EPiC Series in Computing
Volume 102, 2024, Pages 199–214
Proceedings of Conference on Cognitive and Computational Aspects of Situation Management 2023

A Framework for Surface Vessel Nautical-Behaviour Analysis towards Cognitive Situation Awareness *

Ghassan Al-Falouji, Lukas Haschke, Dirk Nowotka, and Sven Tomforde

Kiel University, Kiel, Germany
gaf/lha/dn/st@informatik.uni-kiel.de

Abstract

The rapidly advancing technological development allows for increased automation of traffic behaviour, including nautical systems. Innovative technologies for cognitive situation modelling form the basis for an autonomous assessment of environmental conditions, which can then be used to realise autonomous navigation behaviour. Cutting-edge vehicles designed for autonomous operations are currently in development, demonstrating the capability to operate seamlessly even in busily frequented harbour areas and waterways. In this article, we present a novel framework for analysing the spatio-temporal nautical behaviour of maritime surface vessels. In a scenario, we assess and analyse navigation data gathered from the Kiel Fjord as part of a demonstration use case. The presented framework integrates a variety of different approaches and thus represents the basic technology for assessing and modelling ship behaviour prior to operational use, as well as detecting anomalous events during utilisation.

1 Introduction

Autonomous driving has experienced numerous breakthroughs in the last decade. The first services, such as an autonomous taxi service in San Francisco, USA, are now emerging [1]. The corresponding technology is also becoming increasingly transferable to other areas of application. For instance, autonomous ships are no longer utopian, even in heavily used harbour areas and waterways.

An essential part of autonomous maritime vessel navigation is the ability to locate and identify other ships, including their current and expected behaviour. Large container vessels, for instance, typically have a turning radius of hundreds of metres, significantly constraining their capacity to avoid obstacles. On the other hand, busy waterways and harbour regions are characterised by strong traffic stemming from highly heterogeneous ship classes. Consequently, technology such as AIS (Automatic Identification System) is constantly used to broadcast navigation-related information at regular intervals to other ships nearby.

*This research has been partly funded by the German Federal Ministry for Digital and Transport within the project “CAPTN Förde Areal II - Praxisnahe Erforschung der (teil)autonomen, emissionsfreien Schifffahrt im digitalen Testfeld” (45DTWV08D).

As a basis for cognitive situation modelling, on which autonomous navigation behaviour can then be established, efficient processing, preparation, and analysis of observed historical data are necessary [2]. In this article, we present a *Surface Vessel Nautical Behaviour Analysis* (SV-NBA) framework that combines and compares different approaches for nautical situation awareness and modelling. In addition, we provide the following contributions: (i) An efficient pipeline for comparative analysis of observational data, (ii) a sample data set based on AIS observations in the Kiel Fjord, (iii) an exploration of relevant metrics especially suitable for anomaly detection of navigation behaviour.

The remainder of this article is organised as follows. Section 2 describes the technological background of this article, with a focus on AIS data and its use. Subsequently, Section 3 introduces our novel framework for surface vessel nautical behaviour analysis. Section 4 demonstrates the experimental analysis of the framework. Finally, Section 5 presents a summary and gives an outlook on future work.

2 Background

The last decade has witnessed strong efforts towards autonomous vessel navigation. Currently, existing technical studies and prototypical solutions for autonomous and semi-autonomous behaviour are primarily confined to standard assistance systems, such as open-sea autopilot, or focused on short-distance cargo transportation within coastal areas [3]. Many decisions regarding the behaviour of ships are already supported or implemented by technical systems (cf. [4, 5]). It is expected, that full autonomy will shortly become achievable, albeit initially in secure environments.

Autonomous surface vessel navigation behaviour is based on continuous observation and assessment of the current environmental conditions–which is then turned into concrete navigation actions by an autonomous controller (see e.g., [6]). Consequently, situation awareness is a prerequisite for autonomous navigation. Various contributions have considered aspects of cognitive situation modelling and awareness problems in the last decade, including approaches to reconstruct ship trajectories from spatio-temporal historical data (i.e., based on AIS data) [7], modelling of ship trajectories in safety-critical conditions–such as collision avoidance manoeuvres–using optimisation heuristics [8], a prediction of trajectories for the own and other ships in the vicinity based on recent machine learning technology such as artificial neural networks [9], modelling of current constellations based on observational data [2] or an analysis of ship trajectories [10]. However, there is no unified approach for maritime navigational situation modelling and analysis yet established, which is partly driven by the high heterogeneity of available input data.

The various maritime test environments oriented towards autonomous navigation (i.e., prototype test ships for experimentation purposes such as the Wavelab in Kiel, Germany [11]) have different sensor constellations, characteristics, and capabilities. The only reference data that is common in all use cases is AIS, which is an integral component in today's maritime navigation and safety. This system allows real-time tracking of ships, providing vital data that encompass their identity, position, speed, and course, thus aiding vessel operators in collision avoidance strategies [12]. Consequently, the International Maritime Organisation (IMO) has mandated the adoption of AIS for vessels exceeding a certain size [13, 14].

Despite its invaluable utility, AIS is not without vulnerabilities, including the potential for errors and susceptibility to cyberattacks. These concerns have prompted examinations into the accuracy and integrity of position data transmitted via AIS, as in [15]. Nevertheless, AIS remains a prevailing choice within the research community for tracking maritime vessels efficiently and cost-effectively [16, 17].

2.1 AIS-based Safety Approaches

In terms of safety, AIS data have been used for traffic management, identifying potentially dangerous locations for maritime transportation, and providing relevant information for traffic control. One of the critical applications of AIS is traffic management, which involves spatial analysis of near collisions to identify potentially hazardous locations in maritime transportation. As an example, Wu et al. [18] investigated vessel conflicts in the Southeast Texas waterway, examining features such as vessel size, spatial distribution, time of day, and collision risk levels using the Vessel Conflict Ranking Operator (VCRO) model. This research identified hotspots with high frequencies of vessel conflicts and evaluated the impact of time of day on conflict density in each hotspot. As an alternative, Yoo et al. [19] estimated near-collision density in coastal areas based on AIS data, pinpointing high collision-risk locations using parameters such as DCPA (Distance at Closest Point of Approach) and TCPA (Time at Closest Point of Approach) derived from ship location, speed, and course data.

As a second major aspect, anomaly and novelty detection (e.g., [20, 21]) has been performed on AIS data with goals such as the identification of conspicuous ships, the detection of deviation from expected manoeuvres, or an online observation and assessment system. For instance, Chen et al. [22] developed a method to detect anomalies in the AIS data, considering ship position and speed. This approach classifies abnormal AIS points and processes them based on longitude (lon), latitude (lat), speed, acceleration, and direction information. The method identifies anomaly drift track points by assessing ship manoeuvrability, including parameters such as maximum acceleration, minimum acceleration, maximum distance, and maximum angular displacement. The effectiveness of this method was verified using a data set comprising 156 AIS messages with a cruise length of 110 metres and a 10-second interval recording. Recent work considered AIS data in the Kiel Fjord to identify abnormal ship behaviour, which then serves as a basis for increasing safety in autonomous navigation [23].

2.2 AIS-based Monitoring and Cognitive Reasoning

In terms of monitoring and cognitive reasoning, AIS data have been used for behaviour analysis, reasoning the solution to various decision-related applications. For example, Shelmerdine et al. [24] leveraged AIS data to generate vessel tracks and density maps, illustrating temporal variations between months and distinguishing between different vessels around the Shetland Islands. The analysis revealed variations in vessel routes, especially around island groups. As an alternative, Jensen et al. [25] conducted an analysis of cargo traffic outside of San Francisco Bay, identifying dynamic changes in shipping traffic across seasons, both annually and daily. The study highlighted the challenge of documenting shifts in shipping traffic effectively.

Further work focused on the aspect of situational awareness and reasoning. For example, Kaiser et al. [26] quantified the activity of the service vessels in the US Gulf of Mexico, mapping traffic in different channels for comparative analysis. As an alternative, Gao et al. [27] proposed a framework based on genetic programming to predict vessel positions. This predictive model relies on the vessel's past positions within a specific time interval, enabling rapid and effective responses during emergencies.

2.3 AIS as a Basis for Navigation Decisions

AIS data has been used for motion prediction and in autonomous navigation and assistant systems. For example, Li et al. [28] conducted a comprehensive comparison of trajectory prediction methods, including traditional techniques (e.g., Kalman Filter, Support Vector Progression, etc.) and deep learning methods (e.g., Recurrent Neural Network, Long Short-Term Memory, etc.) on three AIS data sets, representing normal channel, complex traffic, and port areas. In an illustrative example, Kayano and Imazu [29] developed a collision avoidance support system to assist watch officers. The system operates in three phases: firstly, it warns the officer of potential risks; secondly, it selects a safe route; and finally, it recommends this improved route. The subsequent application is traffic flow modelling. For example, Jiang et al. [30] used Monte Carlo simulation to model probabilistic distributions of vessel size, vessel speed, and traffic volume for inland water transport. This research sheds light on the challenges of extracting knowledge from AIS data, which arise from factors such as data volume, incompleteness, noise, and unobserved vessels. As an alternative, Al-Falouji et al. [2] developed a navigation action-conditional forward model and a state-action policy network, utilising the Prediction and Policy-learning Under Uncertainty (PPUU) technique [31]. These models were trained on a state-action data set of ship trajectories constructed from AIS messages.

3 SV-NBA: Spatio-temporal Nautical Behaviour Analysis Framework

3.1 Approach

Many existing research efforts in the field of maritime situation analysis tailor their developments to specific use cases under investigation. In addition to the limited availability of open-source tools for maritime navigation analysis, it is noteworthy that these tools currently provide functionalities that are somewhat constrained, and the integration of existing modelling frameworks presents considerable challenges. PyVT [32] for instance focuses on trajectory processing and analysis, while [33] specialises in maritime activity recognition and forecasting. Recognising the need for a versatile and multifunctional framework for spatio-temporal maritime analysis, the SV-NBA framework was developed to connect various tools including *MovingPandas (MPD)* [34] for efficient large scale trajectory management operations, *tslearn* [35] for time-series analysis, and *scikit-learn* [36] for machine learning functionalities; initialising a comprehensive platform for a wide range of applications in maritime data analysis. The evolving SV-NBA framework is designed to perform critical tasks for maritime situational awareness and cognitive management. These tasks include decoding AIS messages, processing them into usable features, detecting anomalous behaviour, and clustering navigational patterns.

3.2 Challenges of AIS-based Data Reasoning

The frequency of transmitted AIS messages varies widely, ranging from once every second to as infrequent as one message every three minutes. Non-commercial vessels often use unreliable AIS transponders, leading to erratic signal emission or extended signal gaps. Consequently, exchanged AIS signals commonly exhibit noise in terms of both transmission loss and quality due to faulty transceivers. This noise necessitates careful data selection and pre-processing techniques when analysing AIS data for safety-critical applications such as autonomous navigation.

In our analysis of AIS data, we address the following primary challenges:

(i) Data quality: AIS data may be afflicted with errors, omissions, or falsifications in transmissions.

(ii) Data volume: In regions of dense maritime activity, AIS generates a large amount of data, straining both data storage and processing systems.

(iii) Data interpretation: The presence of noise in transmitted data and interruptions in the time-series can lead to a misinterpretation of actual ship movements.

3.3 Trajectory Retrieval

In the following, we outline our approach of extracting meaningful trajectories from real-world noisy AIS data.

AIS Data – Structure, Decoding, and Categorisation: AIS signals follow a structured format designed to convey relevant information about maritime vessels. These signals are transmitted at the bit level in accordance with the *NMEA 0183* Standard [37] while the information encoding adheres to the *ITU-R M.1371* standard [38]. Refer to Table 1 for an overview of the features collected in our work.

Input Data – Formatting and Categorisation: Our framework assumes the prior storage of all AIS information. Currently, in SV-NBA, we do not include real-time recording capabilities. We also expect timestamps to be associated with the raw messages. Basic message decoding of AIS data strings is accomplished using the open-source software *PyAIS* [39].

Extracting Data and Crawling by MMSI: Position reports are, by default, labelled with a unique identifier— their 9-digit Maritime Mobile Service Identity (MMSI) number. Additionally, AIS signals broadcast to any interested party, enabling various public domain websites to track AIS messages and share their information with other providers and the public. We have implemented a straightforward web crawling functionality that facilitates querying one of these public AIS data providers [40] for static ship information. The queries are simple HTTPS-based requests to download HTML files containing all available static vessel information for an MMSI number. This information includes retrieval time (Unix epoch), message type, ship type, ship dimensions, draught, length, and width.

Base Trajectory Reconstruction: Once all the AIS messages have been loaded, decoded, categorised and appended with publicly available vessel information, the actual trajectories can be retrieved. Note that by design, all relevant AIS messages include the unique MMSI number of the sender. Thus, we group and sort the position reports for one day building rudimentary base trajectories supplied with additional information from voyage-related data for 24-hour time intervals. We implemented these trajectories using the *Trajectory* class from MPD, using the timestamps for each position report as an index, and all additional information as data columns in the underlying *dataframe*.

3.4 Trajectory Assembly

Handling trajectory data presents substantial challenges when confronted with imprecise position and time values. In this section, we highlight some specific challenges encountered with the

AIS data collected in the Kiel Fjord: (i) The accuracy of position values may be compromised due to sensor errors, noise, and missing data points. (ii) Trajectory data can be information-rich, particularly in scenarios involving many moving objects. This abundance of data can make computations of ship-to-ship combinatorial metrics, such as CPA or similarity metrics like dynamic time wrapping (DTW), exceedingly computationally expensive. (iii) Trajectories often exhibit gaps or contain missing data points. These gaps can lead to faulty inferences about vessels' movement, particularly when attempting to derive continuous movement patterns.

In the following, we present our approach to tackle these challenges for the goal of constructing meaningful trajectories from imperfect and extensive real-world movement data.

Geofence – Removing Signals of Vessels at Berth: The AIS transceiver system operates with high automation, sending position reports automatically at its base frequency when the radio equipment is active, even if the broadcasting vessel remains stationary. This automation can introduce data anomalies when analysing vessel movements, especially when ships are at anchor on the water. Similarly, vessels berthed in locations such as marinas or shipyards can generate problematic outliers in trajectory time-series data if not properly managed. While these scenarios contribute noise to the data, they still provide valid position information. In our framework, we distinguish them from the data of ships underway.

One common approach to address this issue is geofencing, which involves removing all position reports originating from within recognised berthing areas. Our framework offers automatic geofencing capabilities when provided with machine-readable maps of the environment. We have implemented geometric calculations based on the *shapely* framework [41], allowing for well-established tests of geometric inclusion, such as identifying points within polygon boundaries.

Splitting Trajectory Data: Detecting or extracting stops from movement trajectories poses challenges due to tracking inaccuracies, where movement speed rarely reaches true zero. Ships under anchor or moored conditions, for instance, appear to often move around the stop location. Additionally, trajectory interruptions, or gaps, are a frequent occurrence in AIS-based tracks. These interruptions can result from transmission halts or reception losses. Also, in certain scenarios, objects in motion might exit the observation area for extended periods before reappearing and continuing their recorded trajectory.

To tackle these challenges, a systematic two-step approach is utilised to break down trajectories into coherent and uninterrupted trips. Initially, stops are pinpointed if movement stays within an area for a predetermined minimum period. Trajectories are divided into trips, excluding these stops to enhance the precision of the analysis. In a following step, the resulting trajectories undergo further refinement by segmenting them at observation gaps that exceed a minimum duration limit.

Outlier Removal: In AIS data analysis, the presence of outliers can significantly undermine the quality of results. This issue is particularly critical when dealing with geospatial time-series data. The interconnectedness of the data dimensions introduces the potential for spatial anomalies to coexist with missing values in the time dimension, ultimately resulting in trajectories that are entirely implausible. For example, historical ship trajectories might appear to display sudden speed variations in cases of significant time gaps between data points. Inaccuracies in latitude and longitude measurements can exert a significant adverse influence, particularly in clustering algorithms employed for traffic flow modelling.

To address these challenges, several well-established methods are available for trajectory analysis, including robust statistical anomaly detection methods (as discussed in [42]). In this context, our focus is on unsupervised statistical techniques designed to identify outliers within trajectories. The implemented anomaly detection supports two main approaches: (i) The *Z-score* algorithm (as introduced in [43]), and (ii) A variation of the *Generalised Extreme Studentised Deviate* method (outlined in [44]).

Trajectory Smoothing: Various filtering techniques have been employed to address the trajectory smoothing challenge for AIS data. These attempts include classical methods like the constant velocity linear Kalman Filter. Additionally, more complex solutions involving non-linear filters have been explored (as discussed in [45]).

In the SV-NBA framework, we applied the *Stone Soup*'s Kalman Filter implementation, specifically using their *nearly constant velocity model* (as detailed in [46]).

Trajectory Interpolation: Due to the asynchronous nature of AIS signal transmission, ships broadcast signals at varying rates, making it challenging to generate features for analysis.

To address this issue, we employ the *Piecewise Cubic Hermite Interpolating Polynomial* (PCHIP) algorithm (see [47]). This algorithm constructs a polynomial approximation of the movement data, optimised to closely match the original data points. It maintains monotonicity, preventing spurious oscillations in the interpolated results. This is particularly advantageous when dealing with ship trajectories, as it ensures more accurate and realistic representations of the vessel's movements. The resulting polynomial representation is then sampled at discrete time-steps to fill gaps in the time-series, ensuring synchronised data indexing for ship-to-ship feature calculations. These polynomial splines have demonstrated strong performance in testing and exhibit the efficiency required for use with large data sets. They provide an approximation of the actual vessel path, including details on speed, acceleration, and changes in direction. Moreover, these features offer greater robustness compared to the information transmitted solely within AIS messages, making them valuable additions to the trajectory data.

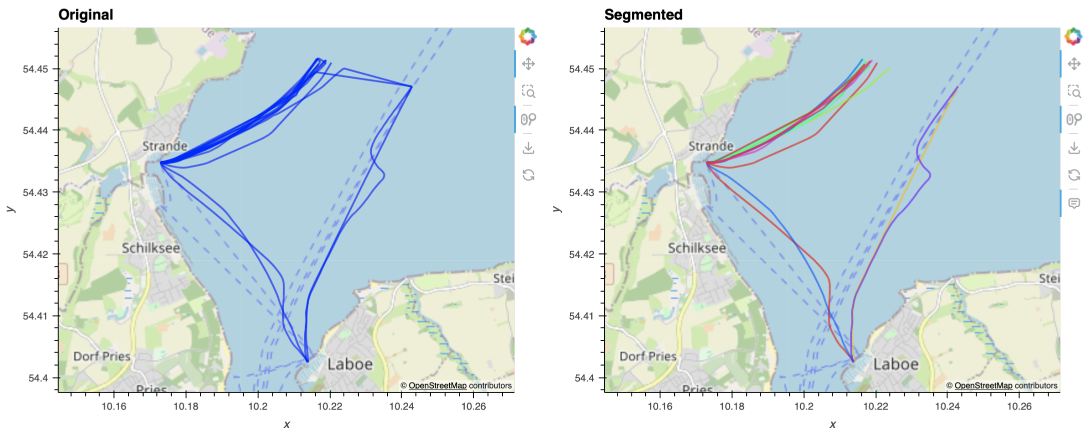
Trajectories Assembly – distinct and meaningful Parts: The methods and procedures mentioned earlier address challenges in AIS data analysis by filtering out noise and correcting trajectory errors. Figure 1 illustrates the impact of the pre-processing steps on a selected ship trajectory. The developed pre-processing significantly improves the usability and analytical power of movement data. Thus, ensuring a continuous representation of movement with reliable values for position, speed, and direction. All of which are essential for accurate analysis, modelling, and informed decision-making.

3.5 Trajectory Mining

In the context of maritime situational awareness–especially when considering autonomous vessels–there are some key differentiating factors compared to other applications like in automotive: The governing navigational rules are harder to interpret, the available information on the water is sparse, and long deceleration times and restricted manoeuvrability need for a much wider decision-making time horizon.

Careful and precise modelling of navigational situations, in combination with high-quality movement data of vessels in demanding situations, allows for a meaningful insight into these decision-making processes.

Here, we focus on mining high-level nautical situation behaviour by involving multi-dimensional feature clustering, traffic flow modelling, and geographical density map visualisation of the pre-processed trajectory data. The first two branches of analysis supported by the SV-NBA framework utilise straightforward analysis of trajectory data, with various filter



(a) (b)

Figure 1: (a) Unprocessed trajectory of a pilot ship from 24th March 2022, (b) Segmented trajectory of the same voyage into distinct trips (colour-coded), retrieved by removing stops, time gaps, and waypoint outliers.

possibilities to control the selection of the time-series to be analysed.

First, the set of all observed ship trajectories can be grouped into their distinct AIS ship types, allowing for a fine control of the analysis level of the heterogeneous data to be mined. Secondly, the length of the analysed time-series can be adjusted by a sliding window approach of variable size. The distinct analysis capabilities and the processing order are illustrated in Figure 2.

Pipeline A: After selecting the movement data, clustering algorithms are applied to the features extracted from the vessels' trajectories. These features, including geographical positions, speed, and direction, are grouped into clusters within a multi-dimensional feature space. Our framework provides wrappers for commonly used clustering algorithms such as *HDBScan*, *K-Means*, and *GMM-Clustering*.

The clustering methods can utilise standard distance metrics such as *Manhattan* and *City-Block*, or nautical-oriented metrics such as *haversine* distance for spherical coordinates, and *DTW* or *Hausdorff* distances for time-series.

Pipeline B: The next branch assesses traffic flow. We employ an algorithm for spatial generalisation that aggregates movement data and extracts flow information from the underlying trajectories based on [48]. The algorithm allows tuning of the grid areas in which the aggregation is performed and determines significant points of the trajectories, that represent the essential characteristics of the movement patterns–capturing the traffic flow. There are open-source implementations available, here we used the one provided by MPD.

Pipeline C: The targeted geographical area for analysis is partitioned into equal-sized squares (cells), transforming it into a grid map. Subsequently, an analysis time window is applied to select active ship trajectories, from which features are generated for each cell. These features serve a dual purpose: (i) They enable effective statistically-based comparison of geospatial sub-regions, such as median speed, and they can be visually represented through plots. Heat maps of this nature prove invaluable when examining risk zones or handling vast datasets that would be otherwise challenging to interpret. (ii) The generated grid-based features simplify the integration of machine learning algorithms, such as Deep Neural Networks (DNN).

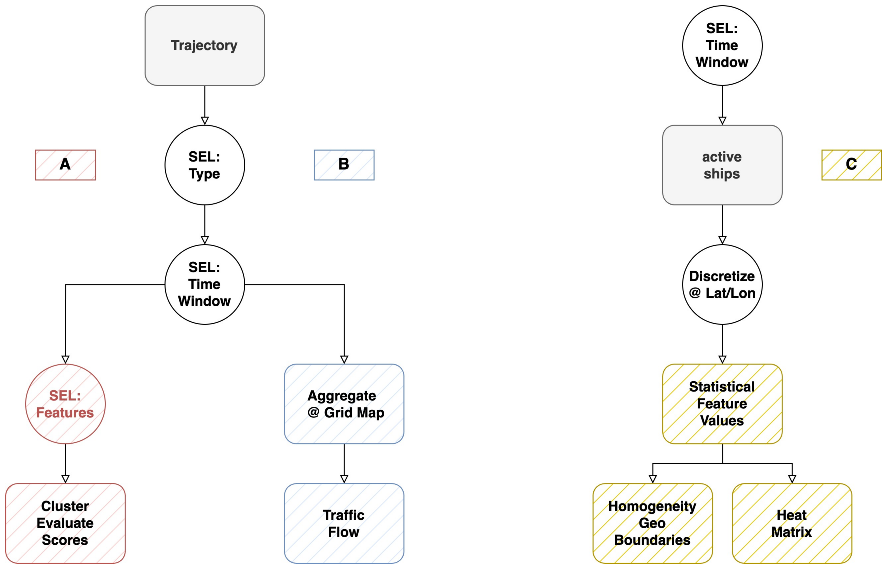


Figure 2: The distinct Trajectory Mining approaches are highlighted as A, B and C, to illustrate the processing steps.

4 Experiments

We selected the Kiel Fjord marine environment as our evaluation use case due to its distinctive maritime characteristics. This area features a narrow navigation waterway with a dense flow of maritime traffic, resulting in a multitude of dynamic scenarios that involve mutual interactions among ships' navigation behaviours. Moreover, the presence of diverse vessel types, including cargo ships, ferries, and leisure boats, makes it an ideal location for analysing the functionality of the SV-NBA framework in terms of situational awareness and behaviour analysis.

4.1 Setup

We collected a whole day's worth of AIS data from 24th March 2022 using an antenna located next to the Kiel Fjord. The dataset comprises a high traffic volume and diverse ship activities.

The implementation of the SV-NBA framework, along with the three designated pipelines described in Section 3.5, has been made publicly available[1] to facilitate greater access and transparency.

For data analysis, geofencing is used to ensure a focused data set for the Kiel Fjord region. Furthermore, we generated a comprehensive map covering various aspects of Kiel's maritime topology, including the shoreline, waterways marked by buoys, restricted zones, and marinas.

[1] https://github.com/CAPTN-sh/SV-NBA

These polygons serve as essential spatial filters for our analyses, allowing us to refine and tailor our investigations to specific areas of interest within the Kiel Fjord.

The hyperparameters selected for the pre-processing of AIS geospatial data are crucial in shaping the initial data for further analysis. We use the *WGS84* ellipsoidal map projection for the geographical coordinates. To calculate distances between two geographical points, the *vicenty* method is applied. In terms of anomaly detection, the *z-score* test is utilised, with a threshold set at 2σ. For trajectory splitting, various parameters come into play. The maximum time gap allowed for the gap splitter is 10 minutes. To distinguish between underway and stopped vessels, the speed splitter considers a minimum underway speed of 0.1 knots and a minimum duration threshold of 10 minutes. The stop splitter segments trajectories based on a diameter of the stop region of 50 metres and a minimum duration threshold of ten minutes. The resulting trajectories are then smoothed using the *Kalman filter with nearly-constant velocity* method, with a process noise $\sigma_p = 0.5$ and measurement noise $\sigma_m = 1$. For interpolation, the *PCHIP* method is employed with a step size of six seconds. Further details and explanations for these methods can be found in the Approach Section 3.4.

4.2 Evaluation

Pipeline A – Navigational Behaviour Clustering: In the experimental evaluation of Pipeline A (see Section 3.5), we create a feature set that includes SOG, COG, and minimum distance to the shoreline.

For clustering trajectories, we employ two distinct methods based on these three features:

1. HDBSCAN clustering: We apply the HDBSCAN clustering algorithm with the following parameters: a minimum cluster size of 100 instances, $\epsilon = 5$, and the L_1 norm as the distance metric.

2. K-Means clustering: We use the K-Means clustering algorithm with a specified number of clusters, $k = 5$.

The result of the K-Means clustering is illustrated in Figure 3. Notably, distinctive features, are clustered in green (*cluster 1*), exhibiting a median COG of approximately 48°, in contrast to the vessels in red (*cluster 0*), which have a median COG of approximately 218°.

Pipeline B – Traffic Flow Analysis: In this experiment, We apply the procedure outlined in Section 3.5, Pipeline B, to aggregate trajectories and generate a traffic flow model. This pipeline is designed to summarise and extract traffic flow information from the trajectories, providing insights into movement patterns, intensity, and interactions within a geographic area. We have coupled this flow analysis with a clustering step that groups similar trajectories. These clusters help categorise navigated routes based on their distinct movement patterns. By combining both of these analysis methods, we gain deeper insights into the movement behaviour of vessels, both within and between clusters.

The resulting traffic flow model is visually represented in Figure 4. In this figure, the centroids of the clusters are marked in red, with the size indicating the number of significant trajectory points grouped within each cluster. Additionally, the flow lines are highlighted in blue, with varying widths representing the number of trajectories moving from one group to another. It is noticeable that within the geographical coordinates *lat* $\in [54.38 - 54.4]^\circ$ and *lon* $\in [10.19 - 10.21]^\circ$, the blue flow representation line appears wider than others, indicating a higher flow of trajectories in that region, likely due to the narrow passage.

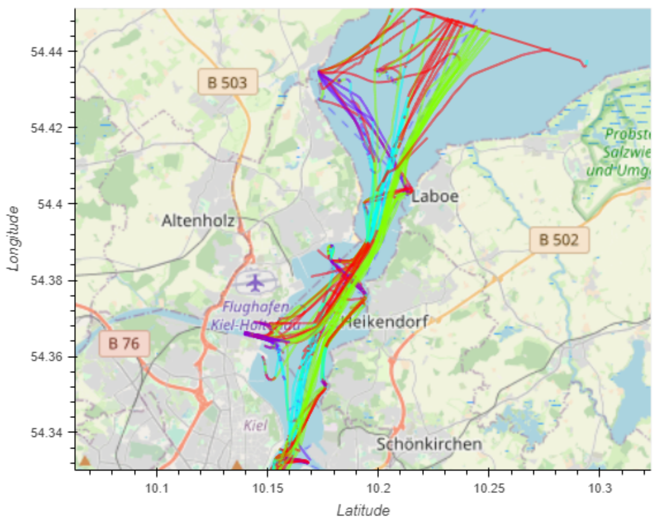


Figure 3: Cluster of distinct movement patterns of sailing ships for a 24h time interval. The five clusters of K-Means are distinguished by colour. The feature set combines ship speed, COG, and distance to the shore.

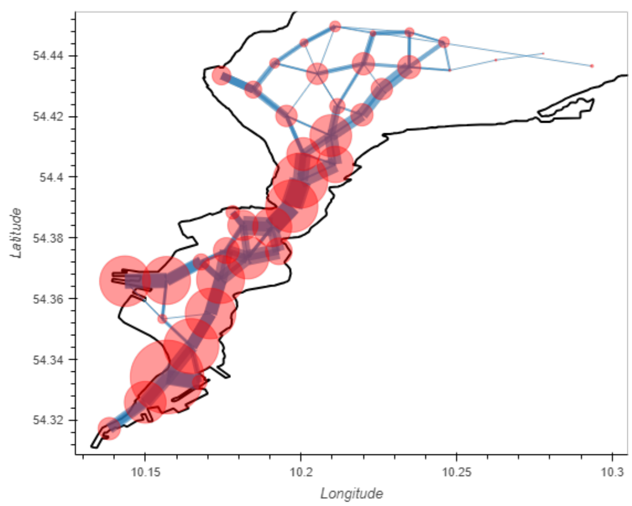


Figure 4: Traffic Flow density (blue) between the centroids of the critical point groups (red). Based on sailing vessels' movement for a 24h time interval.

Pipeline C – Grid Map Analysis: In our final showcase, we concentrate on the visual analysis of short to medium-term trends in navigational behaviour using the methods described in Section 3.5, Pipeline C. We generate discretised, geospatial and statistical features from movement data. We partition the geofencing rectangle, into a grid of 100×100 individual cells. Within this grid, we extract interpolated speed data from the trajectories and link them to their respective grid cells. To capture a statistical measure of the speed distribution, we apply the *median*.

The resulting geospatial feature map, see Figure 5, allows for the visualisation of risk zones in terms of fast-moving vessels.

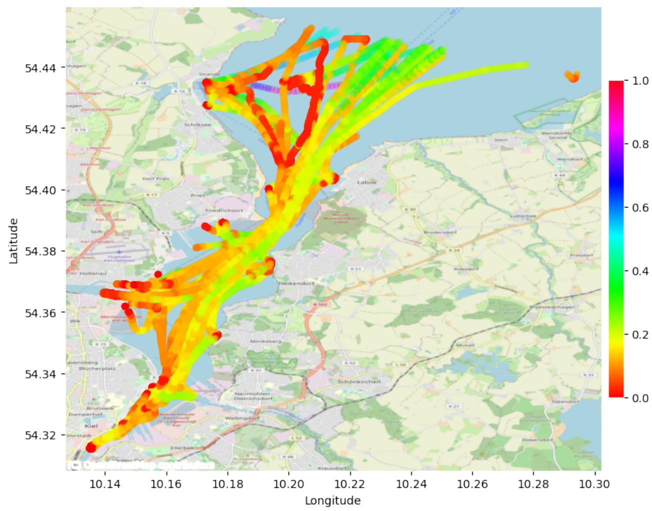


Figure 5: Heat map generated for the median speed feature value of ships in a 100×100 grid of the Kiel Fjord.

This approach enables a comprehensive analysis of navigational behaviour trends, providing insights into both spatial and statistical aspects of vessel movement over time.

5 Conclusion

We presented a novel framework for analysing the navigational behaviour of surface vessels using historical AIS data. SV-NBA framework integrates a variety of different approaches, including trajectory reconstruction, trajectory processing, feature generation from spatio-temporal data, behaviour analysis, and outlier detection, facilitating cognitive decision-making and situational awareness. We evaluated our framework using a sample AIS data set from the Kiel Fjord area, comprising all AIS messages captured for one day.

We highlighted some main capabilities and demonstrated that our framework can be utilised for various cognitive situational modelling and awareness tasks, including:

- Model normal navigational behaviour, which can be used to avoid collisions and other dangerous situations.
- Understand the overall flow of traffic on waterways, which can be used to improve maritime safety and efficiency.
- Identify anomalous behaviour, such as ships that are deviating from their planned course or sailing at unsafe speeds.

Our framework is designed to incorporate a digital twin of maritime environments, allowing the simulation of the behaviour of ships in different conditions, and to develop more robust and generalised ship behaviour models. Additionally, our framework can be used to detect anomalous navigational behaviour, adding more dimensions to maritime situational awareness. These research points are planned as future work, based on the developed SV-NBA framework.

In summary, our framework promises to advance surface vessel behaviour analysis, benefiting maritime safety and efficiency. Its extensibility and simulation capabilities open novel avenues for further research.

References

[1] Aidan H While, Simon Marvin, and Mateja Kovacic. Urban robotic experimentation: San francisco, tokyo and dubai. *Urban Studies*, 58(4):769–786, 2021.

[2] Raphael Schwinger, Ghassan Al-Falouji, and Sven Tomforde. Autonomous ship collision avoidance trained on observational data. In *Architecture of Computing Systems*, pages 296–310. Springer Nature Switzerland, 2023.

[3] Andrzej Felski and Karolina Zwolak. The ocean-going autonomous ship-challenges and threats. *Journal of Marine Science and Engineering*, 8(1):41, 2020.

[4] Sheng Liu, Bowen Xing, Bing Li, and Mingming Gu. Ship information system: overview and research trends. *Int. J. Nav. Archit. Ocean Eng.'*, 6(3):670–684, 2014.

[5] LP Perera, JM Rodrigues, R Pascoal, and C Guedes Soares. Development of an onboard decision support system for ship navigation under rough weather conditions. *Sustainable Maritime Transportation and Exploitation of Sea Resources*, 837:837–844, 2012.

[6] Nikita Smirnov and Sven Tomforde. Navigation support for an autonomous ferry using deep reinforcement learning in simulated maritime environments. In *2022 CogSIMA*, pages 142–149. IEEE, 2022.

[7] Liye Zhang, Qiang Meng, Zhe Xiao, and Xiuju Fu. A novel ship trajectory reconstruction approach using ais data. *Ocean Engineering*, 159:165–174, 2018.

[8] Roman Smierzchalski and Zbigniew Michalewicz. Modeling of ship trajectory in collision situations by an evolutionary algorithm. *IEEE Transactions on Evolutionary Computation*, 4(3):227–241, 2000.

[9] Yongfeng Suo, Wenke Chen, Christophe Claramunt, and Shenhua Yang. A ship trajectory prediction framework based on a recurrent neural network. *Sensors*, 20(18):5133, 2020.

[10] H Rong, AP Teixeira, and C Guedes Soares. Ship trajectory uncertainty prediction based on a gaussian process model. *Ocean Engineering*, 182:499–511, 2019.

[11] Ghassan Al-Falouji, Lukas Haschke, Dirk Nowotka, and Sven Tomforde. Self-explanation as a basis for self-integration – the autonomous passenger ferry scenario. In *ACSOS'23 Companion*. IEEE, 2023.

[12] IMO. Resolution A.1106(29). `https://bit.ly/3EIC7H3`, 2001. [Online; accessed 3-Oktober-2022].

[13] IMORULES. SOLAS regulation V/19. `https://bit.ly/3PJQBMY`, 2022. [Online; accessed 3-Oktober-2022].

[14] Bundesamt für Seeschifffahrt und Hydrographie. German Traffic Regulations for Navigable Maritime Waterways. `https://bit.ly/3PHn3ib`, 2022. [Online; accessed 3-Oktober-2022].

[15] Andrzej Felski and Krzysztof Jaskólski. The integrity of information received by means of ais during anti-collision manoeuvring. *TransNav*, 7(1):95–100, 2013.

[16] Kurt D Schwehr and Philip A McGillivary. Marine ship automatic identification system (ais) for enhanced coastal security capabilities: an oil spill tracking application. In *OCEANS 2007*, pages 1–9. IEEE, 2007.

[17] Duong Nguyen, Rodolphe Vadaine, Guillaume Hajduch, René Garello, and Ronan Fablet. A multitask deep learning architecture for maritime surveillance using ais data streams. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 331–340. IEEE, 2018.

[18] Xing Wu, Aesha L Mehta, Victor A Zaloom, and Brian N Craig. Analysis of waterway transportation in southeast texas waterway based on ais data. *Ocean Engineering*, 121:196–209, 2016.

[19] Sang-Lok Yoo. Near-miss density map for safe navigation of ships. *Ocean Engineering*, 163:15–21, 2018.

[20] Ghassan Al-Falouji, Christian Gruhl, Torben Neumann, and Sven Tomforde. A heuristic for an online applicability of anomaly detection techniques. In *2022 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion (ACSOS-C)*, pages 107–112. IEEE, 2022.

[21] Ghassan Al-Falouji, Christian Gruhl, and Sven Tomforde. Digital shadows in self-improving system integration: A concept u sing generative modelling. In *2021 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion (ACSOS-C)*, pages 166–171. IEEE, 2021.

[22] Shuguang Chen, Yikun Huang, Wei Lu, et al. Anomaly detection and restoration for ais raw data. *Wireless Communications and Mobile Computing*, 2022, 2022.

[23] Martin Goller, Ingo Thomsen, Ghassan Al-Falouji, and Sven Tomforde. Abnormal behaviour detection of self-adaptive agents in traffic environments. In *ACSOS'23 Companion*, page 1–8. IEEE, 2023.

[24] Richard L Shelmerdine. Teasing out the detail: How our understanding of marine ais data can better inform industries, developments, and planning. *Marine Policy*, 54:17–25, 2015.

[25] Caitlin M Jensen, Ellen Hines, Barbara A Holzman, Thomas J Moore, Jaime Jahncke, and Jessica V Redfern. Spatial and temporal variability in shipping traffic off san francisco, california. *Coastal Management*, 43(6):575–588, 2015.

[26] Mark J Kaiser. Service vessel activity in the us gulf of mexico in support of the oil and gas industry using ais data, 2009–2010. *Marine Policy*, 63:61–80, 2016.

[27] Miao Gao and Guo-You Shi. Ship-collision avoidance decision-making learning of unmanned surface vehicles with automatic identification system data based on encoder—decoder automatic-response neural networks. *Journal of Marine Science and Engineering*, 8(10):754, 2020.

[28] Huanhuan Li, Hang Jiao, and Zaili Yang. Ais data-driven ship trajectory prediction modelling and analysis based on machine learning and deep learning methods. *Transportation Research Part E: Logistics and Transportation Review*, 175:103152, 2023.

[29] Jun Kayano and Hayama Imazu. A study of a collision avoidance support system. In *2009 IEEE International Conference on Systems, Man and Cybernetics*, pages 244–249. IEEE, 2009.

[30] Zhonglian Jiang, Zhen Yu, Daiyong Zhang, Xiumin Chu, and Qiong Yang. Characteristics of vessel traffic flow during waterway regulations: A case study in the yangtze river. In *2019 5th International Conference on Transportation Information and Safety (ICTIS)*, pages 364–368. IEEE, 2019.

[31] Mikael Henaff, Alfredo Canziani, and Yann LeCun. Model-predictive policy learning with uncertainty regularization for driving in dense traffic. *arXiv preprint arXiv:1901.02705*, 2019.

[32] Ye Li, Hongxiang Ren, and Haijiang Li. Pyvt: A toolkit for preprocessing and analysis of vessel spatio-temporal trajectories. *SoftwareX*, 21:101316, 2023.

[33] Manolis Pitsikalis, Alexander Artikis, Richard Dreo, Cyril Ray, Elena Camossi, and Anne-Laure Jousselme. Composite event recognition for maritime monitoring. In *Proceedings of the 13th ACM International Conference on Distributed and Event-based Systems, DEBS 2019, Darmstadt, Germany, June 24-28, 2019*, pages 163–174. ACM, 2019.

[34] Pipy - movingpandas. https://pypi.org/project/movingpandas/. Online; accessed 2023-09-15.

[35] Pipy - tslearn. https://bit.ly/3PJgmNK. Online; accessed 2023-09-15.

[36] sklearn - cluster. https://bit.ly/45WSqfe. Online; accessed 2023-09-15.

[37] National marine electronics association - nmea 0183. https://bit.ly/3sWoRMv. Online; accessed 2023-09-15.

[38] M.1371: Technical characteristics in the vhf maritime mobile frequency band. https://bit.ly/3Pl02j0. Online; accessed 2023-09-15.

[39] Pypi - pyais. https://bit.ly/3LovJsj. Online; accessed 2023-09-15.

[40] My ship tracking free realtime ais vessel tracking vessels finder map. https://bit.ly/3PqtWnw. Online; accessed 2023-09-15.

[41] Pypi - shapely. https://bit.ly/46kuanj. Online; accessed 2023-09-15.

[42] Peter J. Rousseeuw and Mia Hubert. Anomaly detection by robust statistics. *WIREs Data Mining and Knowledge Discovery*, 8(2):e1236, 2018.

[43] Schyler W. Huck, Tracey L. Cross, and Sheldon B. Clark. Overcoming misconceptions about z-scores. *Teaching Statistics*, 8(2):38–40, 1986.

[44] S. R. Paul and Karen Y. Fung. A generalized extreme studentized residual multiple-outlier-detection procedure in linear regression. *Technometrics*, 33(3):339–348, 1991.

[45] Krzysztof Jaskólski, Lukasz Marchel, Andrzej Felski, Marcin Jaskólski, and Mariusz Specht. Automatic identification system (ais) dynamic data integrity monitoring and trajectory tracking based on the simultaneous localization and mapping (slam) process model. *Sensors*, 21, 2021.

[46] Pypi - stone soup. https://bit.ly/3Luf2Ma. Online; accessed 2023-09-15.

[47] Nobili F., Fritscht, , E. R., and Carlson. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17:238–246, 1980.

[48] Natalia Adrienko and Gennady Adrienko. Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics*, 17(2):205–219, 2011.

Appendix

Table 1: List of features. Basic AIS Features are marked by $*$.

	feature	description	unit
static	mmsi*	unique MMSI number	int
	imo	unique IMO number	alph./int
	ship_type*	type, e.g., 'sailing'	int
	length*	dimension	m
	width*	dimension	m
	draught*	dimension	m
	tonnage	gross tonnage	m^3
	dwt	dead weight tonnage	t
ship	epoch	unix time	s
	lat*	latitude	degree dec.
	lon*	londitude	degree dec.
	status*	navigational status	int
	heading*	nautical heading	deg.
	COG*	Course Over Ground	deg.
	speed*	speed	kn
	ROT*	Rate Of Turn	deg./s
	man*	nav. manoeuvre	int
	ROA	Rate Of AIS-signals	t
	i-speed	interp. speed	kn
	i-COG	interp. COG	deg.
	i-ROT	interp. ROT	deg./s
	i-accel.	interp. accel.	kn/s^2
	i-lat	interp. lat	deg. dec.
	i-lon	interp. lon	deg. dec.
	DTS	distance to shore	m
	IWW	vessel in waterways	bool
ship/ship	CPA	Closest Point of Appr.	m
	TCPA	time to CPS	s
	BEA	bearing	degrees
	RBEA	rel. bearing	degrees
	RSPE	rel. speed	kn
	dist	distance	m