Automated Scene Generation for Testing COLREGS-Compliance of Autonomous Surface Vehicles (ASVs)

Domink Frey, Ulf Kargén and Daniel Varró

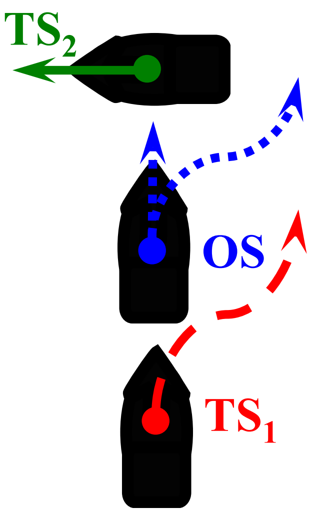
LINKÖPING UNIVERSITY

WASP | WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

Office of Naval Research Science & Technology

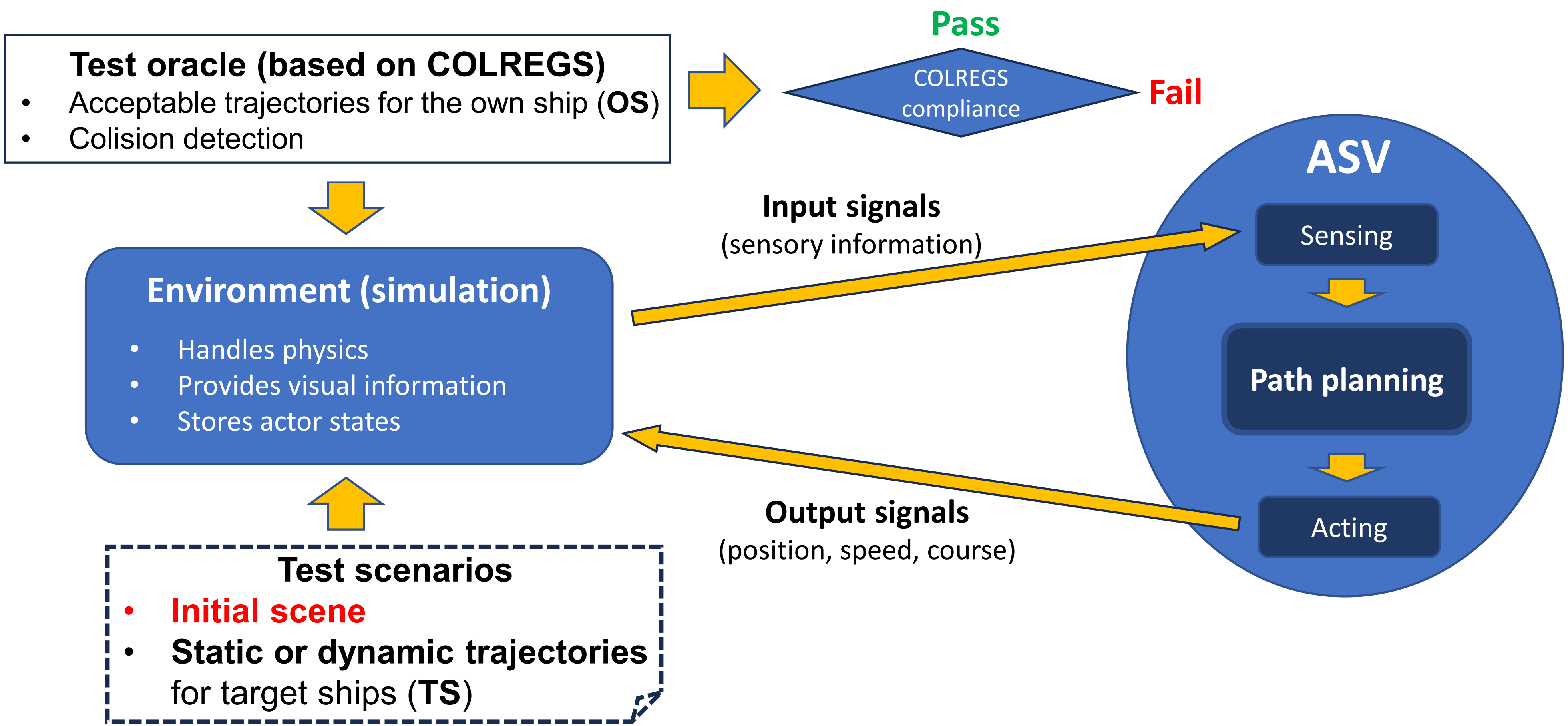
Motivation

- **Autonomous surface vehicles** (ASVs) increasingly important for both civilian and military applications
    - Needs to complete their mission autonomously in the presence of other maritime traffic
- **International Regulations for Preventing Collisions at Sea (COLREGs)** (by the International Maritime Organization)
- COLREGs compliance is critical for the safe operation of ASVs
- But COLREGS are …
    - **underspecified**
    - formulated with **human operators in mind**
    - **ambiguous** in case of multi-ship encounters



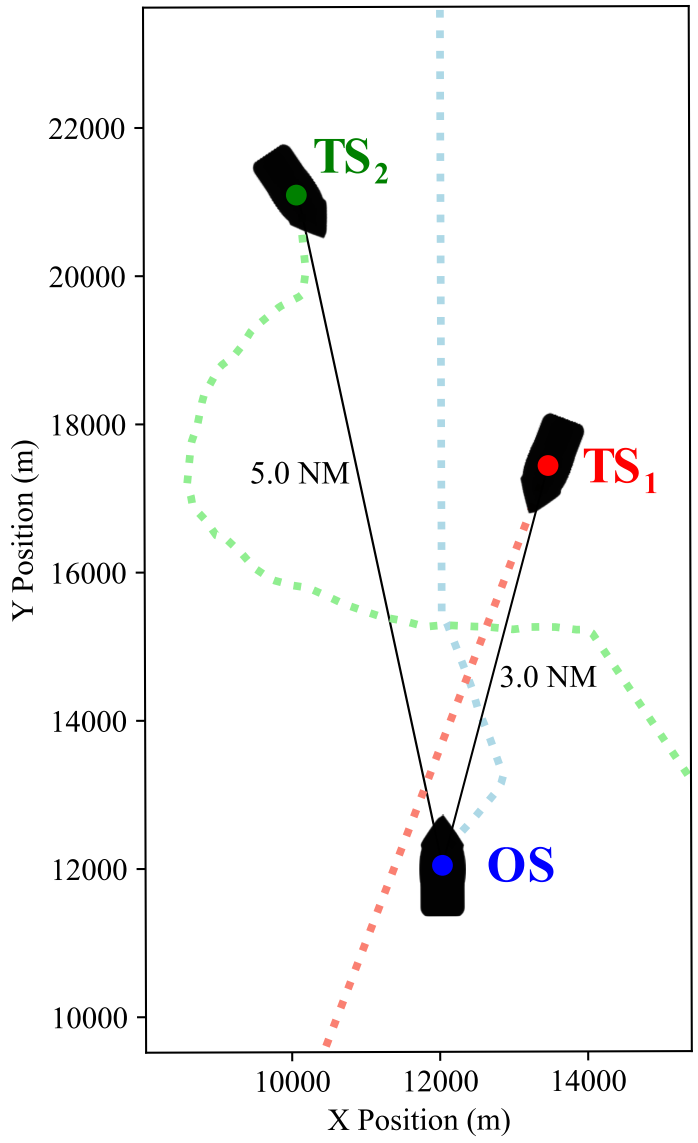
How to ensure safe behavior in such *rare critical scenarios*?

System-level testing of ASVs



Approach

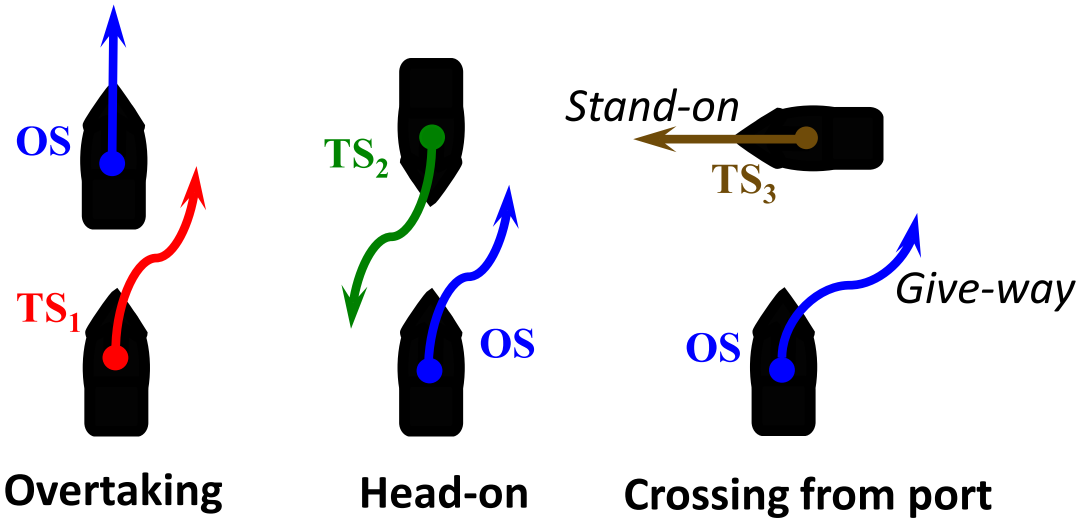
- Random/sampling-based test generation is likely to miss dangerous edge cases.
- Historical traffic data is likely to miss rare scenarios.
- Modelling COLREGS → systematically generating
  - **Diverse, dangerous, multi-vessel** (3–6) encounters in a **short time.**

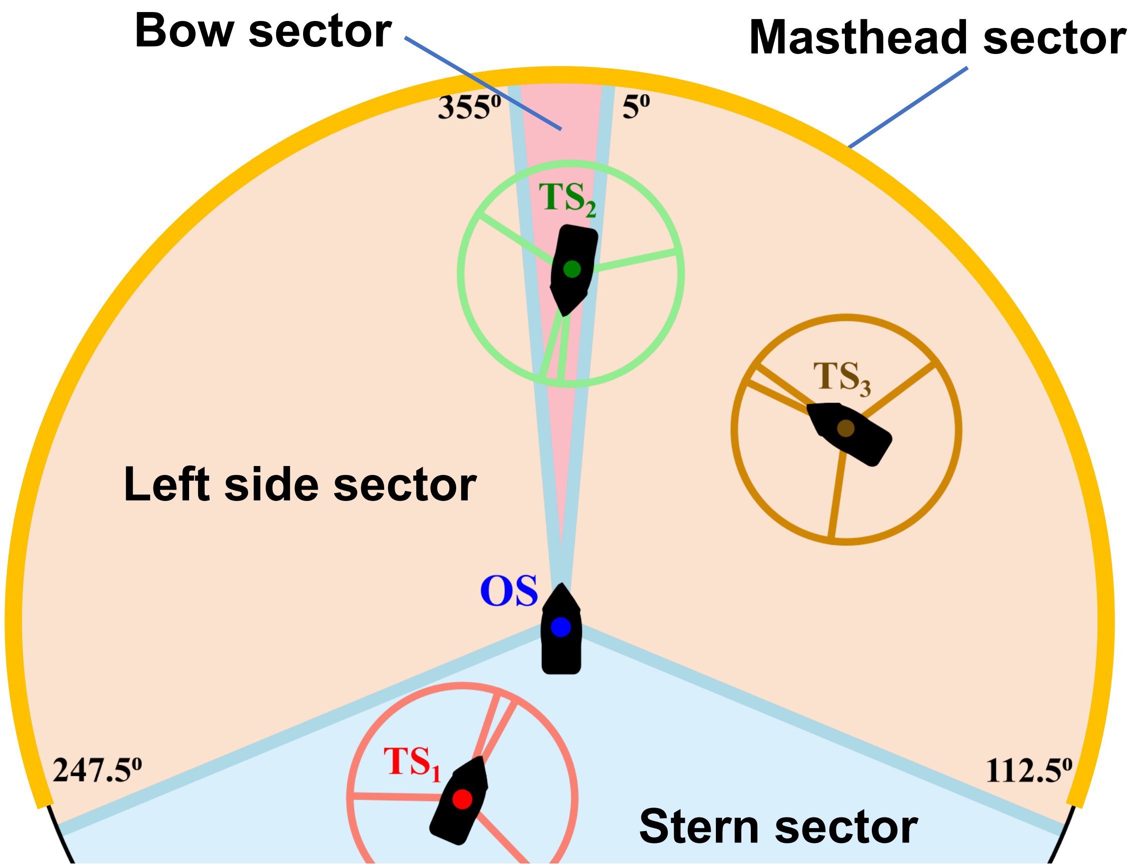


Dominik Frey, Ulf Kargén, Dániel Varró. 2024. Automated Scene Generation for Testing COLREGS-Compliance of Autonomous Surface Vehicles. ©

Background: COLREGS situations

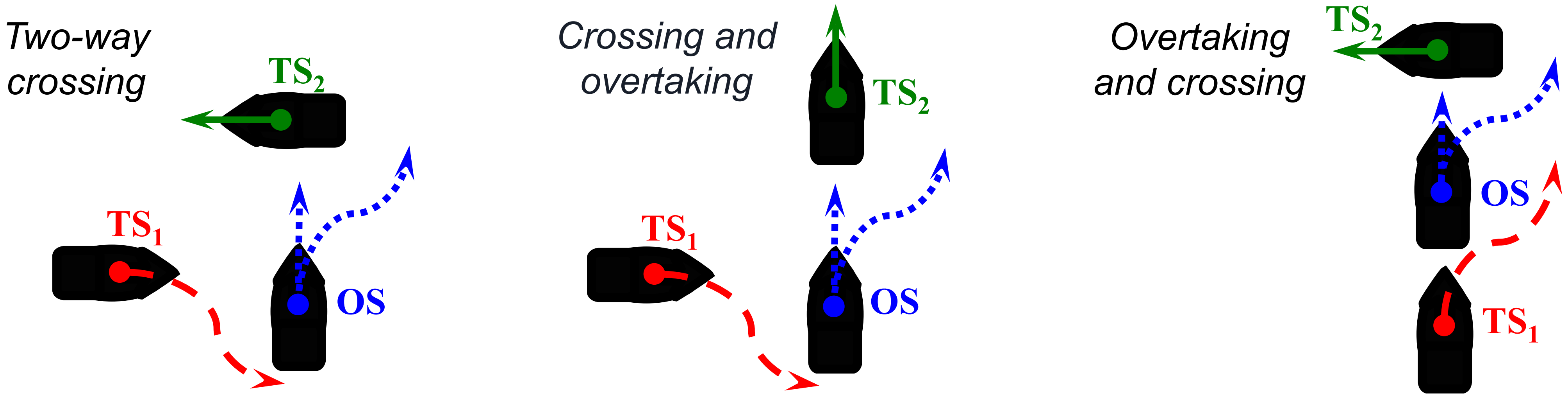
- COLREGS apply when
  1. collision without evasive action,
  2. in visibility distance,
  3. one of the following relative bearings:





Multi-vessel COLREGS scenarios

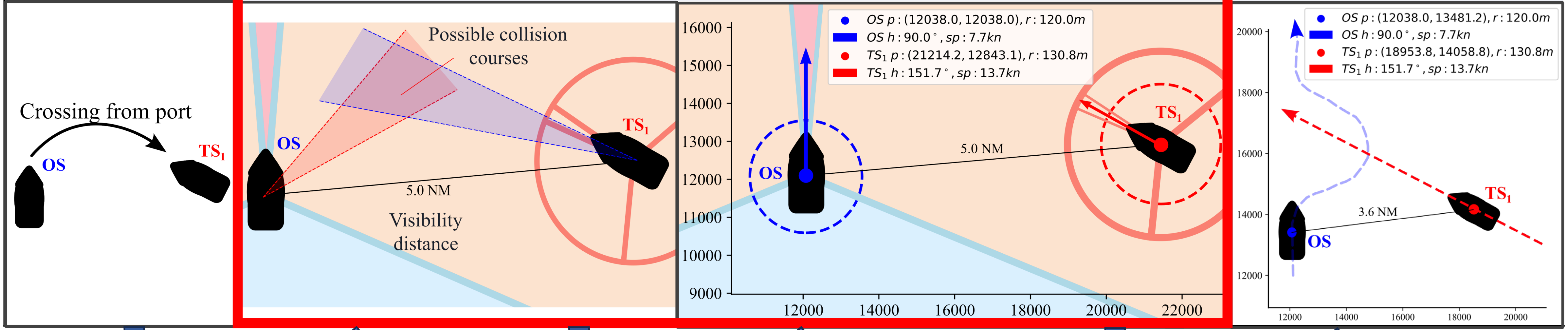
- **COLREGS scenario:** set of COLREGS situations
- **Ambigious situation:** Give-way + Stand-on



COLREGS scenarios on multiple abstraction levels

Menzel et al. 2018.

Functional scenarios	Logical scenarios (numerical constraints)	Concrete scenarios (Example crossing scenario)	
		Initial scenes	Trajectories



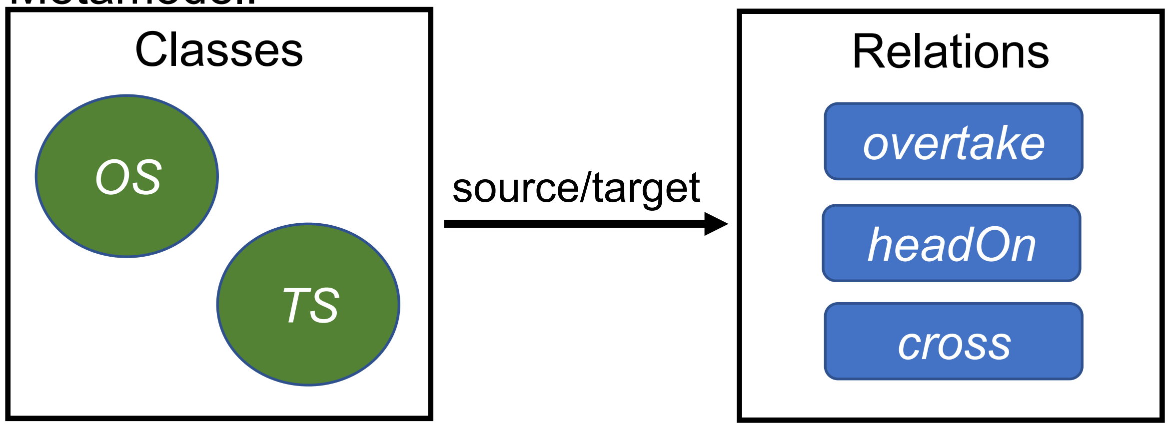
Mapping abstract relations to geometrical constraints.

Solving the constraints ⟼ Initial actor placements and velocities.

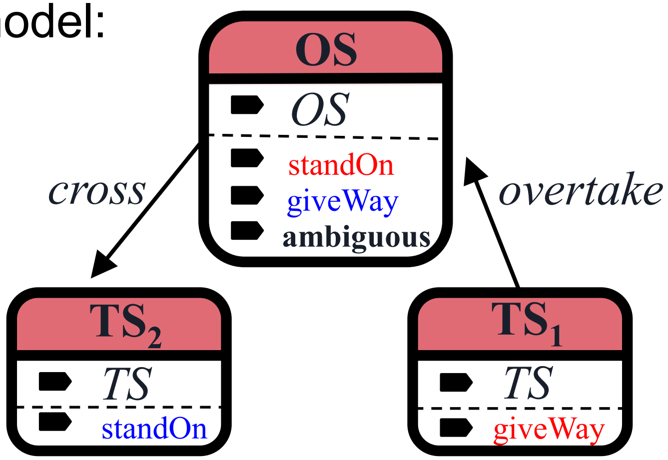
Ongoing future work

Functional scenarios and equivalence classes

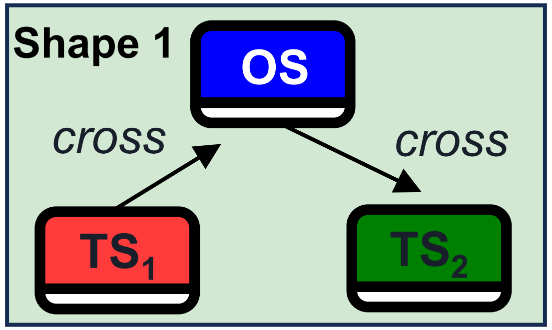
Metamodel:



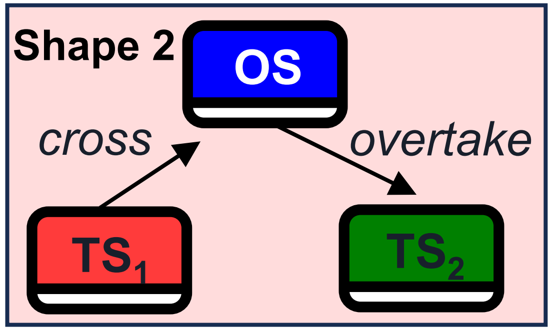
Instance model:



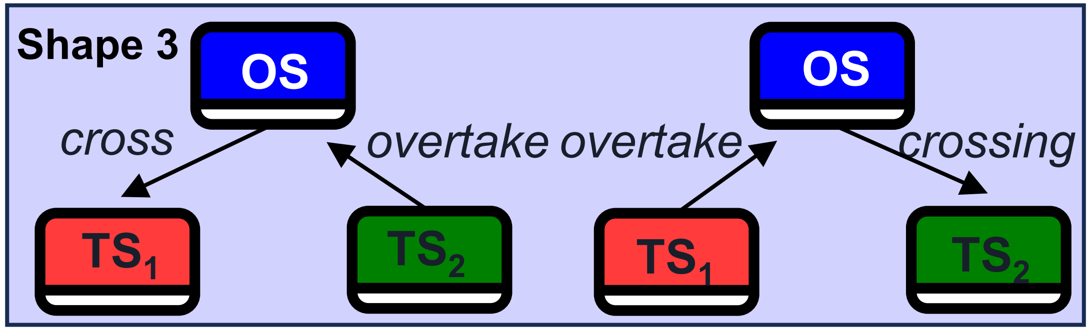
Two-way crossing



Crossing and overtaking



Overtaking and crossing

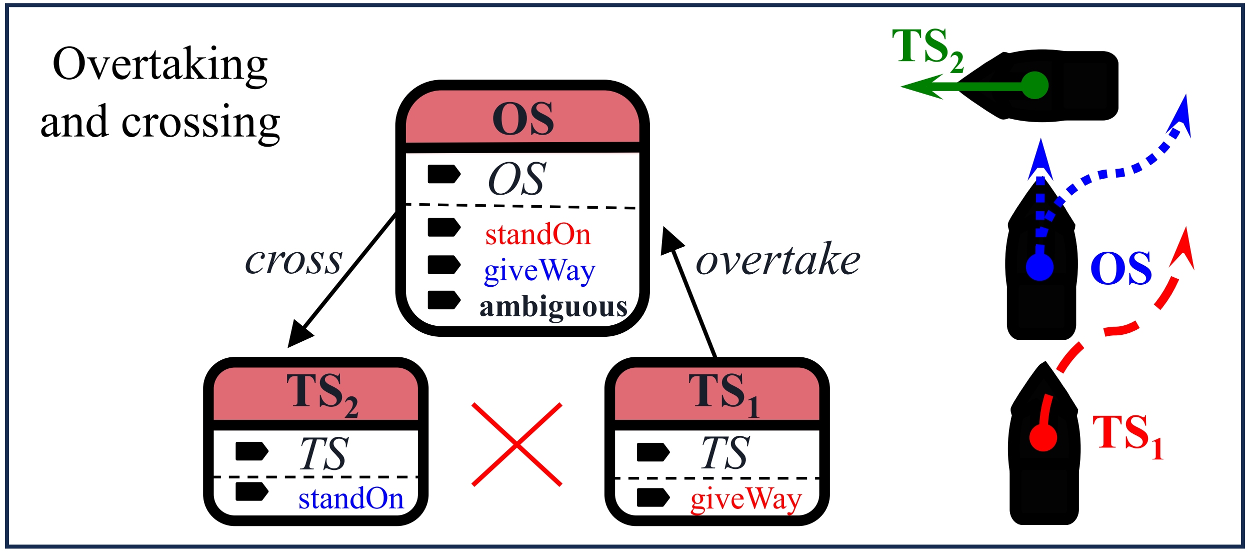


LINKÖPINGS UNIVERSITET

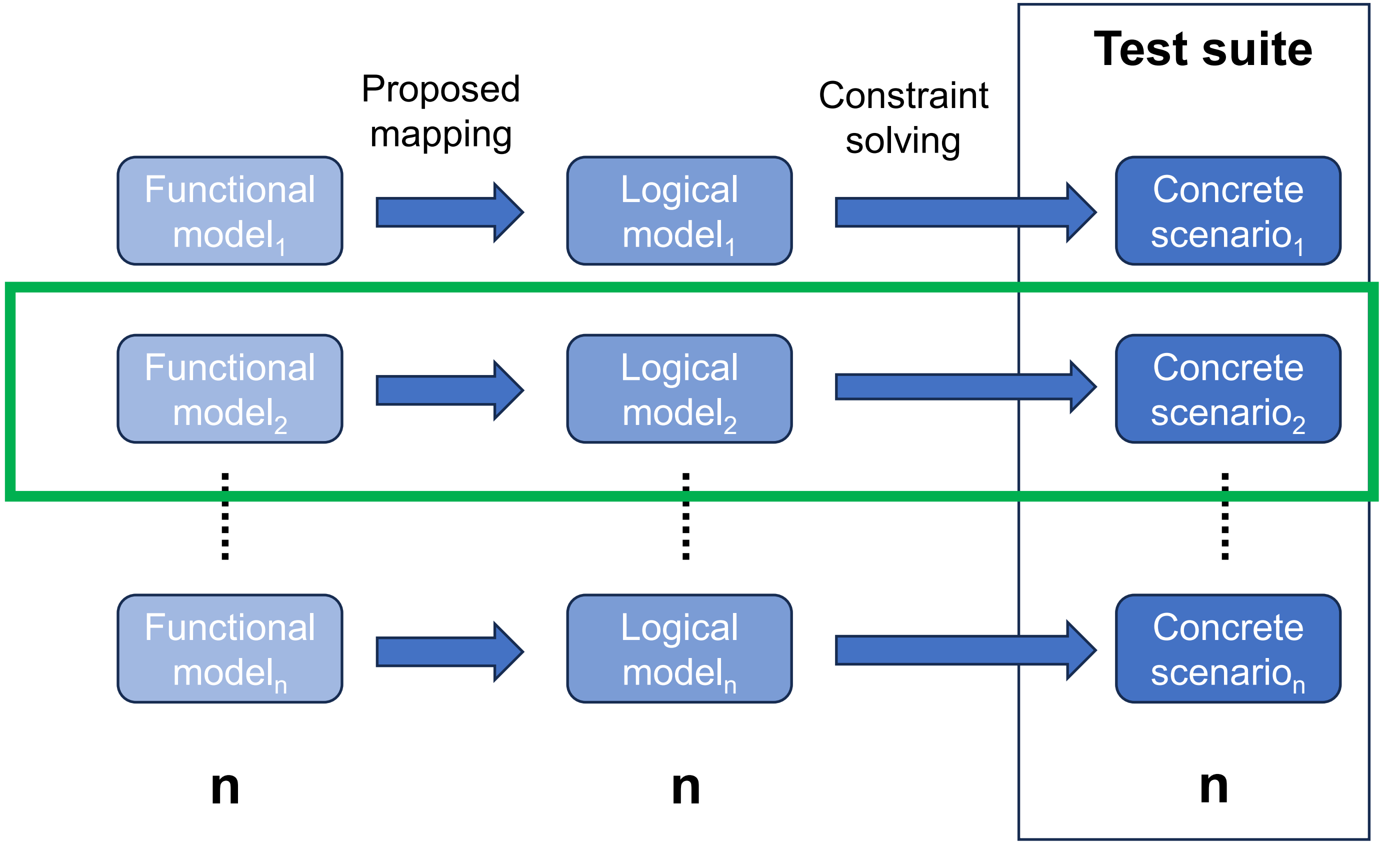
Dominik Frey, Ulf Kargén, Dániel Varró. 2024. Automated Scene Generation for Testing COLREGS-Compliance of Autonomous Surface Vehicles. ©

Functional constraints

1. There is exactly one own ship *(OS)*, while all other ships are target ships *(TS)*.
2. Each *OS*, *TS* pair must be in a COLREGS situation.
3. The *OS* must be in an **ambiguous** scenario.
4. Two *TS*s cannot be in a COLREGS situation.

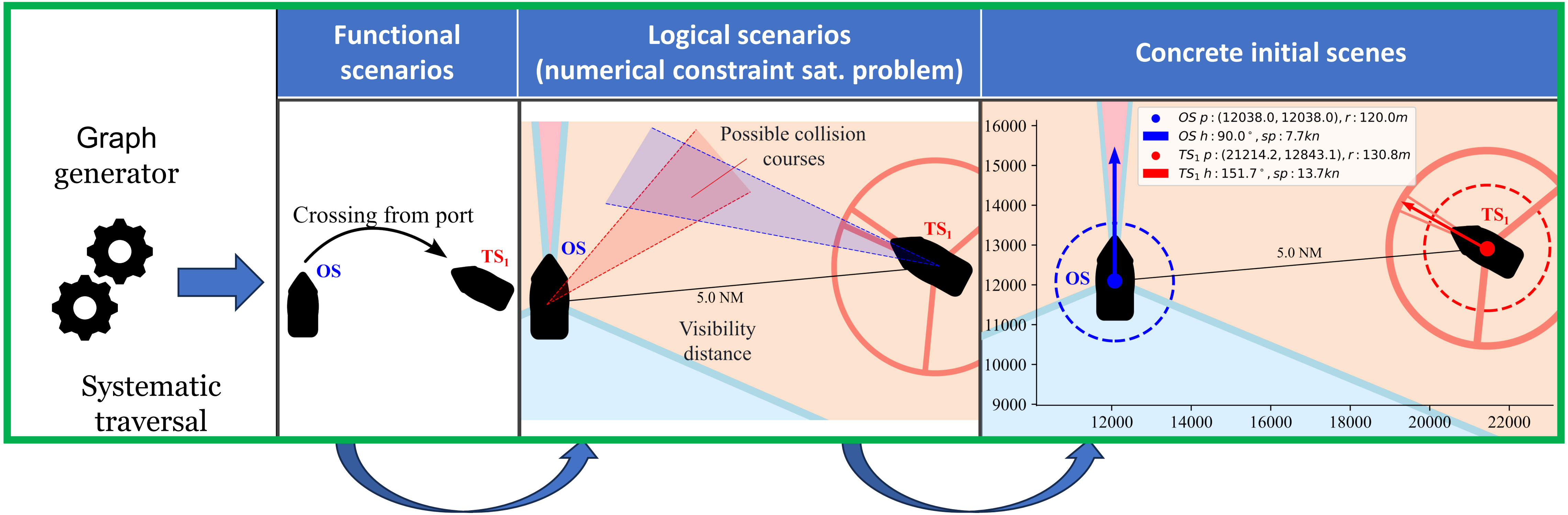


Multi-Step Refinement (MSR)



- Providing a scene **for each equivalence class** of functional scenarios.
- Constraints may be infeasible $\Longrightarrow$ wasting time when solving such cases.

Automated generation of initial scenes from functional scenarios using multi-step refinement

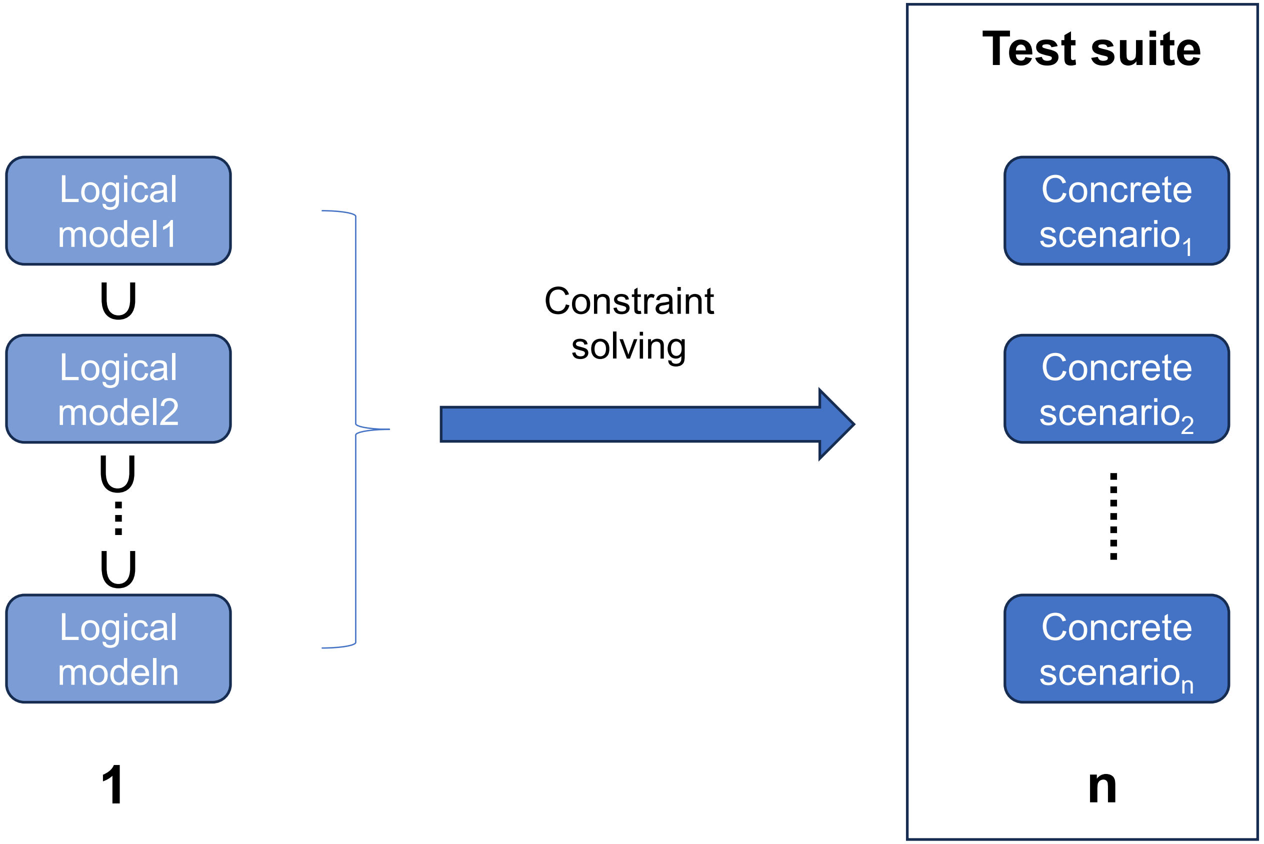


Mapping abstract relations to geometrical constraints.

1. Constraint satisfaction problem ⟼ optimization problem.
2. Using NSGAII to find an optimal solution.

Dominik Frey, Ulf Kargén, Dániel Varró. 2024. Automated Scene Generation for Testing COLREGS-Compliance of Autonomous Surface Vehicles. ©

Search-Based Only (SBO)



- Solving the **disjunction** of all logical models.
- Satisfiability is not an issue ⇒ faster runtime expected.
- **No completeness guarantees** for the coverage of semantic equivalence classes.

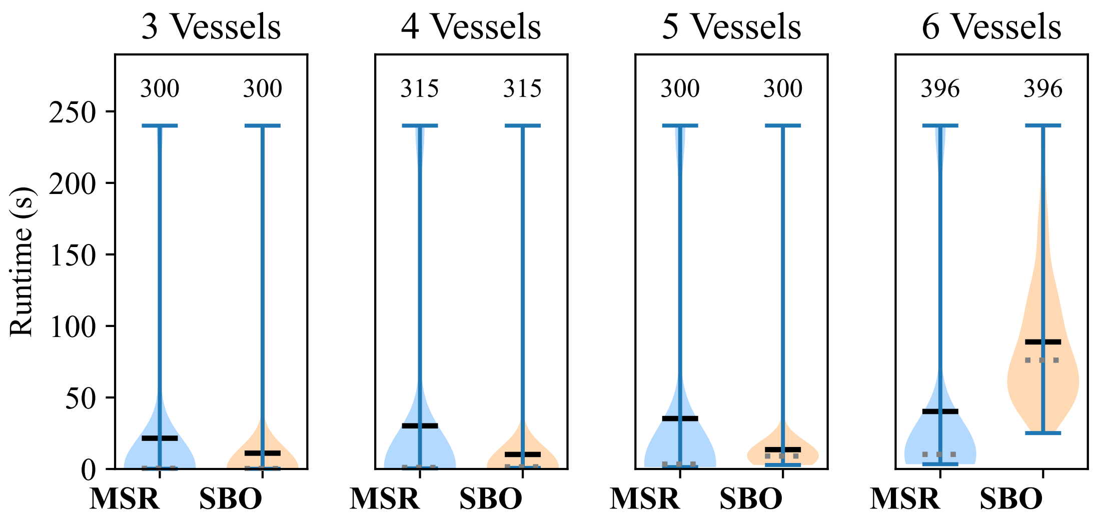
Empirical evaluation

How do the two approaches MSR and SBO perform wrt. (a) **runtime** and (b) **success rate**?

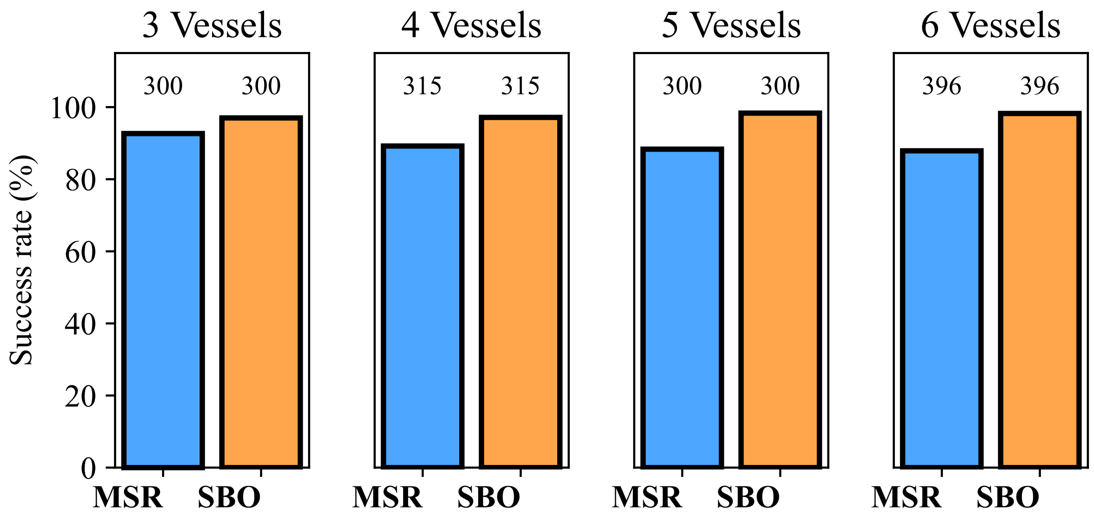
How do the two approaches perform wrt. **structural diversity** of test scenes?

How do the two approaches perform wrt. (a) runtime and (b) success rate?

- Testing of ASV behaviour requires a **large number of test cases** ⇒ scenario generation efficiency is important.
- No guarantees for convergence ⇒ we need to verify sufficient success rate.
- Setup:
  - >300 runs, for 3, 4, 5, and 6 participating vessels, with **MSR** and **SBO.**
  - 4-minute timeout.



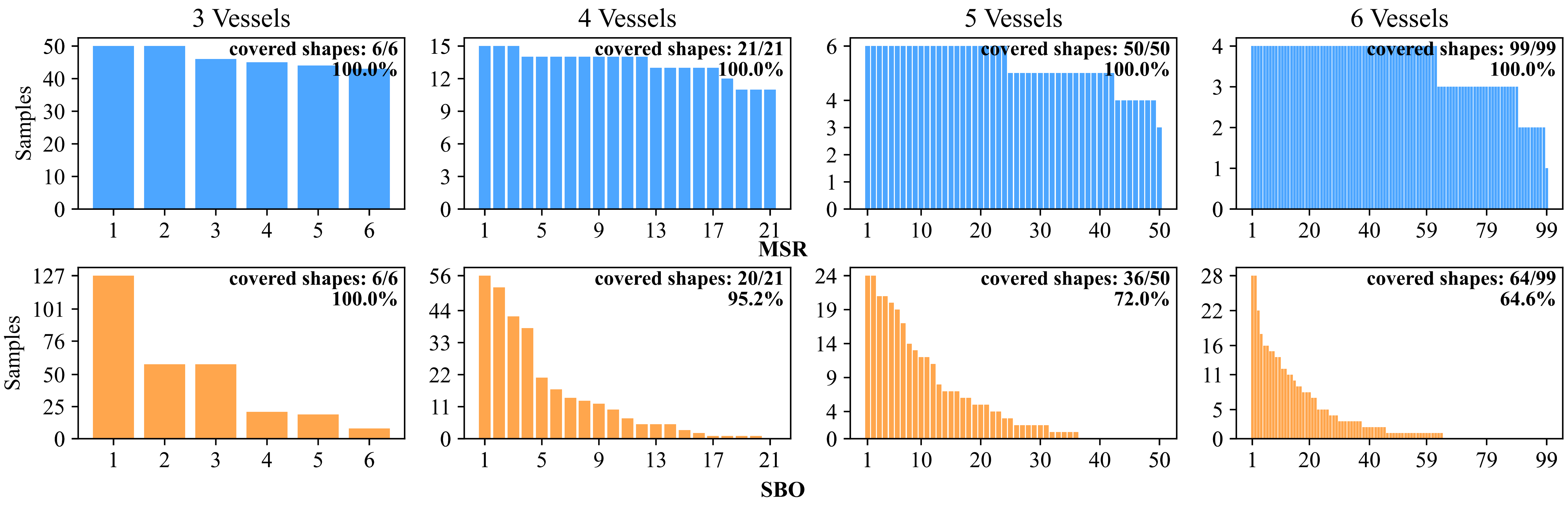
MSR is faster and scales well for larger problem sizes (K = 6) while SBO performs considerably better than MSR wrt. average runtime for K < 6.



Both approaches converge before the 4-minute timeout in no less than 93% of the time.

How do the two approaches perform wrt. structural diversity of test scenes?

- Assessing ASV behaviour in semantically diverse scenarios is essential for safety assurance.
- Setup:
  - Abstracting the concrete scenes back to the functional level.
- Structural coverage metric: $$\frac{|\textit{covered functional equivalence classes}|}{|\textit{total functional equivalence classes}|}$$



MSR is able to cover 100% of the total equivalence classes across all problem sizes with a more even distribution of samples along classes. However, SBO shows a linear decrease in its coverage with uneven distributions. Both approaches reach 100% coverage in 3 vessel scenarios.

Conclusion

Two novel approaches for generating challenging scenes to test ASV COLREGS compliance.

MSR: Multi-step scenario refinement – functional $\longmapsto$ logical $\longmapsto$ concrete scenes.

- 100% equivalence class coverage, near-linear scalability for problems of 3–6 vessels.

SBO: Concrete scenes directly from single joint numerical problem.

- Shorter runtimes on smaller problems BUT worse scalability and coverage.

Future work

Extending metamodel complexity: Diverse ship types, such as limited-maneuverability vessels.

Extending environment complexity: Static objects and complex maps (e.g., ports, coastal areas).

Trajectory Generation: Different methods for trajectory generation, dynamic trajectories.