401	0.847448	Douarnenez	Brest	Brest	Ocean
407	0.605311	Douarnenez	Brest	Douarnenez	Ocean
409	0.716976	Brest	Douarnenez	Ocean	Douarnenez
460	0.941618	Brest	Douarnenez	Ocean	Douarnenez
475	1.000000	Douarnenez	Brest	Douarnenez	Brest
487	0.969345	Brest	Douarnenez	Ocean	Douarnenez
498	0.554150	Douarnenez	Brest	Douarnenez	Ocean
504	0.507506	Douarnenez	Brest	Douarnenez	Unknown
540	0.664743	Brest	Douarnenez	Ocean	Douarnenez
604	0.730807	Douarnenez	Brest	Le Conquet	Camaret
633	0.929956	Douarnenez	Brest	Douarnenez	Ocean
650	0.677828	Douarnenez	Brest	Douarnenez	Brest
684	0.467977	Douarnenez	Brest	Douarnenez	Ocean
694	0.391742	Douarnenez	Brest	Douarnenez	Fishing
731	0.949339	Douarnenez	Brest	Ocean	Brest
770	0.368832	Brest	Douarnenez	Fishing	Douarnenez
772	0.990857	Brest	Douarnenez	Brest	Ocean

Table 2: True and estimated origins and destinations for tracklets assigned to the hand-crafted routes Brest-Douarnenez and Douarnenez-Brest

	Origin			Destination				
	True			True				
	Douarnenez	Brest	else	Douarnenez	Brest	else		
Estimated	Douarnenez	9	0	2	Douarnenez	5	0	1
	Brest	0	1	5	Brest	0	3	8

Table 3: Results for association with both Brest-Douarnenez and Douarnenez-Brest, split by origin and destination ports

## 647 6. Conclusions

648 The work presented in this paper is part of the research in the fields of maritime  
649 transportation and fuzzy logic. We proposed a fuzzy logic approach to vessel-  
650 to-route association, which relies on the very notion of maritime routes and its  
651 fuzzy constructs.

652 We presented a detailed geometrical approach to compute the distance be-  
653 tween a vessel and a maritime route valid when segments of trajectories span  
654 over very large distance. The approach is original with respect to data-centric  
655 methods where statistical motion models help compensating the possible lack  
656 of data. Focused on geometrical features only, our method for distance compu-  
657 tation is agnostic to the way the maritime route is obtained, and is valid for

658 hand-crafted routes.

659 The distance of a vessel to a maritime route is used to define the two fuzzy  
660 concepts of “vessel close to the route corridor”, and “vessel traveling in the same  
661 direction than the maritime route”. Two main features are considered: the  
662 position and the course over ground. It is expected that if the vessel is both  
663 in the vicinity of the spatial extent of a given route and exhibits a course over  
664 ground corresponding to the direction of the route, the vessel is likely to follow  
665 that specific maritime route. As a consequence, the approach is also able to  
666 detect vessels far from some maritime routes of interest, vessels close or within  
667 the route corridor but in reverse or perpendicular direction.

668 Membership scores along the two features of position and course over ground  
669 are combined through a  $t$ -norm (conjunction) to obtain a membership score  
670 of the vessel to a given maritime route. This aggregated score is interpreted  
671 as a likelihood degree that the vessel is actually following that route. The  
672 decision step further allows selecting the subset of routes possibly followed by  
673 the vessel. Two decision methods have been proposed which both enable non-  
674 specific answers as subsets of routes rather than single ones. The “top-k decision”  
675 selects the  $k$  most likely routes, while the “threshold decision” selects only routes  
676 which scores exceeds a given threshold.

677 A series of experiments has been conducted on real data excerpt from the  
678 AIS in the Brest area (France). A dataset of 800 maritime tracklets was pre-  
679 viously labeled with the route the corresponding vessels were actually followed,  
680 providing thus some ground truth and enabling the assessment of the correct-  
681 ness of our method. This ground truthed dataset along with the maritime routes  
682 of interest have been made available in a companion data publication. Those  
683 experiments include the assessment of association to pairwise maritime routes  
684 (*e.g.* corresponding to ferry trips between close ports) and hand-crafted mar-  
685 itime routes in addition to assessment of association to classical data extracted  
686 routes. The main interest in the fuzzy rule-based classifier proposed is its inter-  
687 pretability and flexibility, possibly at the expense of the classification accuracy.  
688 Rather than a unique solution to route association, we defined a framework to  
689 associate vessels to maritime routes which considers users’ needs and knowledge.  
690 The framework allows customising the antecedent fuzzy membership functions

691 according to specific routes geometry or user needs, while optimising some pa-  
692 rameters such as the weights of the fuzzy rules by means of a training data set  
693 with ground truth labels, and achieve a higher accuracy. Typically, the param-  
694 eters for defining the route membership should reflect the specific area under  
695 surveillance as well as the specific route. The aggregation operator ( $t$ -norm)  
696 could reflect more or less optimistic (or pessimistic) approach of the operator,  
697 and can change given the mission context. In future work, we would address  
698 the optimisation of this set of parameters to fit both the data (and capture the  
699 actual patterns of life) and the users' knowledge and information needs.

700 With little improvement, the approach proposed could be used in the scope  
701 of the development of future maritime green routes [53, 54]. Indeed, in order to  
702 navigate in areas with low past traffic where no or few data is available, hand-  
703 crafted routes could be used as precise guides that our computation can make  
704 the vessel follow. This could be beneficial to navigation software applications, in  
705 offering assistance in navigational choices, and helping to reduce the economic  
706 or ecological impacts of voyage paths.

## 707 Acknowledgments

708 The authors wish to thank the NATO Allied Command Transformation (NATO-  
709 ACT) for supporting this work through the funding of the Centre for Maritime  
710 Research and Experimentation.

## 711 References

- 712 [1] C. Iphar, A. Napoli, C. Ray, Integrity assessment of a worldwide maritime  
713 tracking system for a geospatial risk analysis at sea, in: Proceedings of the  
714 20th AGILE International Conference on Geographic Information Science  
715 (AGILE 2017), 2017.
- 716 [2] C. Kooij, M. Loonstijn, R. Hekkenberg, K. Visser, Towards autonomous  
717 shipping: Operational challenges of unmanned short sea cargo vessels, in:  
718 P. Kujala, L. Lu (Eds.), Marine Design XIII, 2018, pp. 871–880.
- 719 [3] J. Montewka, K. Wróbel, E. Heikkilä, O. Valdez-Banda, F. Goerlandt,

- 720 S. Haugen, Challenges, solution proposals and research directions in safety  
721 and risk assessment of autonomous shipping, in: Proceedings of PSAM 14  
722 - Probabilistic Safety Assessment and Management Conference, 2018.
- 723 [4] M. D. Robards, G. Silber, J. Adams, J. Arroyo, D. Lorenzini, K. Schwehr,  
724 J. Amos, Conservation science and policy applications of the marine vessel  
725 Automatic Identification System (AIS), *Bulletin of Marine Science* 92 (1)  
726 (2016) 75–103. doi:10.5343/bms.2015.1034.
- 727 [5] O. Bodunov, F. Schmidt, A. Martin, A. Brito, C. Fetzner, Grand challenge:  
728 Real-time destination and ETA prediction for maritime traffic, in: Proceed-  
729 ings of the DEBS’18 conference, 2018. doi:10.1145/3210284.3220502.
- 730 [6] T. Wahl, G. K. Høy, A. Lyngvi, B. T. Narheim, New possible roles of  
731 small satellites in maritime surveillance, *Acta Astronautica* 56 (1-2) (2005)  
732 273–277. doi:10.1016/j.actaastro.2004.09.025.
- 733 [7] M. Fournier, R. Casey Hilliard, S. Rezaee, R. Pelot, Past, present, and  
734 future of the satellite-based automatic identification system: areas of ap-  
735 plications (2004–2016), *WMU Journal of Maritime Affairs* 17 (3) (2018)  
736 311–345. doi:10.1007/s13437-018-0151-6.
- 737 [8] L. Zhen, Y. Wu, S. Wang, G. Laporte, Green technology adoption for  
738 fleet deployment in a shipping network, *Transportation Research Part B:  
739 Methodological* (2020) 388–410doi:10.1016/j.trb.2020.06.004.
- 740 [9] H. Rong, A. Teixeira, C. Guedes Soares, Ship collision avoidance behaviour  
741 recognition and analysis based on AIS data, *Ocean Engineering* (2022).  
742 doi:10.1016/j.oceaneng.2021.110479.
- 743 [10] V. Prochazka, R. Adland, F.-C. Wolff, Contracting decisions in the crude  
744 oil transportation market: Evidence from fixtures matched with AIS data,  
745 *Transportation Research Part A: Policy and Practice* 130 (2019) 37–53.  
746 doi:10.1016/j.tra.2019.09.009.
- 747 [11] M. Tichavska, B. Tovar, Port-city exhaust emission model: An application  
748 to cruise and ferry operations in las palmas port, *Transportation Research*

- 749 Part A: Policy and Practice 78 (2015) 347–360. doi:10.1016/j.tra.2015.  
750 05.021.
- 751 [12] L. Styhre, H. Winnes, J. Black, J. Lee, H. Le-Griffin, Greenhouse gas emis-  
752 sions from ships in ports – case studies in four continents, *Transporta-*  
753 *tion Research Part D: Transport and Environment* 54 (2017) 212–224.  
754 doi:10.1016/j.trd.2017.04.033.
- 755 [13] D. Toscano, F. Murena, F. Quaranta, L. Mocerino, Assessment of the im-  
756 pact of ship emissions on air quality based on a complete annual emission  
757 inventory using ais data for the port of naples, *Ocean Engineering* (2021).  
758 doi:10.1016/j.oceaneng.2021.109166.
- 759 [14] C. Liu, J. Liu, X. Zhou, Z. Zhao, C. Wan, Z. Liu, AIS data-driven approach  
760 to estimate navigable capacity of busy waterways focusing on ships entering  
761 and leaving port, *Ocean Engineering* (2020). doi:10.1016/j.oceaneng.  
762 2020.108215.
- 763 [15] J. Cai, G. Chen, M. Lutzen, N. Gorm Maly Rytter, A practical AIS-based  
764 route library for voyage planning at the pre-fixture stage, *Ocean Engineer-*  
765 *ing* (2021). doi:10.1016/j.oceaneng.2021.109478.
- 766 [16] T. P. Zis, H. N. Psaraftis, L. Ding, Ship weather routing: A taxonomy  
767 and survey, *Ocean Engineering* (2020). doi:10.1016/j.oceaneng.2020.  
768 107697.
- 769 [17] K. Kepaptsoglou, G. Fountas, M. G. Karlaftis, Weather impact on con-  
770 tainership routing in closed seas: A chance-constraint optimization ap-  
771 proach, *Transportation Research Part C: Emerging Technologies* (2015)  
772 139–155doi:10.1016/j.trc.2015.01.027.
- 773 [18] V. N. Armstrong, Vessel optimisation for low carbon shipping, *Ocean En-*  
774 *gineering* (2013) 195–207doi:10.1016/j.oceaneng.2013.06.018.
- 775 [19] M. Bentin, D. Zastrau, M. Schlaak, D. Freye, R. Elsner, S. Kotzur, A new  
776 routing optimization tool-influence of wind and waves on fuel consumption  
777 of ships with and without wind assisted ship propulsion systems, *Trans-*  
778 *portation Research Procedia* (2016) 153–162doi:10.1016/j.trpro.2016.  
779 05.051.

- 780 [20] V. Windeck, *A Liner Shipping Network Design*, Springer Gabler, Wies-  
781 baden, Germany, 2013.
- 782 [21] M. Grifoll, F. Martinez de Osés, M. Castells, Potential economic benefits of  
783 using a weather ship routing system at short sea shipping, *WMU Journal*  
784 *of Maritime Affairs* (2018) 195–211doi:10.1007/s13437-018-0143-6.
- 785 [22] L. Perera, C. Guedes Soares, Weather routing and safe ship handling in  
786 the future of shipping, *Ocean Engineering* (2017) 684–695doi:10.1016/j.  
787 *oceaneng*.2016.09.007.
- 788 [23] V. Kotovirta, R. Jalonen, L. Axell, K. Riska, R. Berglung, A system for  
789 route optimization in ice-covered waters, *Cold Regions Science and Tech-*  
790 *nology* (2009) 52–62doi:10.1016/j.coldregions.2008.07.003.
- 791 [24] C. P. Padhy, D. Sen, P. K. Bhaskaran, Application of wave model for  
792 weather routing of ships in the north indian ocean, *Natural Hazards* (2008)  
793 373–385doi:10.1007/s11069-007-9126-1.
- 794 [25] G. Pallotta, M. Vespe, K. Bryan, Traffic Route Extraction and Anomaly  
795 Detection from AIS Data, in: *Proceedings of the COST MOVE Workshop*  
796 *on Moving Objects at Sea*, 2013.
- 797 [26] G. Pallotta, M. Vespe, K. Bryan, Vessel Pattern Knowledge Discovery from  
798 AIS Data: A Framework for Anomaly Detection and Route Prediction,  
799 *Entropy* 15 (6) (2013) 2218–2245. doi:10.3390/e15062218.
- 800 [27] H. Rong, A. Teixeira, C. Guedes Soares, Data mining approach to shipping  
801 route characterization and anomaly detection based on ais data, *Ocean*  
802 *Engineering* (2020). doi:10.1016/j.oceaneng.2020.106936.
- 803 [28] P. Coscia, P. Braca, L. Millefiori, F. Palmieri, P. Villet, Multiple ornstein-  
804 uhlenbeck processes for maritime traffic graph representation, *IEEE trans-*  
805 *actions on Aerospace and Electronic Systems* (2018). doi:10.1109/TAES.  
806 2018.2808098.
- 807 [29] Z. Sun, J. Zheng, Finding potential hub locations for liner shipping, *Trans-*  
808 *portation Research Part B: Methodological* (2016) 750–761doi:10.1016/  
809 *j.trb*.2016.03.005.



- 810 [30] W.-K. Tseng, J.-L. Guo, C.-P. Liu, A comparison of great circle, great  
811 ellipse, and geodesic sailing, *Journal of Marine Science and Technology*  
812 21 (3) (2013) 287–299. doi:10.6119/JMST-012-0430-5.
- 813 [31] C. S. Jensen, D. Lin, B. Chin Ooi, Continuous clustering of moving ob-  
814 jects, *IEEE Transactions on Knowledge and Data Engineering* 19 (9) (2007)  
815 1161–1174. doi:10.1109/TKDE.2007.1054.
- 816 [32] F. Boem, F. A. Pellegrino, G. Fenu, T. Parisini, Multi-feature trajectory  
817 clustering using Earth Mover’s Distance, in: *Proceedings of the 2011 IEEE*  
818 *Conference on Automation Science and Engineering*, 2011. doi:10.1109/  
819 CASE.2011.6042423.
- 820 [33] G. Spiliopoulos, K. Chatzikokolakis, D. Zisis, E. Biliri, D. Papaspyros,  
821 G. Tsapelas, S. Mouzakis, Knowledge extraction from maritime spa-  
822 tiotemporal data: An evaluation of clustering algorithms on big data, in:  
823 *Proceedings of the 2017 IEEE International Conference on Big Data (BIG-*  
824 *DATA)*, 2017, pp. 1682–1687. doi:10.1109/BigData.2017.8258106.
- 825 [34] S. Qi, P. Bouros, D. Sacharidis, N. Mamoulis, Efficient point-based trajec-  
826 tory search, in: C. Claramunt, M. Schneider, R. C.-W. Wong, L. Xiong,  
827 W.-K. Loh, C. Shahabi, K.-J. Li (Eds.), *Advances in Spatial and Temporal*  
828 *Databases*, Springer, 2015, pp. 179–196. doi:10.1007/978-3-319-22363-  
829 6\_10.
- 830 [35] Y. Liu, X. Li, W. Hu, Semi-supervised trajectory learning using a multi-  
831 scale key point based trajectory representation, in: *Proceedings of the*  
832 *2010 20th International Conference on Pattern Recognition*, 2010. doi:  
833 10.1109/ICPR.2010.860.
- 834 [36] F. Shao, S. Cai, J. Gu, A modified Hausdorff distance based algorithm  
835 for 2-dimensional spatial trajectory matching, in: *Proceedings of the 5th*  
836 *International Conference on Computer Science & Education*, 2010, pp. 166–  
837 172. doi:10.1109/ICCSE.2010.5593666.
- 838 [37] R. Laxhammer, G. Falkman, Sequential conformal anomaly detection in  
839 trajectories based on Hausdorff distance, in: *Proceedings of the 14th Inter-*  
840 *national Conference on Information Fusion*, 2011.

- 841 [38] J. P. Bustos, F. Donoso, A. Guesalaga, M. Torres, Matching radar and  
842 satellite images for ship trajectory estimation using the Hausdorff distance,  
843 IST Radar Sonar Navigation 1 (1) (2007) 50–58. doi:10.1049/iet-rsn:  
844 20060025.
- 845 [39] B. Guan, L. Liu, J. Chen, Using relative distance and Hausdorff distance  
846 to mine trajectory clusters, TELKOMNIKA 11 (1) (2013) 115–122. doi:  
847 10.11591/telkonnika.v11i1.1877.
- 848 [40] M. Siljander, E. Venäläinen, F. Goerlandt, P. Pellikka, GIS-based cost  
849 distance modelling to support strategic maritime search and rescue plan-  
850 ning: A feasibility study, Applied Geography 57 (2015) 54–70. doi:  
851 10.1016/j.apgeog.2014.12.013.
- 852 [41] C. Zhang, AIS data-driven general vessel destination prediction: A tra-  
853 jectory similarity-based approach, Ph.D. thesis, University of British  
854 Columbia (2019).
- 855 [42] L. Etienne, T. Devogele, A. Bouju, Spatio-temporal trajectory analysis  
856 of mobile objects following the same itinerary, in: In Proceedings of the  
857 Joint International Conference on Theory, Data Handling and Modelling  
858 in GeoSpatial Information Science, 2010, pp. 86–91.
- 859 [43] H. Rong, A. Teixeira, C. Guedes Soares, Ship trajectory uncertainty pre-  
860 diction based on a gaussian process model, Ocean Engineering 182 (2019)  
861 499–511. doi:10.1016/j.oceaneng.2019.04.024.
- 862 [44] L. A. Zadeh, Fuzzy sets, Information and Control 8 (1965) 338–353.
- 863 [45] L. A. Zadeh, Outline of a new approach to the analysis of complex systems  
864 and decision processes, IEEE Transactions on Systems, Man, and Cyber-  
865 netics SMC-3 (1) (1973) 28–44.
- 866 [46] D. Dubois, H. Prade, A review of fuzzy set aggregation connectives, Infor-  
867 mation Sciences 36 (1-2) (1985) 85–121.
- 868 [47] I. Bloch, Information combination operators for data fusion: A compara-  
869 tive review with classification, IEEE Transactions on Systems, Man and  
870 Cybernetics - Part A: Systems and Humans 26 (1) (1996) 52–67.

- 871 [48] D. Dubois, H. Prade, Formal representations of uncertainty, Vol. Decision-  
872 making - Concepts and Methods, ISTE, London, UK & Wiley, Hoboken,  
873 N.J. USA, 2009, Ch. 3, pp. 85–156, invited paper.
- 874 [49] A.-L. Jusselme, G. Pallotta, Dissecting uncertainty handling techniques:  
875 Illustration on maritime anomaly detection, *Journal of Advances in Infor-*  
876 *mation Fusion* 13 (2) (2018) 158–178.
- 877 [50] C. Iphar, M. Zocholl, A.-L. Jusselme, Semantics of maritime routes: Con-  
878 ciliating complementary views, in: *Proceedings of the OCEANS 2021 San*  
879 *Diego Conference, 2021*. doi:10.23919/OCEANS44145.2021.9705934.
- 880 [51] C. Iphar, A.-L. Jusselme, Maritime routes and vessel tracklet dataset for  
881 vessel-to-route association (version 1.0), data set. Licence CC-BY-NC-SA-  
882 4.0. Zenodo (2022). doi:10.5281/zenodo.6402160.
- 883 [52] C. Iphar, A.-L. Jusselme, G. Pallotta, Maritime route and vessel tracklet  
884 dataset for vessel-to-route association, *Data in Brief* (2022). doi:10.1016/  
885 j.dib.2022.108513.
- 886 [53] W. Ma, D. Ma, Y. Ma, J. Zhang, D. Wang, Green maritime: a routing and  
887 speed multi-objective optimization strategy, *Journal of Cleaner Production*  
888 (2021). doi:10.1016/j.jclepro.2021.127179.
- 889 [54] H. Psaraftis, C. Kontovas, *Green Maritime Transportation: Speed and*  
890 *Route Optimization*, Springer, 2016, Ch. 9. doi:10.1007/978-3-319-  
891 17175-3\_9.