CHALMERS
UNIVERSITY OF TECHNOLOGY

Data-driven Development Framework for ADAS and Automation for Marine Applications

From On-Vessel Data Logging to KPI-Based Verification: An End-to-End Framework for Marine ADAS/AD Development

Master's thesis in Systems, Control and Mechatronics

Adithyaa Ramesh | Prasanth Balaji Pollachi Malaiyalaswamy

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025
www.chalmers.se

Master's thesis 2025

Data-driven Development Framework for ADAS and Automation for Marine Applications

From On-Vessel Data Logging to KPI-Based Verification: An End-to-End Framework for Marine ADAS/AD Development

Adithyaa Ramesh | Prasanth Balaji Pollachi Malaiyalaswamy

Automation
Department of Electrical Engineering
Motion Control, Volvo Penta
Chalmers University of Technology
Gothenburg, Sweden 2025

Data-driven Development Framework for ADAS and Automation for Marine Application
From On-Vessel Data Logging to KPI-Based Verification: An End-to-End Framework for Marine ADAS/AD Development
Ramesh, Adithyaa and Pollachi Malaiyalaswamy, Prasanth Balaji

Supervisor: Ricardo Graef Papritz, Volvo Penta
Examiner: Martin Fabian, Electrical Engineering

Master's Thesis 2025
Department of Electrical Engineering
Motion Control, Volvo Penta
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2025

Data-driven Development Framework for ADAS and Automation for Marine Applications
From On-Vessel Data Logging to KPI-Based Verification: An End-to-End Framework for Marine ADAS/AD Development
Adithyaa Ramesh
Prasanth Balaji Pollachi Malaiyalaswamy
Department of Electrical Engineering
Chalmers University of Technology

Abstract

The maritime industry is increasingly adopting Advanced Driver Assistance Systems (ADAS) and automation, drawing inspiration from progress in the automotive sector. Applying these technologies to marine environments, however, introduces unique challenges such as sensor limitations, unpredictable conditions, and the lack of standardized validation methods.

This thesis presents a data-driven framework to support the testing and validation of marine ADAS/AD (Autonomous Driving) systems. The framework focuses on data logging, structured data handling, and the use of Key Performance Indicators (KPIs) as objective measures of performance. In collaboration with Volvo Penta and Volvo GTT (Group Trucks Technology), a proof of concept was developed around two representative features, with relevant KPIs defined. The framework was partially implemented on a test vessel equipped with LiDAR (Light Detection and Ranging) and cameras for perception, a Dynamic Positioning System (DPS) for positioning, and a high-bandwidth logger to capture raw sensor data during real operations.

While post-test data ingestion, KPI calculation, and KPI-driven refinement were not completed within the scope of this thesis, these stages are outlined as future extensions. The work provides a foundation for a systematic, data-driven development methodology in the marine ADAS/AD domain, bridging the gap between conceptual design and a fully operational validation pipeline.

Keywords: Data-driven, Marine ADAS/AD, Key Performance Indicators, Data Pipeline, Data collection, Re-simulation, Docking assistance, Verification & Validation

Acknowledgements

Embarking on this thesis has been a journey of discovery, collaboration, and growth. We are deeply grateful to all those who guided and supported us along the way.

First, our sincere thanks to **Ricardo Graef Papritz**, our **supervisor** at Volvo Penta. As the originator of this thesis topic, his vision, deep marine expertise, and ability to connect us with the right people across departments made this work possible. We wish him the very best of luck for the future.

We also owe a special debt of gratitude to **Martin Fabian**, our **examiner** at Chalmers University of Technology, for his steady guidance, thoughtful feedback, and for consistently steering us in the right direction. Good luck in the future.

We want to thank **Amar Shati**, **Therese Gardell**, **Anders Jellbin**, and **Binay Mishra** from Volvo GTT for their invaluable support in shaping and implementing the data pipeline. Their input on KPI definition, validation strategy, and many incisive discussions greatly strengthened the verification and validation aspects of this work. Their generosity in providing access to Volvo GTT's Data Science Laboratory (DSL) and server space allowed us to fast-track our progress. Finally, we are grateful for their hands-on help in setting up the logging equipment and for patiently responding to what must have seemed like an endless stream of messages at all hours. We wish them all the best for the future.

Heartfelt thanks to **Jonatan Bergenwall**, CPAC Systems AB, for his help on the test boat, equipping the logger, running dry runs, and making sure we captured the data we needed. Good luck in the future.

We would also like to thank **Anders Ivarsson** and the rest of the Motion Control team. Your Friday fikas and camaraderie were always a welcome break, and your thought-provoking questions helped us sharpen the focus of this thesis. Good luck in the future.

Finally, thank you to everyone at **Volvo Penta**, **Volvo GTT**, and **CPAC Systems AB** who generously shared their time and expertise. Your support transformed this project from a concept into what it is. Good luck in the future.

Adithyaa Ramesh and Prasanth Balaji Pollachi Malaiyalaswamy, Gothenburg, September 2025

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis, listed in alphabetical order:

AD	Autonomous Driving
ADAS	Advanced Driver Assistance Systems
AIS	Automatic Identification System
AVI	Audio Video Interleave
CAN	Controller Area Network
COLREGs	Convention on the International Regulations for Preventing Collisions at Sea, 1972
DSL	Data Science Laboratory
DPS	Dynamic Positioning Systems
EVC	Electronic Vessel Control
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GTT	Group Trucks Technology
IMU	Inertial Measurement Unit
INS	Inertial Navigation System
IP	Internet Protocol
KPI	Key Performance Indicator
LiDAR	Light Detection and Ranging
MASS	Maritime Autonomous Surface Ship
MF4	Measurement File Version 4
ODD	Operational Design Domain
OE	Operational Envelope
PCAP	Packet Capture
PPS	Pulse Per Second
PTP	Precision Time Protocol, IEEE 1588
RADAR	Radio Detection and Ranging
SONAR	Sound Navigation and Ranging
SOTIF	Safety of the Intended Functionality
UDP	User Datagram Protocol
VAS	Volvo Active Safety
V&V	Verification and Validation

Contents

List of Acronyms ix

List of Figures xiii

List of Tables xvii

1 Introduction 1
- 1.1 Background 1
  - 1.1.1 Volvo Penta 2
  - 1.1.2 Aims and Objectives 2
    - 1.1.2.1 Aim 2
    - 1.1.2.2 Objectives 2
  - 1.1.3 Research Questions 3
- 1.2 Delimitation 3
- 1.3 Ethical and Sustainability Aspects 4
- 1.4 Usage of generative Artificial Intelligence 5

2 Background 7
- 2.1 Equipment 8
  - 2.1.1 Onboard Computer 8
  - 2.1.2 Mechanical LiDAR 8
  - 2.1.3 Solid-State LiDAR 9

3 Methodology 11
- 3.1 State of the Art 11
- 3.2 Feature Investigation 12
- 3.3 KPI Formulation 12
- 3.4 Developing Scenarios 13
- 3.5 Equipment Research 13
- 3.6 Data collection 14
  - 3.6.1 Dry Run 14
  - 3.6.2 On-Site Testing 14
- 3.7 Data Handling 14

4 State of the Art 15
- 4.1 Development Frameworks in Marine domains 15
- 4.2 Autonomy Level Classifications 16

4.3 Data Logging and Pipelines 18
4.3.1 Marine Domain 18
4.3.2 Automotive Domain 18
4.3.3 Cross-Domain Takeaways for This Thesis 18
4.4 KPI for Validation and Performance Measurement 19
4.4.1 Automotive KPIs 19
4.4.2 Marine KPIs 19
4.5 V&V Frameworks for ADAS/AD 20
4.5.1 Automotive V&V Frameworks 20
4.5.2 Maritime V&V Frameworks 21
4.6 Volvo Group Toolchains (GTT/VAS) 22
4.6.1 ADAS Verification & Validation at GTT 22
4.6.2 The Data Science Lab 22
4.7 Conclusion and Outlook 23

5 Key Performance Indicators 25
5.1 Feature Determination 25
5.2 Defining KPIs 25
5.3 Refinement and Prioritization 26
5.4 Implementation Considerations 27
5.5 Selected KPIs 27
5.5.1 Surround Sense 27
5.5.1.1 False Positives for Detection of Obstacles 27
5.5.1.2 False Negatives for Detection of Obstacles 28
5.5.1.3 Object Tracking Consistency 28
5.5.1.4 Docking Spot Determination Rate 29
5.5.2 Docking Assist 30
5.5.2.1 Final Orientation of the Vessel 30
5.5.2.2 Speed of Approach to the Dock 31
5.5.2.3 Collision Speed to an Obstacle 31
5.5.2.4 Total Docking Time 32
5.6 Scenarios 32
5.7 Dataset Determination 33

6 Data collection 35
6.1 Data Logging Equipment 35
6.1.1 Logger 36
6.1.2 Capture Modules 36
6.1.3 Ethernet switch 37
6.1.4 Media Convertors 38
6.1.5 Cameras 38
6.1.6 Trigger/Display/Dashboard 38
6.2 Equipping the Boat 38
6.3 Data Collection 40

7 Data Handling 41
7.1 Data Ingestion 41

7.2 Data Preprocessing and KPI Computation 42

8 Testing and Validation **45**
8.1 Resimulation 45
8.2 Simulation 46

9 Conclusions **49**
9.1 Results 49
9.1.1 Framework Significance and Outlook 50
9.2 Answers to the Research Questions 51
9.3 Shortcomings 53
9.3.1 Data Ingestion 54
9.3.2 Data Preprocessing 54
9.3.3 KPI Computation 54
9.3.4 Iterative Development 54
9.4 Future Work 55
9.4.1 Time synchronization 55
9.4.2 Simulation and Resimulation 56
9.4.3 Ground truth and references 56
9.4.4 Edge case identification 56
9.4.5 Integration with safety standards 56
9.5 Conclusion 57

Bibliography **59**

A Appendix **I**
A.1 Switch Console Setup VI

List of Figures

2.1 Typical vessel and sensor layout (example) 8

4.1 Levels of Autonomy in Shipping, as defined by One Sea [8] 16

5.1 False Positives for Obstacle Detection. 28
5.2 False Negatives for Obstacle Detection. 28
5.3 Object Tracking Consistency 29
5.4 Docking Spot Determination. 30
5.5 Final Orientation of the Boat. 30
5.6 Speed Profile Approaching the Dock. 31
5.7 Collision Speed of Boat. 32

6.1 System with logging equipment 35
6.2 Initial Logging Setup 37
6.3 Ethernet Switch Wetermo Lynx 3510-F2G-T8G-LV. 37
6.4 Logging Wiring Diagram 39

8.1 Resimulation Process 46

9.1 Overview of the proposed data-driven V&V framework for marine ADAS/AD, linking scenario definition, instrumentation, raw-data logging, and post-analysis planning. 50

List of Tables

2.1 Specifications of Mechanical LiDAR 8
2.2 Specifications of Solid-State LiDAR 9

A.1 Initial KPI Set I
A.2 Selected KPI Set II
A.3 Console Port Parameter Settings VI

1 Introduction

This chapter introduces the thesis and frames it within the context of ADAS/AD for marine systems. It outlines the relevant background, reviews current testing and validation practices, surveys what the market offers, and highlights the gaps that remain. The core problem is then defined alongside the project's aims and scope limitations. The chapter concludes with a brief overview of the thesis structure.

The maritime industry is witnessing a transformative wave as autonomous systems become integral to vessel operations, driven by advancements in automation and autonomy across transportation sectors. Technologies for ADAS/AD, widely implemented in automotive and aerospace industries, are now being adapted to improve the safety and efficiency of marine navigation and management, while also reducing operational costs and supporting environmental sustainability. However, the marine sector faces distinct challenges that complicate the adoption of these technologies. Variable sea conditions, such as currents and unpredictable weather, challenge system reliability. Sensor performance is often hindered by water surface reflections or fog, and compliance with diverse global maritime regulations adds complexity. Additionally, the marine industry has seen less investment in autonomous technologies compared to other sectors, slowing progress. These obstacles highlight the need for a data-driven framework to support the development, testing, and validation of marine ADAS/AD systems, enabling efficient testing and safe deployment in real-world marine environments.

1.1 Background

This thesis proposes a data-driven framework tailored for marine ADAS/AD applications, designed to address these challenges by providing a standardized platform for evaluating autonomous technologies. By incorporating KPIs, the framework offers a consistent method to monitor system performance throughout its development, from initial testing to deployment. It employs real-world test scenarios and performance metrics to address emerging challenges in marine automation, ensuring autonomous vessels can operate reliably in complex settings like crowded ports or stormy seas. The framework's development follows a structured approach, using sensors such as LiDARs and DPS to collect critical data for assessing metrics like obstacle detection, navigation precision, and speed control. Built on an iterative testing model, it continuously gathers and analyzes data to refine system performance throughout its development, from initial testing to deployment.

1.1.1 Volvo Penta

This thesis was conducted in collaboration with Volvo Penta during Spring 2025 and draws methodological inspiration from prior initiatives within Volvo Group Trucks Technology (GTT). Volvo Penta specializes in maritime propulsion and control, supplying systems for commercial and recreational vessels that increasingly enable ADAS/AD capabilities. Collaboration with CPAC Systems focusing on marine electronics and software integration provided practical input on instrumentation constraints, test-boat access, and integration practices relevant to data capture.
An additional reference point was Volvo GTT's established ADAS/AD V&V practice and the Data Science Lab (DSL) platform. While this thesis did not realize a full marine adaptation or integration of DSL, the framework concept and proof-of-concept activities were informed by GTT's scenario-based testing, disciplined data capture, and large-scale analytics principles. The collaboration therefore centered on knowledge exchange and methodology review, with preliminary demonstrations of how marine test data can be captured and structured for future ingestion.

1.1.2 Aims and Objectives

1.1.2.1 Aim

The aim of the thesis is to *outline and demonstrate a proof of concept* for a data-driven development framework that supports verification and validation (V&V) of marine ADAS/AD features. The goal is not to deliver a finalized product, but to establish a practicable approach grounded in real vessel logging that can be taken forward.

1.1.2.2 Objectives

To realize this aim, the work pursues the following objectives:

1. **Framework concept and scope:** Specify the core components and interfaces of a data-driven V&V framework for marine applications (scenario definition, instrumentation and logging, data handling, KPI calculation, and reporting), indicating how established on-road practices may be adapted to marine constraints.
2. **KPI definition:** Define KPI sets suitable for selected marine functions (e.g., docking support, surround awareness), including detection and timing measures as well as motion-control outcomes (e.g., approach speed profile, final orientation).
3. **Scenario design:** Formulate realistic, repeatable test scenarios aligned with intended operating contexts (harbor/docking and low-speed maneuvers), to enable consistent data capture and later KPI evaluation.
4. **Instrumentation and logging:** Outline an instrumentation plan and implement a proof-of-concept logging setup on a test vessel (e.g., LiDAR, IMU, GNSS/DPS and a compatible logger), establishing data formats and metadata sufficient for subsequent ingestion.
5. **Data handling approach:** Propose an ingestion and preprocessing strategy

compatible with enterprise analytics platforms (e.g., DSL principles), including timestamping assumptions, quality checks, and schemas for KPI computation acknowledging that a full platform integration was out of scope for this thesis period.

This structure avoids promising a finished product while providing a concrete pathway from concept and scenarios to logging and KPI design that can be extended in future work.

1.1.3 Research Questions

This research addresses three key questions to guide the framework's development:

- **Which KPIs best measure the system's performance, considering the end goal of the feature, and how effective are they at validating test results?** This question identifies which metrics would best reflect the performance of the ADAS/AD functionalities, ensuring reliable validation through a robust data collection pipeline.
- **How scalable is the framework, given that the number, configuration, and type of sensors will increase in complexity as Volvo Penta advances its automation roadmap?** How does this scalability support their automation journey? This question explores the framework's adaptability to evolving systems and sensor setups, aiding Volvo Penta's long-term marine autonomy goals.
- **What challenges arise when developing a framework for marine applications, given the established use of development frameworks for land vehicles?** This question examines marine-specific obstacles, such as environmental variability and regulatory compliance, compared to automotive frameworks.

1.2 Delimitation

This thesis focuses only on testing out the framework on two features on small to medium-sized vessels in controlled and semi-controlled environments, such as harbors or coastal waters. As such, it excludes deep-sea operations and large-scale commercial shipping due to their distinct requirements. Testing is limited to scenarios accessible within the collaboration with Volvo Penta, Volvo GTT, and CPAC systems.

1.3 Ethical and Sustainability Aspects

Sustainable Development Goal 3: Good Health and Well-Being - *Ensure healthy lives and promote well-being for all ages*
Contribution: Improving maritime safety directly impacts the health and well-being of crew members and passengers by reducing the risk of accidents. A safer working environment leads to better mental and physical health outcomes for those on board.

Sustainable Development Goal 9: Industry, Innovation, and Infrastructure - *Build resilient infrastructure, promote inclusive and sustainable industrialization, and foster innovation*
Contribution: Developing a framework for ADAS exemplifies innovation in the maritime industry. It leads to the creation of resilient infrastructures capable of withstanding various operational challenges. By fostering technological advancements, our project contributes to sustainable industrialization and positions the maritime sector to adapt to future demands.

Sustainable Development Goal 11: Sustainable Cities and Communities - *Make cities and human settlements inclusive, safe, resilient, and sustainable.*
Contribution: Ports and coastal cities benefit from safer and more efficient maritime operations. Reduced accident rates and environmental hazards contribute to the sustainability and resilience of these communities, ensuring they remain safe and thriving hubs of economic activity.

Sustainable Development Goal 13: Climate Action - *Take urgent action to combat climate change and its impacts.*
Contribution: Enhanced navigation and vessel control systems contribute to optimized routes and speeds, leading to reduced fuel consumption and lower greenhouse gas emissions. By minimizing the carbon footprint of maritime operations, our project supports global efforts to mitigate climate change.

Sustainable Development Goal 14: Life Below Water - *Conserve and sustainably use the oceans, seas, and marine resources for sustainable development.*
Contribution: By integrating advanced technologies such as LiDAR, IMUs, and DPS, our framework enhances maritime safety and operational efficiency. This reduces the likelihood of accidents that can lead to oil spills, chemical discharges, and physical damage to marine ecosystems. Improved navigation and collision avoidance

systems help in preserving marine biodiversity and maintaining healthier ocean environments.

1.4 Usage of generative Artificial Intelligence

This thesis makes use of AI tools in accordance with Chalmers' *Regulations for the Use of AI Tools in Thesis Work* [1]. In particular, OpenAI's ChatGPT, along with Volvo AB's proprietary VolvoGPT, was used for this purpose. The usage of AI in this thesis was strictly limited to

1. Understanding complex topics and processes
2. Summarising and drawing inferences from existing literature
3. Improving clarity of language and correcting grammar in the text

2
Background

Docking Support (low-speed maneuvering) and Surround Sense (obstacle awareness) are implemented on a vessel stack using LiDAR and cameras for perception, DPS for positioning, and an onboard computer that fuses sensor signals, including CAN, to command actuators. Marine operation involves full 6-DoF motion, making perception and localisation more challenging than in road settings. The sensors, i.e., mechanical LiDARs for 360° coverage and solid-state LiDARs for near-field detail, and cameras for context, motivates the KPI selection and informs the validation workflow that follows.

To understand the decisions made about the current implementation of the framework, it is important to understand the features themselves and the system they are a part of. They are part of Penta's automation journey and are built on Volvo Penta's Electronic Vessel Control (EVC) 2 as a foundation. They mainly focus on Low Speed Maneuvering and Docking.

In collaboration with Volvo Penta, the framework was tested with two technologies currently under development:

1. **Surround Sense:** This system enhances situational awareness by detecting nearby obstacles such as other vessels, piers, or floating debris and issuing warnings or recommendations to the operator.

2. **Docking Support:** A low-speed docking feature built on Volvo Penta's assisted docking capability, supporting precise approach, alignment, and station-keeping near docks.

The system under development includes perception sensors, which include mechanical and solid-state LiDARs, along with the positioning sensor, DPS, which consists of the Inertial Measurement Unit (IMU) and a Global Navigation Satellite System (GNSS), which collectively provide geographic orientation and position. This must be especially robust because, unlike on-road vehicles, which have 2 degrees of freedom, the movement of a marine vessel's perception has 6 degrees of freedom. This results in complex perception and localization computations.

2.1 Equipment

2.1.1 Onboard Computer

The onboard computer takes the information from the sensors like LiDARs, DPS, and CAN (Controller Area Network) data, along with human input, processes it, and sends out control signals to the actuators and the rudders.

2.1.2 Mechanical LiDAR

Mechanical LiDARs, or spinning LiDARs, shown in Fig. 2.1, use moving parts to steer laser beams across a wide area. They provide wide coverage and dense point-cloud data, and make it easier to estimate the distance to obstacles, and thus making them suitable for perception in dynamic environments.

Table 2.1: Specifications of Mechanical LiDAR

Parameter	Specification
Horizontal FOV	360°
Vertical FOV Options	70°, 31°, 26°
Frame Rate	5, 10, or 20 Hz
Point Cloud Output	576,000 points per second

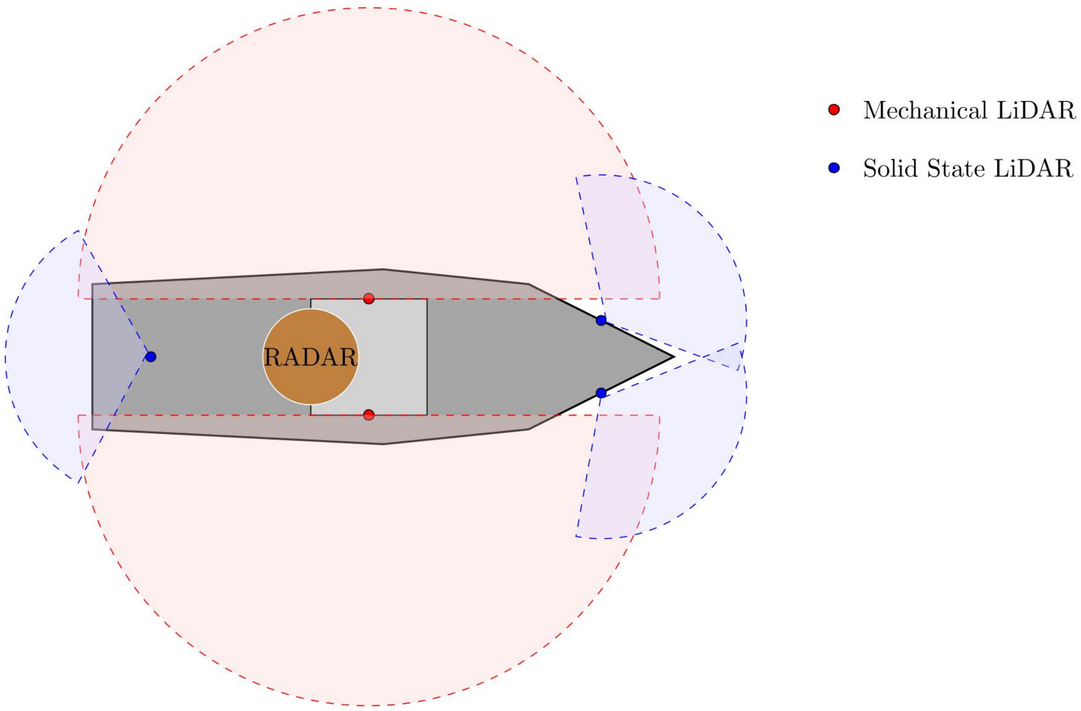


Figure 2.1: Typical vessel and sensor layout (example)

2.1.3 Solid-State LiDAR

Solid-state LiDARs, also shown in Fig. 2.1, have no moving parts and typically provide limited fields of view, but they are compact and reliable for short-to-medium range coverage.

Table 2.2: Specifications of Solid-State LiDAR

Parameter	**Specification**
Horizontal FOV	120°
Vertical FOV	90°
Frame Rate	10 Hz
Point Cloud Output	260,000 points per second

3
Methodology

This chapter details the approach for developing and validating a data-driven framework for ADAS/AD, and automation in marine applications. The methodology consists of several essential steps: defining KPIs, designing test scenarios, selecting suitable sensor equipment, and establishing procedures for data collection and processing. A preliminary test, referred to as a dry run, was conducted to confirm the equipment's functionality and ensure a dependable setup for subsequent testing. The methods for collecting, preprocessing, and analyzing data are described, followed by an explanation of how these data support the evaluation of system performance. The chapter concludes with an overview of the testing and validation process, emphasizing how the framework was iteratively improved based on real-world data and performance feedback.

3.1 State of the Art

The development of the framework started with gathering insights from stakeholders within the industry, particularly from the Volvo GTT, where automation frameworks are already established for on-road vehicles. A central aim was to explore whether these frameworks could be adapted to address the specific requirements of marine automation, such as navigating variable sea conditions or complying with maritime regulations. This process involved close collaboration with experts across different divisions of the Volvo Group to identify applicable technologies and methodologies.

The research also investigated automation efforts by other companies, particularly in the shipping industry, where autonomous systems are making significant advancements. This external review provided valuable perspectives on the challenges of implementing automation in marine environments, such as ensuring sensor reliability in adverse weather, and opportunities, such as improving navigational safety and operational efficiency. By analyzing these examples, the study gained a deeper understanding of current trends and practical approaches to maritime automation.

Furthermore, the study reviewed existing standards and regulations in the automotive and marine sectors to ensure the framework aligns with industry requirements. This included considering environmental factors unique to marine settings, such as unpredictable weather, sea states, and sensor limitations in conditions like fog or high waves. By drawing on successful case studies and established practices, the framework seeks to bridge the gap between automotive and maritime automation, focusing on improving the safety and performance of ADAS/AD systems in marine contexts.

3.2 Feature Investigation

The framework developed in this thesis was intended to be applicable across a range of ADAS/AD functionalities in marine applications. Rather than being limited to a single capability, the design goal was to establish a process that could be applied to multiple features, thereby supporting standardized evaluation across different automation domains.

For the purpose of scoping and demonstrating the framework, the two representative features were chosen because both rely heavily on accurate sensor data, real-time interpretation, and decision-making, which are critical challenges in marine automation. By focusing on these cases, the intention was to explore how the framework could support testing and validation of functions that combine situational awareness with vessel maneuvering tasks.

It should be emphasized, however, that the complete validation of these features within the framework could not be achieved in the course of this thesis. While initial efforts successfully instrumented the vessel and enabled the collection of raw data, no KPI computation or subsequent evaluation of feature performance was carried out. As a result, the usefulness of the collected data for full performance assessment remains to be verified in future work. The present study, therefore, provides a scoped foundation: the framework has been outlined, representative features identified, and data successfully logged, but the end-to-end validation pipeline remains to be completed.

3.3 KPI Formulation

The formulation of KPIs constitutes a fundamental element of the proposed framework, as it enables objective and traceable measurement of system performance throughout the development lifecycle. KPIs provide quantifiable evidence that can be used not only to verify functionality during testing but also, more importantly, to ensure that the implemented features achieve the required levels of safety, reliability, and usability when introduced into production.

Within this thesis, KPIs were defined to address two complementary dimensions. At the feature level, KPIs capture the operational effectiveness of specific functionalities, such as docking support and surround sense, by assessing aspects including detection accuracy, response latency, and maneuvering precision. At the system level, KPIs serve to monitor development progress across different stages, allowing consistency and comparability between test campaigns and scenarios. This dual purpose ensures that performance evaluation is both function-specific and aligned with the broader objective of validating the system for deployment.

The selection process for KPIs was informed by consultation with Volvo Penta, which identified critical operational challenges characteristic of the marine domain. From the wide range of possible indicators, the scope was narrowed to those directly reflecting core performance objectives and feasible to evaluate given the available test setup. Although the actual computation of KPIs could not be performed within the timeframe of this thesis, due to the absence of synchronized ingestion and preprocessing capabilities, the KPI definitions and their intended applications were established. These definitions constitute the basis for future analyses, where they will enable consistent and reproducible evaluation of logged data against well-defined performance targets.

3.4 Developing Scenarios

An essential element of the framework design was the development of representative test scenarios. Scenarios provide the necessary link between abstract KPIs and the practical conditions under which they can be measured. By defining structured and repeatable situations, they enable data to be captured consistently and ensure that KPI calculations remain relevant to operational challenges.

The scenarios considered in this thesis reflect typical and safety-critical conditions in marine operations, such as docking maneuvers, obstacle encounters, and navigation in constrained or dynamic environments. The intention was not only to mimic real-world variability, but also to provide a systematic way of testing feature performance across comparable trials.

Due to project constraints, these scenarios were not executed to their full extent within the scope of this work. However, their formulation establishes the foundation for future testing campaigns, where they will be used to generate traceable evidence of ADAS/AD performance through KPI-based evaluation.

3.5 Equipment Research

Following the formulation of KPIs and scenarios, the next methodological step was to identify the equipment required to capture the necessary data. The guiding principle was to ensure that the instrumentation could provide sensor streams of sufficient fidelity and relevance to enable future KPI calculation. In some cases, the existing sensors on the vessel could provide adequate coverage, while in other cases, supplementary equipment was considered to address gaps in measurement.

The final equipment configuration had to balance technical requirements with practical constraints such as data volume, cost, and availability. Particular emphasis was placed on the logging hardware, which needed to accommodate high-bandwidth inputs from LiDAR and CAN without data loss. As the instrumentation was temporarily borrowed from Volvo GTT, availability and compatibility with the test vessel were further considerations.

By grounding equipment selection in the data requirements of defined KPIs and scenarios, the framework ensures that the collected datasets are both relevant and reusable for future validation efforts.

3.6 Data collection

Data collection was structured in two stages to balance risk management with practical access to the test vessel. The rationale was to first validate the equipment setup under controlled conditions before moving to the more constrained environment of vessel testing.

3.6.1 Dry Run

Initial dry runs were conducted on a test rig at the Lundby campus after installing the equipment borrowed from Volvo GTT. The purpose of these trials was to confirm that the logging hardware could reliably capture high-bandwidth sensor streams and that the connections between devices functioned as intended. Performing these checks in a controlled environment minimized the risk of integration failures during vessel testing, where access time was limited. Only after the system demonstrated stable performance on dry runs was the proposal made to proceed with full-scale testing on the boat.

3.6.2 On-Site Testing

The on-site vessel tests were carried out in collaboration with CPAC and conducted in multiple stages to maximize the value of limited boat time. The initial focus was on ensuring that the equipment could be integrated with the vessel systems with minimal intrusion. Once feasibility was established, subsequent runs were directed toward capturing meaningful raw sensor data for later processing. This phased approach ensured that vessel testing time was used efficiently and that the collected datasets would be suitable for demonstrating the framework's data-driven workflow.

3.7 Data Handling

While the data collection process was underway, training for access to the data analysis platform was initiated. Since the platform included tools for everything from ingesting to processing data, it was considered an integral part of the framework. After reviewing the platform, a Jupyter Notebook-based tool was identified as the most suitable option for the framework. Although access to the server space was secured, the data could not be uploaded for analysis due to time constraints. And due to time delays, the final KPIs could not be calculated and would have to remain an incomplete objective.

4
State of the Art

This chapter surveys development and V&V practices for autonomy in marine and automotive domains. It sketches autonomy taxonomies (SAE, One Sea, IMO, class societies), then contrasts data logging and pipelines at scale in autos with emerging, simulation-heavy workflows at sea. KPI use is framed around safety, rule compliance (including COLREGs), efficiency, and quality of operation. Scenario-based, simulation-led V&V is highlighted as the common backbone, with domain specifics handled via operating bounds (ODD/OE) and X-in-the-Loop testing. The chapter also outlines Volvo Group toolchains (GTT V&V, DSL) as practical infrastructure that can be adapted for marine validation.

4.1 Development Frameworks in Marine domains

ADAS/AD technologies are evolving rapidly in both automotive and maritime contexts, enabled by advances in sensing, data processing, and AI. In the automotive industry, large fleets have accumulated millions of miles of real-world data supplemented with extensive simulation, while the maritime sector is advancing Maritime Autonomous Surface Ships (MASS) [19] that integrate similar technologies for navigation and control. A data-driven approach to development and validation is now essential in both domains, since relying solely on large-scale road tests or sea trials is impractical. For example, a RAND study [3] estimated that proving the safety of fully autonomous vehicles through mileage alone would require hundreds of millions of miles, an infeasible proposition. Similarly, maritime experts agree [4] that autonomy cannot be validated by real-world trials alone and will require systematic simulation-based testing alongside controlled experiments. Consequently, both domains emphasize data-centric development: logging sensor data at scale, processing it through structured pipelines, measuring performance with KPIs, and embedding these steps within V&V frameworks.

This literature survey reviews the state of the art in data-driven development and validation of ADAS/AD systems, with a focus on maritime applications and comparisons to established automotive practices. The review addresses (1) data logging and collection practices, (2) data pipelines for processing and utilization, (3) KPIs for evaluating autonomous system performance, and (4) existing and emerging V&V frameworks. Both academic research and industrial reports are surveyed, with an emphasis on maritime-specific literature and selective inclusion of automotive studies to highlight transferable methods and lessons.

4.2 Autonomy Level Classifications

The concept of autonomy has been formalized through classification frameworks that define the balance of human involvement and automated control. In the automotive domain, the SAE J3016 standard [6] provides the most widely adopted taxonomy, defining six levels of driving automation (0–5). Level 0 represents no automation, with the human driver responsible for all tasks. Level 1 denotes driver assistance, where automation manages either steering or acceleration/braking under specific conditions. Level 2 provides partial automation, with the system controlling both steering and speed but requiring continuous driver supervision. Level 3 introduces conditional automation, where the system performs all driving tasks within a defined operational domain, though the driver must take over when requested. Level 4 represents high automation, where the system manages all tasks in its operational domain without expecting human intervention. Level 5 denotes full automation, with the system assuming complete responsibility under all conditions.

For the maritime domain, several parallel classification systems exist. The One Sea framework [8] defines six levels of automation (0–5), ranging from manual operation to full autonomy. Level 0 represents manual control with basic automated assistance, while Level 5 indicates a fully autonomous vessel operating without human involvement.

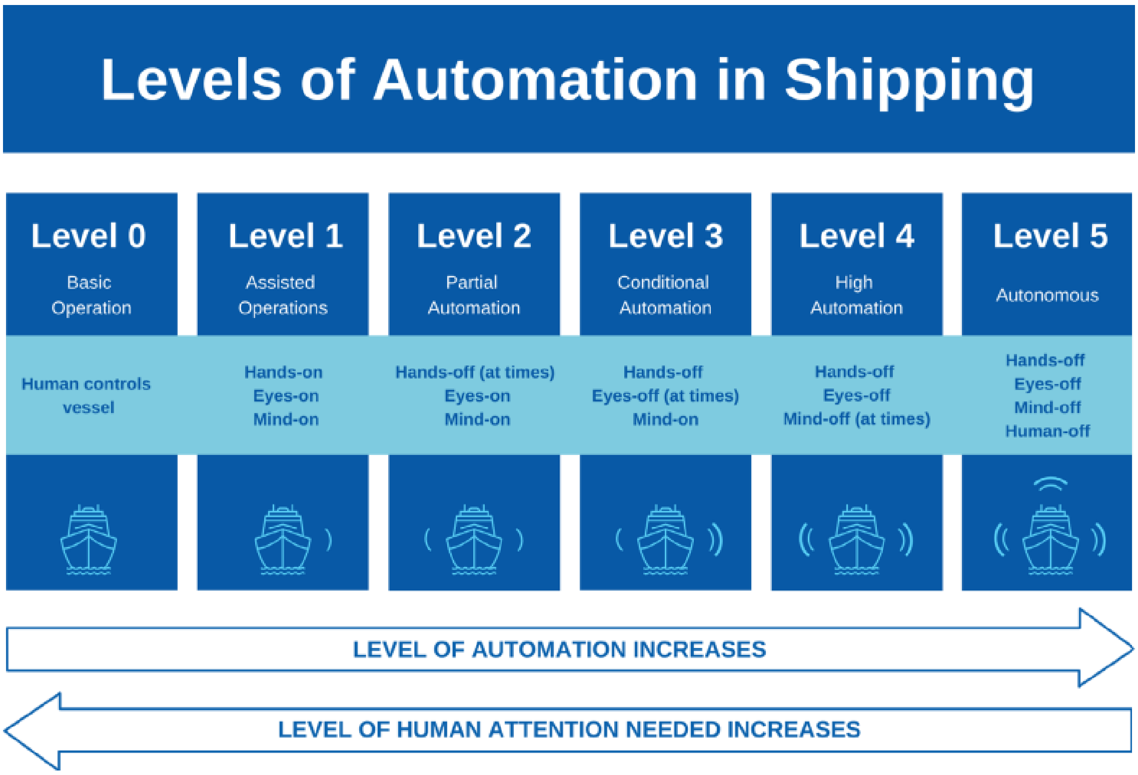


Figure 4.1: Levels of Autonomy in Shipping, as defined by One Sea [8]

Unlike SAE J3016, which classifies automation by the extent to which an ADS performs the Dynamic Driving Task (DDT) and who provides the fallback (human vs.

system), the One Sea levels are tailored to the maritime context and place greater emphasis on *mode of operation* (onboard vs. remote), watchkeeping responsibilities, and integration with COLREGs. In other words, SAE J3016 differentiates primarily along the dimensions of sustained lateral/longitudinal control and fallback responsibility (e.g., Levels 2–5), whereas One Sea distinguishes stages of maritime automation that include manual operation with decision support, supervised autonomy, and increasing degrees of remote operation. The two taxonomies are therefore not isomorphic: a given One Sea level may cut across multiple SAE notions of control/-fallback because it additionally encodes where control is exercised (bridge or shore), how watchkeeping is ensured, and what role the crew retains. This domain-specific focus reflects fundamental differences between road traffic and shipping, including vessel manning, shore-control options, and rule compliance at sea.

The International Maritime Organization (IMO) has defined four degrees of autonomy for MASS [9]. Degree 1 refers to ships with automated processes and decision support where crew remain on board. Degree 2 describes remotely controlled ships with crew present. Degree 3 covers remotely controlled ships without crew on board. Degree 4 represents fully autonomous ships operating without human intervention.

Industry classification societies have also contributed. Lloyd's Register defines six autonomy levels (AL1–AL6) [10], spanning from minimal automation (AL1) to fully autonomous ships requiring no onboard human access (AL6). Det Norske Veritas (DNV) has introduced the AROS (Autonomous and Remotely Operated Ships) class notations [7], which distinguish autonomy modes such as remote control, decision support, supervised autonomy, and full autonomy. These notations also categorize control location as onboard, remote, or hybrid.

Together, these frameworks illustrate structured approaches to defining and regulating autonomy across both automotive and maritime sectors. Each provides a graded scale for distinguishing levels of human involvement, remote operation, and automation, forming a foundation for research, development, and standardization.

Table A.3 summarizes the principal autonomy classification frameworks across domains, including SAE (J3016) for road vehicles and several maritime counterparts such as IMO, DNV, Lloyd's Register, Bureau Veritas, and One Sea. These taxonomies are not presented for completeness alone; they delimit the expected *human role*, *fallback responsibility*, and *operational envelope* for automated functions, which in turn shape verification and validation (V&V) evidence. Because this thesis targets assistance/perception functions in operator-in-the-loop settings (i.e., below fully autonomous operation), the scenario catalogue and KPIs are aligned with those levels: near-harbor maneuvers, obstacle awareness, approach speed control, and final orientation are prioritized over fully autonomous watchkeeping or remote-operation handover. Conversely, for higher degrees of autonomy (e.g., IMO MASS 3–4 or One Sea 4–5), the V&V burden would shift toward broader scenario coverage (COLREGs encounters at scale), independence from human fallback, and additional KPIs (e.g., supervisory takeover latency, remote-control link integrity). Thus, the classification frameworks inform the scope of this work (assistance-grade capabilities), the KPI set (safety and controllability under pilot supervision), and the scenario design (harbor/docking contexts), providing traceability between intended autonomy level and required evidence.

4.3 Data Logging and Pipelines

This section outlines how sensor data is captured and transformed into analysis-ready assets in marine and automotive domains, emphasizing elements relevant to a data-driven V&V framework.

4.3.1 Marine Domain

Modern vessels increasingly carry multi-modal sensors (e.g., radar, LiDAR, cameras), navigation sources (GNSS/INS), and maritime feeds (AIS, ECDIS) [19]. In current practice, logs primarily originate from pilot projects and controlled trials, with digital twins used to complement limited real-world coverage [4]. AIS is particularly valuable for broad coverage and encounter mining [19]. Typical pipeline steps include: (i) integrity checks and metadata attachment (time, vessel state, environment), (ii) time alignment across streams, (iii) enrichment with maritime context (charts, AIS traffic, wind/wave/current estimates), and (iv) scenario extraction (e.g., docking approaches, encounter types) [21]. Because public fleets at autonomy-ready maturity are small, careful curation and reuse of each log is essential; simulation outputs are often co-processed with trials to maintain consistent formats and facilitate replay [4, 21].

4.3.2 Automotive Domain

Automotive programs operate at larger scale via test fleets and production logging [20]. DataOps-style pipelines upload vehicle logs to centralized storage where automated ETL(Extract–Transform–Load) performs integrity checks, synchronization across modalities (camera, radar, LiDAR), and structuring into searchable repositories [18]. Annotation (manual or assisted) and event mining segment logs into scenarios (e.g., near-misses, cut-ins) for training and V&V [20]. At scale, selective logging, compression, and trigger-based capture manage cost while preserving safety-critical coverage [20]. The end result is a traceable flow from raw logs to reusable scenarios and KPI dashboards for regression across software releases [18, 20].

4.3.3 Cross-Domain Takeaways for This Thesis

Three practices are directly applicable to the proposed marine framework: (i) early, consistent time alignment and metadata so streams are analysis-ready; (ii) scenario-centric indexing so logs are discoverable and reusable for KPI computation; and (iii) selective capture and enrichment to maximize the value of limited vessel time [18, 21]. The automotive experience demonstrates that these steps enable reproducible, KPI-driven validation; the marine context adapts them with maritime-specific enrichment (AIS/ECDIS (Electronic Chart Display and Information System), sea state) and trial/simulation co-processing [19, 4].

4.4 KPI for Validation and Performance Measurement

Defining suitable KPIs is essential for quantifying ADAS/AD performance and tracking improvements over time. KPIs condense complex behaviors into measurable indicators that can be compared against requirements or benchmarks such as human performance or regulatory thresholds. Both automotive and maritime domains use KPIs spanning safety, comfort, efficiency, and rule compliance, though the specific metrics differ.

4.4.1 Automotive KPIs

In automotive programs, scenario- and risk-based KPIs are commonly used to quantify safety and rule compliance. Core surrogate safety measures include *time-to-collision* (TTC), *(time) headway*, minimum distance/gap, and lane-keeping/centering error [22]. Comfort-related metrics such as longitudinal/lateral acceleration and jerk are tracked because they correlate with ride comfort and driver/passenger acceptability [24, **?**]. Higher-level safety formalisms, notably Responsibility-Sensitive Safety (RSS), define minimum safe distances and proper responses that can be monitored as KPI violations during tests [25]. In parallel, functional-safety and SOTIF standards (ISO 26262 and ISO 21448) require measurable acceptance criteria and pass/fail definitions across verification campaigns, thereby shaping KPI design and evidence collection [26, 27]. Reliability and operational KPIs such as interventions or disengagements per mile and mean time between hazardous events are reported in industry practice to characterize maturity, while acknowledging their limitations as aggregate indicators [28]. Feature-specific KPIs are also defined; for instance, emergency braking evaluations consider residual impact speed, speed reduction, and stopping distance per established protocols [29]. Finally, many programs automate KPI computation and regression tracking in dashboards that aggregate simulation and road-test results to support continuous V&V [30, 20].

4.4.2 Marine KPIs

In the maritime domain, KPI portfolios emphasize COLREGs compliance, navigation precision, and operational performance. Collision-avoidance performance is commonly evaluated by replaying canonical encounters (head-on, crossing, overtaking) and assessing both closest point of approach (CPA) and rule conformance [16, 36, 37]. Simulation-based verification frameworks explicitly encode such criteria, combining COLREGs-based maneuvering obligations with CPA thresholds to derive pass/fail outcomes [4, 44].
Beyond rule conformance, *navigation- and control-quality* indicators are widely used in constrained waters. Route-following accuracy (cross-track error, XTE) in track-/waypoint control and *station-keeping* performance (position/heading error within a hold region) are particularly relevant for harbor transits and docking operations [21, 17, 33]. These KPIs complement COLREGs: while COLREGs govern right-of-way and maneuvering obligations during encounters, XTE and station-keeping

quantify how precisely the vessel executes intended motion and maintains position in tight operational envelopes.
At a broader program level, initiatives such as AEGIS propose KPIs covering economics (transit time, fuel consumption, operating cost), environment (emissions), and safety (incident and near-miss rates) [31, 42]. Finally, benchmarking against human performance is a common expectation in maritime autonomy; autonomous functions are typically required to meet or exceed safety and operational records of comparable manned operations [38, 39]. This mirrors practices in road-vehicle programs, where human-driver baselines are likewise used to contextualize KPI targets and acceptance thresholds.[27].

4.5 V&V Frameworks for ADAS/AD

Developing confidence in autonomy requires robust verification and validation (V&V) frameworks that combine simulation, real-world testing, and analytical methods. Because exhaustive physical testing is infeasible, contemporary practice emphasizes *scenario-based* and *data-driven* validation, guided by domain-specific requirements and standards (e.g., COLREGs in maritime, functional safety per ISO 26262 and Safety of the Intended Functionality (SOTIF) per ISO 21448 in automotive) [26, 27, 32, 44].

4.5.1 Automotive V&V Frameworks

In automotive, there is broad consensus around *scenario-based testing* as a central V&V strategy: rather than relying on aggregate mileage, systems are evaluated against curated catalogs of scenarios derived from real data and risk analysis to ensure coverage of edge cases [32, 35]. Standards and guidance (e.g., ISO 21448 on SOTIF) support building and maintaining such scenario databases and testing automated driving systems (ADS) systematically against them [27].

- **Simulation-based testing:** High-fidelity environments execute large numbers of scenarios per software release, including log-replay of real events and parameterized synthetic variants. Simulation enables safe discovery of failure modes and accelerated exposure to rare events [32].
- **Closed-course testing:** Once simulation confidence is gained, specific scenarios are validated on proving grounds with controlled conditions (e.g., robotic targets, scripted actors). This real-world component verifies aspects that simulation might miss (vehicle dynamics limits, environmental sensing effects) [26].
- **Limited on-road trials:** Instrumented pilots with safety drivers provide operational evidence in naturalistic conditions, with all interventions feeding back into scenario coverage [27].

Formal and analytical methods are sometimes applied at component boundaries (e.g., supervisory safety logic), while full formal proofs for ML-intensive components remain out of reach. The *Operational Design Domain (ODD)* is pivotal: V&V articulates the conditions under which the ADS is intended to operate and then demonstrates adequate coverage and performance within that ODD [27]. Emerging practice integrates X-in-the-Loop (MiL/SiL/HiL/ViL) to surface issues early and supports *data-driven scenario generation*, where field logs are mined to discover untested but safety-relevant situations [32]. Alongside this, functional safety (ISO 26262) and SOTIF (ISO 21448) provide process requirements for hazard analysis, verification planning, and evidence traceability across the V-model [26, 27].

4.5.2 Maritime V&V Frameworks

Maritime V&V is converging on principles similar to those in automotive while adapting to domain specifics such as environmental stochasticity, longer maneuver time-scales, and COLREGs compliance. The literature and current prototypes indicate a simulation-driven, scenario-based approach for MASS.
A representative example is the digital-twin-centric framework in [4]: a high-fidelity vessel model (dynamics, sensors, actuators) operates within a virtual environment (sea state, weather, traffic), while a test manager generates encounter scenarios, runs the autonomy stack in-the-loop, and automatically evaluates outcomes against safety and rule-conformance criteria (e.g., collision-avoidance success, closest-point-of-approach thresholds, COLREGs adherence). Scenario search can be adaptive, stressing regions where performance deteriorates to improve coverage.
Regulatory and classification activities increasingly emphasize bounded operation. Guidance highlights the explicit definition of a vessel's operational conditions, extending the automotive ODD toward an *Operational Envelope (OE)* that also considers modes, transitions, and human/remote supervision, and requires V&V evidence of safe operation within those bounds [33]. Beyond simulation, maritime practice employs scenario generators (including AIS-derived encounter clusters) and scaled or controlled trials as intermediate steps when full-scale testing is costly or risky [21]. Standards work is ongoing; interim guidance and research point toward safety-case-oriented assurance where simulation evidence, scenario test results, and KPI trends are assembled to argue acceptable risk for defined operations [34].
Across domains, contemporary V&V aligns on a few core principles: (i) clearly define operational bounds (ODD/OE); (ii) maximize high-fidelity simulation with scenario catalogs informed by real data; (iii) use quantitative KPIs (safety, rule compliance, performance) to evaluate outcomes; and (iv) incrementally validate via X-in-the-Loop and targeted field trials. Automotive offers mature processes and standards; maritime adapts these patterns to COLREGs-constrained navigation and ocean-environment variability. These principles directly motivate the framework in this thesis data logging and ingestion, scenario design, and KPI-based evaluationwith a path toward SiL resimulation and reproducible evidence generation.

4.6 Volvo Group Toolchains (GTT/VAS)

This section outlines Volvo Group toolchains that inform the proposed marine V&V framework by providing mature, field-tested practices from on-road ADAS development. The intent is not to replicate automotive processes verbatim, but to identify reusable concepts scenario-based testing, disciplined data capture, and analytics-at-scale that can be adapted to marine constraints. First, we summarize the ADAS Verification & Validation (V&V) approach at Volvo Group Trucks Technology (GTT), emphasizing how scenarios, triggers, and reference signals structure evidence generation. We then describe the DSL platform used across Volvo for ingestion, curation, and analysis of high-volume test data. Together, these elements motivate concrete design choices for the marine framework, including scenario driven logging, metadata conventions, and a path toward scalable KPI computation.

4.6.1 ADAS Verification & Validation at GTT

Within Volvo GTT, the ADAS V&V process provides a structured and proven workflow that spans from data collection to final validation. Real-world data is gathered through both controlled test track experiments and field trials, where scenarios are explicitly designed around predefined KPIs to measure system performance in a repeatable manner. Deterministic testing on dedicated tracks allows the validation team to control environmental variables, ensuring that specific functionalities can be assessed against targeted goals. Field validation complements this by fine-tuning and confirming system behavior under naturalistic conditions. During these trials, CAN logging, reference video, and event-triggered data capture are used to create rich datasets that support annotation and facilitate accurate bug reporting, including feedback loops to sensor suppliers when issues are identified. Dedicated loggers are used to capture both CAN signals and sensor outputs in a consistent format, making it possible to systematically detect, reproduce, and resolve errors. Once collected, these logs are transferred to the Data Analytics team, which processes and analyzes the data to generate actionable insights for development. This end-to-end V&V process has already demonstrated its maturity and efficiency by enabling Volvo trucks to bring ADAS features from concept through validation and ultimately into production.

4.6.2 The Data Science Lab

Complementing the V&V workflow, the Data Science Lab (DSL) serves as an advanced analytics platform built on GTT's High Performance Computing infrastructure. DSL plays a central role in supporting the development of next-generation vehicles, including electromobility, ADAS, hydrogen internal combustion, and autonomous solutions, by providing a unified environment for data storage, management, and analysis. The platform is widely used across the Volvo Group, with more than 500 engineers and analysts from GTT, Volvo Buses, and Volvo Active Safety (VAS) relying on its capabilities. DSL integrates multiple data sources, including fleet data collected from test and customer vehicles equipped with loggers, the Volvo Active Safety Platform, and various simulation rigs, while also incorporating open-source datasets such as weather records. Its functionality extends from data ingestion and metadata management to batch and stream processing, enabling downstream applications in artificial intelligence, machine learning, simulation and re-simulation, interactive analytics, and visualization. Data can be ingested manually through physical storage media or automatically through upload stations, after which metadata management and data engineering pipelines prepare it for analysis. By coupling scalable storage with structured ingestion and advanced analytics tools, DSL provides a one-stop solution for handling both physical test data and simulation outputs. Importantly, its integration into the same pipelines ensures tight coordination between experimental data collection and computational analysis, ultimately improving efficiency and accelerating feedback into the development cycle.

4.7 Conclusion and Outlook

This review of the state of the art demonstrates that both the automotive and maritime sectors are converging toward data-driven, scenario-based approaches for the verification and validation of ADAS/AD technologies. In both domains, it has become evident that exhaustive physical testing alone is neither feasible nor sufficient, and that the integration of simulation, structured data logging, and KPI-based evaluation is essential to ensure safety and performance. The automotive industry, supported by large-scale fleets and established pipelines, has reached a high level of maturity in this regard. Practices such as automated scenario extraction, regression testing, and continuous KPI monitoring have already been proven effective in bringing systems to production. By contrast, the maritime sector is at a comparatively earlier stage of development, relying largely on smaller pilot projects, testbeds, and digital twins to compensate for limited real-world data. Although progress is visible, significant challenges remain in terms of reproducibility, access to data, and regulatory acceptance of simulation-based evidence.

For this thesis, the most important insight is that many of the established methods and infrastructures developed for automotive applications can be directly adapted to the maritime domain. The challenges faced by both sectors are structurally similar, revolving around the need to capture uncertainty in complex environments, to quantify safety through KPI-based arguments, and to use simulation to supplement real-world trials. The divergence lies primarily in domain-specific aspects such as compliance with COLREGs, the influence of sea states and weather conditions, and the longer temporal dynamics of vessel maneuvers. From this observation, a preliminary conclusion can be drawn that existing toolchains within the Volvo Group, specifically those developed by Volvo GTT and Volvo Active Safety, represent strong candidates for reuse. The GTT ADAS V&V process has already demonstrated efficiency in bringing features into production by using KPI-based scenario design, structured logging, and systematic fault management. In parallel, the Data Science Lab has established itself as a scalable platform for data ingestion, storage, and analytics, which could directly support the requirements of a marine validation framework.

The outlook for future work, therefore, suggests that progress in marine ADAS/AD validation is best achieved not through the invention of entirely new processes, but by adapting and extending proven automotive methodologies. Leveraging the maturity of GTT and VAS infrastructures provides a pragmatic and efficient path forward, while further work must focus on defining domain-specific KPIs, constructing COLREGs-compliant scenario databases, and implementing resimulation capabilities such as software-in-the-loop. In this way, the cross-domain transfer of knowledge and tools can accelerate the establishment of a robust framework for the safe and reliable deployment of ADAS/AD technologies in maritime applications.

5
Key Performance Indicators

This chapter defines a focused KPI set to validate perception (Surround Sense) and control (Docking Assist), and details how the features were chosen, how the KPIs are tied to onboard signals, and how the set was narrowed to metrics that capture behavior rather than status. Emphasis is on practical, low-overhead computation. A compact, repeatable set of scenarios and a structured dataset plan (with conditions, flags, and brief annotations) preserve comparability over time and support decision-making.

KPIs form the backbone of this data-driven framework, providing a consistent and objective way to measure performance across the entire development cycle. Each KPI is defined based on how a particular feature is supposed to behave, ensuring that it captures meaningful aspects of performance rather than general system statistics. This helps move away from subjective judgments toward measurable, repeatable verification.

5.1 Feature Determination

Because the goal of this framework was to remain feature-independent, the process of defining KPIs started by first identifying a set of representative features that could demonstrate its functionality. The two chosen features were Surround Sense and Docking Support, both under active development at Volvo Penta. Both rely heavily on real-time sensor data, accurate positioning, and decision-making, which makes them ideal for testing and refining the framework. These operational goals directly informed the selection of measurable KPIs.

5.2 Defining KPIs

The process of defining KPIs followed a logical flow based on how each feature functions and how its performance can be represented through existing data. It began with understanding the intent of the feature: what defines good performance, what outcomes matter most, and how those outcomes could be observed through sensor signals. From there, measurable indicators were identified using available onboard sensors. This mapping ensured that the KPIs remained practical to calculate and relevant to the vessel's capabilities.

Each KPI was therefore designed around three main criteria:

- It should be computable from the vessel's recorded data (LiDAR, CAN, etc.). External aids (e.g., a temporary reference camera or short manual annotation) may be used to validate or calibrate a KPI.
- It must capture a concrete aspect of feature behavior that can improve or deteriorate over time (e.g., accuracy, timing, stability), not just general system status.
- It should have a clear reference or baseline, whether that is through comparing different sensors (like LiDAR and camera), using known physical limits, or setting defined thresholds, so that the results make sense and can be compared across different tests.

For Surround Sense, detections can be checked against a multi-sensor consensus (camera + radar, aligned via DPS) to score false positives/negatives. In Docking Support, docking angle error and approach speed were defined by using DPS and LiDAR data, which provided measurable indicators of control precision and safety.

In cases where reference data or ground truth were required, different strategies can be applied depending on the situation. If direct measurement were possible, higher-precision sensors or manual annotation could be used to validate results. However, since marine testing environments rarely allow perfect ground truth due to dynamic sea states and limited calibration, approximate references such as cross-checking LiDAR data with DPS readings or verifying obstacle detections across multiple sensors are good alternatives. This ensured that each KPI still carried quantitative meaning, even when ideal references were not available.

5.3 Refinement and Prioritization

Once a broad list of KPIs was generated, the next step was refinement. The initial list included a large number of possible metrics covering nearly every aspect of vessel operation. However, not all of them directly reflected the performance of the selected features. Many were operational or environmental indicators rather than performance measures. Therefore, the list was narrowed to retain only those KPIs that best captured the functional behavior and supported meaningful interpretation.

Relevance was another main criterion for reduction. KPIs that did not clearly link to feature outcomes or decision-making were removed. Some overlapping metrics were merged for clarity (for example, different angle-based alignment metrics were combined into "Final Orientation Accuracy"). Others, while not directly useful for feature validation, were moved to a diagnostic list for investigating unexpected behavior during development.

5.4 Implementation Considerations

Every KPI relies on data from the vessel, and with a finite sensor set, practicality drove the final choices. The framework leans on what is already on board to keep the data collection simple and avoid extra hardware. Where it meaningfully boosts confidence, short use of a temporary reference sensor (e.g., a camera) or a small manual labeling pass can be worthwhile, but these are optional. By default, the setup is self-contained and repeatable across vessels without major rework.

This process yields a trimmed set of KPIs that reflect how the system senses, decides, and behaves, covering both perception and control. Table A.1 shows a subset of the broader initial list before refinement, and the following sections describe the final KPIs used for validation, with diagnostics kept for investigating unexpected behavior during development.

5.5 Selected KPIs

Table A.2 lists the KPIs used for the target features. Each KPI is computed over short windows of LiDAR frames with the associated CAN and DPS signals, rather than single-frame samples, to reduce noise.

5.5.1 Surround Sense

These KPIs focus on how effectively the system perceives its environment, identifying obstacles, maintaining consistent detections, and recognizing valid docking spots.

5.5.1.1 False Positives for Detection of Obstacles

Definition: A false positive occurs when the system detects an obstacle that is not really there.

Detections are checked frame by frame and grouped into short events. To filter out brief flickers from spray or wakes, an event is only counted if it lasts at least N_{fp} consecutive frames (e.g., $N_{\text{fp}} = 3$). Each event is then compared against reference data, either a camera label, an offline pass, or a quick manual review. If no matching object exists, this is counted as a false positive.

$$\text{False Positive Rate}(\%) = \frac{\text{False Positives}}{\text{False Positives} + \text{Correct Detections}} \times 100$$

In Figure 5.1, the vessel (blue) detects real targets (orange) but also flags an empty area as an obstacle, i.e., a false positive.

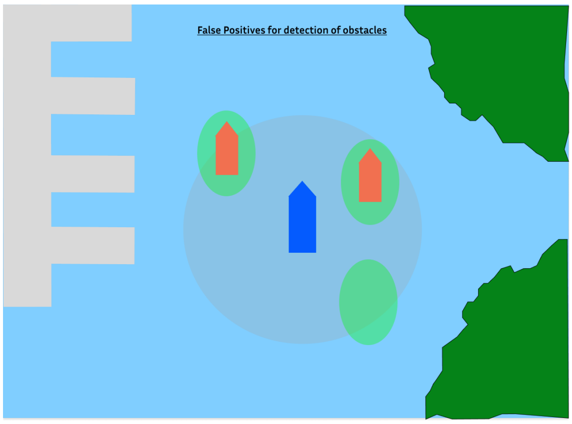


Figure 5.1: False Positives for Obstacle Detection.

5.5.1.2 False Negatives for Detection of Obstacles

Definition: A false negative occurs when a real, visible obstacle is not detected by the system.

Reference obstacles are tracked over time. If one remains visible but is not detected for N_{fn} consecutive frames (e.g., $N_{\text{fn}} = 5$), this is considered a false negative. This avoids counting minor one-frame losses as full detection failures.

$$\text{False Negative Rate}(\%) = \frac{\text{False Negative}}{\text{False Negative} + \text{Correct Detections}} \times 100$$

Figure 5.2 shows a false negative: a real target (orange) inside the coverage zone is not detected by the vessel (blue).

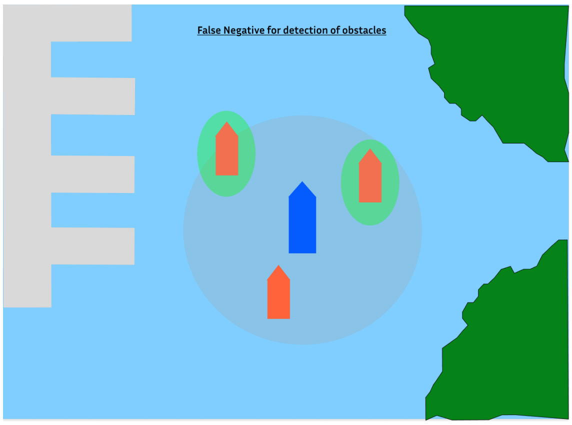


Figure 5.2: False Negatives for Obstacle Detection.

5.5.1.3 Object Tracking Consistency

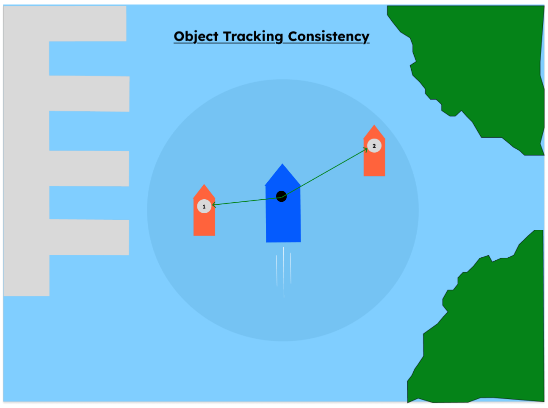
Definition: It is the share of real objects that keep a single, continuous track ID across frames, no ID swaps, no duplicate tracks for the same object, and correct re-linking after brief dropouts.

Each object is followed over time; short misses are fine if the tracker reconnects to the same object on return. A track is inconsistent if the ID changes mid-sequence, if two tracks cover the same object, or if a missed object comes back with a new ID.

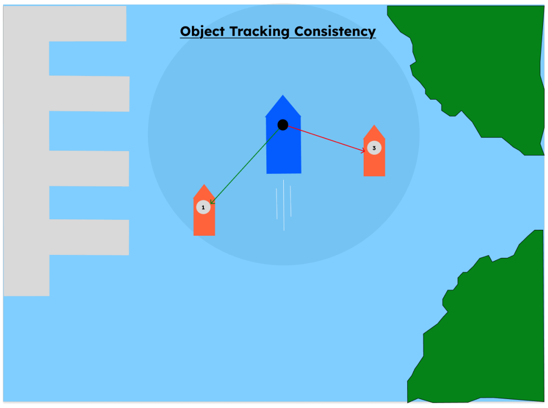
Example: a buoy seen in frame 1, briefly lost in frame 2, and seen again in frame 3 should retain the same ID; a new ID indicates unstable tracking.

$$\text{Consistency Rate}(\%) = \frac{\text{Consistent Tracks}}{\text{Total Tracks}} \times 100$$

In Figure 5.3a, the vessel (blue) correctly assigns IDs 1 and 2 to the two nearby vessels; in Figure 5.3b, the second vessel is misidentified as ID 3, an ID switch that breaks tracking continuity and is undesirable.



(a) Obstacles 1 and 2 are correctly identified



(b) Obstacle 1 correct; Obstacle 2 misidentified as 3

Figure 5.3: Object Tracking Consistency

5.5.1.4 Docking Spot Determination Rate

Definition: Measures how reliably valid docking spots are identified during approach.

Suitable spots are tracked over the approach window. A spot is only counted if it is correct and remains stable for at least Δt_{spot} (e.g., 3) seconds;

$$\text{Docking Spot Accuracy}(\%) = \frac{\text{Correctly Identified Spots}}{\text{Total Available Spots}} \times 100$$

In Figure 5.4, the vessel (blue) detects three of the four available docking spots (light green) but misses one, reducing overall docking-spot determination accuracy.

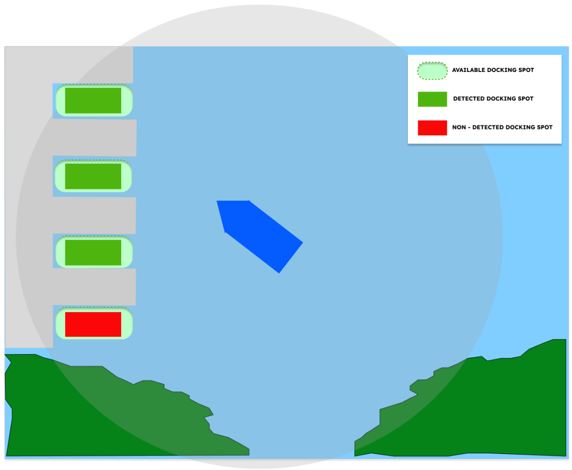


Figure 5.4: Docking Spot Determination.

5.5.2 Docking Assist

These KPIs capture how precisely and smoothly the vessel performs during docking, from approach speed and alignment to overall docking stability.

5.5.2.1 Final Orientation of the Vessel

Definition: After docking, the vessel should sit parallel to the dock, with the average orientation error $\theta(t)$ over a short hold window Δt_{hold} (e.g., last 5 s) staying within a small limit $\theta_{\max}$ (e.g., 2°). This windowed average indicates a stable final pose rather than a one-off alignment. As shown in Eq. 5.1, t_f is the final time at which steadiness and alignment are evaluated.

$$\bar{\theta} = \frac{1}{\Delta t_{\text{hold}}} \int_{t_f - \Delta t_{\text{hold}}}^{t_f} \theta(t)\, dt, \qquad \bar{\theta} \leq \theta_{\max} \text{ (e.g., } 2^\circ\text{).} \tag{5.1}$$

In Figure 5.5, the vessel is misaligned by 5° relative to the dock, exceeding the tentative 2° limit, hence a fail.

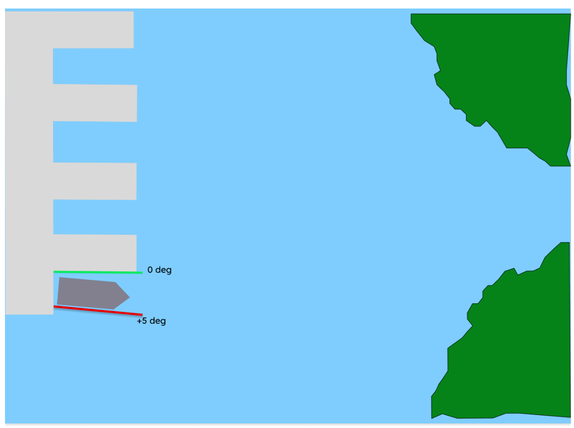


Figure 5.5: Final Orientation of the Boat.

5.5.2.2 Speed of Approach to the Dock

Definition: As the boat gets closer to the dock (distance $d(t)$ decreases), its speed $v(t)$ should taper smoothly without surges. Within the near-dock zone ($d \leq d_{\text{near}}$), target a small speed cap (e.g., $v_{\text{cap}} = 0.5\,\text{kn}$).
A steady, predictable slowdown keeps docking safe and avoids harsh corrections. Checks near the dock (inside d_{near}) include:

- Monotonic slowdown: $\dfrac{dv}{dd} \leq 0$ for $d \leq d_{\text{near}}$.
- Speed cap: $v(t) \leq v_{\text{cap}}$ (e.g., $0.5\,\text{kn}$) for $d \leq d_{\text{near}}$.
- Smoothness: avoid sudden jerks; quantify with integrated absolute jerk $J = \int_{t_0}^{t_f} \left| \dfrac{d^3x}{dt^3} \right| dt$.

Passing Criteria: No sustained speed increases for $d \leq d_{\text{near}}$, $v(t) \leq 0.5\,\text{kn}$ in the final approach, and $J \leq J_{\text{max}}$.

In Figure 5.6, the vessel's speed decreases to less than a preset limit, ie, 0.5 knots, within a preset distance, d_{near}

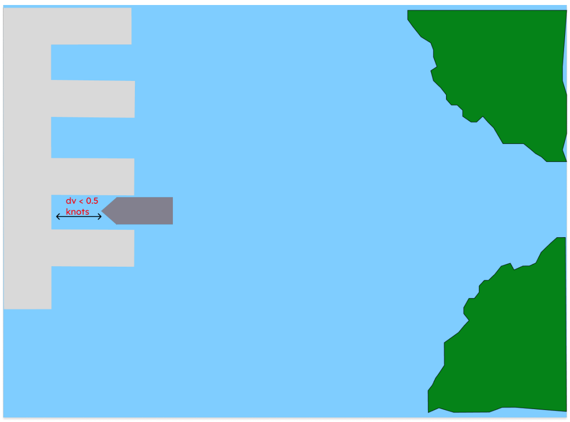


Figure 5.6: Speed Profile Approaching the Dock.

5.5.2.3 Collision Speed to an Obstacle

Definition: At the instant of contact or minimum range (t_c), the vessel's contact speed should remain below a safe cap (e.g., 1 knot for low-speed docking). The relative speed is computed from the vessel and target velocity vectors at that instant:

$$v_{\text{rel}}(t_c) = \left\| \mathbf{v}_{\text{vessel}}(t_c) - \mathbf{v}_{\text{target}}(t_c) \right\|,$$

where $\|\cdot\|$ denotes vector magnitude. For a fixed dock, $\mathbf{v}_{\text{target}}(t_c) = \mathbf{0}$, so $v_{\text{rel}}(t_c) = \|\mathbf{v}_{\text{vessel}}(t_c)\|$.
Passing criterion: $v_{\text{rel}}(t_c) \leq v_{\text{safe}}$ (e.g., 1 knot).
In Figure 5.7, we see the speed before the collision is less than a safe threshold (e.g., 1 knot for low-speed docking)

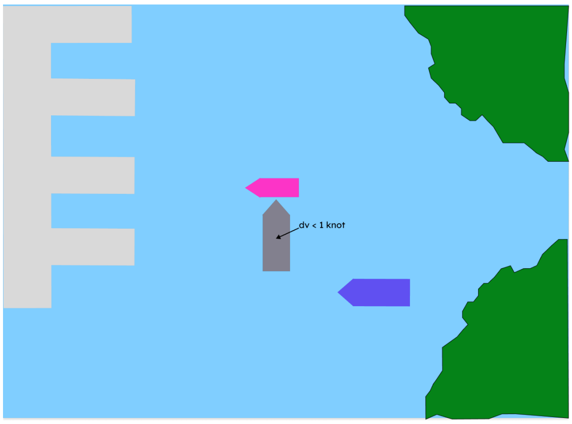


Figure 5.7: Collision Speed of Boat.

5.5.2.4 Total Docking Time

Definition: Total time from docking start until the vessel holds a stable, final position:

$$\text{Total Docking Time } = t_{\text{stable}} - t_{\text{start}}.$$

Here, t_{start} marks Docking support engagement (or the first entry into the docking zone), and t_{stable} is the first time a short hold window meets all limits for orientation, speed, and distance. Lowering the final/relative speed tends to lengthen total docking time, while raising it tends to shorten the time. Plotting a graph with the total docking time and v_{rel} chart makes the balance clear and helps choose practical targets.

5.6 Scenarios

Reliable KPI calculation starts with simple, repeatable scenarios that are easy to run on the water and directly tied to the KPIs. Each scenario is defined by (i) what is being tested, (ii) how the run starts and stops, and (iii) which signals and flags are recorded so the data can be aligned later.

To keep runs comparable, the same scenario is repeated several times under similar conditions. A single repeatable test often feeds multiple KPIs in one go, which is far more efficient than running separate tests for each metric.

What the spec includes (kept minimal):

- *KPI link:* which KPIs the scenario is meant to exercise (e.g., false positives/negatives, approach speed, final orientation, docking time).
- *Start/stop and flags:* when logging begins/ends and the triggers used to drop event flags (e.g., “feature on/off”, “enter docking”, “candidate spots shown”).
- *Signals to record:* the existing sensor outputs used in the framework (e.g., LiDAR, DPS/GNSS/IMU, camera video for context).

The final set of scenarios is outlined below, with the understanding that each is executed under varied environmental conditions (e.g., day/night, rain, fog, snow, and different sea states):

- **Obstacle Detection:** The vessel navigates through areas with known obstacles to evaluate detection accuracy. False negatives are measured when real objects are missed, and false positives are measured by driving through empty areas to detect incorrect obstacle recognition. The same data also supports calculating tracking consistency, verifying that each obstacle is recognized as a single object across frames.
- **Docking Support:** Conducted in a marina, this scenario activates the docking assist feature to identify and approach available docking spots. The sequence provides data for multiple KPIs: docking accuracy, final orientation, speed during approach, and total docking time, capturing how the vessel aligns and stabilizes near the dock. Repeating this scenario under slightly different conditions (e.g., wind or starting angle) helps confirm robustness.
- **Collision Avoidance:** The vessel and an obstacle (static or moving) are placed on a near-collision path to test if the system reacts by reducing speed before impact. The data collected is used to calculate relative collision speed and check whether it remains below the safety threshold, as well as to observe any delay between detection and response.
- **Docking Spot Determination:** The perception system identifies potential docking locations, which are then compared to a manually verified list of valid spots. The results show how accurately the system distinguishes real, accessible spaces from false candidates and how stable those identifications remain throughout the test.

These scenarios were chosen because they capture the essential behaviors of the features. Running them repeatedly across varying conditions builds a consistent and comparable dataset, ensuring that the derived KPIs reflect how the system performs in realistic, everyday marine operations.

5.7 Dataset Determination

After the KPIs are defined, it is equally important to decide what kind of data needs to be collected, under which conditions, and how it will be labeled for analysis. The goal is to capture data that reflects realistic operating conditions across the intended ODD. Doing so helps ensure that the KPIs represent how the system performs in the real world, not just in ideal conditions.

Each test scenario is repeated across selected conditions to build a dataset that is both representative and consistent. For example, docking can be tested in calm and wavy conditions, during daytime and low-light, to observe how approach speed or detection rates change. This structured repetition makes it possible to analyze KPI stability and reliability across different environments.

In parallel, triggers can be used to mark key events inside the recorded data, such as when a feature is activated, when the vessel enters a scenario, or when unusual perception behavior occurs. These flags do not define the dataset itself but make it easier to later align the data with real events during KPI computation.

Finally, environmental cameras or short synchronized video clips can be added for visual context, providing quick confirmation of conditions (e.g., rain, fog, or nearby traffic). Combined with metadata on scenario type, time, and conditions, this approach gives a well-structured dataset that supports accurate, repeatable KPI evaluation.

6

Data collection

This chapter presents a pragmatic data-logging architecture for validating marine ADAS/AD KPIs. It motivates dedicated instrumentation, details the selected logger alongside a managed switch with port mirroring and required media conversion, and records why active capture modules were evaluated but not adopted. A compact camera suite and a trigger/dashboard provide human-interpretable context and lightweight event annotation. Vessel integration is addressed through wiring layout, sensor/computer placement, and power management to preserve data integrity and reproducibility. The campaign comprises bench verification and on-vessel trials conducted under practical constraints (e.g., port-mirroring limits, weather windows), yielding multi-modal logs (PCAP (Packet Capture), AVI (Audio Video Interleave), MF4 (Audio Video Interleave) suitable for downstream KPI computation and analysis.

6.1 Data Logging Equipment

Once the KPIs and the test scenarios have been finalized, the equipment required to collect the data needs to be investigated and determined. This would involve considering several different equipment and finally settling on solutions that are cost-effective while still ensuring that they are effective for any future testing that Penta has planned.

The data logging setup is illustrated in Fig. 6.1. It highlights the equipment used to capture sensor outputs.

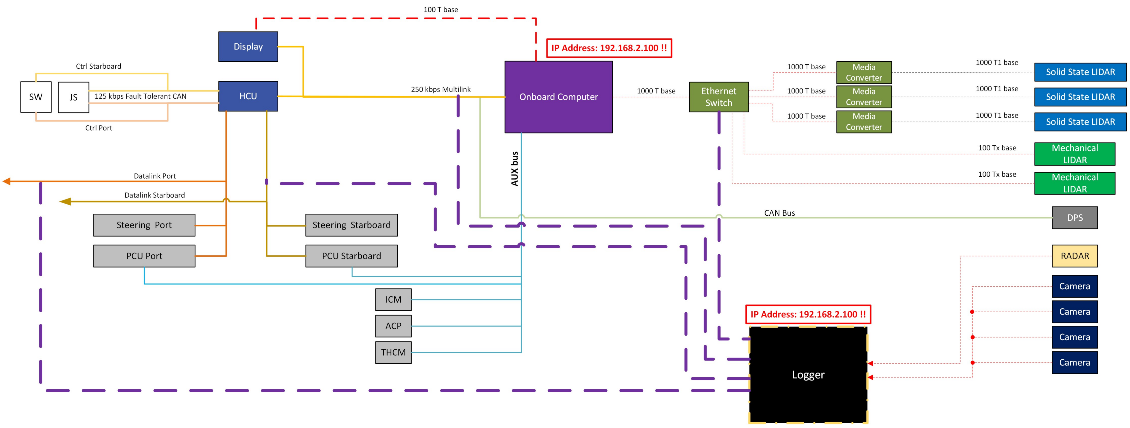


Figure 6.1: System with logging equipment

6.1.1 Logger

To measure the system's performance, relevant data first need to be collected. This involves logging data from the vessel in operation. If the computer cannot handle the incoming data volume, it may drop samples, produce inaccurate timestamps, and overload CPU/memory resources. These failures corrupt the log and render it unreliable.

For this reason, dedicated data loggers are used to log data for sensitive data collection tasks. Because these devices accommodate high-throughput streams in multiple formats and continue operating reliably across wide temperature, vibration, and humidity ranges, built-in features such as power-failure buffering, checksum verification, and secure shutdown procedures preserve data integrity even under adverse conditions. They are external, non-intrusive devices that can be added to or removed from the system without major changes.

Conventional wisdom suggests that the data logger must be able to handle 1.5 times the maximum throughput of the data flowing through it to ensure data fidelity [46]. The calculated output from the LiDARs alone would be around 110 GB per hour. The onboard computer could not reliably handle that level of throughput while simultaneously running onboard processing. A cost-effective and flexible logger was identified as suitable for the current use case as well as future upgrades in Penta's automation roadmap. It is a highly customizable device that supports automotive communication formats such as CAN, FlexRay, LIN, and Ethernet/SOME-IP, covering both protocol and traffic-level communication. The device records outputs in `.pcap` format, ensuring compatibility with standard network analysis tools. It has a bandwidth of 4 Gbit/s, which is well above what was calculated for this application. Further research led to a logger, which was more readily available in inventory but offered lower bandwidth. However, testing was nonetheless continued with the initially selected one.

6.1.2 Capture Modules

These loggers are further enhanced when paired with capture modules, either passive or active, that tap the data lines close to the source. By buffering data bursts and applying hardware timestamps at the capture point, the modules ensure that, even if the main logger momentarily misses packets, each event is accurately timestamped and can be fully reconstructed during post-processing. Since these active capture modules do onboard processing, they introduce lag to the data stream. Passive capture modules, on the other hand, just tap into the data feed and send it to the logger without processing anything.

Suitable active capture modules supporting CAN and Ethernet were identified. However, due to limited availability, high demand, and prohibitive cost, they were removed from the data-collection strategy. The initial logging strategy is shown in Fig. 6.1.2.

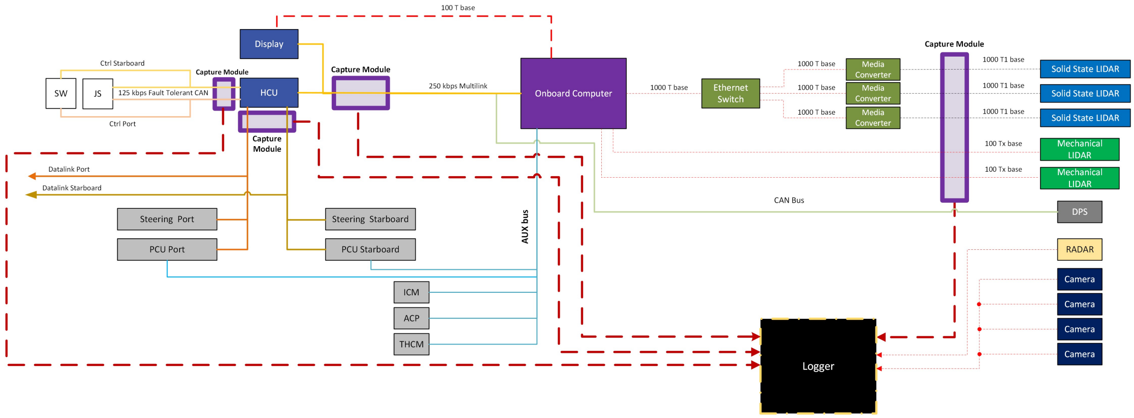


Figure 6.2: Initial Logging Setup

6.1.3 Ethernet switch

A basic Ethernet switch was used for the current system, which did the basic task of forwarding the data from the LiDARs to the computer. Since the LiDARs used in the system are programmed to send their data to a specific IP address, the logger must be configured with that same address to receive the data. However, this leads to an IP address conflict, which can cause packets to be misrouted, dropped, or received inconsistently, especially if multiple devices on the network share that address.

By using port mirroring on a manageable Ethernet switch like Wetermo Lynx 3510-F2G-T8G-LV, Fig. 6.3, the data flowing through to the computer can copy the data and send it simultaneously to the logger. However, this again adds a bit of lag to the input to the computer.

Figure 6.3: Ethernet Switch Wetermo Lynx 3510-F2G-T8G-LV.

The switch can be configured using the PuTTY software with a few simple steps and commands as seen in section A.1.

6.1.4 Media Convertors

The switch used here only supports classical Ethernet (1000T base & 100Tx base), so some output from LiDARS in automotive Ethernet (1000 T1 base) needs to be converted to classical Ethernet. Media converters are used for this purpose. They are networking devices that convert one type of communication to another so that otherwise incompatible devices can connect. This again adds another bit of lag to the system.

6.1.5 Cameras

It is integral for the developers of the system to see and record the surroundings of the boat to get an idea of what is happening around it. Using only the perception sensors, such as LiDAR, is unreliable because, like most sensors, it is not perfect and can produce false positives and false negatives. Also, videos from cameras are easier for humans to interpret since they have higher resolution and include color. For this reason, cameras were integrated into the logging system. These are low-resolution Axis cameras (800x600, 10fps) that provide enough information to the developers while generating small file sizes. Four strategically placed cameras were thought to be enough to cover the surrounding waters. The cameras were directly connected to the logger using Ethernet cables via a camera hub that could accommodate all four cameras.

6.1.6 Trigger/Display/Dashboard

A touchscreen display was connected to the logger to serve as an ON/OFF switch for the logger. It was configured to show the type and size of the data flowing through the logger at any given time, ensuring that the data is complete and not corrupted. It was also equipped with a customizable trigger that can be programmed to a wide array of actions because of the logger's flexibility. For the proof of concept, the display was configured to record audio from the cabin for 30 seconds when triggered, in addition to the already logged data. This will be useful in the field when a test driver wants to provide additional context for what they would be experiencing. This can involve unusual weather patterns, system glitches, unresponsive controls, and other issues. This would help improve the system developers' ability to pinpoint the exact time when a problem arises. Another potential use case for the trigger would be to record 30 seconds before and after the trigger push. This could also be sent over the air to the developers immediately. This would be useful for developers to rectify and possibly push updates remotely, or even if local data storage is a concern.

6.2 Equipping the Boat

Once the equipment is determined, the best method to integrate this system into the vessel is analysed. The equipment cannot be too intrusive and must be relatively easy to install and remove. The positioning of the sensors and the equipment also makes a difference. Placing the equipment far apart would lead to longer cables, and length limits can disrupt the logging system and also introduce lag. So ideally, the logging equipment would be spread out with the logger and the onboard computer in a central position. The power management of the setup also needs to be reliable and energy efficient. Although power interruptions would not damage the equipment, they could corrupt the data. The trigger must be securely placed near the helmsman. The vessel is under the purview of the team at the test site, in this case, Krossholmen. To ensure that they understand the requirements, a proposal needs to be sent with all the information needed for them to effectively assist with the equipping. Fig.6.4 shows the wiring schematics of the vessel for the logging campaign.

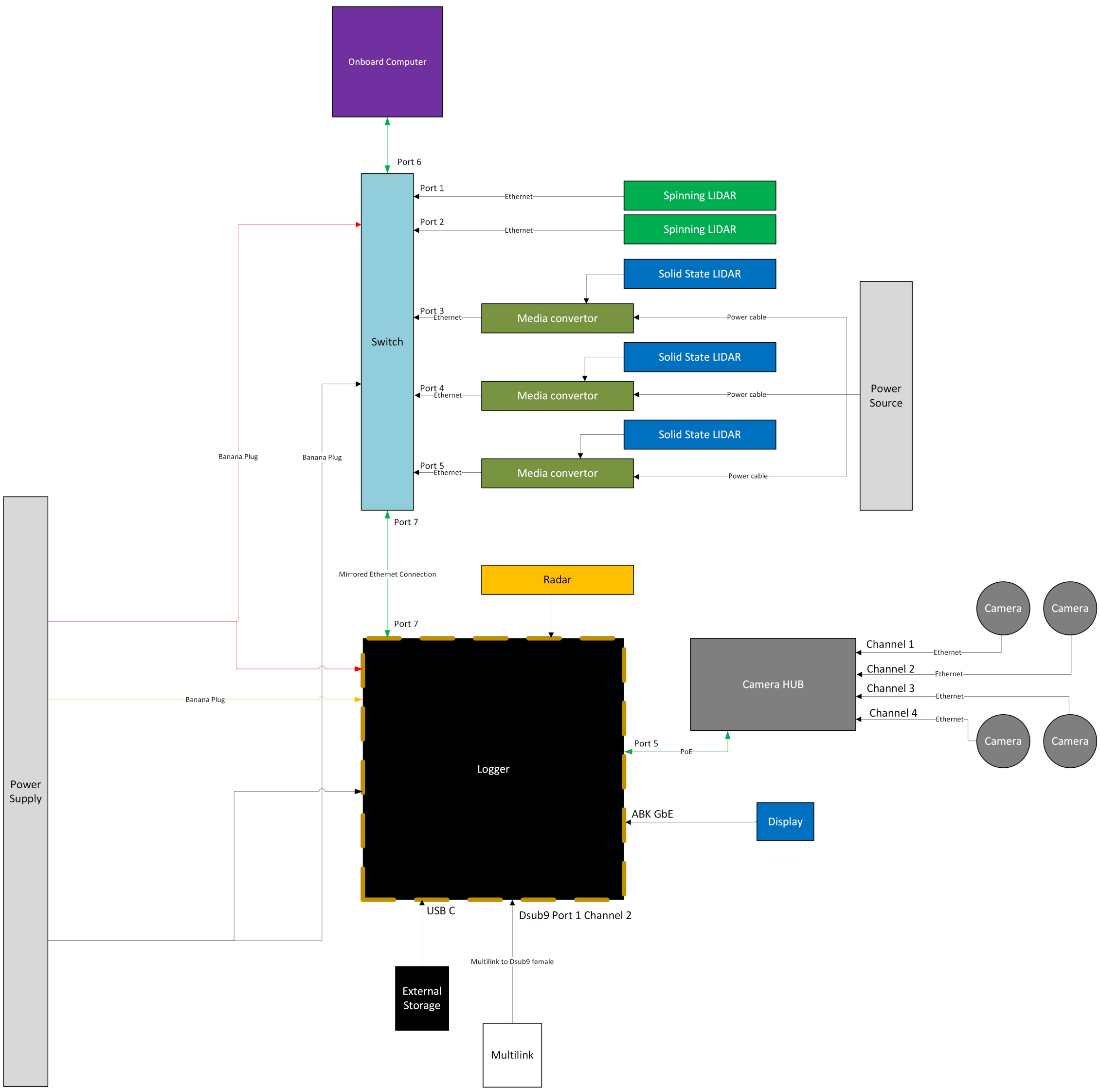


Figure 6.4: Logging Wiring Diagram

6.3 Data Collection

Before the actual data collection with the vessel, a preliminary data collection session was planned using the test rig at Volvo Penta's Lundby campus. The goal here was to make sure that the logging equipment worked as intended and could record the data reliably. Data in three formats, `.pcap` (LiDAR), `.avi` (Camera), and `.mf4` (CAN), were sent to the logger to check if it could reliably log and store the data flowing through it. The data was also logged for longer durations to make sure that the files were formatted as expected without any errors.

This was followed by a secondary data collection session on the actual test vessel to make sure that the equipment all communicated with each other and the logging was successful. In this setup, all devices, including all five LiDARs, were connected simultaneously to the logger and the onboard computer via a mirrored port on the manageable switch. During testing, data reached either the logger or the onboard computer, but never both at once which indicated a problem with the port mirroring. But upon further investigation, port mirroring was made possible, but with only one LiDAR. Unfortunately, since the time with the test vessel was limited, data was collected with the data from the five LiDARs flowing through only the logger. Furthermore, bad weather conditions limited the testing time, but some basic maneuvers were performed, and the data collected. Around 13 GBits of data were collected, which were then transferred to a portable HDD for easy transportation.

7
Data Handling

This chapter outlines how test data are moved, prepared, and converted into KPIs. It covers ingestion to centrally accessible storage (with DSL as the primary path), options for local/remote upload and over-the-air transfer, and security considerations for protected assets. On the processing side, it describes parsing LiDAR, UDP, PCAPs into point clouds, aligning video and CAN/DPS streams, correcting drops and clock drift, and segmenting runs into analysis windows (e.g., docking, obstacle events). KPI computation is implemented as repeatable notebooks in DSL (Spark-backed when needed), producing structured outputs with rich metadata (test ID, sensor/feature versions, conditions) to support comparison, reproducibility, and scale.

7.1 Data Ingestion

Once the data is collected, before it can be processed, it needs to be ingested into a central location made remotely accessible to all the people working on the project. This would also serve as a backup by storing the data at a secure location. There are a variety of methods to achieve this, depending on your application and physical location. Since almost all of this work was done in Gothenburg, access to the central storage at Lundby was preferable. This would allow us to physically upload the data to the CampX servers. Since Penta is not part of the system created by Volvo GTT, it does not have such a data pipeline in place. So the already existing data pipeline from Volvo GTT's ADAS team was considered. Another method would be to upload it directly to their storage space using DSL. For this work, DSL proved to be an ideal solution for handling test data and enabling KPI calculations. Access required completing a series of trainings to ensure familiarity with the tools and adherence to data-handling standards, which extended the timeline.

A common challenge for projects operating away from the home base is data transfer. For example, moving large volumes of vessel data from a distant test campaign to Penta's servers can be cumbersome. Conventional transfer methods may risk data corruption or raise security concerns (e.g., unsecured links), and requiring physical presence at the servers to upload is impractical. In such cases, setting up an on-site upload station is worthwhile. Data can be encrypted and sent automatically to the main servers, reliably and securely, without being present at headquarters [43].

Another potential challenge would be simultaneously, or failing that, periodically sending data over the air from customer tests back to the developers at Penta or CPAC safely and reliably. The logger has the option to have a modem connected to it that can send the generated data to the intended target and vice versa. But even then, since the data is generated by marine vessels operating in different parts of the world, it becomes a challenge to establish a secure connection between it and the servers before transmitting the data. It must also be secure enough that the data and Volvo's intellectual property behind the data are inaccessible to the customer.

7.2 Data Preprocessing and KPI Computation

Once the raw data is received from the logger, it must be processed before it can be useful for KPI calculation. The preprocessing steps depend on the type of data collected and its intended use. Since the logged data mainly consists of LiDAR point clouds, video recordings, and CAN or DPS signals, each source goes through a different treatment pipeline.

LiDAR data is sent as Ethernet frames over UDP and stored in `.pcap` format. These files are parsed using vendor tools such as RSView or the RoboSense SDK, or with open-source utilities like Wireshark, to extract 3D point clouds. The resulting data can be converted into `CSV`, `LAS`, or `PCD` format for further use. Video and CAN/DPS streams could be synchronized using timestamps, and missing packets, duplicates, or clock drifts corrected. The cleaned data is then grouped into event-based segments such as "Docking Approach," "Obstacle Detection," or "Collision Avoidance."

After preprocessing, the data is fed into Python-based KPI scripts. Each KPI is tied to a specific combination of sensor signals:

- **Perception KPIs** (e.g., false positives, false negatives, tracking consistency) compare LiDAR detections with reference data from cameras or short manual labeling. Time-series logic ensures that detections persist across consecutive frames before being classified.
- **Docking KPIs** (e.g., orientation, speed profile, total docking time) use position and velocity data from DPS and CAN logs. These signals are interpolated and filtered to remove noise, and then analyzed over time windows to compute performance metrics.

All KPI computations can be implemented as repeatable Jupyter notebooks within the DSL JupyterHub environment, which connects to a Spark backend for scalable data handling. Each notebook would load the preprocessed datasets, run KPI formulas (e.g., FPR, approach-speed gradient, orientation error), and output structured results in `CSV` or database format.

To improve reliability, each KPI run should store not only the computed values but also its metadata (test ID, sensor version, environmental conditions, and feature version) so that results can be directly compared between scenarios or across future trials. This metadata tagging also enables cross-validation between runs to identify deviations caused by weather, time of day, or test setup.

In short, preprocessing cleans and aligns the raw data, while the KPI computation stage transforms it into quantitative performance measures that can be analyzed, visualized, and compared over time. Together, they bridge the gap between raw logs and meaningful validation results.

8

Testing and Validation

This chapter sketches the iterative loop from trials to improvements: test, compute KPIs, review, and update. Logged data then drives resimulation (SIL/HIL) for open-loop features, enabling quick comparisons across software versions before going back to the vessel. Simulation supports early exploration and safe edge-case coverage, with models checked against field data. Together, these steps balance speed and realism, improving repeatability and confidence in KPI outcomes.

After the KPIs are calculated, their results are reviewed to determine whether performance is satisfactory. If not, the system goes through another iteration. This cycle of testing, KPI evaluation, and system updates is rarely a one-time process; it typically involves multiple rounds, with each round revealing new insights and shortcomings that guide the next update.

At times, the KPIs themselves may need to be reconsidered to ensure that the right metrics are being measured. For example, a KPI based on joystick sensitivity was replaced with one that measures approach speed relative to the distance from an obstacle, which proved both easier to calculate and more representative of actual system performance. Additional KPIs can also be introduced for diagnostic purposes, such as monitoring data throughput to track packet loss and latency to detect communication bottlenecks in the logging chain.

Each test generates large volumes of raw sensor data that capture real-world conditions such as weather, lighting, obstacle placement, and human input. This data forms the basis for **resimulation**, a critical step in the framework that allows developers to re-evaluate system performance offline using previously logged data.

8.1 Resimulation

In resimulation, the recorded raw data is replayed through a hardware-in-the-loop or software-in-the-loop version of the real system. The data is divided into two parts: a development or training set and a test set. The resimulation is tuned and validated on the first, and verified on the latter. This approach is particularly effective for evaluating open-loop features, where outputs from the system do not directly influence incoming sensor data. By running the same dataset through successive versions of the control or perception algorithms, developers can quantify improvements efficiently. Figure 8.1 provides a schematic overview of this workflow.

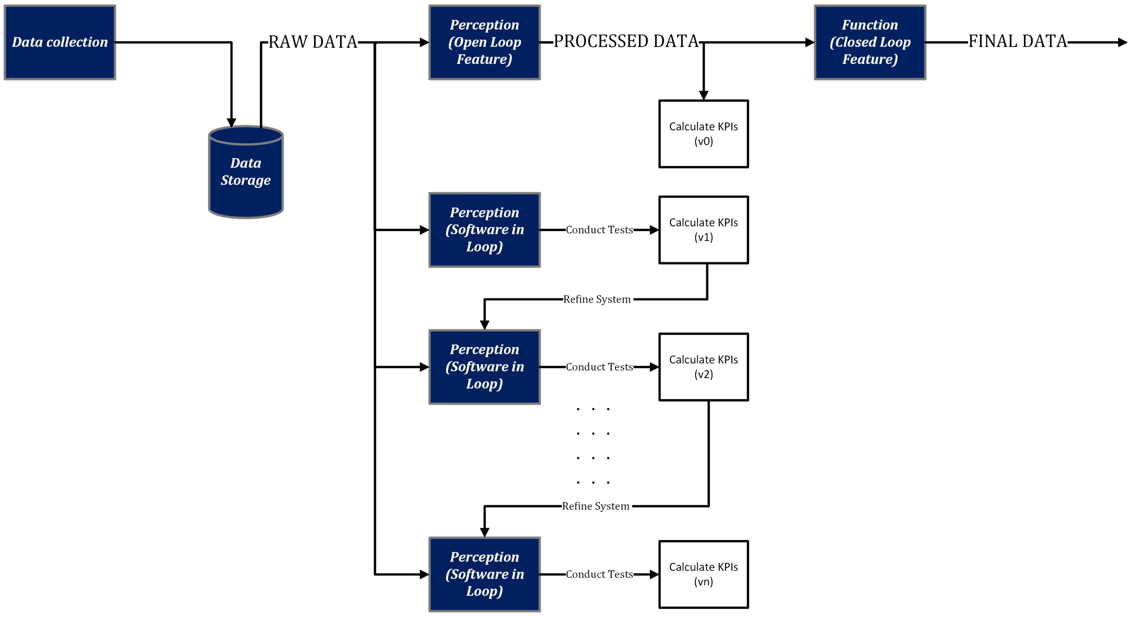


Figure 8.1: Resimulation Process

Once performance in resimulation reaches an acceptable level, the updated software is deployed back to the vessel for real-world validation. This process accelerates development and reduces dependence on vessel availability or external factors such as weather and scheduling.

However, resimulation has its limits. It is best suited for open-loop features, where the data can be replayed independently of live feedback. In closed-loop features, such as joystick docking assist, the system's output affects sensor readings, which invalidates replayed data and makes resimulation unreliable. These cases must still be tested in real-world or high-fidelity simulated environments to ensure accurate results.

8.2 Simulation

Simulation plays an important role in the development process as well. It can be used even before real-world data collection begins, providing an early indication of how the system performs in a safe, controlled setting. This allows developers to test new ideas and catch fundamental issues well before taking the vessel to sea. Closed-loop features can also be explored in simulation, provided the plant model is sufficiently accurate. Simulation can also be used to create synthetic datasets, which can then be fed into the resimulation software.

While simulation can save time and cost, all simulation models have limitations [45] [44]. They must be validated against real-world data to ensure that the behavior they represent is realistic. Consequently, simulation should be used where its accuracy has been confirmed and complemented with field testing where necessary.

Simulation also excels in exploring edge cases, rare, risky, or impractical situations to test in reality. For example, developers can simulate extreme weather conditions, sensor failures, or potential collisions to observe how the system reacts. Running these scenarios in software not only improves safety but also strengthens confidence in the system's ability to handle unexpected or critical events.

In the broader picture, combining resimulation, virtual testing, and targeted field trials provides a balanced workflow. It saves time and resources, increases repeatability, and enhances confidence in the system's behavior, making the entire development cycle more efficient and resilient.

9
Conclusions

This chapter summarizes the outcomes and reflections of the thesis, presenting the results achieved and discussing their implications for future marine ADAS/AD development. It begins by outlining the main findings and the significance of the proposed framework, followed by an evaluation of how the research questions were addressed. The chapter then highlights the current limitations, proposes directions for future work, including technical enhancements and methodological improvements and reflects on knowledge transfer within the Volvo Group. Finally, it concludes with a synthesis of how the work contributes to the broader goal of establishing a scalable, data-driven approach to verification and validation in the marine domain.

9.1 Results

This thesis proposes a data-driven development framework for the V&V of marine ADAS/AD, aiming to produce KPI-based evidence derived from real test data. The work has demonstrated the feasibility of the core steps, scenario design, instrumentation, data logging, and planning for KPI evaluation, while also highlighting the challenges of applying these methods in the maritime domain.

A key contribution of this work is the framework concept itself, which consolidates knowledge from both automotive and maritime domains into a structured approach that can guide future testing at Volvo Penta. Representative KPIs were defined and paired with test scenarios, and instrumentation with LiDAR, and DPS was deployed on a Volvo Penta test vessel. Dry runs and an initial on-water trial confirmed that logging of raw sensor data is technically feasible within the constraints of the test setup. These steps establish a proof of concept for a framework that can, once fully operationalized, provide consistent and reproducible evidence of feature performance.

Nevertheless, important elements remain incomplete. Ingestion into the DSL platform was not finalized, and the defined KPIs could not be computed from the logged datasets within the project timeframe. As such, the results represent an early stage rather than a finished framework. However, this documentation provides the design rationale, methodological steps, and identified challenges, all of which form a foundation for subsequent development.

9.1.1 Framework Significance and Outlook

Although several integration and evaluation steps remain incomplete, this thesis has proposed and partially demonstrated a V&V framework tailored for marine ADAS/AD development. The work establishes a structured process linking scenario definition, instrumentation, and raw data logging to a data-driven validation methodology. Its primary contribution lies in showing how established automotive verification practices from Volvo GTT can inform and be adapted to the marine context, where operational conditions and testing constraints differ substantially. As shown in Fig. 9.1, the framework organizes activities from scenario design and equipment setup through logging to analysis planning, providing a coherent basis for objective assessment.

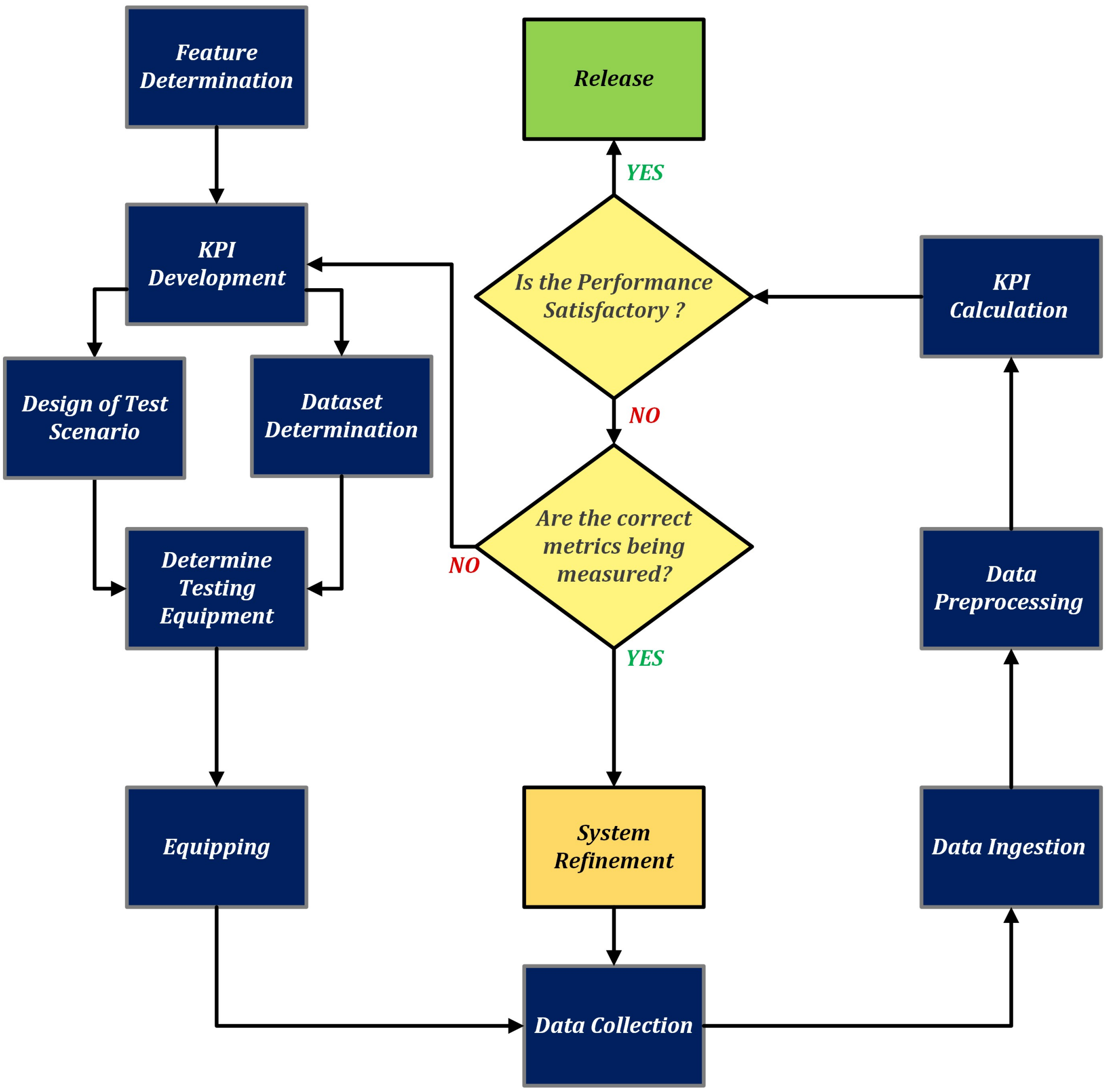


Figure 9.1: Overview of the proposed data-driven V&V framework for marine ADAS/AD, linking scenario definition, instrumentation, raw-data logging, and post-analysis planning.

The framework's significance lies not in completing all technical integrations, but in clarifying how a data-centric approach can be realized for marine automation. By structuring data capture around KPIs and scenario-based testing, the work outlines how objective, repeatable evidence can replace subjective judgments in feature validation. The instrumentation setup and initial dry-run testing verified that multi-sensor data logging on a vessel is feasible and can serve as the foundation for subsequent ingestion and analysis.
Future development should prioritize time synchronization across sensors, ingestion into the DSL environment, and automation of KPI computation, enabling systematic performance tracking and iterative feature improvement. Beyond these immediate tasks, extending the framework to support software-in-the-loop (SiL) evaluation and controlled scenario replays would further align Volvo Penta's validation practices with mature automotive V&V pipelines.
In conclusion, while the present work does not constitute a finalized framework, it provides a proof-of-concept and a clear methodological pathway. The approach offers Volvo Penta and the wider Volvo Group a foundation for scaling marine ADAS/AD validation through data-driven methods, supporting the long-term goal of safe and efficient automation at sea.

9.2 Answers to the Research Questions

Which KPIs best measure the system's performance, considering the end goal of the feature, and how effective are they at validating test results?

The KPIs selected in this thesis were chosen not only for their ability to quantify feature performance but also for their practicality in a development framework, where available sensor data and reproducibility of calculations are critical. The evaluation combines perception-level KPIs and task-level KPIs, thereby linking sensor interpretation with functional outcomes. The final set of KPIs is given in Table A.2.

At the perception level, false positives, false negatives, and object tracking consistency serve as core indicators. These metrics were selected because they are straightforward to calculate from logged sensor data and directly reflect the reliability of obstacle detection, a prerequisite for safe operation. Their interpretability is also a strength: false negatives highlight missed hazards that threaten safety, while false positives expose over-sensitive detections that undermine usability. However, their limitation is that they alone do not measure how the vessel reacts once an obstacle is detected, which is why they must be complemented with higher-level KPIs.

At the task level, KPIs such as final docking orientation error, docking time, and collision speed were defined. Orientation error links directly to user expectations of smooth and precise docking. Docking time and speed reduction profiles capture efficiency and controllability, but also illustrate trade-offs: minimizing time might conflict with ensuring smooth deceleration and safe approach speeds. Collision speed, if measured, provides a direct safety-critical indicator, translating errors in perception or control into an estimate of outcome severity. These task-specific

KPIs are effective because they are calculable from DPS and LiDAR data, which are consistently available in tests, and they tie results directly to safety and usability criteria that operators and regulators care about.

The strength of this KPI set lies in its complementarity. Perception KPIs (FP/FN rates) indicate how accurately the system sees the environment, while task KPIs (orientation, docking time, collision speed) indicate how well the vessel acts within it. Considering them together helps balance system tuning: for example, minimizing docking time must be weighed against keeping approach speed low enough to ensure safety, while maintaining orientation precision. These kinds of tensions between KPIs make them useful in practice, as they guide developers toward acceptable trade-offs rather than optimizing a single dimension in isolation.

In conclusion, the KPIs chosen measure performance effectively because they are (i) aligned with the end goals of safety, precision, and efficiency; (ii) practical to compute with available logged data (LiDAR, DPS); and (iii) complementary, highlighting trade-offs that are central to ADAS/AD tuning. While the framework has not yet demonstrated full computation of these KPIs, their formulation provides a strong basis for objective and reproducible evaluation in future iterations.

How scalable is the framework, given that the number, configuration, and type of sensors will increase in complexity as Volvo Penta advances its automation roadmap?

A central design consideration of the proposed framework is scalability, as sensor configurations and system complexity are expected to increase with Volvo Penta's future automation roadmap. While only a limited set of sensors could be integrated and tested within this thesis, the framework was conceptually structured to accommodate additional modalities such as marine radar, multibeam sonar, stereo vision, or advanced navigation feeds (e.g., AIS, ECDIS). In practice, integrating new sensors would require the development of tailored preprocessing modules for each data stream; however, the overall architecture is designed so that such extensions can be incorporated without restructuring the framework itself. This modularity ensures that the framework can evolve in step with sensor and feature development, provided the necessary preprocessing and quality checks are implemented.

Scalability is also supported at the scenario-definition level. By drawing inspiration from domain-specific scenario languages such as ASAM OpenSCENARIO [35] in the automotive sector, scenarios in this framework are defined at a higher level of abstraction, independent of individual sensor implementations. This makes them modular, reusable, and extendable to new conditions, thereby enabling systematic expansion of test coverage as sensor configurations or functionalities grow more complex. While this capability was not fully operationalized within the timeframe of this thesis, the underlying design principles provide a path toward a framework that remains flexible and extensible, aligning with the long-term needs of marine ADAS/AD validation.

What challenges arise when developing a framework for marine applications, given the established use of development frameworks for land vehicles?

Developing a framework for marine applications presented challenges that differ in important ways from land-based vehicle contexts. One of the most prominent issues was limited access to test platforms: unlike the automotive sector, which can rely on large proving grounds and test fleets, marine development is constrained by the availability of a small number of instrumented vessels. This limited the opportunity for repeated trials and emphasized the need for efficient use of each testing session.

A second challenge concerned sensor integration and mounting. Although the underlying technologies (e.g., LiDAR, DPS) are similar to those used in automotive, their installation on a marine vessel introduces unique complications. These include maintaining stable alignments despite hull motion, reflections from the water surface, and ensuring reliable connections in a less standardized environment compared to road vehicles

A third challenge lies in the relative immaturity of validation ecosystems in the marine domain. Automotive V&V benefits from decades of refinement, with well-established scenario standards (e.g., ASAM OpenSCENARIO [35]), dedicated pipelines for data ingestion and annotation, and proven practices for KPI-based evaluation. The marine sector lacks comparable infrastructures, meaning methods must be adapted or newly developed to ensure systematic and repeatable validation.

Finally, the regulatory framework differs significantly between domains. Whereas automotive development is guided by road traffic laws and UNECE regulations, marine testing must conform to COLREG and maritime safety practices. This requires embedding COLREG-compliant behaviors into test scenarios, a consideration that is domain-specific.

Together, these factors highlight that marine V&V is not simply a matter of transferring existing automotive practices but requires tailoring to domain-specific constraints in logistics, integration, methodological maturity, and regulatory context.

9.3 Shortcomings

This thesis established a proof-of-concept for a data-driven validation framework; however, due to a few constraints, several key elements were intentionally deferred. These elements represent concrete next steps for future work rather than limitations of the current study.

9.3.1 Data Ingestion

The thesis originally planned to perform data ingestion of raw log data into the DSL platform using the provided instructions and codebase. However, repeated JupyterHub environment errors (such as kernel instability and permission issues) prevented the pipeline from running, and the datasets were never ingested into DSL. This represents a significant shortcoming, as successful DSL ingestion would have established a single, versioned source of truth for all recordings. Such a centralized dataset (with standardized manifests and quality control checks) would make data more discoverable, auditable, and comparable over time, thereby streamlining KPI computations and future analyses. Revisiting the ingestion process under more stable conditions or with appropriate support remains an important step for ensuring proper integration of data into the system.

9.3.2 Data Preprocessing

Although extensive data preprocessing was planned, it was not implemented in practice. In real-world datasets, missing values, inconsistencies, and outliers are common, and leaving them unaddressed can complicate or delay subsequent analyses. Steps such as outlier removal, normalization, and labeling, which are typically critical for preparing the data, were not performed. Consequently, the preprocessing stage remained incomplete, representing a notable limitation in the current workflow.

9.3.3 KPI Computation

A detailed set of KPIs was defined to evaluate feature performance, including false positive rate, docking accuracy, and collision approach speed. However, these were not computed in practice, as the logged data was never processed into usable metrics. This left the framework without its central quantitative feedback loop. Implementing automated KPI computation remains essential to make the framework fully operational and to provide objective performance insights.

9.3.4 Iterative Development

The project was limited to a single on-water test, without the iterative test–analyze–refine cycles that are central to system development. As a result, KPIs could not be compared across multiple runs, constraining opportunities to tune parameters, expand scenario coverage, and strengthen logging reliability. More frequent testing, whether through additional vessel trials or complementary simulation, will be necessary to achieve robust refinement. Vessel availability and resource constraints further highlighted the importance of structured iteration for future work.

9.4 Future Work

This section outlines the steps needed to evolve the present proof of concept into a repeatable and scalable V&V process. The emphasis is on closing known gaps in (i) time alignment across heterogeneous sensors, (ii) ingestion and preprocessing suitable for analysis at scale, (iii) automated KPI computation and reporting, and (iv) iterative re-testing with comparable evidence across software versions and operating conditions [20, 43, 18].

9.4.1 Time synchronization

Accurate alignment of heterogeneous sensor streams is foundational for trustworthy correlation, fusion, and KPI calculation. Even small misalignments can affect the temporal ordering of events, bias latency estimates, and degrade comparisons across test runs. In practice, two complementary measures are required: (1) assign a common time base and record hardware or driver-level timestamps as close to signal generation as possible; and (2) persist synchronization metadata (e.g., offsets, lock status, drift/variance) and apply compensations during preprocessing so that all samples are brought to a shared reference before analysis [20, 43].
Within the proposed framework, the recommended approach is to designate a single time master on the vessel network, ensure that each logging interface preserves source timestamps, and verify end-to-end alignment with correlation checks (e.g., matching LiDAR detections against IMU yaw-rate transients during deliberate maneuvers). The ingestion/preprocessing stage should carry synchronization metadata alongside the raw data and make time-alignment a first-class quality check, consistent with data engineering practices used in autonomy DataOps pipelines [18, 43]. Robust time synchronization is thus a prerequisite for reproducible KPI computation and defensible regression testing.
An additional challenge that arises in distributed sensor logging architectures is the potential for time delays between data acquisition and logging. In theory, each intermediary device in the chain and the physical distance between sensors and the central logger can introduce latency. While modern high-speed Ethernet reduces these effects considerably, misalignments across data streams remain a concern when aiming for precise sensor fusion and KPI computation.

Because this thesis did not include timestamp synchronization or post-processing compensation, it was not possible to quantify the extent of such delays in the collected datasets. Nevertheless, prior work in automotive V&V has shown that even modest offsets can complicate the alignment of high-throughput streams such as LiDAR and DPS. To mitigate this, future implementations should incorporate hardware-level timestamping as close as possible to the point of data generation, combined with synchronization protocols such as PTP. Post-processing then becomes essential to compensate and align samples to a common time reference before analysis. Together, these steps ensure that sensor fusion and KPI calculations are not biased by logging-induced latencies.

9.4.2 Simulation and Resimulation

Although simulation and resimulation were introduced earlier in this work, their significance within the broader validation process deserves reflection. Simulation environments can, in the long term, complement real-world testing by expanding scenario coverage and supporting early-stage function evaluation. Resimulation, meanwhile, offers a way to reuse collected data for software comparison and iterative tuning once the logging and ingestion pipelines are fully operational. While these capabilities were not implemented within this thesis, they remain key future enablers that will enhance the efficiency and reproducibility of marine ADAS/AD validation.

9.4.3 Ground truth and references

Establishing reference data is key to objective performance evaluation. High-precision systems such as differential GPS/INS or dense LiDAR can provide centimeter-level positioning and detailed environmental mapping. Supplementary labeled datasets, for instance, annotated objects in a camera imager, would further strengthen the validation process. Together, these ground-truth resources would enable rigorous error analysis, provide benchmarking for system outputs, and form a foundation for both evaluation and future algorithm development.

9.4.4 Edge case identification

Future work should focus on systematically identifying and prioritizing edge cases, rare or safety-critical situations that do not occur often in trials but carry high risk. A practical method is mining logged datasets for statistical outliers and distributional tails (e.g., unusual velocities, near misses, abrupt bearing changes, or sensor dropouts), combined with rule-based detectors for safety surrogates (time-to-collision, minimum passing distance). Machine learning methods such as clustering, novelty detection, or isolation forests can surface atypical sequences for review and codification into scenario templates. Linking these edge cases to targeted KPIs ensures that validation focuses on risk-heavy situations, elevating the framework from general performance assessment to risk-informed verification.

9.4.5 Integration with safety standards

Aligning the framework with recognized standards makes results traceable and usable in regulatory and classification contexts. Concretely, this involves:

- **IMO MASS guidance:** Follow applicable IMO material on Maritime Autonomous Surface Ships (MASS) e.g., guidance documents and interim circulars on degrees of autonomy and operational considerations so that scenarios, controls, and pass/fail criteria can be traced to accepted practice [9, 38].
- **Classification guidance (DNV):** Organize the safety argument to meet autonomy guidance from classification societies; for DNV this includes AROS class notations and related recommendations on simulation-based evidence and structured assurance [7, 44].

- **Data and metadata standards:** Standardize logging/ingestion artefacts using relevant shipboard data standards (e.g., ISO 19847/19848 for data servers and data models) so that KPI computation and audit trails are reproducible across platforms and programs [53, 54].

Together, these elements help ensure that the framework remains focused, auditable, and compatible with regulatory and classification expectations.

9.5 Conclusion

The thesis proposed a practical framework for developing and validating marine ADAS/AD at Volvo Penta, even though some stages could not be completed within the project. The methodology, challenges and next steps have been documented, creating a solid foundation for future work. Beyond proving feasibility, the framework highlights how systematic and KPI-driven testing can close the gap between emerging marine autonomy and the safety, reliability, and trust needed for real-world adoption. It offers both engineers and decision-makers a structured path forward. One that balances innovation with the rigor required for safe deployment.

Bibliography

[1] Chalmers University of Technology, *Regulations for the use of AI tools*. Available at: https://www.chalmers.se/en/education/your-studies/masters-and-bachelors-thesis/regulations-for-the-use-of-ai-tools/ (accessed 15 Oct 2025).

[2] Gustaver, M. (2020). A Chalmers University of Technology Master's thesis template for LaTeX. Unpublished.

[3] RAND Corporation, "Autonomous vehicles would have to be driven billions of miles to demonstrate their reliability," Press release, Apr. 12, 2016. Available at: https://www.rand.org/news/press/2016/04/12.html.

[4] Autonomous navigation, "Towards simulation-based verification of autonomous navigation systems." Available at: https://www.sciencedirect.com/science/article/pii/S092575352030196X?via%3Dihub.

[5] Zereik, E., Bibuli, M., Mišković, N., Ridao, A., & Bruzzone, G. (2018). Challenges and future trends in marine robotics. *Annual Reviews in Control*, 46, 350–368. Available at: https://www.sciencedirect.com/science/article/pii/S1367578818300038.

[6] SAE International, "SAE Levels of Driving Automation." Available at: https://blog.ansi.org/ansi/sae-levels-driving-automation-j-3016-2021/#:~:text=Level%200%20%E2%80%93%20No%20Driving,Automation.

[7] DNV, "The AROS (Autonomous and Remotely Operated Ships) class notations." Available at: https://www.dnv.com/maritime/autonomous-remotely-operated-ships/aros-class-notation/.

[8] One Sea, "Autonomy in the Maritime Industry." Available at: https://b307d3dfff.clvaw-cdnwnd.com/ac8e62b9b6e49d3c818fcd9770855fb0/200000072-b3a53b3a55/POSITION%20PAPER%202023_One%20Sea%20-%20Autonomy%20in%20the%20Maritime%20Industry.pdf?ph=b307d3dfff.

[9] International Maritime Organization (IMO), "The International Maritime Organization defines four degrees of autonomy for Maritime Autonomous Surface Ships (MASS)." Available at: https://www.international-maritime-rescue.org/news/maritime-autonomous-surface-ships-mass-and-sar#:~:text=,board%20ready%20to%20take%20control.

[10] Lloyd's Register, "LR defines 'autonomy levels' for unmanned ships." Available at: https://thedigitalship.com/news/maritime-satellite-communications/lr-defines-autonomy-levels-for-unmanned-ships/#:~:text=,access%20required%20during%20a%20mission.

[11] Parker, A. (2021, Aug. 16). Marine Autonomy: The Future is Being Revealed. *MarineLink*. Available at: https://www.marinelink.com/news/marine-autonomy-future-revealed-489899.

[12] Wang, L., Chen, Y., Wang, H., & Liu, Z. (2023). Water Surface Targets Detection Based on the Fusion of Vision and LiDAR. *Sensors*, 23(4), 1768. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC9967045/.

[13] Li, Y., Zhou, X., & Zhang, J. (2024). Target Detection on Water Surfaces Using Fusion of Camera and LiDAR Based Information. *Computers, Materials & Continua*, 80(1), 467–486. Available at: https://www.techscience.com/cmc/v80n1/57387/html.

[14] Xie, J., et al. (2024). Reliable LiDAR-based ship detection and tracking for Autonomous Surface Vehicles in busy maritime environments. *Ocean Engineering*, 312, 119288. Project page: https://frl-ucl.github.io/.

[15] Shao, Y., et al. (2024). A robust method for multi-object tracking in autonomous ship navigation systems. *Ocean Engineering*, 311, 118560. Available at: https://www.sciencedirect.com/science/article/pii/S0029801824018985.

[16] Zhang, G., Wang, H., & Sun, C. (2021). Collision-avoidance navigation systems for Maritime Autonomous Surface Ships: A state of the art survey. *Ocean Engineering*, 235, 109380. Abstract via TRID: https://trid.trb.org/View/1863845.

[17] *Situational awareness systems and data in smart ships. Journal of Marine Science and Engineering*, 13(1), 122, 2025. Available at: https://www.mdpi.com/2077-1312/13/1/122.

[18] Microsoft, "Autonomous vehicle operations: DataOps reference architecture," 2023. Available at: https://learn.microsoft.com/en-us/industry/mobility/architecture/autonomous-vehicle-operations-dataops-content.

[19] *Maritime Autonomous Surface Ships: Architecture for Autonomous Navigation Systems*. Available at: https://www.mdpi.com/3129822.

[20] dSPACE, "Data-driven development pipelines: why current data logging approaches are not sustainable," webinar/learning page, 2023. Available at: https://www.dspace.com/en/inc/home/learning-center/recordings/video-elearning_datadrivenpipe.cfm.

[21] *Towards the Testing and Validation of Autonomous Ships: Design of a Variable Stability Ship Control System*. Available at: https://www.mdpi.com/2077-1312/11/7/1274#:~:text=autonomous%20ships%20are%20%E2%80%9Cconducted%20safely%2C,sea%20trials%20is%20developed%20in.

[22] SAE, "SAE J2944: Terms and Definitions Related to Driving Performance Measures (covers TTC, headway terminology)." Available at: https://www.auto-ui.org/docs/sae_J2944_PG_13-02-12.pdf.

[23] Survey/review on surrogate safety metrics and KPIs for ADAS/AD. Available at: https://www.researchgate.net/publication/385142308_Surrogate_Safety_Analysis_-_Leveraging_Advanced_Technologies_for_Safer_Roads.

[24] ISO 2631-1, Evaluation of human exposure to whole-body vibration (comfort/acc/jerk rationale). Available at: https://files.core.ac.uk/download/pdf/40036762.pdf.

[25] Shalev-Shwartz, S., et al. (2017). On a Formal Model of Safe and Scalable Self-driving Cars (RSS). Available at: https://arxiv.org/abs/1708.06374.

[26] ISO 26262 (functional safety) — motivates measurable acceptance criteria and evidence. Available at: https://www.diva-portal.org/smash/get/diva2:962406/FULLTEXT01.pdf.

[27] ISO 21448 SOTIF — focuses on performance/safety-of-intended-function, informs KPI definition and scenario adequacy. Available at: https://gooees.com/data/www.fusa-aspice.com.cn/static/upload/file/20240912/1726108615198033.pdf.

[28] California DMV disengagement reports (industry practice for interventions/disengagement KPIs). Available at: https://gooees.com/data/www.fusa-aspice.com.cn/static/upload/file/20240912/1726108615198033.pdf.

[29] Euro NCAP AEB protocol (residual impact speed, speed reduction, stopping distance). Available at: https://gooees.com/data/www.fusa-aspice.com.cn/static/upload/file/20240912/1726108615198033.pdf.

[30] Applied Intuition (2023). Quantifying knowns and unknowns: measuring progress for autonomy. Company blog. Available at: https://www.appliedintuition.com/blog/quantifying-knowns-and-unknowns.

[31] Zis, T. P. V., Psaraftis, H. N., & Reche-Vilanova, M. (2023). Design and application of a key performance indicator (KPI) framework for autonomous shipping in Europe. *Maritime Transport Research*, 5, 100095. Available at: https://doi.org/10.1016/j.martra.2023.100095.

[32] Cai, J., Deng, W., Guang, H., Wang, Y., Li, J., & Ding, J. (2022). A Survey on Data-Driven Scenario Generation for Automated Vehicle Testing. *Machines*, 10, 1101. Available at: https://doi.org/10.3390/machines10111101.

[33] Lloyd's Register, *Maritime Autonomous Surface Ships (MASS) — Creating a framework for efficiency, safety and compliance*. Available at: https://maritime.lr.org/MASS/report.

[34] Maritime Industries UK, *Assuring the successful deployment of Maritime Autonomous Systems — A data-driven approach*. Available at: https://www.maritimeindustries.org/application/files/3316/8078/8113/1530-1545_-_Andre_Burgess.pdf#:~:text=,needed%20for%20Maritime%20AI%2FML%20systems.

[35] ASAM OpenSCENARIO DSL 2.1.0, Dynamic scenario description, 2024-03-07. Available at: https://www.asam.net/standards/detail/openscenario-dsl/#:~:text=validation%20,include%20maneuvers%20that%20involve%20multiple.

[36] Porres, A., et al. (2020). Scenario-based testing method for ship collision avoidance systems. (Cited via JMSE review.)

[37] Lazarowska, D. (2021). Generation of head-on and crossing encounter scenarios for path planning evaluation. (Cited via JMSE review.)

[38] Lloyd's Register (2025). *Maritime Autonomous Surface Ships (MASS): assurance for complex autonomous systems.* Research report. Available at: https://www.lr.org/en/knowledge/research-reports/2025/maritime-autonomous-surface-ships-mass/.

[39] Maritime UK (Maritime Industries) (2023). *Assurance and validation for Maritime AI/ML systems.* Industry report. Available at: https://www.maritimeindustries.org/download_file/view/2726/2054.

[40] Son, Y., et al. (2018). A simulation-based testing and validation framework for ADAS development. (Amesim + PreScan co-simulation). Summary via Scipedia: https://www.scipedia.com/public/Son_2018a.

[41] NASA, "Runtime verification and validation of autonomous flight operations," Technology Transfer (TOP2-320). Available at: https://technology.nasa.gov/patent/TOP2-320.

[42] Nordahl, H., et al. (2023). Competitiveness and societal KPIs for autonomous ship concepts. Available at: https://www.sciencedirect.com/science/article/pii/S2666822X2300014X.

[43] Hilgendorf, M. (2022). *Efficient industrial big data pipeline for lossless transfer of vehicular data.* Master's Thesis, Chalmers University of Technology, Gothenburg, Sweden. Available at: https://odr.chalmers.se/handle/20.500.12380/305885.

[44] DNV, *Proving the case for simulation-based testing.* Available at: https://www.dnv.com/research/review-2023/featured-projects/proving-the-case-for-simulation-based-testing/ (accessed 3 Nov 2025).

[45] Box, G. E. P. (1976). Science and Statistics. *Journal of the American Statistical Association*, 71(356), 791–799. Available at: https://www.tandfonline.com/doi/abs/10.1080/01621459.1976.10480949 (accessed 3 Nov 2025).

[46] Google SRE, *Monitoring Distributed Systems (SRE Book).* Available at: https://sre.google/sre-book/monitoring-distributed-systems/ (accessed 3 Nov 2025). Accessed: 3 Oct 2025.

[47] IEEE, *IEEE 1588: Precision Time Protocol (PTP) — Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems*, 2019.

[48] Industry Consortium, *Best Practices for Deploying IEEE 1588 PTP in Time-Sensitive Networks*, 2022. Available at: https://example.org/ptp-best-practices (accessed 3 Nov 2025).

[49] Timing Systems Inc., *A Field Guide to PTP: Grandmasters, Boundary and Transparent Clocks*, 2021. Available at: https://example.org/ptp-field-guide (accessed 3 Nov 2025).

[50] GNSS Solutions, *GNSS-Disciplined Time: Using 1PPS and Time-of-Day for System Synchronization*, 2020. Available at: https://example.org/gnss-pps (accessed 3 Nov 2025).

[51] Automotive Tool Vendor, *Time-Synchronous Data Logging for ADAS/AD Validation*, Whitepaper, 2023. Available at: https://example.org/adas-logging (accessed 3 Nov 2025).

[52] Research Note, *Time Alignment and Metadata in DataOps Pipelines for Autonomous Systems*, 2024. Available at: `https://example.org/dataops-timesync` (accessed 3 Nov 2025).

[53] ISO, *ISO 19847: Shipboard data servers to share field data on the ship*, 2018. (Specifies requirements and interfaces for shipboard data servers.)

[54] ISO, *ISO 19848: Standard data for shipboard machinery and equipment*, 2018. (Defines a common data model and naming for shipboard data exchange.)

A

Appendix

Table A.1: Initial KPI Set

KPI	**Description**
Tracking Consistency	Ability of the system to reliably track moving objects over time without losing them or drifting off target.
System Uptime	Percentage of time the autonomous system remains active and available for operation.
Movement Tolerance	System's ability to maintain stable performance despite vessel motion under varying sea conditions.
Global Positioning Accuracy	Precision of the vessel's reported position in global (latitude/longitude) or local coordinates.
Sensor Latency	Time delay between an object's appearance and its detection by the system.
Sensor Performance in Low Visibility	Effectiveness of the sensors in adverse conditions such as fog, rain, or low light.
Network Latency	Communication delay between onboard components transmitting data.

Table A.2: Selected KPI Set

KPI	Description
False Positive Rate (FPR)	How often the system reports an obstacle that isn't there (e.g., wake or spray misread as an object).
False Negative Rate (FNR)	How often real, visible obstacles are missed—safety–critical as it can lead to collisions.
Object Tracking Consistency	Whether a single real object stays as one continuous track across frames (no ID flips, duplicates, or lost–then–new IDs after brief gaps).
Docking Spot Determination Rate	Accuracy of identifying valid docking spots and keeping them stable long enough to be usable by the pilot.
Final Orientation at Berth	Alignment to the dock at completion; average orientation error within a short hold window should stay below a small bound (e.g., 2°).
Approach Speed Profile	Speed tapers smoothly as distance to dock decreases (no sustained speed rises near the dock; smoothness checked via jerk limits).
Collision/Contact Relative Speed	Relative speed at first contact or minimum range remains under a safety cap (e.g., 0.5 m/s for low-speed docking).
Total Docking Time (TDT)	Time from docking start to the first stable final pose (meets angle/speed/offset limits over a short hold period).

Table A.3: Autonomy classification frameworks across domains

Table 1: Autonomy classification frameworks across domains

Framework	Level / Code	Summary
	SAE (J3016) — Road Vehicles	
SAE J3016	Level 0	Basic features like emergency braking, blind-spot warning *(Driver Assistance)*.
SAE J3016	Level 1	Lane centering *or* adaptive cruise control.
SAE J3016	Level 2	Lane centering *and* adaptive cruise control.
SAE J3016	Level 3	Drives in limited conditions (e.g., traffic jams); human fallback required *(Automated Driving)*.
SAE J3016	Level 4	Driverless within its ODD (operational design domain); may have no pedals.
SAE J3016	Level 5	Full automation in all conditions.
	IMO — Maritime	
IMO	Degree 1	Crew onboard; some tasks automated.
IMO	Degree 2	Crew onboard; ship can be remotely operated.
IMO	Degree 3	No crew onboard; remotely operated vessel.
IMO	Degree 4	Fully autonomous; no humans needed.
	DNV — Maritime	
DNV	M (Manual)	Manually operated.
DNV	DS (Decision Supported)	Human makes decisions with system support.
DNV	DSE (System decision supported)	System proposes and can conditionally execute; human confirmation required.
DNV	SC (Self-controlled)	System executes and controls itself but can still seek human approval for certain actions.
DNV	A (Autonomous)	Fully autonomous.
	Lloyd's Register — Maritime	
Lloyd's Register	AL 0	Manual control.
Lloyd's Register	AL 1	Human operated; options suggested by onboard systems.
Lloyd's Register	AL 2	As AL 1, but decision support may be offboard.
Lloyd's Register	AL 3	System makes decisions/actions with human supervision.
Lloyd's Register	AL 4	Autonomous with human supervision and possible intervention.
Lloyd's Register	AL 5	Autonomous with very limited human supervision/intervention.
Lloyd's Register	AL 6	Fully autonomous.

Framework	Level / Code	Summary
	Bureau Veritas — Maritime	
Bureau Veritas	Degree A0 (Human operated)	Manual control (humans onboard).
Bureau Veritas	Degree A1 (Human directed)	System processes info; cannot execute actions (humans on/off board).
Bureau Veritas	Degree A2 (Human delegated)	Processes info and executes, but only with human confirmation (on/off board).
Bureau Veritas	Degree A3 (Human supervised)	Processes info and executes; humans can interrupt as needed (on/off board).
Bureau Veritas	Degree A4 (Full Automation)	Broad execution autonomy; humans can still interrupt if needed (on/off board).

	One Sea — Maritime	
One Sea	Level 0	Human controlled.
One Sea	Level 1	**HON–EON–MON**: *Hands ON, Eyes ON, Mind ON.* Human steers, watches, and decides; automation assists only.
One Sea	Level 2	**HOFF–EON–MON**: *Hands OFF, Eyes ON, Mind ON.* System actuates; human monitors and decides.
One Sea	Level 3	**HOFF–EOFF–MON**: *Hands OFF, Eyes OFF, Mind ON.* System actuates and perceives; human remains decision authority/supervisor.
One Sea	Level 4	**HOFF–EOFF–MOFF**: *Hands OFF, Eyes OFF, Mind OFF* within defined scope/area; system handles operation; human not needed in the loop.
One Sea	Level 5	Fully autonomous (all functions, all conditions).

A.1 Switch Console Setup

The steps below describe console access and basic port-mirroring configuration using PuTTY.

Accessing the Switch through the Console Port

1. Connect to the console port of the switch using a USB Type A/C cable.
2. Download and install the required driver.
3. Open PuTTY and configure it with the settings shown in Table A.3.

Table A.3: Console Port Parameter Settings

Parameter	**Value**
Data rate	115200 bits/s
Data bits	8
Stop bits	1
Parity	Off/None
Flow control	Off/None

```
admin
admin

configure
monitor 1
no source port eth6
source port eth6 ingress
destination port eth7
enable
exit
```

In this setup, `eth6` is the source (monitored) port connected to the onboard computer, while `eth7` is the destination port connected to the logger.

Department of Electrical Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY