Ships traffic encounter scenarios generation using sampling and clustering techniques

Victor Bolbot, *University of Strathclyde*, victor.bolbot@strath.ac.uk

Christos Gkerekos, *University of Strathclyde*, christos.gkerekos@strath.ac.uk

Gerasimos Theotokatos, *University of Strathclyde*, gerasimos.theotokatos@strath.ac.uk

ABSTRACT

The Marine Autonomous Surface Ships (MASS) constitute a novel type of systems, which require novel methods for their design and safety assurance. The collision avoidance system is considered one of the most critical systems for MASS. This study aims at developing a process for generating and selecting ship encounter scenarios to test the collision avoidance system in a virtual environment. The proposed process employs sampling techniques for generating encounter scenarios, deterministic criteria for identifying the hazardous scenarios, risk metrics estimation for the classification of the encounter situations, as well as clustering techniques for further downsizing of the scenarios number. This process is applied to a small short-shipping vessel thus demonstrating its applicability.

Keywords: *Marine Autonomous Surface Ships; Collision Avoidance System; Safety; Testing*

1. INTRODUCTION

For enhancing the sustainability of the maritime industry, novel systems and technologies have been developed, including marine autonomous surface ships (MASS) (AUTOSHIP, 2019). In these ships, the collision avoidance system is considered one of the most critical systems (Bolbot et al., 2020). The development, validation and verification of safe and robust collision avoidance systems is a challenging task, which requires the identification of scenarios that need to be tested either by employing a virtual environment (simulator) or during sea trials.

The navigation of ships is primarily regulated by the - International Regulations for Preventing Collisions at Sea (COLREGS) (COLREGS, 1972). However, the COLREGS requirements were developed considering manned ships and not the MASS. COLREGS do not provide quantitative criteria for categorising the MASS navigation actions, whilst their implementation rely on the crew judgement, so they cannot be used to develop testing scenarios for future autonomous ships (Woerner et al., 2019). Data acquired from the Automatic Identification System (AIS) can be used for that purposes (Gao and Shi, 2020; Goerlandt et al., 2017; Kulkarni et al., 2020; Mou et al., 2010), but it also entails a number of limitations(IMO, 2015). Several previous studies investigated the development of collision avoidance system for MASS, e.g. (Brcko et al., 2021; Huang et al., 2020; Huang and van Gelder, 2020; Namgung and Kim, 2021), however, very few studies focused on the testing of autonomous ships and the testing scenarios generation (Pedersen et al., 2020; Woerner, 2014).

This study aims at proposing a process to develop testing scenarios for collision avoidance system of MASS. The developed approach integrates methods from different research areas, namely statistical sampling, ship manoeuvrability studies, big data analytics, and software testing techniques. This is elaborated in more detail in the next sections.

The remainder of this paper is structured as follows. First a brief presentation of the developed process is implemented. Then the investigated case parameters are provided. Lastly the results and considerations for future research are discussed.

2. PROCESS DESCRIPTION

The proposed process consists of five steps as illustrated in the flowchart of **Error! Reference source not found.**. The required input includes the parameters for the ships and operational area, the weather conditions, the size of canals or fjords, as well as the sea depth.

Step 1 employs sampling of the selected parameters to develop the encounter situations. Whilst a plethora of methods can be employed herein, the Sobol sampling technique was selected due its ability to offer an effective coverage of the sampling space and relative results robustness (Bolbot and Theotokatos, 2021; Burhenne et al., 2011; Kucherenko et al., 2015; Qian and Mahdi, 2020). Sampling is implemented considering the parameters range from their minimum to maximum values.

The hazardous situations are identified using a set of deterministic rules in step two. For this purpose, a number of geometrical metrics is employed, such as the geometric distance between the Own Ship (OS) and the k[th] Target Ship (TS_k) ($D_{i,k}$), time to the closest point of approach between the OS and TS_k ($TCPA_{i,k}$), distance at the closest point of approach between the OS and TS_k ($DCPA_{i,k}$) and the safety domain around the OS depicted using a circle with radius $a_{i,1}$. The hazardous scenarios identification procedure is depicted in the pseudocode form provided in Table 1.

The following criteria are employed: (a) $TCPA_{i,k} > 0$, to exclude scenarios where the closest encounter occurred in the past and the vessels are expected to diverge from each other; (b) $DCPA_{i,k} < a_{i,1}$, to identify scenarios where the two vessels will come very close to each other (TS_k in the safety domain of OS_k); (c) $D_{i,k} < d_s$, where d_s is a set threshold, to identify scenarios in which the two vessels are in proximity to each other; (d) $TCPA_{i,k} < t_s$, as the focus is on potential collision scenarios in the near future, depicted using t_s as threshold. The equations for all the employed metrics are

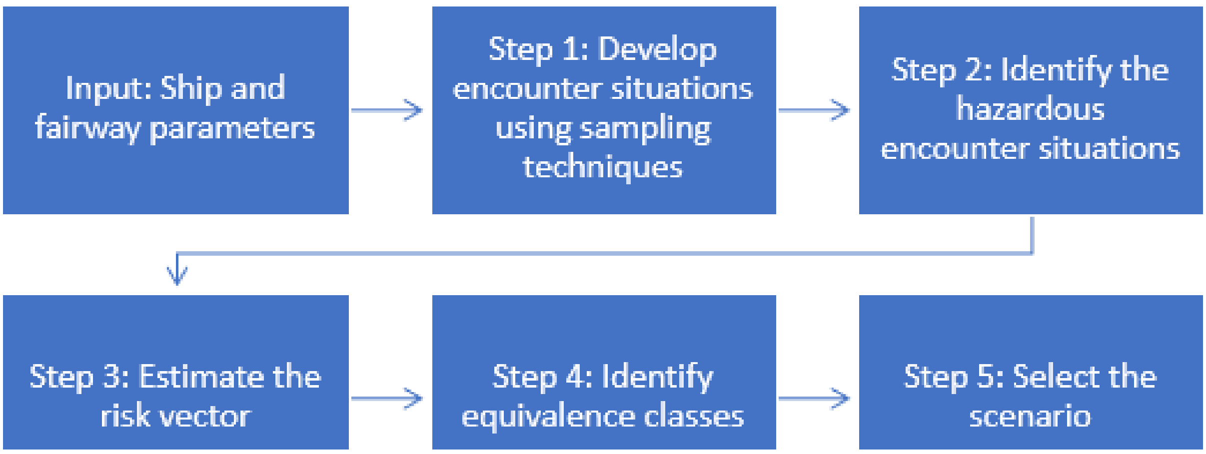


Figure 1 Process overview.

well-known and can be found in (Namgung and Kim, 2021; Pedersen et al., 2020; Woerner, 2014). The radius $a_{i,1}$ is calculated as in (Namgung and Kim, 2021). 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).

Table 1 Pseudocode used for hazardous scenarios identification

	Algorithm: Hazardous scenarios identification
1:	**Procedure:** Hazardous scenarios identification
2:	**Input:** Potential traffic situations 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 & $DCPA_{i,k}$ < $a_{i,1}$ & $D_{i,k} < d_s$ & $TCPA_{i,k} < t_s$ **then** Situation should be considered as hazardous Keep the parameters
8:	End for $D_{i,k}$
9:	End for i
10:	**End** procedure

At step 3, the risk metrics and vector for each scenario are estimated. The risk vector (RV) for each encounter scenario is calculated according to the following equation:

$$RV = [\sqrt{DCPAn_k\, TCPAn_k},\ \sqrt{A},\ h_d,\ W] \quad (1)$$

By setting the elements in a vector, an implicit assumption is introduced, according to which, all the vector elements are equally important.

The $DCPAn_k$ and $TCPAn_k$ represent the normalised versions of DCPA and TCPA for the TS_k vessel (or shore) based on the selected safety domains, and are estimated according to the following equations:

$$DCPAn_k = \begin{cases} 0, & if\ TCPA_k < 0\ or\ TCPA_k > t_s \\ 1, & if\ DCPA_k < a_1 \\ \frac{a_1}{DCPA_k}, & \text{for all other cases} \end{cases} \quad (2)$$

$$TCPAn_k = \begin{cases} 0, & if\ TCPA_k < 0\ or\ TCPA_k > t_s \\ \frac{t_s - TCPA_k}{t_s}, & \text{for all other cases} \end{cases} \quad (3)$$

The physical meaning of the equations (2−3) is as follows. If the closest point of approach with the vessel TS_k is in the past or too far away in the future, the vessel TS_k does not contribute to the risk (relevant value 0). Therefore, if $TCPA_k \rightarrow 0$ from positive values of TCPA, as we already excluded the scenarios with TCPA<0, then the TCPA contributes the most to the risk (relevant value 1). For the intermediate values, linear interpolation is used. If TS_k is in the safety domain, then the risk from DCPA becomes maximum (relevant value 1). The further the DCPA is from the ship, the smaller is its contribution to the risk of collision. The multiplication between $DCPAn_k$ and $TCPAn_k$ is used to emphasise that the closer an encounter is in distance and in time, the higher is the risk that is coming from the vessel 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 vessels, another metric, A, which depicts the area in the $u - \phi$ (speed-angle) space that is not allowed for manoeuvres. The metric A is calculated using concepts from velocity obstacle algorithms (Degre and Lefevre, 1981; Fiorini and Shiller, 1998) as following:

$$A = \frac{AC}{\pi\, V_{max}^2} \quad (4)$$

Where AC is the area in the $u - \phi$ space that is not available for safe navigation, as collision with vessels TS_k can occur according to the holonomic hypothesis. Therefore, AC is defined according to the following equation:

$$AC = \bigcup_{k=1}^{k=N} (u-\varphi)\ space\ Area\ \ where\ collision$$

$$with\ TS_k\ occurs \cap\ (u-\varphi)\ circle \qquad (5)$$

The parameter h_d is estimated as follows:

$$h_d = \begin{cases} \frac{d}{0.25\,V_{max}}, if\ speed\ is\ \text{within}\ AC\ area \\ 0, if\ outside \end{cases} \qquad (6)$$

where d is the minimum distance between $\overrightarrow{u(\phi)}$ and the safe area. This metric is used to depict how effectively the vessel can change speed and end up in a safe combination of speed and direction values. The normalisation with $0.25\,V_{max}$ is employed in line with the assumption that a ship cannot instantly change its speed to the desired level.

The last metric of RV is used to depict the weather conditions prevailing during manoeuvring and is estimated as follows:

$$W = \sqrt[3]{\frac{H_s\,V_c\,V_w}{H_{s,max}\,V_{c,max}\,V_{w,max}}} \qquad (7)$$

Where H_s is the wave height for the considered scenario, V_c is the current speed, V_w is wind speed, and max denotes the maximum values of these parameters (H_s, V_c and V_w).

In step 4, the identified hazardous scenarios from step 2 are grouped into equivalence classes using the risk vector from step 3 with the assistance of clustering techniques. For the purpose of this study, the mean shift clustering is used (Cheng, 1995). This algorithm estimates the extent of similarity between every pair of data using Gaussian kernels and Euclidean distance till the procedure converges according to a predefined bandwidth. In Gaussian kernels, the bandwidth is equivalent to the standard deviation. Code developed in (Finkston, 2021) is used for that purpose.

Once the equivalence classes have been identified, the sample closest to the mean RV value for each class is used as representative.

3. INVESTIGATED CASE STUDY AND SELECTED PARAMETERS

This study considered the case study of a small cargo vessel (OS) from the AUTOSHIP project (AUTOSHIP, 2019), which operating outside the coasts of Norway and is interacting with a sailing boat (TS1) and a high speed craft (TS2). The input parameters for the investigated situations are provided in Table 2. The random parameters with their ranges are provided in Table 3. These 18 parameters are assumed to vary from 0 to their maximum values and are sampled using the Sobol technique. The test area is set to [0 3 nm] x [0 3 nm] in line with (Namgung and Kim, 2021). A shore is also considered to be present in the study, represented by a simple spline line.

Table 2 Input parameters.

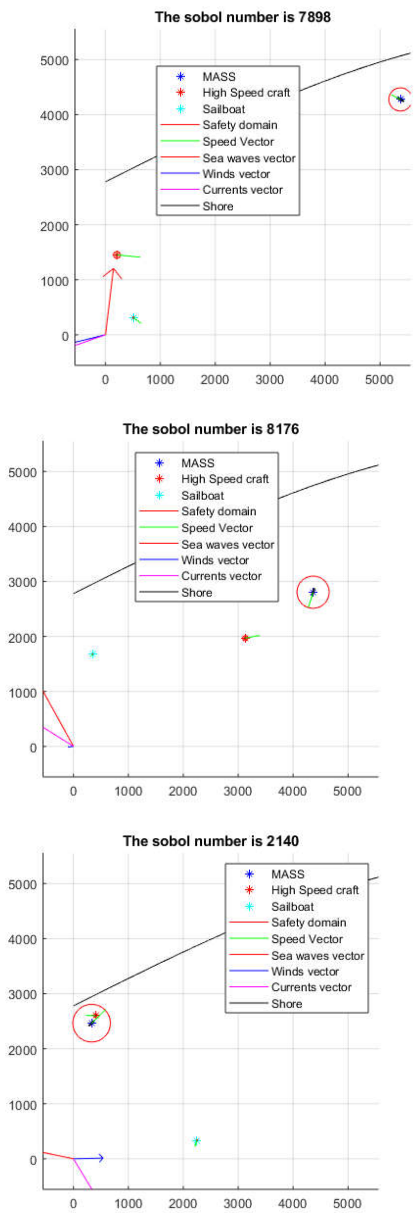
	Own ship SSS Cargo ship	Target ship No 1 (Sailboat)	Target ship No 2 (High speed craft)
Length	74.7 m	6 m	12 m
Beam	13.6 m	2 m	2.5 m
Max speed	15 kn	10 kn	40 kn
Max current	3 m/s		
Max waves height	2 m		
Max wind speed	14 kn		

Table 3 Random parameters.

Random parameters	Range
Fish feeding vessel speed	[0 max]
Fish feeding vessel speed direction	[0 2π] rad
Fish feeding vessel location	[0 3 nm] x [0 3 nm]
Sail boat speed	[0 max]
Sail boat speed direction	[0 2π] rad
Sail boat location	[0 3 nm] x [0 3 nm]
High speed craft speed	[0 max]
High speed craft direction	[0 2π] rad
High speed craft location	[0 3 nm] x [0 3 nm]
Current speed	[0 max]
Current direction	[0 2π] rad
Waves height	[0 max]
Waves direction	[0 2π] rad
Wind speed	[0 max]
Wind direction	[0 2π] rad

4. RESULTS AND DISCUSSION

The selected scenarios for N=10000 (number of Sobol samples) and b=0.5 (selected bandwidth in the clustering) are presented for each of the class/cluster shown in Figure 2. As it can be observed, the selected scenarios depict the following traffic conditions in proximity and away from the shore: (a) collision between OS and sailboat; (b) collision between OS and high speed craft; (c) collision between OS and both the vessels.



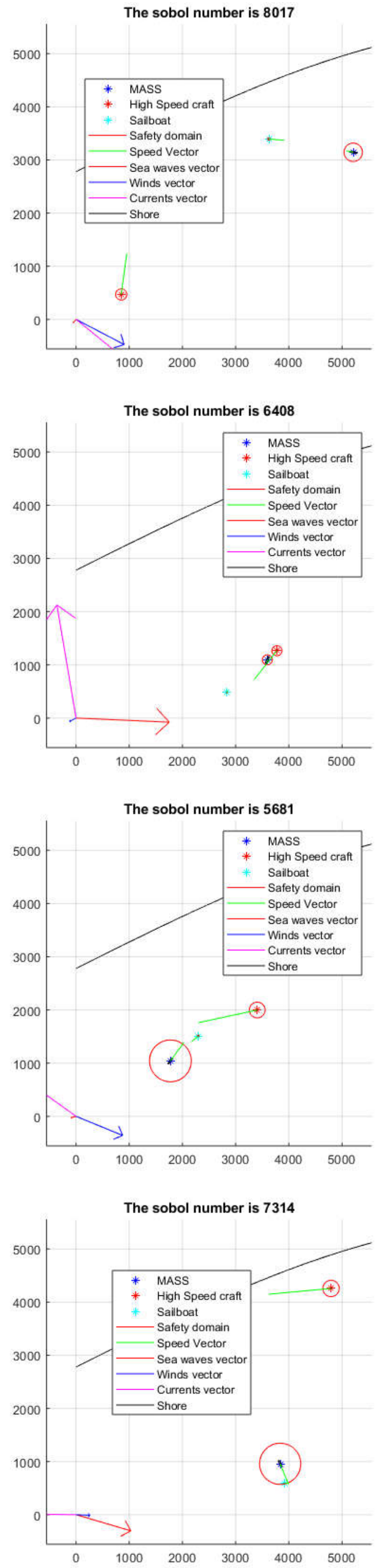


Figure 2 Representatives for each zone

The main advantage of the presented process is that it follows a deductive and not inductive thinking, compared to the traffic situation generated by AIS data. It starts by considering the complete set of the potential encounter conditions and progresses to more specific scenarios. Thus, it is more robust than inferring the potential encounter conditions based on the AIS data, as the scenarios generation proceeds from specific scenarios that already occurred in the past. This will contribute to the identification and testing of scenarios that have not been encountered before, but might occur in the future, belonging to the 'known unknown' region of Johari window(Luft and Ingham, 1961). The consideration of these type of scenarios during testing will contribute to the greater safety of collision avoidance system.

Another advantage of the developed process is that it can generate data for the vessels, for which AIS equipment is not required and therefore, no AIS data exists, e.g. sailboats and leisure high speed crafts. This is important, as these types of vessels constitute an considerable source of hazards for autonomous ships.

The critical items to be controlled in the analysis are the selected safety thresholds, such as d_s, t_s values and the safety domains around the vessel, as they influence the calculation of relevant risk metrics and therefore, the clusters. In the same manner, the selected risk metrics also influence the defined equivalence zones and clusters. When the risk vector varies, then the clustering results and the selected scenarios vary as well. A challenge is that currently there is a plethora of approaches for defining the risk metrics and safety domains without a common agreement or standard. Standardisation in this area is needed to define the COLREGS requirements for the autonomous ships as well as to promote the safe use of autonomous technology.

5. CONCLUSIONS

In this study, a process to develop encountering scenarios for testing the autonomous collision avoidance system of MASS was proposed and implemented for the case study of the SSS next generation autonomous ship of the AUTOSHIP project. The results demonstrated that this approach constitute an effective tool for identifying encounter conditions and replacing/substituting the AIS data.

Future research could focus on the selection of the appropriate safety domain, update the risk metrics as well as on the fundamental questions related to coverage scenarios and to the clustering algorithm convergence.

ACKNOWLEDGEMENTS

The study was carried out in the framework of the AUTOSHIP project , which is funded by the European Union's Horizon 2020 research and innovation programme under agreement No 815012. The authors also greatly acknowledge the funding from DNV AS and RCCL for the MSRC establishment and operation. The opinions expressed herein are those of the authors and should not be construed to reflect the views of EU, DNV AS, RCCL or other involved partners in the AUTOSHIP project.

REFERENCES

AUTOSHIP, 2019. Autonomous Shipping Initiative for European Waters.

Bolbot, V., Theotokatos, G., 2021. Automatically generating collision scenarios for testing ship collision avoidance system using sampling techniques, LEAC.

Bolbot, V., Theotokatos, G., Boulougouris, E., Vassalos, D., 2020. A novel cyber-risk assessment method for ship systems. Safety Science Under review.

Brcko, T., Androjna, A., Srše, J., Boć, R., 2021. Vessel Multi-Parametric Collision Avoidance Decision Model: Fuzzy

Approach. Journal of Marine Science and Engineering 9 (1), 49.

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

Cheng, Y., 1995. Mean shift, mode seeking, and clustering. IEEE transactions on pattern analysis and machine intelligence 17 (8), 790-799.

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

Degre, T., Lefevre, X., 1981. A collision avoidance system. The Journal of Navigation 34 (2), 294-302.

Finkston, B., 2021. Mean Shift Clustering.

Fiorini, P., Shiller, Z., 1998. Motion planning in dynamic environments using velocity obstacles. The International Journal of Robotics Research 17 (7), 760-772.

Gao, M., Shi, G.-Y., 2020. Ship collision avoidance anthropomorphic decision-making for structured learning based on AIS with Seq-CGAN. Ocean Engineering 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. Safety Science 92, 66-84.

Huang, Y., Chen, L., Chen, P., Negenborn, R.R., Van Gelder, P., 2020. Ship collision avoidance methods: State-of-the-art. Safety Science 121, 451-473.

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

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

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. Safety Science 129, 104798.

Luft, J., Ingham, H., 1961. The johari window. Human Relations Training News 5 (1), 6-7.

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

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.

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. Safety Science 129, 104799.

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

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.

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. Autonomous Robots 43 (4), 967-991.