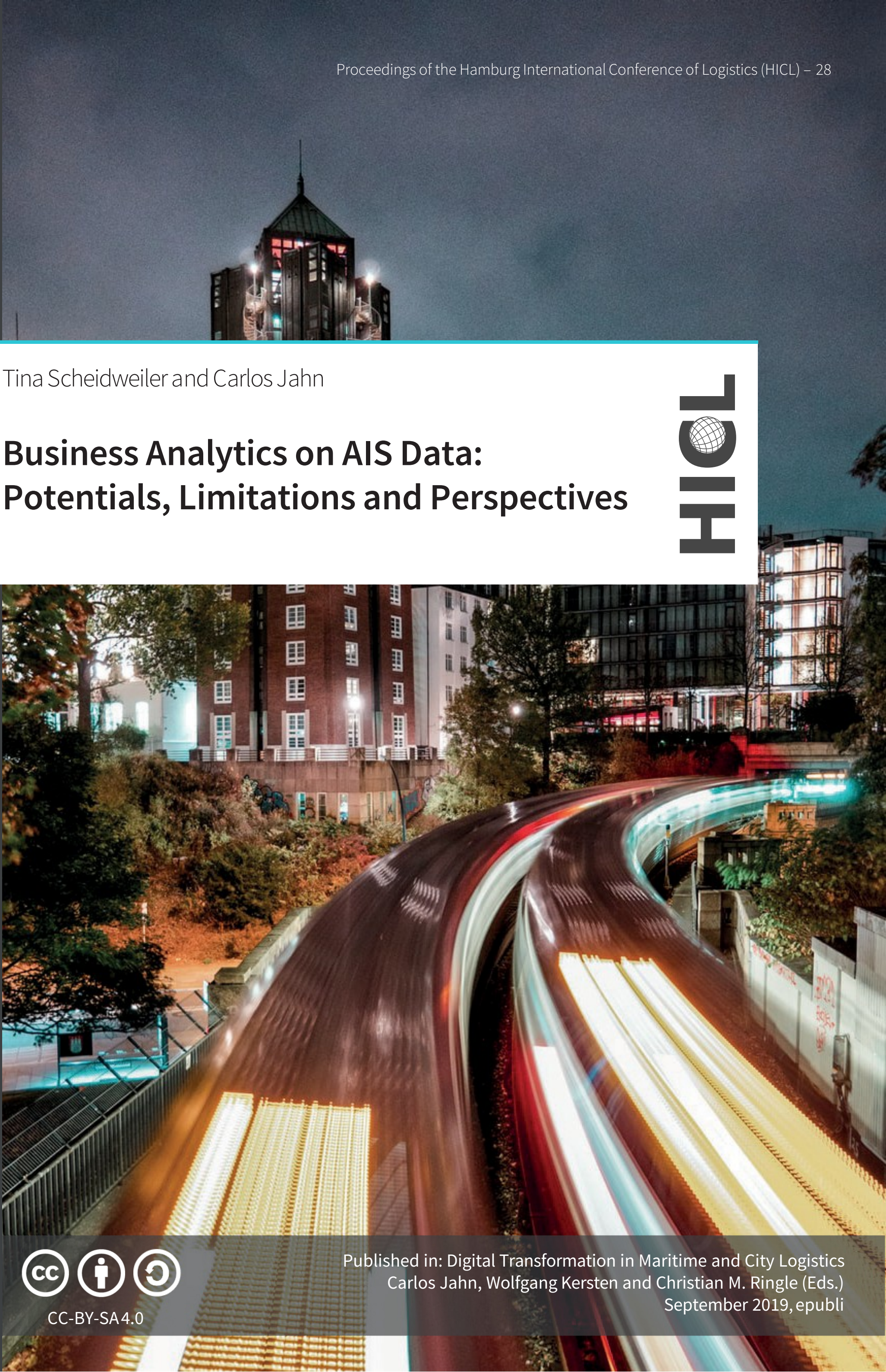
Proceedings of the Hamburg International Conference of Logistics (HICL) – 28



Tina Scheidweiler and Carlos Jahn

Business Analytics on AIS Data: Potentials, Limitations and Perspectives

HICL

Published in: Digital Transformation in Maritime and City Logistics
Carlos Jahn, Wolfgang Kersten and Christian M. Ringle (Eds.)
September 2019, epubli

Business Analytics on AIS Data: Potentials, Limitations and Perspectives.

Tina Scheidweiler[1]*, Carlos Jahn*[1]

1– Fraunhofer Center for Maritime Logistics and Services

Purpose: As maritime digitalization progresses, great opportunities for maritime transport arise: The introduction of the AIS opened up a number of possibilities and perspectives for increasing efficiency, automation and cost reduction using business analytics and machine learning in the supply chain and maritime sector.

Methodology: Various analysis and forecast techniques of machine learning as well as interactive visualizations are presented for the automated analysis of ship movement patterns, risk assessments of encounter situations of two or more ships as well as anomaly detections or performance indicators to quickly extract key figures of certain ships, routes or areas.

Findings: In addition to a comprehensive representation of relevant potentials and business analytics areas of AIS data, the feasibility and associated accuracy of the data mining and machine learning methods used are described. In addition, limitations will be shown and perspectives especially on autonomous surface ships will be discussed.

Originality: At present, there is no information platform that bundles the areas described in the previous sections in a central source. Previous work has either been limited to the visualization of historical and current ship movements or deals with narrowly limited individual questions of isolated applications.

Keywords: Business, Analytics, Maritime, Traffic

First received: 17.May.2019 **Revised:** 29.May.2019 **Accepted:** 03.June.2019

1 Introduction

As maritime digitalization progresses, great opportunities for the shipping and port industries as well as for plant manufacturers and service providers arise: Increased efficiency can be expected from the latest processes in the developments of ship operation and port logistics. Emission and fuel reductions can be achieved by prospective voyage and route planning.

With the introduction of merchant ships being equipped with transmitters of the Automatic Identification System (AIS) in 2004, shipping was able to reach a first milestone on the way to digitalization. The automated exchange of position, speed, course and other data from ship to ship or ship to shore station has increased the efficiency and safety of maritime traffic. In addition to the monitoring of ship movement data, timetables and weather conditions are used for the overall monitoring of maritime traffic, so that over the past few years a wealth of data of unexpected dimensions and possibilities has accumulated.

Shipping companies and seaports face a multitude of challenges: These include in particular the ship size developments of recent years, in which the absolute number of ships calling at German seaports is stagnating, but the gross tonnage of ships, the absolute total handling volumes and the number of unusually large vessels are growing continuously. These facts have the consequence that more and more transshipments are concentrated in relatively small time windows for port operations. Environmental influences have a significant impact on ship travel times and terminal operations. Strong winds cause challenges within the seaports: For example, transshipment facilities have to cease operation and seagoing ships have to adjust their speeds. In the case of shipping companies, emission and associated efficiency regulations currently play a central role - not only due to the obligation on ship-owners to report CO2 emissions from voyages to, from and within European waters since 1 January 2018 - but also within the high-profile context of climate change.

Previous applications based on AIS data focus on the visualization of historical and current ship movements or circumvent the influence of environmental data on ship movements only in a theoretical way. The investigation and inclusion of the effects of environmental influences on the travel times, routes and movement patterns of ships are not given much consideration in previous applications. In addition to a more efficient and optimized control of maritime traffic and the maritime supply chain, the focus is on ensuring safe navigation and thus avoiding collisions - especially on the open sea. Due to the ever-increasing volume of data and the simultaneous exponential growth of computing and storage capacities, the use of machine learning methods also offers immense possibilities in the maritime logistics sector.

After a short introduction to existing techniques on ship motion modelling and analysis, this paper gives an overview of the possibilities for the analysis of AIS data. In addition, limits and further perspectives of the application of AIS data for efficiency enhancement and analysis of business figures are shown.

2 Literature Review

Most of the machine learning methods used in maritime logistics are applied on AIS data. The current state of these applications is depicted in the following sections. A distinction is made between route prediction and the modeling of ship movements, i.e. representing arrival times, in order to increase the planning horizon of maritime stakeholders and thus to act more cost-efficiently and with more foresight. In addition, many algorithms are used for anomaly detection and collision avoidance tasks.

2.1 Route Forecasts

Within the route forecasting waypoints of a ship to a given destination or port of destination are to be predicted. In principle, route forecasting algorithms can be divided into two categories (Lo Duca, et al., 2017): point based and trajectory based. In the first case, the surrounding area of a ship is divided into parcels. For the individual ships in the close range, the algorithms estimate the probabilities of cells being targeted next by the ships under consideration. The probabilities are calculated using different machine learning algorithms. In trajectory-based estimation, clusters are generated from historical data. These contain all routes and serve to structure them.

The most frequently referenced work on the application of machine learning methods on AIS data is the "framework for anomaly detection and route prediction" created by Pallotta et al. (Pallotta, et al., 2013). Based on historically generated routes, algorithms for the classification and prediction of routes and for the subsequent anomaly detection of unnatural ship maneuvers are developed. Given a sequence of state vectors - consisting of position, course and speed values of a certain ship type - each compatible route is assigned a-posteriori probabilities in relation to the ship affiliation of the corresponding route. This allows the future position of the ship to be predicted for a given ship and a pre-specified period. Clustering algorithms are also used (Hexeberg, et al., 2017) to identify historical route patterns, assign ships to these patterns and predict the trajectory based on these patterns. Lo Duca et al. (Lo Duca, et al., 2017) use a k-Nearest-Neighbor classifier for ship route prediction. The point-based probability of reaching a grid point is calculated and the most probable route is given. In accordance to the AIS framework (Pallotta, et al., 2013), there is only a relatively small prediction horizon of 60 minutes. Zhang et al. (Zhang, et al., 2018) describe a method how AIS data can be used to select suitable routes. By analyzing the movement patterns of those ships that have already completed a planned route, the shortest route can be generated algorithmically: In a first step, significant changes in direction of individual trajectories are identified and used as turning points. A clustering method (Density-Based Spatial Clustering of Applications with Noise, DBSCAN) groups common turning points of all routes. This results in the last phase of automatic routing. The shortest route between start and end is searched for from the generated graph (Zhang, et al., 2018), (Dobrkovic, et al., 2015). Vespe et al. (Vespe, et al., 2012) present an unsupervised learning approach for incremental learning of movement patterns without specifying a priori contextual descriptions. The extraction of waypoints is the first step, followed by the definition of sea routes and routes connecting them. The proposed algorithm uses AIS data to detect changes in the course over ground for the proximity of the observed position; if these values reach a certain threshold value, a new turning waypoint is added to the list (Vespe, et al., 2012).

Neural networks are often used for regression tasks, also with regard to large data and deep learning tasks being the current state of the art for speech and image recognition. Mao et al. (Mao, et al., 2018) use a neural single hidden layer feedforward network with random hidden nodes for route prediction. Perera et al. (Perera, et al., 2010), (Perera, et al., 2012), use artificial neural networks for the tasks of classifying and identifying ships as well as tracking multiple ships. By means of an extended Kalman Filter, ship states are estimated and ship trajectories are predicted (Prévost, et al., 2007). In addition to the above, neural networks are also used in (Simsir & Ertugrul, 2007), (Simsir & Ertugrul, 2009), (Xu, et al., 2012) and (Zissis, et al., 2015).

In addition to neural networks, Bayesian networks are often used to generate "normal" behavior, such as, for example, done by Mascaro et al. (Mascaro, et al., 2014), (Mascaro, et al., 2010) or Aoude et al. (Aoude, et al., 2011), particulate filters (Mazzarella, et al., 2015) are used for clustering ship trajectories and Ornstein-Uhlenbeck processes (Pallotta, et al., 2014) and genetic algorithms (Pelizzari, 2015) are used for trajectory prediction as well.

2.2 Estimation of Arrival Times

The speed and arrival time estimations are usually accompanied by the route forecasts, since the speed is included in the route calculation for most approaches. Nevertheless, some methods, in which the speed or arrival time are estimated, are presented below.

Arrival times are an important factor for the handling of arriving ships in ports. Even forecasts with an accuracy of 60-70% can significantly improve process planning at the terminals (Yu, et al., 2018). If the route between two points, for example between two AIS data intervals, is to be predicted, the most common method (Posada, et al., 2011) is the Constant Velocity Model. A linear movement from point A to point B is assumed. Further information on speed and course, possibly contained in the AIS data, will not be taken into account. The speed is set as constant. The temporal/spatial proximity plays a considerable role in the accuracy of the forecast. The greater the distance between the points, the less accurate the interpolation. This problem occurs with predictions of more than one hour, as well as in offshore areas (Mazzarella, et al., 2015). By assuming constant speed, however, environmental influences that could lead to a reduction in speed are not taken into account.

Besides neural networks, support vector machines (SVM) are often used for classification and regression tasks. Using a kernel that calculates distances between two objects, a SVM divides a set of objects into classes so that the widest possible area around the class boundaries remains free of objects (EliteDataScience, 2017). Parolas et al. (Parolas, et al., 2017) use SVMs and neural networks to estimate the time of arrivals for container ships in the port of Rotterdam. The weather and environmental conditions are clustered and used alongside the AIS data for training. Fancello et al. (Fancello, et al., 2011) use artifical neural networks to estimate ships arrival times in ports to allocate human resources in container terminals.

2.3 Collision Avoidance and Anomaly Detection

Besides the use of the AIS data for the modelling of basic ship movements, there are possibilities to AIS for collision avoidance, safety assessment as well as anomaly detection with regard to the early detection of these anomalies to enhance safety. Using a Deep Neural or Bayesian Network, the navigation and movement behaviour of ships during an encounter can be studied and predicted for real-time encounters (Perera, et al., 2010), (Perera, 2018). By adjusting the ship's speed and course, a route recommendation for a collision avoidance manoeuvre can then be given. Same approaches are developed by (Lee, et al., 2004), (Xue, et al., 2008), (Mou, et al., 2010), (Simsir, et al., 2014) or depicted by (Tu, et al., 2017).

In addition to collision avoidance, there are possibilities for anomaly detection of historical and current ship movements using machine learning methods. Examples are Gaussian Processes (Rasmussen, 2006), (Kowalska & Peel, 2012) or Bayesian Networks (Johansson & Falkman, 2007), (Mascaro, et al., 2010) or (Mascaro, et al., 2014).

2.4 Approach

For the above mentioned reasons, in the field of maritime logistics, where highly efficient process control is an important success factor, there are numerous opportunities to apply machine learning methods to the upcoming arising problems to gain efficiency increases. Nevertheless, in contrast to the above-mentioned applications, there exists almost no tool for the first descriptive analysis of movements in real-time. The BigOceanData portal (BigOceanData, 2019), which provides first descriptive analyses of historical data, is the only one to be mentioned here.

The need for an interactive solution for the automated analysis and evaluation of maritime and environmental data is in the limited range of information available from providers already operating on the market for the processing of AIS data. An exemplary study of information providers already operating on the market showed that data collected using AIS receivers is usually mapped only.

The innovative value of the presented research project can be seen in the automated consolidation, processing and provision of data relevant to shipping from various sources. To the knowledge of the project participants no comparable data processing system exists - neither nationally nor internationally. However, in isolated cases and with different methods AIS data again and again for risk and safety analyses is used.

3 Data Source

In addition to AIS data, other data sources were identified and subsequently used. These are described below.

3.1 Automatic Identification System (AIS)

The Automatic Identification System, which was introduced by the International Maritime Organization (IMO) to increase safety in shipping, provides the basis for the following analyses and methodologies. The data transmitted by AIS transmitters for the exchange of nautically relevant information between different ships and shore stations can be classified into three categories: Static data, dynamic data and Voyage-related data.
According to the recommendations published by the International Telecommunication Union (International Telecommunication Union, 2014), AIS data exchange consists of 27 different messages. The most relevant for navigation are the position reports (messages 1, 2 and 3) and the static and voyage-related ship data (message 5). For further analysis, the Fraunhofer internal AIS dataset for the period 01.02.2016 - 30.04.2018 of the North Sea and Baltic can be used to validate and check the theoretical approaches. Nevertheless, it should be noted within this paper, and in particular in the presentation, that various methods have been performed on a smaller, generalized dataset in order to ensure the real-time capability. The following parameters are used:

ShipID: In order to comply with data protection guidelines, each ship will be assigned an independent identification number (ID)
ShipType: Integer in accordance with (International Telecommunication Union, 2014)
Length: The overall length of the vessel in meters
Breadth: The breadth of the vessel in meters
Draught: The maximum current draught in meters
Latitude: latitude in 1/10 000 min (90°, north = positive (according to 2er-complement), South = negative (according to 2er-complement)
Longitude: longitude in 1/10 000 min (180°, east = positive (according to 2er-complement), West = negative (according to 2er-complement)
SOG: Speed over ground in 1/10 knots
COG: Course over ground in 1/10 = (0-3 599) °
TH: True Heading from 0 to 359°

3.2 Environmental Data

In addition to the ship movement data of the AIS, weather and environmental data is used for further analyses, such as the determination of the resistance of a ship, to estimate emissions. In particular, the parameters wind, wave and current, which are often depicted on a $m \times n$ grid, must be used for this purpose. For the correlation of the environmental and AIs data, the projected AIS data on a $1000\ m \times 1000\ m$ grid is given by
$x_i = \frac{\lambda_i \cdot 1852 \cdot 60}{100}$ and $y_j = \frac{\varphi_j \cdot 1852 \cdot 60 \cdot \cos\left(\lambda_i \cdot \frac{\pi}{180}\right)}{100}$
with λ_i and φ_j being longitude and latitude for $i = 1, \dots, n$ and $j = 1, \dots, m$.

4 Potentials: Analysis Framework

In the following, the results of the analyses and, in a first step, the data flow are presented. During the presentation of the paper the Jupyter Notebook (Project Jupyter, 2018) framework will be presented.
AIS data is checked for plausibility within a first step and outliers are removed. In addition, the existing position data of the AIS of individual ships are merged into trips from a port of departure to a port of destination. For details on the methodology extracting trips, reference is made to (Jahn & Scheidweiler, 2018). The data is then processed within a Jupyter Notebook (Project Jupyter, 2018) framework using Python.
Within the framework, investigations are made on the automated analyses of historical and current ship movements in specific sea areas. In addition, potentials for risk and safety assessments of different encounter situations of ships or areas, for the anomaly detection of historical and current ship movements are examined. Within Python, descriptive analyses, regressions as well as clustering or supervised machine learning algorithms such as neural networks or decision trees are used.

4.1 Movement Patterns

A central aspect of the analysis of movement patterns should be the automated representation of historical ship movements for the identification of distributions and movement patterns of ships as well as route patterns and traffic densities of different ship classes and sea areas. In addition to the classical motion pattern analysis, the potentials of automated frequency analyses of ships as well as distribution analyses of speeds and courses along predefined areas are to be investigated and the corresponding methods developed. Methods for calculating the time ships spend at berths and on anchorage will also be presented.

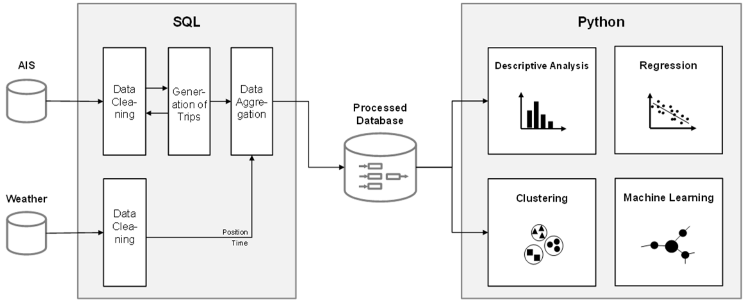


Figure 1: Visualization of the data flow of AIS and weather data. Preparation and consolidation of these within SQL and following analyses in Python.

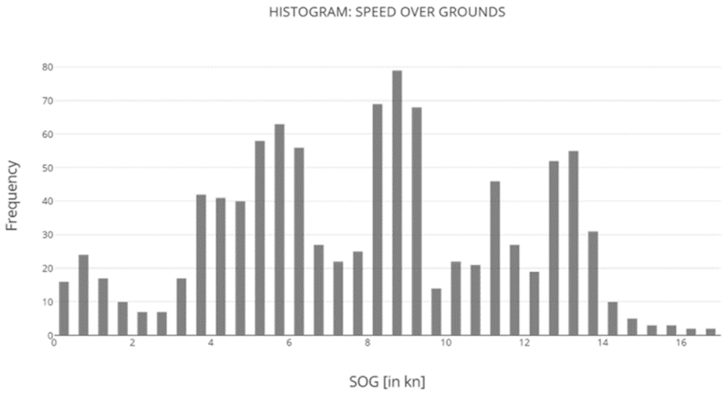


Figure 2: Histogram of the speeds over the ground of the study area

4.1.1 Motion pattern of a predefined area

For the identification of historical movement patterns within an area, this region has to be specified. Using the boundaries given by φ_{min} , φ_{max}, λ_{min} and λ_{max} ship movement data can be automatically extracted from the database. In addition to chart displays, histograms can be generated for the initial analysis of the speeds or courses, but also of the static data of the ships. Figure 2 shows an example of the speeds prevailing in the study area. In addition to histograms, other descriptive tools such as box plots can of course be used for visual analysis of the position distributions of the measurements of the parameters and edited interactively.

4.1.2 Motion Patterns of Predefined Routes

With a given departure and destination port, ship movement data of the AIS is extracted from the specified area of the database. In the following, the movements from Rotterdam to Hamburg are considered.

In addition to the data described in section 3 transmitted via the AIS, trip data is used within the following analyses. In addition to motion pattern representations, histograms and scatter plots are created for visual analysis. The diagram below depicts the travel times from Rotterdam to Hamburg. With the help of this information, possible ship times can be determined in real time and thus further planning of the hinterland or other resources can be carried out.

For port and shipping authorities, the automated speed analysis of voyages is of particular relevance in order to trace incidents, for example, as depicted in Figure 4.

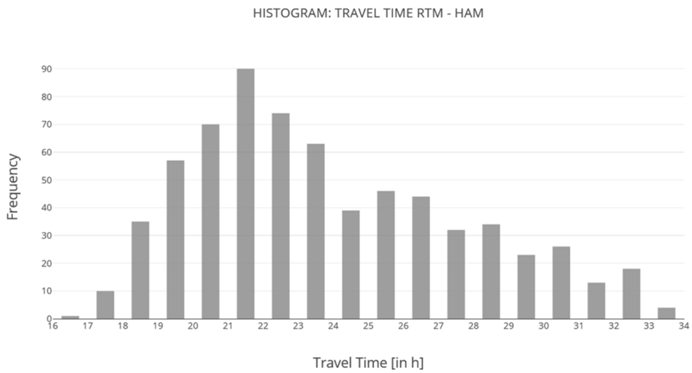


Figure 3: Histogram of the total travel times of the investigation route

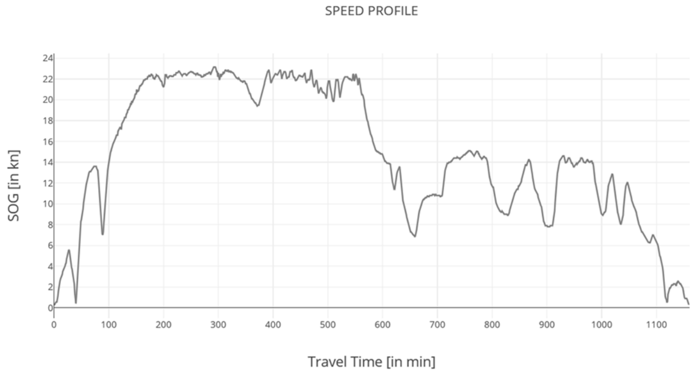


Figure 4: Average speed profile of trips from Rotterdam to Hamburg

The energy efficiency of ships is significantly influenced by the speed-dependent fuel consumption. To assess the efficiency, the travel times and speeds of ships of different sizes on specified routes are examined. For a more detailed analysis, the frequency of speeds along the route is determined for selected ship size classes. For this purpose, the velocities are derived as Froude numbers:

$$Fn = \frac{SOG}{\sqrt{g * length}}$$

where g is the gravitational force and SOG the current speed over the ground of a ship.

Froude's number relates inertia forces to gravity forces and is used in shipbuilding as an indicator of the energy loss caused by the radiation of waves in relation to the total energy loss. For regular services on a route with similar or even identical ships (e.g. ferry connections), the absolute speeds can be compared. This allows conclusions to be drawn about the efficiency of individual ships within a fleet or about the efficiency of individual operators in a direct competition situation.

4.1.3 Frequency and Distribution Analysis

In addition to the motion pattern analysis, potentials of automated frequency analyses of ships as well as distribution analyses of speeds and courses along predefined ranges are investigated and the corresponding methods developed. Figure 5 depicts the crossing positions of ships in the Kiel Canal at Sehestedt. Based on these crossing positions, lateral distributions can be determined for a later risk analysis which provides ship-owners and shipping companies with a rapid and efficient risk assessment of their own ships.

4.1.4 Berthing and Anchoring Times

Methods are developed for calculating the time ships spend at berth and at anchorage in order to draw conclusions about inefficient and thus costly waiting times for ships. The first step in determining anchoring times is to determine where vessels are at anchor assuming that the speed is larger than a specified threshold c. As an example, the average anchoring time of vessels entering the port of Hamburg in the German Bight can be determined based on the areas: 5.023 hours.

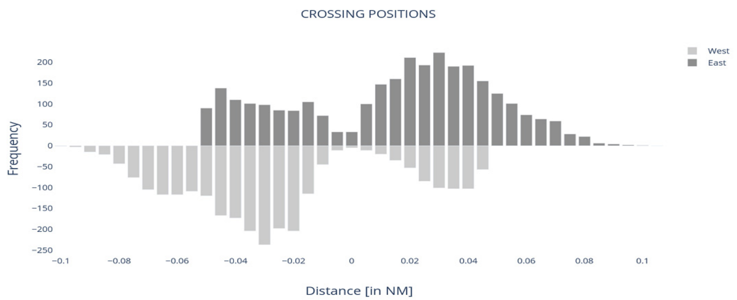


Figure 5: Histogram of the crossing positions of the study area

4.2 Risk Assessment

For the risk and safety assessment the passing distances of two ships during encounters, as well as the distances to fairway edges and barrel lines are examined. Furthermore the assessment of maximum and minimum angles of approach which a pre-defined group of ships has steered in a certain fairway are defined. The encounter situations of two pre-selected ships can be evaluated with regard to collisions and fairways with regard to their safety levels by using accident statistics. Both the current position data of the AIS and historical movement data will be used to provide a risk profile. Encounter situations are to be assessed by determining the number of encounters in a user-defined fairway area within a defined time period. Again, the passing distance of the ships are analysed. Finally, it should be possible to assign a safety level to each fairway and a risk level to each encounter situation.

The risk indicates a numerical value which is determined by the probability of occurrence (also called frequency) and the average amount of damage. This corresponds to determining the probability of occurrenceP_i, which results in the consequence C_i from an identified hazard H_i. For the risk analysis, this is then linked with the monetary valuation U_i of the resulting consequence, so that the overall risk applies (Pedersen, 2010), (International Maritime Organization, 2007):

$$Risk = \sum_i P_i(H_i, C_i) \cdot U(C_i)$$

The following passages are components of a possible risk assessment.

4.2.1 Passing Distances

The Closest Point of Approach (CPA) between ships is the smallest distance they are likely to have without changing course or speed. Given a standard traffic situation, the CPA and the Time to Closest Point of Approach (TCPA) can be calculated based on the course over ground of the own ship x, the course over ground of the traffic ship y, the speed over ground of the own ship v, the speed over ground of the traffic ship w as well as the longitudinal distance t and lateral distance u by

$$CPA = \sqrt{\frac{(u \cdot w \cdot \sin y - t \cdot w \cdot \cos y - u \cdot v \cdot \sin x + t \cdot v \cdot \cos x)^2}{v^2 + w^2 - 2 \cdot v \cdot w \cdot (\sin x \cdot \sin y + \cos x \cdot \cos y)}}$$

For the risk and safety assessment, the distances between two ships during encounters, as well as the distances to fairway edges and barrel lines, can be recorded, analysed and visualised. The TCPA can be used to determine the urgency of an evasive action and is given by

$$TCPA = \frac{t \cdot v \cdot \sin x - t \cdot w \cdot \sin y + u \cdot v \cdot \cos x - u \cdot w \cdot \cos y}{v^2 + w^2 - 2 \cdot v \cdot w \cdot (\sin x \cdot \sin y + \cos x \cdot \cos y)}$$

For the discussion of the passing distances, those of certain areas are determined. Figure 6 presents the passing distances on the basis of percentages.

In addition, the encounter situations of two pre-selected ships can be evaluated with regard to collisions. Encounter situations shall be assessed by determining the number of encounters in a user defined fairway area within a defined time period. The passing distance of the ships is also analysed here. Finally, it should be possible to assign a safety level to each fairway and a risk level to each encounter situation.

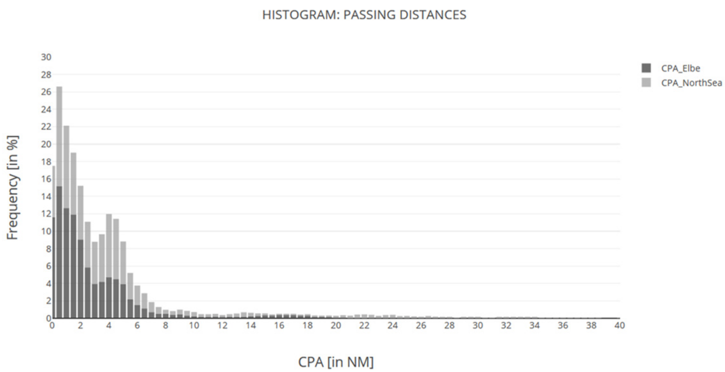


Figure 6: Representation of the CPAs in the Elbe River and the North Sea

Using the IWRAP method recommended by the International Association of Lighthouse Authorities, the frequency of collisions and groundings in a given area can be specified on the basis of information on traffic volume and composition, route geometry and bathymetry. As an example, the collision risk for the Elbe at the level of the Hamburg suburb Blankenese is 1.4%. For further information on the modelling tool, please refer to the IALA documentation (IALA, 2014).

4.2.2 Safety levels of fairways

Fairways can be assessed in terms of their safety levels by discussing domain violations from ships. Both the current position data of the AIS and historical movement data will be used to provide a risk profile.
The ship domains indicate the area around a ship which should be kept free from other ships at all times. Since a violation of this area is the prerequisite for a possible collision, encounters between two ships that do not violate the other domain can be classified as comparatively low-risk. Moments in which at least one of the domains is violated require special attention and an analysis of the expected time of a possible collision (TCPA) as well as the speeds and courses of the ships against the background of historical data.

4.3 Anomaly Detection

In addition to the general movement pattern and behaviour analyses of ships, the potentials of anomaly detections of historical and current traffic and traffic movements are to be identified and examined using maritime traffic data. For example, strong course deviations can be investigated here, but also the abandonment of fairways, main shipping lanes or traffic separation areas due to large draughts. Anomaly detection offers the potential of an early intervention on the ship's command, for example to prevent collisions or groundings.

For the analysis of anomalies on the basis of course deviations within a defined area, the eastern section of the Kiel Canal is used. Based on this, an Isolation Forest (scikit-learn developers, 2018) provided by scikit-learn (Géron, 2018) will be used for outlier detection (Lewinson, 2018).

With the help of the trained Isolation Forest, an anomaly value is provided for each data point and displayed in color. In addition to the anomalies of specified areas, anomalies of routes can also be evaluated with the help of the Isolation Forest. If only position data and thus point densities are taken into account, unusual behavior of ships - such as slow-moving ships - cannot be detected. Conversely, fast sailing is more often regarded as unusual. In addition, in this case areas are marked as anomalies that are far away from the center of the point (e.g. in ports). For the training the input parameters latitude, longitude, speed as well as course over ground were used and an Isolation Forest returning an anomaly score for each data point the area of the Kiel Canal was trained. With a number of 1000 estimators, an accuracy of 90 % was achieved.

Collisions can be roughly divided into two categories: Collisions between two or more ships or the collision of a ship with an immovable structure such as offshore wind farms or bridges. For the detection of anomalies to prevent collisions, encounter situations can be evaluated with respect to the compliance of typical and thus normal evasive maneuvers of avoidable ships.

In a first step, encounter situations are extracted from the AIS data. For the extraction of these, the assumption is made that an encounter situation exists when the minimum distance of two ships in the same period on the open sea is less than 3 nautical miles. With the help of the trips described in section 3 and the CPA, all ship encounters and corresponding AIS messages

can be extracted. It has to be ensured that always the same times of two ships meeting each other are combined.
Based on the encounter data, the type of encounter situation is determined depending on the maneuver of the first moment. A distinction is made between the situations of frontal encounter, overtaking and intersection. In addition, the system detects which ship should initiate an evasive maneuver and which should maintain course and speed.
In order to determine the behavior of two ships encountering each other, a confidence interval of the speed and course change values of ships within the same area and considering the same encounter situations is then used. If μ_{TH} is the mean value of the course change rate, σ_{TH} the corresponding standard deviation and x_{TH} the course change rate of any evasive ship at the time of the assessment, then a normal evasive behavior exists for a significance level α if

$$x_{TH} \in \left[\mu_{TH} - z_{\left(1-\frac{\alpha}{2}\right)} \frac{\sigma_{TH}}{\sqrt{n}}; \mu_{TH} + z_{\left(1-\frac{\alpha}{2}\right)} \frac{\sigma_{TH}}{\sqrt{n}}\right],$$

with n being the number of observations and $z_{\left(1-\frac{\alpha}{2}\right)}$ the quantile of standard normal distribution.

4.4 Conclusion

To assess the feasibility of the AIS data analysis, the developed methods were validated on the basis of historical ship movement and environmental data and the values were checked for plausibility. With regard to methods that were developed using machine learning methods, the accuracy of these methods was also determined. It was found that the methods, especially the Isolation Forest for anomaly detection, used had good accuracy.

In addition to the potentials identified in the presented paper, the analysis and prediction of encounter situations between two or more ships can be used to determine the changes in speed and/or course of given encounter situations. In this context, neural networks and agent-based learning could be used to learn speed and course changes to avoid collisions. Deep learning methods can also be used for the surface and underwater detection of ships, coastlines or buoys and thus for object detection.

5 Limitations

It can be concluded that the identified potentials can be implemented within a system for business analytics on AIS data. With regard to a possible real-time implementation, the extraction of the mass data for motion pattern analysis and the offline training of the models in the case of anomaly detection must be taken into account.

In this context, it is also often necessary to take into account missing static ship data such as engine values or insufficient data storage capacities and computational powers. In addition, incorrectly captured values in the AIS and data consistencies result in an increased processing effort. Furthermore, especially in the context of machine learning algorithms, the loss of control as well as the lack of theory and the high degree of trial and error should be mentioned.

Despite the consistently provided data, they leave room for analysis within the human decision-making process, depending on the viewer. In the course of a real-time online tool, the topic of data security must also be taken into account as well.

6 Perspectives and Outlook

According to the port and water authorities, up to 30% efficiency savings can be achieved through automated analysis of movement data. In addition, the number of direct connections and transshipments can be determined for shipping companies and ship owners in particular or to determine congestion issues of ports and terminals. In this context, efficiency gains can be achieved through the selective retrieval of relevant information or complete situational awareness from a single source. There is the possibility that further or more precise outputs could improve the processes or expand the planning horizon, since the available information is currently considered and evaluated separately. Apart from this information, conclusions on the speed of port operations of different shipping companies can be discussed, berth availabilities or schedule integrities can be determined to extend the planning horizon of maritime stakeholders.

Since 01.01.2018, the obligation to report CO2 emissions is mandatory for ships travelling to and from Europe as well as for intra-European traffic. In this context, the first emission report has to be delivered by 30.04.2019 to the EU. The amount of CO2 emitted must be transmitted for the entire voyage, from the port of departure via European ports to the port of destination. In order to meet these requirements and to enable and provide the maritime stakeholder with performance monitoring of fuel consumption and associated emissions, Fraunhofer CML together with the Wismar University of Applied Sciences, the JAKOTA Design Group, the German Aerospace Center (DLR) and the project manager JAKOTA Cruise Systems is developing a software prototype for calculating CO2-emissions within the project EmissionSEA. The data of the ships are used as well as information from the weather service to derive fuel consumption and emissions from speed and external influences. These assessments can also be used to draw conclusions about compliance with slow steaming measures.

In addition, CML is planning a follow-up project based on the results of the potential analysis presented within the paper containing the technical implementation and development of a suitable software solution for the interactive analysis of the above-mentioned data.

Acknowledgements

The TINA project, which led to the results presented here, was funded by the Federal Ministry of Transport and Digital Infrastructure within the framework of the mFUND (Modernitätsfonds) under the funding number 19F1043A. The CML researchers worked in close cooperation with the associated partners, the Federal Maritime and Hydrographic Agency and the Directorate-General for Waterways and Shipping.

The project EmissionSEA, which is mentioned in the article, is funded by the Federal Ministry of Transport and Digital Infrastructure within the framework of the mFUND (Modernitätsfonds) under the funding number 19F2062C.

References

Aoude, G. S., Joseph, J., Roy, N. & How, J. P., 2011. Mobile Agent Trajectory Prediction using Bayesian Nonparametric Reachibility Trees. American Institute of Aeronautics and Astronautics.

BigOceanData, 2019. BigOceanData.com. [Online]
Available at: www.bigoceandata.com [Accessed 12 03 2019].

Dobrkovic, A., Iacob, M.-E. & van Hillegersberg, J., 2015. Using Machine Learning for Unsupervised Maritime Waypoint Discovery from Streaming AIS Data. i-KNOW.

EliteDataScience, 2017. Modern Machine Learning Algorithms: Strengths and Weaknesses. [Online]
Available at: https://elitedatascience.com/machine-learning-algorithms [Accessed 11 07 2018].

Fancello, G. et al., 2011. Prediction of arrival times and human resources allocation for container terminal. Maritime Economics & Logistics, 13(2), pp. 142-173.

Géron, A., 2018. Praxiseinstieg Machine Learning mit Scikit-Learn und TensorFlow: Konzepte, Tools und Techniken für intelligente Systeme. s.l.:O´Reilly.

Hexeberg, S., Flaten, A. L., Eriksen, B.-O. H. & Brekke, E. F., 2017. AIS-based Vessel Trajectory Prediction. 20th International Conference on Information Fusion.

IALA, 2014. IALA - IWRAP MkII. [Online]
Available at: https://www.iala-aism.org/wiki/iwrap/index.php/Theory

International Maritime Organization, 2007. Formal Safety Assessment. Consolidated text of the Guidelines for Formal Safety Assessment (FSA) for use in the IMO rule-making process (MSC/Circ.1023−MEPC/Circ.392), s.l.: s.n.

International Telecommunication Union, 2014. Technical characteristics for an automatic identificatio nsystem using time division multiple access in the VHF maritime mobile frequency band; Recommendation ITU-R M.1371-5, Genua: M Series; Mobile, radiodetermination, amateur and related satellite services.

Jahn, C. & Scheidweiler, T., 2018. Port Call Optimization by Estimating Ships' Time of Arrival. Dynamics in Logistics.

Johansson, F. & Falkman, G., 2007. Detection of vessel anomalies - a bayesian network approach. 3rd International Conference on Intelligent Sensors, Sensor Networks and Information, ISS-NIP, IEEE, pp. 395-400.

Kowalska, K. & Peel, L., 2012. Maritime anomaly detection using gaussian processes active learning. 15th International Conference on Information Fusion (FUSION), IEEE, pp. 1164-1171.

Lee, S.-M., Kwon, K.-Y. & Joh, J., 2004. A Fuzzy Logic for Autonomous Navigation of Marine Vehicles Satisfying COLREG Guidelines. International Journal of Control, Automation and Systems.

Lewinson, E., 2018. Outlier Detection with Isolation Forest. [Online] Available at: https://towardsdatascience.com/outlier-detection-with-isolation-forest-3d190448d45e

Lo Duca, A., Bacciu, C. & Marchetti, A., 2017. A k-nearest neighbor classifier for ship route prediction. OCEANS 2017, pp. 1-6.

Mao, S. et al., 2018. An Automatic Identification System (AIS) Database for Maritime Trajectory Prediction and Data Mining. Proceedings of ELM-2016, Springer International Publishing (Proceedings in Adaptation, Learning and Optimization, 9), pp. 241-257.

Mascaro, S., Korb, K. B. & Nicholson, A. E., 2010. Learning Abnormal Vessel Behaviour from AIS Data with Bayesian Networks at Two Time Scales. Clayton: Clayton School of Information Technology, Monash University.

Mascaro, S., Nicholson, A. E. & Korb, K. B., 2014. Anomaly Detection in Vessel Tracks using Bayesian networks. International Journal of Approximate Reasoning, pp. 84-98.

Mazzarella, F., Fernandez Arguedas, V. & Vespe, M., 2015. Knowledge-Based Vessel Position Prediction using Historical AIS Data. Sensor Data Fusion: Trends, Solutions, Applications.

Mou, J. M., van der Tak, C. & Ligteringen, H., 2010. Study on collision avoidance in busy waterways by using AIS data. Ocean Engineering, pp. 483-490.

Pallotta, G., Horn, S., Braca, P. & Bryan, K., 2014. Context-enhanced vessel prediction based on ornstein-uhlenbeck processes using historical ais traffic patterns: Real-world experimental results. 17th International Conference Information Fusion (FUSION).

Pallotta, G., Vespe, M. & Bryan, K., 2013. Vessel pattern knowledge discovery from AIS data: a framework for anomaly detection and route prediction. In: Entropy, vol. 15. La Spezia, Italy: Centre for Maritime Research and Experimentation, pp. 2218-2245.

Parolas, I., Tavasszy, L., Kourounioti, I. & van Duin, R., 2017. Prediction of Vessel's Estimated Time of Arrival (ETA) Using Machine Learning - a Port of Rotterdam Case Study. 96th Annual Meeting of the Transportation Research Board.

Pedersen, P. T., 2010. Review and application of ship collision and grounding analysis procedures. Marine Structures, 11 10, pp. 241-262.

Pelizzari, A., 2015. Genetic Algorithm for Shipping Route Estimation with Long-Range Tracking Data. NOVA Information Management School.

Perera, L., Carvalho, J. P. & Guedes Soares, C., 2010. Bayesian Network based sequential collision avoidance action execution for an Ocean Navigational System. 8th IFAC Conference on Control Applications in Marine Systems, pp. 266-271.

Perera, L. P., 2018. Autonomous Ship Navigation under Deep Learning and the Challenges in COLREGS. Proceedings of the 37th International Conference on Ocean, Offshore and Arctic Engineering.

Perera, L. P., Oliveira, P. & Guedes Soares, C., 2012. Maritime Traffic Monitoring Based on Vessel Detection, Tracking, State Estimation, and Trajectory Prediction. IEEE Transactions on Intelligent Transportation Systems, 13(3), pp. 1188-1200.

Posada, M. et al., 2011. Maritime Awareness For Counter-Piracy in the Gulf of Aden. European Commission - Joint Research Centre.

Prévost, C. G., Desbiens, A. & Gagnon, E., 2007. Extended Kalman Filter for State Estimation and Trajectory Prediction of a Moving Object Detected by an Unmanned Aerial Vehicle. Proceedings of the 2007 American Control Conference.

Project Jupyter, 2018. Jupyter. [Online]
Available at: http://jupyter.org/

Rasmussen, C. E., 2006. Gaussian processes for machine learning. s.l.:s.n.

scikit-learn developers, 2018. sklearn.ensemble.IsolationForest. [Online]
Available at: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html

Simsir, U. et al., 2014. Decision support system for collision avoidance of vessels. Applied Soft Computing, Issue 25, pp. 169-378.

Simsir, U. & Ertugrul, S., 2007. Prediction of Position and Course of a Vessel Using Artifical Neural Networks by Utilizing GPS/Radar Data. 3rd International Conference on Recent Advances in Space Technologies, Volume 3, pp. 579-584.

Simsir, U. & Ertugrul, S., 2009. Prediction of manually controlled vessels' position and course navigating in narrow waterways using Artificial Neural Networks. Applied Soft Computing , Volume 9, pp. 1217-1224.

Tu, E. et al., 2017. Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey. IEEE Transactions on Intelligent Transportation Systems, Issue 19, pp. 1559-1582.

Vespe, M., Visentini, I., Bryan, K. & Braca, P., 2012. Unsupervised learning of maritime traffic patterns for anomaly detection. 9th IET Data Fusion & Target Tracking Conference: Algorithms & Applications, pp. 1-5.

Xue, Y., Lee, B. S. & Han, D., 2008. Automatic collision avoidance of ships. Proceedings of the IMechE (Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment), pp. 33-46.

Xu, T., Liu, X. & Yang, X., 2012. A novel approach for ship trajectory online prediction using bp neural network algorithm. Advances in Information Sciences & Service Sciences, 4(11).

Yu, J. et al., 2018. Ship arrival prediction and its value on daily container terminal operation. Ocean Engineering 157, pp. 73-86.

Zhang, S.-k.et al., 2018. Data-driven based automatic maritime routing from massive AIS trajectories in the face of disparity. Ocean Engineering 155, pp. 240-250.

Zissis, D., Xidias, E. K. & Lekkas, D., 2015. Real-time vessel behavior prediction. Evolving systems, pp. 1-12.