

Model-Driven Scenario-Based Assurance for the COLREGS Compliance of Autonomous Surface Vehicles

Dominik Frey
Linköping University
Linköping, Sweden
dominik.frey@liu.se

Abstract

Autonomous surface vehicles (ASVs) must operate safely in open-sea maritime traffic while complying with the International Regulations for Preventing Collisions at Sea (COLREGS). However, the COLREGS rely on human judgment and contain ambiguities, making the compliance assurance of ASVs a challenging problem. Existing simulation-based testing approaches generate realistic scenarios but provide limited guarantees regarding the completeness of assurance suites and often struggle with scalable scenario generation, simulation execution, and compliance evaluation. This research proposes a model-based assurance methodology for ASVs that formally models open-sea encounters and systematically generates simulation-ready assurance suites that are relevant, functionally complete, diverse, and robust. We then execute these scenarios at scale and provide simulation-based evidence of COLREGS compliance through runtime monitoring.

ACM Reference Format:

Dominik Frey. 2026. Model-Driven Scenario-Based Assurance for the COLREGS Compliance of Autonomous Surface Vehicles. In *ACM/IEEE 29th International Conference on Model Driven Engineering Languages and Systems (MODELS Companion 2026)*, October 04–09, 2026, Málaga, Spain. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3837062.3839545>

1 Introduction

Significant efforts are directed towards developing *autonomous surface vehicles* (ASVs) for civilian and military applications. A critical requirement for ASVs in these contexts is the ability to complete missions without posing risks to other maritime traffic. To prevent collisions, the International Maritime Organization (IMO) has established the *International Regulations for Preventing Collisions at Sea* (COLREGS) [6], a set of 41 rules that guide vessel maneuvers in open-sea close-encounter situations to avoid collisions. However, since these regulations were designed with human operators in mind, many rules are subject to subjective interpretation and rely on human judgment. This presents a major challenge when it comes to ensuring that ASV control systems comply with the COLREGS. Moreover, in May 2026, IMO adopted a new *International Code of Safety for Maritime Autonomous Surface Ships* (MASS Code)¹, which highlights the need for assurance methods capable of systematically demonstrating the safety and COLREGS compliance of ASVs.

¹<https://www.imo.org/en/mediacentre/hottopics/pages/autonomous-shipping.aspx>



This work is licensed under a Creative Commons Attribution 4.0 International License.
MODELS Companion 2026, Málaga, Spain
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2903-4/2026/10
<https://doi.org/10.1145/3837062.3839545>

An *effective* assurance method requires *functionally complete* and *diverse* assurance suites [8] that cover all semantically distinct open-sea scenarios, each with uniformly *robust* attention. Moreover, the included scenarios must be *relevant*, i.e., meaningful and challenging from a COLREGS testing perspective [15].

Problem statement. Simulation-based testing is widely used to find safety violations in ASV systems operating in simulated open-sea encounters. Existing approaches rely on manually designed scenarios [1, 29], search- or sampling-based simulation [21, 24], or historical AIS traffic data [3, 20]. While these techniques can derive realistic test cases, they provide limited guarantees of the functional completeness of the resulting test suite. Consequently, they often fail to cover rare, safety-critical edge-case scenarios that are essential for assurance. Moreover, their simulation setups do not scale efficiently to large assurance suites and offer limited solutions for the oracle problem [11] for evaluating compliance when COLREGS become underspecified or ambiguous.

To address current gaps in safety and COLREGS assurance at open sea, we investigate the following research questions:

RQ1: How to formally capture open-sea encounters?

RQ2: How to automatically and efficiently generate simulation-ready assurance suites for ASVs that are relevant, functionally complete, diverse, and robust?

RQ3: How to obtain assurance evidence of the COLREGS compliance of an ASV?

Objectives. This research aims to provide a model-driven approach for generating effective assurance suites for ASV systems, executing them at scale in realistic simulators, and addressing the oracle problem of the COLREGS for compliance evaluation.

By doing this, we can provide simulation-based assurance evidence, which is a strong guarantee that an ASV system can safely complete its mission autonomously in maritime traffic.

2 Expected Contributions

To address the research questions, this work is expected to provide the following contributions to the fields of model-based engineering and maritime safety assurance:

C1: A domain-specific modeling method for maritime encounter scenarios to represent various types of vessels, static obstacles, environmental conditions, and COLREGS-specific relations between actors. The method provides abstractions for representing both COLREGS situations and high-level geometric relations.

C2: A formal technique relying on model-based abstractions to provide metrics for assessing the functional completeness, diversity, and relevance of assurance suites.

C3: A model-driven test generation approach that efficiently derives a functionally complete and diverse assurance suite of relevant scenes. The approach derives multi-vessel scenarios with static obstacles. It can detect geometrically infeasible scenes early to avoid wasting time on trying to generate such cases.

C4: A technique to derive variations of a relevant assurance scenario that preserve the same applicable COLREGS rules. This enables assessing the robustness of an ASV’s decision-making by evaluating whether it consistently identifies and executes the correct actions, even in scenarios close to semantic boundaries.

C5: An efficient approach to derive simulation-ready test scenarios with complete, COLREGS-compliant mission plans for non-ASV vessels. The approach supports dynamic path planning during simulation to keep the vessels that encounter the ASV COLREGS-compliant, even in the presence of unpredictable ASV behavior.

C6: A runtime monitor that identifies, classifies, and explains safety and COLREGS violations. The monitor considers standard safety metrics and implements temporal COLREGS awareness, tracking both maneuvers and COLREGS situations as they emerge, evolve, and disappear over time. The monitor outputs execution reports that provide assurance evidence of ASV compliance.

C7: A workflow to execute simulation-ready scenarios in advanced maritime simulators capable of modeling realistic physics and environmental conditions. The workflow supports faster-than-real-time, large-scale, and parallel execution of assurance suites. The workflow is agnostic to the ASV system under test and supports hybrid setups in which both real and simulated vessels participate.

C8: A modular end-to-end framework for automated and systematic ASV assurance. The framework integrates COLREGS modeling and assurance suite generation approaches with simulation execution and runtime monitoring for compliance assessment. It allows precise configuration of the scenario domain, including feasible ship dimensions, physics constraints, and regulatory parameters.

3 COLREGS Concepts

COLREGS scenarios. The COLREGS applies when two vessels are within visibility range and on a potential collision course at open sea. An encounter like this is classified as a COLREGS *situation* based on the relative bearing of the vessels. The COLREGS distinguish three situations: *overtaking* (Rule 13), *head-on* (Rule 14), and *crossing* (Rule 15), each assigning different maneuvering responsibilities where the typical roles involve *giving way* (i.e., executing an evasive maneuver) and *standing on course*. Multiple situations may coexist in a multi-vessel encounter, forming a COLREGS *scenario*. The right side of Figure 1 illustrates a four-vessel example adapted from [8], where the OS as the *own ship* (OS) is simultaneously in distinct COLREGS situations with *target ships* (TSs) ts1, ts2, and ts3. In this case, OS has two contradicting responsibilities (give-way and stand-on at the same time), resulting in an *ambiguous* scenario. These cases are particularly relevant from an ASV testing perspective, as resolving these hazardous cases involves complex reasoning [15].

In a *relevant* test scenario, the OS is in a COLREGS situation with each TS, while no COLREGS situations exist among the TSs themselves [15]. This assumption focuses testing on interactions involving the ASV and simplifies scenario construction.

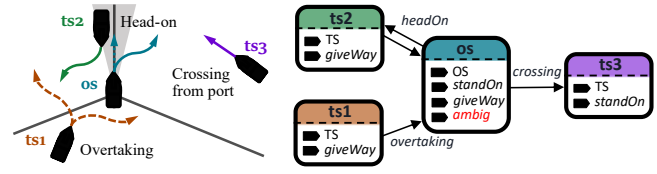


Figure 1: Example 4-vessel scenario adapted from [8].

COLREGS on a functional level. To represent maritime encounters at a higher level of abstraction, a metamodel was introduced in [15] where COLREGS situations are represented as abstract binary relations and roles are represented as labels. A *functional scenario* is an instance graph over this metamodel specified by first-order logic constraints. The right side of Figure 1 shows the functional representation of the example scenario. This representation allows automated reasoning over a COLREGS scenario, e.g., by inferring stand-on or give-way responsibilities and ambiguities. Moreover, functional scenarios capture the underlying semantics of a scenario instead of concrete positions and headings. We apply this to define *functional equivalence classes* (FECs) that are the unique graph shapes of functional models and represent semantically distinct scenarios. A functionally complete assurance suite should cover all finitely enumerable FECs on the functional level.

4 Related Work

Scenario-based assurance for ASVs. Scenario-based testing is widely used to evaluate ASVs in safety-critical maritime situations [1, 24, 29]. Existing approaches primarily rely on manually designed scenarios, historical AIS traffic data [18, 32], or stochastic scenario generation techniques [5, 8]. While these methods can generate realistic test cases, they are likely to miss rare edge case scenarios [8]. Hence, they provide limited guarantees regarding the semantic completeness of test suites that are relevant for assuring COLREGS compliance. In contrast, our work investigates systematic generation of assurance suites for strong guarantees driven by model-based abstractions of COLREGS scenarios.

Safety and COLREGS compliance assessment. Maritime safety research has proposed models for safety domains [25, 28], collision risk assessment, and encounter classification [13], but these methods do not perform complete COLREGS reasoning.

Existing COLREGS compliance approaches include decision trees [16], optimization-based methods [4, 33], and logic-based frameworks [17, 19]. However, they typically cover only subsets of the rules or are limited to pairwise encounters. More recent work extends compliance assessment to multi-vessel scenarios [11] or predicts vessel roles during encounters [12], yet no existing approach provides a complete compliance oracle for dynamic multi-vessel encounters while addressing the ambiguities in the COLREGS [15]. Our work aims to address this oracle problem by developing a comprehensive COLREGS compliance assessment framework.

Simulation-based frameworks for testing. Simulation is the standard approach for testing ASV safety and COLREGS compliance.

Maritime simulators [3, 22, 24] provide realistic execution environments but require users to manually construct scenarios, limiting scalability and reproducibility. More integrated approaches combine scenario generation, simulation, and COLREGS monitoring [20, 21], yet rely on simplified simulators or scenarios extracted from historical encounters, making them unsuitable for systematic scenario-based assurance [8]. Generally, these tools are not designed to provide complete assurance workflows. A model-driven framework [14] developed for autonomous road vehicles provides end-to-end integration, but does not adapt to the maritime domain.

We aim to provide an end-to-end framework that systematically generates functionally complete assurance suites, executes them in a realistic maritime simulator, and performs runtime COLREGS compliance monitoring in a scalable and reproducible workflow.

5 Approach

We apply model-driven techniques for providing scenario-based assurance. To achieve this, our approach realizes multi-level test scenario synthesis, in which FECs should be successively refined into *initial scenes* that describe the concrete placement, heading, and speed of each vessel at the start of the scenario. Then COLREGS-compliant *trajectories* for TSs should be derived, represented as a series of positions over time. Finally, the generated scenarios are executed in simulation to assess ASV behavior. Expanding on this agenda, Figure 2 presents the main stages of our research together with the corresponding research questions and contributions.

RQ1. We extend the metamodel proposed in [15] to represent more expressive functional models capturing heterogeneous vessel types, static obstacles, and geometric relations such as relative bearing, visibility, and collision courses (C1). The extended metamodel also formally captures scenarios where the COLREGS are underspecified [8], e.g., when two vessels within visibility are on a collision course but not in any of the COLREGS situation types.

To quantify the level of semantic guarantee of an assurance suite on the functional level, we define predicates for COLREGS relevance over FECs. We then enumerate all valid FECs with different graph shapes [27] using efficient graph solvers. The resulting set of relevant FECs serves as a functional-level baseline for functionally complete assurance suites (C2).

RQ2. Our model-driven assurance scenario generation approach consists of three refinement stages: (1) translating FECs into sets of geometric constraints referred to as *logical models*, (2) solving logical models as constraint satisfaction problems to obtain a robust set of concrete initial scenes, and (3) generating complete COLREGS-compliant trajectories for TSs.

First, each FEC is transformed into a logical model in which functional objects become variables and functional relations become geometric constraints. Second, search-based and sampling-based constraint solvers compute concrete solutions that instantiate the logical models (C3). Since not every FEC is geometrically feasible, we aim to investigate efficient constraint-solving strategies to identify infeasible models and minimize generation time.

To derive scenarios for robustness testing (C4), we will develop search heuristics that generate different concrete scene variants of the same FEC. To achieve this, we will refine the design space of

FECs into disjoint subspaces and apply the multi-step refinement approach to the refined models. During concretization, we plan to use metaheuristic search, initialized by the first scene variant. Furthermore, we will perform boundary analysis by traversing the boundaries of FEC solution spaces to evaluate an ASV’s COLREGS awareness in scenarios that are close to semantic transitions.

Finally, we investigate motion-planning approaches (C5) that generate COLREGS-compliant trajectories initialized from the generated scenes. Safety and COLREGS rules are monitored and enforced as hard constraints during generation. Ultimately, we aim to run the trajectory generation together with simulation execution. As a result of ASV navigation, static TS trajectories can become invalid. We plan to regenerate those trajectories from a calculated point in time to ensure that they continue to comply with the COLREGS.

RQ3. In a scenario execution context, we aim to develop a runtime monitoring framework (C6) that continuously reconstructs the global maritime scene from the published states of heterogeneous actors. The actors can be simulated or real, centralized or distributed, and can have different sensory capabilities.

The monitor considers standard safety metrics such as distance at *closest point of approach* (DCPA), and *time to closest point of approach* (TCPA), the *stages of risk*, and encounter type-specific *safety domains* [13]. Furthermore, it acts as an oracle adjudicating ASV COLREGS compliance over the whole time frame for each formalized rule using temporal logic [11, 19]. The monitor also accounts for the emergence and disappearance of COLREGS situations and updates the applicable set of rules accordingly.

For the simulation execution of assurance suites (C7), we identified the following practical requirements:

- *Fast, scalable, and reproducible*, enabling repeatable, faster-than-real-time execution of large assurance suites.
- *Realistic*, in terms of (1) physics by accurately capturing ASV behavior, environmental effects such as waves, wind, and restricted visibility, and (2) architecture by separating the physics world from the ASV control system.
- *ASV-system agnostic*, allowing straightforward integration of any external ASV control systems.
- *Supports hybrid simulation*, enabling experiments with both simulated and real vessels for field test compatibility.

To achieve fast and scalable simulations, we plan to execute them in a distributed fashion across multiple machines. We use *lockstep* for synchronizing simulation and control loops for deterministic behavior, which ensures reproducibility [26]. To achieve realism, we plan to use high-fidelity external simulators instead of a simplified in-house approach. This introduces additional complexity in terms of interfacing and scalability, which we must account for. To make our simulations ASV-system agnostic with hybrid setups, we integrate the simulated (or real) vessels through a common integration platform that provides standard interfaces for emitting control commands and status updates.

Finally, we will develop tooling (C8) that automates the complete assurance workflow, from functional scenario specification to simulation execution and evidence collection. Users specify assurance suites at the functional level together with configurable parameters of the assurance domain, such as vessel dimensions and COLREGS

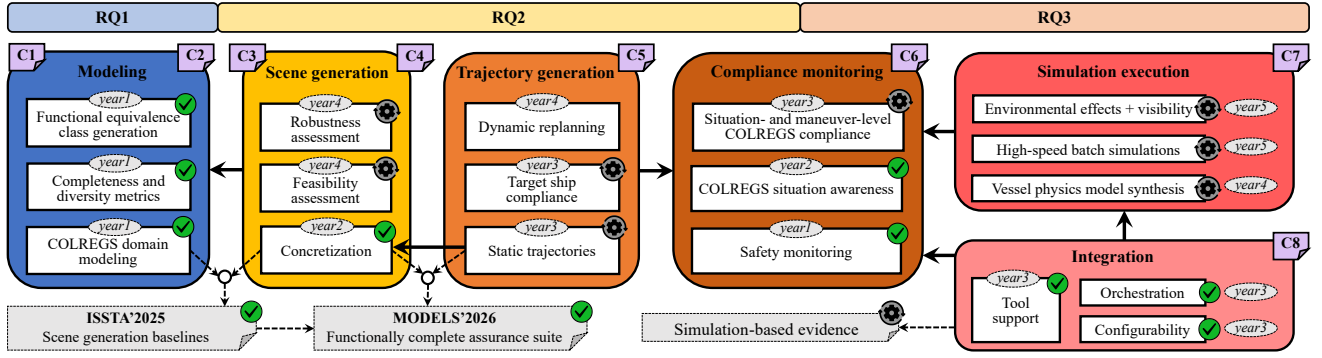


Figure 2: Overview of the main stages of our research, along with their motivating research questions and related contributions.

constraint parameters. The framework orchestrates scenario generation, simulation, and runtime monitoring, enabling automated extraction of assurance evidence.

Ethical considerations. As this research focuses on model-driven scenario generation, simulation, and assurance of ASV systems, no significant ethical concerns are anticipated.

6 Planned Evaluation

Investigating RQ1 and RQ2. We conduct an experimental study to assess the effectiveness of existing scenario generation techniques for testing COLREGS compliance. We derive hazardous test suites involving 2–6 vessels using sampling-based and search-based scenario generation adapted from the automotive domain, together with historical maritime encounter extraction [18]. These test suites are abstracted into FECs (C1) and evaluated using the functional-level metrics (C2) proposed for relevance, ambiguity, and completeness. This study demonstrates how our model-based abstractions can assess the assurance guarantees of a test suite.

We evaluate our model-driven scene generation approach (C3) in terms of generation efficiency and functional completeness. Then, we compare it with the approaches from the previous study that we use as baselines. To evaluate the effectiveness of our scene generation for robustness testing (C4), we measure the solution space coverage within each FEC.

We then evaluate the scalability and efficiency of extending initial scenes into complete TS trajectories (C5) that satisfy safety and COLREGS compliance constraints. Compliance is assessed using our runtime monitor (C6), which reports both quantitative measures of rule violations and qualitative explanations linked to the violated COLREGS rules. Existing path-planning approaches either rely on soft constraints or fail to produce COLREGS-compliant plans efficiently in multi-vessel encounters, motivating this evaluation.

For scene and trajectory generation approaches, each generation is evaluated in 30 independent runs with fixed random seeds to ensure statistical significance and reproducibility. We perform warm-up runs and invoke the garbage collector explicitly to improve measurement stability. We set a practical 4-minute timeout per generation to avoid getting stuck in a local optimum, which is a common issue when using stochastic approaches.

Investigating RQ3. We execute several open-source ASV control systems [30, 31] in simulation (C7) using the generated relevant and functionally complete assurance suites. We collect assurance evidence of safety and COLREGS compliance using the runtime monitor (C6). The proposed integration framework for end-to-end assurance (C8) serves as the evaluation pipeline. The tool will be demonstrated at the WARA-PS Demonstration Week² in September 2026, a major annual event in Sweden bringing together researchers and practitioners of AI, autonomous systems, and robotics.

7 Current Status

Figure 2 shows the current status and expected completion of each stage of the research plan. To date, we have addressed RQ1 and partially addressed RQ2 and RQ3.

We have developed a metamodel for *functional scenario models* (FSMs). While FECs are COLREGS specific representations, FSMs capture a richer set of functional properties of open-sea encounters, including various types of vessels, static obstacles, and high-level geometric relations such as relative bearing and visibility.

We generated the complete set of relevant FECs using the *Refinery* graph solver [23]. These FECs were then used to evaluate the relevance and functional completeness of test suites produced by existing approaches by abstracting and classifying each test case into one of the FECs. The work has been published at ISSTA’2025 [8]. We plan to extend these metrics to incorporate FSMs by year 4.

In our recent work accepted at MODELS’2026, we proposed *multi-step refinement* [7], an approach for generating functionally complete assurance suites [9, 10]. The work also presents how a simplified version of FSMs can be used in scene generation. Furthermore, we conducted a preliminary investigation regarding the robustness and the geometric feasibility of functional models on the generated artifacts. The completion of these investigations and the complete use of FSM for scene generation is planned for year 4.

We investigated several works on ASV path-planning and formal oracles for maritime safety and COLREGS compliance. Building on these works, we developed a prototype planner that generates static TS trajectories and even incorporates an incomplete COLREGS monitor. Fully compliant trajectory generation is planned in late year 3, while dynamic trajectory replanning during scenario execution is planned for year 4.

²<https://portal.waraps.org/page/event-demonstration-week-2026>

A prototype simulation framework has been developed that integrates scenario generation and monitoring with *Gazebo*³, a high-fidelity physics simulator. The framework is currently capable of running faster-than-real-time simulations with simple physics models and configurable environmental effects. Simulated or real vessels are integrated through the WARA-PS Core System [2], which serves as the common integration platform for heterogeneous autonomous *agents* deployed in public safety missions. The tool was conceptually validated through discussions with practitioners and presented as a poster at the WARA-PS Demonstration Week.

Future work will include running simulations automatically in batches, improving physics models, and finally, a thorough evaluation of open-source ASV control systems in our assurance framework. Completing the simulation-based assurance framework is planned for late *year 4* and *year 5*.

Acknowledgments

This work relates to the Department of Navy Award N62909-24-1-2006 issued by the Office of Naval Research. The work has been supported by the Linköping University Graduate School in Computer Science (CUGS) and the Wallenberg AI, Autonomous Systems, and Software Program (WASP), which is funded by the Knut and Alice Wallenberg Foundation.

References

- [1] Arnold Akkermann and Axel Hahn. 2020. Comparing Simulation with Physical Verification and Validation in a Maritime Test Field. *International Journal of Systems Engineering* 4, 2 (2020), 18. doi:10.11648/j.ijse.20200402.12
- [2] Olov Andersson, Patrick Doherty, Mårten Lager, Jens-Olof Lindh, Linnea Persson, Elin A. Topp, Jesper Tordenlid, and Bo Wahlberg. 2021. WARA-PS: a research arena for public safety demonstrations and autonomous collaborative rescue robotics experimentation. *Auton. Intell. Syst.* 1, 1 (2021). doi:10.1007/s43684-021-00009-9
- [3] Mustafa Bayrak and Haluk Bayram. 2023. COLREG-Compliant Simulation Environment for Verifying USV Motion Planning Algorithms. In *Proc. IEEE OCEANS*. IEEE, 1–10. doi:10.1109/oceanslimerick52467.2023.10244676
- [4] Hiba Ben Lahbib, M. Taha Bennani, and Jüri Vain. 2025. *Autonomous Vessels Collision Verification: Geometric Optimization*. Springer Nature Switzerland, 132–145. doi:10.1007/978-3-031-96239-4_10
- [5] Victor Bolbot, Christos Gkerekos, Gerasimos Theotokatos, and Evangelos Boulougouris. 2022. Automatic traffic scenarios generation for autonomous ships collision avoidance system testing. *Ocean Engineering* 254 (2022), 111309. doi:10.1016/j.oceaneng.2022.111309
- [6] UCG Commandant. 1999. International regulations for prevention of collisions at sea, 1972 (72 COLREGS). *US Department of Transportation, US Coast Guard, Commandant Instruction M 16672* (1999).
- [7] Dominik Frey, Ulf Kargén, and Dániel Varró. 2026. Automated Generation of Functionally Complete Assurance Suites for COLREGS-Compliance of Autonomous Surface Vehicles. In *Proc. ACM/IEEE MODELS*. Accepted for publication.
- [8] Dominik Frey, Ulf Kargén, and Dániel Varró. 2025. Assessing Scene Generation Techniques for Testing COLREGS-Compliance of Autonomous Surface Vehicles. *Proc. ACM Softw. Eng.* 2, ISSTA (2025), 985–1008. doi:10.1145/3728919
- [9] Dominik Frey, Ulf Kargén, and Dániel Varró. 2026. Measurement Data for Functionally Complete Assurance Suites for COLREGS-Compliance of Autonomous Surface Vehicles - MODELS26. doi:10.5281/ZENODO.20792733
- [10] Dominik Frey, Ulf Kargén, and Dániel Varró. 2026. Software Artifact for Functionally Complete Assurance Suites for COLREGS-Compliance of Autonomous Surface Vehicles - MODELS26. doi:10.5281/ZENODO.19246756
- [11] Jordan Gleeson, Matthew Dunbabin, and Jason J. Ford. 2024. COLREG Scenario classification and Compliance Evaluation with temporal and multi-vessel awareness for collision avoidance systems. *Ocean Engineering* 313 (2024), 119552. doi:10.1016/j.oceaneng.2024.119552
- [12] P.N. Hansen, D. Papageorgiou, M. Blanke, R. Galeazzi, M. Lützen, J. Mogensen, M. Bennedsen, and D. Hansen. 2020. COLREGS-based Situation Awareness for Marine Vessels - a Discrete Event Systems Approach. *IFAC-PapersOnLine* 53, 2 (2020), 14501–14508. doi:10.1016/j.ifacol.2020.12.1453
- [13] Yixiong He, Yi Jin, Liwen Huang, Yong Xiong, Pengfei Chen, and Junmin Mou. 2017. Quantitative analysis of COLREG rules and seamanship for autonomous collision avoidance at open sea. *Ocean Engineering* 140 (2017), 281–291. doi:10.1016/j.oceaneng.2017.05.029
- [14] Jerry Hou-Liu, Zhekai Jiang, and Aren A. Babikian. 2024. Concretize: A Model-Driven Tool for Scenario-Based Autonomous Vehicle Testing. In *Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems (MODELS Companion '24)*. ACM, 66–70. doi:10.1145/3652620.3687793
- [15] Ulf Kargén and Dániel Varró. 2024. Towards Automated Test Scenario Generation for Assuring COLREGS Compliance of Autonomous Surface Vehicles. In *Proc. ACM/IEEE MODELS*, Vol. 222. ACM, 249–256. doi:10.1145/3640310.3674098
- [16] Inchul Kim. 2023. Systematization of Legal Procedures for Collision Avoidance between a Fully Autonomous Ship and a Traditional Manned Ship. *Journal of Marine Science and Engineering* 11, 10 (2023), 1850. doi:10.3390/jmse11101850
- [17] Hanna Krasowski and Matthias Althoff. 2021. Temporal Logic Formalization of Marine Traffic Rules. In *Proc. IEEE ITSC*. IEEE, 186–192. doi:10.1109/iv48863.2021.9575685
- [18] Hanna Krasowski and Matthias Althoff. 2022. CommonOcean: Composable Benchmarks for Motion Planning on Oceans. In *Proc. IEEE ITSC*. IEEE, 1676–1682. doi:10.1109/itsc55140.2022.9921925
- [19] Hanna Krasowski and Matthias Althoff. 2024. Provable Traffic Rule Compliance in Safe Reinforcement Learning on the Open Sea. *IEEE Transactions on Intelligent Vehicles* (2024), 1–18. doi:10.1109/tiv.2024.3400597
- [20] Hanna Krasowski and Stefan Schäringer. 2025. CommonOcean-Sim: A Traffic Simulation Environment for Unmanned Surface Vessels. In *Proc. IEEE Int. Conf. Intell. Transp. Syst. (ITSC)*. IEEE, 966–972. doi:10.1109/itsc60802.2025.11423548
- [21] Eunkyun Lee, Junaid Khan, Umar Zaman, Jaebin Ku, Sanha Kim, and Kyungsup Kim. 2024. Synthetic Maritime Traffic Generation System for Performance Verification of Maritime Autonomous Surface Ships. *Applied Sciences* 14, 3 (Jan. 2024), 1176. doi:10.3390/app14031176
- [22] Ivan Loncar, Juraj Obradovic, Natko Krasevac, Luka Mandic, Igor Kvasic, Fausto Ferreira, Vladimir Slosic, Dula Nad, and Nikola Miskovic. 2022. MARUS - A Marine Robotics Simulator. In *Proc. IEEE OCEANS*. IEEE, 1–7. doi:10.1109/oceans47191.2022.9976969
- [23] Kristóf Marussy, Attila Ficsor, Oszkár Semeráth, and Dániel Varró. 2024. Refinery: Graph Solver as a Service: Refinement-based Generation and Analysis of Consistent Models. In *Proc. IEEE/ACM Int. Conf. Softw. Eng. (ICSE), Companion Proc. (ICSE-Companion '24, Vol. 21)*. ACM, 64–68. doi:10.1145/3639478.3640045
- [24] Tom Arne Pedersen, Jon Arne Glomsrud, Else-Line Ruud, Aleksander Simonsen, Jarle Sandri, and Bjørn-Olav Holtung Eriksen. 2020. Towards simulation-based verification of autonomous navigation systems. *Safety Science* 129 (2020), 104799. doi:10.1016/j.ssci.2020.104799
- [25] Zbigniew Pietrzykowski and Mirosław Wielgosz. 2021. Effective ship domain – Impact of ship size and speed. *Ocean Engineering* 219 (2021), 108423. doi:10.1016/j.oceaneng.2020.108423
- [26] Simon Sagmeister, Marcel Weinmann, Phillip Pitschi, and Markus Lienkamp. 2026. RSLCPP—Deterministic Simulations Using ROS 2. *IEEE Robot. Autom. Pract.* 1 (2026), 118–122. doi:10.1109/rap.2026.3704080
- [27] Oszkár Semeráth, Rebeka Farkas, Gábor Bergmann, and Dániel Varró. 2019. Diversity of graph models and graph generators in mutation testing. *Int. J. Softw. Tools Technol. Transf.* 22, 1 (2019), 57–78. doi:10.1007/s10009-019-00530-6
- [28] Rafal Szlapczynski and Joanna Szlapczynska. 2017. Review of ship safety domains: Models and applications. *Ocean Engineering* 145 (Nov. 2017), 277–289. doi:10.1016/j.oceaneng.2017.09.020
- [29] Anete Vagale, Ottar L. Osen, Andreas Brandsæter, Marius Tannum, Christian Hovden, and Robin T. Bye. 2022. On the use of maritime training simulators with humans in the loop for understanding and evaluating algorithms for autonomous vessels. *Journal of Physics: Conference Series* 2311, 1 (2022), 012026. doi:10.1088/1742-6596/2311/1/012026
- [30] Andrew Alexander Vekinis and Stavros Perantonis. 2023. Aeolus Ocean – A simulation environment for the autonomous COLREG-compliant navigation of Unmanned Surface Vehicles using Deep Reinforcement Learning and Maritime Object Detection. (2023). arXiv:2307.06688 [cs.RO]
- [31] Gang Wang, Jingheng Wang, Xiaoyuan Wang, Quanzheng Wang, Longfei Chen, Junyan Han, Bin Wang, and Kai Feng. 2024. Local Path Planning Method for Unmanned Ship Based on Encounter Situation Inference and COLREGS Constraints. *J. Mar. Sci. Eng.* 12, 5 (2024), 720. doi:10.3390/jmse12050720
- [32] Weiqiang Wang, Liwen Huang, Kezhong Liu, Yang Zhou, Zhitao Yuan, Xuri Xin, and Xiaolie Wu. 2024. Ship encounter scenario generation for collision avoidance algorithm testing based on AIS data. *Ocean Engineering* 291 (2024), 116436. doi:10.1016/j.oceaneng.2023.116436
- [33] Kyle Woerner, Michael R. Benjamin, Michael Novitzky, and John J. Leonard. 2018. Quantifying protocol evaluation for autonomous collision avoidance: Toward establishing COLREGS compliance metrics. *Autonomous Robots* 43, 4 (2018), 967–991. doi:10.1007/s10514-018-9765-y

³<https://gazebo.org/docs/harmonic>