Automatic traffic scenarios generation for autonomous ships collision avoidance system testing

Victor Bolbot [a,b], Christos Gkerekos [a], Gerasimos Theotokatos [a,*], Evangelos Boulougouris [a]

[a] *Maritime Safety Research Centre, Department of Naval Architecture, Ocean and Marine Engineering, University of Strathclyde, 100 Montrose Street, G4 0LZ, Glasgow, United Kingdom*
[b] *Research group on Safe and Efficient Marine and Ship Systems, Department of Mechanical Engineering (Marine Technology), Aalto University, Espoo, Finland*

ABSTRACT

The Collision Avoidance (CA) system constitutes a key enabling technology for the Maritime Autonomous Surface Ships (MASS), the appropriate functionality of which is critical for assuring the navigation safety. Although several techniques including testing of the collision scenarios in a virtual environment can be employed, the trust of testing phase results depends on the number of tested scenarios and their coverage. This study aims at proposing a systematic and automatic process for the generation of the traffic scenarios that can be employed for the CA system testing. First, the range of the investigated parameters is defined, and samples of potential traffic parameters are generated using Sobol sequences. Subsequently, hazardous traffic scenarios are identified from the initially generated scenarios by using predefined rules. For these hazardous scenarios, a risk vector considering weather conditions and traffic conditions is calculated. A clustering algorithm is employed to identify the groups of traffic conditions that can be encountered based on each scenario risk vector and COLREGs traffic scenarios. For each of these groups, the riskiest scenario is provided as input for the test cases development, thus, simplifying the selection process of testing scenarios. The process is applied to a theoretical Short Sea Shipping autonomous vessel, whereas the derived results are employed to discuss the advantages and disadvantages of the developed process.

1. Introduction

To address the marine industry imense challenges, novel systems and technologies have been introduced, including Maritime Autonomous Surface Ships (MASS). Major industrial initiatives, such as Yara Birkeland (Yara, 2018) and ASKO (Smartmaritime, 2020), have already launched commercial MASS along with projects such as, MUNIN (MUNIN, 2016), AAWA (AAWA, 2016), SISU, SVAN (Daffey, 2018), AUTOSHIP (Bolbot et al., 2020).

The MASS autonomy degree can be classified into four types according to International Maritime Organisation (IMO), ranging from ships with automated processes to fully autonomous ships (IMO, 2020). The research and development initiatives pertinent to MASS extend to multiple directions. Several research studies focused on the development of intelligent machinery systems in the context of MASS, such as in (Abaei et al., 2021; BahooToroody et al., 2022; Bolbot et al., 2021b; Tsoumpris and Theotokatos, 2022). The required regulatory aspects for

MASS were investigated in (Bačkalov, 2020; Erceg, 2018; Nzengu et al., 2021; Ringbom, 2019). The aspects related to MASS remote control were studied in (Basnet et al., 2022), whilst m the Collision Avoidance systems were investigated in (Huang et al., 2020; Utne et al., 2020).

The Collision Avoidance (CA) system is considered a critical system for MASS, as it is expected to make decisions, affecting the safety of the own and surrounding ships (Bolbot et al., 2021a). Several approaches have been demonstrated for the design of CA systems with a variety of algorithms and models being considered, such as holonomic (Degre and Lefevre, 1981), kinematic (Vincent, 1977), dynamic models (Fossen, 2002) for ship motion modelling in CA manoeuvres, physics-based (Fossen, 2018), manoeuvre-based (Peel and Good, 2011), interaction-based (Hu et al., 2008) for ship trajectory prediction, CA based on collision risk estimation (Kearon, 1997), and artificial intelligence algorithms (Yu-Hong and Chao-Jian, 2005). A considerable number of the previous studies focused on the development of CA systems for conventional ships and MASS, as for example in (Brcko et al.,

---

2021; Huang and van Gelder, 2020b; Mizythras et al., 2021; Namgung and Kim, 2021), or on the estimation of collision risk metrics (Goerlandt and Kujala, 2014; Huang and van Gelder, 2020a; Liu et al., 2019; Silveira et al., 2021; Tam and Bucknall, 2010; Zhang et al., 2017).

The existence of a plethora of advanced design methods and CA system solutions is an indicator of complexity and challenges associated with the safety assurance for CA systems. This statement is supported by the real-world accidents, which have occurred in other industries involving similar CA functions, due to defects in such systems (Guiochet et al., 2017; National Transportation Safety Board (NTSB), 2017). A prerequisite for avoiding malfunctions is to ensure adequate coverage of the potential scenarios that may be encountered by the ship during testing. This requirement stems from the problem of environmental complexity (Alexander et al., 2015), which indicates that MASS should interact properly with all the anticipated objects, under various conditions and traffic scenarios (Bolbot et al., 2019; Sørensen and Ludvigsen, 2017).

The exhaustive testing of all the potential traffic conditions to determine proper functionality of the CA seems to be a challenging endeavour, which is almost impossible to be achieved under the present state of technology (Torben et al., 2022). The number of scenarios increases exponentially with the number of parameters, therefore, there exists a critical threshold beyond which, exhaustive testing cannot be performed (Torben et al., 2022). The testing procedure is also restricted by the fact that the CA designers would prefer to conceal the details of their CA system, to avoid leakages of proprietary information (Pedersen et al., 2020).

This dual problem can be addressed by reducing the required number of tested scenarios, use of gradual approaches for the verification of the CA system, as well as by developing black box testing and verification techniques that would provide sufficient confidence to the CA system without the disclosure of proprietary information. In other words, only once an adequate number of scenarios are tested in a virtual or real environment, it can be assured that the CA system will not jeopardise safety when operating. Moreover, black-box functional or performance testing is desirable, as these techniques do not require the release of sensitive intellectual property information (Nidhra and Dondeti, 2012; Pedersen et al., 2020).

The navigation of ships is primarily regulated by the Convention on the International Regulations for Preventing Collisions at Sea (COL-REGs) (COLREGS, 1972). However, as COLREGs requirements were developed considering crew on-board (and not autonomous systems), they do not provide numerical criteria for crew actions, and their implementation relies on the crew judgement. Therefore, they cannot be used to develop a comprehensive set of testing scenarios (Porathe, 2019; Woerner et al., 2019), they do not include information on the frequency of traffic scenarios, or provide a comprehensive list of objects with which the MASSs will interact. These COLREGs limitations in relation to MASSs have been discussed at International Maritime Organisation (IMO) level (IMO, 2021), whilst some of the national authorities have already issued recommendations for application of specifically adapted for a MASS version of COLREGs (Russian Federation, 2021).

Automatic Identification System (AIS) data received from ship traffic systems can be a valuable data source for identifying such testing scenarios. It has been widely used for the analysis of traffic conditions, grouping of scenarios, identification of the most probable collision scenarios and development of safety domain as reported in (Gao and Shi, 2020; Goerlandt et al., 2017; Han et al., 2021; Kulkarni et al., 2020; Mieczyńska and Czarnowski, 2021; Mou et al., 2010). However, the quality of AIS data exhibit some limitations, as the ships with switched off their transponder and small recreational ships that are not required to have AIS, are not visible (IMO, 2015). Moreover, objects other than ships and buoys are not included in these data sets. Testing scenarios generated based on AIS data will inherit these limitations, resulting in incomplete number of scenarios being finally considered.

Several studies focused on testing cases generation, such as in

automotive (Fremont et al., 2020; Khastgir et al., 2021; Li et al., 2016; Riedmaier et al., 2021), aviation (Lindvall et al., 2017), and other industries (Arora and Bhatia, 2018; Clark et al., 2021). Relatively few studies focused on both the testing process, testing cases generation, or formal verification in the maritime industry, as reported in (Fedorowski et al., 1979; Foster et al., 2020; Minne, 2017; Park and Kim, 2020; Pedersen et al., 2020; Rokseth et al., 2019; Shokri-Manninen et al., 2020; Stankiewicz and Mullins, 2019; Torben et al., 2022; Woerner, 2014; Yang et al., 2007). To the best of authors' knowledge, only (Torben et al., 2022) proposed a way to automatically generate testing scenarios. However, this approach requires the use of simulators and application of search algorithms resulting in a significant number of simulations for testing of the CA systems. This is an important limitation that need to be addressed.

This study aims to develop a process that specifies traffic scenarios for the ship CA system testing in automatic fashion, whilst reducing the required number of scenarios for testing. The novel contribution of this research stems from the integration of methods used in statistics, ship hydrodynamics and big data analytics for solving a test scenarios generation problem, and includes: (a) the development of a novel automatic process for traffic scenarios generation independent from CA system with focus on scenarios that need to be tested; (b) the recommendation of novel risk metrics for use in a CA system testing; (c) the process application to the Short Sea Shipping autonomous vessel considering operations close to shore.

The remaining of the article is organised as follows. Section 2 presents the developed process methodology. Section 3 provides the investigated case study parameters. Section 4 reports and discusses the derived results. Finally, Section 5 summarises the main findings from this study and provides recommendations for future studies.

2. Methodology

2.1. Rationale, assumptions, and steps

Considering that the potential traffic scenarios space is infinitely large, it is necessary to identify representative scenarios as well as to demonstrate how these representative traffic scenarios depict the traffic scenarios that were not accounted for. The intention of this study is also to recommend testing scenarios without considering the CA system properties under testing and the CA system evaluation. Another objective is to identify testing scenarios, which would require as limited as possible customisation of the CA system.

To achieve the set aim and objectives, this study follows the steps depicted in Fig. 1. During step 1, uniform sampling of the potential traffic scenarios over the potential traffic space parameters values is conducted. These traffic encounter scenarios can be considered as representative for all the potential traffic encounters based on the efficiency of sampling technique.

Subsequently, an automatic risk assessment is conducted in steps 2 and 3, with the support of several selected risk metrics and rule-based criteria. The risk assessment is employed to identify those traffic scenarios, which are the riskiest in terms of navigational safety, and to filter out the safe scenarios. Apparently, there is no value in testing safe traffic scenarios, which does not require actions from the CA system. It should be noted though that "safe" traffic scenarios can be selected for testing if the CA system overreacts to safe traffic scenarios, however, this is outside the scope of the present research.

In step 4, the various collision risk metrics organised in a risk vector are analysed using a clustering technique to identify similarities between the selected traffic scenarios of step 3 in terms of risks. Some additional criteria are also employed to distinguish the different traffic scenarios, such as overtaking/overtaken, crossing, and head on. Based on these similarities, different groups of traffic scenarios can be identified, which are similar in terms of safety and traffic conditions.

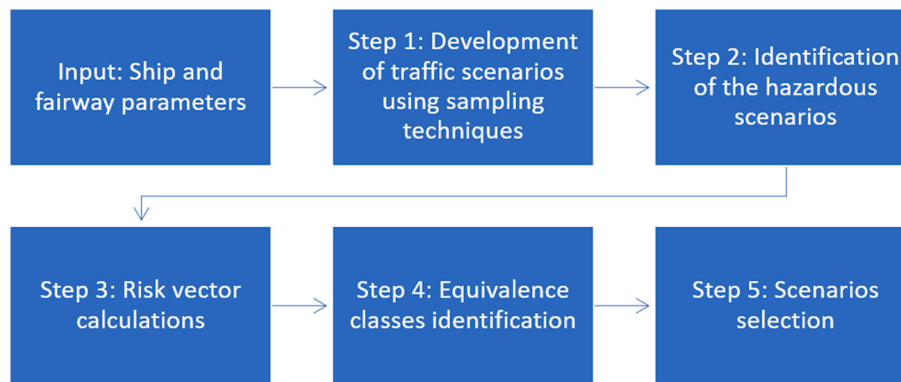During the last step, a representative scenario is selected for each

Fig. 1. Process overview.

identified traffic scenarios group. The main criteria for selecting a representative scenario include the values in the risk vector. It is assumed that the riskiest scenario in a group/cluster is adequate to depict the CA system behaviour in other traffic scenarios belonging to the same group. Therefore, if the CA system passes the test criteria for the riskiest traffic scenario, it will also pass the criteria for the less risky traffic scenarios. In this way, reduction of the considered test scenarios is ensured, without omitting a specific category of traffic encounter scenarios.

The effectiveness of the derived results depends on the assumptions reported in the preceding paragraphs, which are typically employed in pertinent research studies. For clarity, these assumptions are summarised below to ensure better understanding of the limitations of this study and the derived results.

- The sampled traffic scenarios offer a comprehensive coverage of the potential traffic scenarios.
- The selected rules for filtering the traffic scenarios are adequate and rational.
- The selected risk metrics accurately represent the safety of the encounter scenarios.
- Clustering and grouping of traffic scenarios are accurate.
- Holonomic models, which function under the assumption that the ships retain their speed and direction (Huang et al., 2020), are considered for the identification of the hazardous scenarios.
- Use of a circular domain is considered as a reasonable approximation of the ship domain.
- The riskiest scenario in each group is considered representative for each group.
- The elements of the risk vector as elaborated in section 2.5 are considered as equivalently important to each other.
- A ship cannot instantly change its speed to the desired level, and an arbitrarily threshold is considered as elaborated in section 2.5.

2.2. Identification of parameters influencing safety of navigation

The collision between the ships is usually an emergent event dependent on a large number of risk influencing factors that are related to route, human factors, ship design, equipment characteristics and organisational factors (Chen et al., 2019; Pedersen, 2010; Puisa et al., 2018). As the CA system reacts to the navigational factors of the route, only these factors are considered in this study.

The lists of navigational factors and pertinent metrics are provided in Table 1. These factors were derived based on the information reported in (Abebe et al., 2021; Bakdi et al., 2020; Bye and Aalberg, 2018; COL-REGS, 1972; Goerlandt et al., 2015; Li et al., 2021; Mizythras et al., 2021; Silveira et al., 2021; Woerner et al., 2019). The identified metrics that can be used to represent these navigational factors are derived based on the pertinent literature (Chen et al., 2019). These metrics and

Table 1
Navigational factors.

	Navigational factors	Pertinent metrics
1	How close the ships are/expected to be to each other and to the shore	Distance between ships/to shore
		Distance to Closest Point of Approach (Hilgert and Baldauf, 1997)
		Degree of Domain Violation (Szlapczynski, 2006)
		Vessel collision risk operator (Zhang et al., 2017)
		Minimum distance to collision (Montewka et al., 2012)
2	Time allowed to implement the CA manoeuvre	Time to Closest Point of Approach (Hilgert and Baldauf, 1997)
		Time to Domain Violation (Lenart, 2015; Szlapczynski and Szlapczynska, 2016)
3	Space required/available for ship manoeuvres	Ship domain (Fujii and Tanaka, 1971; Szlapczynski and Szlapczynska, 2017)
		Velocity space (Degre and Lefevre, 1981; Fiorini and Shiller, 1998)
		Collision Threat Parameter Area (Lenart, 1983, 2015)
		Velocity obstacle derivatives (Huang et al., 2018, 2019; Huang and Gelder, 2017)
4	Environmental conditions	Beaufort and direction of winds, Current speed and direction Waves significant height and direction
5	Visibility	Distance (The weather window, 2022)
6	Draught against seabed depth	Ship draft seabed depth
7	Type of COLREGs scenario	Ships' locations, speeds and headings
8	Other objects/agents in the area	Location, size, speed, type of movement, number of crew, passengers

their ranges can be determined with the assistance of AIS data, meteorological data, geospatial services, operator prior knowledge, according to the methodology described in (Bolbot et al., 2022).

The first three metrics of Table 1 can be considered as outcome of several basic parameters, such as, the ships location, heading, and speed. Hence, the latter parameters (ships location, heading, and speed) are selected for the sampling, instead of the metrics of Table 1. For demonstration purposes, we also selected the environmental conditions in our sampling. For the sake of this study brevity, we omitted the other remaining parameters/metrics, which can be considered in follow up studies.

2.3. Step 1: Development of traffic scenarios using sampling techniques

Once the sea traffic parameters are specified, the traffic scenarios can be generated using any sampling technique, e.g., Sobol sequences (Sobol', 1967), random Monte Carlo sampling (Metropolis et al., 1953), Latin hypercube sampling (McKay et al., 1979). In this study, Sobol

sequences were selected, as they demonstrate a number of useful properties. First, Sobol sequences provide an uniform coverage of the parameters space (Sobol', 1967). Second, they provide greater repeatability in results compared to the random and Latin hypercube sampling techniques without sacrificing accuracy, as demonstrated by several previous studies (Bolbot and Theotokatos, 2021; Burhenne et al., 2011; Kucherenko et al., 2015; Qian and Mahdi, 2020). Third, they do allow for a better repeatability of the results due to their quasi-random nature (Sobol', 1967).

The sampling is implemented over a continuous and not discrete space. Therefore, for the estimation of the sampling effectiveness, it is required to discretise the space of the potential values. This is implemented with the assistance of parameter dif_{min}, depicting the average normalised minimum difference in the parameter p_l for the range $(p_{l,max} - p_{l,min})$ with sampling over a number of p parameters. The physical meaning of dif_{min} is also illustrated in Fig. 2. There exist several other measures for assessing the sampling uniformity, such as, discrepancy-based, point-to-point based and volumetric (Ong et al., 2012; Zaremba, 1968); however, parameter dif_{min} was selected, as its physical meaning is more readily perceived.

The number of potential combinations T that can be sampled for p parameters, by considering the parameter dif_{min} is estimated according to the following equation:

$$T = \left(\frac{1}{dif_{min}} + 1 \right)^p \tag{1}$$

The sampling ratio (R) (Analytic Technologies, 2010) is defined according to the following equation, as the ratio between the N sample points and the total number of points that can be generated:

$$R = \frac{N}{T} = \frac{N}{\left(\frac{1}{dif_{min}} + 1 \right)^p} \tag{2}$$

For an existing sample, the estimation of dif_{min} can be approximated using the following formula:

$$dif_{min} = \frac{1}{N} \sum_{i=1}^{N} \frac{\sum_j^{p+1} x_{p,i,n} - x_{p,j,n}}{\sqrt{p}\,(p+1)} \; for \; i \neq j \tag{3}$$

where p is the total number of parameters, $x_{p,i,n}$ is the i normalised sample over p parameters and j is used to denote the p points having the least distance from the ith point.

For the Sobol sequences, due their pseudorandom nature, eq. (3) can be simplified as follows (which was determined after a series of

Fig. 2. Sampling space for two parameters (p_1, p_2) and physical meaning of parameter dif_{min} with uniform sampling.

numerical tests):

$$dif_{min} = \frac{\sum_j^{p+1} x_{p,i,n} - x_{p,j,n}}{\sqrt{p}\,(p+1)} \; for \; i \neq j, \; \forall \; i \neq 1 \tag{4}$$

2.4. Step 2: Identification of the hazardous scenarios

The identification of hazardous scenarios is implemented by analysing the traffic scenarios generated using the sampling techniques from the previous step.

For this purpose, a number of metrics from Table 1 is employed, such as, the geometric distance ($D_{i,k}$) between the Own Ship (OS) and kth Target Ship (TS$_k$), the time to the closest point of approach between the OS and TS$_k$ ($TCPA_{i,k}$), the distance at the closest point of approach between the OS and TS$_k$ ($DCPA_{i,k}$), and the safety domain around the OS that is represented by a circle with radius $a_{i,1}$. These are the metrics that were employed in a considerable number of previous studies (Szlapczynski and Szlapczynska, 2017).

However, some more advanced and effective metrics, such as, the Degree of Domain Violation (Szlapczynski, 2006) and Time to Domain Violation (Lenart, 2015; Szlapczynski and Szlapczynska, 2016), which were included in Table 1 could be also employed in a follow up study. In this study, due to the use of circular domain, these metrics (Degree of Domain Violation and Time to Domain Violation) practically give the same information with $TCPA_{i,k}$ and $DCPA_{i,k}$, as indicated by (Szlapczynski, 2006). This is not valid for cases where different domains were employed, such as, elliptical domains (Szlapczynski, 2006). This is one of the limitations of the study.

It should be noted that the use of circular domain constitutes an oversimplification and a very conservative approach to the safety domain definition. More advanced approaches define the safety domain as an ellipse (Namgung and Kim, 2021) (Coldwell, 1983; Davis et al., 1982; Fujii and Tanaka, 1971; Hansen et al., 2013; Liu et al., 2016; Pietrzykowski and Wielgosz, 2021; Szlapczynski and Szlapczynska, 2021), as block areas (Kijima and Furukawa, 2003), as quaternion (Wang, 2010), or as polygons (Bakdi et al., 2020; Hansen et al., 2013; Pietrzykowski and Uriasz, 2009; Szlapczynski and Szlapczynska, 2015). This simplification is adopted to facilitate the implementation and investigation of the overall automatic scenarios generation process. The consideration of alternative representations for the safety domain and the selection of the most appropriate is proposed as a subject for future research. A comprehensive review of safety domains and their characteristics can be found in (Du et al., 2021a; Szlapczynski and Szlapczynska, 2017).

All these metrics are valid under the assumption that the ship movements are holonomic. This assumption is not valid in real conditions, as the ships will eventually interact and communicate with each other and change their route based on the manoeuvres of the other ships (Huang et al., 2020). However, this assumption can be considered as plausible and sound approximation of traffic scenarios and traffic risks, as (a) the ships exhibit rather slow dynamics, and (b) the distance between the ships considered during the sampling is small relative to their dynamics. The calculation of more detailed safety metrics, considering ships interactions and simulations, would be rather too computationally expensive (potentially causing memory overflow issues), therefore, it would be challenging to implement the proposed analysis for a large sampling number.

The hazards identification is based on a set of criteria/rules, as presented in the pseudocode form provided in Table 2. First, for each traffic scenario generated using samples, the previously mentioned metrics are estimated. Then, for the ship that is the closest to the OS, the violation of all safety criteria/rules is examined. The following criteria/rules are considered:
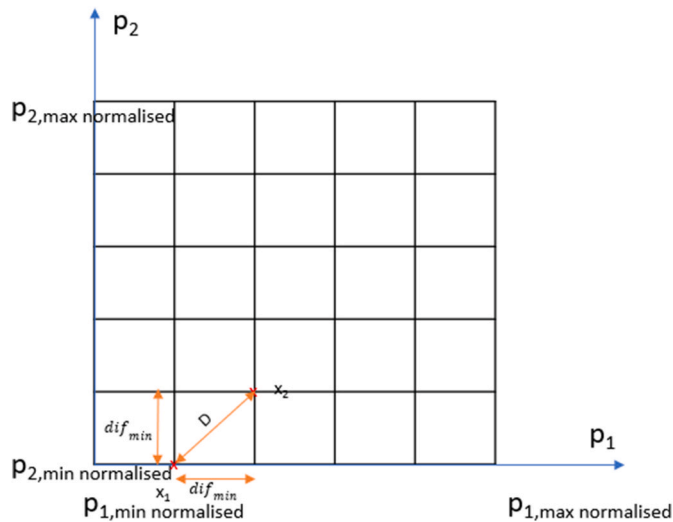
Table 2

Pseudocode for hazardous scenarios identification.

	Algorithm: Hazardous scenarios identification
1:	**Procedure:** Hazardous scenarios identification
2:	**Input:** Potential traffic scenarios generated using Sobol sequences in previous step
3:	For i = 1:N% for all the sample points
4:	Estimate
	$D_{i,k}$ %distance between OS and TS_k
	$TCPA_{i,k}$ %time of closest approach
	$DCPA_{i,k}$ %distance of closest approach
	$a_{i,1}$ %estimation of safety domain radius for OS
5	Find min $(D_{i,k})$, i = constant
6:	For $(D_{i,k})$ = min $(D_{i,k})$: max $(D_{i,k})$, i = constant
7:	If $TCPA_{i,k} > 0$ & $DCPA_{i,k} < a_{i,1}$ & $D_{i,k} < d_s$ & $TCPA_{i,k} < t_s$ & $D_{i,k} > a_{i,1}$ **then**
	Scenario should be considered as hazardous
	Store the scenario parameters
8:	End for $D_{i,k}$
9:	End for i
10:	**End** procedure

- Firstly, $TCPA_{i,k} > 0$, as we are not interested in scenarios where the closest encounter occurred in the past, and the ships are expected to diverge from each other, according to the holonomic assumption.
- Secondly, $DCPA_{i,k} < a_{i,1}$, as we are interested in scenarios where the two ships are anticipated to be close enough to each other in the future (TS_k in the safety domain of OS_k).
- Thirdly $D_{i,k} < d_s$, where d_s is a set distance threshold, as we are interested in ships which are in proximity to each other.
- Fourthly, $TCPA_{i,k} < t_s$, as the focus is on potential collision scenarios in the near future, depicted using t_s as threshold.
- Lastly, $D_{i,k} > a_{i,1}$, as we are not interested in the scenarios, where TS_k is already in the safety domain of OS_k and the collision is considered as unavoidable.

The use of these rules is quite similar to the ones employed in (Goerlandt et al., 2015; Szlapczynski and Szlapczynska, 2017) for the risk assessment purposes; it is much simpler though herein. This can be attributed to the fact that the rules are used to filter out the scenarios on a high level, whilst a more detailed analysis and grouping is implemented in step 4.

The values of d_s and t_s are set to 1 nm and to 10 min in line with the information provided in (Namgung and Kim, 2021) and (Rødseth et al., 2020). The d_s can be set to 3 nm as proposed in (Namgung and Kim, 2021), 2 nm as reported in (Federation, 2021), or 12 nm, which is the normal range of radar operation (Du et al., 2021b) and is also influenced by the operational area size and characteristics, such as the presence of shore. Other studies proposed the use of 15 min (Federation, 2021), 20 min, or 30 min for the t_s, and 3 nm whereas 6 nm for d_s depending on the traffic scenarios (Wang et al., 2021). A constraint that needs to be considered is the satisfaction of the holonomic assumption, therefore the values of d_s and t_s should not be too large.

When the shore is considered, the same set of criteria can be employed. The difference is that the shore is considered as an aggregation of different points with zero speed and the compliance against the set criteria is verified by considering if any of the shore points satisfies these criteria. The shore characteristics are represented using the shore point with the least DCPA in the next step. A graphical explanation is also provided in Fig. 3.

The equations for estimating the employed safety metrics are provided in eqs. (5)–(9), where (x, y) denote the location coordinates for each ship, u is speed in m/s, φ is the ship direction in degrees, 0 and k denotes the OS and TS_k, respectively based on (Namgung and Kim, 2021; Pedersen et al., 2020; Woerner, 2014).
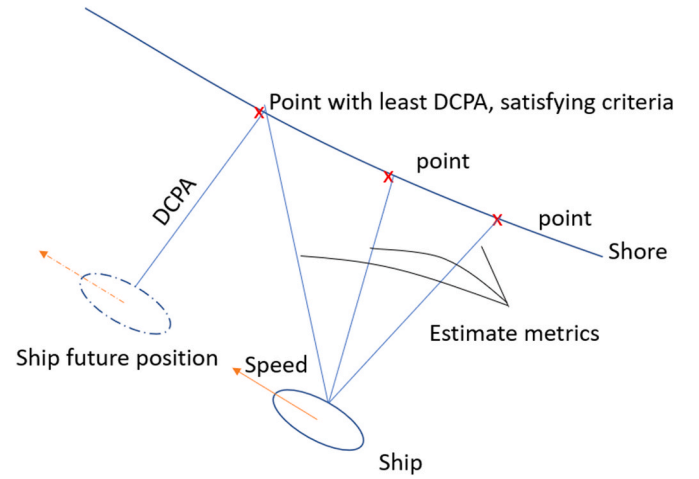

Fig. 3. Explanation for shore treatment.

$$D_k = \sqrt{(x_0 - x_k)^2 + (y_0 - y_k)^2} \tag{5}$$

$$k_{2,k} = u_0^2 - 2u_0 u_k \cos(\varphi_0 - \varphi_k) + u_k^2 \tag{6}$$

$$k_{1,k} = 2(u_0 \cos(\varphi_0) - u_0 y_k \cos(\varphi_0) - u_k y_0 \cos(\varphi_k) + u_k y_k \cos(\varphi_k) + u_0 x_0 \sin(\varphi_0) \\ - u_0 x_k \sin(\varphi_0) - u_0 x_k \sin(\varphi_0) - u_k x_0 \sin(\varphi_k) + u_k x_k \sin(\varphi_k)) \tag{7}$$

$$TCPA_k = -\frac{k_{1,k}}{2 k_{2,k}} \tag{8}$$

$$DCPA_k = \sqrt{k_{2,k} TCPA_k^2 + k_{2,k} TCPA_k + D_k^2} \tag{9}$$

The radius of the safety domain is calculated by the following equation (Namgung and Kim, 2021):

$$a_{i,1} = \begin{cases} \left(4 - 0.3\left(10 - \left|u_0/0.5144\right|\right)\right)L, & u \le 10\ kn \\ \left(4 + 0.3\left(\left|u_0/0.5144\right| - 10\right)\right)L, & u > 10\ kn \end{cases} \tag{10}$$

where L is ship length in (m).

The advantage of eq. (10) is that it renders the safety domain dependent not only on generic manoeuvrability characteristics expressed by the ship length, but also on the ship speed.

2.5. Step 3: Risk vector calculations

For the hazardous scenarios identified in the previous step, a risk vector with relevant risk metrics is calculated. The purpose of the risk vector is to characterise the risk type of the investigated hazardous scenarios, to support the scenarios clustering in the next step, and to rank the different hazardous scenarios. The navigational risk metrics estimated in this section include the normalised versions of the metrics presented in Table 1 to allow for the comparison of the selected scenarios on the same scale. The normalisation is implemented to have values in the range from 0 (if the risk metric does not reveal any hazardous condition) to 1 (indicating that collision is imminent). Exponential relationships between the metrics of Table 1 and the used risk metrics was preferred. This is implemented to assign higher values to those scenarios, which are closer to collision.

Based on the above considerations, the risk vector (*RV*) for each traffic scenario includes the components according to eq. (11).

$$RV = \left[\sqrt{DCPAn_k TCPAn_k},\ A_n,\ h_n,\ W_n \right] \qquad (11)$$

By setting the elements in a vector, an implicit assumption is introduced, which is that the elements of vector are equally important with each other.

The $DCPAn_k$ and $TCPAn_k$ represent the normalised version of DCPA and TCPA for TS_k ship (or shore) based on the selected safety domains and are estimated according to eqs. (12) and (13).

$$DCPAn_k = \begin{cases} 0, & \text{if } TCPA_k < 0 \text{ or } TCPA_k > t_s \\ 1, & \text{if } DCPA_k < a_1\ (TS_k \text{ inside safety domain}) \\ \dfrac{e^{\frac{d_r - DCPA_k}{d_r - a_1}} - 1}{e - 1}, & \text{otherwise} \end{cases} \qquad (12)$$

$$TCPAn_k = \begin{cases} 0, & \text{if } TCPA_k < 0 \text{ or } TCPA_k > t_s \\ \dfrac{e^{\frac{t_s - TCPA_k}{t_s}} - 1}{e - 1}, & \text{otherwise} \end{cases} \qquad (13)$$

The physical meaning of eq. 12 and 13 is as follows. If the closest point of approach with the ship TS_k is in the past or too far away in the future, ship TS_k does not contribute to the risk (relevant value 0). If the $TCPA_k \rightarrow 0$ from positive values of TCPA, as we already excluded TCPA<0, then the TCPA contributes the most to the risk (relevant value 1). For the intermediate values an exponential relationship is used to put greater emphasis to the conditions where the TS_k is closest to the OS. If TS_k lies within the safety domain, then the risk from $DCPAn_k$ becomes maximum (relevant value 1). The larger the $DCPAn_k$ value, the smaller is its contribution to the risk of collision. The parameter d_r is a parameter depicting the distance for which the risk metric is estimated, which is set to 3 nm in line with (Wang et al., 2021). The multiplication between $DCPAn_k$ and $TCPAn_k$ is implemented to emphasise that the closer an encounter is in distance and in time, the higher is the risk of collision with the ship TS_k.

It should be noted that the above risk metrics are relevant only when considering ships independently from each other. To address potential interactions between the ships, the risk metric, A, is calculated, which depicts the percentage of the area in the $u - \varphi$ (speed-bearing) space that is not allowed for manoeuvres. A is calculated using concepts from velocity obstacle algorithms (Degre and Lefevre, 1981; Fiorini and Shiller, 1998) according to the following equations:

$$A = \frac{AC}{\pi\, u_{0,max}^2} \qquad (14)$$

where AC is the area in the $u - \varphi$ space that is not available for safe navigation as collision with the ship TS_k can occur according to the holonomic hypothesis. Therefore, AC is defined as:

$$AC = \bigcup_{k=1}^{k=N_v} A_k \cap u - \varphi\ circle \qquad (15)$$

where N_v denotes the number of ships interacting with the own ship, whereas A_k denotes the area in the u–φ space where collision of the OS and the TS_k occurs.

This area is also graphically depicted in Fig. 4 as the intersection between the highlighted triangles and the circle. The metric A is used to depict the traffic complexity and the area that is not available for the safe manoeuvring of the own ship.

The normalised A_n is calculated according to:

$$A_n = \frac{e^A - 1}{e - 1} \qquad (17)$$

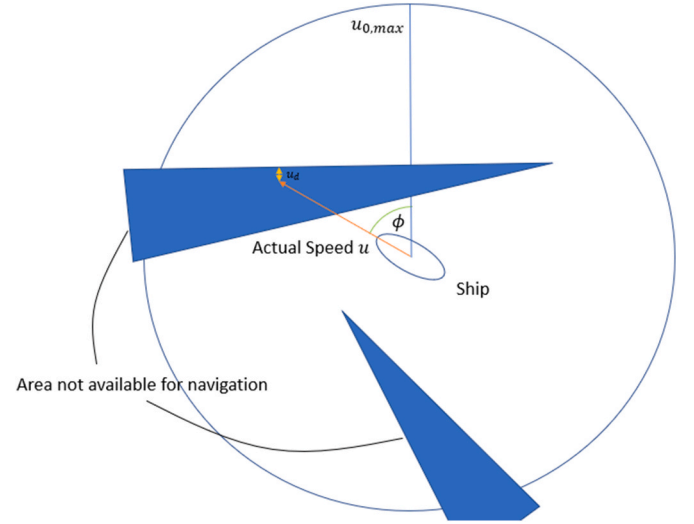The exponential normalisation is used to give greater emphasis on

Fig. 4. Description of $u - \varphi$ area.

values of A_n closer to 1.

The parameter h_{u_d} is estimated according to:

$$h_{u_d} = \begin{cases} \dfrac{u_d}{0.25\, u_{0,max}}, & \text{if the OS speed vector lies within the AC area} \\ 0, & \text{otherwise} \end{cases} \qquad (18)$$

where u_d is the minimum distance between $\overrightarrow{u(\varphi)}$ and the safe area, as depicted in Fig. 4. This metric is used to depict the easiness with which the ship can change its speed and find itself in a safe combination of speed and direction values. The normalisation with $0.25\, u_{0,max}$ is implemented in line with the assumption that a ship cannot instantly change its velocity to the desired level. Herein, this threshold is set arbitrarily for simplicity reasons. Typically, it depends on the ship manoeuvrability characteristics (rudder size, engine power, length, etc.) and available manoeuvring time.

The normalised parameter h_n is calculated by the following equation:

$$h_n = \frac{e^{h_{u_d}} - 1}{e - 1} \qquad (19)$$

The last metric of the risk vector is used to account for the weather conditions during manoeuvring and is estimated according to:

$$W = \sqrt[3]{\frac{H_s V_c V_w}{H_{smax} V_{cmax} V_{wmax}}} \qquad (20)$$

where H_s is the wave height that is considered for the selected scenario, V_c is the current speed, V_w is wind speed, and max denotes the maximum value of H_s, V_c and V_w.

The normalised parameter W_n is calculated by:

$$W_n = \frac{e^W - 1}{e - 1} \qquad (21)$$

The risk metrics presented herein are novel and to the best of the authors' knowledge, have not been employed in previous studies. These metrics could be also employed in real CA system for the risk estimation. However, other metrics potentially could be considered, such as obedience to COLREGs of the present encounter traffic scenario (Du et al., 2021b), metrics having sinusoidal and parabolical relationship with DCPA, TCPA, distance (Abebe et al., 2021), metrics having polynomial relationship with DCPA and TCPA (Kearon, 1997; Park and Jeong, 2021), metrics having exponential relationship with DCPA and

TCPA (Mou et al., 2010), risk metrics for the Degree of Domain Violation (Szlapczynski, 2006), the Time to Domain Violation (Lenart, 2015; Szlapczynski and Szlapczynska, 2016), or alternative approaches for risk estimation (Chun et al., 2021; Ha et al., 2021). However, they are outside the scope of this study, as the aim is to demonstrate the functionality of the proposed process for identify testing scenarios. As there are no definite guidance with respect to which risk metrics must be selected in the maritime community, and it constitutes an area of an intense research efforts, the conclusive selection of the appropriate risk metrics is left as a subject for future work.

Since the risk vector is multi-dimensional, it cannot easily be visualised. In order to enable its visualisation, the t-distributed stochastic neighbour embedding (t-SNE) algorithm is used (Hinton and Roweis, 2002; Van der Maaten and Hinton, 2008). This algorithm first calculates a probability distribution that represents similarities between neighbouring points, (the conditional probability that a point would pick another point as its neighbour), and subsequently defines a similar probability distribution over the points in the low-dimensional mapping. Finally, the t-SNE algorithm minimises the Kullback-Leibler divergence (Kullback, 1968) between the two distributions to eventually derive the final mapping.

2.6. Step 4: Equivalence classes identification

In this step, mathematical tools of unsupervised clustering analysis are employed for the determination of equivalence classes (Bhat and Quadri, 2015). For the purpose of this study, the k-means clustering is selected (Cheng, 1995; Fukunaga and Hostetler, 1975; MacQueen, 1967). This algorithm aims to partition any available observations into k clusters, in which each observation belongs to the cluster with the nearest mean. The parameter k, (denoting the number of clusters) is user selectable.

The following two approaches were used to identify the optimal number of clusters in the available dataset: the elbow method (MacQueen, 1967) and the silhouette method (Steinhaus, 1956). The elbow method involves the calculation of within-cluster error sum of squares (SSE) and identifying the number of clusters after which the SSE stops reducing rapidly. The relevant equations for estimating SSE can be found in (Stanford, 2022).

The silhouette method follows a similar working principle, however it employs the silhouette score to quantify the similarity of data points within a cluster and their distance to data points that belong to other clusters. In this way, the silhouette method calculates how similar an object is to its own cluster (cohesion) compared to other clusters (separation) (de Amorim and Hennig, 2015; Rousseeuw, 1987). The silhouette score ranges from −1 to +1, where a high value indicates that the object matches to the dedicated cluster (de Amorim and Hennig, 2015; Rousseeuw, 1987). For a clustering method to be considered effective, most of the points must have a high value. The mean of silhouette score is used to assess the clustering technique quality. The silhouette score can be calculated as in (de Amorim and Hennig, 2015; Rousseeuw, 1987).

The number of cluster zones is determined according to the following formula:

$$C_N = \left(2^{k+s} - 1\right)3^k \tag{22}$$

where s denotes the number of employed shore points.

The rationale behind eq. (22) is as follows. When a number of k TSs are in proximity to the OS, the OS may collide either with one TS, or with any combination of two or more TSs. Eq. (22) is developed by considering that the potential to collision is expressed using binary values (Yes or No). The subtraction of 1 is used to address the scenario when there is

no collision with any TS or shore (all values are No). The multiplication with 3^k is followed to distinguish if the scenario with the kth TS is classified as crossing, following or heading. This distinction of traffic conditions is reported in (Namgung and Kim, 2021).

Another indication of the potential cluster number is associated with the risk vector (RV) length. By considering that each risk value in the RV can either have a high or low value (binary assumption) and are independent from one another, the total number of clusters can be calculated by:

$$C_N = \left(2^{dim(RV)} - 1\right)3^k \tag{23}$$

The subtraction of 1 is employed to account for conditions where all the RV elements exhibit low values. These conditions are already filtered using the rules in step 2 of the proposed process.

2.7. Step 5: Scenarios selection

Once the equivalence classes (clusters or groups) have been identified, for each class (cluster/group), the sample with its RV closest to the mean RV value of the whole class (cluster/group) is used as representative. However, this does not guarantee to provide the riskiest scenario in the cluster. For this reason, in line with the boundary value analysis (Bhat and Quadri, 2015), the most critical scenario is determined by considering the total risk (TR), which is calculated by:

$$TR = \frac{\sum RVelevements}{dim(RV)} \tag{24}$$

Practically, the average of the RV elements in eq. (11) is considered as the risk measure for each scenario, without assigning any weight to the risk vector elements. In this way, more complex scenarios are associated with higher risk. It is assumed that if the collision is avoided for a scenario with high risk, it will also be avoided for the scenarios with lower risk in the same group.

3. Investigated case and selected parameters

This study considers the case of the sort ship shipping (SSS) cargo ship as the own ship (OS), which is employed at the AUTOSHIP project (AUTOSHIP, 2019). The OS is considered to operate outside the coasts of Norway, and interacts with a high speed craft (TS$_1$) and a sailing boat (TS$_2$).

The input parameters for the investigated scenarios are provided in Table 3. The random parameters with their ranges are provided in Table 4. These 18 parameters are assumed to vary from 0 to their maximum value and are sampled using the Sobol technique as described in section 2. The test area is set to 3 nm × 3 nm in line with (Namgung and Kim, 2021). A shore is also considered, which is represented by a simple spline line. It should be noted that both the sailing boat and the high-speed craft are not required to have an AIS transponder. Hence, they are not visible on the AIS data, which justify the developed approach use.

Table 3
Input parameters.

	Own ship (OS)	1st target ship (TS$_1$)	2nd target ship (TS$_2$)
Type	Cargo ship	High speed craft	Sailing boat
Length	74.7 m	12 m	6 m
Beam	13.6 m	2.5 m	2 m
Max speed	15 kn	40 kn	10 kn
Max current	3 m/s		
Max waves height	2 m		
Max wind speed	14 kn		

Table 4
Random parameters.

Random parameter	Range
Own ship (OS) speed	0–15 kn
OS speed direction	0–2π rad
OS location	0–3 nm × 0–3 nm
Sail boat speed	0–10 kn
Sail boat speed direction	0–2π rad
Sail boat location	0–3 nm × 0–3 nm
High speed craft speed	0–40 kn
High speed craft direction	0–2π rad
High speed craft location	0–3 nm × 0–3 nm
Current speed	0–3 m/s
Current direction	0–2π rad
Waves height	0–2 m
Waves direction	0–2π rad
Wind speed	0–14 kn
Wind direction	0–2π rad

4. Results and discussion

4.1. Step 1: Development of traffic scenario using sampling techniques

A typical sampled traffic scenario is provided in Fig. 6. This is a non-hazardous traffic scenario for the OS, which sails far away from the high-speed craft (TS$_1$) whereas the sailing boat (TS$_2$) slowly approaches the OS. The shore boundary was not shown for demonstration purposes. This scenario can potentially occur during navigation. Therefore, it can be concluded that the artificial sampling can generate realistic traffic conditions, which can be categorised as both safe and unsafe (several unsafe scenarios are shown in the next figures). As the OS is sampled uniformly over the investigated location space (the same applies for both target ships), the process generates scenarios, in which no hazardous situations are observed. This can be viewed as a drawback of the process, as it results in a certain degree of overhead, which can be addressed in a future study.

The sampled positions for the OS (without presenting the target ships TS$_1$ and TS$_2$) with N = 100 (Sobol samples number) are presented in
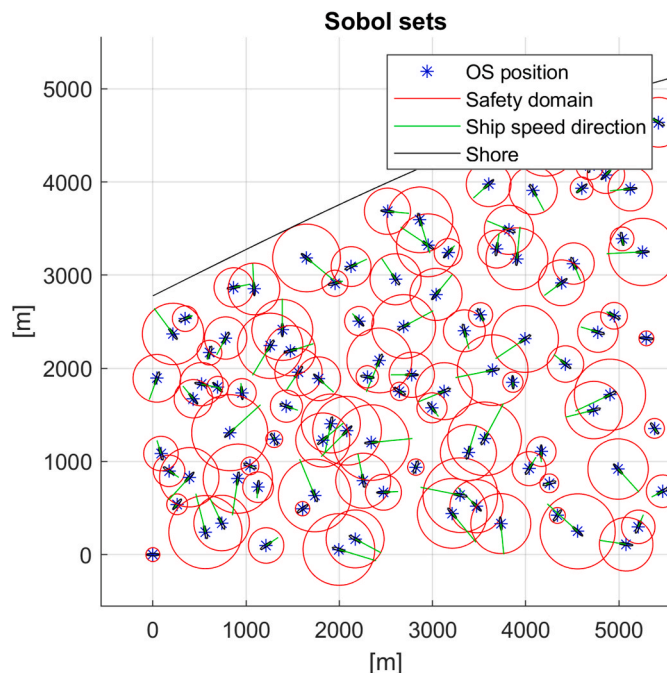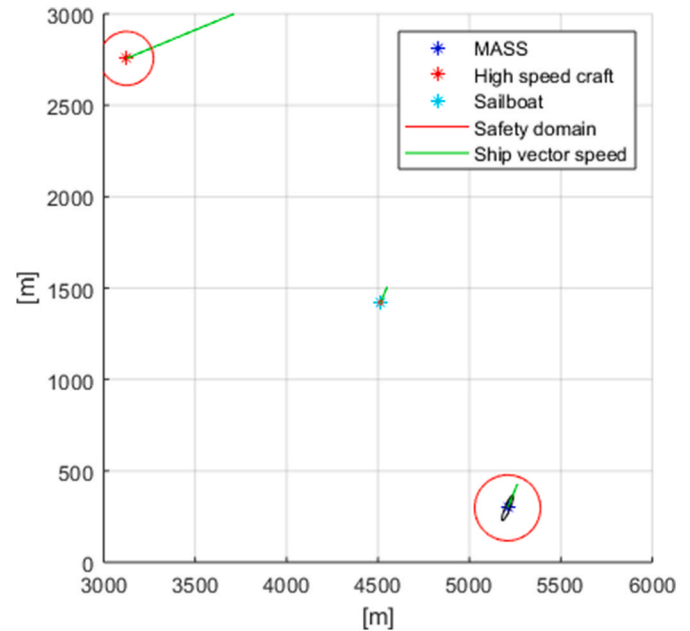

Fig. 6. A typical sampled encounter scenario.

Table 5
Estimation of dif_{min}, T and R, when p = 18 for several Sobol sample numbers.

Number of Sobol samples (N)	Average minimum difference (dif_{min})	Total scenarios number with dif_{min} according to crude sampling (T)	Sampling ratio (R)
100	0.2535	3.12 10^{12}	3.21 10^{-11}
1000	0.2132	3.93 10^{13}	2.55 10^{-11}
10,000	0.1797	5.15 10^{14}	1.94 10^{-11}
100,000	0.1619	2.55 10^{15}	3.95 10^{-11}
1,000,000	0.1428	1.82 10^{16}	5.49 10^{-11}
10,000,000	0.1228	1.99 10^{17}	5.02 10^{-11}

Fig. 5. This figure was developed by illustrating only the OS positions, speed vectors and safety domains for the sampled scenarios. As it can be observed, the Sobol sampling method provided a uniform coverage of the OS positions, speeds and speed directions in the selected area. This is in line with findings from previous studies (Bolbot and Theotokatos, 2021; Burhenne et al., 2011; Kucherenko et al., 2015; Qian and Mahdi, 2020), which justifies the use of the Sobol sampling compared to alternative sampling techniques. The estimations for the parameter dif_{min}, the total number of scenarios, and the sampling ratio for several values of the samples number are provided in Table 5. As it can be observed, the average normalised minimum difference between points (dif_{min}) reduces with increasing samples numbers, however, the sampling ratio remains approximately the same. This also means that with a specified dif_{min}, the Sobol sequence method requires far less samples

Fig. 5. Sampled OS positions, safety domain and speed for N = 100.

Table 6
Estimation of dif_{min}, T and R, with p varying and N = 10,000,000.

Number of parameters	Average minimum difference (dif_{min})	Total scenarios number with dif_{min} according to crude sampling (T)	Sampling ratio (R)
2	2.03 10^{-5}	2.41 10^9	0.0041
3	0.0023	8.79 10^7	0.1138
5	0.0147	1.59 10^9	0.0063
10	0.0717	5.59 10^{11}	1.78 10^{-5}
15	0.1023	3.06 10^{15}	3.27 10^{-9}
18	0.1228	1.99 10^{17}	5.02 10^{-11}

Table 7
Number of hazardous scenarios and clustering zones.

Number of Sobol samples (N)	Number of hazardous scenarios (N_H)	Optimal number of clusters (N_C) and N_C/N_H percentage in parentheses	Calculation time (s)
100	18 (18%)	11 (61%)	1.2
1000	178 (18.7%)	43 (24%)	16.6
2000	367 (18.4%)	79 (22%)	45
5000	898 (18.0%)	97 (11%)	246
10,000	1810 (18.1%)	62 (3%)	958

than the crude sampling of all potential combinations for the desired parameters, thus considerably reducing the number of scenarios. Due to the computational limitations of the employed hardware, this study considers the Sobol samples generated using N = 100, 1000, and 10,000.

It should be also noted that the estimated sampling ratio, dif_{min} and T depend on the number of parameters p, as was revealed based on the calculations presented in Table 6. The sampling ratio R decreases whereas the parameter dif_{min} increases with increasing number of parameters. This results in the following two conclusions. First, the benefit from using sampling increases with the number of parameters, as the sampling ratio decreases for the same dif_{min} compared to the exhaustive combinations number. Secondly, the coverage of the potential parameters becomes scarcer, as p and dif_{min} increase, which means that the sampling results in higher uncertainty.

4.2. Step 2: Identification of the hazardous scenarios

The number of hazardous scenarios found using the rules from the second step is provided in Table 7. As the number of samples increases, the identified number of hazardous scenarios increases accordingly. However, it can be observed that the percentage of identified hazardous scenarios does not significantly change with the sample size. This indicates that there should be a similarity between the different scenarios, therefore their grouping in clusters is possible. However, it should be noted that the required computational time on a standard desktop computer increases exponentially, with the number of samples, which constitutes an important limitation. For this reason, also the estimation of safety metrics based on the holonomic assumption is necessary, whereas the interactions between ships need to be left outside the analysis scope, unless a supercomputer or work/station is employed.

4.3. Step 3: Risk vectors calculations

As mentioned in methodology section, for each hazardous scenario a risk vector is calculated. The risk vector has 6 elements, as two ships and one shoreline are considered. As the visualisation of the estimated risk vector in a multi-dimensional space is impossible, the t-distributed stochastic neighbour embedding (t-SNE) algorithm is applied to reduce the dimensionality of the data and allow for their plotting. The results of this algorithm are presented in Fig. 7. The colour scale represents the normalised Euclidean distance of each point from the origin, prior to the application of the t-SNE algorithm. This is done to enable the visual evaluation of the algorithm results. The visualisation demonstrates that there are distinguished groups of potential traffic scenarios, which can be identified using the clustering techniques.

4.4. Step 4: Equivalence clusters identification

For the range of the considered sample sizes, it was investigated how the selected cluster number affects the performance metrics, in specific the sum of squared error (SSE) and the silhouette score, described in section 2.6. These results are provided in Fig. 8. The SSE metric rapidly reduces with the cluster number, whereas the relationship between cluster number and SSE is monotonic. The SSE also asymptotically tends to 0 with increasing cluster size, which indicates that after a specific cluster number, the improvement in SSE is limited.

The response is different for the silhouette score, where its optimal or convergent values are exhibited. This can be justified by the fact that the grouping converges to well-defined groups. This is anticipated as described in section 2.6, and the expected cluster number for the investigated case study is between 63 according to eq. (22) and 576 according to eq. (23). The silhouette score is constantly on the positive metric side, which indicates that the convergence progresses relatively well.

Based on the optimal values of the silhouette score for the minimum number of clusters, the selected cluster number is provided in Table 7. The results demonstrate that as the number of samples increases, the optimal number of clusters for each sample does not increase in the same manner, but instead also demonstrates a trend towards convergence. This constitutes another evidence that the grouping of scenarios in clusters is possible. However, further investigation is required by using a
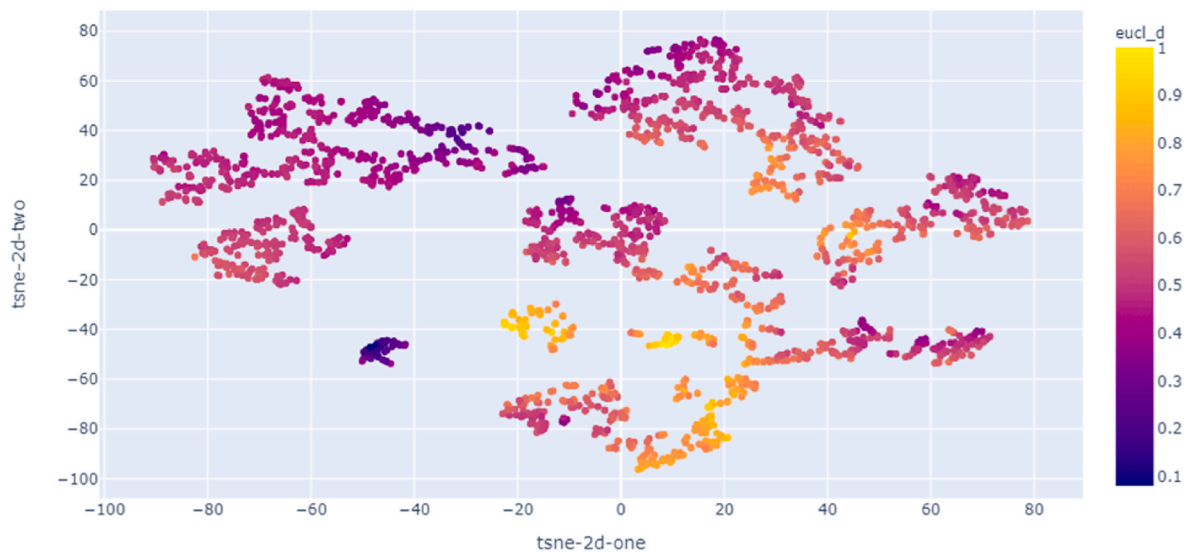

Fig. 7. 2D visualisation of the risk vectors of each scenario through the t-SNE algorithm.
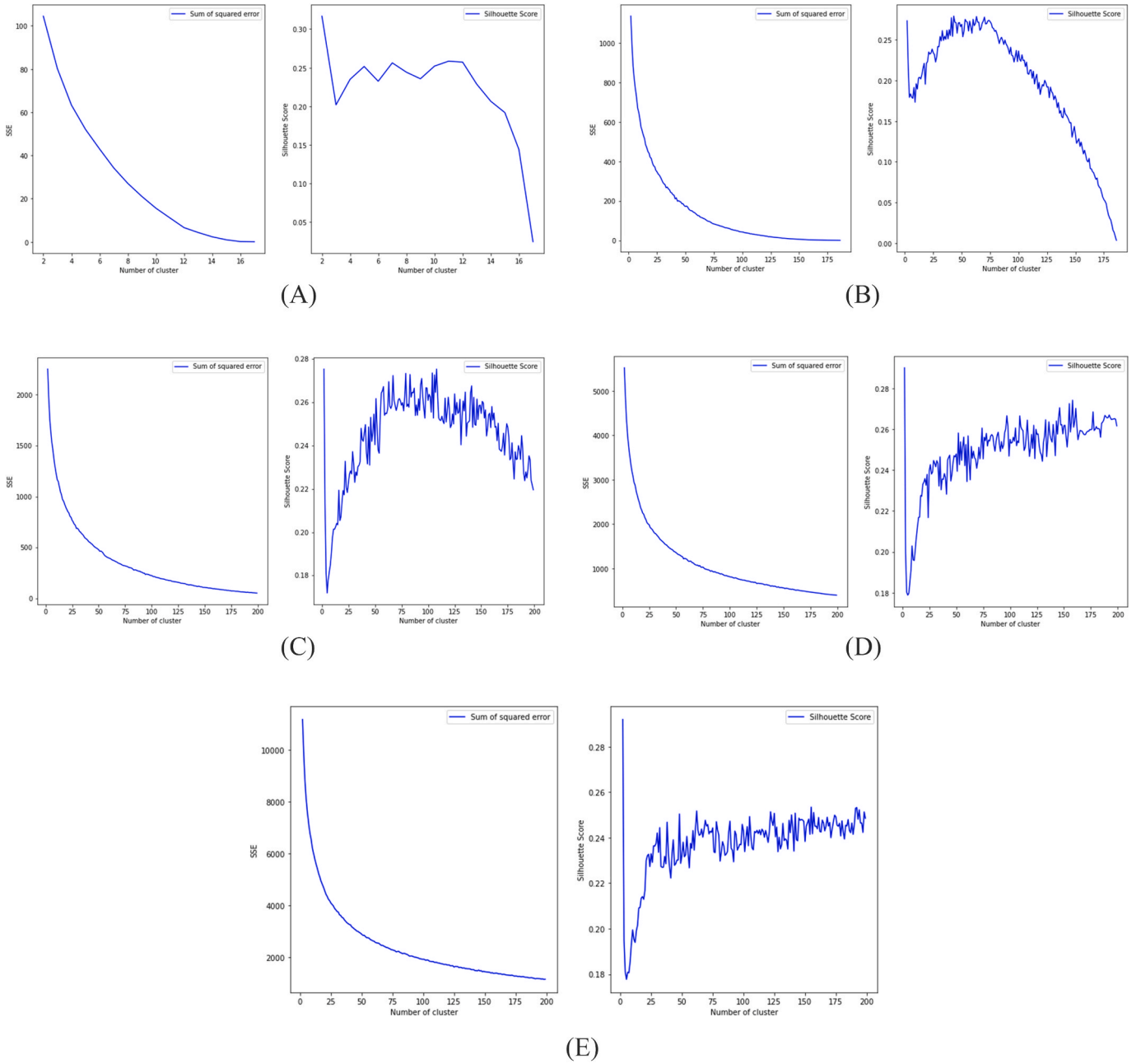
Fig. 8. Derived performance metrics with the cluster number (a) N = 100; (b) N = 1000; (c) N = 2000; (d) N = 5000; (e) N = 10000.

work station of greater computational capacity, so that increased sample sizes are evaluated.

4.5. Step 5: Scenarios selection

The selected scenarios for N = 10,000 are presented for some of the equivalent classes/clusters/groups in Fig. 9. This clustering is selected, as it is the closest to the minimum number of classes/clusters according to eq. (22) (C_N = 63). The selected scenarios depict the different traffic conditions. Some of the traffic encounters are simple, as depicted for the case of Sobol numbers of 7080 and 4640. However, some include both target ships in the proximity to OS, such as, the traffic conditions

generaed using Sobol numbers of 795, 831, and 4183. This is in line with our expectations, since, as elaborated in Section 2.6, we anticipate interactions with both TS. Some of these scenarios demonstrate high complexity and total risk (TR) value, whilst others exhibit lower risk and complexity, however they are different in nature. It should be noted that the present classes/clusters/equivalence groups have been identified based on the considered risk vector and risk metrics. It is expected that the inclusion of additional risk metrics would result in different identified scenarios. This constitutes a subject for future research. Additionally, the Sobol numbers are included in Fig. 9 plots for enhancing traceability. In another case study considering other ships (types and numbers), different Sobol samples are required.
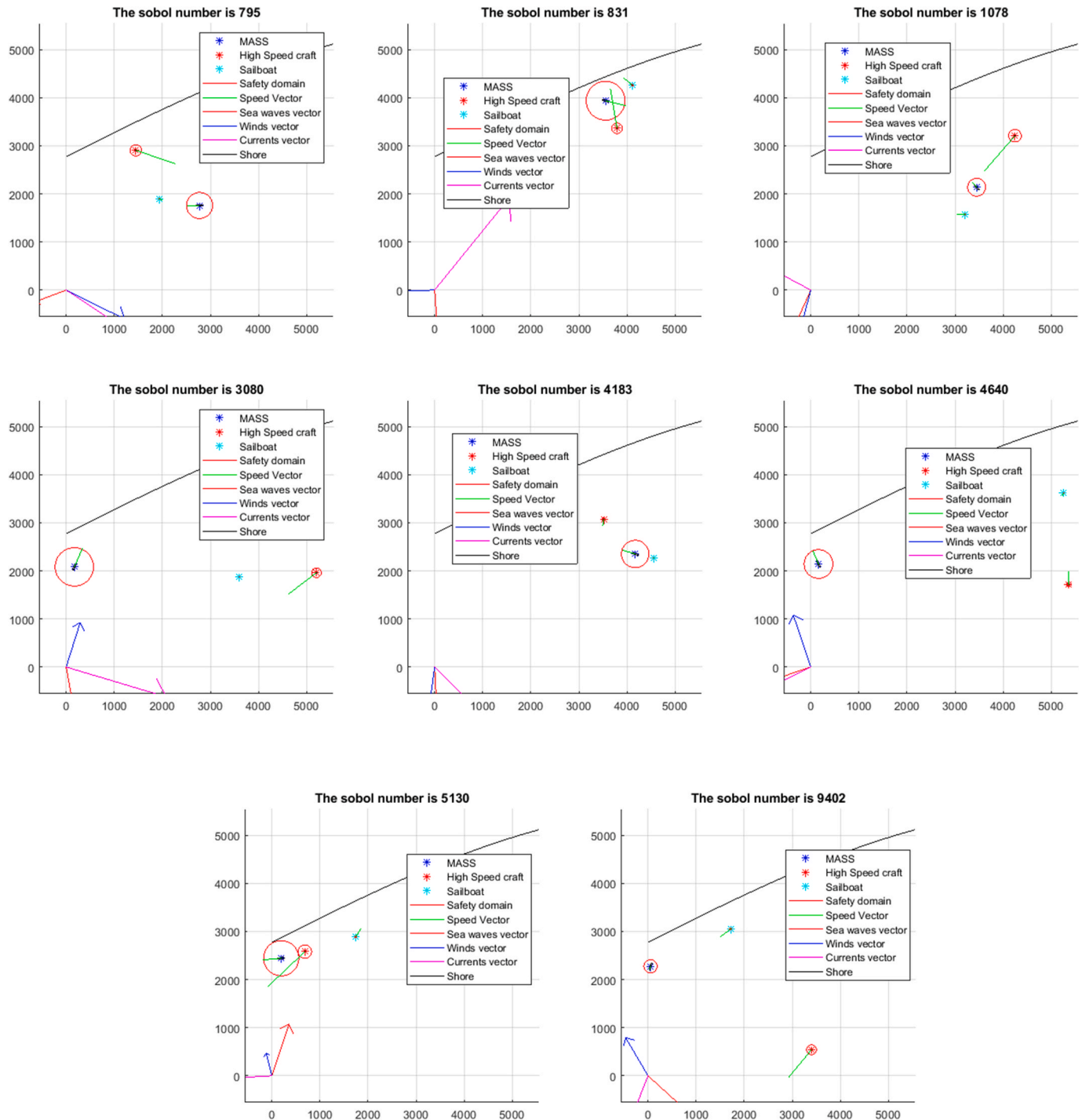
Fig. 9. Most critical scenarios from some equivalence classes/groups with N = 10000 and 62 clusters.

5. Discussion and limitations

The main advantage of the presented process is that it follows a deductive, and not inductive, logic compared to the traffic scenarios generated by using the AIS data. It starts by considering the whole set of potential traffic scenarios, and progresses to more specific scenarios, which are not expected to be the most frequently occurring, as it is expected when AIS data is used. This is a more robust process compared to the case of deriving the potential traffic scenarios based on AIS data. The identification and testing of scenarios that have not been encountered in the past, but might occur in the future, belongs to the 'known unknowns'

region of the Johari window (Luft and Ingham, 1961). The consideration of these types of scenarios during testing will contribute to the safety assurance of the CA system.

The reduction of the potential scenarios is achieved by using more effective coverage of the traffic scenarios (use of sampling techniques during step 1), by excluding non-hazardous scenarios (application of rules in step 2), by grouping the scenarios based on similarity, as well as by using clustering algorithms and selecting the riskiest scenario (steps 3–5). A similar approach, with some modifications could be applied by using AIS data, as demonstrated in (Gao and Shi, 2020; Goerlandt et al., 2017; Han et al., 2021; Kulkarni et al., 2020; Mieczyńska and

Czarnowski, 2021; Mou et al., 2010) for other types of problems; however, it would result in identifying different scenarios due to the use of different input. An important advantage of the proposed process is that it can generate data for the ships, for which AIS equipment is not required, and therefore, no AIS data are collected, e.g., sailboats and leisure high speed crafts. This is important, as these types of ships constitute an important source of hazards for MASS (Bolbot et al., 2021a, 2021c).

The proposed process was demonstrated by considering only 3 ships. Essentially, the analysis needs to be repeated by considering a number of other ships, ship types and shore types to derive a set of testing scenarios. This alone contributes to the complexity of the problem, and the reduction in the considered scenarios is therefore useful. Testing of 60 scenarios for a combination of OS with two TS and shore is far less stricter compared to the requirements reported in (Torben et al., 2022), where more than 250 testing scenarios had to be simulated for a simple head-on encounter situation with one ship, without considering the impact of weather conditions and the presence of other ships/shore.

The process is rather flexible and allows for using alternative methods. Advanced sampling techniques can be employed in step 1, more elaborate sets of rules compared to the ones employed in step 2 can be considered, advanced risk metrics based on the Degree of Domain Violation (Szlapczynski, 2006) and the Time to Domain Violation (Lenart, 2015; Szlapczynski and Szlapczynska, 2016) can be included in step 3, more effective clustering techniques can be employed in step 4, whereas and multiple criteria can be employed in step 5 for selecting the scenarios for testing. The presented process can be further enhanced, and its adaptations constitute a subject of further research. However, we would advocate strongly against the use of Latin Hypercube samples and random samplings, as the Sobol sampling offers much better results performance and stability (Bolbot and Theotokatos, 2021; Burhenne et al., 2011; Kucherenko et al., 2015; Qian and Mahdi, 2020).

The selected safety parameters, such as, d_s, t_s, the safety domain and the associated risk metrics, constitute critical decisions for the proposed approach, as they influence the identified scenarios and their clusters. The plethora of approaches for defining the risk metrics and safety domains reported on the pertinent literature poses a considerable challenge. Standardisation in this area is needed for defining internationally agreed COLREGs requirements for the MASS as well as to promote the use of MASS technologies.

Further investigation is required for defining acceptable values for the parameter dif_{min} and the sampling number N, as they influence the number of identified traffic scenarios. As was demonstrated in the preceding sections, there is an indication that the optimal zones number/metrices converge with the increased sample size, which constitutes an important factor for selecting the sample number in future studies.

An important weakness to the analysis stems from the holonomic assumption use, as the previous ship track and the potential future interactions are not considered during all the process steps, such as during the risk metrics calculations. The use of holonomic assumption was necessitated by the limitations in the computational capacity. It is left as a subject for future research to consider the incorporation of these factors to greater detail. It is expected though that a dense sampling, as demonstrated herein, might be sufficient for effectively identifying the critical scenarios.

Lastly, the application of the process revealed that there are fundamental mathematical problems that need to be addressed. Firstly, we could not locate an analytical expression for calculation of the parameter dif_{min} in the Sobol sequences, although we could identify some other metrics representing the sampling uniformity coverage. Additionally, there is a need for a rigorous mathematical proof that the successful testing of the riskiest scenarios can be extrapolated with a specific confidence level to the less risky scenarios. Proving mathematically that the process converges and identifying the conditions for which the process converges are areas for future research.

6. Conclusions

In this study, a process for identifying a reduced number of traffic scenarios to be considered during the testing of CA system was proposed. This process was based on several steps that include sampling, filtering, and clustering. The process was applied to a cargo ship operating in close proximity to shore, demonstrating a significant reduction in the identified number of scenarios that can be selected for testing either in a virtual environment or full-scale trials. The main findings of this study are summarised as follows.

The proposed process constitutes an effective tool for identifying traffic scenarios and replacing the AIS data, especially for testing the CA system interactions with target ships, AIS data for which is not available.

The process offers a significant reduction in the considered traffic scenarios compared to the crude testing, through the use of sampling techniques especially with increasing number of simulated parameters, without sacrifice in sampling coverage. The process results in much fewer scenarios for testing compared to some other automatic scenario generation schemes.

The categorisation of the traffic scenarios based on their navigational risk metrics using clustering techniques into well distinguished groups was effectively demonstrated.

Identification of complex traffic scenarios that might not be considered by a human, where multiple ships are on a collision track was demonstrated.

The limitations of this study include the following: (a) The derived results depend on the selected criteria/rules, the navigational risk metrics and the clustering algorithm settings; careful selection of these parameters is required to avoid inconsistency of the identified scenarios; (b) Computational limitations necessitate the use of holonomic assumption for the ship motions; advanced computers are needed to facilitate cases with considerable sample sizes.

It is expected that the proposed process for the automatic traffic scenarios identification constitutes a useful tool in the hands of the CA systems designers and certifiers. However, as also highlighted in the preceding sections, there is a potential for the process further improvement. Future studies are recommended to explore: (a) the standardisation of the navigational risk metrics and the considered safety thresholds, (b) the consideration of alternative safety domains (e. g., COLREGs-reflecting domains) and advanced metrics (e.g., time to domain violation and degree of domain violation), (c) complex scenarios with more than two target ships.

CRediT authorship contribution statement

Victor Bolbot: Conceptualization, Methodology, Data curation, Writing – original draft, Writing – review & editing, Visualizzation, Project administration. **Christos Gkerekos:** Conceptualization, Methodology, Data curation, Writing – original draft, Writing – review & editing, Visualizzation. **Gerasimos Theotokatos:** Methodology, Resources, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. **Evangelos Boulougouris:** Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Appendix A. Nomenclature and abbreviation list

AIS	Automatic Identification System
MASS	Maritime autonomous surface ships
OS	Own Ship
RV	Risk vector
SSE	Within-cluster error sum of squares
TS	Target Ship

Nomenclature

$a_{i,1}$	Radius of safety domain around the OS (m)
A	% of area in the $u - \varphi$ (speed-bearing) space (-)
AC	area in the $u - \varphi$ (speed-bearing) space (m/s degree)
C_N	Minimum number of cluster zones (-)
u_d	Minimum distance between $\overrightarrow{u(\varphi)}$ and the safe area in the $u - \varphi$ (speed-bearing) space (m/s)
d_s	Distance threshold (nm)
$D_{i,k}$	Distance between OS and TS_k ship for i sampled scenario (m)
$DCPA_{i,k}$	Distance at the closest point of approach between the OS and TS_k (m)
$DCPAn_k$	Normalised version of $DCPA_{i,k}$ (-)
dif_{min}	Average normalised minimum difference in parameter p_l of the range $(p_{l,max} - p_{l,min})$
h_{u_d}	Normalised version of u_d (-)
H_s	Wave height (m)
L	Ship length (m)
N	Number of sampled points (-)
N_C	Number of clusters (-)
N_H	Number of hazardous scenarios (-)
p_l	Parameter (-)
p	Total number of parameters (-)
R	Reduction ratio (-)
SSE	Sum of Squares Error (-)
t_s	Time threshold (min)
T	Potential number of scenarios considering all possible combinations (-)
$TCPA_{i,k}$	Time to the closest point of approach between the OS and TS_k (min)
$TCPAn_k$	Normalised version of $TCPA_{i,k}$ (-)
TR	Total risk (-)
u	Speed (m/s)
$u_{0,max}$	Maximum speed for OS (m/s)
V_c	Current speed (m/s)
V_w	Wind speed (m/s)
W	Metric is used to depict the weather conditions (BF)
φ	Ship speed direction (degree)
$x_{p,i}$	Value of sampled parameter p_l in scenario i (depending on the type)
$(x, y$	) Location for each ship (m,m)
Index	
0	Own ship
i	Sampled scenario
k	TS_k ship
j	Other sampled scenario than i
n	Normalised parameter

References

AAWA, 2016. AAWA Project Introduces the Project's First Commercial Ship Operators.

Abaei, M.M., Hekkenberg, R., BahooToroody, A., 2021. A multinomial process tree for reliability assessment of machinery in autonomous ships. Reliab. Eng. Syst. Saf. 210, 107484.

Abebe, M., Noh, Y., Seo, C., Kim, D., Lee, I., 2021. Developing a ship collision risk index estimation model based on Dempster-Shafer theory. Appl. Ocean Res. 113, 102735.

Alexander, R., Hawkins, H.R., Rae, A.J., 2015. Situation Coverage–A Coverage Criterion for Testing Autonomous Robots.

Analytic Technologies, 2010. Introduction to Sampling.

Arora, P.K., Bhatia, R., 2018. A systematic review of agent-based test case generation for regression testing. Arabian J. Sci. Eng. 43 (2), 447–470.

AUTOSHIP, 2019. Autonomous Shipping Initiative for European Waters.

Bačkalov, I., 2020. Safety of autonomous inland vessels: an analysis of regulatory barriers in the present technical standards in Europe. Saf. Sci. 128, 104763.

BahooToroody, A., Abaei, M.M., Banda, O.V., Kujala, P., De Carlo, F., Abbassi, R., 2022. Prognostic health management of repairable ship systems through different autonomy degree; from current condition to fully autonomous ship. Reliab. Eng. Syst. Saf. 221, 108355.

Bakdi, A., Glad, I.K., Vanem, E., Engelhardtsen, Ø., 2020. AIS-based multiple vessel collision and grounding risk identification based on adaptive safety domain. J. Mar. Sci. Eng. 8 (1), 5.

Basnet, S., Bahootoroody, A., Chaal, M., Valdez Banda, O.A., Lahtinen, J., Kujala, P., 2022. A decision-making framework for selecting an MBSE language–A case study to ship pilotage. Expert Syst. Appl. 193, 116451.

Bhat, A., Quadri, S., 2015. Equivalence class partitioning and boundary value analysis-A review. In: 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, pp. 1557–1562.

Bolbot, V., Theotokatos, G., 2021. Automatically Generating Collision Scenarios for Testing Ship Collision Avoidance System Using Sampling Techniques. LEAC.

Bolbot, V., Theotokatos, G., Andreas Wennersberg, L., Faivre, J., Vassalos, D., Boulougouris, E., Jan Rødseth, Ø., Andersen, P., Pauwelyn, A.-S., Van Coillie, A., 2021a. A novel risk assessment process: application to an autonomous inland waterways ship. Proc. Inst. Mech. Eng. O J. Risk Reliab. 0 (0), 1748006X211051829.

Bolbot, V., Theotokatos, G., Boulougouris, E., Wennersberg, L., Nordahl, H., Rødseth, Ø. J., Faivre, J., Colella, M.M., 2020. Paving the Way toward Autonomous Shipping Development for European Waters–The AUTOSHIP Project.

Bolbot, V., Theotokatos, G., Bujorianu, L.M., Boulougouris, E., Vassalos, D., 2019. Vulnerabilities and safety assurance methods in Cyber-Physical Systems: a comprehensive review. Reliab. Eng. Syst. Saf. 182, 179–193.

Bolbot, V., Theotokatos, G., Hamann, R., Psarros, G., Boulougouris, E., 2021b. Dynamic Blackout probability monitoring system for Cruise ship power plants. Energies 14 (20).

Bolbot, V.T., Theotokatos, G., Wennersberg, L.A., 2022. A method to identify and rank objects and hazardous interactions affecting autonomous ships navigation. Journal of Navigation.

Bolbot, V., Theotokatos, G., Andreas Wennersberg, L., et al., October 2021. A novel risk assessment process: Application to an autonomous inland waterways ship. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability. https://doi.org/10.1177/1748006X211051829.

Brcko, T., Androjna, A., Srše, J., Boć, R., 2021. Vessel multi-parametric collision avoidance decision model: fuzzy approach. J. Mar. Sci. Eng. 9 (1), 49.

Burhenne, S., Jacob, D., Henze, G.P., 2011. Sampling based on Sobol'sequences for Monte Carlo techniques applied to building simulations. In: Proc. Int. Conf. Build. Simulat, pp. 1816–1823.

Bye, R.J., Aalberg, A.L., 2018. Maritime navigation accidents and risk indicators: an exploratory statistical analysis using AIS data and accident reports. Reliab. Eng. Syst. Saf. 176, 174–186.

Chen, P., Huang, Y., Mou, J., van Gelder, P.H.A.J.M., 2019. Probabilistic risk analysis for ship-ship collision: state-of-the-art. Saf. Sci. 117, 108–122.

Cheng, Y., 1995. Mean shift, mode seeking, and clustering. IEEE Trans. Pattern Anal. Mach. Intell. 17 (8), 790–799.

Chun, D.-H., Roh, M.-I., Lee, H.-W., Ha, J., Yu, D., 2021. Deep reinforcement learning-based collision avoidance for an autonomous ship. Ocean Eng. 234, 109216.

Clark, A., Walkinshaw, N., Hierons, R.M., 2021. Test Case Generation for Agent-Based Models: A Systematic Literature Review. Information and Software Technology, p. 106567.

Coldwell, T.G., 1983. Marine traffic behaviour in restricted waters. J. Navig. 36 (3), 430–444.

COLREGS, 1972. International Regulations for Preventing Collisions at Sea - Articles of the Convention on the International Regulations for Preventing Collisions at Sea.

Daffey, K., 2018. Technology Progression of Maritime Autonomous Surface Ships.

Davis, P.V., Dove, M.J., Stockel, C.T., 1982. A computer simulation of multi-ship encounters. J. Navig. 35 (2), 347–352.

de Amorim, R.C., Hennig, C., 2015. Recovering the number of clusters in data sets with noise features using feature rescaling factors. Inf. Sci. 324, 126–145.

Degre, T., Lefevre, X., 1981. A collision avoidance system. J. Navig. 34 (2), 294–302.

Du, L., Banda, O.A.V., Huang, Y., Goerlandt, F., Kujala, P., Zhang, W., 2021a. An empirical ship domain based on evasive maneuver and perceived collision risk. Reliab. Eng. Syst. Saf. 213, 107752.

Du, L., Valdez Banda, O.A., Goerlandt, F., Kujala, P., Zhang, W., 2021b. Improving near miss detection in maritime traffic in the Northern Baltic sea from AIS data. J. Mar. Sci. Eng. 9 (2), 180.

Erceg, B.Č., 2018. Legal framework of European inland waterways and Croatian legislation on inland waterways navigation-problems of non-harmonized rules. Eur. Sci. J. 14, 37–57.

Federation, F.a.f.m.a.r.t.M.o.T.o.t.R., 2021. Recommendations for Application of COLREGs 1972 in MASS.

Fedorowski, J., Galor, W., Hajduk, J., 1979. Testing the Sperry collision avoidance system. J. Navig. 32 (1), 84–90.

Fiorini, P., Shiller, Z., 1998. Motion planning in dynamic environments using velocity obstacles. Int. J. Robot Res. 17 (7), 760–772.

Fossen, S., 2018. Visualization of Ships in a Mixed-Reality Environment and Automated Situational Awareness Using Live AIS Data. NTNU.

Fossen, T.I., 2002. Marine Control Systems–Guidance. Navigation, and Control of Ships, Rigs and Underwater Vehicles. Marine Cybernetics, Trondheim, Norway, Org., ISBN 82 92356 00 2. Number NO 985 195 005 MVA. www.marinecybernetics.com

Foster, S., Gleirscher, M., Calinescu, R., 2020. Towards deductive verification of control algorithms for autonomous marine vehicles. In: 2020 25th International Conference on Engineering of Complex Computer Systems (ICECCS). IEEE, pp. 113–118.

Fremont, D.J., Kim, E., Pant, Y.V., Seshia, S.A., Acharya, A., Bruso, X., Wells, P., Lemke, S., Lu, Q., Mehta, S., 2020. Formal Scenario-Based Testing of Autonomous Vehicles: from Simulation to the Real World. IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), pp. 1–8, 2020.

Fujii, Y., Tanaka, K., 1971. Traffic capacity. J. Navig. 24 (4), 543–552.

Fukunaga, K., Hostetler, L., 1975. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans. Inf. Theor. 21 (1), 32–40.

Gao, M., Shi, G.-Y., 2020. Ship collision avoidance anthropomorphic decision-making for structured learning based on AIS with Seq-CGAN. Ocean Eng. 217, 107922.

Goerlandt, F., Goite, H., Valdez Banda, O.A., Höglund, A., Ahonen-Rainio, P., Lensu, M., 2017. An analysis of wintertime navigational accidents in the Northern Baltic Sea. Saf. Sci. 92, 66–84.

Goerlandt, F., Kujala, P., 2014. On the reliability and validity of ship–ship collision risk analysis in light of different perspectives on risk. Saf. Sci. 62, 348–365.

Goerlandt, F., Montewka, J., Kuzmin, V., Kujala, P., 2015. A risk-informed ship collision alert system: framework and application. Saf. Sci. 77, 182–204.

Guiochet, J., Machin, M., Waeselynck, H., 2017. Safety-critical advanced robots: a survey. Robot. Autonom. Syst. 94, 43–52.

Ha, J., Roh, M.-I., Lee, H.-W., 2021. Quantitative calculation method of the collision risk for collision avoidance in ship navigation using the CPA and ship domain. J. Comput. Des. Eng. 8 (3), 894–909.

Han, X., Armenakis, C., Jadidi, M., 2021. Modeling vessel behaviours by clustering AIS data using optimized DBSCAN. Sustainability 13 (15), 8162.

Hansen, M.G., Jensen, T.K., Lehn-Schiøler, T., Melchild, K., Rasmussen, F.M., Ennemark, F., 2013. Empirical ship domain based on AIS data. J. Navig. 66 (6), 931.

Hilgert, H., Baldauf, M., 1997. A common risk model for the assessment of encounter situations on board ships. Deutsche Hydrografische Zeitschrift 49 (4), 531–542.

Hinton, G., Roweis, S.T., 2002. Stochastic Neighbor Embedding. NIPS. Citeseer, pp. 833–840.

Hu, Q., Yang, C., Cheng, H., Xiao, B., 2008. Planned route based negotiation for collision avoidance between vessels. TransNav: Int. J. Mar. Navigat. Saf. Sea Transport. 2 (4), 363–368.

Huang, Y., Chen, L., Chen, P., Negenborn, R.R., Van Gelder, P., 2020. Ship collision avoidance methods: state-of-the-art. Saf. Sci. 121, 451–473.

Huang, Y., Chen, L., Van Gelder, P., 2019. Generalized velocity obstacle algorithm for preventing ship collisions at sea. Ocean Eng. 173, 142–156.

Huang, Y., Gelder, P., 2017. Non-linear Velocity Obstacles with Applications to the Maritime Domain. Maritime Transportation and Harvesting of Sea Resources, pp. 999–1007.

Huang, Y., van Gelder, P., 2020a. Collision risk measure for triggering evasive actions of maritime autonomous surface ships. Saf. Sci. 127, 104708.

Huang, Y., Van Gelder, P., Wen, Y., 2018. Velocity obstacle algorithms for collision prevention at sea. Ocean Eng. 151, 308–321.

Huang, Y., van Gelder, P.H.A.J.M., 2020b. Collision risk measure for triggering evasive actions of maritime autonomous surface ships. Saf. Sci. 127, 104708.

IMO, 2015. Revised guidelines for the onboard operational use of shipborne Automatic Identification Systems (AIS). Resolution A 1106 (29).

IMO, 2020. Regulatory Scoping Exercise.

IMO, 2021. Outcome of the Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships. MASS), London, UK.

Kearon, J., 1997. Computer Programs for Collision Avoidance and Traffic Keeping Conference on Mathematical Aspects on Marine Traffic. Academic press, London.

Khastgir, S., Brewerton, S., Thomas, J., Jennings, P., 2021. Systems Approach to Creating Test Scenarios for Automated Driving Systems. Reliability Engineering & System Safety, p. 107610.

Kijima, K., Furukawa, Y., 2003. Automatic collision avoidance system using the concept of blocking area. IFAC Proc. Vol. 36 (21), 223–228.

Kucherenko, S., Albrecht, D., Saltelli, A., 2015. Exploring Multi-Dimensional Spaces: A Comparison of Latin Hypercube and Quasi Monte Carlo Sampling Techniques arXiv preprint arXiv:1505.02350.

Kulkarni, K., Goerlandt, F., Li, J., Banda, O.V., Kujala, P., 2020. Preventing shipping accidents: past, present, and future of waterway risk management with Baltic Sea focus. Saf. Sci. 129, 104798.

Kullback, S., 1968. Information Theory and Statistics. Dover Publications, New York, United States.

Lenart, A.S., 1983. Collision threat parameters for a new radar display and plot technique. J. Navig. 36 (3), 404–410.

Lenart, A.S., 2015. Analysis of collision threat parameters and criteria. J. Navig. 68 (5), 887–896.

Li, L., Huang, W., Liu, Y., Zheng, N., Wang, F., 2016. Intelligence testing for autonomous vehicles: a new approach. IEEE Trans. Intel. Veh. 1 (2), 158–166.

Li, M., Mou, J., Chen, L., He, Y., Huang, Y., 2021. A rule-aware time-varying conflict risk measure for MASS considering maritime practice. Reliab. Eng. Syst. Saf. 215, 107816.

Lindvall, M., Porter, A., Magnusson, G., Schulze, C., 2017. Metamorphic model-based testing of autonomous systems. In: 2017 IEEE/ACM 2nd International Workshop on Metamorphic Testing (MET). IEEE, pp. 35–41.

Liu, J., Zhou, F., Li, Z., Wang, M., Liu, R.W., 2016. Dynamic ship domain models for capacity analysis of restricted water channels. J. Navig. 69 (3), 481–503.

Liu, Z., Wu, Z., Zheng, Z., 2019. A cooperative game approach for assessing the collision risk in multi-vessel encountering. Ocean Eng. 187, 106175.

Luft, J., Ingham, H., 1961. The johari window. Hum. Relat. Train. News 5 (1), 6–7.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Oakland, CA, USA, pp. 281–297.

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21 (2), 239–245.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21 (6), 1087–1092.

Mieczyńska, M., Czarnowski, I., 2021. K-means clustering for SAT-AIS data analysis. WMU J. Marit. Aff. 20 (3), 377–400.

Minne, P.K.E., 2017. Automatic Testing of Maritime Collision Avoidance Algorithms. NTNU.

Mizythras, P., Pollalis, C., Boulougouris, E., Theotokatos, G., 2021. A novel decision support methodology for oceangoing vessel collision avoidance. Ocean Eng. 230, 109004.

Montewka, J., Goerlandt, F., Kujala, P., 2012. Determination of collision criteria and causation factors appropriate to a model for estimating the probability of maritime accidents. Ocean Eng. 40, 50–61.

Mou, J.M., Van der Tak, C., Ligteringen, H., 2010. Study on collision avoidance in busy waterways by using AIS data. Ocean Eng. 37 (5–6), 483–490.

MUNIN, 2016. Maritime Unmanned Navigation through Intelligence in Networks.

Namgung, H., Kim, J.-S., 2021. Collision risk inference system for maritime autonomous surface ships using COLREGs rules compliant collision avoidance. IEEE Access 9, 7823–7835.

National Transportation Safety Board (NTSB), 2017. Collision between a Car Operating with Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida May 7, 2016 Highway Accident Report. National Transportation Safety Board, United States, Washington D.C.

Nidhra, S., Dondeti, J., 2012. Black box and white box testing techniques-a literature review. Int. J. Embed. Syst. Appl. (IJESA) 2 (2), 29–50.

Nzengu, W., Faivre, J., Pauwelyn, A.-S., Bolbot, V., Lien Wennersberg, L.A., Theotokatos, G., 2021. Regulatory Framework Analysis for the Unmanned Inland Waterway Vessel. WMU Journal of Maritime Affairs.

Ong, M.S., Kuang, Y.C., Ooi, M.P.-L., 2012. Statistical measures of two dimensional point set uniformity. Comput. Stat. Data Anal. 56 (6), 2159–2181.

Park, J., Jeong, J.-S., 2021. An estimation of ship collision risk based on relevance vector machine. J. Mar. Sci. Eng. 9 (5), 538.

Park, J., Kim, J., 2020. Autonomous Docking of an Unmanned Surface Vehicle Based on Reachability Analysis, 2020 20th International Conference on Control. Automation and Systems (ICCAS), pp. 962–966.

Pedersen, P.T., 2010. Review and application of ship collision and grounding analysis procedures. Mar. Struct. 23 (3), 241–262.

Pedersen, T.A., Glomsrud, J.A., Ruud, E.-L., Simonsen, A., Sandrib, J., Eriksen, B.-O.H., 2020. Towards simulation-based verification of autonomous navigation systems. Saf. Sci. 129, 104799.

Peel, D., Good, N.M., 2011. A hidden Markov model approach for determining vessel activity from vessel monitoring system data. Can. J. Fish. Aquat. Sci. 68 (7), 1252–1264.

Pietrzykowski, Z., Uriasz, J., 2009. The ship domain – a criterion of navigational safety assessment in an open sea area. J. Navig. 62 (1), 93–108.

Pietrzykowski, Z., Wielgosz, M., 2021. Effective ship domain–Impact of ship size and speed. Ocean Eng. 219, 108423.

Porathe, T., 2019. Maritime Autonomous Surface Ships (MASS) and the COLREGS: Do We Need Quantified Rules or Is "The Ordinary Practice of Seamen" Specific Enough?.

Puisa, R., Lin, L., Bolbot, V., Vassalos, D., 2018. Unravelling causal factors of maritime incidents and accidents. Saf. Sci. 110, 124–141.

Qian, G., Mahdi, A., 2020. Sensitivity analysis methods in the biomedical sciences. Math. Biosci. 323, 108306.

Riedmaier, S., Schneider, J., Danquah, B., Schick, B., Diermeyer, F., 2021. Non-deterministic model validation methodology for simulation-based safety assessment of automated vehicles. Simulat. Model. Pract. Theor. 109, 102274.

Ringbom, H., 2019. Regulating autonomous ships—concepts, challenges and precedents. Ocean Dev. Int. Law 50 (2–3), 141–169.

Rødseth, Ø.J., Faivre, J., Hjørungnes, S.R., Andersen, P., Bolbot, V., Pauwelyn, A.-S., Wennersberg, L.A., 2020. AUTOSHIP Deliverable D3.1 Autonomous Ship Design Standards. Revision 2.0.

Rokseth, B., Haugen, O.I., Utne, I.B., 2019. Safety Verification for Autonomous Ships, MATEC Web of Conferences. EDP Sciences, 02002.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65.

Shokri-Manninen, F., Vain, J., Waldén, M., 2020. Formal verification of COLREG-based navigation of maritime autonomous systems. In: de Boer, F., Cerone, A. (Eds.), Software Engineering and Formal Methods. Springer International Publishing, Cham, pp. 41–59.

Silveira, P., Teixeira, A., Figueira, J., Soares, C.G., 2021. A Multicriteria Outranking Approach for Ship Collision Risk Assessment. Reliability Engineering & System Safety, p. 107789.

Smartmaritime, 2020. ASKO to Build Two Autonomous Vessels for Oslo Fjord Operations.

Sobol', I.y.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. Zh. Vychislitel noi Mat. Mat. Fiz. 7 (4), 784–802.

Sørensen, A.J., Ludvigsen, M., 2017. Underwater technology platforms. In: Encyclopedia of Maritime and Offshore Engineering, pp. 1–11.

Stanford, 2022. Error Sum of Squares (SSE).

Stankiewicz, P.G., Mullins, G.E., 2019. Improving Evaluation Methodology for Autonomous Surface Vessel Colregs Compliance, OCEANS 2019-Marseille. IEEE, pp. 1–7.

Steinhaus, H., 1956. Sur la division des corps matériels en parties. Bull. Acad. Polon. Sci. 1 (804), 801.

Szlapczynski, R., 2006. A unified measure of collision risk derived from the concept of A ship domain. J. Navig. 59 (3), 477–490.

Szlapczynski, R., Szlapczynska, J., 2015. A simulative comparison of ship domains and their polygonal approximations. TransNav, Int. J. Mar. Navigat. Saf. Sea Transport. 9 (1), 135–141.

Szlapczynski, R., Szlapczynska, J., 2016. An analysis of domain-based ship collision risk parameters. Ocean Eng. 126, 47–56.

Szlapczynski, R., Szlapczynska, J., 2017. Review of ship safety domains: models and applications. Ocean Eng. 145, 277–289.

Szlapczynski, R., Szlapczynska, J., 2021. A ship domain-based model of collision risk for near-miss detection and Collision Alert Systems. Reliab. Eng. Syst. Saf. 214, 107766.

Tam, C., Bucknall, R., 2010. Collision risk assessment for ships. J. Mar. Sci. Technol. 15 (3), 257–270.

The weather window, 2022. Visibility Measurements at Sea.

Torben, T.R., Glomsrud, J.A., Pedersen, T.A., Utne, I.B., Sørensen, A.J., 2022. Automatic simulation-based testing of autonomous ships using Gaussian processes and temporal logic. Proc. Inst. Mech. Eng. O J. Risk Reliab. 1748006X211069277.

Tsoumpris, C., Theotokatos, G., 2022. Performance and reliability monitoring of ship hybrid power plants. J. ETA Marit. Sci. 10 (1), 29–38.

Utne, I.B., Rokseth, B., Sørensen, A.J., Vinnem, J.E., 2020. Towards supervisory risk control of autonomous ships. Reliab. Eng. Syst. Saf. 196, 106757.

Van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9 (11).

Vincent, T.L., 1977. Collision Avoidance at Sea. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 205–221.

Wang, B., He, Y., Hu, W., Mou, J., Li, L., Zhang, K., Huang, L., 2021. A decision-making method for autonomous collision avoidance for the stand-on vessel based on motion process and COLREGs. J. Mar. Sci. Eng. 9 (6), 584.

Wang, N., 2010. An intelligent spatial collision risk based on the quaternion ship domain. J. Navig. 63 (4), 733.

Woerner, K., 2014. COLREGS-compliant Autonomous Collision Avoidance Using Multi-Objective Optimization with Interval Programming. MASSACHUSETTS INST OF TECH CAMBRIDGE.

Woerner, K., Benjamin, M.R., Novitzky, M., Leonard, J.J., 2019. Quantifying protocol evaluation for autonomous collision avoidance. Aut. Robots 43 (4), 967–991.

Yang, S., Li, L., Suo, Y., Chen, G., 2007. Study on Construction of Simulation Platform for Vessel Automatic Anti-collision and its Test Method, 2007 IEEE International Conference on Automation and Logistics. IEEE, pp. 2414–2419.

Yara, 2018. Yara Birkeland press kit.

Yu-Hong, L., Chao-Jian, S., 2005. A fuzzy-neural inference network for ship collision avoidance. In: 2005 International Conference on Machine Learning and Cybernetics, vol. 4758, pp. 4754–4759.

Zaremba, S., 1968. The mathematical basis of Monte Carlo and quasi-Monte Carlo methods. SIAM Rev. 10 (3), 303–314.

Zhang, W., Kopca, C., Tang, J., Ma, D., Wang, Y., 2017. A systematic approach for collision risk analysis based on AIS data. J. Navig. 70 (5), 1117–1132.