APPLICATION OF A SIMULATION-BASED V&V APPROACH TO AUTOMATIC VOYAGE PLANNING

Arnold Akkermann[1] and Bjorn Age Hjollo[2]

[1]German Aerospace Center, Institute of Systems Engineering for Future Mobility, Department Application and Evaluation Oldenburg, Germany
[2]NAVTOR AS, Egersund, NORWAY

ABSTRACT

Automatic route planning for seagoing vessels, generated by a shore-based service provider, should be highly reliable. Even if this route planning has to be adapted to the specifications of the ship on board, the route should always be navigable and safe. Depending on the relevant parameters such as estimated time of arrival (ETA), speed, etc., it should therefore not deviate significantly from the usual routes for tankers, ferries, etc. between two ports. In the past, the quality of the route planning carried out by the on-board navigation depended, among other things, on the experience and training of the personnel. As a result, these plans did not necessarily represent the optimal route. The automatic route generated by NAVTOR starts with the length of the route, which is intended to be as short as possible while still being fundamentally safe to navigate, thus helping to reduce CO2 emissions.

KEYWORDS

AIS data cleaning, Creation of two reference routes to be used for comparison, Trustworthiness of generated Auto-Routes, Verification and Validation (V&V) in the virtual world.

1. INTRODUCTION

Reducing fuel consumption and hence CO2 emissions from shipping has become an increasingly dominant issue in maritime transport. The factors involved are many and highly complex. One key factor is the route taken by ocean-going vessels between ports scattered around the world. These routes should be as short as possible, but also (in principle) safe for navigation. In Use Case (UC) 2 of the TRANSACT project (Transform safety-critical cyber-physical systems into distributed solutions for end-users and partners), the company NAVTOR is developing a machine learning (ML)-based approach for the generation of worldwide routes for seagoing vessels. As a starting point for the implementation, NAVTOR is using its e-Navigation Suite. Figure 1 shows the state of the NAVTOR e-Navigation framework prior to TRANSACT. Today, routes in this context are planned manually using the NavStation on board or ashore and subjected to a safety check in accordance with the IMO (International Maritime Organisation) specification [1].

David C. Wyld et al. (Eds): SIPR, NCWC, BIBC, SOFEA, DSML, ARIA, NLP, CSEN -2023
pp. 109-129, 2023. CS & IT - CSCP 2023 DOI: 10.5121/csit.2023.131909

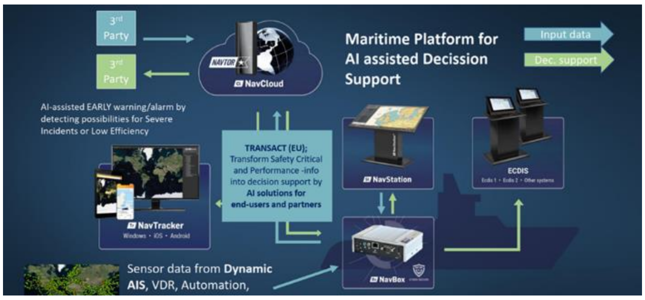


Figure 1: NAVTOR e-Navigation framework prior to TRANSACT (Source NAVTOR)

Within this test, the navigational safety of a planned route can be determined or established. Once the route has been established ashore, it is transmitted to the vessel via the NavBox and a satellite link, where it must be checked and/or confirmed by the navigator. This route can then be transferred to an autopilot, for example. The route is monitored on shore using the NavTracker application. Today, ship routes in this context are planned manually using the NavStation on board or ashore and subjected to a safety check in accordance with IMO specifications. During this check, the navigational safety of a planned route can be determined or established. Once the route has been established ashore, it is transmitted to the vessel via the NavBox and a satellite link, where it must be checked and/or confirmed by the navigator. This route can then be transferred to an autopilot, for example. The route is monitored ashore using the NavTracker application. As part of this project, NAVTOR is developing an edge cloud-based infrastructure that will enable machine learning (ML)-based automatic route generation. NAVTOR has selected three consulting services to be integrated into the Maritime e-Navigation suite: An external advisory service hosted on an external cloud accessible from the cloud layer, an ML-enhanced advisory service to be integrated into the e-Navigation suite, and an internal advisory service hosted on both the edge and cloud layers. These three advisory services are distributed across the edge-cloud tier. The generated route is proposed to the on-board navigator and must be verified by him to ensure safe navigation. The aim of these auto-routes is to suggest a short distance between two ports that is generally safe from a navigational point of view. The ML-based route suggestions are based on the analysis of historical and real-time AIS data from NAVTOR. Our work consists of verifying and validating the generated NAVTOR Auto-Routes. Our research question based on this task is: How can an ML-based auto-route between two ports be verified and validated? (RQ).

2. RELATED WORK

We see the adaptation of appropriate methodologies from other domains (automotive, aerospace) as an appropriate way to answer the RQ, given the high investment and innovative nature of the research in these areas. In [2] four qualification methods are described to achieve a certain level of confidence. However, not all methods are required: different methods are recommended based on the target ASIL (Automotive Safety Integration Level). This standard is still the basis for a safe system, safe hardware and safe software that can operate independently and safely in the event of a fault. Brat and Jonsson [3] discussed the challenges in V&V of autonomous systems designed for space exploration. This includes the space domain. In [4], Zou et al. discussed safety evaluation of probabilistic airborne collision avoidance systems and proposed a genetic algorithm to search for undesirable situations. Included to cover probabilistic approaches. Virtual testing is one of the key steps in assessing the performance and safety of autonomous vehicles (AVs). The core component of this step is a high-fidelity virtual simulator. These simulators contain a world model and a physic engine in which a virtual vehicle can be navigated and in which the dynamics of the vehicle can be modelled [5]. Assuming that this vehicle is equipped with the same

Automated Driving System (ADS) software [6] as a real vehicle, it is possible to run different scenarios of how the AV would react to certain traffic situations. However, the design of traffic scenarios is a complex, lengthy and non-standard process. In a further study by [7], six influencing parameters, namely relative distance, relative speed, temperature, humidity, weather event and visibility, were included in the generation model of test scenarios for an autonomous emergency braking system. Similarly, in [8], the values of eight demonstrative influence parameters related to the kinematic state of the own and target vehicles were varied to generate test scenarios for virtual ADAS (Advanced Driver Assistance Systems) verification and validation. With the exception of [7], other researchers have focused only on the parameters related to the "objects" layer in the PEGASUS model [9]. The SOTIF standard [10], which is currently under development, provides many methods and guidelines for the inclusion of environmental scenarios for preliminary analysis of the concept and final validation. The absence of unacceptable risk from hazards arising from functional inadequacies of the intended functionality or from reasonably foreseeable human misuse is referred to as "Safety of The Intended Functionality" (SOTIF).

3. IMPLEMENTATION OF THE CONCEPT

Existing work on V&V in the automotive domain emphasises the need to use a large number of specially designed scenarios to find some bugs in autonomous functions. We believe that the same criteria should apply to the maritime domain and that there is a need for domain-specific methods for systematic verification and validation of autonomous functions in ships. The majority of the articles defined safety based on the values of either Time to Closest Point of Approach (TCPA) or Distance to Closest Point of Approach (DCPA) [11]. In the context of our RQ, scientific articles dealing with the navigational safety of automatically generated routes for ship navigation are not freely available. Real-world testing in the maritime domain is challenging. This is because the maritime domain is characterized by domain-specific requirements such as the hydrodynamic environment, ship dynamics according to the six degrees of freedom (6DoF) model, exclusive environmental sensors such as radar or AIS (Automatic Identification System), and navigation support systems such as ECDIS (Electronic Chart Display and Information System) or VTS (Vessel Traffic Service). Reproducible test cases for larger electronic components, such as ship bridges, are therefore difficult to test economically. In order to perform acceptable tests, real data must be used as ground truth in the simulation as much as possible. The use of ground truth is essential and must be adapted in its scope to the respective tests in the simulation. In other words, not all real data (such as weather data, traffic data, etc.) must be used for each test in the simulation. In our case, the ground truth is obtained from real-time and historical AIS data. This AIS data is processed to create a route between two ports. This route is then stored in the scenario database (Figure 2) as the basis for a scenario.

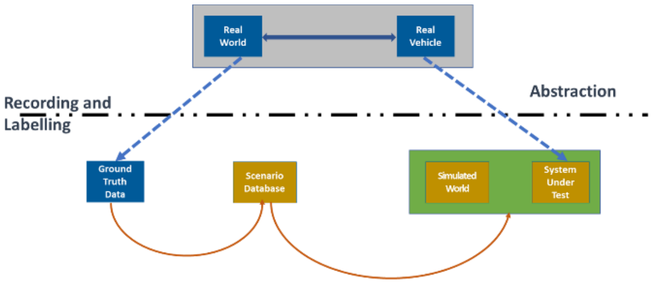
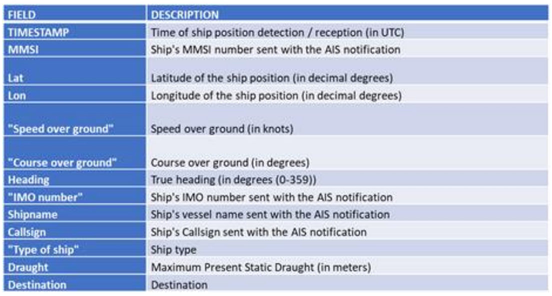


Figure 2: General process abstraction Real World- to Virtual-Testing

Another challenge is the integration of the system under test (SuT) into the simulation. Depending on the actual SuT, this can be physically integrated or has to be simulated by models. The simulated world should allow us to replicate possible real-world events. First, it is important to clarify that a virtual world is not dichotomous to the real, physical or material world that we inhabit as humans [12]. Instead, a virtual world is a subset of the world within the physical world.

3.1. Importance of Ais Data

Today's applications of AIS data [13] have shifted from collision avoidance, identification and tracking to the monitoring of shipping routes, maritime traffic trends, risk analysis and maritime accident investigations, etc. [14]. This paper describes how generated auto-routes are verified and validated against two reference routes generated for a specific vessel and a frequently sailed reference route between two ports. The two reference routes are also generated from historical and real-time AIS data. However, we do not use the AIS data provided by NAVTOR, but a variety of freely available and commercial AIS data. In this way we want to achieve the best possible solution, although not completely independent of the Auto-Route generated by NAVTOR. The methods and models we use to generate the two reference routes are completely different to NAVTOR's AI-based approach. This is because only a comparison of independent models/methods can identify differences in route guidance that need to be investigated in more detail. Both reference route generation involves the application of several mathematical models, some of which overlap, to achieve a high level of confidence. AIS data provide real-time trajectory data that can be used to monitor the navigational status of ships and trigger alerting mechanisms for ship collision avoidance, maritime surveillance, trajectory clustering, vessel traffic flow prediction and marine casualty investigations [15]. The AIS communication network can be divided into two parts, the ship borne AIS station and the shore station. The ship borne AIS stations transmit ship related static and dynamic information in the very high frequency (VHF) band at regular frequency according to the International Telecommunication Union (ITU) specification. The static information, which is manually entered or updated, mainly includes identification, call sign, ship name, dimensions, type, etc. [16]. Dynamic information, e.g. from the Global Navigation Satellite System (GNSS), includes time stamp, speed over ground (SOG), course over ground (COG), position (in latitude and longitude), etc. (Figure 3).



FIELD	DESCRIPTION
TIMESTAMP	Time of ship position detection / reception (in UTC)
MMSI	Ship's MMSI number sent with the AIS notification
Lat	Latitude of the ship position (in decimal degrees)
Lon	Longitude of the ship position (in decimal degrees)
"Speed over ground"	Speed over ground (in knots)
"Course over ground"	Course over ground (in degrees)
Heading	True heading (in degrees (0-359))
"IMO number"	Ship's IMO number sent with the AIS notification
Shipname	Ship's vessel name sent with the AIS notification
Callsign	Ship's Callsign sent with the AIS notification
"Type of ship"	Ship type
Draught	Maximum Present Static Draught (in meters)
Destination	Destination

Figure 3: Main AIS static and dynamic information

Shore-based stations are mainly set up by the authorities on shore. While the ship borne AIS station is transmitting messages, ships and shore stations in the vicinity can receive these messages and display them on the Electrical Navigation Chart (ENC). In this way, AIS enables communication between ships and shore-based authorities and helps maritime administrations, ship owners and ship pilots to determine the position and status of ships. However, the detection probability of AIS is not a purely geometric matter, but depends on the VHF propagation, which is affected by ground conductivity, atmospheric conditions, receiver sensitivity, antenna attenuation, signal shadowing, radio interference and more. Nominal reporting intervals for data transmission vary from 2 s to 6 min and depend on the type of AIS station, the group of messages, the navigational status, the speed and the course change of the vessels [13]. Slower vessels transmit kinematic data every 10 seconds, medium-speed vessels every 6 seconds and high-speed vessels every two seconds. If the ship changes course, the transmission intensity increases by a factor of 3 (for slow and medium speed ships).

3.2. Scenario Database

There are millions of shipping routes, ports and different classes of ships and vessels around the world. This almost unmanageable number and variety has to be reduced without giving up the representativeness of reality or limiting it in the long term. In order to achieve an acceptable representativeness of the real-world within the simulation, the simulation scenarios are defined by a high diversity of the mapping of the real-world in the route sailed by seagoing vessels. To this end, a global statistical analysis of oil, grain, iron ore, passenger and container routes is used to filter out the essential routes from the millions of possible shipping routes. The statistical analysis was also carried out to ensure that the most navigable routes through the major shipping lanes (straits) were included. For example, the following straits are selected:

- Bosporus Strait and Strait of Dover,
- Hormuz and Bab-el-Mandeb Strait,
- Gibraltar Strait and Malacca Strait.

In addition, routes have been selected that are navigationally complex and pass through a large number of TSS (Traffic Separation Schemes). A TSS is an area of the sea where shipping is highly regulated. In addition, the routes of different classes of vessels such as containers, tankers, bulk carriers and passenger ships are compared to maintain realism. The result of the static analysis is 386 routes between two ports. These routes are planned with different ship classes/ships and are available as reference routes for comparison with the NAVTOR Auto Route. Statistically, these results cover, on average, the most travelled routes in the world. This test area in the virtual world thus covers a flexible, wide range of routes, ships and ports, resulting in a high degree of representativeness for the real world. The actual reference routes between the ports are created by evaluating historical and real-time AIS data. Each reference route for a given vessel and the most frequently sailed route between the two ports is stored as a scenario in the scenario database.

3.3. Generating of Reference-Routes

Raw AIS data is used to create a reference dataset for the identified routes and to define a matrix of historical AIS data as follows [17]:

$X_{Tj} = [X_1, X_2, \ldots.., X_N]^T$ (1)

where *N* is the total number of AIS messages and

$X_i = [MMS_{Ii}, ti, piT, ci, \upsilon i]$ (2)

$i \varepsilon [1, 2, \ldots..., N]$ is a vector where MMS_{Ii}, ti, piT and υi respectively represent the Maritime Mobile Service Identity (MMSI), time stamp, location (WGS 84 longitude and latitude), course over ground (COG) and speed over ground (SOG). Each vessel is identifiable by its MMSI.

3.4. General Process of Development AIS Reference Routes

The process shown in Figure 4 is to develop two reference routes for comparison with the NAVTOR Auto-Route calculation.

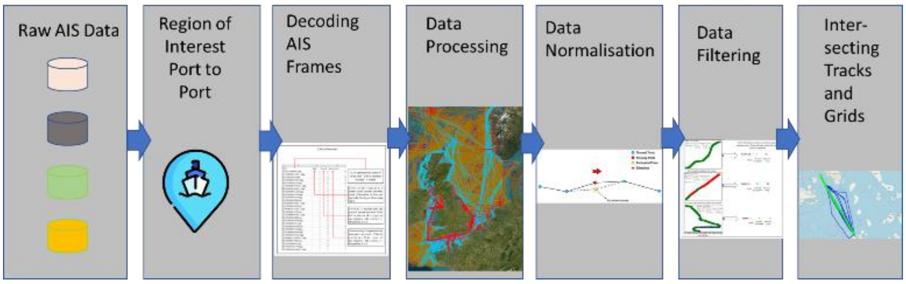


Figure 4: General Process for Deployment Reference Routes

To do this, the reference route must be of very high quality and reliability. These characteristics are achieved by using different mathematical models to build the reference route, which are independent of each other and partly overlapping. In addition, different models have been used for the same task and their results have been compared and evaluated against state-of-the-art techniques. The first step is to extract the relevant data for the route between 2 ports from the available historical AIS data. This step leads from a decision boundary to a validity region. This is our ROI (Region of Interest) (Figure 5).

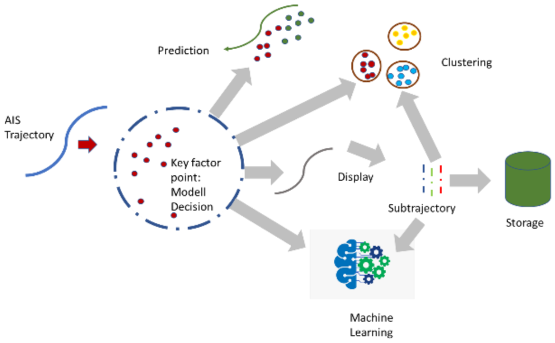


Figure 5: From Decision Boundary to Validity Domain to Model Decision

The data provided by AIS is in the form of an ASCII string like the following:

!AIVDM,1,1,,B,13cita001N0uC7\<hva8qqaif0\<0@,0*1f <=”” p=””>

In order to process this information digitally, it must be decoded. Accurately parsing a complete AIS message is challenging for several reasons. Firstly, the International Telecommunications Union (ITU) Recommendation M.1371-5, the standard that governs the operation of AIS devices, defines 27 different AIS message types, each with its own unique set of approximately 15 data fields [18]. Within each of these different message field sets, the individual message fields themselves often have variable bit structures that require unique logic rules for parsing. Four open source AIS decoders [19] were considered for this work (Table 1).

	Language	**Message Types decoded**	**OS**
libais	C++/Python	1-27	any
AIS Parser	C/Python	1-24	any
NMEA Plus	Ruby	1-16, 18-21, 24, 26	any
AisLib	Java	1-14, 17-19, 21, 24, 27	Windows

Table 1: Summary of Open Source AIS Decoders [19]

A thorough comparison of each is not intended in NMEA (National Marine Electronics Association), but rather to show the limited availability of open source software for advanced AIS decoding, we have included general details about each parser in Table 1 for reference. For our work we mainly chose Kurt Schwehr's libais software [20] for three main reasons: it was able to decode all 27 message types, its Python interface made it easy to implement, and it's C++ code base suggested efficient performance on large datasets. In addition, the documentation and discussion group support for libais seemed superior to other options.

3.5. Data Processing

Pre-processing is a basic step performed at the beginning to improve the quality of the trajectory data and to generate sub-trajectories. This process consists of the following three steps:

- Trajectory extraction and separation: This step is divided into two sub-steps: separation of data from different vessels and separation of data from different trajectories of the same vessel. Vessels are identified by their MMSI. The trajectory of the vessel is discontinuous in different time periods and different trajectories of the same vessel can be separated according to the time stamp of the AIS data.
- Time interval standardization: AIS data transmission time interval has a standard specification; however, a large number of AIS data acquisition time intervals do not conform to the standard [21].
- Data cleaning: Longitudinal and latitudinal data of ship AIS are derived from GPS (Global Positioning System) [21]. The relative positioning of GPS is subject to errors due to atmospheric delay, multipath and diffraction. Data cleaning aims to discard impossible positions or trajectories using specific constraints.

3.5.1. Ship trajectory extraction and separation

Vessel trajectory extraction can be divided into the separation of trajectories from different vessels (MMSI selection) and the removal of discontinuous vessel trajectories. The raw AIS samples are then sorted by timestamp in ascending order. It is found that an AIS sample may be recorded several times in the database. To solve this problem, the repetition samples are removed to avoid further processing if the constraints in equation (3) are satisfied. The outputs of the above step are the raw ship trajectories. In the detailed analysis of the raw data, we found that, based on the time stamps, there were periods of 2 to 4 hours of no data available for each vessel. This could be an indication of data loss. Such discontinuity of AIS data poses a great challenge for the detailed analysis of the kinematic motion state of ships. To fill the gaps in the data, we divided the raw ship trajectory into segments and used Equation 4 to fill the gaps as a function of a threshold:

$$\begin{aligned} T_a &= T_b, \\ Lat_a &= Lat_b, \\ Lon_a &= Lon_b, \end{aligned} \qquad (3)$$

$$T_i > T_{th}, \tag{4}$$

Where T_a and T_b are the time stamps of two AIS records. The vessel positions at timestamp T_a are denoted Lat_a and Lon_a respectively, Lat_b and Lon_b are the counterparts at timestamp T_b. T_i is the default threshold. Sometimes ships change their IMO and MMSI numbers. This may be because they change flag, ownership and type of AIS, radio equipment or supplier. To avoid multiple AIS tracks for a vessel, we map changes and combine automatic tracking with manual work. One way to check the accuracy of MMSI numbers is to compare them with the ITU database: Here we manually enter an MMSI number, call sign or vessel name and check the existence of the vessel in the ITU database. To increase the reliability of the two reference routes, the MMSI numbers were randomly checked. Figure 6 shows an example of MMSI cleaning. This figure shows a ferry sailing from Eemshaven in the Netherlands to Kristiansand in Norway. This ferry currently operates exclusively on this route. The duplicate MMSI of the ferry off the Spanish and French coasts are incorrect and have been removed.

Figure 6: Example of double MMSI Numbers (left) and after cleaning (right side)

3.5.2. Time interval standardization

According to ITU requirements, AIS is not transmitted continuously but at intervals. Therefore, AIS data received from different vessels is asynchronous and must be transmitted separately. When a vessel is underway, AIS data typically has a 30 second interval between positions from each Class B AIS and a 10 second interval between positions from each Class A AIS [18]. If the shore observer receives the latest positions of the own vessel at time t_0, t_1 and t_2, the received positions of the target vessel are at time t'_0, t'_1 and t'_2 before t_0, t_1 and t_2 respectively ($t'_0 < t_0 < t'_1 < t_1 < t'_2 < t2$). Figure 7 illustrates this relationship.

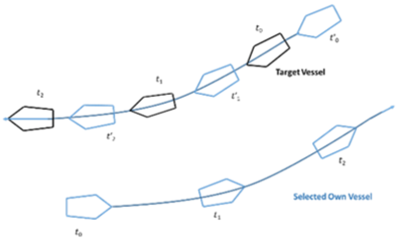


Figure 7: Asynchronous AIS positions

From the position of the two ships, the distance between them can be calculated directly for comparison. If the positions of two ships are (lat_1, lon_1) and (lat_2, lon_2), the distance between them is:

$$*l = \arccos[\sin(lat_1) \sin(lat_2) + \cos(lat_1) \cos(lat_2) \cos(lon_1 - lon_2)] \times R*$$

where l is the distance between the two positions and R is the radius of the Earth. Calculate the distance l' between two sets of predicted positions of two approaching vessels in order, using the

AIS message times of the closest match. To standardise the time intervals, we used the hill-climbing algorithm. The hill-climbing algorithm starts from the first two positions of the two vessels. If the change in distance dl' is less than zero, the shortest distance can be accepted:
$dl'<0$

If the change in distance dl' is greater than zero, it means that the values of the distance obtained are increasing and the vessels are moving away from each other. As shown in Figure 8, the closest predicted approach point is the current position and can be referred to as unsynchronised pseudo-CPA.

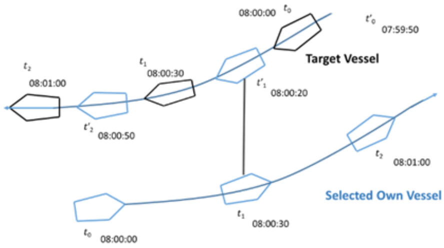


Figure 8: Time interval standardization

3.5.3. Data cleaning

Measurement and data collection errors are common in the real world. Noise is an unavoidable problem with AIS as it relies on many sensors in addition to manual input. A change in the original value can be defined as noise. Random position errors in ship tracks, AIS speed not matching ship speed and unstable speed and course values were also found in the AIS sensor data, as noted in [22]. Our observations are broadly consistent with [23]. The removal of noise is usually difficult, but we have addressed it using the Point-based Similarity Search Prediction (PSSP) model [24] for efficient results in the presence of noise. In Figure 9, lat$_{TP,t1}$ and lon$_{TP,t1}$ are the current known coordinates of the target point. Lat$_{TP,t2}$ and lon$_{TP,t2}$ are the future unknown coordinates of the target point to be predicted. In the same way, lat$_{SP,t1}$ and lon$_{SP,t1}$ and lat$_{SP,t2}$ and lon$_{SP,t2}$, which are both known, represent the coordinates of a similar point at the present time and at the future time respectively.

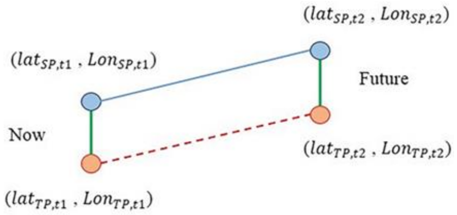


Figure 9: The PSSP model's approach to location prediction [25]

The PSSP model performs appropriately when the target and the most similar extracted points move in the same direction [25]. The prediction accuracy of the PSSP model drops dramatically when these objects move in different directions. Thus, in order to improve the accuracy of the PSSP model, a constraint is imposed which specifies the absolute difference between the azimuthal angle of the path of the target point and the azimuthal angle of the path of the similar point, which must be less than 22.5° [25]. To confirm the results obtained and to increase the reliability, the generated noise-free tracks of the reference routes were evaluated with the cubic spline interpolation. It was found that the time interval between different data samples can vary, which prevents the ship trajectory reconstruction model from accurately extracting the AIS intrinsic patterns. Thus, it is difficult to predict the ship trajectory in real applications. To overcome this problem, we use cubic spline interpolation and moving average models to normalise the AIS data series. This interpolation scheme is based on [26]. It requires a priori knowledge of the signal domain so that the interpolant is consistent not only at the sampling points, but also with respect to the given samples at intermediate points between samples. As an output of this trajectory reconstruction model, we obtain normalized ship trajectory samples, which we fed into an AIS data reconstruction model to obtain normalized AIS data samples as an output. For this we used a moving average model [27]. In AIS data, missing values are a recurring problem that affects every attribute. If the time interval between successive observations is very long, missing reports may result in discontinuous trajectories or large geographical coverage gaps.An example of an incomplete trajectory is shown in Figure 10, where interpolation methods can be used to complete missing trajectories [28].

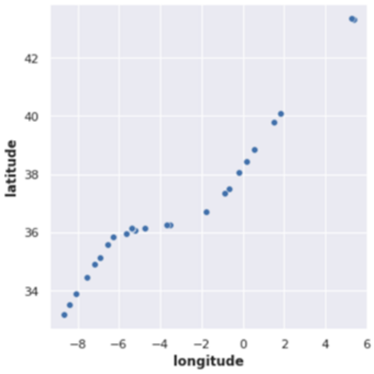


Figure 10: Example Missing Position Values [28]

However, there are a number of challenges in analysing marine trajectory data. 1) Ships move relatively freely in the maritime environment compared to the restricted movement of vehicles on roads. It is difficult to define the normal movement of ships, although there are recommended main shipping routes for ships to follow. 2) the refresh rate of AIS is every few seconds or minutes depending on the speed of movement. To evaluate the results of the PSSP model, we used the Lagrange interpolation [29]. This is a well-known and established algorithm that is preferred to the Newton interpolation because it gives better approximations. To calculate the number of interpolation points to be added to the trajectory, we take-into account the minimum transmission frequency of the AIS (T_f =3 minutes). Dividing the temporal distance with T_f results in the number of interpolated positions k. Dividing the spatial distance with k results in the spatial increment x_{incr}. This increment indicates the vessels movement in the x-axis at each time-frame of the missing trajectory. Knowing that the ship moves by x_{incr} every 3 minutes and that the ship has transmitted a total of k positions during the communication gap, we calculate the corresponding values (y, speed and heading) of the interpolated positions [30]. At each step of the iteration the new interpolating timestamp and longitude are calculated by adding the corresponding increment values, while the new latitude value is calculated using the Lagrange formula [30]. To calculate the speed, we first need to calculate the distance between the previous and the current position. We use the Haversine formula because it takes into-account the earth’s curvature [30]. Then, the new heading is calculated by using simple trigonometric functions. Finally, the newly created interpolating position is added to a list and becomes the previous position as the iteration proceeds. To constrain the interpolation from creating unrealistic positions, we apply the interpolation to each route $R_{A\rightarrow B}$, after positions inside waypoints have been removed. Waypoints contain a large amount of positions that are anchored or moored, which means that positions will be scattered in a small area inside the waypoint, rather than follow a specific kinematic path, introducing extreme curvatures and extreme speed values. Moreover, routes that have a gap equal to or more than 24 hours are removed from the process [30]. Figure 11 visualizes the result of the Lagrange interpolation applied to one distinct trajectory with large communication gaps. Figure 11(a) illustrate a trajectory with a communication gap. This Figure visualizes a vessel that is traveling from the Ionian Sea towards the Aegean Sea, when a large communication gap appears at the south side of the Peloponnese. Figure 11(b) illustrate the same trajectories after the Lagrange interpolation. In this case, our approach at interpolating missing positions manages to fill in the missing segments efficiently.

Figure 11: Example of Lagrange interpolation [30]

3.6. Data Normalization

Normalization refers to the adjustment of measured values to make them suitable for comparison. Out-of-range values can occur in many AIS data attributes and can be easily handled using filters. For example, latitude greater than 90° or longitude greater than 180° [31] can be easily detected. Another example is negative or very large speed values, e.g. a bulk carrier doesn't normally exceed 15 knots. Speed jumps can be detected and corrected by evaluating the distances between successive observations and setting a threshold [32]. Another example is the course and heading values of ships, which should be between 0° and 359°. A heading value of 511 means not available and a heading value of 360 means unknown. However, values between 360 and 511 may appear although they are not allowed. In AIS messages, ship specific information should remain consistent. For example, there should be no change in ship type, size or cargo type during the same voyage [33]. Also, the ship's beam and length should match the draught values. Another issue is the destination text format, which can have any structure. A vessel can use the name of the port, the code of the port, the country of the port, the port of origin and destination, an unknown abbreviation or leave it empty.

3.7. Data Filtering

The filtering of AIS data is primarily used to find and remove data problems that are very different from most of the AIS data (data outliers and numerical redundancy in the sources). We use this step for further evaluation and post-processing. To further evaluate the consistency of the individual AIS messages with each other, it is important to understand that there are 2 different AIS messages. 1) AIS position reports, which report the position, speed, heading, rate of turn and status of vessels, and 2) AIS static voyage reports, which report the identity, type, size and voyage information [18]. These two types of data are linked only by the vessel's MMSI number, which is used as the vessel identification for both. Message types 1, 2, 3, 18, 19, 27 are position messages and contain values in the position related fields. Message types 5 & 24 are static voyage messages and contain values in the static details and voyage related fields [18]. In order to identify the name, IMO number or type of vessels associated with the AIS position reports, the MMSI number must be used to link the two different sets of AIS data. For example, if position reports to MMSI 636018333 are recorded in AIS message type 1, then the IMO and name of that vessel can be determined by looking for AIS message type 5 reported using the same MMSI number as shown in Figure 12. With this data consistency, the AIS data can be used and visualized as a Reference-Route for a particular vessel between two ports.

Position					Both			Static				
speed	heading	rot	latitude	longitude	timestamp	msg_type	mmsi	imo	name	callsign	eta	destination
0	511	0	-86.67	21.8061	2020-04-23T15:38:10.37	1	636018333					
					2020-04-23T15:58:40.38	5	636018333	9821299	SOUTHERN SHA	D5PG4	2020-04-25T09:00:00	MX COA
13.5	288	-11	-86.702	21.8523	2020-04-23T16:03:11.23	1	636018333					
12.8	290	0	-86.786	21.919	2020-04-23T17:30:50.34	27	636018333					
14.4	296	-128	-86.81	21.922	2020-04-23T17:38:00.41	1	636018333					
14.3	295	-43	-86.966	21.9533	2020-04-23T17:48:51.11	1	636018333					
					2020-04-23T18:01:10.43	5	636018333	9821299	SOUTHERN SHA	D5PG4	2020-04-25T12:00:00	MX COA
14.4	290	8	-87	22.0199	2020-04-23T18:20:44.34	1	636018333					
14.1	300	0	-87.068	22.0678	2020-04-23T18:28:00.45	1	636018333					

Figure 12: Distribution of AIS message key field values

3.8. Trajectory Clustering for the Frequently Most Sailed Reference-Route

The last step of our approach is related to the clustering of the trajectories that have the same itinerary $R_{A \rightarrow B}$. Clustering analysis is a way to group data into similar attributes where similarity is based on a distance metric between points. In Figure 13 each dot represents a collection of AIS signals contained within the area occupied by the dot.

Figure 13: Clustering for the identification of boundaries

The typical algorithm for clustering the points of one or more trajectories is DB-Scan [34], which is used as a density-based spatial clustering method. Our version of DB-Scan uses 3 parameters to specify the proximity of candidate vessel AIS signals (positions) [35]. Therefore, for each position in a ship's trajectory, we keep: i) the ship's speed, ii) the ship's heading (course over the ground), and iii) its latitude and longitude coordinates [35]. Three different thresholds are then used to decide if the positions of the two different vessel trajectories are neighbours: the speed difference threshold s, the heading difference threshold h and the distance threshold eps. If two positions from the same or different ship trajectories have an absolute difference in all three dimensions that is below the threshold, the trajectory points are clustered together [35]. Figure 14 illustrates the differences between the two variants of the DB-Scan algorithm. 14(a) on the left shows the original DB-Scan algorithm, which uses only spatial distance as a neighbourhood criterion. 14(b) on the right shows the modified DB-Scan for moving objects, which considers two (multi-dimensional) points (including position, direction and velocity dimensions) to be close to each other ('neighbours' in DB-Scan notation) if the positions are spatially close to each other and there are small differences in direction and velocity [35]. Each colour in Figure 14(b) represents a different cluster. Red arrows belong to the same cluster because they are close together and have approximately the same speed and direction. Dashed thin blue arrows indicate a different cluster because they have a higher speed than the red arrows. Dotted thick blue arrows indicate outliers that are either far away or have a different speed or direction from all their neighbouring vectors. Specifically, the dotted thick blue arrows at the bottom right have a different direction from the rest, while the dotted thick blue arrow at the top is spatially away [35].

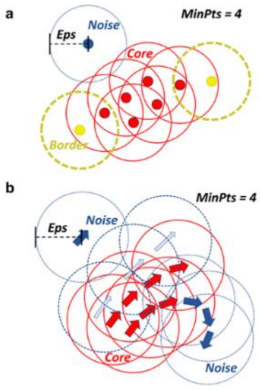


Figure 14: Trajectory Clustering for the frequently most sailed Reference-Route [35]

Before applying the modified DB-Scan algorithm, we first segment the monitoring area into 0.2 degrees tiles. The choice of this grid resolution was made because most weather services use the same resolution, making it easier to integrate with such services and allowing us to extend our approach in the future (e.g. enriching each tile with weather information to explain unexpected vessel movement behaviour). A database stores the trajectory samples of different vessels (i.e. AIS data) via the data sending/receiving timestamp, and thus we need to cluster trajectory data for the same vessel for the purpose of further trajectory analysis task. We remove the repeated trajectory sample considering that data duplication is common (i.e. vessel with different MMSI with the same latitude, longitude and time stamp). We divide the initial ship trajectory into two sub-trajectories if the neighbouring time difference meets certain criteria of Equation 2 in [36].Further we evaluate the success of trajectory compression in three ways. First, is to count the number of points. Second, using the Dynamic Time Warping (DTW) algorithm to find out changes in the original trajectory with the compressed one. The closer to 0 the value of the distance generated by the DTW shows that the compression results are similar to the original trajectory. Third, to determine the effect of compression with the time complexity of the similarity measurement process, the DTW processing time is calculated for each trajectory. Finally, after the trajectory clustering process completes, the clusters are converted into convex hulls [30]. The result is a set of convex hulls, each one annotated with the itinerary $R_{A\rightarrow B}$ it belongs to. Therefore, each itinerary can be represented as a set of convex hulls indicating the spatial boundaries the vessels must move in when following $R_{A\rightarrow B}$.

4. RESULT OF THE CONCEPT IMPLEMENTATION

Figure 15 illustrates the implemented framework for performing V&V. The basis of the scenarios for performing these activities are the 383 routes between the globally distributed ports of the statistical analysis. These routes form the basic ground truth, which is underpinned by the cleaned and processed AIS data. The resulting virtual routes, together with the virtual ports and vessels, are stored as scenarios in the scenario database. The generation of the two Reference-Routes in Figure 15 represents the first test run. Test run 2 is currently still in progress and is not explained in detail in this paper. It serves to validate the NAVTOR Auto-Route against the HAGGIS world model.

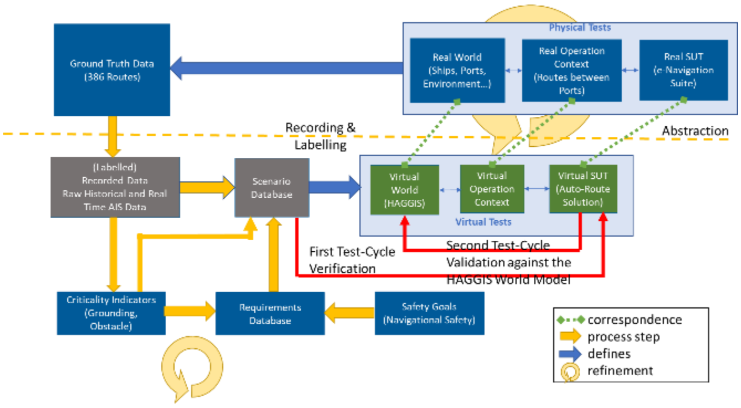


Figure 15: Result of the Concept Implementation

The following criteria have been defined as safety goals to assess the basic navigational safety of the generated NAVTOR Auto-Routes for vessels:

- The route must be accessible through the advisor service and follows the path between the two designated ports. The advisory service should be stable and reliable.
- The blocker used to set up navigational safety active.
- The autoroute must include the relevant prescribed TSS and these must be planned in the correct direction.
- The route must include the relevant prescribed deep- water routes, if required by the draught of the vessel, and these must be planned in the correct direction.
- The route of the vessel must follow the fairways in the correct direction. Buoyancy must be maintained in accordance with the regulations.
- The generated NAVTOR Auto-Route should not deviate by more than 10% in distance from the most frequently used reference route for a given vessel type (saving fuel and reducing CO2 emissions).
- Failure of the advisory service should be clearly indicated to the user.

Particularly critical indicators are the avoidance of ground contact and contact with objects such as buoys or moorings. Our framework allows two test cycles to be performed (Figure 17): In the first test cycle, the NAVTOR Auto-Route is manually compared with the Reference Route of a given vessel and the most frequently used Reference-Route between the two respective ports. This comparison requires the availability of up-to-date paper charts or ENCs. The main purpose of this test cycle is to verify the requirements. During the second test cycle, the HAGGIS (Hybrid Architecture for Granularly, Generic and Interoperable Simulations) maritime simulation of the German Aerospace Centre is used to validate the generated NAVTOR Auto-Routes. The HAGGIS virtual co-simulation infrastructure includes AI-based environmental, traffic and vessel simulators. These are used to assess the risks, efficiency and navigational safety of the NAVTOR Auto-Routes for ships during product development.

5. EVALUATION OF THE NAVTOR AUTO-ROUTES

When using the NAVTOR advisory service, a disclaimer is included stating that the user of the service is solely responsible for determining navigational safety and that NAVTOR accepts no responsibility. The user of the service is therefore responsible for determining the navigational safety of the route and must adapt the proposed Auto-Route to his vessel, for example by selecting a so-called blocker. These blockers prevent certain routes, such as those through the Kiel Canal, from being included in the Auto-Route calculation.

Our first task was to determine, if any depth information was included in the Auto-Route. This issue is explained below using the example of an Auto-Route from Abbot Point in Australia to Mundra Port in India. One ship that operates in the real world between these two ports is the MV Balzani. This vessel has a draught of 13.5m. We plan to sail the virtual route at a speed of 15 knots (kn). Our generated Reference-Route sailed by MV Balzani (Figure 16) between these two ports is approximately 6825 nautical miles (nm) in length. Due to its draught, the MV Balzani takes a (simulated) route around Papua New Guinea.

Figure 16: Reference-Route of the MV Balzani between the port of Abbot Point to Mundra Port

The inner routes around Australia are not navigable for vessels of this draught. The most frequently sailed Reference-Route between these two ports is approximately 5837 nm in length.(Figure 17). The green line in Figure 19 represents the frequently most sailed Reference-Route between the two ports, while the yellow and blue triangles represent vessels of different classes under way.

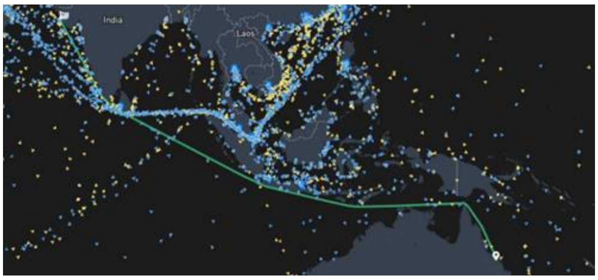


Figure 17: Most frequent sailed Reference-Route between the port of AUABP Abbot Point to INMUN Mundra Port

At a speed of 15 knots, the most frequently sailed Reference-Route takes approximately 16 days and 5 hours. The most frequently sailed Reference-Route is 988 nm shorter than the Reference-Route of MS Balzani. The most frequently sailed routes pass through the inner passages of Australia. Ships using these passages have a shallower draught than MS Balzani. If no blocker is used in the Auto-Route service, then the NAVTOR Auto-Route also suggests the inner route to Australia (Figure 18) and gives a calculated length of 5976 nm. At a speed of 15 knots, the duration is 16 d, 14 h and 24 min.

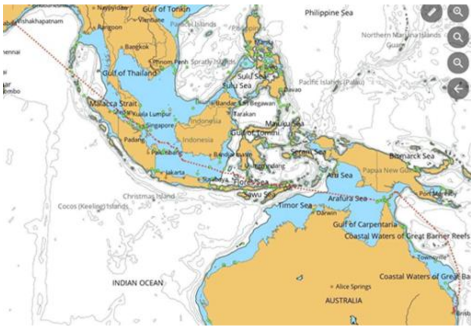


Figure 18: Auto-Route without blocker between the port of Abbot Point to Mundra Port

In this example, the inner route refers specifically to the use of the inner route "Northern Passage" with the use of the Gannet and Varzin Passage (Figure 19). The blue line shows our manual navigational/mathematical calculation of the course and position required to safely navigate this route. The red line illustrates the Southern Passage.

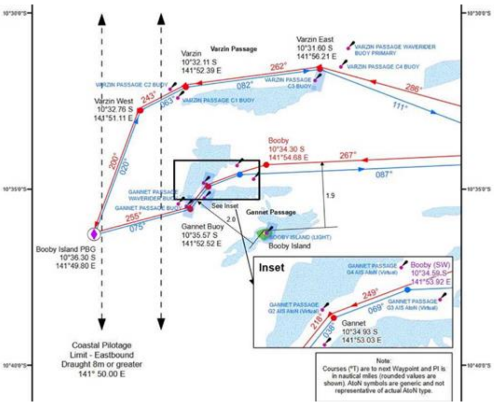


Figure 19: The Garnnet and Varzin Passage of the Inner Route North Passage

It is clear from this example that the draught of the vessel is not taken into-account by NAVTOR Auto-Route. What is the solution in the Auto-Route calculation? Vessels with larger draughts must use the blockers Gannet and Varzin Passage for a voyage between the port of Abbot Point and the port of Mundra. Repeating the test with the Gannet and Varzin Passage blockers produced the result shown in Figure 20. As can be seen from this figure, the generated auto-route no longer crosses the Great Barrier Reef - Inner Route North, Northbound - but bypasses Papua New Guinea.

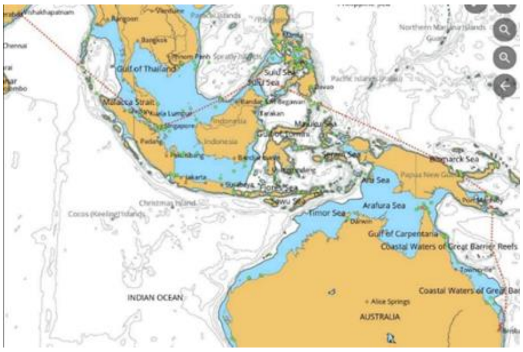


Figure 20: NAVTOR Auto-Route Abbot Point to Mundra Port with blocker Gannet and Varzin Passage

The result of the Auto-Route calculation is a route length of 6830 nm. The NAVTOR Auto-Route is 5 nm longer than the Reference-Route of the MV Balzani (6825 nm). However, the route is almost identical. The verification of navigational safety in accordance with the requirements of the IMO has shown that this route is basically navigationally safe. The Auto-Route follows the fairways and passes the buoys on the correct side and danger spots are effectively avoided. The adaptation effort for on-board navigation is > 70% compared to manually planning the route with a NavStation or other navigational equipment.

As a further evaluation in the form of simulation of so-called special routes (not included as a result of the static analysis), we compared the results of manually navigated/mathematically calculated ship routes between two ports with the Reference-Route of a given ship and with the NAVTOR Auto-Route. The following example illustrates this evaluation: The route from Busan to Taichung Harbour was manually calculated by navigation/mathematics on nautical paper charts and manuals. This route is basically safe in terms of navigation and has a length of 795 nm. The King Men Express with a draught of 11,9 m was used as a simulated ship with a cruising speed of 12 kn. The calculated NAVTOR Auto-Route between the port of Busan and the port of Taichung has a length of 799 nm. With a speed of 12 kn the duration is about 2days, 18 hours and 32 minutes. The Auto-Route is only 4 nm longer than the manual calculated route. Our Reference-Route of the King Men Express between these ports has a length of about 829 nm. The Figure 21shows a basic illustration of this.

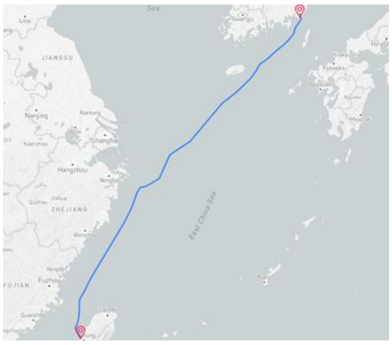


Figure 21: Reference-Route King Men Express between Port of Busan to Taichung Port

The length of NAVTOR Auto-Route is 30 nm shorter than the Reference-Route of the King Men Express. The most frequently sailed Reference-Route between these two ports has a length of about 791 nm. This route includes ships with shallower draughts. While these vessels can sail closer to the coast, shortening the route (Figure 22).

e

Figure 22: Most frequently sailed Reference-Route between Port of Busan to Taichung Port

The concrete comparison of the individual WPs of the 4 routes showed an almost identical route guidance with only gradual differences. This example illustrates the high quality of the NAVTOR Auto-Route and also why comparisons with the reference routes are useful because of their minimum equivalent quality. With the implemented V&V architecture, 93 of the 386 statistically determined routes have been checked so far. The summarized results are as follows:

- The calculation of the NAVTOR Auto-Route is stable.
- The blockers used to ensure navigational safety work perfectly.
- Most NAVTOR Auto-Routes are shorter than the Reference-Route of the specified ship. This reduces CO2 emissions and fuel consumption.
- Some of the calculated NAVTOR Auto-Routes are longer than the most frequently sailed Reference-Route. The main reason is that the ships on this route sometimes have a shallow draft and sail closer to shore to shorten the route. However, the distance differences at long distances (10,000 nm) are very small.
- The NAVTOR Auto-Routes are fundamentally safe from a navigational point of view. The route follows the fairways and passes the buoys on the correct side. Danger spots are effectively avoided.
- A comparison of the time taken to plan a conventional route (including a safety check) and using a NAVTOR Auto-Route with a safety check showed a saving in planning time of up to 70 per cent (manually stopped times within the relevant comparisons).

6. CONCLUSION AND FUTURE WORK

One of the main objectives of a shipping company to be competitive is to optimize the voyage. It is largely driven by the need to operate a ship as cost-effectively, energy-efficiently and safely as possible on every voyage of its lifetime. Voyage planning is now supported by weather routing systems on a large number of ships. NAVTOR is currently working on integrating a weather forecast into their Auto-Route. Depending on the requirements of the ship operator and the shipping business, the main objective may be to optimize the voyage in terms of energy efficiency, voyage duration, safety or a combination of these aspects. In the case of a flexible arrival time, a minimum time or minimum total cost optimization problem has to be solved. In addition to time constraints, other constraints include vessel characteristics, safety considerations, and geographical conditions, which are mainly route restrictions due to shore, shallow water, icebergs, mines, or traffic separation schemes. Ship characteristics are usually taken into-account by a hydrodynamic model that reflects the behaviour and reactions of the ship, its speed profile and fuel consumption in the face of wind, waves, currents and other environmental conditions. Our method of comparison in the form of two reference routes is applicable to ports of any size or, more generally, to Areas of Interest of any size. It must be admitted that the generation of the Reference Routes was not carried out in a coherent framework, but several independent methods and models were used to increase the reliability of the references. For the same reason, NAVTOR AIS data was not used. The variance of the worldwide route guidance achieved by the static analysis is extended by the retrieval of so-called special routes. These special routes present special challenges in terms of further comparison possibilities and navigational peculiarities. An essential key to the generation of reference routes is the preparation of AIS data. As this is mainly done manually using various methods, it is very time consuming. It often takes up to 10 hours to complete a reference route. Part of this effort is certainly due to the fact that no closed framework was used to generate the reference routes, but this should not hide the fact that the preparation of the AIS data is time consuming. Unclean, incomplete or even contradictory content in AIS data prevents the generation of high-quality reference routes. The generation of 2 Reference Routes for comparison with a NAVTOR Auto-Route has proved to be useful in our experience. The most frequently sailed Reference-Route should be almost shorter on average than the NAVTOR Auto-Route, as this route is also used by vessels with a shallower draught. The Reference-Route of the specified vessel will be longer on average than the calculated NAVTOR Auto-Route, as we have generated this route from the AIS data of a real vessel with its current draught. These vessels sail deep-water routes further away from the coasts, thus increasing the distance between the two ports. The NAVTOR Auto-Route takes into-account the gross tonnage of the vessel, but not the actual draught. The navigational safety of the Auto-Route must be guaranteed by navigation on board the real ship. The available blockers must be used for this purpose. The quality of the Auto-Route is determined by comparing the individual waypoints of the three routes. The summary result is that the quality can be set as high as none of the previously tested routes are nonsensical, as the Auto-Route follows a short route between the two ports. In 93 static analysis and special routes tested to date, 98% of the safety targets defined in Chapter 4 have been met. The potential for optimisation identified within individual routes is continually incorporated into the product, leading to continuous improvement.

Our next work is to continue the previous V&V activities until a NAVTOR Auto-Route version with weather forecast is implemented. With this in mind, we are working on a concept to extend / optimize the methods used so far to be able to generate Reference-Routes that also include a highquality weather forecast.

ACKNOWLEDGEMENTS

This work has been conducted within the TRANSACT project that has received funding from the ECSEL Joint Undertaking under Grant Agreement no. 101007260-2. This Joint Undertaking receives support from the European Union's HORIZON 2020 research and innovation programme and Austria, Denmark, Germany, Finland, Spain, Poland, Belgium, Netherlands, Norway.

REFERENCES

[1] PERFORMANCE STANDARDS FOR ELECTRONIC CHART DISPLAY AND INFORMATION SYSTEMS (ECDIS). [IMO Resolutions A.817 (19), MSC.64 (67) and MSC.86 (70)]. March 1999.
[2] ISO 26262-1:2018. Road vehicles — Functional safety — Part 1: Vocabulary. 2018.
[3] Brat, G and Jonsson, A. Challenges in verification and validation of autonomous systems for space exploration. In Proc. of the IEEE International Joint Conference on Neural Networks, volume 5, pages 2909–2914 vol. 5, 2005.
[4] X. Zou, X., Alexander, R. and McDermid, J. On the Validation of a UAV Collision Avoidance System Developed by Model-Based Optimization: Challenges and a Tentative Partial Solution. In Proc. of the 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshop, pages 192–199, 2016.
[5] Piazzoni, A.; Cherian, J.; Azhar, M.; Yap, J. Y.; Shung, J. L. W.; Vijay, R.: ViSTA: a Framework for Virtual Scenario-based Testing of Autonomous Vehicles. arXiv:2109.02529v2 [cs.AI]. 7 Sep 2021.
[6] SAE, "J3016 standard: Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," 2021.
[7] Chelbi, N. E., Gingras, D., and Sauvageau, C. Proposal of a new virtual evaluation approach of preventive safety applications and advanced driver assistance functions – application: AEB system. IET Intell. Transp. Syst. 12 (9), 1148–1156. doi:10.1049/iet-its.2018.5269. 2018.
[8] Kluck, F., Zimmermann, M., Wotawa, F., and Nica, M. "Genetic algorithm-based test parameter optimization for ADAS system testing," in 2019 IEEE 19th international conference 22.07.2019, 418–425. doi:10.1109/QRS.2019.00058. 2019.
[9] Guo, F.; Fuchs, A.; Kirschbichler, S.; Sinz, W.; Tomasch, E.; Steffan, H.; Moser, J.: Collection and classification of influence parameters for safety effectiveness of ADAS. Front. Future Transp., Sec. Transport Safety. Volume 4 – 2023. https://doi.org/10.3389/ffutr.2023.94559903. April 2023.
[10] ISO/PAS 21448:2019. Road vehicles — Safety of the intended functionality. 2022.
[11] Porres, I.; Azimi, S.; Lafond, S.; Lilius, J.; Salokannel, J.; Salokorpi, M.: On the Verification and Validation of AI Navigation Algorithms. arXiv:2101.06091v1 [cs.AI]. 15 Jan 2021.

[12] Girvan, C.: What is a virtual world? Definition and classification. Springer Link. Educational Technology Research and Development. 66, pages1087–1100. 2018.

[13] IALA Guideline-An Overview of AIS, Edition 2.0. Available online: https://www.navcen.uscg.gov/pdf/IALA_Guideline_1082 An_Overview_of_AIS.pdf (accessed on 15 March 2023).

[14] Lee, E.S.; Mokashi, A.J.; Moon, S.J.; Kim, G.S. The Maturity of Automatic Identification Systems (AIS) and Its Implications for Innovation. J. Mar. Sci. Eng. 7, 287. 2019.

[15] Perera, L.P.; Oliveira, P.; Soares, C.G. Maritime traffic monitoring based on vessel detection, tracking, state estimation, and trajectory prediction. IEEE Trans. Intell. Transp. Syst. 13, 1188–1200. 2012.

[16] He; W.; Li, Z.; Malekian, R.; Liu, X.; Duan, Z.: An Internet of Things Approach for Extracting Featured Data Using AIS Database: An Application Based on the Viewpoint of Connected Ships. July 30, 2017. https://zenodo.org/record/1042036

[17] Suo,Y.; Chen, W.; Claramunt, C.; Yang, S.: A Ship Trajectory Prediction Framework Based on a Recurrent Neural Network. Sensors (Basel). Sep; 20(18): 5133. 2020. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7570964/.

[18] [n. d.]. M.1371: Technical characteristics for an automatic identification system using time-division multiple access in the VHF maritime mobile band. ([n. d.]). http://www.itu.int/rec/R-REC-M.1371-5-201402-I/en, Last accessed: 2022-10-22.

[19] Blauwkamp, D.; Nguyen, T. D.; Xie, G. G.: Toward a Deep Learning Approach to Behavior-based AIS Traffic Anomaly Detection. 2019. https://faculty.nps.edu/xie/papers/ais_analysis_18.pdf.

[20] Kurt Schwehr. [n. d.]. libais: C++ decoder for Automatic Identification System for tracking ships and decoding maritime information. ([n. d.]). https://github.com/schwehr/libais, Last accessed: October, 2022.

[21] Shuang, S.; Yan, C.; Jinsong, Z.: Trajectory Outlier Detection Algorithm for ship AIS Data based on Dynamic Differential Threshold. Journal of Physics Conference Series 1437(1):012013. DOI:10.1088/1742-6596/1437/1/012013. January 2020.

[22] Zhao, L., Shi, G., and Yang, J., Ship trajectories pre-processing based on AIS data, The Journal of Navigation, vol. 71, no. 5, pp. 1210-1230, 2018.

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

[24] Wijaya, W. M. and Nakamura, Y.: Predicting Ship Behavior Navigating Through Heavily Trafficked Fairways by Analyzing AIS Data on Apache HBase. 2013 First International Symposium on Computing and Networking. IEEE, pp. 220–226. 2013.

[25] Alizadeh, D.; Alesheikh, A. A.; Sharif, M.: Vessel Trajectory Prediction Using Historical Automatic Identification System Data. Published online by Cambridge University Press: 26 August 2020. https://www.cambridge.org/core/journals/journal-of-navigation/article/vessel-trajectory-prediction-using-historical-automatic-identification-system-data/64D2DB412029364590EC071B92E71A6B.

[26] Behjat, H., Dogan, Z., Van De Ville, D. and Sornmo, L. “Domain-informed spline interpolation,” IEEE Transactions on Signal Processing, vol. 67, no. 15, pp. 3909–3921, 2019.

[27] Chen, X. “Anomaly detection and cleaning of highway elevation data from google earth using ensemble empirical mode decomposition,” Journal of Transportation Engineering, Part A: Systems, vol. 144, no. 5, Article ID 04018015, 2018.

[28] Mao S., Tu E., Zhang G., Rachmawati L., Rajabally E., and Huang G.B., An Automatic Identification System (AIS) Database for Maritime Trajectory Prediction and Data Mining, Proceedings of the ELM-2016, vol. 9, Springer, Cham, 2018.

[29] Berrut, J.P. and Trefethen, L.N., Barycentric Lagrange interpolation. Society for Industrial and Applied Mathematics, 46 (3), 501–517. 1984.

[30] Kontopoulos, I.; Varlamis, I.; Tserpes, T.: A distributed framework for extracting maritime traffic patterns. International Journal of Geographical Information Science. 2020. https://www.tandfonline.com/doi/pdf/10.1080/13658816.2020.1792914.

[31] Harati-Mokhtari, A., Wall, A., Brooks, P., and Wang, J., Automatic Identification System (AIS): Data Reliability and Human Error Implications, Journal of Navigation, vol. 60, no. 3, pp. 373-389, 2007.

[32] Mao S., Tu E., Zhang G., Rachmawati L., Rajabally E., and Huang G.B., An Automatic Identification System (AIS) Database for Maritime Trajectory Prediction and Data Mining, Proceedings of the ELM-2016, vol. 9, Springer, Cham, 2018.

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

[34] Ester, M. A density-based algorithm for discovering clusters a density- based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the second international conference on knowledge discovery and data mining, KDD'96. Portland: Oregon AAAI Press, 226–231. 1996.

[35] Kontopoulos, J.; Varlamis, I.; Tserpes, T.: Uncovering Hidden Concepts from AIS Data: A Network Abstraction of Maritime Traffic for Anomaly Detection. Springer Link. 04.January 2022. https://link.springer.com/chapter/10.1007/978-3-030-38081-6_2.

[36] Yang Sun, Y.; Xinqiang Chen, X.; Jun, L.; Jiansen Zhao, J.; Hu, Q.; Fang, X.; Ying Yan, Y.: SHIP TRAJECTORY CLEANSING AND PREDICTION WITH HISTORICAL AIS DATA USING AN ENSEMBLEANN FRAMEWORK. International Journal of Innovative Computing, Information and Control. Volume 17, Number 2, April 2021. http://www.ijicic.org/ijicic-170206.pdf.

AUTHORS

Arnold Akkermann, German Aerospace Center since 2013 Research Associate at the Oldenburg Institute of Computer Science and German Aerospace Center. Scientific focus: RE-Engineering, Quality Assurance, V&V and Maritime Systems.

Bjørn Åge Hjøllo, NAVTOR AS. Chief Sustainability Officer.