

# Container ship fleet route evaluation and similarity measurement between two shipping line ports

---

Šakan, Davor; Žuškin, Srđan; Rudan, Igor; Brčić, David

Source / Izvornik: **Journal of marine science and engineering, 2023, 11, 1 - 16**

**Journal article, Published version**

**Rad u časopisu, Objavljena verzija rada (izdavačev PDF)**

<https://doi.org/10.3390/jmse11020400>

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:187:935218>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-07-06**



**Sveučilište u Rijeci, Pomorski fakultet**  
University of Rijeka, Faculty of Maritime Studies

Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Maritime Studies - FMSRI Repository](#)



Article

# Container Ship Fleet Route Evaluation and Similarity Measurement between Two Shipping Line Ports

Davor Šakan \* , Srđan Žuškin \* , Igor Rudan and David Brčić 

Faculty of Maritime Studies, University of Rijeka, Studentska ulica 2, 51000 Rijeka, Croatia

\* Correspondence: davor.sakan@uniri.hr (D.Š.); srdan.zuskin@uniri.hr (S.Ž.)

**Abstract:** The characterization of ship routes and route similarity measurement based on Automatic Identification System (AIS) data are topics of various scientific interests. Common route research approaches use available AIS identifiers of ship types. However, assessing route and similarity profiles for individual fleets requires collecting data from secondary sources, dedicated software libraries or the creation of specific methods. Using an open-source approach, public AIS and ship data, we evaluate route characteristics for the container ships of a single fleet in a six-month period, calling on two selected ports of the shipping line on the USA East Coast. We evaluate the routes in terms of length, duration and speed, whereas for the similarity measurement we employ the discrete Fréchet distance (DFD). The voyage length, duration and average speed distributions were observed to be moderately positive (0.77), negative (−0.62), and highly positively skewed based on the adjusted Fisher–Pearson coefficient of skewness (1.23). The most similar voyages were from the same ships, with the lowest discrete Fréchet distance similarity value (0.9 NM), whereas 2 different ships had the most dissimilar voyages, with the highest DFD value (14.1 NM). The proposed methodology enables assessment of similarities between individual ships, or between fleets.

**Keywords:** route similarity; similarity measures; discrete Fréchet distance; fleet analysis; container fleet



**Citation:** Šakan, D.; Žuškin, S.; Rudan, I.; Brčić, D. Container Ship Fleet Route Evaluation and Similarity Measurement between Two Shipping Line Ports. *J. Mar. Sci. Eng.* **2023**, *11*, 400. <https://doi.org/10.3390/jmse11020400>

Academic Editors: Marko Perkovic, Lucjan Gucma, Sebastian Feuerstack and Zaili Yang

Received: 27 December 2022

Revised: 30 January 2023

Accepted: 7 February 2023

Published: 11 February 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction and Background

Route, path, line or way are various terms describing series of interconnected way-points that a movable entity will traverse or has traversed. For inanimate entities, including land vehicles, aircrafts or ships, routes must be created and adapted to the particulars of the transport mode and voyage objectives. On board ships, the creation of intended routes is still the duty of the navigator(s) and is a part of voyage planning. This is a process, subject to numerous constraints, rules and regulations in which the navigator uses skills, knowledge and experience to create a safe and efficient route for the forthcoming voyage [1]. During the voyage, the vessel movement, traversed positions and other voyage-related data is saved in predetermined intervals by the onboard equipment, and, if necessary, route retrieval is simple on board. For shipping companies, such route data from ships can be useful for various purposes, including safety of navigation, individual voyages, or even more, fleet efficiency metrics based on length of the route, average speed, traversal time or fuel consumption. Using such data for the wider research community, notably in fleet analysis, is not so common since it is rarely public. However, there are other sources of vessel data, such as data broadcasted from the Automatic Identification System (AIS). AIS data is collected and made available by several public and free sources, including the one used in the conducted research, as will be presented in the following chapter. From the sources, such data can be used for various ship route data purposes.

Route extraction in the context of safe route design, based on AIS data and available categories of ships using Kernel Density Estimation with route boundary extraction and centerline extraction, was considered in [2]. In [3], spatial and temporal vessel trajectory clustering in the context of extracting traffic flow information, customary route discovery

and anomaly detection was carried out based on AIS data. The trajectories have been investigated in terms of their similarity to Merge Distance. In [4], computer vision techniques were employed for navigation patterns estimation in the context of traffic statistics and a simplified ship maneuvering model, with traffic group collecting based on geometric similarity using AIS data. In [5], the Equivalent Passage Plan method, based on the notion that the vessel in navigation follows a prepared voyage plan, was used as a framework for trajectory simplification and event identification in the context of ship behavior extraction. Further, the Hausdorff distance was used to measure similarity between the simplified and original AIS trajectories. Hausdorff and other similarity methods can be broadly grouped into warping-based methods and shape-based methods. The warping methods include Dynamic Time Warping (DTW), the Longest Common Subsequence (LCSS), and the Edit Distance on Real sequence (EDR). Shape-based methods include the Hausdorff distance, One Way Distance (OWD), and the Fréchet distance [6]. Similarity measures can be used in spatial, temporal and spatiotemporal contexts. For spatial contexts, the measures can be compared topologically and quantitatively, in terms of global and local path similarity, travelled distance, range and shape [7]. After assessing the available various route research interests and similarity methods, it is important to present a brief overview of AIS, the data source on which route analysis and similarity methods are employed in the maritime context.

AIS is a digital short-range automatic data exchange system operating at a very high frequency (VHF) band. Although it was devised as a terrestrial system, AIS signals can be received by satellites as well. Its use is mandatory for all ships under the requirements of the International Convention for the Safety of Life at Sea (SOLAS), determined by the International Maritime Organization (IMO). Although used primarily in maritime contexts such as ship identification, collision avoidance or traffic monitoring [8], the available AIS data is a valuable source for various research interests [9]. AIS data is static (e.g., ship identification, basic dimensions, type), dynamic (e.g., position, speed over ground, heading), voyage-related (e.g., navigation status, draft, cargo type) or in form of short safety-related messages. SOLAS-regulated Class A ships send their dynamic data messages in intervals dependent on ship speed and course change. Intervals start from between 2 and 12 s, when the ship is underway, and are 3 min when it is anchored or moored [10]. Further, AIS provides several identifiers for special crafts (e.g., pilot vessels and tugs) and ships such as high-speed crafts (HSC), cargo ships, tankers, passenger ships and ships from the other ships category [11]. Although beneficial for broad ship-type categories, the shortcoming of AIS categorization appears when the research focus is on different categories of ships, such as general cargo ships and container ships. They are both cargo ships; however, they lack separate AIS categories. This makes it more difficult, for example, for evaluation comparing single or multiple fleets, or evaluation of previously stated different cargo ship classes. This was confirmed in the research literature, since the available articles on the subject are rather limited. With the observed gap, we decided to evaluate if the available approaches using AIS default categories in route analysis could be further extended for single or multiple ship fleet analysis.

When assessing ship route research objectives using AIS data [12] and including similarity measures, frequently applied approaches consider general AIS ship categories or created subcategories for route and traffic analysis [13]. In [14], authors examined and delineated maritime routes between ports along the Atlantic coast of the USA using AIS data from 2010 to 2012. Commercial tracks were generated for commercial vessels, with a focus on cargo vessels. The routes were delineated with a 95% boundary of all traffic transiting each route. In [15], authors analyzed ship traffic demand and the spatial-temporal dynamics in Singapore port waters using AIS data. They found that origin-to-destination pairs and navigation routes in the area were stable and found several hotspots with high speeds. In [16], authors evaluated trajectory clustering for marine traffic pattern recognition around the Chinese Zhoushan Islands from January to February 2015. A Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm with

improved parameters was proposed. In [17], authors devised a method to transform the ship trajectory into a ship trip semantic object (STSO) with semantic information. Authors used graph theory, and integrated STSO into the nodes and edges of a directed maritime traffic graph for route extraction. The AIS data was data collected from 1 January to 31 December 2017 from crude oil tankers in the Asia-Pacific and Indian Ocean regions. In [18], a ship AIS trajectory clustering method based on the Hausdorff similarity distance and a Hierarchical DBSCAN (HDBSCAN) was used. AIS data on the estuary waters of the Yangtze River in China was used and compared with k-means, spectral clustering and DBSCAN.

As we have presented, there are various route research approaches and interests using AIS data for route and trajectory research. Similarity measurement is used in numerous contexts of AIS data, route and trajectory analysis. Although diverse types of ships were covered in the presented papers, route comparison between other groups of ships besides general AIS ship types is not very well researched. With such observations, we focused our selection on the cargo ship subcategory, which could be suitable for fleet selection and analysis.

Accounting for 13.34% in 2022 of total world fleet by deadweight, container ships are the third largest principal vessel type, behind bulk carriers and oil tankers [19]. Further, the assessment of container ship routes and fleets has several possible advantages over the assessment of other principal vessel types. Vessels are employed in liner services, with companies usually providing public data on their services and vessels. In regular operation, vessels call on the same service ports according to their schedules, thus enabling route comparison in both spatial and temporal domains. Moreover, 4 large carriers (MSC, APM-Maersk, CMA CGM and the COSCO group), accounting for more than 50% of capacity, dominate the market [19], facilitating the collection of significant route data, which can be interpreted for individual fleets or used to make inferences about total container fleet insights.

A broad overview of container ship fleet research topics includes various levels of planning and decision-making, liner shipping networks, scheduling, fleet compositions and deployment, air emissions, optimization and routing, resulting in a considerable body of literature [20–22]. Further, ships and fleet route elements are commonly evaluated in terms of speed and speed optimization [23–25] or trajectory and speed distribution [26]. Speed accuracy is evaluated as well, using the collected speed through water (STW) data of 190 container vessels from Maersk [27]. However, literature on container ship fleet route analysis is scarce. An operational analysis of container ships using AIS and environmental data was carried out [28] based on Hortonworks' big data framework. Trajectories and speeds from several voyages were compared to evaluate voyage energy efficiency for a single 10,000 TEU ship and four 13,000 TEU sisterships. A speed pattern analysis was carried out for 8600, 13,000 and 18,000 TEU ships from several companies, whereas route similarity was not considered. Container ships and fleets were analyzed, using AIS data collected in a database from along the Portuguese coast for top container carriers, services, and vessels [29]. Routes were characterized for individual ships, indicating the sequence of visited ports, respective travel times, the total number of trips and carrying capacity. However, spatial distributions of routes and their similarities were not investigated. In [30], vessel speed, course and path analysis in the Botlek area of the port of Rotterdam was assessed for five container ship categories along with other vessel types. This research was further extended in [31], where a comparison study on AIS data with a characterization of ship traffic behavior was carried out. The data was evaluated for traffic modelling and simulation in restricted waterways. Research areas were a narrow waterway in the port of Rotterdam and the Yangtze wide waterway area near the Su-Tong Bridge. Spatial distributions perpendicular to the water flow and to the channel, sailing speeds, average speeds and course distributions were analyzed. Again, route similarities were not assessed in either of the articles. As can be observed, route similarity measures are usually not considered in either general container or individual fleet analysis.

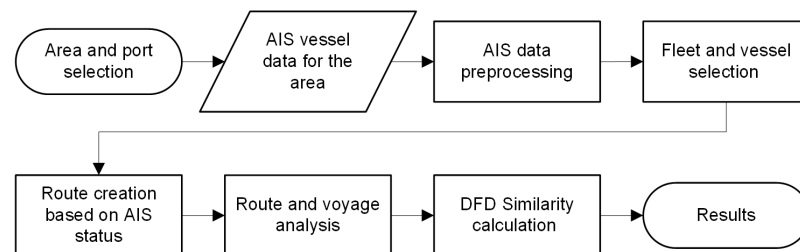
With that notion in mind, we selected container ships as a subcategory of interest extracted from the total cargo ships category and examined it in terms of route characteristics and similarity.

A fleet of container ships from a single shipping company employed in regular service calling on the ports of Savannah and Charleston on the east coast of the USA was chosen. The selection was chosen due to the similar constrained conditions of the selected ports and possible routes and expected navigational decisions made when planning and executing the voyage.

The remainder of this paper is structured as follows. In Section 2, we present the methods used in the research; we introduce the research area and describe the pre-processing of AIS data, route creation and similarity measurement. In Section 3, we present the research results and findings. Section 4 is the discussion, and we present the conclusions with future research suggestions in Section 5.

## 2. Methods

To evaluate the differences between the ship routes, several conditions were considered. The ports and area of interest should be located in relationship to one to another to provide modest route choice variability, within a somewhat constrained area. Further, the ports should be called on regularly by major companies of interest to facilitate fleet creation from available collected and processed data. From several candidate pairs, the ports of Savannah and Charleston were chosen. Their respective details and further methodology will be described in the following paragraphs along with a prior general overview, presented in Figure 1.



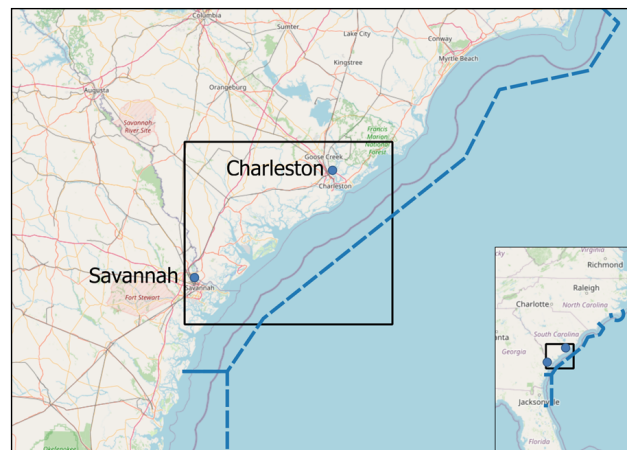
**Figure 1.** Methodology overview.

The data collected and used is publicly available AIS data from the NOAA AccessAIS portal [32]. The observed period starts on 1 January 2019 and ends on 30 June 2019. The area of interest (AOI) is bounded by 31.791° N, 81.224° W, 33.102° N and 79.453° W, as presented in Figure 2. This is the coastal area between two major USA East Coast ports of interest, Savannah and Charleston. The respective ports, besides handling several other types of cargo, are major hubs for containerized cargo.

During the fiscal year 2019 (ending 30 June 2019), 1848 ships called on the Savannah Garden City Terminal with 4.48 million twenty-foot equivalent units (TEUs) handled. The projected increase in cargo volume for the year 2030 was 48% [33]. In the port of Charleston, the Wando Welch and North Charleston terminals handled 2.93 million TEUs. At all the Charleston cargo terminals, a total of 1696 vessels called in the 2019 fiscal year [34]. Further, vessels from several major container carriers and alliances call on the ports regularly as a part of their services.

According to [35], the distance from Savannah to Charleston is 104 nautical miles (NM). The calculated distance is measured along navigable tracks as the shortest route for safe navigation between the two ports. As we present in the results section, the actual distances for the selected ships are greater due to several reasons, including the size of the ships and the characteristics of the navigational area.

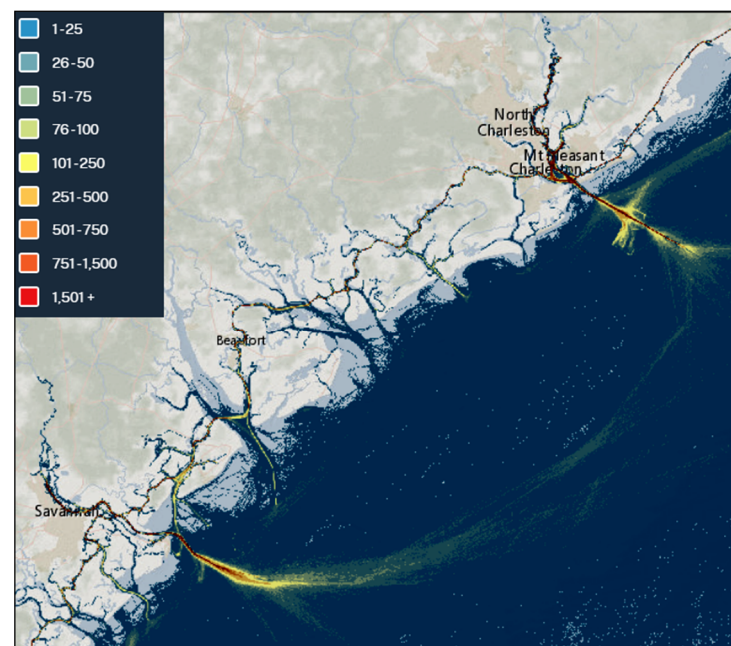




**Figure 2.** Area of interest with AIS data boundary marked in rectangle with ports of Savannah and Charleston. Right whale Seasonal Management Area (SMA) boundary marked with dashed line.

Besides the importance of the respective ports and the significant vessel traffic they generate, we must emphasize that the coastal area is part of the North Atlantic right whale Seasonal Management Areas (SMA). All vessels with length overall (LOA) of 65 feet (19.81 m) or greater and subject to the jurisdiction of the United States are restricted to speeds of 10 knots or fewer in a continuous 20 NM Seasonal Management Area between 1 November and 30 April annually [36]. The areas were established to prevent collisions of vessels with endangered right whales.

To identify the patterns of interest for route creation, available vessel density data for the AOI was investigated. The data is available as a publicly available annual AIS vessel count dataset for the USA and international waters. The annual data is summarized using 100 m by 100 m geographical grid cells with single transits counted each time a vessel track passes through, starts or stops within the grid cell [37]. A simplified geographical area subset transit density graph can be obtained through the AccessAIS portal as well, as presented in Figure 3.



**Figure 3.** Annual AIS transit vessel counts for the area of interest in the observed period.

We conducted data preparation and preprocessing using the Python programming language; version 3.10.4, the Pandas data analysis library [38], version 1.4.2 [39]; the Shapely package for computational geometry [40] and the NumPy array programming library [41], version 1.21.6. To create routes, we used the MovingPandas library [42], version 0.9.rc3 [43]. The library is based on Pandas, GeoPandas, an extension to Pandas enabling spatial operations on geometric types [44], and HoloViz high-level visualization tools [45]. A trajectory in MovingPandas is defined as a time-ordered series of geometries and can be point or line based.

From the whole dataset, we selected cargo ships by the AIS ship code (70 to 79) used for cargo ships [46]. The cargo ship subset included 724 unique entries. To create a container ship fleet subset, we collected publicly available fleet data about ships and selected ships by their names and identities. The container fleet from the selected company included 22 unique ships with 107,977 cumulative AIS observations and data entries. From the ship identities, a trajectory collection was created for all 22 unique ships. A single trajectory entry in the collection accounts for all ship observations throughout the six-month period. There are several approaches available for the creation and separation of trajectory entries into individual voyages and routes with available MovingPandas functions. They are based on regular time intervals, speed, stopping and observation gaps. Since the AIS data includes the navigational status of the ship—which is 0 for the ship when underway using its engine—the individual voyages were split at entries when the ship status changed from underway to moored or anchored. We must emphasize that the change of a ship’s navigational status in AIS must be changed manually on board. If not done in a timely manner, this could lead to false interpretations of a ship’s stopping or movement and result in erroneous statuses. However, in the observed dataset, this was not the case, and there were no significant discrepancies between ship status and individual voyage delineation. Further, compared with other available MovingPandas splitting approaches, the preselection by status yielded better results in the splitting up of individual voyages (or trajectories, as implemented in MovingPandas). To reduce the number of unnecessary voyages and split the voyages accordingly, the observational gap value between observations was set to 30 min and the minimum length of the trajectory was set to 200 km (to approximate the distance between ports). With these constraints, the subset was reduced to 13 ships with 23 individual voyages. We created the final subset by choosing the general direction (between 050° and 060°) from individual trajectory start and end points from Savannah to Charleston. The final subset therefore includes 12 voyages (trajectories) from 8 ships. Basic ship particulars are presented in Table 1. The ships in the subset are sorted by overall length and principal dimensions. The abbreviations and acronyms are described as follows: Length Over All (LOA), Breadth extreme ( $B_{ext}$ ), Draught maximum ( $D_{max}$ ), Draught per voyage in meters ( $D_{voy}$ ) and deadweight (DWT) in metric tons. IDs are designated as letters with added numbers for individual voyages.

**Table 1.** Container ship fleet subset.

Year Built	ID	LOA (m)	$B_{ext}$ (m)	$D_{max}$ (m)	$D_{voy}$ (m)	DWT (t)	GT
2015	G	299.00	48.20	14.80	14.8	113,800	96,253
2015	F1, F2	299.95	48.20	14.80	14.8	112,729	95,263
2010	D	334.00	42.80	15.00	15.0	109,056	89,787
2005	E1, E2, E3	334.07	42.80	14.52	14.5	101,779	91,410
2008	A	347.00	45.20	15.50	15.5	130,700	128,600
2008	C	349.98	42.87	15.00	12.0	120,892	111,249
2008	B	359.99	42.80	15.00	15.0	120,944	111,249
2009	H1, H2	363.61	45.60	15.50	15.5	131,292	131,332

After the creation of individual voyages, we evaluated the route lengths, voyage durations, average speed in knots (kt) and similarity between individual routes. Among numerous available similarity measures, Fréchet distance (FD) or, formally,  $\delta_F$  is often

used [47]. It is described intuitively as a leash length between a person and a dog walking on their respective curves. They can have independent velocities, stop, and start; however, they cannot go backwards. Therefore,  $\delta_F$  can be described as the shortest possible leash length for the completion of the walk. In a formal way, we can then define a curve as continuous mapping  $f : [a, b] \rightarrow V$  where  $a, b \in \mathbb{R}$  and  $a \leq b$  are in metric space  $V$ . Two continuous curves,  $f : [a, b] \rightarrow V$  and  $g : [a', b'] \rightarrow V$  with their  $\delta_F$  can be defined as [47]

$$\delta_F(f, g) = \inf_{\alpha, \beta} \max_{t \in [0, 1]} d(f(\alpha(t)), g(\beta(t))), \tag{1}$$

where continuous nondecreasing functions  $\alpha : [0, 1]$  and  $\beta : [0, 1]$  map to  $[a, a']$  and  $[b, b']$ . The continuous curves can be further approximated as discrete polygonal curves. Therefore, an approximative discrete Fréchet distance (DFD) or coupling distance  $d_{dF}$  can be calculated. The polygonal curves can then be defined as  $P : [0, n] \rightarrow V$  and  $Q : [0, n] \rightarrow V$  with corresponding sequences  $\sigma(P) = (u_1, \dots, u_p)$  and  $\sigma(Q) = (v_1, \dots, v_q)$  with coupling  $L$ , which is a sequence of distinct pairs of points of curves  $P$  and  $Q$ , respecting the point order  $(u_{a_1}, v_{b_1}), (u_{a_2}, v_{b_2}), \dots, (u_{a_m}, v_{b_m})$ . The first and last point pairs must be ordered as  $a_1 = 1, b_1 = 1, a_m = p, b_m = q$ . For all subsequent points  $i = 1, \dots, q$ , we have  $a_{i+1} = a_i$  or  $a_{i+1} = a_i + 1$ , and  $b_{i+1} = b_i$  or  $b_{i+1} = b_i + 1$ . The maximum distance  $\|L\|$  between the point pairs for a given coupling  $L$  is therefore the coupling distance. We can then define the discrete Fréchet distance as the minimum distance between all possible couplings between  $P$  and  $Q$  [47].

$$d_{dF}(P, Q) = \min\{\|L\|\} \tag{2}$$

Therefore, using the previous person and dog or frog hopping analogy—since we consider discrete movement—we can describe the pair matching and discrete distance calculation. The person and dog can stay on their current vertices, while the other advances or they both advance to the next vertex on their respective paths. The DFD is then the smallest distance for the whole sequence of jumps to the last vertices.

The advantage of DFD over continuous FD is that it has a reduced computational cost while being a good approximation of the continuous solution. Further, provision of an upper bound on the continuous FD is provided, and the deviation is not greater than the longest edge of the trajectory [48]. Finally, the chosen DFD implementation [49] used in our research is based on the algorithm from [47].

Compared to other similarity measures such as Hausdorff distance, FD (and variants) is more useful for comparing curves, since it considers location and ordering of points of the curve [50]. The chosen DFD implementation is available in the Python similarity measures library, which includes Dynamic DTW, Partial Curve Mapping (PCM), Curve Length (CL) and other similarity measures [51].

The basic implementation of the similarity measures library considers the calculation of similarity between two curves only, so we had to extend the approach as follows. To calculate DFD values, we created 2D arrays of latitude and longitude coordinate pairs from individual ship position entries. The position coordinates, which were defined using the angular World Geodetic System (WGS84), were transformed into local planar-projected Universal Transverse Mercator (UTM) zone 17N system coordinates to improve the accuracy of the implemented DFD approach. Then, a square  $n \times n$  mesh grid consisting of individual route arrays with coordinate pairs was created. A vectorized DFD method was then applied to obtain the similarity results, which are presented in the results section.

### 3. Results

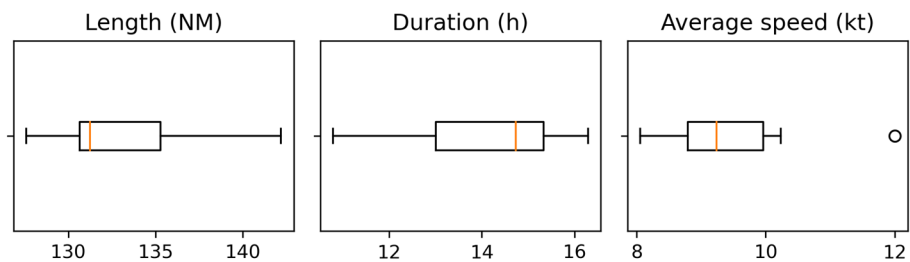
Basic features are summarized in three voyage values: length in nautical miles (NM), duration in hours (h) and average speed in knots (kt). These values, in the form of statistical summary, are presented in Table 2, along with boxplots in Figure 4. The values represent



length in nautical miles (NM), voyage duration in hours (h) and average ship speeds in knots (kt).

**Table 2.** Statistical summary of voyage lengths.

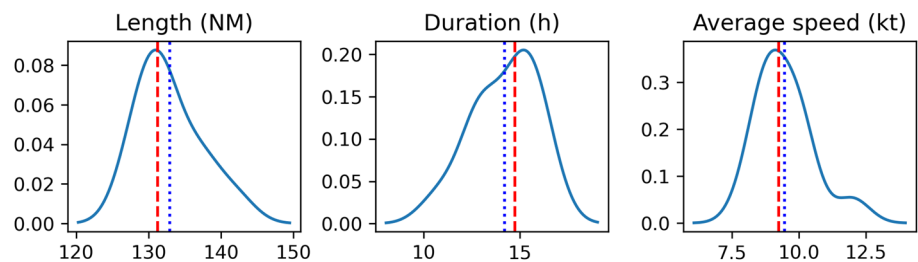
Count (12)	Length (NM)	Duration (h)	Average Speed (kt)
Mean	132.9	14.2	9.5
Std	4.4	1.7	1.0
Min	127.6	10.8	8.1
25%	130.7	13.0	8.8
50%	131.1	14.7	9.2
75%	135.3	15.3	10.0
Max	142.2	16.3	12.0



**Figure 4.** Boxplots for voyage length (NM), duration (h) and average speed (kt).

As is observable, the mean value of voyage length is 132.9 NM with a standard deviation of 4.4 NM, whereas the difference from the shortest to the longest voyage is 14.6 NM. The mean and median length values are close; however, the distribution is moderately positively skewed with a value of 0.77. The skewness value is calculated as an adjusted Fisher–Pearson coefficient of skewness with classification as in [52]. This classification is as follows: normal (−0.5 to +0.5), moderately skewed (−1.0 to −0.5 and +0.5 to +1.0) or highly skewed (<−1.0 or >+1.0).

The next value is the distribution for the voyage duration in hours. The difference is 5.2 h between minimum and maximum value, whereas the mean value is 14.2 h. The distribution is moderately negatively skewed (−0.62) as is observable in Figures 4 and 5, in which other distribution characteristics are visible as well.

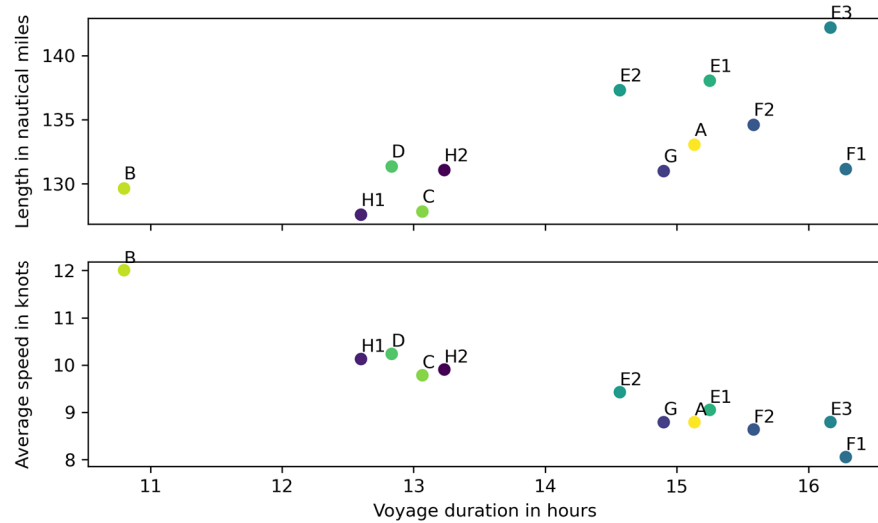


**Figure 5.** Distribution plots for voyage length, duration and average speed. Dashed vertical line represents median value and dotted line represents mean value.

The last distribution is for the average speed with a difference between minimum and maximum of 3.9 kt. Further, the maximum value can be considered an outlier (outside the interquartile range multiplied by 1.5), which is observable in the boxplot in Figure 4. However, since this not an erroneous value and the sample size is small, we did not remove it from analysis. Further, the distribution is highly positively skewed (1.23), which can be seen in the respective plot in the Figure 5.

Since the number of voyages is small—compared to overall voyages conducted over several years, or total numbers that include other ships and fleets on this service—we did not make further inferences by using methods to test distributions, either for normality

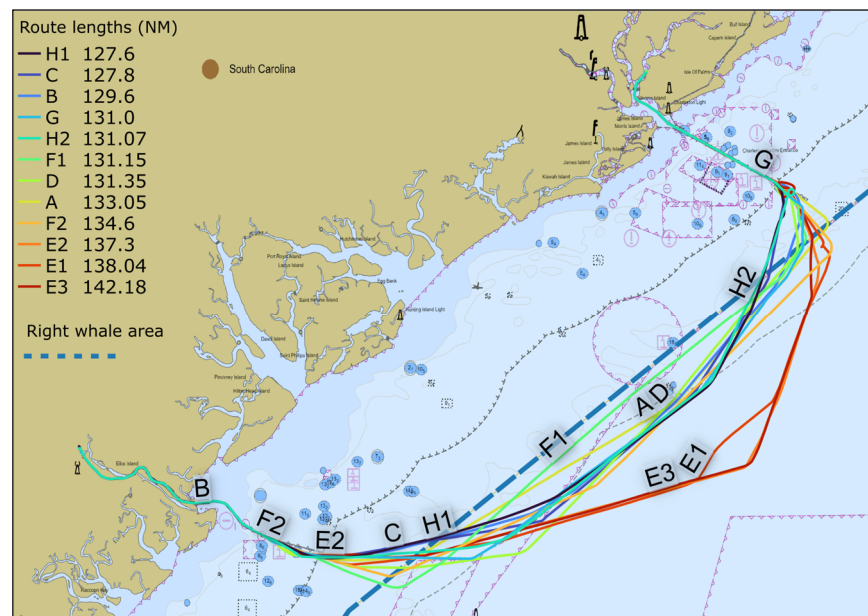
or for underlying distribution determination. This will be addressed further in future research. The actual distribution of individual voyage lengths and speeds can be observed in scatterplots in Figure 6.



**Figure 6.** Scatterplot (top) of voyage duration (h) and average speed (kt) compared with voyage duration (bottom). The letters correspond to ship identification as stated in Table 1.

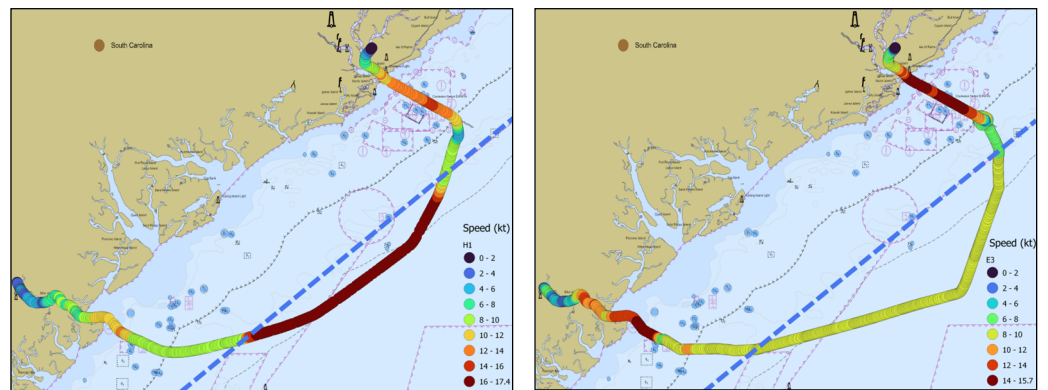
There are three clusters observable. First, a single entry for ship B, with the shortest voyage length and highest average speed. Second, centered on about 13 h of voyage length, including ships C and D with 2 individual voyages of ship H (H1, H2). The third cluster spreads out between 14 h and 16 h of voyage duration and includes the remaining ships: A, E, F and G.

The characteristics of individual ship routes and their spatial distribution can be observed in Figure 7. Seen beside the visible individual ship routes, the dashed line represents the Seasonal Management Area boundary, and we can see that the ships were sailing outside the area after passing the port approach areas for Savannah and Charleston.



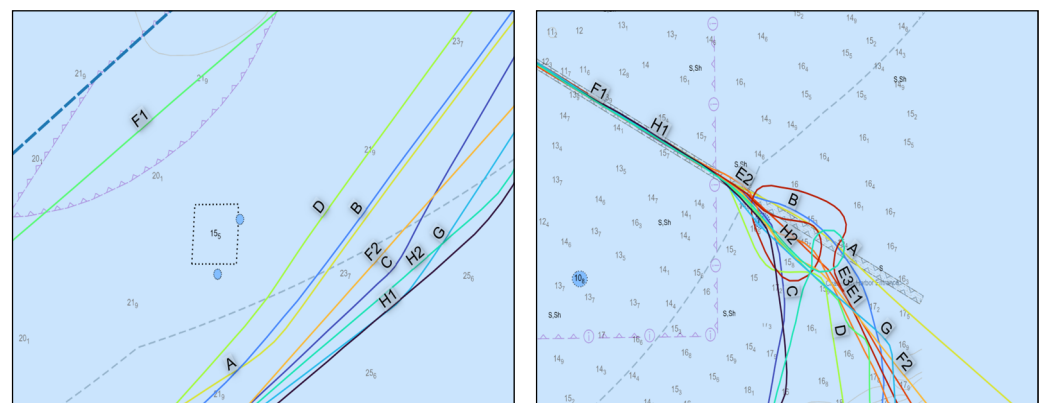
**Figure 7.** Route lengths for individual ships and voyages (scale 1:600000). Dashed line represents right whale Seasonal Management Area boundary.

Further, we examined the trajectory profiles for the longest and shortest route (H1 and E3) presented in Figure 8. The route length for ship H in voyage H1 (27 January 2019) was 127.6 NM with a passage time of 12.6 h, which was accounted for by an average speed of 10.1 kt. On the other hand, the voyage with the highest average speed of 12 kt, for ship B, was completed in 10.8 h with a total length of 129.6 NM. Ship E, in voyage E3 (14–15 April 2019), completed the respective voyage in 16.2 h with a total length of 142.2 NM with average speed of 8.8 kt. The voyage with the lowest average speed of 8 kt was conducted by ship F in voyage F1, with a total route length of 131.1 NM. The following route comparison could be expressed in terms of highest and lowest speeds, although we present the cases by route lengths only.



**Figure 8.** Trajectory profile for shortest (H1–(left)) and longest route (E3–(right)) with Speed Over Ground (SOG) in knots for individual AIS observations throughout the voyage (scale 1:600000).

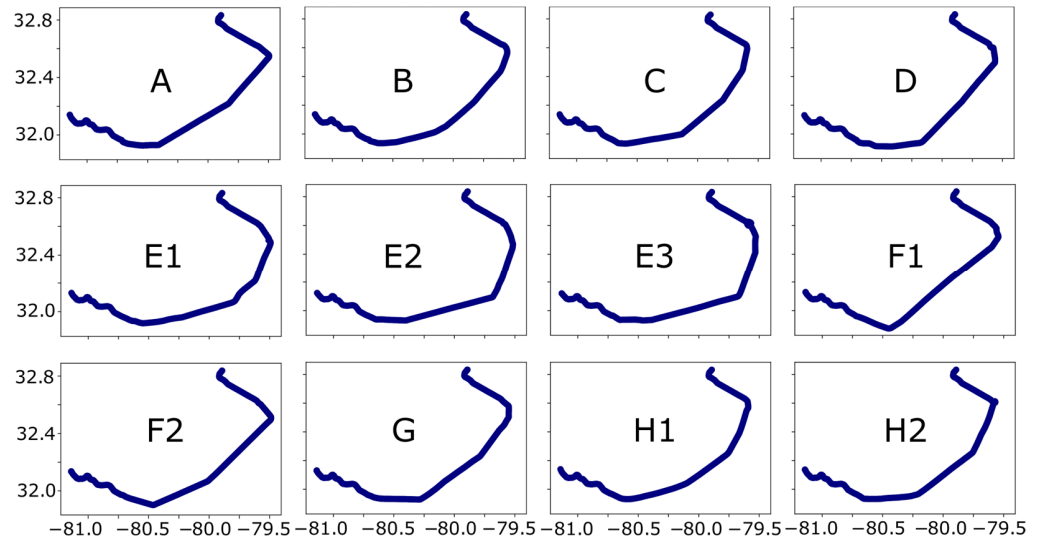
Besides the route selection and length, there are differences in speed profiles as well. Ship H1 had a significantly higher speed in the central part of the route between the Savannah and Charleston approaches. For ship H, the speed was over 16 kt to a maximum of 17.4 kt, as was reported in SOG values in AIS data. Ships sailed with required speeds in the SMA and stayed outside the coastal areas in which there are obstructions and fishing facilities which could be dangerous to navigation, as is visible in Figure 9 (left). Further, we can observe differences in the Charleston approach and pilot boarding area (in vicinity of Charleston Entrance; buoys not visible on presented figures). Some of the ships proceeded directly, while two ships (H2, E3) made turns while presumably picking up the pilot.



**Figure 9.** Fish haven obstructions (buoys not shown) with depth of 15.5 m (left) and approach area (right) for port of Charleston (scale 1:50000).

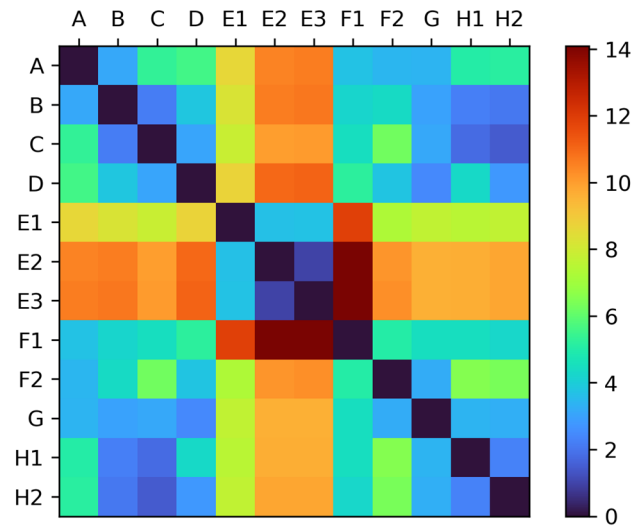
Finally, we evaluated the routes in terms of similarity. The individual routes are presented in Figure 10 as a reference. We can observe that route parts in constrained areas of navigation such as port approaches or river passages were similar. The differences arose

with choices about whether to pass northwards or southwards to avoid several obstructions, which can be observed in previous figures.



**Figure 10.** The routes for individual voyages. The  $x$ -axis represents longitude,  $y$ -axis latitude, for each of the subplots.

The calculated DFD similarity values are presented in Figure 11 as a heatmap. The lower values and darker color intensity indicate higher similarity. We can observe several routes with higher similarity. The calculated values are the distances in nautical miles.



**Figure 11.** Heatmap plot for discrete Fréchet distance route similarity results (NM). Lower value and darker color intensity indicates higher similarity.

There were similarities among the different voyages from the same ships (E1–E2–E3, F1–F2 and H1–H2). Further, the highest route similarity between different ships was for ship C and ship H, whereas the lowest similarity was between routes from ships E and F. The calculated DFD similarity value complements the interpretation of similarities of visual or geographical routes and the statistical analysis of voyage data.

#### 4. Discussion

The aim of our research was to analyze general route characteristics and evaluate similarity for a container ship fleet of a single company. As the preliminary research literature

showed, route and similarity analysis considering fleets is not common. Reasons might include the perception of the availability of data and tools for such analyses. Formatted public AIS data records are available from NOAA, other national providers, and at times as sample or test datasets from private providers. The other challenge is collecting the ship data on which fleet assessment is based. This includes information on ship fleets and categories other than general AIS ship type categories. Nevertheless, the data can be collected from shipping companies and other reliable online sources such as shipping vessel registers and official international or national vessel data providers. However, this is a time-consuming process. Although there are many private providers which can provide aggregated ship data on companies and fleets with detailed ship types, the financial aspect of acquiring data might be worth consideration or even be a barrier for those interested.

For route characterization, we selected the two ports because of their proximity and constraints in the number of possible routes. As expected, the sailed routes taken by ships of different sizes were mostly similar, notably for the same ships. The route differences arose after the port approach areas were passed and the right whale protection area, which all the ships sailed outside, was exited. This was presumably due to the right whale avoidance and speed constraints required during the period of the SMA (from 1 November to 30 April 2019). Further, even after the end of the SMA period, the vessels sailed outside the area. It must be seen in future research if this occurs during the rest of the year. Further, as we presented regarding ships' particulars, the maximum and reported voyage drafts were large, ranging from 14.5 to 15.5 m, with only one vessel (C) reporting a lower voyage draft (12 m) than the maximum (15 m). Therefore, to keep a safe depth below keel, referred to as Under Keel Clearance (UKC), ships must stay in the deeper waters with depths of 20 m or more, which are outside the SMA area. In the shallower area near the coast, there are obstructions and fishing facilities as well, so even in that context it is safer to stay in the deep-water area.

Route characteristics do not differentiate substantially. There are differences in length which we can attribute to collision avoidance, besides route choice by the navigator. However, to confirm that decisively, we should evaluate the surrounding traffic as well. Looking in terms of route length, voyage duration and speed, most of the observations are not very dispersed either from the mean or median values. The distribution shapes reveal moderate to high skewness; however, the number of voyages and routes is too low to give decisive conclusions, and further tests should be conducted regarding the type and shape of the underlying distribution. Further, assessment of the individual voyages reveals there is some clustering of individual routes, considering both voyage length and duration. The shortest voyage duration was 10.8 h, compared to 16.3 h for the longest voyage. The mean and median values of the average speed are 9.5 kt and 9.2 kt, with only one vessel sailing with an average speed of 12 kt. This along with trajectory analysis reveals differences in the speed profiles of the vessels, at least during parts of their routes. The fastest-average-speed vessel had much more variable speed (H1) throughout the voyage than the vessel with the lowest average speed (E3). To better understand behavior in parts of the voyage, additional trajectory analysis is required, and identification of regularities in behavior, if they exist.

The similarity measurement revealed that the same vessels take more similar routes in different voyages compared to the rest of the vessels. This is observable both in visual shape evaluation (vessels E, F and H), which was applicable in our study, and with DFD similarity measurement. This similarity should be assessed against a baseline route which either is derived as a generalization from numerous routes or is a route devised by a navigator in accordance with established voyage-planning procedures. The DFD pairwise similarity measurement, visual inspection for a small number of routes and methodology, is applicable. However, there is an issue of computational cost for DFD calculation of substantial numbers of routes. We did not generalize the routes due to the small number (several hundred for each route) of AIS data entries—points regularized at approximately one-minute intervals by default in NOAA AIS datasets. The number of entries can be reduced even more with route generalization and the extraction of important route points



(e.g., when the vessel alters the course) and omission of the rest. This would reduce computation time when dealing either with substantial numbers of points per voyage or substantial numbers of voyages. Further, other similarity measures should be used as well for comparison. To evaluate usability, in forthcoming research we will assess DFD against other similarities in other shipping areas and with other ships and fleets.

Our methodology builds on established tools and practices for data assessment and similarity measurement, along with our own approach to similarity comparison between the individual ships or entire fleets. Since we conducted the analysis with open-source software based on the Python programming language, other interested researchers can apply the described methodology. Finally, the usage of discrete Fréchet distance to measure similarity of the routes is a valuable addition to route knowledge interpretation. It formalizes the route similarity in a single value; likewise with the route length, which is one of the values with which we compare the routes. The use of DFD values can be applicable in a variety of contexts, such as limiting values for the route selection from a trajectory collection. It can be used for similarity values when comparing planned routes on board ships, besides visual inspection and leg by leg comparison. Further, it can be used as an indication of possible route deviations in the context of the safety of navigation. This can be extended to creating profiles of various ship categories based on different criteria such as fleet, class and sizes.

Finally, we must state that we expected high similarity due the area characteristics, fleet selection, route possibilities and expected ship behavior. Further, this methodology opens new possibilities of knowledge discovery between diverse groups of ships, such as fleets, which are not observable when considering general ship groups.

## 5. Conclusions

Ship route analysis often considers general route and knowledge extraction in spatial and temporal dimensions. With AIS historical data availability, such a process is simpler than it used to be. However, since the available AIS ship types are limited, analysis for various ship groups has its challenges. With our methodology, we examined route characteristics both in statistical terms and using discrete Fréchet distance for the measurement of similarity for a single container ship fleet. This opens the possibility for the comparison of route creation and execution, not only between similar types of ships, but with fleets as well.

The investigated routes shared similarities in length, duration and average speed; however, there are differences between route legs, where we observed differences in speed profiles. Due to ship size, draft, speed limitations and the avoidance of obstacles in the coastal right whale protection area, ships stayed farther from the coast and outside of the area.

In terms of visual and numerical similarity, we observed that the same ships in different voyages tended to have similar routes. The route similarity between other vessels was also observable, as we expected, due to the relative proximity of, and limited number of potential routes between, the ports. The DFD similarity measure adds value to route interpretation with common values such as route length and duration. We extended the route and similarity analysis from general research on ship categories to fleets, which is not as common. Further, we extended the methods from known software libraries, which were not designed specifically for AIS data analysis, and therefore do not have all the specific functions that could apply when dealing with AIS data. The proposed methodology is based on open-source libraries and data, and can be adapted for further applications. It is our belief that our methodology can improve the reproducibility of similar research approaches and future frameworks.

For future research, we will evaluate longer time periods and datasets to obtain more routes for the comparison of ships in a single fleet and between fleets. We will evaluate methodologies for more distant ports and use other similarity measures. Trajectory analysis will be considered as well, to evaluate the speeds between different route legs.

**Author Contributions:** Conceptualization, D.Š., S.Ž., I.R. and D.B.; methodology, D.Š. and S.Ž.; software, D.Š.; validation, D.Š., S.Ž., I.R. and D.B.; formal analysis, D.Š.: writing—original draft preparation, D.Š. and S.Ž.; writing—review and editing, I.R. and D.B.; visualization, D.Š.; supervision, S.Ž. and I.R.; project administration, S.Ž. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partially funded by European Union’s Horizon Europe under the call HORIZON-CL5-2022-D6-01 (Safe, Resilient Transport and Smart Mobility services for passengers and goods), grant number 101077026, project name SafeNav. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Code and data used in the research are openly available in the Zenodo repository [10.5281/zenodo.7573544] under the license CC BY 4.0.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Šakan, D.; Žuškin, S.; Rudan, I.; Brčić, D. Static Maritime Environment Representation of Electronic Navigational Charts in Global Path Planning. In Proceedings of the 9th International Conference on Maritime Transport, Barcelona, Spain, June 27–28 2022; pp. 1–14. [CrossRef]
- Lee, J.-S.; Son, W.-J.; Lee, H.-T.; Cho, I.-S. Verification of Novel Maritime Route Extraction Using Kernel Density Estimation Analysis with Automatic Identification System Data. *J. Mar. Sci. Eng.* **2020**, *8*, 375. [CrossRef]
- Li, H.; Liu, J.; Wu, K.; Yang, Z.; Liu, R.W.; Xiong, N. Spatio-temporal vessel trajectory clustering based on data mapping and density. *IEEE Access* **2018**, *6*, 58939–58954. [CrossRef]
- Gunnar Aarsæther, K.; Moan, T. Estimating Navigation Patterns from AIS. *J. Navig.* **2009**, *62*, 587–607. [CrossRef]
- Sánchez-Heres, L.F. Simplification and Event Identification for AIS Trajectories: The Equivalent Passage Plan Method. *J. Navig.* **2019**, *72*, 307–320. [CrossRef]
- Nie, P.; Chen, Z.; Xia, N.; Huang, Q.; Li, F. Trajectory similarity analysis with the weight of direction and k-neighborhood for ais data. *ISPRS Int. J. Geo.-Inf.* **2021**, *10*, 757. [CrossRef]
- Ranacher, P.; Tzavella, K. How to compare movement? A review of physical movement similarity measures in geographic information science and beyond. *Cartogr. Geogr. Inf. Sci.* **2014**, *41*, 286–307. [CrossRef]
- Šakan, D.; Rudan, I.; Žuškin, S.; Brčić, D. Near real-time S-AIS: Recent developments and implementation possibilities for global maritime stakeholders. *Pomor.-Sci. J. Mar. Res.* **2018**, *32*, 211–218. [CrossRef]
- Tu, E.; Zhang, G.; Rachmawati, L.; Rajabally, E.; Bin Huang, G. Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey from Data to Methodology. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 1559–1582. [CrossRef]
- International Maritime Organization (IMO). Resolution A.1106(29): Revised Guidelines for the Onboard Operational Use of Shipborne Automatic Identification Systems (AIS), London. 2015. Available online: <https://edocs.imo.org/Final> (accessed on 7 August 2022).
- ITU-R. *Technical Characteristics for An Automatic Identification System Using Time Division Multiple Access in the VHF Maritime Mobile Frequency Band*; M Series Mobile, radiodetermination, amateur and related satellite services; Recommendation ITU M.1371-5: Geneva, Switzerland, 2014; pp. 1371–1375.
- Lu, N.; Liang, M.; Yang, L.; Wang, Y.; Xiong, N.; Liu, R.W. Shape-Based Vessel Trajectory Similarity Computing and Clustering: A Brief Review. In Proceedings of the 2020 5th IEEE International Conference on Big Data Analytics, ICBDA 2020, Harbin, China, 8–11 May 2020; pp. 186–192. [CrossRef]
- Cai, J.; Chen, G.; Lützen, M.; Rytter, N.G.M. A practical AIS-based route library for voyage planning at the pre-fixtured stage. *Ocean Eng.* **2021**, *236*, 109478. [CrossRef]
- Breithaupt, S.A.; Copping, A.; Tagestad, J.; Whiting, J. Maritime Route Delineation using AIS Data from the Atlantic Coast of the US. *J. Navig.* **2017**, *70*, 379–394. [CrossRef]
- Zhang, L.; Meng, Q.; Fang Fwa, T. Big AIS data based spatial-temporal analyses of ship traffic in Singapore port waters. *Transp. Res. E Logist. Transp. Rev.* **2019**, *129*, 287–304. [CrossRef]
- Zhao, L.; Shi, G. A trajectory clustering method based on Douglas-Peucker compression and density for marine traffic pattern recognition. *Ocean Eng.* **2019**, *172*, 456–467. [CrossRef]
- Yan, Z.; Xiao, Y.; Cheng, L.; He, R.; Ruan, X.; Zhou, X.; Li, M.; Bin, R. Exploring AIS data for intelligent maritime routes extraction. *Appl. Ocean Res.* **2020**, *101*, 102271. [CrossRef]

18. Wang, L.; Chen, P.; Chen, L.; Mou, J. Ship AIS Trajectory Clustering: An HDBSCAN-Based Approach. *J. Mar. Sci. Eng.* **2021**, *9*, 566. [[CrossRef](#)]
19. United Nations. *UNCTAD Review of Maritime Transport 2022; Review of Maritime Transport/United Nations Conference on Trade and Development, Geneva*; United Nations: Geneva, Switzerland, 2022; ISBN 978-92-1-002147-0.
20. Ksciuk, J.; Kuhlemann, S.; Tierney, K.; Koberstein, A. Uncertainty in Maritime Ship Routing and Scheduling: A Literature Review. *Eur. J. Oper. Res.* **2022**. [[CrossRef](#)]
21. Tran, N.K.; Haasis, H.-D. Literature Survey of Network Optimization in Container Liner Shipping. *Flex. Serv. Manuf. J.* **2015**, *27*, 139–179. [[CrossRef](#)]
22. Meng, Q.; Wang, S.; Andersson, H.; Thun, K. Containership Routing and Scheduling in Liner Shipping: Overview and Future Research Directions. *Transp. Sci.* **2014**, *48*, 265–280. [[CrossRef](#)]
23. Wang, S.; Meng, Q. Sailing Speed Optimization for Container Ships in a Liner Shipping Network. *Transp. Res. Part E Logist. Transp. Rev.* **2012**, *48*, 701–714. [[CrossRef](#)]
24. Gao, C.-F.; Hu, Z.-H. Speed Optimization for Container Ship Fleet Deployment Considering Fuel Consumption. *Sustainability* **2021**, *13*, 5242. [[CrossRef](#)]
25. Li, X.; Sun, B.; Jin, J.; Ding, J. Speed Optimization of Container Ship Considering Route Segmentation and Weather Data Loading: Turning Point-Time Segmentation Method. *J. Mar. Sci. Eng.* **2022**, *10*, 1835. [[CrossRef](#)]
26. Gao, X.; Makino, H.; Furusho, M. Ship Behavior Analysis for Real Operating of Container Ships Using AIS Data. *TransNav Int. J. Mar. Navig. Saf. Sea Transp.* **2016**, *10*, 213–220. [[CrossRef](#)]
27. Ikonomakis, A.; Nielsen, U.D.; Holst, K.K.; Dietz, J.; Galeazzi, R. How Good Is the STW Sensor? An Account from a Larger Shipping Company. *J. Mar. Sci. Eng.* **2021**, *9*, 465. [[CrossRef](#)]
28. Oh, M.-J.; Roh, M.-I.; Park, S.-W.; Chun, D.-H.; Son, M.-J.; Lee, J.-Y. Operational Analysis of Container Ships by Using Maritime Big Data. *J. Mar. Sci. Eng.* **2021**, *9*, 438. [[CrossRef](#)]
29. Botter, R.; Santos, T.; Soares, C.G. Characterizing Container Ship Traffic along the Portuguese Coast Using Big Data. In *Progress in Maritime Technology and Engineering*; CRC Press: Boca Raton, FL, USA, 2018; pp. 93–100.
30. Shu, Y.; Daamen, W.; Ligteringen, H.; Hoogendoorn, S.P. AIS Data Based Vessel Speed, Course and Path Analysis in the Botlek Area in the Port of Rotterdam. In Proceedings of the International Workshop on Next Generation of Nautical Traffic Model, Shanghai, China, 21–24 September 2012.
31. Xiao, F.; Ligteringen, H.; van Gulijk, C.; Ale, B. Comparison Study on AIS Data of Ship Traffic Behavior. *Ocean. Eng.* **2015**, *95*, 84–93. [[CrossRef](#)]
32. BOEM; NOAA. MarineCadastre.gov., Automatic Identification System Data 2019 from 2019-01-01 to 2019-06-30. for the Area of Interest [(31.791, −81.224), (33.102, −79.453)]. 2022. Available online: <https://marinecadastre.gov/accessais/> (accessed on 17 July 2022).
33. G. F. Department. Georgia Ports Authority Comprehensive Annual Financial Report For the Fiscal Years Ended June 30, 2019 and 2018, Savannah, GA 31402 USA, 2019. Available online: <https://gaports.com/wp-content/uploads/2021/10/CAFR-FY19-Final.pdf> (accessed on 19 July 2022).
34. Comprehensive Annual Financial Report for Fiscal Year Ended June 30, 2019, Mount Pleasant, South Carolina 29464, USA, 2019. Available online: <https://scspa.com/wp-content/uploads/cafr-fy19-final.pdf> (accessed on 19 July 2022).
35. National Oceanic and Atmospheric Administration (NOAA). *Distances Between United States Ports 2019*, 13th ed.; NOAA: Washington, DC, USA, 2019; p. 10.
36. Raimondo, G.M.; Spinrad, R.W.; Leboeuf, N.R. *National Oceanic and Atmospheric Administration (NOAA) National Ocean Service Coast Pilot 4 Atlantic Coast: Cape Henry, Virginia to Key West, Florida*, 53rd ed.; NOAA: Washington, DC, USA, 2021. Available online: [https://nauticalcharts.noaa.gov/publications/coast-pilot/files/cp4/CPB4\\_WEB.pdf](https://nauticalcharts.noaa.gov/publications/coast-pilot/files/cp4/CPB4_WEB.pdf) (accessed on 1 August 2022).
37. AIS Vessel Transit Counts 2019 | InPort. Available online: <https://www.fisheries.noaa.gov/inport/item/61037> (accessed on 19 July 2022).
38. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June 28–3 July 2010; Volume 445, pp. 56–61. [[CrossRef](#)]
39. Reback, J.; McKinney, W.; den Bossche, J.V.; Augspurger, T.; Roeschke, M.; Hawkins, S.; Cloud, P.; jbrockmendel; gyoung; Sinhrks999; et al. *Pandas-Dev/Pandas*; Pandas 1.4.2; Zenodo, 2022. [[CrossRef](#)]
40. Gillies, S.; van der Wel, C.; Van den Bossche, J.; Taves, M.W.; Arnott, J.; Ward, B.C. Shapely (2.0). Zenodo, 2022. [[CrossRef](#)]
41. Harris, C.R.; Millman, K.J.; Van Der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [[CrossRef](#)] [[PubMed](#)]
42. Graser, A. MovingPandas: Efficient Structures for Movement Data in Python. *GI Forum* **2019**, *1*, 54–68. [[CrossRef](#)]
43. Graser, A.; Bell, R.; Boiko, G.P.; Parnell, A.; Vladimirov, L.; Theodoropoulos, G.S.; Richter, G.; Elcinto, J.B.; Rodriguez, J.L.C.; Lovelace, R.; et al. *Anitagraser/Movingpandas*; v0.9.rc3; Zenodo, 2022. [[CrossRef](#)]
44. Jordahl, K.; den Bossche, J.V.; Fleischmann, M.; McBride, J.; Wasserman, J.; Badaracco, A.G.; Gerard, J.; Snow, A.D.; Tratner, J.; Perry, M.; et al. *Geopandas/Geopandas*; v0.10.2; Zenodo, 2021. [[CrossRef](#)]
45. Rudiger, P.; Stevens, J.-L.; Bednar, J.A.; Nijholt, B.; Mease, J.; Andrew; Liquet, M.; B, C.; Randelhoff, A.; Tenner, V.; et al. *Holoviz/Holoviews*; Version 1.14.9; Zenodo, 2022. [[CrossRef](#)]

46. MarineCadastre.gov, (NOAA). AIS Vessel Type and Group Codes Used by the Marine Cadastre Project 2018. Available online: <https://coast.noaa.gov/data/marinecadastre/ais/VesselTypeCodes2018.pdf/> (accessed on 2 September 2022).
47. Eiter, T.; Mannila, H. *Computing Discrete Fréchet Distance*; Technical Report CD-TR 94/64; Technische Universität Wien: Wien, Austria, 1994.
48. Tang, B.; Yiu, M.L.; Mouratidis, K.; Zhang, J.; Wang, K. On discovering motifs and frequent patterns in spatial trajectories with discrete Fréchet distance. *Geoinformatica* **2022**, *26*, 29–66. [[CrossRef](#)]
49. Jekel, C.F.; Venter, G.; Venter, M.P.; Stander, N.; Haftka, R.T. Similarity measures for identifying material parameters from hysteresis loops using inverse analysis. *Int. J. Mater. Form.* **2019**, *12*, 355–378. [[CrossRef](#)]
50. Agarwal, P.K.; Avraham, R.B.; Kaplan, H.; Sharir, M. Computing the discrete Fréchet distance in subquadratic time. *SIAM J. Comput.* **2014**, *43*, 429–449. [[CrossRef](#)]
51. Jekel, C.F. Similaritymeasures PyPI. Available online: [https://github.com/cjekel/similarity\\_measures](https://github.com/cjekel/similarity_measures) (accessed on 21 July 2022).
52. Piovesana, A.; Senior, G. How Small Is Big: Sample Size and Skewness. *Assessment* **2018**, *25*, 793–800. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.