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

Clément Iphar, Anne-Laure Jousselme

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

Clément Iphar[a,c], Anne-Laure Jousselme[b,c]

[a] *Université de Brest, UMR 6554 LETG*
*Brest, France*
[b] *CS GROUP Research Lab*
*La Garde, France*
[c] *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.
 doi: 10.1016/j.oceaneng.2023.114755
*Email addresses:* clement.iphar@univ-brest.fr (Clément Iphar), anne-laure.jousselme@csgroup.eu (Anne-Laure Jousselme)

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

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

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

The notion of maritime route is vague in itself. Although it indeed denotes the path to be followed by a vessel between two ports, its data-driven estimation is an object with fuzzy boundaries. The generation of maritime routes from a dataset (typically the AIS) is usually based on clustering algorithms [25, 26], providing a synthetic representation of the routes followed by vessels over a past period, also enabling anomaly detection from those maritime routes [27]. In a similar fashion than highways link cities in a road network, data analytics draws a network of routes between the main maritime areas of interest [28], some of the ports being hubs, *i.e.* major nodes of the network [29]. The main difference with highways or railroads is that the maritime network is neither closed nor constrained to physical paths, but open (although subject to bathymetric, legal and geopolitical restrictions), making the maritime routes estimated from data fuzzy objects.

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

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

The paper is organised as follows. We first provide some background in Section 2, where we present basic geometric concepts pertaining to the definition of maritime trajectories with long distances, basics on fuzzy set theory and the AIS dataset used in this study. In Section 3, we further detail the geometry of maritime routes and provide detailed computation of the distance between a vessel and a maritime route along the two features of position and course over ground. We present in Section 4 a semantic framework for vessel to route association, where membership degrees of vessel to routes are computed through several aggregation operators. The approach is illustrated in Section 5 on a real AIS dataset, enriched with ground truth labeled tracklets, while results are presented along the three quality dimensions of correctness, precision and confidence. Finally, conclusions and future work are discussed in Section 6.

## 2. Background

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

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

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

[1] http://www.amiwx.com/seawareenroute.html
[2] https://www.dtn.com/weather/shipping/spos/
[3] https://www.stormgeo.com/products/s-suite/s-planner/bonvoyage-system-bvs/

### 2.1. Geometrical considerations for maritime routes

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

#### 2.1.1. Approximations

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

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

where $\rho_{eq} = 6378137.0$ m is the equatorial radius and $\rho_{pol} = 6356752.3$ m is the polar radius of the Earth[4].

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 $(A, B)$

[4]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:

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

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

#### 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:

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

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

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

**Distances to and between trajectories** In the literature, distance computation between two trajectories mostly consider trajectories as a set of points, either under the form of a cluster, or under the form of a series of points (that we defined as a track, when the points originate from sensor data). This reduces the distance between two trajectories to the distance between two sets of points, and the Hausdorff (or similar) distance [34] is thus often applied [35], sometimes modified [36] for a measure of dissimilarity [37], as a matching degree of trajectories between radar and satellite imagery [38], for the mining of clusters [39]. Other distances such as cost distance [40], or similarity measures Fréchet or discrete Fréchet [41] are also used.

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

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

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

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

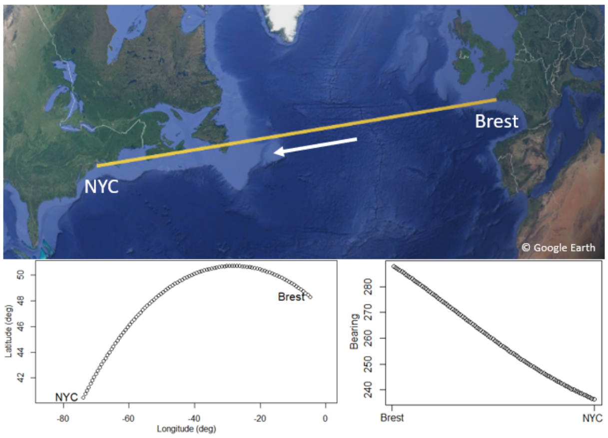



Figure 1: Illustration of the variation of the bearing along one segment. Top: geodesic line linking Brest and New York City (great circle arc). Bottom left: representation of the segment on a traditional planimetric map. Bottom right: evolution of the value of the local bearing of a mobile along the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

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

## 2.2. Fuzzy logic

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

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

$$\mu_A = \{(x, \mu_A(x))\,|x \in \mathcal{X}\} \tag{2.2}$$

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

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

Compared to probabilities which define degrees of belief regarding the occurrence (or truth) of an event, being itself either true or false, fuzzy sets define degrees of truth for events which are thus allowed to be more or less true. Fuzzy sets capture thus the notion of *graduality* [48] and will be used in the following to define fuzzy concepts related to the association of vessels to maritime routes such as "*vessel close to a maritime corridor*" or "*vessel travelling in the direction of the maritime corridor*" (Section 2.2). Fuzzy concepts allow to express that vessels can be more or less close to a route, or travel more or less in the direction of the route. This semantics contrasts with a probabilistic approach which expresses that vessels are probably on the route or are probably travelling 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:

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

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:

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

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:

$$C_k = \arg\max_k \left(t(\mu_A(x_1), \mu_B(x_2))w_k\right) \tag{2.5}$$

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

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

# 3. Geometry of maritime routes

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

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

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

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

## 3.1. Contribution of the whereabouts

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

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

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

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

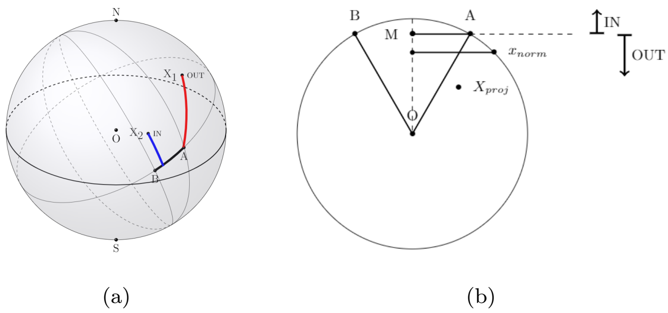



Figure 2: The position of the vessel relatively to the trajectory segment drives different computational cases. (a) Distance from a point to the smallest great circle arc. The great circle $(AB)$ is divided into two arcs: the smallest one (in bold) is the arc of interest in our study, while the largest 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 $\overline{AB}$

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,

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

is necessarily inside the circle of centre $O$ and of radius $\rho$, because of the properties of spheres and great circles. The normalisation enables to bring the point on 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 of the scalar (or dot) products of $(\vec{x}_{norm} \cdot \vec{M})$ and $(\vec{A} \cdot \vec{M})$ enables to decide. Indeed, the only case where $(\vec{A} \cdot \vec{x}_{norm})$ can be greater than $(\vec{A} \cdot \vec{M})$ is when the normalised projection of $X$ is between $A$ and $B$ on the great circle. This 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:

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

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)):

$$D_{outside}(X, \overline{AB}) = \min\Big(H(A, X), H(B, X)\Big) \tag{3.4}$$

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

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

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

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

## 3.2. Contribution of the Course Over Ground

**Principles** The course over ground provides directional information which is precious in the association of a tracklet to a given route, enabling to distinguish between the vessels sailing alongside the route and those sailing either in the opposite direction or perpendicularly to that route, while still being on the spatial extent of the route (*i.e.*, the route area).

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

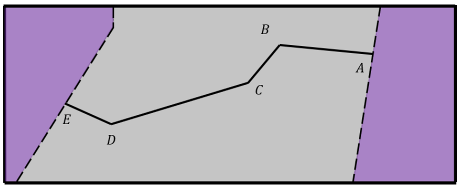



Figure 3: Representation of a synthetic maritime route under the form of a 4-segment route prototype with respect to its environment. In purple, the "outside area", which is defined as the area "before" the first segment and "after" the last segment, and in grey, the "inside area".

Out of the segments, two of them are of particular interest for the computation of the course over ground:

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

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

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

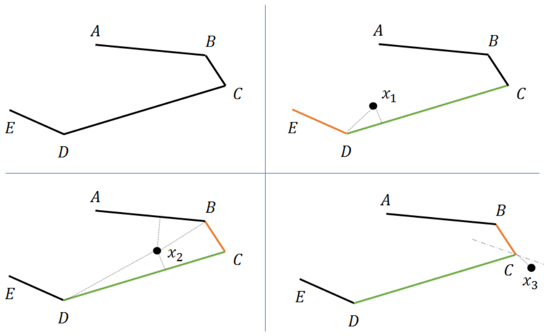



Figure 4: Closest (green) and second closest (orange) segments of a route prototype with respect to a point of interest. Top-left: a 4-segment prototype, from $A$ to $E$. Top-right: classical case. Bottom-left: $\overline{BC}$ is the second segment, because adjacent to $\overline{CD}$ (the closest),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

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

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

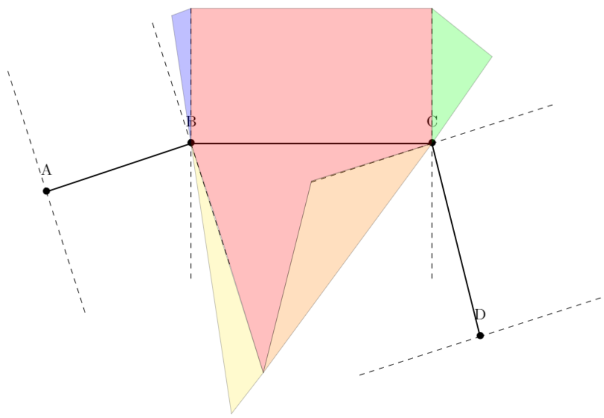



Figure 5: Different areas taken into consideration for the COG computation

We distinguish between five cases of point location, which share three computational cases:

- The point is in the orthogonal area of the closest segment and not in the orthogonal area of the second closest (red area) [Computational case 1]

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

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

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

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

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

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

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

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

$$\lambda_k = \arctan 2(K_y, K_x) \tag{3.6}$$

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

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

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

As in this case, both projections are located on the two great circles along the two considered segments, it is possible to compute a heading for both, using the same formulas as in computational case 1. Let $K_1$ be the closest point to $X$ in the closest segment $\overline{BC}$, and let $K_2$ be the closest point to $X$ in the second closest segment, so either $\overline{AB}$ or $\overline{CD}$. Let us denote the headings in $K_1$ and $K_2$ as $G_1 = G_{K_1\longrightarrow C}$, and either $G_2 = G_{K_2\longrightarrow B}$ or $G_2 = G_{K_2\longrightarrow D}$, 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 $G_X = \underset{c}{\Delta}(G_1, G_2)$. However, if $\overline{KK_2} > 2 \cdot \overline{KK_1}$, the circular mean is computed recursively as:

$$G_X = G(p) = \underset{c}{\Delta}(G_1, G(p-1)), \text{ with } G(0) = G_2 \tag{3.9}$$

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

Whether the second closest segment is located before (blue case) or after (green case) the closest one does not change the computation method. As in this case $X$ is in-between two segments, we call those two segments the *former segment* (which is the second closest in the blue case and the closest in the green case) and the *next segment* (which is the closest in the blue case and the second closest in the green case) respectively. Two headings are computed: the initial heading of the next segment and the final heading of the former segment[5].

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

$$\mathrm{G}_1 = \begin{cases} G_{B \longrightarrow A} + \pi & \text{, if blue case} \\ G_{C \longrightarrow B} + \pi & \text{, if green case} \end{cases} \qquad \mathrm{G}_2 = \begin{cases} G_{B \longrightarrow C} & \text{, if blue case} \\ G_{C \longrightarrow D} & \text{, if green case} \end{cases}$$

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

$$G_X = \underset{c}{\Delta}(G_1, G_2) \tag{3.10}$$

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

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

[5] The final heading of a segment is the initial heading of the same segment considered backwards, to which is added $\pi$ rad.

# 4. Semantic framework for vessel to maritime route association

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

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

## 4.1. A fuzzy logic approach to route association

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

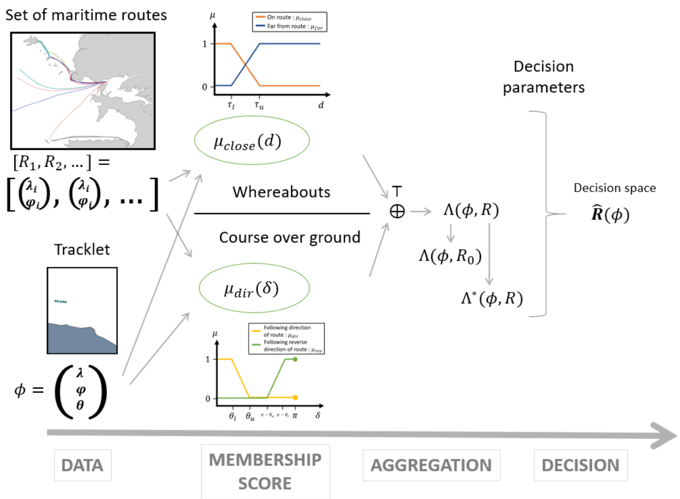



Figure 6: General framework for vessel-to-route association

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

#### 4.1.1. Fuzzy membership functions

The translation of distances along the whereabouts and course over ground respectively, into membership scores is done through the definition of fuzzy membership functions which express fuzzy events such as "Vessel close to a route" or "Vessel in the same direction of the route". The fuzzy sets thus defined allow expressing the fuzziness of the maritime route, allowing vessels to be "more or less on a route", to be "more or less in the same direction than the route", or 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 $\mu^{(w)}_{\text{close}}(d) : \mathbb{R}^+ \longrightarrow [0,1]$ defined as:

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

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

**Score from the course over ground** Similarly, from the angular difference $\delta$ between the course over ground of the point and the local course over ground of the route, we define a membership function where 1 translates into "the vessel travels in a parallel direction to the route" and 0 translates into "the vessel travels in the opposite direction to the route". A score between 0 and 1 translates into 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 $\hat{D} =\mid \hat{D}_0 \mid$. 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 $\mu^{(c)}_{\text{dir}}(\delta) : [0, \pi] \longrightarrow [0,1]$:

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

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

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

### 4.1.2. Inference and aggregation

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

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

Thus, we will combine conjunctively the scores on the individual features of position $\mu^{(w)}_{\text{close}}$ and course over ground $\mu^{(c)}_{\text{dir}}$ 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:

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

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

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

### 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 $R_i$, $i = 1, \ldots, m$. Hence, this is expressed by the disjunction of the previous fuzzy events and defined as:

$$\Lambda(\phi, R) = s\left(\Lambda(\phi, R_1), \ldots, \Lambda(\phi, R_m)\right) \tag{4.5}$$

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

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

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

$$\Lambda(\phi, R) = \max_{i \in [\![1,m]\!]} \Lambda(\phi, R_i) \tag{4.7}$$

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

$$\Lambda(\phi, R_0) = 1 - \Lambda(\phi, R) \tag{4.8}$$

$R_0$ 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, $\forall i \in [\![1, m]\!]$, we define

$$\Lambda^*(\phi, R_i) = \frac{\Lambda(\phi, R_i)}{\sum\limits_{j=1}^{m} (\Lambda(\phi, R_j))} \tag{4.9}$$

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

## 4.2. Route association

Let us denote by $\mathcal{R} = \{R_1, \ldots, R_m\}$ 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:

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

We present below two decision methods for route association.

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

#### 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 $\forall i, j \in [\![0, m]\!]^2, i \neq j$,

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

Then, the set of retrieved routes is:

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

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

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

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

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

#### 4.2.2. Threshold

In the threshold decision method, we fix instead a threshold value $\varepsilon_i \in [0; 1]$, for each $i = 0, \ldots, m$, 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:

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

The advantage of the threshold method is to be able discard classes with scores too low to be relevant. Moreover, the thresholds can be set individually to the different routes and thus somehow capture their geometry. The drawback however is to define properly the thresholds. Although expert knowledge elicitation methods can be used, we will rather set in this work arbitrary thresholds based on our own knowledge of the area.

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

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

# 5. Illustration on real data

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

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

We name those problems the "18-class problem" and "2-class problem", respectively. For the 18-class problem, we use the dataset denoted by $I^*$ of 400 on-route tracklets with 17 labels, while for the 2-class problem we use the whole dataset of 800 tracklets, with only two labels, denoted by $I$.

## 5.1. Association quality dimensions

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

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

#### 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 is 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:

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

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

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

where $TP$, $TN$, $FP$ and $FN$ are elements of the confusion matrix (see Table 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

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

#### 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:

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

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

### 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:

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

where $\Lambda(\phi_j, r)$ is the aggregated membership score for route $r$ assigned to tracklet $\phi_j$.

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

## 5.2. Vessels travelling on maritime routes

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

#### 5.2.1. Route association: 18-class problem

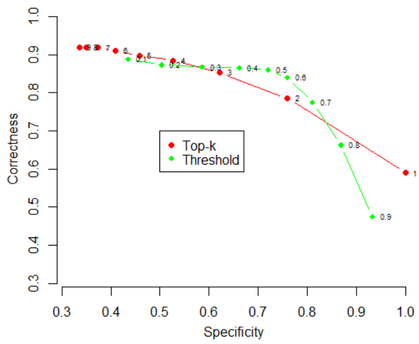
We use here the dataset of the 400 “on-route” tracklets (denoted dataset $I^*$) with 17 possible labels corresponding to the 17 routes of $\mathcal{R}$. The output of the classifier will be provided over $\Omega_R = \{R_0, R_1, \ldots, R_{17}\}$, where a rejection class $R_0$ is added for “off-route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

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

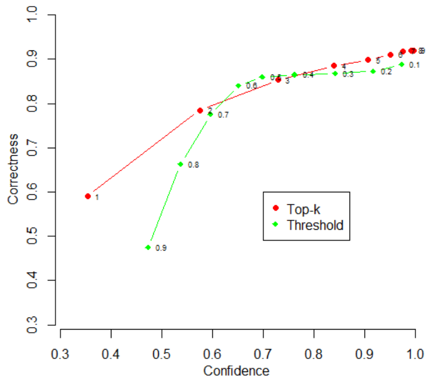
We observe a natural decrease of the correctness as the specificity increases (Fig. 7(a)), a natural increase of correctness as the confidence increases (Fig. 7(b)) as well as a natural decrease of specificity as the confidence increases (Fig. 7(c)). An optimum seems to be reached for a threshold around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

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

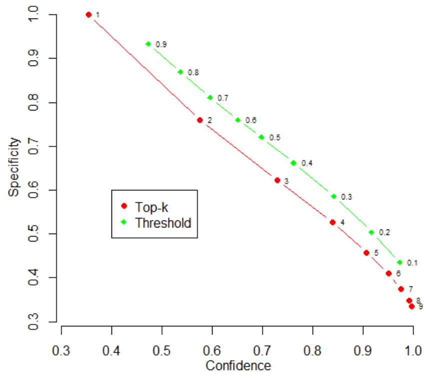
**Remark 4.** We observe a maximum correctness at 0.91, which means that in about 9% of the cases, some routes truly assigned to vessels are assigned null




(a) Correctness versus specificity




(b) Correctness versus confidence



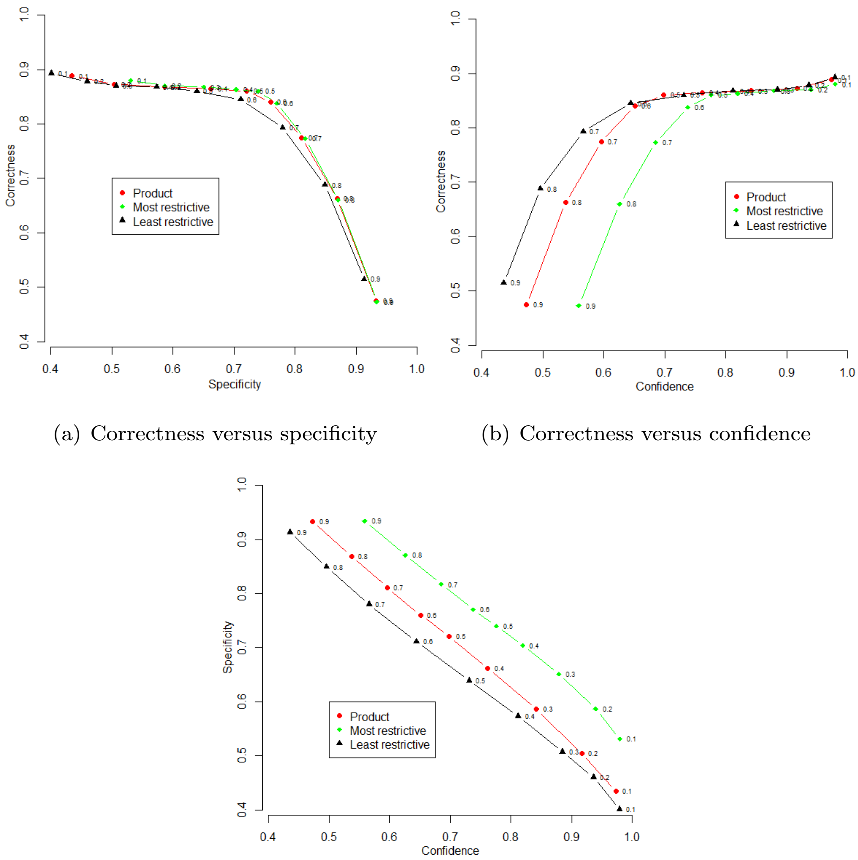

(c) Specificity versus confidence

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

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

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

**Comparison of aggregation methods** We set now the decision method to the Threshold method with varying parameter from 0.1 to 0.9, with 0.1 steps. 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




(a) Correctness versus specificity

(b) Correctness versus confidence

(c) Specificity versus confidence

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

that the optimum at parameter values between 0.6 and 0.7 is consistent across the aggregation methods used, while the aggregation method does not impact much the results in terms of specificity-correctness. Figure 8(b) shows that correctness reaches a plateau for values of the threshold between 0.2 and 0.6, although the confidence is higher for the stronger $t$-norm. Figure 8(c) shows that for both specificity and confidence, the weakest the $t$-norm, the better the outcome, which is a result that is expected by the very nature of those two quality measures. Best results are thus obtained with $t_L$.

### 5.2.2. Association to reverse route

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

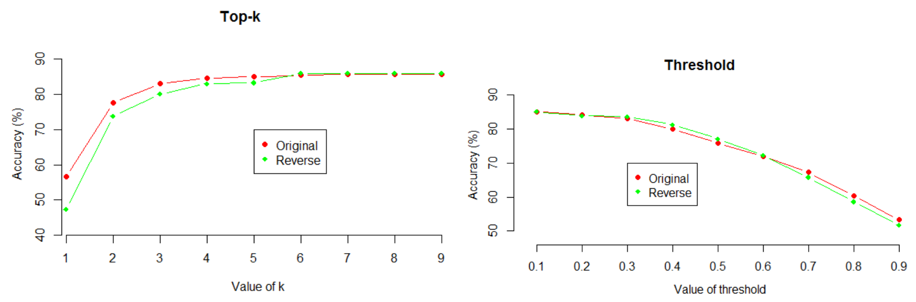
We would like to test the robustness of our classifier to correctly associate vessels travelling within the corridor of the opposite route. To such an aim, we modified the dataset reversing the COG of all tracklets on one route of the pair together with the label. We thus obtain sets of tracklets on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

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

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

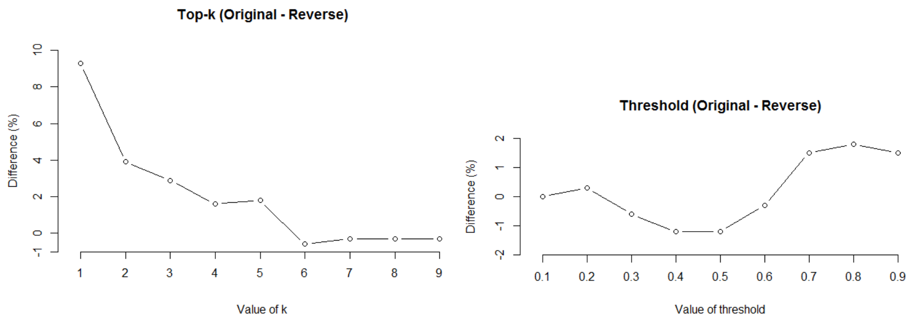
### 5.2.3. Hand-crafted maritime route: Brest-Douarnenez example

One original feature of the proposed approach is to be agnostic to the representation 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

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

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

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

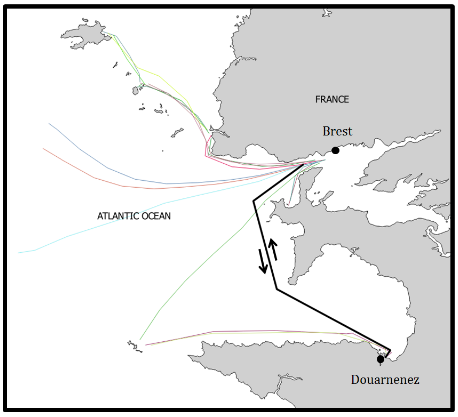



Figure 10: In black, the bidirectional hand-crafted route between Brest and Douarnenez together with the 17 routes of the dataset

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

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

Table 3 shows the confusion matrix for the origins (on the left-hand side) and the destinations (on the right-hand side) of the tracklets for which one of the two hand-crafted routes were output by the classifier. It is worth noticing that out of the 17 tracklets which had Douarnenez as either origin or destination), 14 were actually estimated correctly (82%). Instead, Brest was correctly estimated for only 4 tracklets. This can be explained by the relative isolation of Douarnenez compared to the other maritime routes, while Brest is the nodal point of the local network. Also, out of the 17 tracklets, only 2 tracklets were actually corresponding to Douarnenez to Brest journeys while none of the actual routes that the vessels followed were one of the other 17 routes of the dataset. An interesting result is that one of these two tracklets was the only one to be assigned a confidence score of 1, proving the utility of including hand-crafted routes and 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

# 6. Conclusions

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

We presented a detailed geometrical approach to compute the distance between a vessel and a maritime route valid when segments of trajectories span over very large distance. The approach is original with respect to data-centric methods where statistical motion models help compensating the possible lack of data. Focused on geometrical features only, our method for distance computation is agnostic to the way the maritime route is obtained, and is valid for hand-crafted routes.

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

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

A series of experiments has been conducted on real data excerpt from the AIS in the Brest area (France). A dataset of 800 maritime tracklets was previously labeled with the route the corresponding vessels were actually followed, providing thus some ground truth and enabling the assessment of the correctness of our method. This ground truthed dataset along with the maritime routes of interest have been made available in a companion data publication. Those experiments include the assessment of association to pairwise maritime routes (*e.g.* corresponding to ferry trips between close ports) and hand-crafted maritime routes in addition to assessment of association to classical data extracted routes. The main interest in the fuzzy rule-based classifier proposed is its interpretability and flexibility, possibly at the expense of the classification accuracy. Rather than a unique solution to route association, we defined a framework to associate vessels to maritime routes which considers users’ needs and knowledge. The framework allows customising the antecedent fuzzy membership functions according to specific routes geometry or user needs, while optimising some parameters such as the weights of the fuzzy rules by means of a training data set with ground truth labels, and achieve a higher accuracy. Typically, the parameters for defining the route membership should reflect the specific area under surveillance as well as the specific route. The aggregation operator ($t$-norm) could reflect more or less optimistic (or pessimistic) approach of the operator, and can change given the mission context. In future work, we would address the optimisation of this set of parameters to fit both the data (and capture the actual patterns of life) and the users' knowledge and information needs.

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

# Acknowledgments

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

# References

[1] C. Iphar, A. Napoli, C. Ray, Integrity assessment of a worldwide maritime tracking system for a geospatial risk analysis at sea, in: Proceedings of the 20th AGILE International Conference on Geographic Information Science (AGILE 2017), 2017.

[2] C. Kooij, M. Loonstijn, R. Hekkenberg, K. Visser, Towards autonomous shipping: Operational challenges of unmanned short sea cargo vessels, in: P. Kujala, L. Lu (Eds.), Marine Design XIII, 2018, pp. 871–880.

[3] J. Montewka, K. Wróbel, E. Heikkilä, O. Valdez-Banda, F. Goerlandt,

S. Haugen, Challenges, solution proposals and research directions in safety and risk assessment of autonomous shipping, in: Proceedings of PSAM 14 - Probabilistic Safety Assessment and Management Conference, 2018.

[4] M. D. Robards, G. Silber, J. Adams, J. Arroyo, D. Lorenzini, K. Schwehr, J. Amos, Conservation science and policy applications of the marine vessel Automatic Identification System (AIS), Bulletin of Marine Science 92 (1) (2016) 75–103. doi:10.5343/bms.2015.1034.

[5] O. Bodunov, F. Schmidt, A. Martin, A. Brito, C. Fetzer, Grand challenge: Real-time destination and ETA prediction for maritime traffic, in: Proceedings of the DEBS'18 conference, 2018. doi:10.1145/3210284.3220502.

[6] T. Wahl, G. K. Høye, A. Lyngvi, B. T. Narheim, New possible roles of small satellites in maritime surveillance, Acta Astronautica 56 (1-2) (2005) 273–277. doi:10.1016/j.actaastro.2004.09.025.

[7] M. Fournier, R. Casey Hilliard, S. Rezaee, R. Pelot, Past, present, and future of the satellite-based automatic identification system: areas of applications (2004–2016), WMU Journal of Maritime Affairs 17 (3) (2018) 311–345. doi:10.1007/s13437-018-0151-6.

[8] L. Zhen, Y. Wu, S. Wang, G. Laporte, Green technology adoption for fleet deployment in a shipping network, Transportation Research Part B: Methodological (2020) 388–410doi:10.1016/j.trb.2020.06.004.

[9] H. Rong, A. Teixeira, C. Guedes Soares, Ship collision avoidance behaviour recognition and analysis based on AIS data, Ocean Engineering (2022). doi:10.1016/j.oceaneng.2021.110479.

[10] V. Prochazka, R. Adland, F.-C. Wolff, Contracting decisions in the crude oil transportation market: Evidence from fixtures matched with AIS data, Transportation Research Part A: Policy and Practice 130 (2019) 37–53. doi:10.1016/j.tra.2019.09.009.

[11] M. Tichavska, B. Tovar, Port-city exhaust emission model: An application to cruise and ferry operations in las palmas port, Transportation Research

Part A: Policy and Practice 78 (2015) 347–360. doi:10.1016/j.tra.2015.05.021.

[12] L. Styhre, H. Winnes, J. Black, J. Lee, H. Le-Griffin, Greenhouse gas emissions from ships in ports – case studies in four continents, Transportation Research Part D: Transport and Environment 54 (2017) 212–224. doi:10.1016/j.trd.2017.04.033.

[13] D. Toscano, F. Murena, F. Quaranta, L. Mocerino, Assessment of the impact of ship emissions on air quality based on a complete annual emission inventory using ais data for the port of naples, Ocean Engineering (2021). doi:10.1016/j.oceaneng.2021.109166.

[14] C. Liu, J. Liu, X. Zhou, Z. Zhao, C. Wan, Z. Liu, AIS data-driven approach to estimate navigable capacity of busy waterways focusing on ships entering and leaving port, Ocean Engineering (2020). doi:10.1016/j.oceaneng.2020.108215.

[15] J. Cai, G. Chen, M. Lutzen, N. Gorm Maly Rytter, A practical AIS-based route library for voyage planning at the pre-fixture stage, Ocean Engineering (2021). doi:10.1016/j.oceaneng.2021.109478.

[16] T. P. Zis, H. N. Psaraftis, L. Ding, Ship weather routing: A taxonomy and survey, Ocean Engineering (2020). doi:10.1016/j.oceaneng.2020.107697.

[17] K. Kepaptsoglou, G. Fountas, M. G. Karlaftis, Weather impact on containership routing in closed seas: A chance-constraint optimization approach, Transportation Research Part C: Emerging Technologies (2015) 139–155doi:10.1016/j.trc.2015.01.027.

[18] V. N. Armstrong, Vessel optimisation for low carbon shipping, Ocean Engineering (2013) 195–207doi:10.1016/j.oceaneng.2013.06.018.

[19] M. Bentin, D. Zastrau, M. Schlaak, D. Freye, R. Elsner, S. Kotzur, A new routing optimization tool-influence of wind and waves on fuel consumption of ships with and without wind assisted ship propulsion systems, Transportation Research Procedia (2016) 153–162doi:10.1016/j.trpro.2016.05.051.

[20] V. Windeck, A Liner Shipping Network Design, Springer Gabler, Wiesbaden, Germany, 2013.

[21] M. Grifoll, F. Martinez de Oses, M. Castells, Potential economic benefits of using a weather ship routing system at short sea shipping, WMU Journal of Maritime Affairs (2018) 195–211doi:10.1007/s13437-018-0143-6.

[22] L. Perera, C. Guedes Soares, Weather routing and safe ship handling in the future of shipping, Ocean Engineering (2017) 684–695doi:10.1016/j.oceaneng.2016.09.007.

[23] V. Kotovirta, R. Jalonen, L. Axell, K. Riska, R. Berglung, A system for route optimization in ice-covered waters, Cold Regions Science and Technology (2009) 52–62doi:10.1016/j.coldregions.2008.07.003.

[24] C. P. Padhy, D. Sen, P. K. Bhaskaran, Application of wave model for weather routing of ships in the north indian ocean, Natural Hazards (2008) 373–385doi:10.1007/s11069-007-9126-1.

[25] G. Pallotta, M. Vespe, K. Bryan, Traffic Route Extraction and Anomaly Detection from AIS Data, in: Proceedings of the COST MOVE Workshop on Moving Objects at Sea, 2013.

[26] G. Pallotta, M. Vespe, K. Bryan, Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction, Entropy 15 (6) (2013) 2218–2245. doi:10.3390/e15062218.

[27] H. Rong, A. Teixeira, C. Guedes Soares, Data mining approach to shipping route characterization and anomaly detection based on ais data, Ocean Engineering (2020). doi:10.1016/j.oceaneng.2020.106936.

[28] P. Coscia, P. Braca, L. Millefiori, F. Palmieri, P. Villett, Multiple ornstein-uhlenbeck processes for maritime traffic graph representation, IEEE transactions on Aerospace and Electronic Systems (2018). doi:10.1109/TAES.2018.2808098.

[29] Z. Sun, J. Zheng, Finding potential hub locations for liner shipping, Transportation Research Part B: Methodological (2016) 750–761doi:10.1016/j.trb.2016.03.005.

[30] W.-K. Tseng, J.-L. Guo, C.-P. Liu, A comparison of great circle, great ellipse, and geodesic sailing, Journal of Marine Science and Technology 21 (3) (2013) 287–299. doi:10.6119/JMST-012-0430-5.

[31] C. S. Jensen, D. Lin, B. Chin Ooi, Continuous clustering of moving objects, IEEE Transactions on Knowledge and Data Engineering 19 (9) (2007) 1161–1174. doi:10.1109/TKDE.2007.1054.

[32] F. Boem, F. A. Pellegrino, G. Fenu, T. Parisini, Multi–feature trajectory clustering using Earth Mover's Distance, in: Proceedings of the 2011 IEEE Conference on Automation Science and Engineering, 2011. doi:10.1109/CASE.2011.6042423.

[33] G. Spiliopoulos, K. Chatzikokolakis, D. Zissis, E. Biliri, D. Papaspyros, G. Tsapelas, S. Mouzakitis, Knowledge extraction from maritime spatiotemporal data: An evaluation of clustering algorithms on big data, in: Proceedings of the 2017 IEEE International Conference on Big Data (BIGDATA), 2017, pp. 1682–1687. doi:10.1109/BigData.2017.8258106.

[34] S. Qi, P. Bouros, D. Sacharidis, N. Mamoulis, Efficient point-based trajectory search, in: C. Claramunt, M. Schneider, R. C.-W. Wong, L. Xiong, W.-K. Loh, C. Shahabi, K.-J. Li (Eds.), Advances in Spatial and Temporal Databases, Springer, 2015, pp. 179–196. doi:10.1007/978-3-319-22363-6_10.

[35] Y. Liu, X. Li, W. Hu, Semi-supervised trajectory learning using a multiscale key point based trajectory representation, in: Proceedings of the 2010 20th International Conference on Pattern Recognition, 2010. doi:10.1109/ICPR.2010.860.

[36] F. Shao, S. Cai, J. Gu, A modified Hausdorff distance based algorithm for 2-dimensional spatial trajectory matching, in: Proceedings of the 5th International Conference on Computer Science & Education, 2010, pp. 166–172. doi:10.1109/ICCSE.2010.5593666.

[37] R. Laxhammer, G. Falkman, Sequential conformal anomaly detection in trajectories based on Hausdorff distance, in: Proceedings of the 14th International Conference on Information Fusion, 2011.

[38] J. P. Bustos, F. Donoso, A. Guesalaga, M. Torres, Matching radar and satellite images for ship trajectory estimation using the Hausdorff distance, IST Radar Sonar Navigation 1 (1) (2007) 50–58. doi:10.1049/iet-rsn:20060025.

[39] B. Guan, L. Liu, J. Chen, Using relative distance and Hausdorff distance to mine trajectory clusters, TELKOMNIKA 11 (1) (2013) 115–122. doi:10.11591/telkomnika.v11i1.1877.

[40] M. Siljander, E. Venäläinen, F. Goerlandt, P. Pellikka, GIS-based cost distance modelling to support strategic maritime search and rescue planning: A feasibility study, Applied Geography 57 (2015) 54–70. doi:10.1016/j.apgeog.2014.12.013.

[41] C. Zhang, AIS data-driven general vessel destination prediction: A trajectory similarity-based approach, Ph.D. thesis, University of British Columbia (2019).

[42] L. Etienne, T. Devogele, A. Bouju, Spatio-temporal trajectory analysis of mobile objects following the same itinerary, in: In Proceedings of the Joint International Conference on Theory, Data Handling and Modelling in GeoSpatial Information Science, 2010, pp. 86–91.

[43] H. Rong, A. Teixeira, C. Guedes Soares, Ship trajectory uncertainty prediction based on a gaussian process model, Ocean Engineering 182 (2019) 499–511. doi:10.1016/j.oceaneng.2019.04.024.

[44] L. A. Zadeh, Fuzzy sets, Information and Control 8 (1965) 338–353.

[45] L. A. Zadeh, Outline of a new approach to the analysis of complex systems and decision processes, IEEE Transactions on Systems, Man, and Cybernetics SMC-3 (1) (1973) 28–44.

[46] D. Dubois, H. Prade, A review of fuzzy set aggregation connectives, Information Sciences 36 (1-2) (1985) 85–121.

[47] I. Bloch, Information combination operators for data fusion: A comparative review with classification, IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans 26 (1) (1996) 52–67.

[48] D. Dubois, H. Prade, Formal representations of uncertainty, Vol. Decision-making - Concepts and Methods, ISTE, London, UK & Wiley, Hoboken, N.J. USA, 2009, Ch. 3, pp. 85–156, invited paper.

[49] A.-L. Jousselme, G. Pallotta, Dissecting uncertainty handling techniques: Illustration on maritime anomaly detection, Journal of Advances in Information Fusion 13 (2) (2018) 158–178.

[50] C. Iphar, M. Zocholl, A.-L. Jousselme, Semantics of maritime routes: Conciliating complementary views, in: Proceedings of the OCEANS 2021 San Diego Conference, 2021. doi:10.23919/OCEANS44145.2021.9705934.

[51] C. Iphar, A.-L. Jousselme, Maritime routes and vessel tracklet dataset for vessel-to-route association (version 1.0), data set. Licence CC-BY-NC-SA-4.0. Zenodo (2022). doi:10.5281/zenodo.6402160.

[52] C. Iphar, A.-L. Jousselme, G. Pallotta, Maritime route and vessel tracklet dataset for vessel-to-route association, Data in Brief (2022). doi:10.1016/j.dib.2022.108513.

[53] W. Ma, D. Ma, Y. Ma, J. Zhang, D. Wang, Green maritime: a routing and speed multi-objective optimization strategy, Journal of Cleaner Production (2021). doi:10.1016/j.jclepro.2021.127179.

[54] H. Psaraftis, C. Kontovas, Green Maritime Transportation: Speed and Route Optimization, Springer, 2016, Ch. 9. doi:10.1007/978-3-319-17175-3_9.