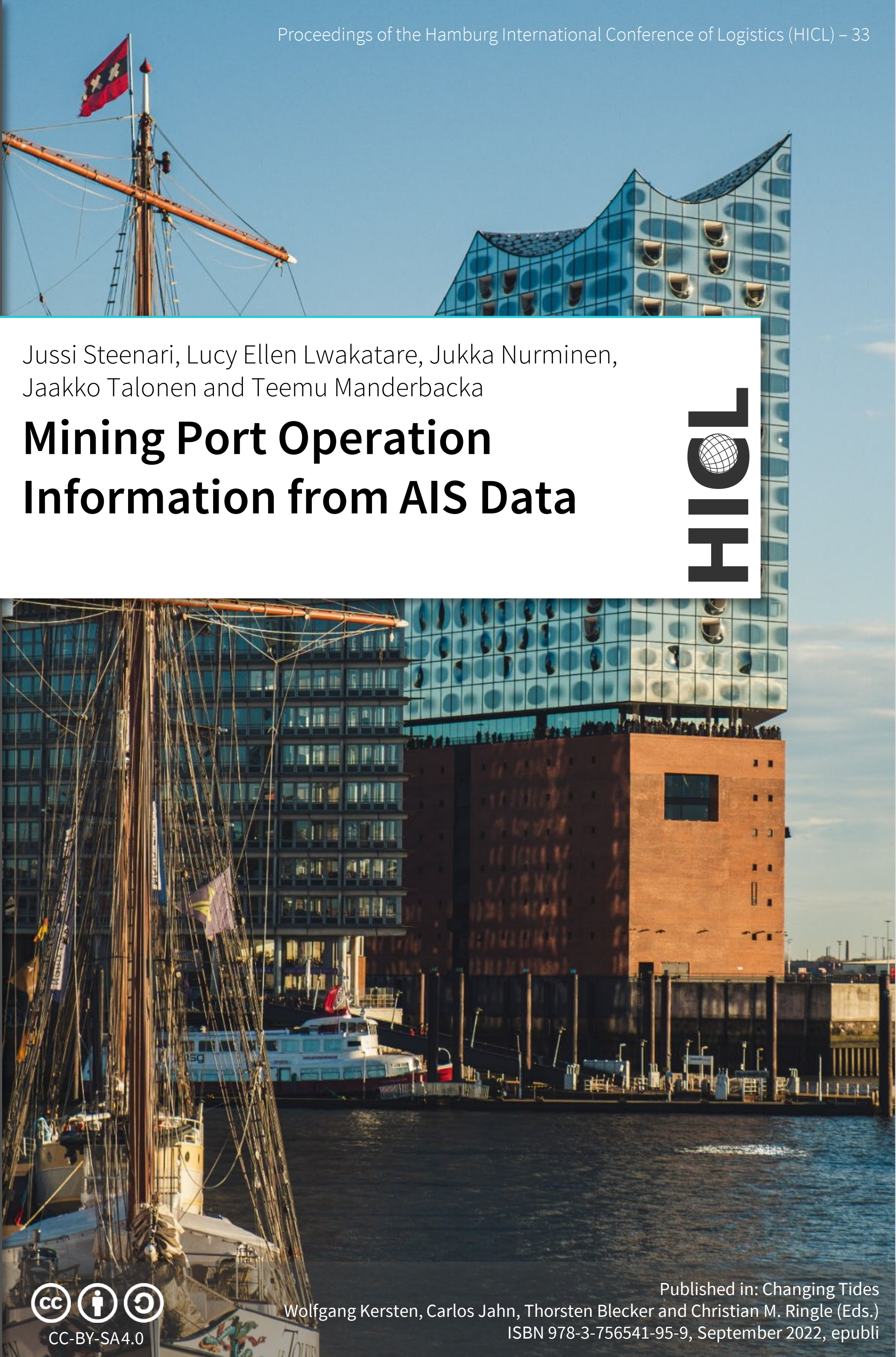
Proceedings of the Hamburg International Conference of Logistics (HICL) – 33



Jussi Steenari, Lucy Ellen Lwakatare, Jukka Nurminen, Jaakko Talonen and Teemu Manderbacka

Mining Port Operation Information from AIS Data

Published in: Changing Tides
Wolfgang Kersten, Carlos Jahn, Thorsten Blecker and Christian M. Ringle (Eds.)
ISBN 978-3-756541-95-9, September 2022, epubli

Mining Port Operation Information from AIS Data

Jussi Steenari [1], Lucy Ellen Lwakatare[1], Jukka Nurminen[1], Jaakko Talonen[2] and Teemu Manderbacka[3]

1 – University of Helsinki 2 – NAPA Ltd.

3 – VTT Technical Research Centre of Finland Ltd.

Purpose: *Ports play a vital role in global trade and commerce. While there is an abundance of analytical studies related to ship operations, less work is available about port operations and infrastructure. Information about them can be complicated and expensive to acquire, especially when done manually. We use an analytical machine learning approach on Automatic Identification System (AIS) data to understand how ports operate.*

Methodology: *This paper uses the DBSCAN algorithm on AIS data gathered near the Port of Brest, France to detect clusters representing the port's mooring areas. In addition, exploratory data analyses are per formed on these clusters to gain additional insights into the port infrastructure and operations.*

Findings*: From Port of Brest, our experiment results identified seven clusters that had defining characteristics, which allowed them to be identified, for example, as dry docks. The clusters created by our approach appear to be situated in the correct places in the port area when inspected visually.*

Originality: *This paper presents a novel approach to detecting potential mooring areas and how to analyse characteristics of the mooring areas. Similar clustering methods have been used to detect anchoring spots, but this study provides a new approach to getting information on the clusters.*

First received: 17. Mar 2022 Revised: 22. Aug 2022 Accepted: 25. Aug 2022

1 Introduction

Efficient maritime logistics is extremely valuable and can be improved by adequate planning of operations. This requires up-to-date and complete information on the state of the global fleet; including details on destination port and terminal of other ships on the voyage, their estimated times of arrival (ETAs), number of ships waiting to enter ports, either drifting or at anchor, port capacities (how many ships and at which pace the terminals can handle). Generally, the operator has information and knowledge of the fleet under its control; however, the past operations of the global and competing fleet are difficult to gather.

Global shipping is a game of several stakeholders. Vessels and terminal infrastructures are expensive. Ports are owned or controlled by cities or countries renting areas for companies operating terminals, with their loading and offloading equipment and berth arrangements. The different stakeholders involved have their own interests in the business, where some are collaborating and the others competing with each other (Stopford, 2009). There is no single or even one publicly open data system on past port and, more specifically, terminal calls.

Some ship types generally operate on scheduled liner traffic, such as passenger and container vessels. Whereas the most significant proportion of ship types, bulkers (both dry bulk and tankers), generally operate on the spot market, i.e., voyages are chartered individually for different cargo owners according to the market requirements. For an operating company, it is valuable to be able to know transportation patterns and situations on a global level. The AIS data reveals the location of all the vessels in the global fleet, but meaningful information on the location can be obtained only when sufficient data on port, terminal and berth locations are combined with it.

A difficulty in maritime logistics planning is knowing how the ports operate: when are they busy, what kind of services they provide, what is their capacity, etc. Some of this information can be found by contacting the port directly. However, it is not an option when we think of the planning of international supply chains where the language, continuously changing situations, and other things form barriers. Looking at the web material, e.g. the port of Brest (which has been the focus of our study), shows that it is difficult to find all the relevant information.

Although AIS data has primarily helped increase maritime safety (Silveira et al., 2015; Montewka et al., 2022), it can be used for other purposes, such as understanding port operations. For the latter, there is a need to understand fleet or even single vessel operations more accurately. It isn't even possible to know the port of departure without additional data sources. World Port Index (WPI) (National Geospatial-Intelligence Agency, 2019), and United Nations Economic Commission for Europe (UNECE) provide data sets containing global port locations. The United Nations Code for Trade and Transport Locations (UN/LOCODE) includes over 100,000 locations (UNECE Trade Division, 2021). This makes it possible to convert raw AIS data to organized and structured facts, e.g. voyages.

In this paper, we investigate the use of AIS data to understand port operations. Since 2004 all passenger ships, and all ships over the size of 300 gross tonnage on international voyages, and cargo ships over 500 gross tonnage on national and international operation are required to be equipped with an AIS transponder (International Maritime Organization, 2004) according to the SOLAS regulations (International Maritime Organization, 2014) set forth by the International Maritime Organization (IMO). Using primarily the location data of the AIS messages, which are not always very reliable (Silveira et al., 2015), we can e.g. detect the mooring areas in a harbor area, the port opening hours (via the arrival and leave times of vessels), the peak hours harbour load, and typical times different types of vessels spend at port.

Given the amount of AIS data is vast, efficient algorithms to filter, aggregate, and analyze the data are needed. This paper describes the algorithms we use to analyze the port operations and illustrates their possibilities by studying the functions of the Port of Brest in France. The port area of Brest is geometrically quite simple, all the berths are located on one shore and they are mostly oriented in a south-north axis. This makes the area quite uniform and thus easy to cluster compared to, for example, the port of Rotterdam, which is a labyrinthine collection of canals where terminals are located close to each other. In future work, we plan to implement this approach to other port areas by either tailoring the methods to the specific port area or creating a general approach that works on all port areas.

2 Related work

AIS data analysis is conducted in numerous studies to extract useful information regarding vessel movement (Zhang et al., 2020) as well as port operational patterns and performance statistics (Millefiori et al., 2016). However, to the best of our knowledge, very few studies have used AIS data to benchmark ports with various performance metrics, such as operational capacity and vessel time at anchorages (Millefiori et al., 2016). In the latter studies, AIS data is used to gain insights into specific port operations, such as bunkering (Fuentes, 2021), understand ship maneuvering (Lee et al., 2021) and congestion (Rajabi et al., 2018) in ports, and identify ports' locations and operational boundaries (Millefiori et al., 2016).

Millefiori et al., 2016 proposed a methodology that uses AIS data to define the exact seaport location and its operational boundaries. According to the authors (Millefiori et al., 2016), an accurate definition of a port's location and operational boundaries using a data-driven approach is essential when calculating the port's capacity and efficiency. The proposed method was applied to a dataset of more than 57 million AIS messages focusing on the port of Shanghai.

Lee et al., 2021 studied ship maneuvering in port based on AIS data of vessels arriving at and departing from Busan New Port in Korea collected for four months. From this data, the authors analyzed predominant ship trajectory patterns using DBSCAN algorithm. Given that accidents involving ship collisions with terminals or gantry crane collision can occur, their results were useful when developing port maneuvering guidelines.

Rajabi et al., 2018 studied how the port of Le Havre operates by analyzing AIS data recorded over a one-year interval. Specifically, they analyzed AIS data to determine the number of different types of vessels at each terminal. The authors identified the terminal that served most of the port traffic, and popular quay positions for different types of vessels. According to the authors(Rajabi et al., 2018), since the allocation of vessels at the berth is subject to the rules between the port and shipping companies, this practice results in a high concentration of vessels in some wharves compared to others. Furthermore, it was noticed that the traffic of small vessels increased continuously, which increased the waiting time of big vessels .

Compared to the prior studies, our work addresses some limitations related to the lacking description of the methodology used to study port operations (Rajabi et al., 2018). In addition, this work builds on the preceding study (Lee et al., 2021) to showcase the use of a similar approach (DBSCAN algorithm) to analyse port operations based on AIS data relevant for understanding port infrastructure.

3 Methods

Clustering algorithms have been used in data mining to extract hidden and interesting patterns from massive datasets. Specifically, density-based spatial clustering of applications with noise (DBSCAN) is useful in determining arbitrary shaped clusters in spatial databases that contain noisy data (Khan et al., 2014). DBSCAN was first proposed by Ester et al., 1996 and since then the algorithm has been used and improved extensively(Khan et al., 2014). In short, DBSCAN works by clustering points based on two parameters, minimum points and maximum distance, also called epsilon. The minimum points indicate the minimum number of points that need to be within epsilon distance of a point for it to be classified as a core point for a cluster. Points found this way are further tested if they have enough points within epsilon distance for them to be classified a core point. If not enough points are found they are classified as a border point of the cluster. This process repeats until each point is part of a cluster as a core or border point or classified as noise.

3.1 Detecting Port Mooring Areas with DBSCAN

The clustering is done on points derived from the AIS data. The clustering points are calculated by taking the median coordinates from each mooring event. A mooring event is a continuous time series of AIS messages from a single ship that has the navigational status ”moored” attached. Mooring areas are defined as either berths or quays. A berth is a place for a single ship to moor, and a quay is a structure on the shore of a harbor that can contain several berths.

In our experiment the minimum points parameter for the DBSCAN algorithm is configured to be three. This parameter is chosen since it is a minimum amount of points to form a polygon shape. Additionally setting the parameter to three decreases the probability of false positives in the data compared to having the parameter to be two.

In deciding the most optimal epsilon parameter, two things have to be taken into account. First, the epsilon parameter is expressed in coordinate degrees. This means that the epsilon corresponds to different lengths in different latitudes. A degree of change in latitude corresponds to the same distance at all latitudes. However, the distance covered by a degree of longitude changes drastically the closer to the poles the points are. To mitigating this transformation can be done by either doing a equidistant projection change for the coordinates, or configuring the DBSCAN algorithm to use the Haversine distance. The latter will increases the computational complexity, but this is negligible if relatively few points are used in the clustering process.

The second thing to take into account when selecting the epsilon is to decide whether to detect individual berths or complete quays. The detection of quays with several berths is relatively straightforward and can be easily fine-tuned by changing the parameters. However, this method produces rather large clusters that can contain multiple berths. This, in turn, can affect the analysis of the clusters if berths with differing characteristics are clustered together into a single cluster. Detecting smaller clusters would mean that the epsilon has to be defined much more robustly to detect smaller clusters. This would also most likely result in loss of certain areas, when enough points are not available in a given area. Most likely in this approach the selection of epsilon configured in a way that takes into account different positions of the transponders aboard the ships. Also, given the smaller clusters have less data points each, making meaningful analysis would be more difficult using smaller clusters.

In this study the validation of formed clusters was done manually, i.e. the clusters were overlaid on a map and checked for their location and size. The small area encompassing the dataset makes this approach feasible, but for larger datasets and automatic processes this issue would need to be streamlined. A robust validation method for the clusters would also allow for more nuanced selection of the epsilon and minimum points parameters.

4 Experiments

The experiment will be run by following the framework depicted in Figure 1. The process starts by doing preprocessing steps for the raw AIS data to both improve the quality of the data as well as improving the computational performance in the following steps. After preprocessing individual mooring events are detected and median coordinates of each mooring event are extracted to use in the clustering step. Since each mooring event is expected to be stationary (which they sometimes are not in real data), this condenses the data to easily clusterable form. These points are then clustered with the DBSCAN algorithm to produce clusters. The clusters are then combined with the processed AIS data to determine which cluster each AIS data point belongs to. Finally this combined dataset is used to analyze certain aspects of the port area.

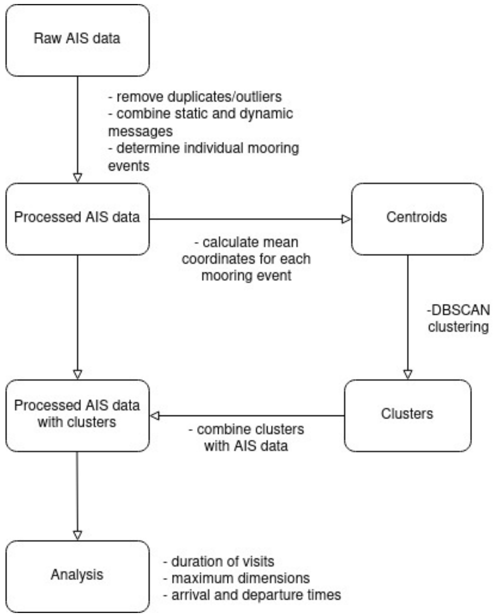


Figure 1: Framework of the experiment

5 Data description

The analysis methods were applied to the AIS dataset collected from areas near the port of Brest in France (Ray et al., 2018). The data covers a time span of six months, from October 1st, 2015 to March 31st, 2016. The data includes both the static and dynamic AIS messages. The dataset also includes the data provided by the WPI that contains geographical information, such as the coordinates of each port. This can be used to filter points from the data set to only contain coordinates near a specific port.

The dynamic messages in the dataset contain a total of 19,035,630 rows which are collected from 5055 different Maritime Mobile Service Identity (MMSI) numbers. The MMSI number is an identifier to distinguish different ships from each other. However, while at a given time an MMSI should be linked to a single ship, over time the MMSI can be reused between different ships, and a ship might change its MMSI for numerous reasons. So while the MMSI is not strictly speaking a unique identifier, in this project this is used as an indicator of individual ships. There also exists an immutable identifier called IMO number that is unique for each ship. However this identifier is missing from smaller ships and it needs to be manually inserted to the system, thus making it susceptible to data errors.

5.1 Data preparation

The AIS data is known to contain a large amount of data errors (Emmens et al., 2021). Before running the DBSCAN algorithm on the data, steps are taken to clean and filter the data. First, all the (ship-id, timestamp) duplicates are removed from the data. Second, new columns are created by parsing the time stamp to its components such as the time of day and the day of the week. Third, all the points outside a certain radius of the port to be analyzed are removed from the data. This radius is centred on the port coordinates from the WPI data set. Finally, the speed between sequential points is calculated. If a point has navigational status set as ”moored” and a speed above a certain threshold, these points are removed from the data.

The dynamic messages need to be combined with the static messages to complete some fields. This is done by using the MMSI number as a key for a joining operation. The following fields are added to the dynamic data: ship type, ship length, ship width and ship draft. Since the static messages are transmitted at six minute intervals, each MMSI number has multiple static messages. This means that some of the above mentioned fields can have multiple values for a single MMSI. Some of these changes can be attributed to ship operations, such as the ship’s draft changing after loading or unloading cargo, but most of these changes can not be attributed to anything else than data errors. The static and dynamic messages are joined using an as-of merge operation, which joins the rows on closest key values, in this case timestamps.

The clustering will be done on two different data sets, one with all the ship types and a second that has only specific ship types included. The thought process is that by filtering the data beforehand, the analysis can focus on specific clusters more easily. The risk in this implementation is that some ships might have their ship type configured incorrectly and thus would not be included in the filtered data. The clusters produced by the unfiltered data can be compared with the filtered data to see if any major clusters are missed with the reduced data set.

The clustering points are generated by identifying individual mooring visits. A mooring visit is defined by a ship having the navigational status ”moored” for continuous timestamps. Since the data also contains some incorrect navigational statuses, only mooring visits of that last over an hour are included in the clustering. From each of these mooring visits, the median coordinates are taken as a clustering point.

5.2 Application of DBSCAN Algorithm

The prepared AIS data described in Section 4.2 was entered as a variable to DBSCAN algorithm employed from Scikit-learn in Python. In this study, two different epsilon parameters were used on the DBSCAN algorithm. A larger epsilon is used to detect whole quays that can contain multiple berths. For detecting individual berths, a smaller Epsilon parameter is defined. The smaller epsilon is defined to be 50 meters while the larger epsilon parameter is set to 100 meters. Further, the dataset variable is filtered to only contain the preprocessed data, or the preprocessed data with selected ship types, in our case cargo ships and tankers.

The resulting clusters with all ship types is presented in Figure 2 and Figure 3. In the figures clusters created by all of the ship types are marked as red polygons. Clusters created by data set containing only tanker ships have green outlines and clusters created by data set containing only cargo ships have blue outlines. When visual analysis is done on these clusters it would seem that all the tanker and cargo ships are contained within the clusters made by all the data. In other words filtering the data did not result in clusters in new areas. Also, we take note that when increasing the Epsilon parameter, this also does not produce clusters in new areas. In some cases the larger epsilon combined existing clusters while in some cases it produced larger clusters. A smaller Epsilon parameter is capable of producing more granular clusters and thus can bring about more precise analysis of the mooring areas. However, a larger Epsilon parameter produces clusters that have a bigger area, which can also bring new aspects to the analysis.

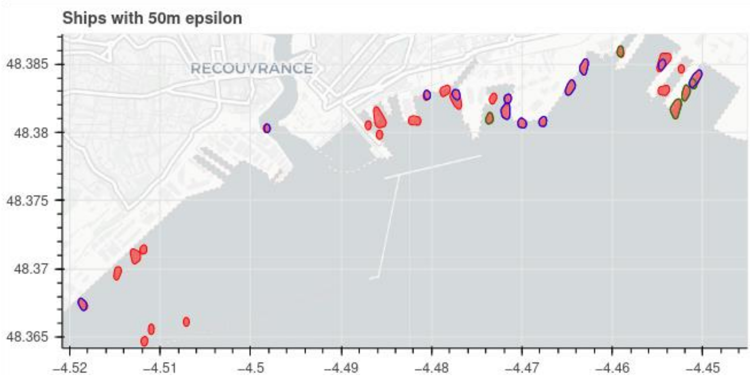


Figure 2: DBSCAN clusters created with 50 m epsilon

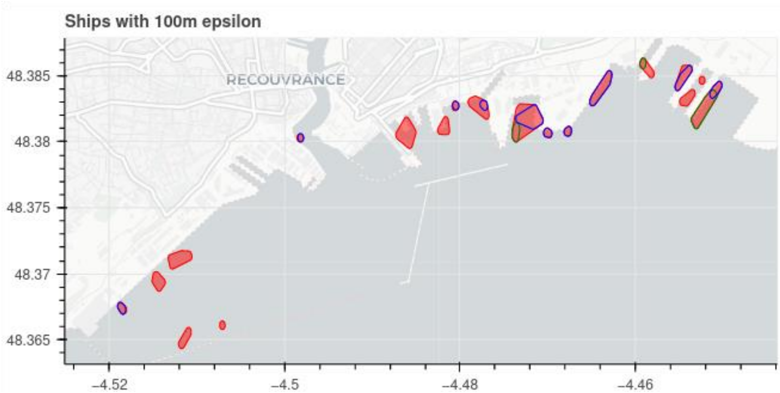


Figure 3: DBSCAN clusters created with 100 m epsilon

A too large Epsilon can produce clusters that combine multiple distinct clusters. There is also a risk that a too large Epsilon can detect clusters that go over land areas or quays. An example of an cluster where the epsilon is too big is given in Figure 4. However with a too small epsilon, bigger ships can produce two clusters when the AIS transponder is situated at a different parts of the ship or when different ships are moored with opposite directions. An example of a too small epsilon is demonstrated in Figure 3. In this picture the clusters are clearly in an area that is meant for mooring a single ship. It appears that ships moor in this place either their bow towards east or west, producing two clusters when in fact it is a single berth. Some of these issues can be counteracted by filtering the data more thoroughly. For example the cluster in Figure 5 does not go over the quay if the data is filtered to include only tanker ships as seen in Figure 6.

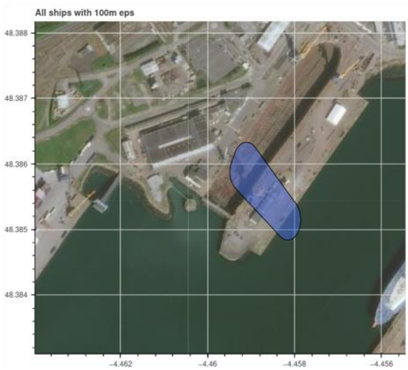


Figure 4: Cluster going over a quay when epsilon is too large

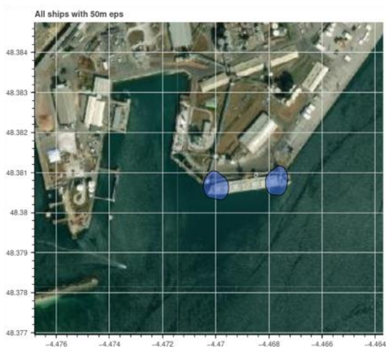


Figure 5: Dividing a single cluster in two when the epsilon is too small

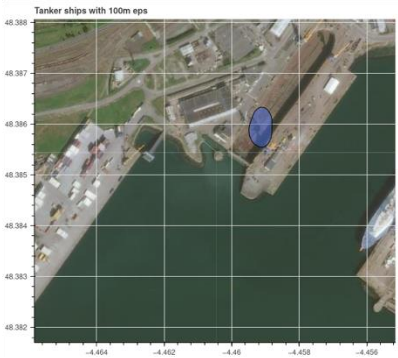


Figure 6: Cluster created after filtering to include just tanker ships

Further analysis can be done on the generated clusters to gain additional insights on the identified terminals. For example, dimensional characteristics for To achieve this, the cluster data is added to the original unfiltered AIS messages. This is done by checking each AIS message whether the coordinates are inside a given cluster. This can be done relatively fast if the messages are coded to a geospatial dataframe e.g. such as GeoPandas, thus allowing the use of spatial clustering.

5.3 Analysis of the Clusters

Further analysis was performed on selected clusters to find the following statistics for each cluster: number of visits, number of unique ships, maximum dimensions of visiting ships, mean time spend in cluster and the hours of the day that ships are entering or leaving the clusters. Additionally, a breakdown of the different types of ships that visit each cluster is calculated. The breakdown of different ship types can show whether a terminal is meant to service just one type of ship or multiple ship types such as in case of maintenance areas.

For number of visits, an individual cluster visit is defined as a continuous time series of AIS messages from a single ship that are positioning inside a cluster. The number of visits and number of unique ships can be used to determine terminals that have high traffic.

The dimensional data is collected from a separate data set that contain the static AIS messages. The dimensional data include the length and beam (width) of a ship as well as its draft. The ship dimensions tell what kind of ships each terminal can service. The amount of time spend in a cluster needs to be calculated by subtracting the arrival timestamp from the departing timestamp for each individual cluster visit.

6 Results

In this results section several clusters will be analysed to get further insight on the data. The clusters are chosen based on showing some interesting metrics that call for additional visual analysis. The visual analysis is done using aerial and satellite images provided by ESRI (ESRI, 2022). Some clusters showed interesting characteristics, but they were situated in an area that is pixelated in satellite images due to close proximity to military installations. This prevents further visual analysis of the cluster and because of this these clusters were not included into the analysis.

The analysis was done on the clusters formed by using the whole data set with a small epsilon. This was as the preferred approach since this minimizes the amount of overlapping clusters, while also at the same time gives insight on various aspects on the port infrastructure. The results of the chosen clusters are shown in Table 1 and Table 2. The AIS points are filtered to include just tankers and cargo ships (ship type numbers 70 to 89). This is done to reduce the noise in the data. Preliminary analysis indicates that some ships such as tugs have irregular and movement patterns that can affect the analysis. The AIS data is spatially joined with the clusters to include which cluster each point belongs to. Insights are analyzed from the data to study which clusters have ships staying in them for extended periods of time. In this data set the longest median visiting times are in cluster 28 (median time 14 days 21:31), cluster 27 (median time 7 days 12:00). These clusters are visualized in Figure 7. Both of these clusters appear to be situated in dry docks.

These findings would indicate that by looking for clusters that have unusually long stays in them, it could be possible to detect areas that have some sort of maintenance capacity such as dry docks.

Table 1: Analysis of Selected Clusters

Cluster ID	Berth type	Data subset	N ships	N visitits	Max Length (m)	Max draft (m)
27	Dry dock	Tanker/cargo	5	34	291.0	9.8
28	Dry dock	Tanker/cargo	5	6	291.0	8.6
5	Undefi-ned	All	20	43	176.0	8.8
10	Undefi-ned	All	17	45	136.0	7.1
1	Undefi-ned	All	18	106	136.0	6.5

Table 2: Median time in cluster

Cluster ID	Median time in cluster
27	7 days 12:00
28	14 days 21:31
5	0 days 12:10
10	3 days 10:40
1	0 days 13:19

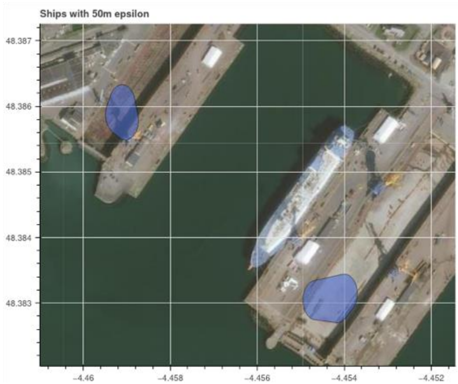


Figure 7: Cluster 27 (bottom right) and 28 (top left)

Another interesting metric to analyse is to see how many different kinds of ships visit each cluster. The ship types are expressed as integers that have two digits, where the first digit indicates the type of ship (Marine Traffic, 2022). The second integer can be used for further specification for some ship types such as the type of cargo the ship is carrying. In analysing these clusters the focus will be mainly on the ship type so for example all cargo ships are grouped into one group. For analyzing this aspect all the ship types are included to the AIS data. The data is then spatially joined with the cluster data similarly as in the previous analysis. The analysis found that clusters with significant distribution of different ship types are: cluster 1 (56% other, 37% tugs, 7% missing value), cluster 5 (75% cargo ships, 25% tanker ships), and cluster 10 (36% tugs, 25% cargo ships, 38% other, 1% missing).

Clusters 1, 5 and 10 are visualized in Figure 8. All of these clusters are situated in areas near dry docks. Additionally clusters 1 and 10 are situated adjacent to each other. This in combination with the variety of ship types visiting these clusters could potentially suggest that these areas have some connection to the dry dock areas. This area could be a waiting are to enter the dry docks or some sort of maintenance could be performed here that does not require the dock to be drained. Alternatively based on the distribution of ship types (lots of ships classified as tugs or ’other’ types) clusters 1 and 10 could be used as a waiting area for ships that help with operations related to the dry dock. Cluster 5 has only tanker and cargo ships so most likely this place is reserved for waiting to enter the dry docks. The time spend in clusters 1 and 5 would indicate that these are for short time mooring needs, such as temporary mooring for maintenance vessels. The time spend in cluster 10 is much longer, indicating that this cluster is also used for longer time mooring needs. So by analysing distribution of different ship types it might be possible to detect maintenance areas near other points of interest such as dry docks. Most notably the dry dock area above clusters 1 and 10 did not produce a cluster from the data. This could be because the AIS transponder was configured incorrectly on the visiting ships or there were no points in this dry dock in the given time frame. In any case analysing these areas that have a high distribution of different ship types might give an indication that there is potentially a dry dock area nearby. Alternatively just the presence of tugs in clusters could be enough to detect areas used for maintenance. In any case this hypothesis would need to be further tested to see if this analysis can give additional insight on the port infrastructure.

By analysing the clusters created by DBSCAN algorithm we can get some insights on the infrastructure of the port. It appears that areas that harbor some sort of dry dock capabilities are relatively easy to identify due to them having significantly longer stays within the clusters. Also by detecting areas that have high distribution of different ship types it might be possible to detect areas that are related to port maintenance.

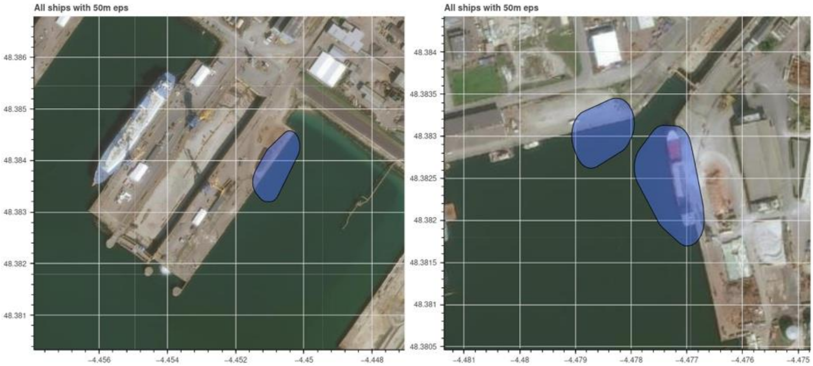


Figure 8: Cluster 5 (left image), 1 (leftmost cluster in right image) & 10 (rightmost cluster in right image)

7 Discussion

The method for mooring areas (berths/quays) identification by aggregating the AIS data were described in Section 3.1. This method solves one of the challenges - is vessel at port or not. When open data sets such as UNECE and WPI are used, the data is not up-to-date. New ports and terminals are being constantly built and old ones discontinued in use. Secondly, location accuracy is rather poor in UNECE where the two last digits refer to minutes and the two or three first digits indicate the degrees. DBSCAN makes it possible to detect and keep automatically mooring areas up-to-date, globally.

In this article we have produced a proof-of-concept approach on how to detect mooring areas based on just AIS data. The DBSCAN algorithm was able to perform quite well even given the noisiness of AIS data. While the initial approach shows promise, there are still some issues to overcome. Firstly the selection of optimal parameters is still done by a trial and error approach. While the chosen parameters uncover interesting insights from Brest harbor area, there is no guarantee that these parameters would work in other port areas. An automated approach to this would need some sort of cluster validation for evaluating the parameters. While some methods have been developed for optimal selection of DBSCAN parameters, these are still pretty much in experimental stages and would most likely not work with data as noisy as AIS messages (Starczewski, Goetzen and Er, 2020). However while it is unlikely that parameters exists that work for all port areas, we are developing methods that would allow us to select optimal DBSCAN parameters for a selected port area.

8 Conclusions

Presented method makes it possible to deduce detailed information on individual berth locations and their capacities covering all global ports and terminals. Such information enables to build up a complete picture of the ongoing operations. As different ports have different number and combination of terminals and berths attending different ship types, it is important to know these capacities, in order to define and predict potential congestion at port and to estimate time for departure after cargo operations. Manually entered fields of port of destination and estimated time of arrival, and on ship status; moored, anchored, or steaming are very useful but often unreliable or updated late due to required human input.

Estimates and predictions of waiting times for ships entering to port can be done when the capacities of the port are defined. The individual berth information together with other ship port calls enables to identify operational patterns. One port may have several terminals attending one ship type. However, these terminals with their own berths may have contracts with different operators or cargo owners, which can be revealed by analysis of operational patterns. Such analysis could be conducted with help of berth information that could be gathered globally with application of the presented method to global AIS data. The importance of detailed location information on berth level manifests especially in case of the busiest ports. These ports are often located in the areas with close passing traffic and they contain a very large number of berths as a more or less continuous line along the coast. The passing traffic often needs to slow down and sometimes the ships stop for bunkering on nearby locations. Without accurate enough berth data, which the publicly available databases do not offer, is would not be possible to define if the ship had really entered the port or not.

The analysis of the mooring area clusters gave some interesting insights, but it might be too early to tell if these methods can be generalized to be applicable to other data sets. In this case further studies are needed to see if these methods are sound. Further addition to the analysis phase could include things such as detecting whether the draft of a ship changes inside a cluster, detecting linkages between anchorage areas and mooring areas and detecting how long on average ships wait to enter a given mooring area. These could give additional benefit to the analysis and produce results that go beyond just detecting the structure of the port.

Acknowledgements

This work was partly funded by the VesselAI project (Horizon 2020 Framework Programme of the European Union under grant agreement No 957237)

References

Emmens, T., Amrit, C., Abdi, A. and Ghosh, M., 2021. The promises and perils of Automatic Identification System data. Expert Systems with Applications, 178, p.114975.

ESRI, 2022. ESRI. [online] Available at: <https://esri.maps.arcgis.com/home/index.html>.[Accessed: 2022-01-10]

Ester, M., Kriegel, H.-P., Sander, J., Xu, X. and others, 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: kdd. pp.226–231.

Fuentes, G., 2021. Generating bunkering statistics from AIS data: A machine learning approach. Transportation Research Part E: Logistics and Transportation Review, 155, p.102495.

International Maritime Organization, 2004. Guidelines for the installation of a shipborne Automatic Identification System (AIS). [online] Available at: <https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/AIS/SN.1-Circ.227.pdf>. [Accessed: 2022-01-07]

International Maritime Organization, 2014. SOLAS Consolidated edition.

Khan, K., Rehman, S.U., Aziz, K., Fong, S. and Sarasvady, S., 2014. DBSCAN: Past, present and future. In: 5th international conference on the applications of digital information and web technologies. pp.232–238.

Lee, H.-T., Lee, J.-S., Yang, H. and Cho, I.-S., 2021. An AIS Data-Driven Approach to Analyze the Pattern of Ship Trajectories in Ports Using the DBSCAN Algorithm. Applied Sciences, 11(2), p.799.

Marine Traffic, 2022. What is the significance of the AIS Shiptype number? [online] Available at: <https://help.marinetraffic.com/hc/en-us/articles/205579997-What-is-the-significance-of-the-AIS-Shiptype-numbe>. [Accessed: 2022-01-07]

Millefiori, L.M., Zissis, D., Cazzanti, L. and Arcieri, G., 2016. A distributed approach to estimating sea port operational regions from lots of AIS data. In: IEEE International Conference on Big Data. pp.1627–1632.

Montewka, J., Manderbacka, T., Ruponen, P., Tompuri, M., Gil, M. and Hirdaris, S., 2022. Accident susceptibility index for a passenger ship-a framework and case study. Reliability Engineering & System Safety, 218, p.108145. https://doi.org/https://doi.org/10.1016/j.ress.2021.108145.

National Geospatial-Intelligence Agency, 2019. WORLD PORT INDEX. [online] Available at: <https://msi.nga.mil/Publications/WPI>. [Accessed: 2022-01-07]

Rajabi, A., Saryazdi, A.K., Belfkih, A. and Duvallet, C., 2018. Towards smart port: an application of AIS data. In: 20th International Conference on High Performance Computing and Communications; 16th International Conference on Smart City; 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS). pp.1414–1421.

Ray, C., Dréo, R., Camossi, E. and Jousselme, A.-L., 2018. Heterogeneous Integrated Dataset for Maritime Intelligence, Surveillance, and Reconnaissance. https://doi.org/10.5281/zenodo.1167595.

Silveira, P., Teixeira, A.P., Guedes Soares, C. and Santos, T.A., 2015. Assessment of ship collision estimation methods using AIS data. Maritime Technology and Engineering, .

Starczewski, A., Goetzen, P. and Er, M.J., 2020. A new method for automatic determining of the DBSCAN parameters. Journal of Artificial Intelligence and Soft Computing Research, 10.

Stopford, M., 2009. Maritime Economics, 3rd Ed. Routledge.

UNECE Trade Division, 2021. UN/LOCODE (Code for Trade and Transport Locations). [online] Available at: <https://service.unece.org/trade/locode/2021-2%20UNLOCODE%20SecretariatNotes.pdf>. [Accessed: 2022-01-07]

Zhang, M., Montewka, J., Manderbacka, T., Kujala, P. and Hirdaris, S., 2020. Analysis of the grounding avoidance behavior of a ro-pax ship in the gulf of Finland using big data. In: Proceedings of the International Offshore and Polar Engineering Conference. pp.3558–3567.