

Data-Driven Vessel Design

Data-driven operational profiles
used for design input for new HTV designs

Floor Hartjes

Delft University of Technology



Thesis for the degree of MSc in Marine Technology in the specialization of Ship
Design

Data-Driven Vessel Design

Data-driven operational profiles
used for design input for new HTV designs

by

Floor Hartjes

Performed at

Ulstein Design & Solutions B.V.

This thesis MT.25/26.025.M is publicly available in accordance with the general conditions for projects
performed by the TU Delft.

To be defended publicly on 17/04/2026.

Company Supervisor

Responsible supervisor: J.D. Stroo
Additional supervisor: J.J. de Ridder

Thesis Exam Committee

Responsible professor: J.L. Gelling
Chair: E. van Hassel
Staff Member: R. de Winter
Staff Member: M. van Koningsveld
Company member: J.D. Stroo

Author details

Student: Floor Hartjes
Student number: 4299438
Faculty: Faculty of Mechanical Engineering
Master: MSc Marine Technology

Cover Image of HX121 by Ulstein Design & Solutions B.V. (HX121 - Ulstein Design & Solutions B.V., 2023)
Style TU Delft Report Style, adapted by Daan Zwaneveld



ULSTEIN®

Abstract

Early-stage vessel design is traditionally based on assumed operational profiles, introducing uncertainty and potentially leading to suboptimal design decisions. This study presents a data-driven approach to vessel design by deriving operational profiles from Automatic Identification System (AIS) data, with a specific focus on Heavy Transport Vessels (HTVs).

A structured methodology was developed to process, clean, and segment AIS data into operational activities, which were subsequently used to construct representative operational profiles. In contrast to existing studies, this research evaluates the impact of data cleaning and repairing procedures on both the resulting operational profiles and their impact on vessel design.

The results show that data cleaning and repairing can significantly influence operational profiles when analysing individual vessels. However, when multiple vessels are combined into fleet-level operational profiles, these effects decrease, indicating that the importance of cleaning and repairing is reduced at fleet level. Furthermore, the impact of data-derived operational profiles on early-stage vessel design is found to be limited. For HTVs in particular, key design parameters are primarily constrained by monopile dimensions, reducing the sensitivity of design outcomes to variations in operational profiles.

This study demonstrates that AIS-based operational profiling is a reliable method for supporting early-stage vessel design. It further shows that increasing dataset size is more effective in improving robustness than increasing data cleaning complexity, thereby providing practical guidance on the required level of data processing precision. Overall, this research provides a framework for integrating AIS data into early-stage vessel design and demonstrates the influence of data cleaning and repairing at multiple levels, as well as the impact of data-derived operational profiles on vessel design.

Preface

This thesis marks the final stage of my Master's degree in Marine Technology at Delft University of Technology. Over the past months, I have worked on developing a data-driven approach to vessel design, focusing on the use of AIS data to construct operational profiles for HTVs. This research allowed me to critically reflect on commonly used assumptions in early-stage design and on the influence of data cleaning and repairing.

I would like to express my sincere gratitude to my daily supervisor, Jaap, for his guidance, support, and constructive feedback throughout this project. I greatly appreciated not only his help on the technical and research aspects, but also his support in navigating the challenges along the way and continuously striving to make the most out of this project.

I would also like to thank Ko for his guidance and for providing the freedom and trust to explore both academic and personal development during this research. In addition, I am grateful to Bert and Rens from Made Smart Group for giving me access to their tools and sharing valuable insights into AIS data, which significantly strengthened this research.

Furthermore, I would like to thank Edwin for his critical perspective as chair, and Jesse for his feedback and support, not only on the blended design aspect but also on the graduation process as a whole.

Finally, I would like to thank my family and friends for their continuous support throughout this period. Their encouragement, patience, and willingness to listen helped me through the more challenging moments of this thesis. I am especially grateful for the balance they provided through both support and distraction.

This thesis represents not only an academic achievement, but also a period of personal growth. I have learned how to deal with setbacks, take ownership of my work, and make the most of the freedom and responsibility given to me. Without the support, trust, and encouragement of the people around me, this would not have been possible.

*Floor Hartjes
Delft, April 2026*

Contents

Preface	ii
Nomenclature	x
I Introduction	1
1 Introduction	1
II Literature	4
2 Data Driven Design	4
2.1 The role of operational profiles in early stage design	4
2.2 Current applications of data within ship design	5
2.3 Definition of Operational Profiles	5
2.4 Conclusion Data Driven Design	5
3 AIS Data	7
3.1 Overview and functioning of AIS	7
3.2 AIS data characteristics	7
3.3 Data quality and reliability	8
3.4 Preprocessing of AIS data	8
3.5 Conclusion AIS data	9
4 MSG Tools	10
4.1 MSG Tools	10
4.2 Data Retrieval	11
4.3 Data Export	12
5 Blended Design	13
5.1 Design Tool	13
5.2 Blended Design Methodology	13
5.3 Operational profile in Blended Design	15
5.4 Opportunities and Challenges	16
5.5 Conclusion Blended Design	16
6 Heavy Transport Vessel Market	17
6.1 HTV compared to Semi-Submersibles	17
6.2 Market Demand	18
6.3 Market Supply	18
6.4 Market Fit	19
6.5 Conclusion Heavy Transport Vessel Market	20
III Method	21
7 Methodology	21
7.1 Fleet selection	21
7.2 Data Collection from Made Smart Group	22
7.3 Data Preparation	24
7.3.1 Interpretation of the AIS data	24
7.3.2 Dataset 1 (All Data)	25
7.3.3 Interpolated Data	26
7.3.4 Comparisons	29
7.4 Operational profile	32
7.4.1 Unsegmented method	33

7.4.2	Segmented method	34
7.5	Conclusion Methodology	39
8	Results	40
8.1	Fleet Selection Results	40
8.2	Data Collection from MadeSmart Group Results	41
8.3	Data Preparation Results	41
8.3.1	Coverage of Dataset 1 results	41
8.3.2	Dataset 2 (Interpolated) procedure results	42
8.3.3	Comparison between datasets results	45
8.4	Operational Profile Results	48
8.4.1	Unsegmented method results	48
8.4.2	Comparison between HTVs and Semi-Submersibles operational profiles with the unsegmented method	53
8.4.3	Segmented method results	53
8.4.4	Comparison between HTVs and Semi-Submersibles operational profiles with the segmented method	56
8.4.5	Trip distance analyses	58
8.4.6	Comparison between methods for data-driven operational profiles	60
8.5	Verification and Validation of the Segmentation Method	62
8.5.1	Verification Results	62
8.5.2	Validation Results	63
8.6	Conclusion Results	63
9	Blended Design as early stage design tool	64
9.1	Operational profiles in Blended Design	64
9.2	Ship design from Blended Design	66
9.3	Conclusion Blended Design as early stage design tool	69
10	Additional ship insight through AIS data	70
10.1	Monopile fabrication site visits by the vessels in this research	70
10.2	Trends over the years for the HTV fleet	71
10.3	Vessel characteristics compared to operations	73
10.4	Ship insights conclusion	74
IV	Conclusion	75
11	Conclusion	75
12	Discussion	78
13	Recommendations	80
	References	81
A	Blended Design additional information	85
B	Settings used in Prospector by MSG to collect the AIS data	90
C	Overview of Python-Based Workflows Supporting the Methodology	92
D	Operations within the offshore wind market	96
E	Speed Distribution Plots	98
E.1	Speed Distribution Audax	98
E.2	Speed Distribution Bigroll Bering	101
E.3	Speed Distribution Bigroll Beaufort	103
E.4	Speed Distribution Biglift Baffin	105
E.5	Speed Distribution Biglift Barentsz	107
E.6	Speed Distribution Dongbang Giant No.6	109
E.7	Speed Distribution Dongbang Giant No.7	111
E.8	Speed Distribution Fan Zhou 6	113
E.9	Speed Distribution Fan Zhou 8	115

E.10 Speed Distribution Mega Caravan	116
E.11 Speed Distribution Zhong Yu 188	118
E.12 Speed Distribution GPO Grace	119
E.13 Speed Distribution Blue Marlin	121
E.14 Speed Distribution Target	123
E.15 Speed Distribution Hua Hai Long	125
E.16 Speed Distribution Fan Zhou 10	127
E.17 Speed Distribution Sun Shine	128
F Interpolation inconsistency example of Biglift Baffin 2019	129
G Speed plot comparison between Dataset 2 and 4	131
H Results for Semi-Submersibles with Dataset 2 as input for the speed bins	144
I Operational Profile Using Segmented Method: Comparison Between Dataset 2 and Dataset 4 for Biglift Baffin	145
J Blended Design Additional Results	147
K Monopile Fabrication Sites	156

List of Figures

1.1	Overview of the workflow from raw AIS data to early-stage vessel design, highlighting existing industry components and the contributions of this research.	2
2.1	Uncertainties for ship design [9]	4
3.1	Functionality of AIS [19]	7
3.2	Data sent and received by AIS	8
3.3	Most common preprocessing topics and techniques for AIS data	9
5.1	Principle idea of Blended Design by J.D. Stroo	14
5.2	Transforming trend lines to probability functions by Zwaginga [1]	14
5.3	Visual guide for interpreting result plots by de Ridder [28]	15
5.4	Typical operations of a HTV	15
6.1	Comparison between HTV and Semi-Submersible vessels	18
6.2	Filtering steps for supply list	19
6.3	Supply of currently operated HTVs by year	19
7.1	Filtering steps for supply list	21
7.2	Preview of a query in Prospector from MADE SMART GROUP [4]	22
7.3	Data variables provided when using the All Data option in Prospector by MADE SMART GROUP [4]	23
7.4	Example of data collection using Plotter from MADE SMART GROUP [4]	24
7.5	Cleaning procedure	26
7.6	Thresholds used to determine the category of gaps in the data	28
7.7	Data stream	30
7.8	Example of cumulative SOG distribution and ranked SOG graphs	32
7.9	Methods used to obtain data-driven operational profiles	32
7.10	Segmentation Rules Level 1	35
7.11	Segmentation Rules Level 2	35
7.12	Segmentation Rules Level 3	36
7.13	Segmentation Rules Level 4	36
7.14	Segmentation rules level 5	37
7.15	Operational profiles with different vessel input	38
7.16	Global Methodology	39
8.1	Consecutive data points in Biglift Baffin 2019 interpolated data	43
8.2	SOG CDF for Biglift Baffin comparing Dataset 2 and Dataset 4	47
8.3	Ranked speed distribution for Biglift Baffin comparing Dataset 2 and Dataset 4	47
8.4	Maximum endurance for the HTVs determined with the unsegmented method	49
8.5	Visual Plotter overview of Fan Zhou 10 in 2022	52
8.6	Distribution of operational modes for HTVs and SS	55
8.7	Distribution of operational modes for HTVs and Semi-Submersibles	57
8.8	Distribution based on segments and duration for HTVs using Dataset 4	59
8.9	Duration spent within distance bins per trip based on Dataset 2 for HTVs	59
8.10	Duration spent within distance bins per trip based on Dataset 4 for Semi-Submersibles	60
8.11	Unsegmented compared to segmented on sailing and non-sailing distribution	61
9.1	OP 2 D4 HTV compared with OP 3 D2 HTV for optimal design speed	67
9.2	OP 2 D4 HTV compared with OP 4 D4 SS for optimal design speed	67

9.3	OP 2 D4 HTV compared with UDSBV 1 and 2 for vessel length	68
9.4	OP 2 D4 HTV compared with UDSBV 1 and 2 for vessel breadth	68
9.5	OP 2 D4 HTV compared with UDSBV 1 and 2 for design speed	69
10.1	Duration spent in EEW Rostock port by GPO Grace	71
10.2	Median SOG for HTVs per year	72
10.3	Average distance sailed per year for the HTV fleet	72
10.4	Comparison between time spent sailing near AIS-derived cruising speed and design speed	73
A.1	Inputs for Blended Design	86
A.2	Example of how ship configurations are generated in Blended Design	87
A.3	Overview of the ship modules in Blended Design	87
A.4	Market module forecast by Zwaginga [1]	88
A.5	A loop trough Blended Design	89
B.1	Download type settings	90
B.2	Download format settings	90
B.3	Download settings	91
B.4	Download filter settings	91
C.1	Workflow for converting raw AIS text files into standardized CSV files and merging vessel data based on the selected years from the input matrix.	92
C.2	Computation of AIS-based metrics, including time differences, travelled distance, and calculated speed between consecutive points.	93
C.3	Three-step cleaning process applied to the merged AIS dataset, including removal of invalid data, filtering based on last valid points, and consistency checks between speed and distance.	94
C.4	Classification of temporal gaps into stationary, moving, and unknown categories based on vessel behaviour before and after the gap.	95
C.5	Gap repair process where stationary gaps are filled with fixed-position points and moving gaps are interpolated between known positions.	95
G.1	CDF comparison between Dataset 2 and 4 for all vessels	131
G.2	Speed ranked plot comparison between Dataset 2 and 4 for all vessels	137
J.1	Comparison between Blended Design results with different endurance	147

List of Tables

7.1	Input settings for data downloads used in this research	23
7.2	Interpretation of the data	25
7.3	Data reliability criteria for the All Data source	25
7.4	Comparison expectations	31
7.5	Comparison of the cleaned datasets	31
7.6	Comparison of gap statistics	31
7.7	Output unsegmented operational profile speed distribution	33
7.8	Output of segmented operational profile based on modes	39
7.9	Distance statistics per trip for the Biglift Baffin in 2019	39
8.1	List of vessels used in this research	40
8.2	Datasets collected for this research	41
8.3	Datasets that do not fulfill the initial thresholds	42
8.4	Distance comparison between Dataset 1 and Dataset 2 for the Biglift Baffin in 2019	42
8.5	Mismatch results for all data combined	43
8.6	Mismatches occurring in certain SOG ranges for all data	44
8.7	Example from Bigroll Bering 2023 showing an unexplained mismatch	44
8.8	Example of a stationary gap of Mega Caravan in 2023	44
8.9	Example of a moving gap in the AIS data of Mega Caravan in 2023	45
8.10	Gap identification	45
8.11	Comparison between all datasets used in this research	46
8.12	Comparison of the cleaned datasets	46
8.13	Comparison between datasets for gap statistics	46
8.14	Average sailed distance comparison between Dataset 2 and Dataset 4 for HTVs (unsegmented)	48
8.15	Maximum endurance comparison between Dataset 2 and Dataset 4 for HTVs (unsegmented)	49
8.16	Comparison of the specific maximum endurance between Dataset 2 and Dataset 4 for HTVs (unsegmented)	49
8.17	Median SOG based on different ranges for HTVs using Dataset 4	50
8.18	Median SOG using 1–20 kn for HTVs comparing Dataset 2 and Dataset 4	50
8.19	HTV Dataset 4 speed distribution	50
8.20	HTV Dataset 2 speed distribution	51
8.21	Average sailed distance comparison between Dataset 2 and Dataset 4 for Semi-Submersibles	51
8.22	Maximum endurance comparison between Dataset 2 and Dataset 4 for Semi-Submersibles	51
8.23	Comparison of the specific maximum endurance between Dataset 2 and Dataset 4 for Semi-Submersibles	51
8.24	Median SOG based on different ranges for Semi-Submersibles using Dataset 4	52
8.25	Median SOG using 1–20 kn for Semi-Submersibles comparing Dataset 2 and Dataset 4	52
8.26	Semi-Submersibles Dataset 4 speed distribution	53
8.27	Comparison between operational profiles for HTVs and Semi-Submersibles using Dataset 4 and the unsegmented method	53
8.28	Comparison for segmented method between Dataset 2 and Dataset 4 for HTVs	54
8.29	Comparison for segmented method between Dataset 2 and Dataset 4 for Semi-Submersibles	54
8.30	Endurance comparison for HTVs between Dataset 2 and Dataset 4	54
8.31	Endurance comparison for Semi-Submersibles between Dataset 2 and Dataset 4	55
8.32	Mean sailing distance comparison for HTVs between Dataset 2 and Dataset 4	55
8.33	Mean sailing distance comparison for Semi-Submersibles between Dataset 2 and Dataset 4	56

8.34	Mean duration comparison for HTVs between Dataset 2 and Dataset 4	56
8.35	Mean duration comparison for Semi-Submersibles between Dataset 2 and Dataset 4 . .	56
8.36	Comparison between HTVs and SS for endurance using the segmented method	57
8.37	Comparison between HTVs and SS for moored duration using the segmented method .	57
8.38	Comparison between HTVs and SS for mean sailing distance using the segmented method	58
8.39	Comparison of sailed distance between the unsegmented and segmented methods . .	61
8.40	Comparison of endurance between the unsegmented and segmented methods	61
8.41	Comparison between the possibilities of the unsegmented and segmented methods . .	62
8.42	Verification of the segmentation method	62
8.43	Snapshot of dataset 2 (Interpolated) from Biglift Baffin 2019 containing interpolation mis- take	63
9.1	Input Blended Design	65
9.2	Ship design output from Blended Design	66
10.1	AIS-derived monopile transport performed by HTVs	71
10.2	AIS-derived monopile transport performed by Semi-Submersibles	71
10.3	Comparison between endurance based on specification sheets and AIS data	74
11.1	Summary of impact of cleaning and repairing and data-derived operational profiles . . .	77
D.1	Operations within the offshore wind market	96
D.2	Operations in the offshore wind market with vessels outside of this research	97
F.1	Snapshot of Dataset 2 (Interpolated) from Biglift Baffin 2019 containing an interpolation mistake	129
F.2	Snapshot of Dataset 1 from Biglift Baffin 2019 at the same timestamp as the interpolation mistake in Dataset 2	130
F.3	Comparison between Dataset 1, Dataset 2, and Dataset 2.2 cleaned only on SOG cal- culation for Biglift Baffin 2019	130
H.1	Semi-Submersibles Dataset 2 speed distribution	144
I.1	Comparison of vessel-level operational profile between Dataset 2 and Dataset 4 for the Biglift Baffin (2019)	145
I.2	Time distribution comparison between Dataset 2 and Dataset 4 for the Biglift Baffin (2019)	145
I.3	Mean value comparison between Dataset 2 and Dataset 4 for the Biglift Baffin (2019)	146
K.1	Monopile Fabrication Site	157

Nomenclature

Abbreviations

Abbreviation	Definition
AIS	Automatic Identification System
B	Breadth
COG	Course Over Ground
CSV	Comma-Separated Values
D	Depth
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DP	Dynamic Positioning
DWT	Deadweight Tonnage
ECDIS	Electronic Chart Display and Information System
ETA	Estimated Time of Arrival
FloFlo	Float-on/Float-off
HDG	Heading
HTV	Heavy Transport Vessel
KML	Keyhole Markup Language (Google Earth)
LOA	Length Overall
MMSI	Maritime Mobile Service Identity
MSG	Made Smart Group
OPEX	Operational Expenditure
QC	Quality Control
ROI	Return on Investment
RoRo	Roll-on/Roll-off
ROT	Rate of Turn
SOG	Speed Over Ground
TXT	Tab-Separated Text
UDSBV	Ulstein Design & Solutions B.V.
UTC	Coordinated Universal Time

Symbols

Symbol	Definition	Unit
d	Haversine distance	[m]
f	Interpolation factor	[-]
R	Earth radius	[m]
λ	Longitude	[rad]
φ	Latitude	[rad]

I. Introduction

1

Introduction

In early-stage vessel design, operational profiles are used to describe how a vessel is expected to operate throughout its lifetime. These profiles typically include parameters such as sailing distance, port time, loading and unloading duration, waiting time, and speed distributions. Such operational characteristics play a critical role in determining key vessel characteristics, including dimensions, speed, and cargo capacity. However, in current practice, these profiles are largely based on assumptions rather than data. This reliance on assumed operations may lead to designs that are not fully aligned with real-world vessel behaviour, resulting in over- or under-engineered vessels.

Ulstein Design & Solutions B.V. (UDSBV), together with several MSc students, developed the early-stage design method *Blended Design* [1], which integrates vessel design with business case evaluation. This approach enables designers to explore vessel concepts under varying market scenarios and assess their performance over the vessel's lifetime. Despite its strengths, the operational inputs within this framework are still primarily assumption-based, introducing uncertainty in the resulting designs and highlighting the need for more realistic, data-driven operational input.

AIS data offers a promising opportunity to capture real vessel behaviour. AIS provides high-resolution information on vessel position, speed, and movement over time and is available for a large number of vessels worldwide. Since 31 December 2004, international regulations have mandated the use of AIS for ships over 300 gross tonnage on international voyages and cargo ships over 500 gross tonnage [2]. While AIS data has been widely applied in fields such as traffic analysis, route prediction, and anomaly detection, its potential for informing early-stage vessel design remains largely unexplored.

A key challenge in utilising AIS data for this purpose is the presence of noise, missing values, and inconsistencies within the dataset. These issues introduce uncertainty, making it unclear how data preprocessing influences the resulting operational profiles and their application in vessel design. In addition, vessel trajectories must be translated into meaningful operational states through segmentation based on operational modes.

To demonstrate the framework of data-driven operational profiles, this research focuses on HTVs. The offshore wind industry has experienced rapid growth in recent years, driven by global ambitions to achieve net-zero emissions by 2050 [3]. HTVs play a key role in this transition by transporting large offshore wind components such as monopiles and transition pieces. As demand for offshore wind increases, operational requirements for HTVs are becoming more complex and dynamic, making it increasingly challenging to design vessels based on static assumptions.

This research follows a structured, end-to-end approach that links AIS data to early-stage vessel design, while building upon existing tools and datasets. Processed AIS data is obtained through Made Smart Group (MSG) [4], which provides interpolated vessel datasets via the Prospector tool. Within this context, this research evaluates data quality by applying additional cleaning and repairing steps to address remaining inconsistencies.

The influence of these preprocessing steps are assessed at multiple levels to determine their impact on the resulting operational profiles and early-stage vessel design.

Subsequently, methods are developed and evaluated to derive operational profiles from AIS data, including both segmentation-based and non-segmentation-based approaches. These methods are compared to assess their ability to capture vessel operations with sufficient detail for design applications. The resulting operational profiles are analysed at both vessel and fleet level and are compared with the assumed operational profiles currently used by UDSBV for the HTV design.

Finally, the derived operational profiles are adapted for use within the Blended Design framework developed by UDSBV, enabling the evaluation of their impact on vessel design characteristics and economic performance, including vessel dimensions and Return on Investment (ROI). Therefore, the research establishes a direct link between AIS data processing and early-stage vessel design.

The main contribution of this research lies in the integration and evaluation of data-driven operational profiles within early-stage vessel design. By analysing the influence of data preprocessing, comparing different methods for deriving operational profiles, and assessing their impact on vessel design, this study provides a structured approach to incorporating AIS data into design practice.

To provide a clear overview of the overall workflow and to position the contribution of this research within the broader design process, Figure 1.1 illustrates the steps from raw AIS data to early-stage vessel design.

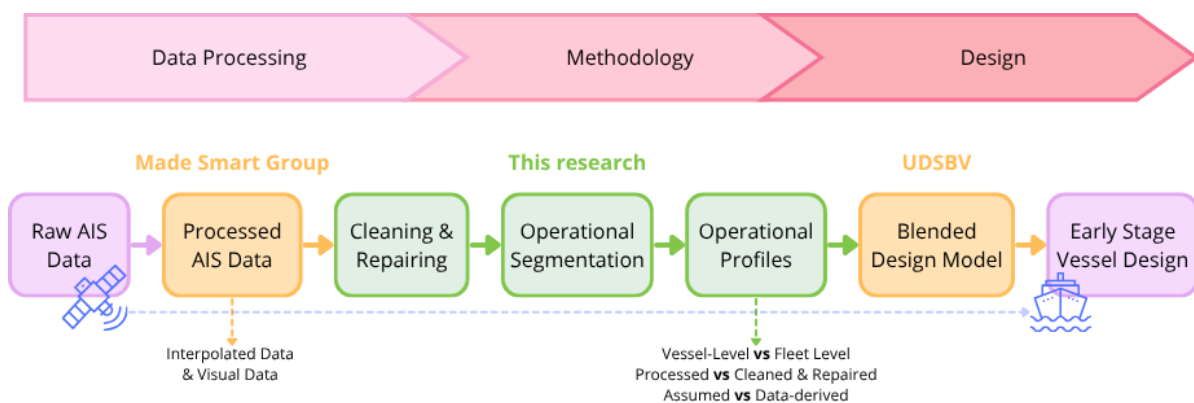


Figure 1.1: Overview of the workflow from raw AIS data to early-stage vessel design, highlighting existing industry components and the contributions of this research.

As shown in Figure 1.1, part of the process relies on existing tools, such as AIS data processing provided by MSG and the blended design model developed by UDSBV. This research focuses on the intermediate steps, including data cleaning and repairing, operational segmentation, and the generation of operational profiles. In addition, the figure highlights how the impact of these steps is evaluated at both vessel and fleet level, and how the resulting operational profiles are integrated into the design process.

The aim of this research is to investigate how AIS data can be used to derive operational profiles for early-stage vessel design, while assessing the sensitivity of these profiles to preprocessing steps and their impact on vessel design outcomes. The following main research question is addressed:

“How can data-driven operational profiles be derived from AIS data for early-stage vessel design?”

To answer this main research question, the following sub-questions are addressed:

- How can AIS data be processed and repaired to enable reliable operational analysis?
- How can AIS trajectory data be segmented to identify different operational states of HTVs?
- How do AIS-derived operational profiles compare to the assumed operational profiles used by UDSBV?

-
- What is the impact of incorporating AIS-derived operational profiles on HTV design within early-stage vessel design?
 - How can AIS-derived operational profiles support early-stage vessel design?

This thesis¹ is structured as follows. First, the main insights from the literature review are presented to position the research within existing work. Next, the methodology is described, including AIS pre-processing and segmentation. This is followed by the results, where operational profiles are analysed at both vessel and fleet level. Subsequently, these results are compared with assumed operational profiles and applied within the Blended Design framework. Finally, the discussion, conclusions, and recommendations are presented.

¹For this research, Generative AI (ChatGPT) was used to support the writing process by improving grammar, style, and clarity of the text, and to assist in the development and structuring of Python code for data processing and analysis. All outputs were reviewed and adapted by the author.

II. Literature

2

Data Driven Design

There are numerous design methods for designing vessels. During ship design, many important decisions must be made in the early stages. In 2016, Keane [5] highlighted the importance of improved decision making support during the early phases of ship design. Similarly, Gaspar [6] emphasized the critical role of data in generating information and knowledge. He argued that data, when effectively processed, can be transformed into actionable knowledge, leading to more efficient task execution and faster, more accurate decision making. Combining the insights of Keane and Gaspar shows that leveraging data in the early design phase can significantly enhance the quality and precision of design decisions.

In section 2.1, the specific areas within early stage ship design where data can provide meaningful support are explored. This is followed by section 2.2, which discusses various current applications of data in ship design. In section 2.3, several definitions of an operational profile from the literature are discussed. Finally, section 2.4 concludes the chapter by presenting key insights and reflections.

2.1. The role of operational profiles in early stage design

Ship design is typically divided into several stages. The first stage, referred to as the preliminary design stage, determines the main technical and economic characteristics of the vessel based on the owner's operational requirements [7, 8].

As illustrated in Figure 2.1, the preliminary design stage, which includes requirement definition and concept design, is characterized by limited knowledge and a high degree of uncertainty. As noted by Gaspar [6], data can be transformed into knowledge when properly processed. Incorporating data at this stage can therefore improve designers' understanding and lead to more informed decision making.

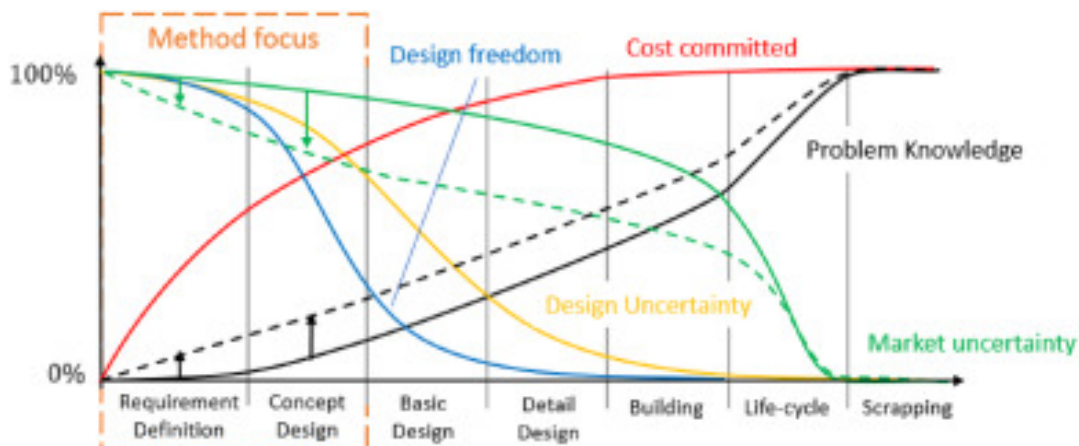


Figure 2.1: Uncertainties for ship design [9]

The first step in ship design is determining the intended use of the vessel in order to define requirements and an operational profile. An operational profile specifies the proportion of time a ship is expected to operate in different modes [10] to fulfill its intended purpose.

Operational profiles must be carefully estimated, as overestimation or underestimation can significantly influence ship design, costs, economic viability, and environmental performance [11]. Leveraging data during the preliminary design stage can reduce assumptions, increase knowledge, and result in more accurate and realistic designs that better reflect actual operational demands.

2.2. Current applications of data within ship design

Research on the use of data in ship design has increased in recent years to improve knowledge and support design decisions. Many studies rely on data from Automatic AIS, which is mandatory for a large number of vessels.

Several authors have shown that operational profiles derived from data can provide valuable insights for new vessel designs, as actual vessel behavior and power demand become better understood. These studies often reveal discrepancies between assumed and actual vessel usage. For example, a study that used data to assess propulsion and energy system performance resulted in more accurate dimensioning of engines, generators, and batteries [12].

Historical vessel data is not only used for more accurate dimensioning of vessel components. Jafarzadeh and Schjøberg [13] used AIS data to develop data driven operational profiles based on load factors of different vessel types to evaluate their suitability for alternative propulsion systems.

Other studies emphasize that AIS data alone is often insufficient to reliably distinguish between operational modes. Mysen [14] and Smith [15] incorporated additional contextual information, such as safety zones, logbooks, and binary structures, to identify vessel activities. These findings highlight the importance of contextual data when deriving operational profiles from AIS data. In addition, external influences such as political events or economic disruptions can significantly affect vessel behavior and should be considered when developing operational profiles from data [16].

Overall, existing research demonstrates that data can provide valuable insights for vessel dimensioning, analysis, and design, while also highlighting limitations related to missing contextual information.

2.3. Definition of Operational Profiles

The term operational profile is widely used not only for ships but also across many different fields, yet there is no single uniform definition in the literature. Operational profiles are interpreted differently depending on the vessel type, research objective, and available data.

Some studies describe operational profiles as speed time distributions. In this approach, the time spent within specific speed ranges is accumulated and analyzed [17]. A similar method focuses on analyzing time spent at certain load factors derived from speed data, vessel data, and propulsion characteristics [13].

Other studies emphasize time spent in contextualized operational modes rather than speed or load ranges. In this case, different operational modes are defined along with the time spent in each mode, such as transit, standby, or dynamic positioning [12]. These modes can still be described using speed or load ranges.

Overall, the literature shares a common objective for operational profiles, which is to describe how a vessel is operated. The specific approach depends on the research objective and the vessel type.

2.4. Conclusion Data Driven Design

This chapter reviewed current research on data driven design, with a focus on the use of data for developing operational profiles. It also addressed the first research question:

What is the current state of research on data driven operational profiles for maritime vessels?

The reviewed research shows a growing interest in using historical data to develop operational profiles

for new vessel designs. Historical data can also be used to evaluate changes in operational profiles over time and to estimate future operational profiles.

Current studies on data driven operational profiles highlight several important considerations for future research. One key finding is the importance of incorporating contextual data and assessing whether the data is significantly influenced by major external factors.

The integration of data derived operational profiles offers a promising opportunity to improve the accuracy and relevance of ship design processes. However, further research is needed to incorporate data driven operational profiles into comprehensive design frameworks and to better understand their influence on vessel design.

3

AIS Data

For data-driven research in the maritime domain, AIS data is widely used because it provides the most extensive vessel position information available through commercial or open sources. Since December 31, 2004, the International Maritime Organization has required AIS technology to be installed on board certain vessels [2].

In section 3.1, the functioning of AIS is discussed, followed by a description of AIS data characteristics in section 3.2. Subsequently, data quality and reliability are addressed in section 3.3, along with pre-processing techniques in section 3.4. Finally, the second research question of the literature study is answered in section 3.5.

3.1. Overview and functioning of AIS

The International Maritime Organization has mandated the use of AIS on board all ships of 300 gross tonnage or more engaged in international voyages [18]. In addition, all cargo ships of 500 gross tonnage or more and all passenger ships are required to carry AIS. Nowadays, many pleasure vessels and other non-mandatory vessels also operate with active AIS systems on board.

An AIS transponder determines its vessel's position, speed, and course using a built-in GPS receiver [20]. This information is automatically communicated to other AIS-equipped vessels, AIS shore stations, and satellites.

An AIS system consists of a Very High Frequency transmitter, two Very High Frequency Time Division Multiple Access receivers, one Very High Frequency Digital Selective Calling receiver, and standard marine electronic communication links to onboard display and sensor systems [21]. AIS operates autonomously and continuously, regardless of the vessel's operating location.



Figure 3.1: Functionality of AIS [19]

3.2. AIS data characteristics

AIS data can be divided into three categories: static data, dynamic data, and voyage-related data [22]. Most dynamic information is automatically transmitted by the AIS system, while some information must be manually entered by crew members, such as navigational status, draught, destination, and estimated time of arrival. Figure 3.2 presents the data transmitted and received by AIS, categorized into these three groups.

There are two AIS classes: Class A and Class B [23, 24]. Class A is required for larger vessels and transmits data every two to ten seconds while moving. Class B transmits less frequently, typically

every thirty seconds to three minutes, and is optional for smaller vessels. Class B systems have lower costs and simpler installation requirements [20]. The main differences between the two classes are transmission frequency and the fact that Class B systems cannot be installed on vessels subject to the Safety of Life at Sea convention.

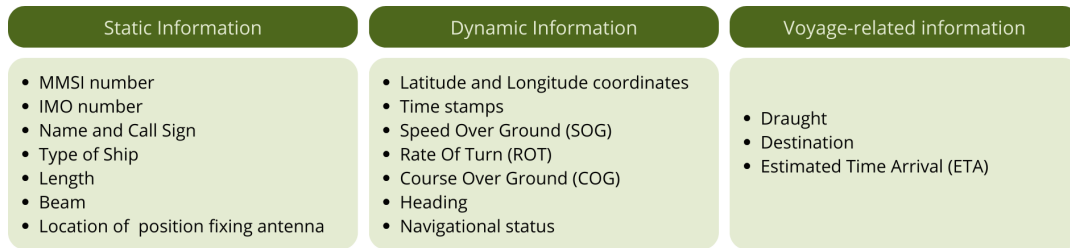


Figure 3.2: Data sent and received by AIS

3.3. Data quality and reliability

AIS provides a large volume of data, part of which is automatically transmitted and part manually entered [25]. Manually entered data is prone to errors, either accidental or intentional. In addition, delays may occur, for example when a vessel changes its navigational status during mooring, as crew members are occupied with operational tasks. Therefore, manually entered AIS data should not automatically be assumed to be reliable.

Automatically transmitted data can also contain errors due to system malfunctions, signal interference, or environmental conditions [26]. This may result in missing or incorrect data, such as unrealistic speeds, abrupt position jumps, or inconsistencies between speed over ground and vessel coordinates. These issues highlight the importance of AIS data preprocessing.

Data coverage also varies by year and geographic location. Initially, AIS coverage relied mainly on shore-based stations. Since 2008, satellite-based AIS has been developed and, from 2010 onward, satellites have increasingly been used to receive and relay AIS data, significantly expanding offshore coverage [27]. However, even with satellite reception, data gaps and errors may still occur, particularly in high-traffic areas where signal congestion can overload receivers.

3.4. Preprocessing of AIS data

Before AIS data can be used for analysis or research, preprocessing is required to obtain consistent and reliable datasets. The literature describes various preprocessing approaches, with the most common techniques summarized in Figure 3.3.

Many studies begin by filtering unreliable data, such as unachievable values of speed over ground, rate of turn, or vessel coordinates. This filtering can be based on expert-defined thresholds or known physical limits. Another recurring issue is the presence of shared Maritime Mobile Service Identity numbers, which are often removed during preprocessing. Outliers caused by noise are also commonly filtered, frequently using Density-Based Spatial Clustering (DBSCAN) methods.

Segmentation is another important preprocessing step and is applied depending on the research objective. In some cases, segmentation is based on known home ports, while in others, moored or anchored periods are identified to separate vessel activities.

Handling missing data is also a common preprocessing task. Various interpolation methods are used, including cubic spline interpolation and adjusted kinematics interpolation. For longer gaps, similarity-based trajectory methods may be applied, although these approaches are mainly effective in areas with high vessel traffic density.

Finally, data compression techniques are sometimes applied due to the large size of AIS datasets. These techniques aim to reduce data volume while preserving essential information, such as maintaining turning points rather than converting curved trajectories into straight lines.

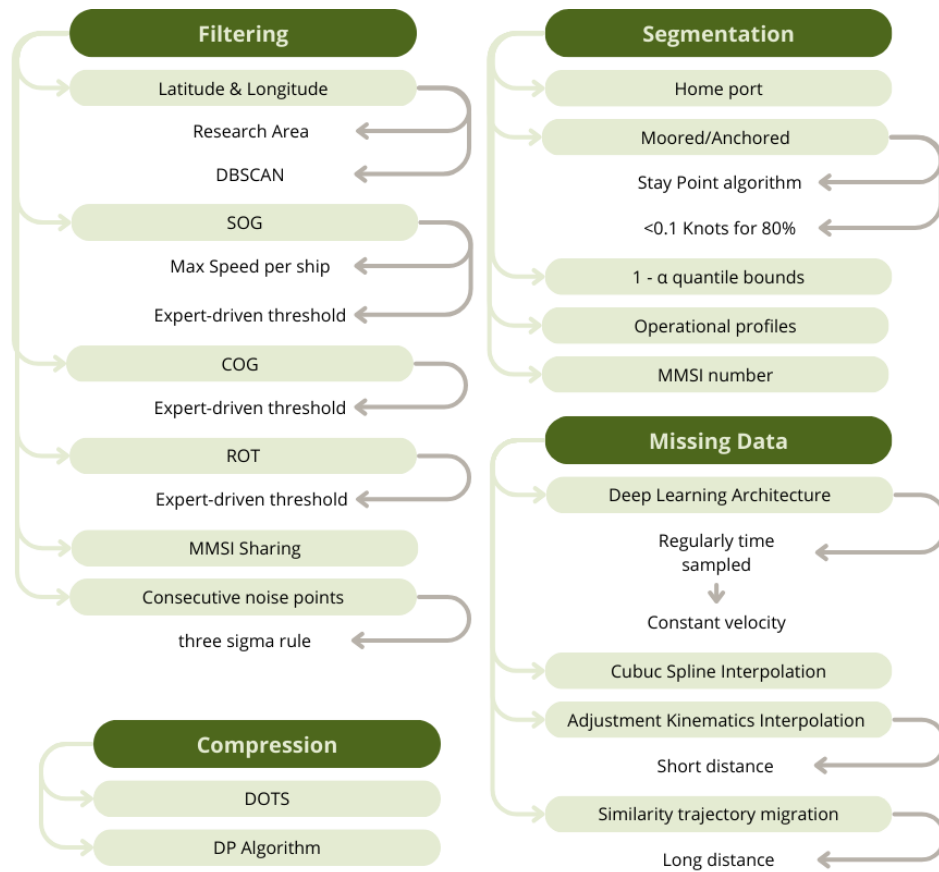


Figure 3.3: Most common preprocessing topics and techniques for AIS data

3.5. Conclusion AIS data

The literature reviewed in this chapter highlights the importance of preprocessing AIS data to ensure reliable and usable datasets. However, it does not explicitly show the impact of preprocessing. A wide range of techniques is available, from basic filtering to more advanced methods, and the selection of appropriate techniques depends on the scope and objectives of the study.

This chapter addressed the second research question:

What are effective methods for processing AIS data for an entire fleet of HTVs?

AIS data should first be checked for unachievable values, particularly when working with raw datasets. For fleet-level analyses, expert-defined thresholds are more practical than vessel-specific thresholds. When using datasets processed by third parties, it is important to understand which preprocessing steps have been applied.

Since AIS data often lacks consistent time intervals, interpolation methods such as cubic spline interpolation or adjusted kinematics interpolation can be used to address missing values. Segmentation based on mooring, anchoring, and port-to-port movements is more feasible than segmentation based on operational profiles. If dataset size becomes limiting, data compression may be considered, provided that essential information is preserved.

Overall, preprocessing AIS data remains a complex and time-consuming task and may still contain inconsistencies. For this research, the use of preprocessed AIS datasets is therefore considered the most efficient approach.

4

MSG Tools

This research was supported in obtaining AIS data by the company Made Smart Group (MSG) [4] was used. MSG provides access to several maritime datasets and offers various tools for processing and analyzing vessel and environmental data. This chapter describes the tools provided by MSG in section 4.1. In section 4.2 some options for data retrieval are discussed and lastly in section 4.3 the options of downloading the data are explained.

4.1. MSG Tools

The MSG granted access to a selection of their analytical tools, which were used extensively throughout this study. The three main tools employed were **Prospector**, **Plotter**, and **Cartographer**. Each of these is described below.

Prospector

The first tool used is *Prospector*, described by the MSG as the “entry point for maritime AIS vessel movements, weather, oceanographic, and tidal data” [4]. *Prospector* provides access to one of the largest and most comprehensive vessel position databases in the world. This database contains raw AIS messages dating back to 2005, covering approximately 230,000 vessels.

Prospector enables the retrieval of not only vessel position data but also historical weather, oceanographic, and tidal information for specific voyages. Weather and oceanographic data are available from 2015 onwards, while tidal data are available from 2019 onwards.

Data can be queried in multiple ways, for instance by selecting specific vessels, groups of vessels, geographical regions, or by applying filters such as destination, speed, or vessel type. Results can be exported in several formats, including Comma-Separated Values (CSV), Tab-Separated Text (TXT), or Google Earth (KML). Initially, positional data are retrieved, after which additional environmental data can be appended.

Another feature of *Prospector* is its integration with *Plotter*, a visualization tool that enables graphical inspection and replay of vessel movements together with associated weather and oceanographic conditions.

The *prospector* database contains historical AIS data. Every few hours a new batch of historical data is inserted into the database. This short delay ensures data quality and reliability, as all information undergoes a rigorous cleaning and quality control (QC) process. During this process, duplicate records are removed and timestamps are standardized to Coordinated Universal Time (UTC). This is particularly important for satellite-received AIS data, where transmission delays may occur if no nearby ground station is available.

Plotter

The second tool provided by the MSG is *Plotter*. Plotter enables visualization of vessel tracks and related information, which can be derived from different data sources: *Track*, *High Detail Track*, or *All Data*. In Plotter, users can view the track alongside static vessel information and display it over several types of background maps, such as Paper, Electronic Chart Display and Information System (ECDIS), OpenStreetMap, or Satellite imagery.

Plotter also supports integration with weather data, allowing visual representation of environmental parameters such as significant wave height along the voyage. Weather can be represented in three ways:

- **Weather overlay** – an additional layer displayed on the map, updated at three-hour intervals.
- **Weather parameters** – viewable via the time slider, alongside static vessel parameters such as speed.
- **Weather time series** – provides access to combinations of different weather parameters for temporal analysis.

Cartographer

The third tool used in this research is *Cartographer*. This supporting tool follows the principle of “create once, use many times.” In Cartographer, all known monopile manufacturers and marshalling ports can be drawn using geometric objects such as points, lines, circles, and polygons. Monopile manufacturers are typically marked in red, while marshalling ports are drawn in blue. These drawings can then be imported into Plotter, where it becomes straightforward to analyze whether and when a vessel visits specific monopile factories or known marshalling ports. This functionality allows for quick assessments of vessel activity between manufacturing sites and ports.

4.2. Data Retrieval

The AIS data used in this study were obtained through Prospector, which offers two primary methods for data retrieval: *Track Data* and *All Data*. Each method serves a different purpose depending on the required level of detail.

Track Data

The *Track Data* option is designed to quickly display the overall vessel track with minimal positional information. It has a spatial accuracy of approximately 50 meters and includes only essential positional data. Speed and course are derived from these positions. However, this dataset does not include detailed parameters such as heading, speed over ground, high-frequency updates, estimated time of arrival (ETA), or destination.

The advantage of using *Track Data* lies in its fast retrieval, making it suitable for efficiently obtaining an overview of a vessel’s operational patterns, for example, to determine whether a vessel was active during a specific year.

All Data

The *All Data* option contains the complete set of AIS messages, including additional information such as vessel heading, speed, draught, navigational status, and destination. This method is therefore used for more detailed analyses requiring high temporal and spatial resolution.

High Detail Tracks

A further option within *All Data* is the *High Detail Tracks* function. These tracks are derived from all available AIS updates and contain approximately four times more positional data than the basic *Track Data*. The function reduces data density while maintaining the accuracy and fidelity of the vessel’s track, providing a precision of approximately 12.5 meters.

4.3. Data Export

The retrieved datasets can be downloaded in several formats, depending on the selected source. The export process involves multiple steps:

1. **Source selection:** Users can choose between three data sources:
 - **Tracks** – contains essential AIS positions with a spatial precision of about 50 meters.
 - **All Data** – includes all available AIS position reports and vessel parameters.
 - **High Detail Tracks** – derived from All Data, providing high positional accuracy (approximately 12.5 meters) while minimizing redundant data points.
2. **Type selection:** After selecting a source, users can specify the type of information to export:
 - **Positions:** Includes UTC timestamps, positions, speed, and course over ground.
 - **Positions per MMSI:** Provides the same information as above, but with each Maritime Mobile Service Identity (MMSI) stored in a separate Excel sheet.
 - **Positions and Statics:** Combines positional information with static vessel data such as name, length, and width.
 - **Positions and Statics per MMSI:** Contains both positional and static data, separated into individual spreadsheets for each MMSI.
 - **Positions with Weather:** Adds weather information from the MSG Weather and Seastate model. This option is only available for Track and High Detail Track data, as weather data are updated at lower temporal resolution than AIS. Providing weather data for the full All Data frequency would place excessive load on the servers.
 - **Plotter:** Opens the retrieved data directly in Plotter.
 - **Plotter with Weather:** Opens both positional and weather data in Plotter for combined visualization.
3. **Format selection:** For all types except Plotter, the export format can be chosen:
 - Excel (*.xlsx)
 - Comma-Separated (*.csv)
 - Tab-Separated (*.txt)
4. **Settings:** Additional export settings allow interpolation and decimation of data:
 - **Interpolation:** Generates regular time intervals. The user can choose whether to preserve original positions. A maximum gap parameter prevents interpolation across excessively long time intervals.
 - **Decimation:** Reduces data density by selecting one data point per specified time interval. Unlike interpolation, gaps remain unfilled, preserving data continuity.
5. **Filters:** Data can be filtered by several criteria, including MMSI, Cartographer-defined areas, course, destination, vessel dimensions, time period, speed, and vessel type.

These exported datasets form the basis for subsequent analyses within this research, enabling the data driven operational profiles.

5

Blended Design

The objective of this research is not only to create operational profiles based on historical AIS data, but also to apply these profiles in an early-stage ship design tool to assess their impact on vessel design. While standalone data can already provide valuable insights, embedding such data into a design tool enables the evaluation of how different data inputs influence vessel performance over its lifetime. This allows for determining the required level of data cleaning and assessing whether data-driven design is beneficial for an HTV fleet.

UDSBV, together with multiple MSc students, has developed its own early-stage design tool, which will be used in this research. Before applying this tool, it is necessary to understand how it functions and whether it is suitable for the objectives of this study.

In section 5.1, the selection of Blended Design for this research is explained. This is followed by a description of the methodology in section 5.2. In section 5.3, the representation of operational profiles within Blended Design is discussed. Next, section 5.4 outlines the opportunities and challenges associated with using Blended Design. Finally, the third research question from the literature is revisited and answered in section 5.5.

5.1. Design Tool

De Winter and Van Veen [16] emphasized the need to use operational data as input for estimating a vessel's future performance and economic value under varying operating conditions. Blended Design, developed by Zwaginga [1] in collaboration with UDSBV, addresses this need. The tool evaluates performance and economic outcomes across a wide range of design alternatives while accounting for uncertainty in market conditions and enabling the exploration of different market scenarios.

Blended Design has been applied and refined through four master's theses and several internal UDSBV projects. Its performance has also been benchmarked against alternative design methods by de Ridder [28]. The tool is well suited for exploring large design spaces while integrating alternative fuels, market uncertainty, and optimization of financial and environmental performance. This makes Blended Design suitable for assessing the effects of data-driven operational profiles.

5.2. Blended Design Methodology

Blended Design [1] was developed to address challenges in the early stages of ship design, particularly the difficulty of accounting for uncertainty in future market demands. By simulating the long-term effects of design parameters on market performance, the method supports strategic decision-making and improves transparency in communication with clients.

The core concept of Blended Design is shown in Figure 5.1. The tool consists of a ship model and a market model. By combining inputs from both models, Blended Design generates a large number of unique vessel configurations and evaluates their financial performance over the vessel's lifetime within a simulated market environment.

Based on a parent hull, the ship model generates a wide range of vessel configurations by varying key design parameters. Naval architectural methods are used to estimate resistance, propulsion, weight, and capacity for each configuration. Infeasible designs are filtered out based on input constraints and the limitations of specific calculation methods, such as Holtrop and Mennen.

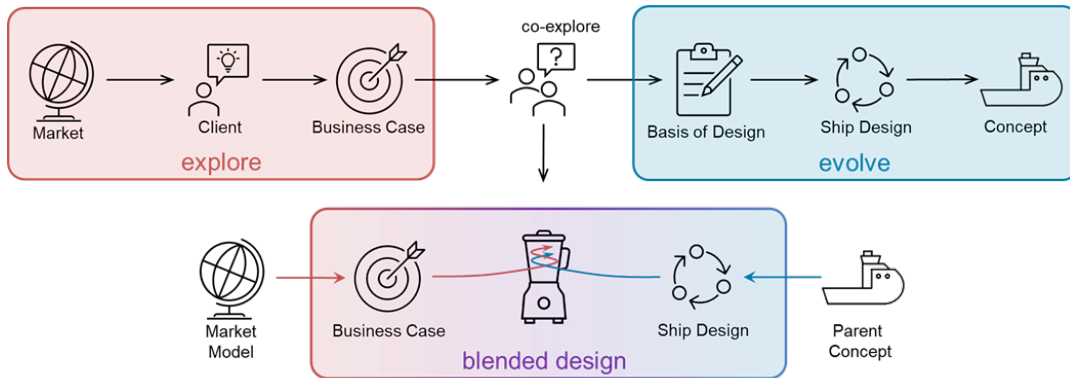


Figure 5.1: Principle idea of Blended Design by J.D. Stroo

The market model simulates future demand for offshore wind foundations based on historical and projected market data. This demand can also be adjusted manually to represent bounded or unbounded market scenarios, depending on assumptions about future developments. Demand is expressed in terms of monopile dimensions. To link the ship and market models, Blended Design constructs a probability matrix.

Figure 5.2 illustrates how trend lines derived from historical and planned market data are converted into probability functions. For each year, the probability of different monopile diameters is estimated, resulting in a probability matrix that represents the likelihood of specific foundation sizes being required in a given year.

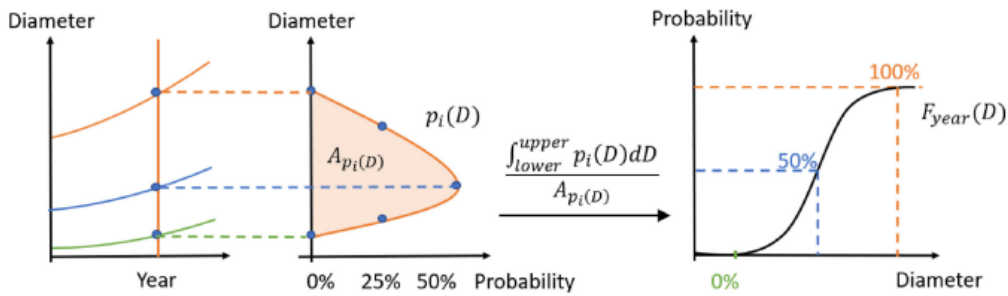


Figure 5.2: Transforming trend lines to probability functions by Zwaginga [1]

This probability matrix is combined with all generated ship configurations. To evaluate these configurations, Zwaginga [1] implemented expected value as a performance metric. This method calculates the probability of different income scenarios and determines the expected return on investment. Discount rates are included to account for depreciation and opportunity costs. Expected value is calculated using a discounted Markov chain with rewards over a finite time horizon. The reward vector can represent costs, revenues, or profits.

In addition to financial metrics, operational performance indicators are considered, such as the number of contracts or foundations a vessel can complete during its lifetime. These indicators depend on operational profiles describing how the vessel operates over time. Return on investment is used as a key evaluation criterion, as it captures both investment costs and generated profits.

The output of Blended Design consists of a large CSV file containing the results for all ship configurations. From this dataset, informative plots can be generated, showing a design parameter on the horizontal axis and a performance indicator on the vertical axis, such as vessel length versus return on investment. An example of such plots is shown in Figure 5.3.

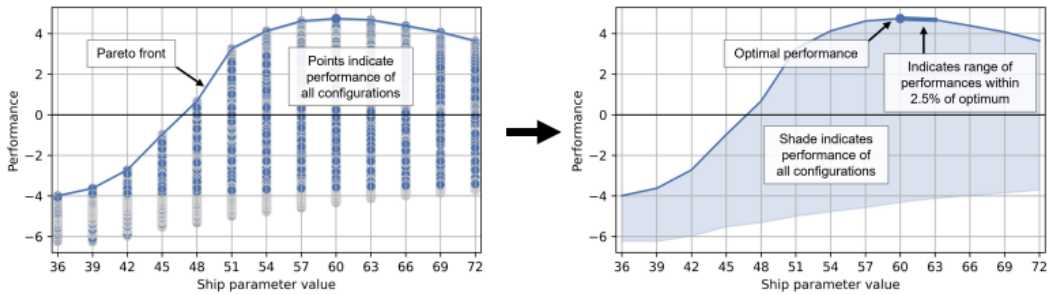


Figure 5.3: Visual guide for interpreting result plots by de Ridder [28]

In these plots, the central line indicates the best-performing designs for a given parameter, while the shaded area represents a range of values close to the optimum. These visualizations help identify trends, trade-offs, and optimal regions within the design space, supporting informed decision-making early in the design process. Blended Design can also be used to investigate alternative scenarios, such as variations in operational profiles or contract types. Some more information about Blended Design can be found in Appendix A.

5.3. Operational profile in Blended Design

As discussed earlier, the primary objective of an operational profile is to describe how a vessel is or will be operated. Figure 5.4 presents the typical operations performed by HTVs, which are also reflected in the operational inputs used in Blended Design.

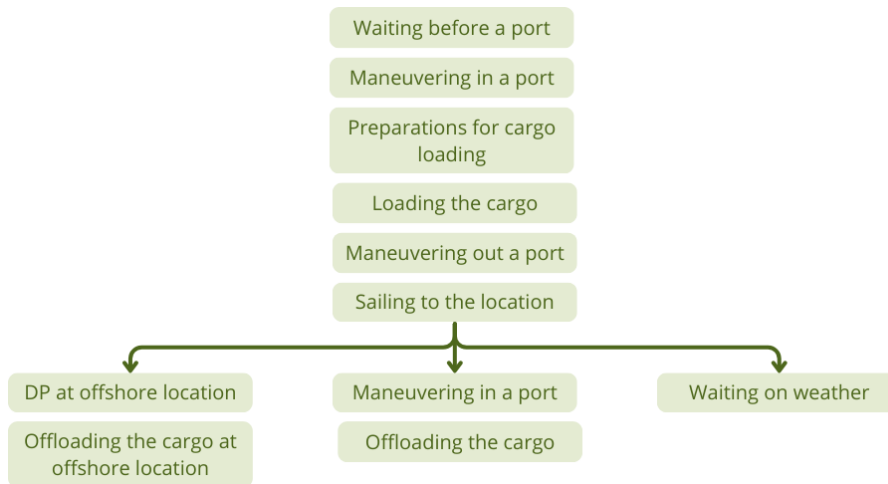


Figure 5.4: Typical operations of a HTV

Currently, an input CSV file describes vessel transit and dynamic positioning activities using parameters such as sailing distance, time in dynamic positioning, and the number of self-supportive days. Cargo loading is handled deeper within the Blended Design code. Once the vessel type and cargo capacity are defined, loading and offloading times are calculated using fixed values per cargo unit. These values could potentially be replaced by data-driven inputs. Additional operational aspects, such as waiting time before port entry and maneuvering within ports, can also be incorporated.

In this research, the operational profile of an HTV is therefore defined by sailing distance, time in dy-

dynamic positioning, self-supportive days, loading and offloading times, and waiting time before entering port.

In the early design stage, design speed is treated as a variable used to generate different ship configurations and is not directly included in the operational profile. Vessels are therefore assumed to sail at design speed at all times. Analyzing actual speed distributions and incorporating them into Blended Design could provide additional insights.

5.4. Opportunities and Challenges

As mentioned before, the real opportunity within Blended Design lies in replacing assumed operational profiles with operational profiles based on real vessel data. De Winter [16] notes that experience-based estimates, although common, can lead to unrealistic scenarios. Assumption-based operational profiles limit prediction accuracy. Replacing these assumptions with data-driven operational profiles could significantly improve realism and reliability.

Blended Design currently assumes that a vessel performs a single type of operation throughout its lifetime, resulting in designs optimized for that specific operation. While this may be representative for some vessels, allowing variation in operational activities could lead to more flexible designs.

Incorporating variable speed patterns for different operations could further improve realism. Although Blended Design allows the evaluation of different sailing speeds, allowing speed to vary over time, for example due to seasonal weather conditions, could provide deeper insight into real vessel behavior.

AIS data presents additional opportunities, such as emissions estimation, extraction of speed profiles per trip or region, and analysis of weather effects when combined with meteorological data. AIS data can also be used to quantify time spent in different load factor ranges and link these to operational modes.

Blended Design can further be used to assess how sensitive vessel designs are to changes in operational profiles. For example, operational profiles previously used by UDSBV may differ from data-driven profiles, while the resulting vessel designs may remain similar. It would also be valuable to study the effect of different levels of data cleaning on both operational profiles and design outcomes.

5.5. Conclusion Blended Design

Blended Design is a comprehensive approach for early-stage ship design that integrates ship, market, and uncertainty models to evaluate vessel performance under a range of future scenarios, with a focus on economic and environmental optimization.

This chapter addressed the following research question:

How can AIS-derived operational profiles be integrated into early-stage design tools such as Blended Design to improve HTV designs?

AIS-derived operational profiles can be integrated into Blended Design by replacing assumption-based inputs with data-driven representations of vessel behaviour. This allows for a more realistic representation of vessel operations within early-stage design, leading to improved evaluation of performance, particularly in terms of economic outcomes such as return on investment and optimal design speed.

It should be noted that Blended Design is a tool developed by UDSBV. Within the scope of this research, the tool has been used as provided, and therefore the level of insight into the internal workings of the model remains limited.

6

Heavy Transport Vessel Market

To effectively analyse operational profiles, it is first necessary to understand the HTV market, as vessels are designed to meet specific market needs. A market analysis forms the foundation of this research by identifying the existing HTV fleet and enabling the selection of the most relevant vessels for AIS data collection and analysis.

The market analysis begins with an explanation of the difference between HTVs and Semi-Submersibles in section 6.1. This is followed by an examination of market demand, describing key drivers for the HTV fleet in section 6.2. Subsequently, the current supply of vessels capable of meeting this demand is assessed in section 6.3. In section 6.4, market demand and supply are compared to evaluate the market fit. Finally, the fourth and last research question of the literature is addressed in section 6.5.

6.1. HTV compared to Semi-Submersibles

Before discussing the HTV market, it is important to define what HTVs and Semi-Submersibles are. Both vessel types have as their main purpose the transportation of heavy and irregularly shaped cargo. Therefore, deck area is one of the most important design features.

Although Semi-Submersibles are only considered to a limited extent in this research, they operate within the same offshore wind market as HTVs, particularly in the transportation of monopiles from fabrication yards to marshalling ports. Therefore, analysing their operational profiles provides relevant context.

The primary difference between HTVs and Semi-Submersibles lies in their loading capabilities. Semi-Submersibles can perform Roll-on/Roll-off (RoRo), Float-on/Float-off (FloFlo), and crane-based loading operations, whereas HTVs are typically limited to RoRo and crane-based methods. The ability of Semi-Submersibles to partially submerge enables FloFlo operations, providing greater flexibility in cargo handling.

However, this increased versatility comes at a cost. Semi-Submersibles generally have higher Capital Expenditures (CAPEX) and Operational Expenditures (OPEX), as illustrated in Figure 6.1. Due to their advanced capabilities, Semi-Submersibles are often considered over-capable for standard HTV operations. Nevertheless, they can still operate within the same market segment when required and therefore are taken into account in this research.



Figure 6.1: Comparison between HTV and Semi-Submersible vessels

6.2. Market Demand

The offshore wind market is growing rapidly, making it an increasingly attractive sector for HTVs. This market requires a wide range of specialized vessels to support the various stages of wind farm development [29]. HTVs play a crucial role during the construction and installation phase. The installation of offshore wind farms requires the fabrication and transport of large components, including foundations, transition pieces, towers, and rotor nacelle assemblies. These components are typically transported from fabrication yards to marshalling ports, where they are stored and preassembled before being delivered to offshore wind farm sites for final installation.

Tjaberings et al. [29] examine several logistics strategies for transporting foundations, including shuttling, feederling, and wet towing. It remains uncertain to what extent these strategies are consistently applied in practice. Gaining insight into whether HTVs primarily operate between ports or also perform feederling tasks would be valuable, not only for the offshore wind market but also for other sectors served by HTVs.

HTVs operate across multiple market segments. In this chapter, market drivers are based on the Clarksons Offshore and Renewables Review 2023 [30]. The primary market driver is the transportation of offshore wind foundations. Clarksons expects a transition from fixed foundations to floating foundations. In addition, the offshore oil and gas sector is expected to continue contributing to demand, particularly for rig removal operations. Another growing market is the transportation of large modules for liquefied natural gas export plants and shipbuilding modules. Beyond these specialized markets, HTVs also operate in the spot market, transporting general cargo such as port cranes, inland barges, and other heavy machinery.

Overall, Clarksons projects increasing demand for HTVs in the transportation of floating offshore wind foundations, fixed foundations, and liquefied natural gas modules. Demand from other sectors is expected to remain stable, resulting in an overall increase in demand for HTVs.

6.3. Market Supply

Market demand for HTVs can also be met by semi submersible vessels. Therefore, semi submersibles are included in the Clarksons overview [30] when assessing market supply. To assess current market supply, a dataset containing 266,096 vessels of various types and operational statuses was obtained [31]. This dataset included vessels ranging from new builds to those marked for scrapping. As no direct filter for semi submersibles or deck cargo carriers was available, a manual filtering process was applied. The filtering steps are illustrated in Figure 6.2. After filtering vessel statuses and types, the dataset was reduced to 1,131 vessels. Further filtering based on vessel length and breadth reduced the number to 230 vessels. A final filter based on crane presence and container capacity resulted in 110 vessels of interest, consisting of HTVs and semi submersibles.

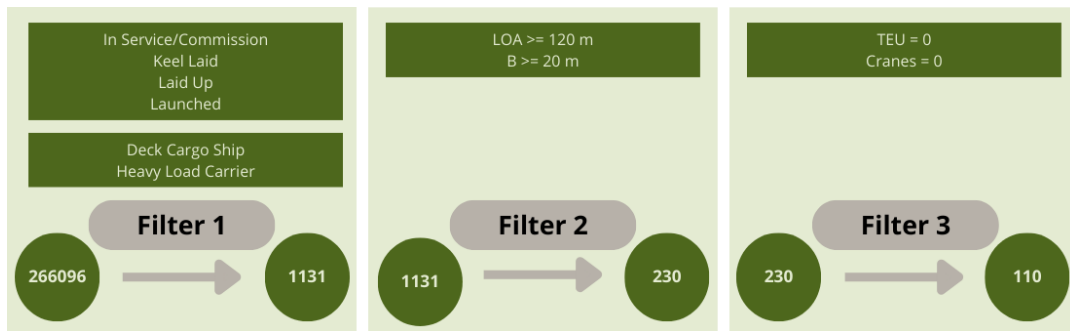


Figure 6.2: Filtering steps for supply list

These vessels were manually classified as HTV, semi submersible, or other, resulting in 61 HTVs and 47 semi submersibles. Shanghai Zhenhua Shipping operates the largest HTV fleet, while COSCO Shipping operates the largest semi submersible fleet, followed by Boskalis. China dominates the construction market for both HTVs and semi submersibles.

The development of fleet supply over time was also analyzed, as shown in Figure 6.3. This analysis indicates a stronger increase in HTV construction in recent years compared to semi submersibles, suggesting that deck cargo carriers are currently better aligned with market demand and logistical preferences.

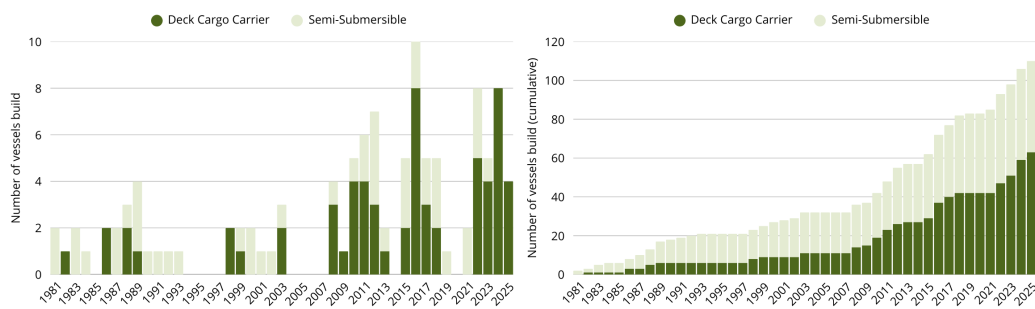


Figure 6.3: Supply of currently operated HTVs by year

This section provides an initial pool of vessels relevant to this research. From this pool, a selection will be made to obtain AIS data for further analysis.

6.4. Market Fit

Assessing whether market demand and supply are balanced remains complex and difficult to quantify accurately. Clarksons [30] projects increasing demand for HTVs, primarily driven by the growth of floating offshore wind foundations. While current supply appears stable, fleet expansion is expected to follow demand trends in the coming years.

An increase in demand within a specific sector does not automatically result in increased demand for HTVs. Semi submersibles offer greater versatility and can operate across multiple markets, but they are often less cost effective for specific tasks. HTVs, while more specialized, are generally better suited for targeted operations within their market niche.

Combining AIS data with market information could provide valuable insights into actual vessel utilization. For design companies, understanding evolving market demands is essential to developing vessels that operate efficiently and cost effectively. Evaluating whether actual vessel usage aligns with design assumptions is therefore crucial. AIS data enables the assessment of such alignment and allows shifts in vessel usage over time to be identified.

To effectively analyze market behavior using AIS data, integration with additional datasets is required. For example, AIS location data could be combined with information on offshore wind foundation fabrication yards, marshalling ports, and offshore wind farm sites. If operational profiles derived from AIS data can be linked to specific market segments, this would provide insight into which vessel designs and operational strategies are most competitive, and how market driven operational demands influence design requirements.

6.5. Conclusion Heavy Transport Vessel Market

The current HTV market is largely evaluated based on assumed demand, which introduces uncertainty in market predictions. AIS data provides an opportunity to validate these assumptions on both the demand and supply sides. To generate meaningful insights, however, AIS data must be combined with additional datasets that allow operational profiles to be linked to specific market segments. This chapter addressed the following research question:

What are the current trends and developments in the HTV market, and how has the market evolved over recent years?

The most prominent trend identified is the growing role of HTVs in transporting offshore wind foundations. A substantial increase in demand is expected for floating offshore wind foundations, although this market remains in a developmental phase and its long term scale is uncertain. This anticipated demand is reflected on the supply side by an increasing number of HTVs entering the market. The focus of this research therefore lies on the HTV segment, as these vessels generally offer lower capital and operating expenditures compared to semi submersibles. Nevertheless, it may also be valuable to examine the extent to which existing semi submersibles participate in this market, and a limited selection of these vessels may therefore be included in further analysis.

III. From AIS Data to Data-Driven Early-Stage Vessel Design for HTVs

7

Methodology

The objective of this research is to evaluate the extent to which AIS data can be used to derive data-driven operational profiles. To achieve this, a structured methodology was applied. This chapter describes and discusses the methods used to obtain the research outcomes.

The first step was to determine the fleet selection; the method and criteria for this will be discussed in section 7.1. Next, the data of these vessels were collected through MSG [4], which will be discussed in section 7.2. In section 7.3, it will be explained how the data will be prepared and what the final datasets are. The method used to create the operational profiles will be discussed in section 7.4.

7.1. Fleet selection

To determine the fleet selection for this research, a dataset containing 266,096 vessels of various types was obtained from UDSBV [31]. The dataset included vessels of all types, ranging from new builds to vessels marked for scrapping. This research focused on the current HTV fleet and, for additional comparison, also included the semi-submersible fleet. Therefore, only HTVs and semi-submersibles were selected from the dataset.

The filtering steps applied are illustrated in Figure 7.1. These filtering thresholds were used to identify the most common HTVs and semi-submersibles.

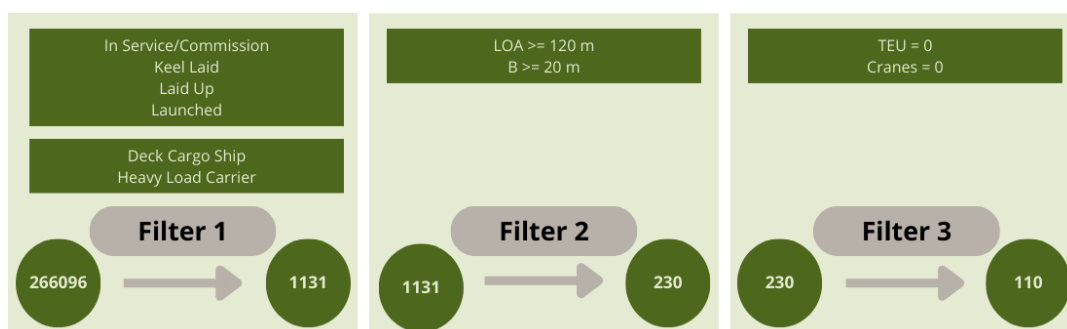


Figure 7.1: Filtering steps for supply list

The selected vessels were manually classified as HTV, semi-submersible, or other, resulting in 61 HTVs and 47 semi-submersibles. This research primarily focused on HTVs and semi-submersibles operating in the offshore wind market. The final selection of vessels was therefore based on their known involvement in offshore wind projects. As it is not publicly documented which vessels transported or supported specific offshore wind components for all projects, the selection was primarily based on visual evidence, such as photographs showing wind farm equipment on board.

A maximum of 20 vessels was selected in agreement with supervisors from UDSBV. This selection included at least 10 HTVs and 5 semi-submersibles known to operate in the offshore wind market, based on visual evidence or information provided by the supervisors. The selected vessels had to have been in operation for at least one full year within the period from 2018 to the present.

7.2. Data Collection from Made Smart Group

From the list of a maximum of twenty vessels, AIS data was collected. All selected vessels are required to carry an AIS Class A system on board, which provides dynamic, static, and voyage-related information.

For this research, preprocessed AIS data provided by MSG[4] was used. The data was obtained using the Prospector tool. This database contains AIS messages dating back to 2005 and covers approximately 230,000 vessels [4]. All data provided by MSG underwent a cleaning and quality control process, during which duplicate records were removed and timestamps were standardized to UTC. In Prospector, vessels were queried using their MMSI numbers.

AIS data was queried per vessel and per year. Before downloading each dataset, a preview was inspected. An example of such a preview is shown in Figure 7.2. The preview was used to assess whether the data was suitable for further analysis. If a vessel showed little to no movement throughout an entire year, this was noted, but the data was not downloaded, as it was not suitable for deriving an operational profile. If the vessel showed sufficient movement during the year, the data was downloaded.

Results preview

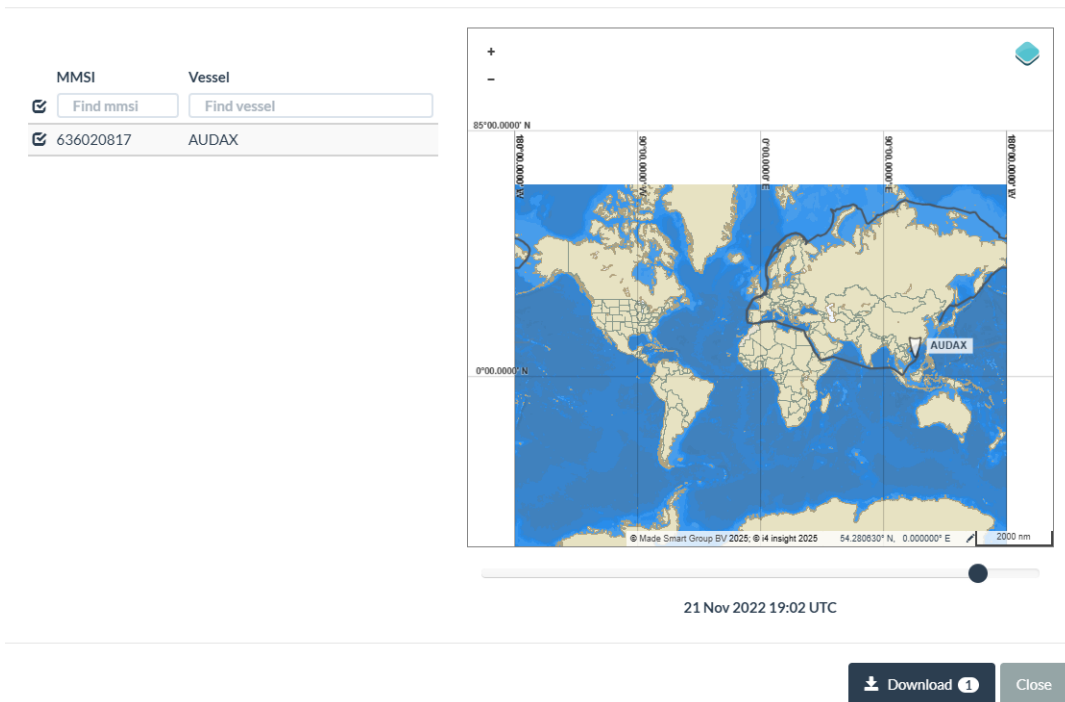


Figure 7.2: Preview of a query in Prospector from MADE SMART GROUP [4]

For this research, three different datasets were downloaded for each vessel and year. The input settings used for these downloads are shown in Table 7.1. Each dataset included all selected vessels and years but differed in data source and processing settings.

Table 7.1: Input settings for data downloads used in this research

	Dataset 1	Dataset 2	Dataset 3
Source	All Data	All Data	High Detail Track Data
Type	Positions and Statics	Positions and Statics	Plotter
Format	txt	txt	N/A
Settings 1	Remove Position Jumps	Remove Position Jumps	Remove Position Jumps
Settings 2	N/A	Interpolate	N/A
Filters	None	None	None
Download Name	VesselName_Year_AD	VesselName_Year_I(300s_30m)	N/A

Dataset 2 was used to derive data-driven operational profiles, while Dataset 3 was used to validate the segmentation method. The All Data source was selected because it includes all AIS transmissions after quality control. Although this option required longer download times, it provided the most complete dataset. Figure 7.3 shows the variables included when using the All Data option.

Bigroll Bering 01/01/24 - 01/06/24				
		Tracks	High Detail Tracks	All Data
Data points		7031	10956	83063
Accuracy		± 50 m	± 12.5 m	
Time		00 min 26 s	16 min 18 s	16 min 18 s
Positions	Date und UTC Time	✓	✓	✓
	MMSI	✓	✓	✓
	Latitude (°)	✓	✓	✓
	Longitude (°)	✓	✓	✓
	SOG (kn)	✓	✓	✓
	HDG (°)			✓
	IG (°)	✓	✓	✓
	ROT (°/min)			✓
	Status			✓
	Distance (nm)	✓	✓	
	Section	✓	✓	
	Statics	Name	✓	✓
IMO Number		✓	✓	✓
Call Sign		✓	✓	✓
Length Bow (m)		✓	✓	✓
Length Stern (m)		✓	✓	✓
Length Overall (m)		✓	✓	✓
Width Port (m)		✓	✓	✓
Width Starboard (m)		✓	✓	✓
Width Overall (m)		✓	✓	✓
Draught (m)				✓
Destination				✓
ETA				✓
Vessel Type				✓
Extra Info			✓	

Figure 7.3: Data variables provided when using the All Data option in Prospector by MADE SMART GROUP [4]

Dataset 3 was obtained using Plotter, another tool provided by MSG [4], which is explained in chapter 4.

Plotter enables visualization of vessel tracks and associated information. This data was derived from the High Detail Tracks source, which is a reduced version of the All Data source that retains only the points necessary to preserve the overall vessel trajectory. This reduced dataset was required because the full All Data source contains too many points for effective visualization. An example of data collection using Plotter is shown in Figure 7.4.

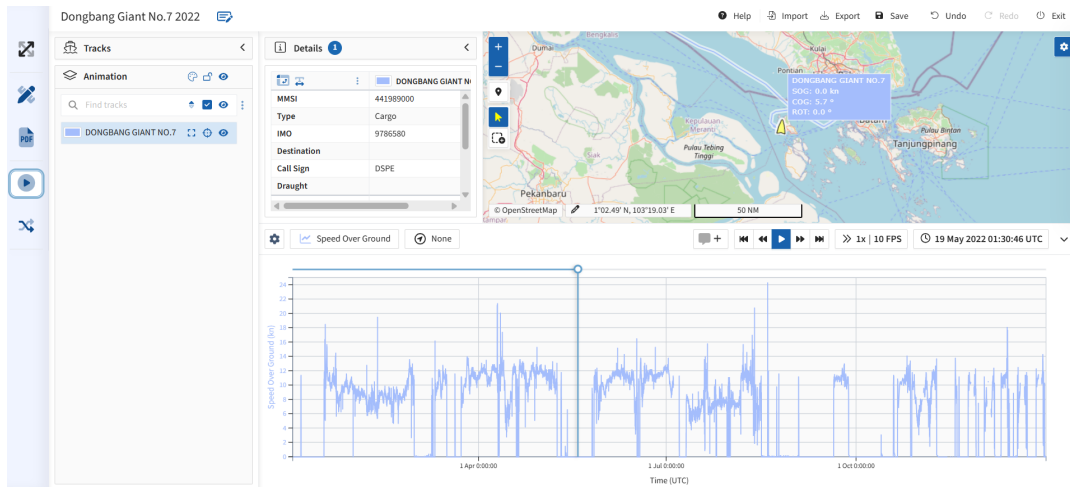


Figure 7.4: Example of data collection using Plotter from MADE SMART GROUP [4]

In summary, AIS data for all vessels in the predefined selection was collected in three different formats. Dataset 1 and Dataset 2 were stored as txt files containing all variables shown in Figure 7.3. Dataset 1 included the complete dataset, while Dataset 2 consisted of interpolated data at five-minute intervals, with gaps filled up to a maximum of thirty minutes. Dataset 3 was collected using Plotter and was used solely for validating the segmentation approach. Visuals of the exact settings used to obtain the three different datasets can be found in Appendix B.

7.3. Data Preparation

The next step after data collection was to prepare the data so that it could be used in this research. Before using the data, it was necessary to define how the AIS data was interpreted within the scope of this study. This is described in subsection 7.3.1. Subsequently, Dataset 1 was assessed, as described in subsection 7.3.2, to determine whether the data was sufficient for use in this research. The method and thresholds applied for this assessment are also presented in that section. In subsection 7.3.3, the cleaning and repairing procedure applied to the datasets is described. Finally, subsection 7.3.4 outlines how the datasets used in this research were compared based on different aspects.

7.3.1. Interpretation of the AIS data

The first step in this research was to define how the AIS data was interpreted and processed. The most relevant AIS variables used in this study were date and UTC time, latitude, longitude, speed over ground (SOG), heading (HDG), and course over ground (COG).

The date and UTC time were used in the format mm/dd/yyyy hh:mm:ss throughout this research. Latitude and longitude were provided in degrees for all data points and represented the geographical position of the vessel at a specific moment in time.

Course over ground and heading were also defined at specific timestamps. Therefore, for each recorded timestamp, the position of the vessel as well as its course over ground and heading were known.

The SOG was calculated by the AIS system based on position changes over a relatively short time interval. As a result, the reported SOG represented the vessel's speed over a short duration rather than over longer time intervals. This implies that sudden position changes or inaccuracies could result in temporarily high SOG values due to large distance changes in comparison to the short time periods.

However, when analyzing the data at five-minute intervals, the distance travelled between two consecutive AIS points could still be relatively small, sometimes even less than 100 meters. This could create a mismatch between the coordinates and the SOG, depending on how the SOG is interpreted within this research.

Table 7.2 illustrates the interpretation applied in this study. For data points with an interval of five minutes and a SOG of 5 knots, it was assumed that during the five minutes preceding that data point, the vessel sailed at a SOG of 5 knots. Thus, the SOG was assigned to the subsequent data point. This interpretation was consistently applied throughout the research, including for distance and duration calculations. Therefore, the distance and duration stored at data point B represent the distance and duration between data point A and data point B.

Table 7.2: Interpretation of the data

	UTC and Time	SOG	Latitude	Longitude	Distance
A	00:05	5	x	y	-
B	00:10	5	x	y	A to B with speed 5 from SOG
C	00:15	6	x	y	B to C with speed 6 from SOG
D	00:20	5	x	y	C to D with speed 5 from SOG
E	00:45	1	x	y	D to E with no speed from SOG
F	00:50	1	x	y	E to F with speed 1 from SOG
G	00:55	2	x	y	F to G with speed 2 from SOG

7.3.2. Dataset 1 (All Data)

Subsequently, it was assessed whether sufficient data was available to represent the operations of HTVs. The criteria used in this research are presented in Table 7.3. Similar to de Winter and van Veen [16], the number of data points per day and the average distance were considered.

A minimum of 120 data points per day was required, corresponding to approximately one data point every 12 minutes. For large vessels such as HTVs, this resolution was considered sufficient to identify operational behaviour. Furthermore, an average distance threshold was defined based on the maximum distance a vessel could travel at 15 knots during the average time gap.

In line with Jafarzadeh and Schjøberg [13], the proportion of data points with time gaps below 30 minutes and below 5 minutes was also evaluated to assess data coverage.

Table 7.3: Data reliability criteria for the All Data source

Dataset	Vessel Name AD Year
Data points	$\geq 43,800$
Avg points per day	≥ 120
Avg distance	$\leq 15 \text{ knots} \times \text{avg time gap}$
Max time gap	$\leq 1 \text{ month}$
Avg time gap	$\leq 10 \text{ min}$

All vessels and years were required to meet these criteria to be included in the analysis. These thresholds were specifically defined for HTVs and semi-submersibles; for other vessel types, different thresholds may be required.

The average number of data points per day was calculated using Equation 7.1. The Python script made for this analyses also accounted for leap years.

$$\text{AvgPoints per Day} = \frac{\#DataPoints}{\#Days} \quad (7.1)$$

The average distance was determined using a similar approach, dividing the total travelled distance by the number of data points. The distance between two coordinates was calculated using the haversine formula (Equation 7.2).

$$a = \sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos(\varphi_1) \cos(\varphi_2) \sin^2\left(\frac{\Delta\lambda}{2}\right), \quad d = 2R \arctan 2(\sqrt{a}, \sqrt{1-a}). \quad (7.2)$$

In Equation 7.2, the following variables were used:

- φ = latitude in radians
- λ = longitude in radians
- R = Earth radius ($R = 6,371,000$ m)
- d = great circle distance in meters

Latitude ranged from -90° to 90° , where negative values indicate the southern hemisphere. Longitude ranged from -180° to 180° , where negative values indicate the western hemisphere. Longitude wrapping was accounted for in the calculation of $\Delta\lambda$ using Equation 7.3.

$$\Delta\lambda_{\text{deg}} = ((\lambda_2^* - \lambda_1^* + 180^\circ) \bmod 360^\circ) - 180^\circ \quad (7.3)$$

7.3.3. Interpolated Data

If Dataset 1 fulfilled all the thresholds described above, it was deemed suitable for use in this research. Subsequently, Dataset 2 (interpolated) was prepared. As Dataset 2 originates from Dataset 1, Dataset 1 was first required to meet the data coverage criteria.

Dataset 2 contained interpolated data at five-minute intervals. If a gap exceeded 30 minutes, it was not interpolated. Despite this interpolation, inconsistencies could still be present in Dataset 2 due to various factors, and the influence of these inconsistencies on the derived operational profiles was not known a priori.

Therefore, the datasets were subjected to a cleaning process. This cleaning procedure is described below. Following this, the data was repaired where possible. The method used for this cleaning and repairing process is also described in this section.

Cleaning method

Dataset 1 was only assessed for coverage and was not yet filtered for inconsistencies in the AIS data. Many inconsistencies that can occur in AIS data were already discussed in the literature study (chapter 3). Most of these inconsistencies had already been addressed by MSG [4] and were therefore not present in the datasets used in this research.

However, during this research and through comparison of different datasets, additional inconsistencies were identified. For each of these, a specific cleaning step was developed. The complete cleaning procedure is shown in Figure 7.5.

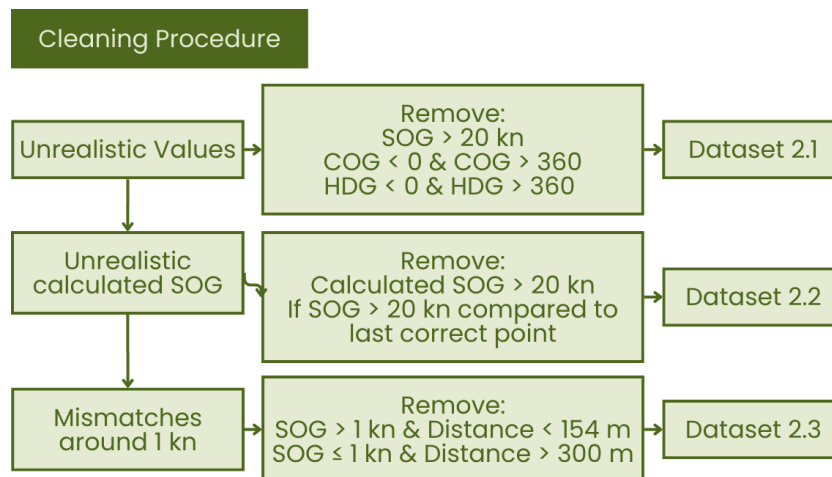


Figure 7.5: Cleaning procedure

The first step was to identify and remove unrealistic values for SOG, COG, and HDG, as these are among the most common inconsistencies in AIS data.

The second step focused on removing interpolation-related inconsistencies. These inconsistencies could occur during interpolation, particularly when switching between positive and negative longitude values. Such inconsistencies could result in unrealistic calculated SOG values. A threshold of 20 knots was applied, as this was considered unrealistic for the vessel types analysed in this research.

Finally, mismatches between the coordinates and the SOG were identified. It was acknowledged that AIS data is inherently imperfect and that coordinates and SOG do not always align exactly. As this research focuses on defining operational profiles, particular attention was given to mismatches around transitions between sailing and non-sailing conditions. Some of these mismatches could be explained by the interpretation of AIS data applied in this research.

To determine whether a vessel was sailing or stationary, both speed over ground and travelled distance were considered. A low speed should correspond to a small travelled distance, whereas a higher speed should correspond to a larger distance. Thresholds were defined to identify data points that did not follow this relationship (see Figure 7.5).

These thresholds were based on the following reasoning. A vessel travelling at 1 knot for five minutes should cover approximately 154 meters. Therefore, if a data point had a SOG greater than 1 knot but a travelled distance smaller than 154 meters, it was identified as a mismatch. The reverse situation was also considered, where the SOG was less than 1 knot but the travelled distance exceeded 154 meters.

However, AIS data is subject to noise, and small position deviations can occur, for example in ports due to interference from cranes or surrounding infrastructure. To account for this, a more conservative threshold of 300 meters was applied. At this distance, the vessel can still be considered stationary while allowing for measurement inaccuracies.

Several studies have identified the issue of unreliable SOG values due to interference, jitter, drift, and variations in reporting frequency. A commonly used method to address this is to calculate the average speed between consecutive data points, as applied by Ma et al. [32]. However, this approach assumes that the positional data is fully reliable. In the presence of small position deviations, this method may incorrectly interpret noise as vessel movement.

For this reason, this method was not applied in this research. Instead, the approach described above was used. In this method, most of the original AIS data was retained, while more attention was paid to data around transitions between sailing and non-sailing conditions. As these transitions are most relevant for defining operational profiles, uncertain data points were removed rather than placing greater trust in either the coordinates or the SOG.

To enable further analysis, all identified mismatches were stored in a separate CSV file. In addition, a cleaned dataset was generated for each vessel, in which implausible data points and mismatches around sailing transitions were removed. The Python workflow used to implement this method is shown in Appendix C.

Repairing method

During the cleaning procedure, data points were removed, resulting in additional gaps in the dataset. Therefore, the next step in the data preparation process was to repair the data. Before performing the repair, the gaps in the dataset were first analysed and classified.

Instead of using the interpolated dataset (Dataset 2), the cleaned dataset (Dataset 2.3) was used as input. This dataset originated from Dataset 2 but had been cleaned of implausible data points, interpolation inconsistencies, and mismatches.

Using the same definitions as described previously, the duration and distance between all consecutive data points were determined. In this phase, the minimum SOG was also introduced. This represents the minimum speed required to travel the calculated distance within the given duration and was used for gap classification.

Gaps in the data were categorized as stationary, moving, or unknown. A stationary gap was defined as a period in which the vessel likely remained in the same location but AIS data was missing. A moving

gap occurred when the vessel appeared to be sailing both before and after the gap. The final category, unknown, was used when the data before and after the gap was inconsistent, making it impossible to reliably estimate vessel behaviour during the missing period. The rules used for this classification are shown in Figure 7.6.

Stationary Gap	Moving Gap
Duration > 300 s	Duration > 300 s
Distance < 300 m	SOG - Min_SOG_kn ≤ 1 kn For 3 points before and after the gap
Min SOG < 0.5 kn	
SOG before and after the gap < 1 kn	

Figure 7.6: Thresholds used to determine the category of gaps in the data

For stationary gaps, the requirement that the minimum SOG remained below 0.5 knots ensured that only very limited movement was allowed. For shorter gaps, this resulted in stricter movement constraints, while for longer gaps slightly more movement was permitted, although the total displacement remained below 300 meters.

Only gaps with consistent behaviour before and after the gap were repaired, as it could reasonably be assumed that the vessel continued operating in a similar manner during the missing period. For gaps with inconsistent surrounding data, no repair was performed due to the high level of uncertainty. The Python workflow used for this classification is shown in Appendix C.

The next step was to fill the stationary and moving gaps. Stationary gaps were filled at five-minute intervals using the coordinates of the data point after the gap and a SOG of 0 knots. Additional parameters such as status, HDG, COG, ROT, draught, and destination were copied from the data point after the gap.

For moving gaps, spherical linear interpolation was applied due to its simplicity. For small gaps, this provides a reasonable approximation, as large vessels are unlikely to deviate significantly over short time intervals. For larger gaps, the vessel trajectory is unlikely to be perfectly linear; however, the reconstructed path represents a plausible minimum trajectory and average speed. A walkthrough of the Python implementation is provided in Appendix C.

For each moving gap, a trajectory was reconstructed on a sphere using a five-minute time interval. Spherical Linear Interpolation [33] was applied between the last data point before the gap ($i - 1$) and the first data point after the gap (i), each defined by latitude and longitude.

Latitude and longitude were first converted to radians (equation (7.4)):

$$\text{radians} = \text{degrees} \cdot \frac{\pi}{180} \quad (7.4)$$

These values were then transformed into 3D Cartesian coordinates (equation (7.5)):

$$\begin{aligned} x_i &= \cos(\text{lat}_i) \cdot \cos(\text{lon}_i) \\ y_i &= \cos(\text{lat}_i) \cdot \sin(\text{lon}_i) \\ z_i &= \sin(\text{lat}_i) \end{aligned} \quad (7.5)$$

The corresponding vectors were defined as (equation (7.6)):

$$\begin{aligned} v_i &= (x_i, y_i, z_i) \\ v_{i-1} &= (x_{i-1}, y_{i-1}, z_{i-1}) \end{aligned} \quad (7.6)$$

The angle between the vectors was calculated as (equation (7.7)):

$$\Omega = \arccos(v_{i-1} \cdot v_i) \quad (7.7)$$

The interpolation factor f was defined as:

$$f = \frac{t - t_{i-1}}{t_i - t_{i-1}} \quad (7.8)$$

Using this factor, interpolation weights a and b were calculated (equation (7.9)):

$$\begin{aligned} a &= \frac{\sin((1-f)\Omega)}{\sin(\Omega)} \\ b &= \frac{\sin(\Omega f)}{\sin(\Omega)} \end{aligned} \quad (7.9)$$

The interpolated Cartesian coordinates were then determined as (equation (7.10)):

$$\begin{aligned} x &= ax_{i-1} + bx_i \\ y &= ay_{i-1} + by_i \\ z &= az_{i-1} + bz_i \end{aligned} \quad (7.10)$$

Finally, the coordinates were transformed back to latitude and longitude (equation (7.11)), followed by conversion from radians to degrees (equation (7.12)):

$$\begin{aligned} lat &= 2(z, \sqrt{x^2 + y^2}) \\ lon &= 2(y, x) \end{aligned} \quad (7.11)$$

$$degrees = radians \cdot \frac{180}{\pi} \quad (7.12)$$

Remaining gaps were not repaired, as the vessel behaviour during these periods could not be reliably estimated. When calculating the percentage of time spent in different operational modes, these gaps were excluded, as their operational state was unknown. Due to the high overall data coverage, it was assumed that the influence of these gaps on the derived operational profiles was limited.

7.3.4. Comparisons

After data preparation, seven different datasets were available for all vessels included in this research. These datasets are listed below. In addition to determining the influence of different input datasets on the derived operational profiles, it was also important to understand how the individual processing steps affected the datasets themselves. Therefore, several comparisons were performed during this stage of the research.

- Dataset 1 = All transmitted data
- Dataset 2 = Interpolated data with a 5-minute interval

- Dataset 2.1 = Dataset 2 cleaned for unrealistic values
- Dataset 2.2 = Dataset 2.1 cleaned for interpolation inconsistencies
- Dataset 2.3 = Dataset 2.2 cleaned for mismatches around 1 knot
- Dataset 3 = Visual data in Plotter, a Made Smart tool [4]
- Dataset 4 = Dataset 2.3 repaired using stationary and moving interpolation

It should be noted that Dataset 1 did not represent raw AIS data, as it had already undergone preprocessing by MSG [4]. The data stream used in this research is shown in Figure 7.7.

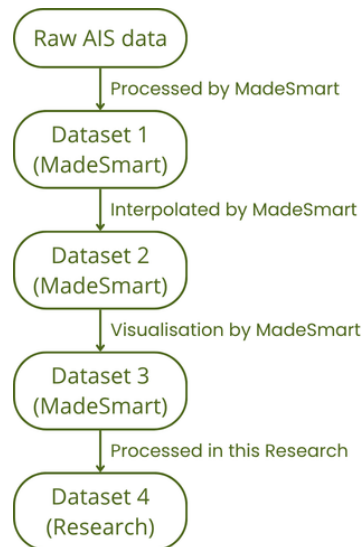


Figure 7.7: Data stream

Dataset 3 was not included in the quantitative comparison, as it only provided a visual representation of the data. The comparisons were also used to verify whether the datasets changed as expected after each processing step and, if not, to identify possible inconsistencies in the code or data. This comparison approach was applied repeatedly during the research to identify, explain, and remove remaining issues, thereby improving the data preparation method.

The datasets were collected per vessel and per year. When the data was analysed per year and subsequently aggregated, the results differed slightly from those obtained by merging all years into a single dataset. This difference arose because the merged dataset also included the duration and distance between the end of one year and the start of the next, whereas this transition was excluded when the years were analysed separately.

Dataset 1 contained too many data points to be merged efficiently into a single dataset and was therefore analysed per year only. For Dataset 2, both methods were applied: yearly totals and totals based on the merged dataset. This comparison was used to verify the correctness of both the data and the script. It was expected that the merged dataset would result in a slightly larger total distance and duration.

For all datasets derived after Dataset 2, only the merged method was used. When merging the data, gaps between years could occur. For these gaps, it was considered unrealistic to assign distance and duration. Therefore, no distance or duration was calculated for the first data point of a new year.

The first comparison was performed for all datasets except Dataset 3. The datasets were compared based on the number of data points, total distance, and total duration.

The number of data points was defined as the number of observations containing a value in the *Date and UTC Time* column. Total distance was calculated using the haversine equation, while total duration was determined as the sum of the durations between consecutive data points. Table 7.4 presents the expected relations between the datasets.

Table 7.4: Comparison expectations

	Data points [#]	Total Distance [nm]	Total Duration [h]
Dataset 1	Largest value	Large	Average
Dataset 2 (Per Year)	Smaller	Large	Smaller
Dataset 2 (Merged)	Smaller	Large	Larger
Dataset 2.1	Smaller	Smaller	Equal
Dataset 2.2	Smaller	Smaller	Equal
Dataset 2.3	Smaller	Smaller	Equal
Dataset 4	Larger	Equal	Equal

Based on the applied methods, Dataset 1 was expected to contain the largest number of data points. During the cleaning steps, the number of data points was expected to decrease, as points were only removed. In Dataset 4, the repair step introduced additional points, resulting in an increase.

For total distance, Dataset 1 or Dataset 2 was expected to contain the highest value, as these datasets had not yet been fully filtered for inconsistencies. After the cleaning steps, total distance was expected to decrease due to the removal of erroneous data points. Dataset 4 was expected to show a total distance similar to that of Dataset 2.3. A large deviation between these datasets would indicate that interpolation inconsistencies had likely been introduced during the repair process.

For total duration, merged datasets were expected to show slightly larger values because they included the transitions between years. However, the duration of all datasets was expected to remain similar as long as the first and last data points remained unchanged.

A second comparison was made between all variations of Dataset 2 to determine which cleaning step had the largest influence on the dataset and which inconsistencies occurred most frequently. Table 8.12 shows the types of inconsistencies that were compared. The results are discussed in the results chapter.

Table 7.5: Comparison of the cleaned datasets

	SOG >20 kn	COG >360 OR <0	HDG >360 OR <0	Calculated SOG >20 kn	Mismatch >1 kn	Mismatch ≤ 1 kn
Dataset 2.1	87	0	0	0	0	0
Dataset 2.2	0	0	0	520	0	0
Dataset 2.3	0	0	0	0	11125	2035

To determine the influence of the repair step, Dataset 2 through Dataset 4 were also compared on gap statistics. This comparison evaluated missing data, number of gaps, average gap duration, maximum gap, and average duration (Table 7.6).

Table 7.6: Comparison of gap statistics

	Duration = 300s [#]	Duration = 300s [h]	Duration >300s [#]	Duration >300s [h]	Average duration gap [h]	Average duration [s]	Max gap [h]	Missing Data [%]
Dataset 2	7,931,338	660,944	49,394	135,637	2.75	359.3	617	17.03
Dataset 2.1	7,931,203	660,933	49,446	135,648	2.74	359.3	617	17.03
Dataset 2.2	7,930,453	660,871	49,676	135,711	2.73	359.4	617	17.04
Dataset 2.3	7,907,733	658,978	59,236	137,604	2.32	359.9	617	17.27
Dataset 4	9,191,861	765,988	10,284	30,593	2.97	311.6	617	3.84

Missing data was calculated by comparing the duration of all data points with the duration of data points having an interval of exactly 300 seconds (Equation 7.13).

$$Data_{missing} = \frac{Duration_{gap}}{Duration_{total}} \cdot 100\% \quad (7.13)$$

To analyse the influence of cleaning and repairing visually, two types of graphs were created to compare Dataset 2 and Dataset 4. The first graph was the cumulative distribution of SOG, which shows how vessel speed was distributed and highlights the most frequently used speed ranges. For this graph, bins from 2 to 20 knots with a width of 0.5 knots were used. An example is shown in Figure 7.8a, where the median of both datasets is indicated by a vertical line.

The second graph was a ranked speed plot, of which an example is shown in Figure 7.8b. In this graph, all SOG values were ranked from low to high and plotted, with the y-axis representing SOG and the x-axis representing the number of data points. This provides an overview of vessel operation based purely on speed and illustrates how long a vessel operated within particular speed ranges.

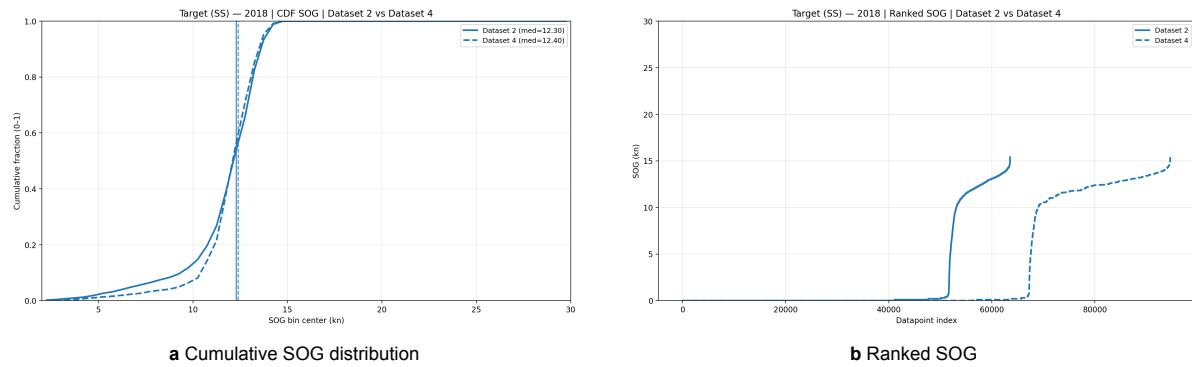


Figure 7.8: Example of cumulative SOG distribution and ranked SOG graphs

The comparison between Dataset 2 and Dataset 4 was performed per vessel and per year to clearly identify the influence of cleaning and repairing on the dataset. A walkthrough of the Python script used for this analysis is shown in Appendix C.

7.4. Operational profile

For the derivation of operational profiles, Dataset 2 and Dataset 4 were used. These datasets were selected to evaluate the influence of the cleaning and repairing procedures, not only on the datasets themselves but also on the resulting operational profiles. Figure 7.9 illustrates the different methods used to generate and compare the operational profiles.

First, the method used to obtain an operational profile without segmentation is described in subsection 7.4.1. Subsequently, the segmentation-based method for deriving operational profiles is presented in subsection 7.4.2.

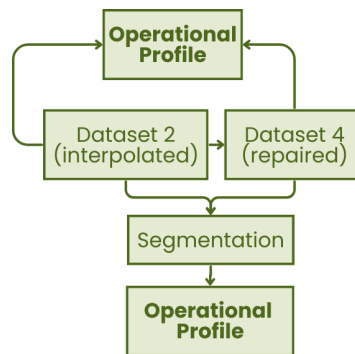


Figure 7.9: Methods used to obtain data-driven operational profiles

7.4.1. Unsegmented method

In the unsegmented method, the data was analysed as a whole, rather than being divided into separate operational segments. Dataset 1 was not used to construct operational profiles, as this dataset was not time-sampled and therefore did not represent vessel operations consistently over time. AIS reporting frequency varies with vessel speed, meaning that higher speeds generally result in more frequent transmissions. As a result, the speed distribution in Dataset 1 would not be representative of actual operational behaviour.

As described in the literature review, operational profiles aim to describe how vessels are operated. In this research, this analysis was performed at multiple levels: per vessel per year, per vessel across all available years, and for all vessels combined across all years. In addition, the results were analysed separately for HTVs and semi-submersibles, allowing differences and similarities in operational behaviour between these vessel categories to be identified.

The unsegmented operational profile described vessel behaviour using the following metrics:

- Total distance (nm)
- Average distance (nm)
- Maximum endurance (dd hh:mm:ss)
- Median SOG (kn)
- Speed bins including: number of data points, duration, distance, and mean SOG

The total distance was calculated as described previously. To determine the maximum sailing endurance, a speed threshold was applied. Consecutive data points with a SOG greater than 1 knot were identified, and their durations were summed. The longest continuous duration was defined as the maximum observed sailing endurance.

Instead of dividing the data into predefined operational states such as moored, anchored, manoeuvring, or sailing, the unsegmented method categorised the data into speed bins. This approach allowed vessel behaviour to be analysed directly based on the speed distribution.

To estimate the operational sailing speed, the median SOG was calculated. As low-speed values can significantly influence the median, multiple thresholds were evaluated. The median SOG was therefore determined for four different speed ranges: 1–25 knots, 2–25 knots, 3–25 knots, and 4–25 knots. This enabled assessment of the influence of low-speed data points on the estimated operational speed.

To further analyse vessel operation, statistics such as number of data points, duration, distance, mean SOG, and median SOG were determined for several SOG bins, as shown in Table 7.7. These statistics provided insight into how HTVs were operated and in which speed ranges most time and distance were accumulated.

Table 7.7: Output unsegmented operational profile speed distribution

Speed Bins for Statistics
0 - 0.5 kn
0.5 - 2 kn
2 - 4 kn
4 - 6 kn
6 - 20 kn
Total for HTV
Median 1 \pm 2 kn
Median 2 \pm 2 kn
Median 3 \pm 2 kn
Median 4 \pm 2 kn

The SOG bins were interpreted as representing different operational states. Speeds between 0 and 0.5 knots were considered stationary behaviour. Speeds between 0.5 and 2 knots were interpreted

as drift or AIS positional jitter rather than active sailing. The range between 2 and 4 knots was interpreted as manoeuvring behaviour. Speeds between 4 and 6 knots likely represented short sailing movements, such as sailing between anchorage and port. Speeds above 6 knots represented regular sailing operations.

In addition, these statistics were also determined for a speed range of ± 2 knots around the median operational speed. This provided insight into how much time and distance the vessel operated close to its typical sailing speed.

7.4.2. Segmented method

For the segmented operational profile, the data was first divided into segments, which were subsequently used to construct the operational profile. This method was applied only to Dataset 2 (interpolated) and Dataset 4 (repaired), as these datasets contained observations at a consistent interval of five minutes.

A review of the literature showed that identifying stopping and underway segments is not straightforward. Different studies apply different segmentation methods, indicating that the choice of method strongly depends on vessel type and research objective.

Several methods from previous studies were tested in this research but did not provide the desired results due to differences in vessel type and operational context. For example, segmentation between ports, as applied by Park and Choi [17], was not suitable for HTVs and semi-submersibles, as these vessels may transfer heavy cargo offshore rather than in ports.

Another commonly used approach is the application of a speed threshold. Wijaya et al. [34] used a threshold of 0.5 knots, since a speed of exactly 0 knots is unrealistic due to AIS jitter. However, relying solely on a speed threshold was found to be insufficient for the vessel types considered in this research. During anchored periods, the SOG could still reach values up to approximately 1.5 knots due to drift, while during manoeuvring or turning the SOG could also fall below this value. Furthermore, AIS jitter combined with a simple speed threshold could result in many small, fragmented segments.

This issue was also identified by Yan et al. [35]. Their method identified candidate stopping points using thresholds for speed, duration, and distance. Although these thresholds reduced the influence of drift and jitter, small segments could still occur. To address this, their method evaluated the distance and duration between segments and applied additional global thresholds, such as a minimum number of data points and a minimum segment duration.

The underlying concept of this method was considered robust. However, the threshold values proposed by Yan et al. [35] were not directly applicable in this research, as the datasets used here had a higher temporal resolution.

Therefore, this study adopted the general concept proposed by Yan et al. [35], while adapting the segmentation procedure by introducing multiple levels of filtering and vessel-specific thresholds.

Segmentation Level 1 – Initial classification The adapted segmentation method started with identifying candidate data points for stopping segments. A data point with a SOG lower than 1 knot was considered a potential stopping point.

Next, candidate points were required to satisfy additional distance and speed criteria. The distance between consecutive points had to remain below 300 meters, and the SOG had to remain below 2 knots. These threshold values were determined through iterative testing of different combinations and by visually comparing the resulting segments with vessel trajectories in Plotter [4].

The distance threshold of 300 meters allowed limited vessel movement while anchored, which could occur due to drift or environmental forces. Movements exceeding this threshold within the five-minute sampling interval generally corresponded to the start of a sailing segment. The SOG threshold of 2 knots allowed for AIS jitter and minor drift while still representing a stopping condition.

A stopping segment was terminated when a data point exceeded either the distance or SOG threshold. In addition, the segment was terminated when three consecutive data points had a SOG greater than 1.1 knots, indicating sustained vessel movement.

All remaining data points were classified as underway. This first level of segmentation is illustrated in Figure 7.10.

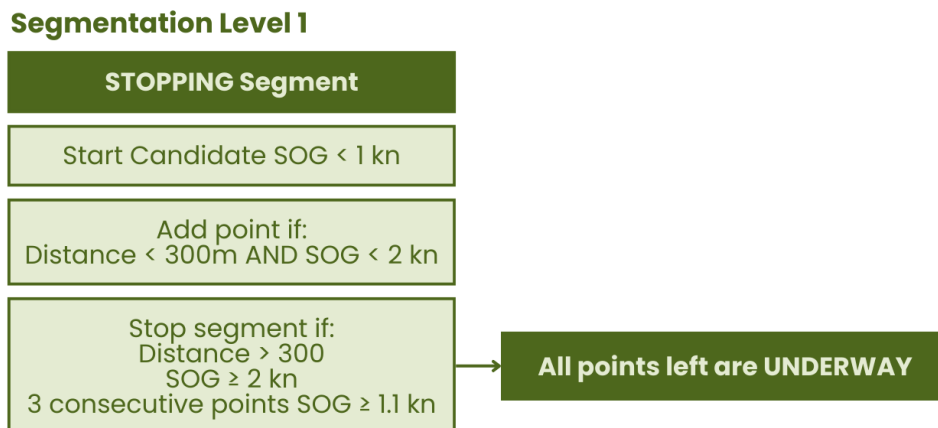


Figure 7.10: Segmentation Rules Level 1

Segmentation Level 2 – Merging small segments The second level aimed to merge small segments that were still created due to drift and jitter after the first segmentation step. Only segments containing three data points or fewer were considered.

If such a segment had a mean SOG lower than 2 knots, it was classified as a stopping segment. Segments with a mean SOG greater than 0.5 knots were classified as underway. Segments with mean SOG values between these thresholds retained their classification from the first segmentation level.

Changing the labels of these small segments could result in consecutive segments with the same classification. Therefore, all consecutive segments with the same label were merged when the time gap between them was less than one hour. This approach was based on the method proposed by Yan et al. [35]. The second segmentation level is illustrated in Figure 7.11.

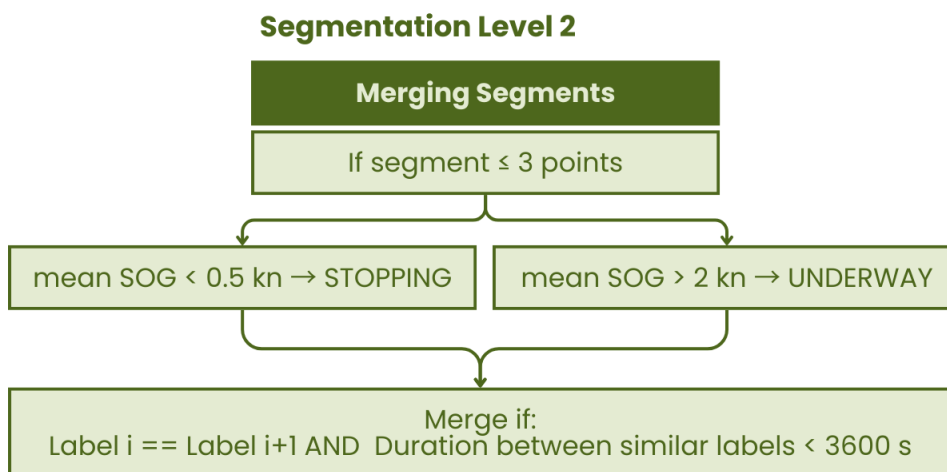


Figure 7.11: Segmentation Rules Level 2

Segmentation Level 3 – Behaviour classification The third level aimed to identify segments that clearly represented sailing or no-sailing behaviour. Segments were relabelled based on predefined thresholds (Figure 7.12).

A segment was classified as sailing when it contained at least 12 data points and had a mean SOG greater than or equal to 4 knots. This corresponded to approximately one hour of sustained sailing.

To capture shorter but clearly moving segments, a second criterion was applied: segments containing at least 4 data points with a mean SOG of at least 6 knots were also classified as sailing.

A segment was classified as no-sailing when it contained at least 72 data points and had a mean SOG less than or equal to 0.5 knots. This corresponded to approximately six hours of stationary behaviour, typical for cargo handling operations.

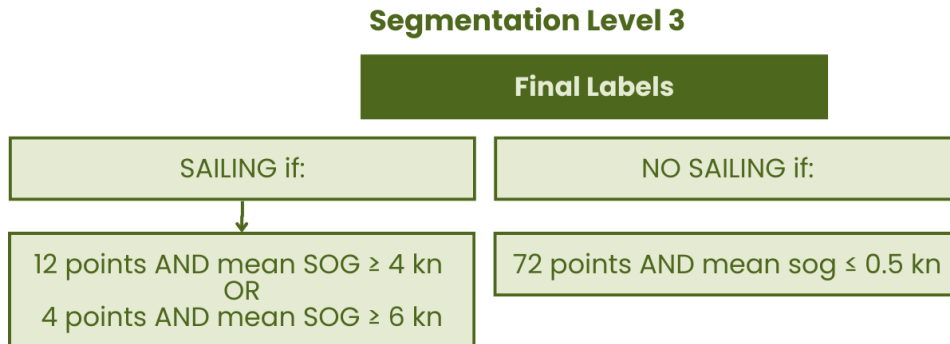


Figure 7.12: Segmentation Rules Level 3

Segmentation Level 4 – Maneuver identification During visual analysis of the AIS data using Plotter [4], vessel behaviour was observed that did not clearly fit within the previously defined sailing or no-sailing categories. Therefore, a fourth segmentation level was introduced. This behaviour was discussed with an operator of the Bigroll Bering [36] and was primarily attributed to Dynamic Positioning (DP) operations or trial runs to test onboard systems.

These maneuvering segments were identified as sequences of alternating stopping and underway behaviour, or as clusters of trajectory points with limited spatial spread but relatively high SOG.

After identifying these segments, their mean behaviour was evaluated. If a segment had a mean SOG lower than 0.1 knots, it was relabelled as no-sailing.

For segments with a mean SOG between 0.1 and 0.5 knots, the spatial spread was analysed by calculating the segment diameter. If the diameter was smaller than 1000 meters, the segment was considered anchored. If the diameter exceeded 1000 meters, the segment was classified as maneuvering.

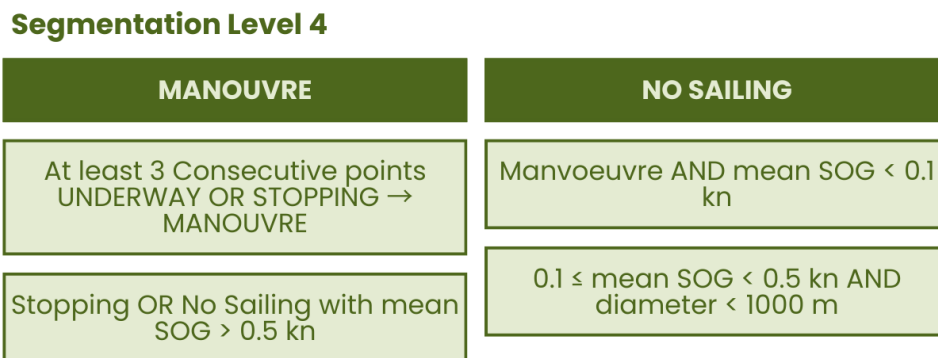


Figure 7.13: Segmentation Rules Level 4

Segmentation Level 5 – Final merging After level 4, sailing and maneuvering segments were strictly separated. However, small maneuvering segments could still occur within sailing trajectories. Therefore, maneuvering segments with a mean SOG greater than 0.5 knots were merged with adjacent sailing segments.

As a final step, all remaining underway segments were classified as sailing, and all stopping segments as no-sailing. Adjacent segments with the same label were merged when the time gap between them was less than one hour.

The resulting segmentation is shown in Figure 7.14.

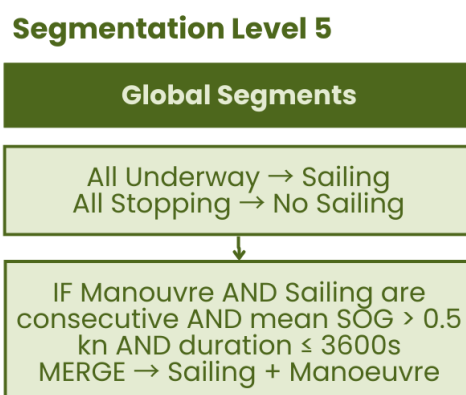


Figure 7.14: Segmentation rules level 5

Subdivision of operational states After segmentation, all data points were classified into sailing, maneuvering, and no-sailing segments. These were further subdivided to better represent operational behaviour.

Sailing segments were divided into long sailing and short sailing, using a threshold of 6 hours. Segments longer than 6 hours were classified as long sailing, while shorter segments were classified as short sailing.

No-sailing segments were divided into moored and anchored. The AIS navigational status was used for this classification, in combination with a mean SOG threshold of 0.5 knots.

Segments with a mean SOG lower than 0.5 knots and a majority AIS status of “moored” were classified as moored. Visual validation using Plotter and OpenStreetMap confirmed that this classification was sufficiently accurate.

Segments with a mean SOG lower than 0.5 knots without a majority “anchored” or “moored” status were classified as stalling, representing temporary waiting behaviour during operations.

After segmenting all vessels, the different operational profiles were constructed. Figure 7.15 shows the operational profiles generated based on the selected vessel datasets.

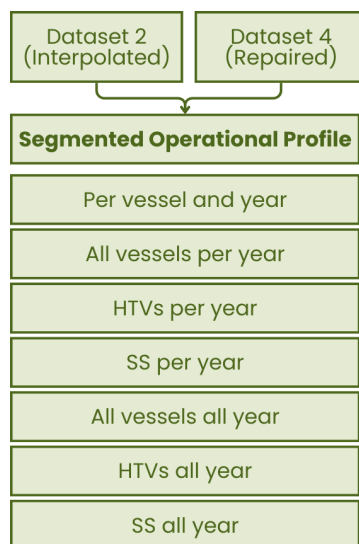


Figure 7.15: Operational profiles with different vessel input

All operational profiles had the same structure but differed in the vessels and datasets used as input. The structure of the operational profiles is described below.

- Total duration
- Number of segments
- Total distance in nm
- Endurance
- Endurance table with reduced duration
- Endurance table with reduced number of segments

The first endurance table evaluated the effect of reducing the maximum endurance duration. Starting from the longest identified sailing duration, the endurance duration was reduced in steps of 10% until 20% of the original duration remained. For each endurance duration, the number of sailing segments that could be completed without stopping was determined.

A second endurance table was constructed based on the percentage of segments covered. In this table, the longest sailing segment represented the endurance required to cover 100% of the segments. Subsequently, the endurance duration required to cover 99% of the segments was determined, followed by 98%, and so on until 90%.

These endurance tables allowed the designer to evaluate different endurance requirements when constructing the operational profile. The endurance was based on historical vessel data and was therefore not overestimated compared to reality. However, designing a vessel to cover 100% of all sailing segments without stopping may not be desirable. For example, a single vessel with exceptionally large endurance or segmentation inconsistencies in the longest sailing segment could result in unrealistic design requirements. Therefore, reduced endurance scenarios were also analysed.

At this stage, all remaining no-sailing segments that were not classified as moored or anchored were labelled as stalling. These segments typically occurred when the vessel temporarily stopped sailing without anchoring or mooring, for example when arriving early at a destination.

For each operational mode listed in Table 7.8, the duration in hours, the percentage of time spent in that mode, the average SOG, the travelled distance in nautical miles, and the corresponding percentage were determined. These statistics provided insight into how much time the HTV fleet operated in each mode, as well as the mean SOG and distance travelled within each operational mode.

Table 7.8: Output of segmented operational profile based on modes

Operational Modes
Long Sailing
Short Sailing
Manoeuvre
Stalling
Moored
Anchored

An additional analysis was performed in which sailing segments were divided into long sailing and short sailing segments, where long sailing was defined as a duration of 6 hours or more. This distinction allowed the analysis to focus on transport voyages, excluding short repositioning movements such as sailing between anchorage and port.

Table 7.9 presents the statistics of the travelled distance per sailing trip. When a vessel consistently sailed similar distances per trip, this was reflected in a relatively small standard deviation. These results will be discussed in the next chapter.

Table 7.9: Distance statistics per trip for the Biglift Baffin in 2019

Distance (nm)	Min	Max	Mean	Median	Standard Deviation
Long Sailing	46.52	13936.34	2695.33	1417.33	3862.10
Short Sailing	10.14	52.13	26.82	24.00	15.95
Manoeuvre	-	-	-	-	-
Long + Short Sailing	10.14	13936.34	1627.93	288.48	3230.26
Long + Short + Manoeuvre	10.14	13936.34	1627.93	288.48	3230.26

A similar analysis was performed for moored, anchored, and stalling segments. Instead of analysing travelled distance, the duration of these segments was considered, as no significant distance was travelled during these operational modes.

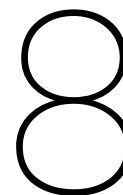
7.5. Conclusion Methodology

The overall methodology used to answer the main research question is illustrated in Figure 7.16. First, a method for vessel selection was applied, for which AIS data was provided by MSG [4]. Subsequently, the data was prepared through a series of cleaning and processing steps to ensure consistency, reliability, and alignment with the interpretation of AIS data.

Finally, a method to derive operational profiles from the processed data at both fleet level and vessel level was applied. These profiles were constructed by categorizing the data into speed bins or operational modes. For each category, the distribution of time spent and distance travelled was determined.

The results of the data-driven operational profiles for HTVs, following the described methodology, are presented and discussed in the next chapter.

**Figure 7.16:** Global Methodology



Results

8.1. Fleet Selection Results

For the fleet selection, the filtering steps described in Figure 7.1 were applied. After filtering based on vessel status and vessel type, the dataset was reduced from 266,086 vessels to 1,131 vessels. Further filtering based on vessel length and breadth reduced this number to 230 vessels. A final filter based on crane presence and container capacity resulted in 110 vessels of interest, consisting of HTVs and semi-submersibles.

After the manual classification of HTVs, semi-submersibles, and other vessel types, the list of currently operating vessels consisted of 61 HTVs and 47 semi-submersibles. The vessels listed in Table 8.1 were selected for this research due to their known involvement in offshore wind operations. Some of these operations are shown in Appendix D.

Table 8.1: List of vessels used in this research

Name	Type	Build Year	LOA (m)	B (m)	D (m)	Deck Space (m)	Trial Speed (kn)	Draft (m)	DWT (t)	Endurance (nm)
Bigroll Beaufort	HTV	2017	173.00	42.00	12.00	5,250	13	6.50	20,675	
Bigroll Bering	HTV	2016	173.00	42.00	12.00	5,250	13	6.50	20,675	
Mega Caravan	HTV	2011	163.80	54.00	10.50	7,560	12	6.00	16,200	
Fan Zhou 8	HTV	2024	256.00	51.00	13.00	11,700	12	8.30	58,405	16,000
Fan Zhou 6	HTV	2018	168.53	36.00	10.20	4,750	12	7.00	19,500	13,000
Dongbang No.6	HTV	2012	152.50	40.00	9.00	5,000		5.25	15,016	
Dongbang No.7	HTV	2016	166.00	42.00	9.50	5,568		6.00	19,400	
Audax	HTV	2016	206.30	43.00	13.50	7,482	13	7.50	33,033	13,000
Biglift Baffin	HTV	2016	173.00	42.00	12.00	5,250	13	6.50	20,675	
Biglift Barentsz	HTV	2016	173.00	42.00	12.00	5,250	13	6.50	20,675	
Zhong Yu 188	HTV	2022	165.00	42.00					25,003	
GPO Grace	SS	2017	225.00	48.00	13.80	8,784	14.5	10.64	65,000	25,000
Blue Marlin	SS	2000	224.80	63.00	13.30	11,227	13	10.24	76,292	
Target	SS	2007	216.79	44.50	14.00	5,785	15	10.44	54,000	
Hua Hai Long	SS	2012	181.90	43.60	11.00	6,278	9	7.80	31,726	
Fan Zhou 10	SS	2019	239.60	48.00		9,523				
Sun Shine	SS	2008	174.00	40.00	10.25	7,104	11.5	5.00	17,824	

The final fleet selection for this research consisted of 11 HTVs and 6 semi-submersibles. All of these vessels were verified to have been in operation for at least one full year within the period from 2018 to the present. However, not all vessels had been in operation since 2018.

8.2. Data Collection from MadeSmart Group Results

All vessels listed in Table 8.1 were reviewed in Prospector, as described in section 7.2. Data previews that did not show sufficient vessel activity during a given year were excluded from further analysis.

The data for *Audax* in the years 2018 and 2019 were removed from the dataset, as the vessel showed almost no operational activity during these years. This can be observed in Appendix E. In addition, the data for *Fan Zhou 8* only showed activity related to trial runs and was therefore also excluded from this research.

This resulted in the final set of datasets shown in Table 8.2.

Table 8.2: Datasets collected for this research

Vessel Name	Years	# Datasets	Vessel Name	Years	# Datasets
Bigroll Bering	2018 - 2024	7	GPO Grace	2018 - 2024	7
Bigroll Beaufort	2018 - 2024	7	Blue Marlin	2021 - 2025	5
Mega Caravan	2018 - 2024	7	Target	2018 - 2024	7
Fan Zhou 8	2025	1	Hua Hai Long	2020 - 2024	5
Fan Zhou 6	2019 - 2024	6	Fan Zhou 10	2022 - 2025	4
Dongbang Giant No.6	2018 - 2024	7	Sun Shine	2024 - 2025	2
Dongbang Giant No.7	2018 - 2024	7			
Audax	2020 - 2024	5			
Biglift Baffin	2018 - 2024	7			
Biglift Barentsz	2018 - 2024	7			
Zhong Yu 188	2023 - 2025	3			

In total, 94 datasets were downloaded as text files. For each dataset, both the original “All Data” (Dataset 1) version and an interpolated version (Dataset 2) were obtained. The interpolated datasets contained a data point every five minutes, with gaps of up to 30 minutes filled through interpolation.

Not all vessels included data for 2025, as this research started in 2025 and the corresponding data was not yet fully available. Since the main focus of this research was on HTVs, data for these vessels was collected for the period from 2018 to 2024. For the additional semi-submersible vessels, only more recent data was used.

Downloading the data for one vessel for a single year took approximately 30 minutes.

8.3. Data Preparation Results

As described in section 7.3, the first step in this research was to evaluate the coverage of Dataset 1, which is discussed in subsection 8.3.1. Next, Dataset 2 (interpolated) was prepared using a cleaning and repairing procedure, which is discussed in subsection 8.3.2. Lastly, the prepared datasets were compared on different levels in subsection 8.3.3 to evaluate the influence on the dataset and to verify whether the data reacted logically to the applied methods.

8.3.1. Coverage of Dataset 1 results

To evaluate Dataset 1, the data needed to fulfill the requirements defined in Table 7.3. The data used in this research was provided by MSG [4], which currently offers one of the best available global AIS datasets. However, the global AIS network based on satellite reception only fully matured between 2020 and 2022. As a result, the coverage in earlier years may be lower compared to more recent years. Therefore, it was not expected that all datasets would fully comply with the defined thresholds, as these thresholds represent very high coverage requirements.

Dataset 1 was analyzed using a Python script, of which the methodology is described in subsection 7.3.2. Out of these 94 datasets, three did not meet the maximum gap threshold. These datasets are listed in Table 8.3.

Table 8.3: Datasets that do not fulfill the initial thresholds

Ship Name	Year	Failed Rule
Audax	2020	Max Gap (1727:38:16)
Biglift Baffin	2021	Max Gap (800:54:19)
GPO Grace (SS)	2020	Max Gap (784:44:11)

The first vessel listed in Table 8.3 that was not fully compliant was the *Audax* in 2020. This dataset contained a gap of approximately 71 days, which corresponded to almost 20% of the entire year. During this period, the vessel sailed from a port in Taiwan to an anchorage in Singapore. Due to this large gap, there were too many uncertainties regarding the exact sailing behaviour and the moment of arrival at the anchorage. These uncertainties might significantly influence the yearly statistics of the vessel. Therefore, it was decided to remove these three datasets from the overall dataset used in this research.

It can be observed that these maximum gaps occurred around the COVID-19 period, which may explain the large gaps in the data. However, this could also be explained by a maintenance period, and the exact reason was not known. Overall, the datasets provided by MSG [4] showed very high coverage. All datasets contained at least 97% of the observations with a gap smaller than 30 minutes. On average, 90.7% of the observations had a duration shorter than 5 minutes.

8.3.2. Dataset 2 (Interpolated) procedure results

Dataset 2 was cleaned and repaired for this research. The method described in subsection 7.3.3 was applied to the vessels discussed above. The results leading to the cleaning method are discussed, as well as the identified gaps in the cleaned dataset.

Dataset 2.1 (Cleaning step 1)

To create Dataset 2.1, all unrealistic values for SOG, COG, and HDG were deleted from Dataset 2. Only 0.001% of all data points in Dataset 2 were removed due to an unrealistic value in SOG. No unrealistic values were found for COG and HDG in this dataset, but these variables were still included in the cleaning procedure because literature has shown that such inconsistencies can also occur.

Dataset 2.2 (Cleaning step 2)

When comparing the travelled distance between Dataset 1 and Dataset 2 for the vessel *Biglift Baffin* in 2019, an unexpected discrepancy was observed. Although Dataset 2 contained fewer data points due to the five-minute interpolation interval, the total travelled distance increased by approximately 56% (Table 8.4). Such a large increase was unrealistic and indicated that interpolation could introduce inconsistencies in the travelled distance.

Table 8.4: Distance comparison between Dataset 1 and Dataset 2 for the *Biglift Baffin* in 2019

	Data points	Total Distance [nm]
Biglift Baffin 2019 Dataset 1 (All)	89883	32,822
Biglift Baffin 2019 Dataset 2 (Interpolated)	68464	51,395

To investigate this behaviour, a small section of the interpolated dataset was analysed. A snapshot of this data can be found in Appendix F. In this section of AIS data, several unrealistic position changes between consecutive interpolated points were observed. These position jumps led to extremely large travelled distances within a short time interval and therefore resulted in unrealistically high calculated speeds.

To determine whether these jumps originated from the interpolation process or from the original AIS data, the same time period was analysed in Dataset 1. In the original dataset, no abnormal position changes were observed. This indicated that the errors were introduced during the interpolation step rather than being present in the original AIS messages.

Plotting the interpolated coordinates in Figure 8.1 on a map clearly showed unrealistic jumps in vessel position occurring within a time span of only five minutes. These jumps occurred when the vessel crossed the longitude boundary around $\pm 180^\circ$, where the interpolation algorithm incorrectly interpolated across the globe.

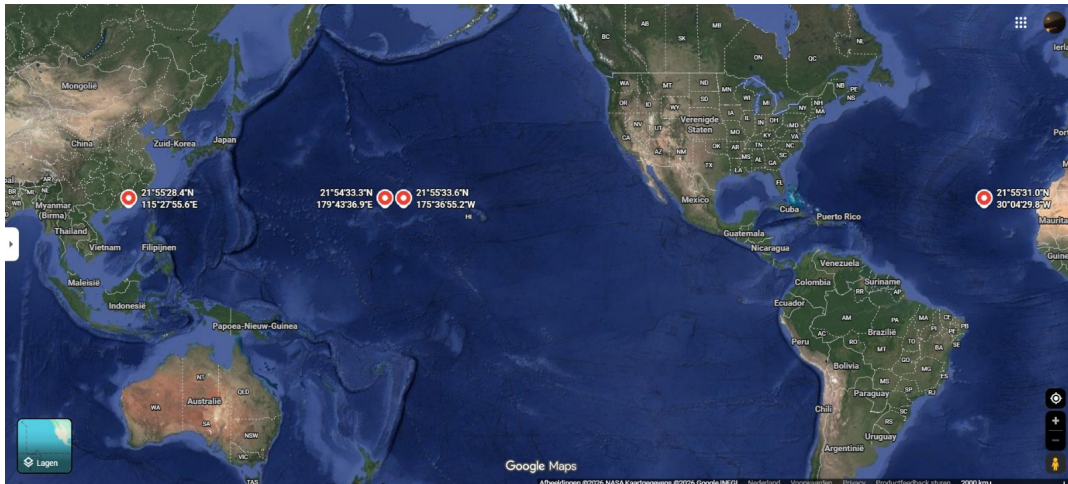


Figure 8.1: Consecutive data points in Biglift Baffin 2019 interpolated data

This example illustrates how interpolation errors could significantly influence the calculated travelled distance. Therefore, the cleaning threshold for the calculated SOG was applied to Dataset 2.1 before further analysis, resulting in Dataset 2.2.

When considering all data, only 0.006% of the data points exceeded the thresholds and were removed from Dataset 2.1. This showed that not every dataset contained this error. However, when fewer datasets were considered together, the influence could still be significant, as shown above.

Dataset 2.3 (Cleaning step 3)

In the final cleaning step, mismatches around the transition from non-sailing to sailing, or vice versa, were identified and removed. In total, 0.16% of the data was identified as a mismatch and removed. Table 8.5 shows that most mismatches exceeded Rule 1, where the SOG is larger than 1 knot but the distance is less than 154 meters. This error can occur due to jitter in the SOG data. When Rule 2 was exceeded, it could be explained by Guo et al. [37], who describe that in port areas positional deviations exceeding 150 meters can occur due to signal obstruction or electromagnetic interference.

Table 8.5: Mismatch results for all data combined

	Count	Percentage	Avg SOG (kn)	Avg Distance (m)
All Mismatches	13,160	0.16 %	1.650	134.80
Rule 1 (SOG > 1)	11,125	84.5 %	1.839	72.16
Rule 2 (SOG ≤ 1)	2,035	15.5 %	0.623	477.27

Overall, it can be seen that the mismatches mostly occurred at low SOG values. As the average SOG of the mismatches was around 1 knot, the majority of these mismatches were relatively small and could be attributed to the inherent imperfections of AIS data. However, given the predefined thresholds used for trajectory segmentation, it was decided to still remove these mismatches. Table 8.6 shows the SOG ranges in which these mismatches occurred. This also shows that some mismatches represented more significant errors, as speeds from 2 to 15 knots should correspond to distances larger than 154 meters.

Table 8.6: Mismatches occurring in certain SOG ranges for all data

SOG	Count	Percentage
0 – 1 kn	2,035	15.46 %
1 – 2 kn	9,259	70.36 %
2 – 6 kn	1,413	10.74 %
6 – 15 kn	441	3.35 %
>15 kn	12	0.09 %

The mismatches with higher SOG values were more difficult to explain. An example of such a mismatch can be seen in Table 8.7. The table shows positions that remained nearly constant over time while the reported SOG remained at approximately 10 knots. At the same time, the vessel status was labelled as moored.

This combination strongly suggested an error in the reported SOG values, likely caused by a malfunction in the onboard measurement system or an issue during the transmission of data from the onboard system to the onshore AIS network.

Table 8.7: Example from Bigroll Bering 2023 showing an unexplained mismatch

Date and UTC Time	Latitude (°)	Longitude (°)	SOG (kn)	Status	Duration (s)	Distance (m)
1/25/2023 13:15:00	44.030633	10.042683	10	Moored	300	4.656
1/25/2023 13:20:00	44.030633	10.042682	10	Moored	300	0.079
...
1/25/2023 13:45:00	44.030630	10.042683	10	Moored	300	1.890
1/25/2023 13:50:00	44.030630	10.042667	10	Moored	300	1.322

Again, the influence of such errors on a single dataset containing such a mistake could be more significant than when many datasets were combined.

Dataset 4 (Repaired)

The cleaning procedure resulted in 0.24% additional missing data. As described in the repairing procedure, all gaps were identified as either stationary, moving, or unknown.

Table 8.8 shows an example of a stationary gap in the dataset for the vessel Mega Caravan. The gap spans almost four hours. During this period, the vessel's position barely changed. The speed was zero and the heading remained constant. The observations before and after the gap were very similar, which strongly suggested that the vessel remained stationary during this time. The vessel status before and after the gap was also similar. Although the status field is manually entered and therefore less reliable, it further supported the assumption that this gap represented a stationary period.

Table 8.8: Example of a stationary gap of Mega Caravan in 2023

Date & UTC Time	Latitude (°)	Longitude (°)	SOG (kn)	HDG (°)	COG (°)	Status (-)	Duration (s)	Distance (m)	Min SOG (kn)
12/01/2023 08:45	34.926298	128.600245	0	233	334.8	Moored	300	0.18	0.001
12/01/2023 12:35	34.926245	128.600245	0	233	175.5	Moored	13800	5.96	0.001

Table 8.9 shows an example of a gap that could be identified as a moving gap. In this research, gaps classified as moving were considered sufficiently reliable to be filled using linear interpolation on a sphere. This assumed that the vessel was sailing mostly in a straight line during the gap, without significant maneuvering.

Table 8.9: Example of a moving gap in the AIS data of Mega Caravan in 2023

Date & UTC Time	Latitude (°)	Longitude (°)	SOG (kn)	HDG (°)	COG (°)	Status (-)	Duration (s)	Distance (m)	Min SOG (kn)
7/3/2023 11:45	36.546743	122.92831	10.3	141	134.9	Under way	300	1652.35	10.7
7/3/2023 11:50	36.536603	122.940883	10.3	141	135.3	Under way	300	1591.52	10.3
7/3/2023 11:55	36.526465	122.953455	10.3	141	135.7	Under way	300	1591.40	10.3
7/3/2023 13:10	36.388128	123.162413	10.4	135	132.5	Under way	4500	24204.48	10.5
7/3/2023 13:15	36.377292	123.173903	10.2	143	139.5	Under way	300	1584.23	10.3
7/3/2023 13:20	36.36626	123.185095	10	147	143	Under way	300	1583.95	10.3

All moving and stationary gaps were repaired, resulting in 13.43% more data being filled. Overall, after the repair only 3.84% of the data remained missing. Table 8.10 shows that most gaps were identified as moving. It also shows that unknown gaps were on average longer, which is consistent with the logic that very long gaps contain more uncertainty and are therefore more difficult to repair accurately.

Table 8.10: Gap identification

	Data points [#]	Percentage [%]	Duration [h]	Average Duration [h]
Total	59,236	100	137,603	2.32
Stationary	21,580	36.43	41,844	1.94
Moving	27,372	46.21	65,166	2.38
Unknown	10,303	17.39	30,593	2.96

Overall, only 3.84% of the data remained missing. These gaps corresponded to time periods where the observations before and after the gap were not consistent enough to reliably estimate vessel behaviour during the missing period. The repaired dataset formed Dataset 4, which was used for deriving the data-driven operational profiles.

8.3.3. Comparison between datasets results

Table 8.11 shows the overall effect of the applied method on all datasets. As expected, Dataset 1 contained significantly more data points than the other datasets. During each cleaning step the number of data points decreased, until Dataset 4 where the repairing procedure introduced additional data points. This behaviour was consistent with the methodology, as the cleaning steps only removed data points while the repair step added new observations through interpolation.

For the total distance it can be observed that Dataset 4, which was both cleaned and repaired, showed the smallest total distance. However, the reduction compared to Dataset 1 was only 2.8%. Considering that Dataset 4 had been cleaned for unrealistic values, interpolation errors, and mismatches, this small reduction indicates that the final dataset still represented the original vessel movements well while removing implausible observations.

Removing unrealistic values resulted in a reduction of 0.66% in the total distance. The removal of interpolation errors led to a reduction of 6.7% in the travelled distance, indicating that this type of error had the largest overall influence on the calculated distance. Removing mismatches resulted in a decrease of only 0.01% in the total distance. Although these mismatches had little influence on the total distance, they could still be important during the segmentation of operational profiles.

Furthermore, it can be observed that the total duration remained nearly constant across all datasets. This indicates that the datasets were consistently imported and processed. The only increase in duration occurred in the merged datasets, where transitions between years were also included, while these

transitions were ignored when yearly results were summed separately.

Table 8.11: Comparison between all datasets used in this research

	Data points [#]	Total Distance [nm]	Total Duration [h]
Dataset 1	30,823,103	3,347,457	795,930
Dataset 2 (Per Year)	7,980,755	3,488,425	795,711
Dataset 2 (Merged)	7,980,755	3,492,007	796,582
Dataset 2.1	7,980,668	3,486,851	796,582
Dataset 2.2	7,980,148	3,252,744	796,582
Dataset 2.3	7,966,988	3,252,355	796,582
Dataset 4	9,202,164	3,252,355	796,582

A more detailed analysis of the cleaning process is shown in Table 8.12. The results indicate that most of the AIS data obtained from MSG [4] was already well processed. Only a very small number of observations contained unrealistic SOG values, although these errors remained important to detect as they are commonly reported in AIS datasets.

During cleaning step 2 only 0.01% of the observations were removed. Although this percentage was very small, these observations could have a large influence on the calculated distance due to unrealistic position jumps. In contrast, the mismatch cleaning step removed approximately 0.4% of the data points, which was about forty times more observations. However, these mismatches had a much smaller influence on the total travelled distance.

Table 8.12: Comparison of the cleaned datasets

	SOG >20 kn	COG >360 OR <0	HDG >360 OR <0	Calculated SOG >20 kn	Mismatch >1 kn	Mismatch ≤ 1 kn
Dataset 2.1	87	0	0	0	0	0
Dataset 2.2	0	0	0	520	0	0
Dataset 2.3	0	0	0	0	11125	2035

Finally, Table 8.13 shows the influence of the cleaning and repairing procedures on the gap statistics of the datasets. Cleaning the data increased the number of gaps, as removing incorrect observations introduced additional small gaps. This effect can be observed in the decreasing average gap duration. Most of these newly created gaps were relatively small and could therefore be repaired through interpolation. After the repairing procedure the number of gaps decreased significantly, as stationary and moving gaps were filled. The maximum gap remained unchanged because very large gaps were not repaired, since interpolation over such long periods would lead to unrealistic vessel behaviour.

Table 8.13: Comparison between datasets for gap statistics

	Duration = 300s [#]	Duration = 300s [h]	Duration >300s [#]	Duration >300s [h]	Average duration gap [h]	Average duration [s]	Max gap [h]	Missing Data [%]
Dataset 2	7,931,338	660,944	49,394	135,637	2.75	359.3	617	17.03
Dataset 2.1	7,931,203	660,933	49,446	135,648	2.74	359.3	617	17.03
Dataset 2.2	7,930,453	660,871	49,676	135,711	2.73	359.4	617	17.04
Dataset 2.3	7,907,733	658,978	59,236	137,604	2.32	359.9	617	17.27
Dataset 4	9,191,861	765,988	10,284	30,593	2.97	311.6	617	3.84

Repairing the data influenced the datasets when looking at the median SOG. Figure 8.2 shows the cumulative distribution of the SOG for the vessel Biglift Baffin. The solid line represents Dataset 2 and the dashed line represents Dataset 4.

The figure shows that cleaning and repairing the dataset could have a significant influence on some datasets. For example, in 2018 there was a difference of approximately 1 knot in the median SOG. However, it can also be observed that not all datasets were affected to the same extent by the cleaning and repairing procedures. For instance, in 2022 only a difference of 0.09 knots was observed in the median SOG.

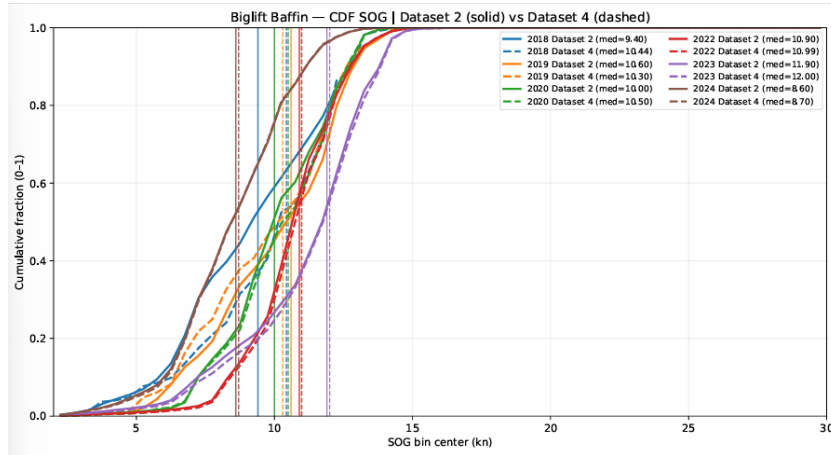


Figure 8.2: SOG CDF for Biglift Baffin comparing Dataset 2 and Dataset 4

When analyzing the ranked speed plot for Biglift Baffin in Figure 8.3, it can be observed that, except for the years 2022 and 2024, a considerable number of low-speed data points had been added in Dataset 4. Such a plot provides a clear indication of how the vessel was operated. If the curve remains long and relatively flat at the beginning, this indicates that the vessel spent a significant amount of time in a non-sailing state. For example, in 2020 the vessel spent the largest amount of time in a non-sailing condition compared to the other years.

Furthermore, it can be observed that the most frequently sailed speeds were higher in 2022 and 2023 compared to 2024. If the line increased very steeply, it indicates that the vessel spent little time within that speed range. In contrast, if the line was relatively horizontal, it indicates that the vessel spent a considerable amount of time operating at that speed. Overall, it can be observed that the operational behaviour of Biglift Baffin differed significantly between years. Therefore, it was expected that a single operational profile might not accurately represent all years of operation. The CDF and ranked speed plots for all the other vessels can be seen in Appendix G.

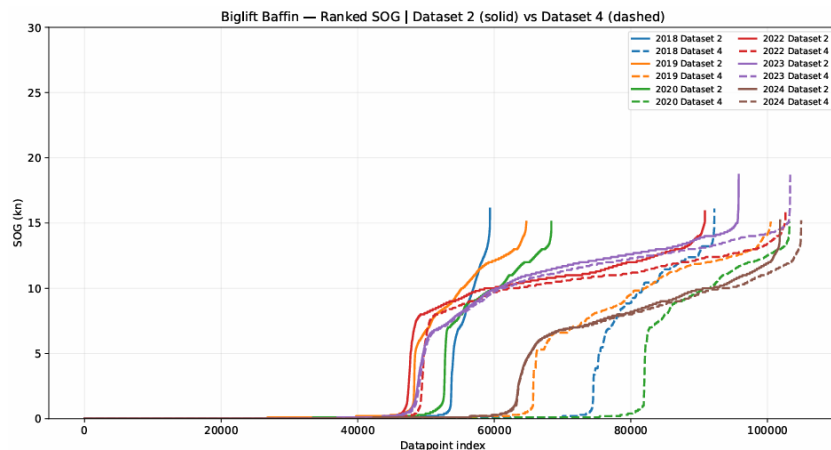


Figure 8.3: Ranked speed distribution for Biglift Baffin comparing Dataset 2 and Dataset 4

Overall, the cleaning and repairing procedures had a noticeable effect on the datasets themselves, particularly when considering the amount of missing data and the distribution of vessel speeds. Whether these changes also significantly influenced the operational profiles and potential ship design implications when combining multiple datasets is discussed in the following section.

8.4. Operational Profile Results

Operational profiles were calculated for each vessel per year. However, the main focus of this research was on deriving an overall operational profile for the HTV fleet and several semi-submersibles known to operate in the offshore wind market.

In subsection 8.4.1, the operational profiles for HTVs and semi-submersibles are discussed using the unsegmented method. These profiles were derived using Dataset 2 and Dataset 4 as input, allowing the influence of the cleaning and repairing procedures on the resulting operational profiles to be evaluated. In subsection 8.4.3, the same comparison is performed using the segmented method. In this approach, operational profiles were determined by first segmenting the data into operational states before analyzing the operational behavior of the HTV and semi-submersible fleet.

8.4.1. Unsegmented method results

As explained in subsection 7.4.1, the operational behavior of the HTV fleet is described by analyzing the total distance, average distance, maximum endurance, median SOG, and speed bins representing the data distribution. This analysis was performed using both Dataset 2 and Dataset 4 as input in order to evaluate the influence of cleaning and repairing on the derived operational profiles.

First, the operational profiles derived for HTVs based on Dataset 2 and Dataset 4 are compared with each other. Next, the same comparison is performed for the semi-submersibles. Finally, the differences between the operational profiles of HTVs and semi-submersibles obtained using the unsegmented method are briefly discussed.

HTVs

In this research, 11 HTVs were analyzed. These vessels had been in operation for at least one year during the period from 2018 to 2025. Table 8.14 shows the average distance sailed by the HTVs, comparing Dataset 2 and Dataset 4.

As expected, Dataset 4 resulted in a lower total sailed distance due to the cleaning and repairing procedures. This effect can clearly be observed in the table. When looking at the average across all vessels and all years combined, the cleaning and repairing procedures resulted in 7.28% less sailed distance.

However, it can also be observed that the influence varied per year. In some years the difference was almost twice as large, indicating that the AIS data for those years likely contained more interpolation errors or mismatches. In contrast, for 2025 no difference between the datasets was observed.

Overall, when using Dataset 2 instead of Dataset 4, the estimated sailed distance per trip would on average be 7.28% higher. This indicates that the influence of the cleaning and repairing procedures on the overall sailed distance was relatively limited.

Table 8.14: Average sailed distance comparison between Dataset 2 and Dataset 4 for HTVs (unsegmented)

Distance [nm]	2018	2019	2020	2021	2022	2023	2024	2025	Average
Dataset 2	28,893	36,631	35,813	40,418	45,838	38,417	34,155	43,671	37,980
Dataset 4	27,554	33,543	34,578	35,679	39,790	34,319	32,595	43,670	35,216
Percentage Difference	4.63%	8.43%	3.45%	11.73%	13.20%	10.67%	4.57%	0.00%	7.28%

For ship design, it is also important to know the maximum endurance. In Table 8.15, the longest sequence of consecutive data points with SOG > 1 knot per year can be seen. Again, it can be observed that the influence of cleaning and repairing was larger when looking at individual years than

8.4. Operational Profile Results

when comparing the maximum values of the datasets as a whole. When comparing the maximum endurance of Dataset 2 and Dataset 4, the endurance based on Dataset 4 was 12.10% higher than that based on Dataset 2. This shows that cleaning and repairing did influence the estimated endurance, although at fleet level across multiple years this influence was less significant than in per-year analyses.

Table 8.15: Maximum endurance comparison between Dataset 2 and Dataset 4 for HTVs (unsegmented)

Maximum Endurance [h]	2018	2019	2020	2021	2022	2023	2024	2025	Maximum
Dataset 2	81.25	212.25	813.92	949.67	406.50	727.50	1229.75	523.42	1229.75
Dataset 4	376.25	438.92	943.08	1069.08	761.58	1186.58	1399.08	779.25	1399.08
Percentage Difference	78.41	51.64	13.70	11.17	46.62	38.69	12.10	32.83	12.10

For Dataset 2, the maximum endurance would be approximately 51 days, whereas for Dataset 4 this would be 58 days. In Table 8.16, the vessel with the maximum endurance within the HTV fleet can be seen. It shows that for Dataset 2 the trip starts one week later and that less distance has been sailed.

Table 8.16: Comparison of the specific maximum endurance between Dataset 2 and Dataset 4 for HTVs (unsegmented)

HTV	Vessel	Start Time	End Time	Mean SOG [kn]	Distance [nm]
Dataset 2	Biglift Barentsz	2024-06-18 19:00:00	2024-08-09 00:40:00	8.5	10,530.98
Dataset 4	Biglift Barentsz	2024-06-11 17:40:00	2024-08-09 00:40:00	8.5	11,928.19

To determine which dataset was more accurate when analyzing the maximum endurance, the trip was checked in Plotter [4]. Figure 8.4 shows that the trip starts at 06/11/2024 17:36 and ends at 08/09/2024 00:40. Therefore, Dataset 4 was verified to be more accurate when determining the maximum endurance using the unsegmented method.

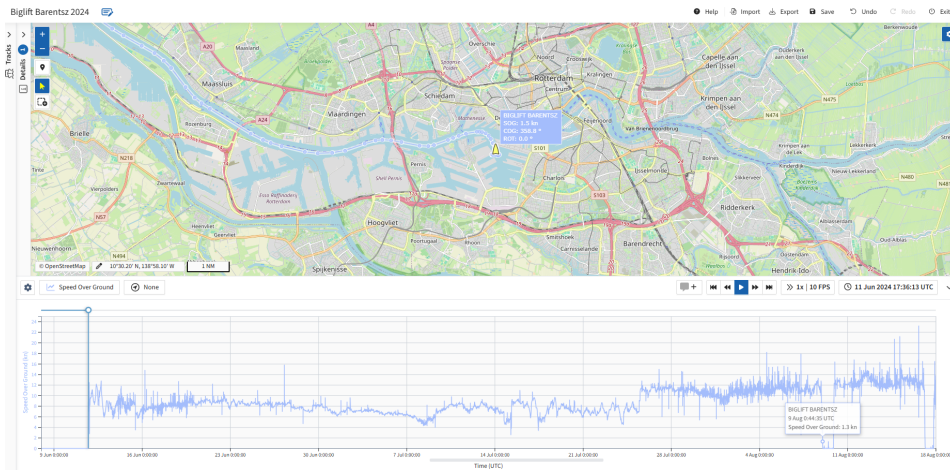


Figure 8.4: Maximum endurance for the HTVs determined with the unsegmented method

As described in subsection 7.4.1, the median SOG is determined for the ranges 1–20 kn, 2–20 kn, 3–20 kn, and 4–20 kn. For both datasets, the influence of changing the lower bound of the range was minimal; at most, the median SOG increased by 0.2 knots. For Dataset 4 this can be seen in Table 8.17.

Table 8.17: Median SOG based on different ranges for HTVs using Dataset 4

Median SOG [kn]	2018	2019	2020	2021	2022	2023	2024	2025	Total
1 - 20 kn	9.6	9.4	10.1	9.4	10.4	10.2	9.7	10.0	9.9
2 - 20 kn	9.6	9.5	10.2	9.5	10.4	10.3	9.7	10.0	10.0
3 - 20 kn	9.7	9.5	10.2	9.5	10.4	10.3	9.7	10.0	10.0
4 - 20 kn	9.8	9.5	10.2	9.6	10.5	10.4	9.7	10.0	10.0

The table above shows that the impact of the selected range for the median SOG calculation was minimal. Therefore, for the comparison between Dataset 2 and Dataset 4, only the median SOG in the range from 1 to 20 knots was considered, as in most literature 1 knot is used as the threshold between sailing and non-sailing conditions. Table 8.18 shows that the difference in median SOG between the datasets was very small. It also shows that there was no clear trend in which one dataset consistently resulted in a higher or lower median SOG. Overall, cleaning and repairing the dataset did not significantly influence the median SOG at fleet level.

Table 8.18: Median SOG using 1–20 kn for HTVs comparing Dataset 2 and Dataset 4

Median SOG [kn]	2018	2019	2020	2021	2022	2023	2024	2025	Total
Dataset 2	9.4	9.1	10.1	9.7	10.4	10.3	9.7	10.0	10.0
Dataset 4	9.6	9.4	10.1	9.4	10.4	10.2	9.7	10.0	9.9

To show how the vessels were operated, the SOG data points were divided into speed bins. For each speed bin, the number of data points, duration, distance, mean SOG, and median SOG were determined. In Table 8.19, the speed distribution for HTVs using Dataset 4 can be seen. It shows that most of the time the vessels are in a non-sailing condition, as SOG values below 0.5 kn indicate that the vessel is not sailing. It also shows that the vessels barely operate at very low sailing speeds, which is in line with the fact that almost all sailed distance is covered in the speed range from 6 to 20 knots. For the range around the median, it can be seen that 63.72% of all distance is sailed around the median SOG.

Table 8.19: HTV Dataset 4 speed distribution

	Data points [#]	Duration [h]	Duration [%]	Distance [nm]	Distance [%]	Mean SOG [kn]	Median SOG [kn]
0 - 0.5 kn	3,734,953	313,593.1	57.70	12,230.67	0.57	0.031	0
0.5 - 2 kn	89,070	7,995.5	1.47	6,058.876	0.28	0.925	0.8
2 - 4 kn	53,511	5,207.167	0.96	14,387.87	0.67	2.975	2.948
4 - 6 kn	97,739	9,520.833	1.75	47,697.22	2.23	5.134	5.2
6 - 20 kn	2,265,117	207,208.6	38.12	2,057,532	96.24	10.079	10.1
Total for HTV	6,240,390	543,525	100	2,137,907	100	-	-
Median 1 ± 2 kn	1,506,990	138,816.3	25.54	1,362,253	63.72	9.956	10

When looking at Table 8.20, the results using Dataset 2 can be seen. The same conclusions can be drawn from this table. The percentages in the different bins are almost the same. However, when looking at the absolute values, it can be seen that all values are slightly lower. Overall, when considering the percentages spent in the different speed bins at fleet level, cleaning and repairing the dataset was not of major relevance.

Table 8.20: HTV Dataset 2 speed distribution

	Data points [#]	Duration [h]	Duration [%]	Distance [nm]	Distance [%]	Mean SOG [kn]	Median SOG [kn]
0 - 0.5 kn	3,296,534	308,611	56.78	13,167	0.57	0.034	0
0.5 - 2 kn	78,320	7,947	1.46	5,095	0.22	0.894	0.8
2 - 4 kn	36,878	4,264	0.78	11,281	0.48	2.967	3
4 - 6 kn	79,244	9,683	1.78	53,199	2.28	5.087	5.1
6 - 20 kn	1,815,087	213,015	39.19	2,245,451	96.45	10.152	10.2
Total for HTV	5,306,063	543,520	100	2,328,193	100	-	-
Median 1 ± 2 kn	1,226,240	146,818	27.02	1,570,516	67.45	10.071	10.1

Semi-Submersibles

When comparing the operational profile results for semi-submersibles between Dataset 2 and Dataset 4, similar conclusions could be drawn as for the HTVs. For the average sailed distance per year, it can be seen in Table 8.21 that the influence of cleaning and repairing could again be significant when looking at the operational profile per year. However, when combining all years, this influence became less significant.

Table 8.21: Average sailed distance comparison between Dataset 2 and Dataset 4 for Semi-Submersibles

Distance [nm]	2018	2019	2020	2021	2022	2023	2024	2025	Average
Dataset 2	36505	47636	44844	27409	37582	40819	45932	42287	40042
Dataset 4	36385	47513	40755	20824	27463	38264	45925	42275	38429
Percentage Difference	0.33	0.26	10.03	31.62	0.32	6.68	0.01	0.03	4.20

As for the HTVs, the effect of cleaning and repairing could also be seen in the maximum endurance for semi-submersibles when analyzed in more detail (Table 8.22). When comparing the maximum endurance over all combined data, the difference between Dataset 2 and Dataset 4 was 4 days.

Table 8.22: Maximum endurance comparison between Dataset 2 and Dataset 4 for Semi-Submersibles

Maximum Endurance [h]	2018	2019	2020	2021	2022	2023	2024	2025	Maximum
Dataset 2	124.42	287.08	512.42	437.17	988.50	701.50	766.50	619.25	988.50
Dataset 4	335.42	836.17	858.50	573.00	988.50	777.25	1093.25	718.92	1093.25
Percentage Difference	62.91	65.67	40.31	23.71	0.00	9.75	29.89	13.86	9.58

Table 8.23 shows the two different maximum endurance cases. It can be seen that Dataset 2 had a higher mean SOG and a larger travelled distance despite the shorter duration.

Table 8.23: Comparison of the specific maximum endurance between Dataset 2 and Dataset 4 for Semi-Submersibles

SS	Vessel	Start Time	End Time	Mean SOG [kn]	Distance [nm]
Dataset 2	Fan Zhou 10	2022-01-06 21:05:00	2022-02-17 01:30:00	12.96	12,884.25
Dataset 4	GPO Grace	2024-11-14 19:45:00	2024-12-30 08:55:00	11.77	12,860.07

Figure 8.5 shows that the trip identified as the maximum endurance in Dataset 2 started earlier. It starts on 01/03/2022 15:47 and continues until 02/17/2022 01:39, which indicates an endurance of approximately 45 days. This again shows that Dataset 2 was not reliable for determining the maximum endurance.

8.4. Operational Profile Results

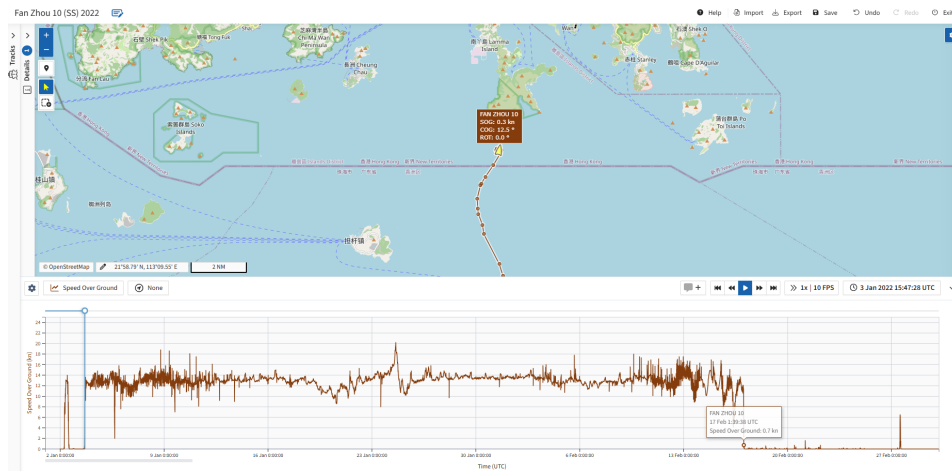


Figure 8.5: Visual Plotter overview of Fan Zhou 10 in 2022

When analysing the median SOG for semi-submersibles, it can again be seen that the selected range for the median was not very important. Table 8.24 shows that from 2020 onwards the median SOG was lower. This was due to the fact that *Blue Marlin* and *Hua Hai Long* have a SOG around 9 knots, and because the number of semi-submersible vessels in the dataset is limited, such lower speeds had a relatively strong influence.

Table 8.24: Median SOG based on different ranges for Semi-Submersibles using Dataset 4

Median SOG [kn]	2018	2019	2020	2021	2022	2023	2024	2025	Total
1 - 20 kn	11.30	12.10	9.20	8.21	9.50	9.70	9.80	9.00	9.70
2 - 20 kn	11.30	12.10	9.24	8.60	9.70	9.80	9.90	9.10	9.90
3 - 20 kn	11.37	12.10	9.30	8.70	9.90	9.90	10.00	9.10	9.90
4 - 20 kn	11.38	12.10	9.40	8.80	9.90	9.90	10.00	9.10	10.00

When comparing Dataset 2 and Dataset 4, it could again be seen that there was only a very limited difference between the results. Therefore, cleaning and repairing did not appear to be very influential at fleet level based on data from MSG [4].

Table 8.25: Median SOG using 1–20 kn for Semi-Submersibles comparing Dataset 2 and Dataset 4

Median SOG [kn]	2018	2019	2020	2021	2022	2023	2024	2025	Total
Dataset 2	11.10	12.20	9.60	8.20	9.70	9.80	9.80	9.00	9.70
Dataset 4	11.30	12.10	9.20	8.21	9.50	9.70	9.80	9.00	9.70

When dividing all semi-submersible data into speed bins, almost no difference in percentages could be observed between Dataset 2 and Dataset 4. Therefore, the results for Dataset 2 can be found in Appendix H and will not be separately discussed. Table 8.26 shows that the vessels were 52.17% of the time in a non-sailing condition and 40.50% of the time in regular sailing operations. This indicates that the vessels mostly operated in normal sailing conditions.

Table 8.26: Semi-Submersibles Dataset 4 speed distribution

	Data points [#]	Duration [h]	Duration [%]	Distance [nm]	Distance [%]	Mean SOG [kn]	Median SOG [kn]
0 - 0.5 kn	1,574,972	132,027	52.17	8,205	0.74	0.041	0
0.5 - 2 kn	99,532	8,463	3.34	8,161	0.73	0.984	0.9
2 - 4 kn	40,820	4,128	1.63	9,898	0.89	2.792	2.7
4 - 6 kn	69,631	5,958	2.35	30,628	2.75	5.093	5.1
6 - 20 kn	1,176,819	102,478	40.50	1,057,554	94.98	10.246	10.1
Total for SS	2,958,774	253,054	100	1,114,446	100	-	-
Median 1 +/- 2 kn	662,034	57,290	22.64	555,406	51.36	9.640	9.6

8.4.2. Comparison between HTVs and Semi-Submersibles operational profiles with the unsegmented method

Overall, at fleet level the difference between Dataset 2 and Dataset 4 was most visible in the maximum endurance. Using Plotter it was verified that Dataset 4 provided a more accurate estimation of the maximum endurance. Therefore, it was decided to compare the operational profiles of HTVs and Semi-Submersibles only based on the results of Dataset 4.

Table 8.27 shows that, overall, Semi-Submersibles covered slightly more distance during the year and spent 2.38% more time in steady sailing conditions. It can also be observed that the SOG of Semi-Submersibles was slightly lower than that of HTVs. However, this may also be due to the higher influence of individual vessels, as less Semi-Submersible data was used for this research. For HTVs, the influence of a single outlying vessel was smaller.

HTVs appeared to have a longer endurance and spent less time in very low-speed sailing operations. Semi-Submersibles may have spent more time in maneuvering operations than HTVs, although this is difficult to conclude based solely on AIS data. Overall, the results of the unsegmented method indicate that HTVs and Semi-Submersibles were operated in a broadly similar way.

Table 8.27: Comparison between operational profiles for HTVs and Semi-Submersibles using Dataset 4 and the unsegmented method

	HTV	SS	Difference
Average distance [nm]	35,216	38,429	8.36 %
Maximum Endurance [h]	1,399	1,093	21.87 %
Average SOG [kn]	9.9	9.7	0.2
Time spent [%] 0 - 0.5 kn	57.70	52.17	5.53
Time spent [%] 0.5 - 2 kn	1.47	3.34	1.87
Time spent [%] 2 - 4 kn	0.96	1.63	0.67
Time spent [%] 4 - 6 kn	1.75	2.35	0.60
Time spent [%] 6 - 20 kn	38.12	40.50	2.38
Time spent [%] Median 1 ± 2 kn	25.54	22.64	2.90

8.4.3. Segmented method results

For the segmented method, all data was divided into operational segments. The methodology used to create these segments was described in subsection 7.4.2. The segmentation was performed on

multiple levels in order to handle noise and jitter as effectively as possible.

Table 8.28 shows that, for the HTV fleet, Dataset 4 resulted in slightly fewer segments than Dataset 2, although this difference is minimal. This indicates that the segmentation method is able to handle datasets containing inconsistencies relatively well. The inconsistencies in Dataset 2 can still be observed in the total distance.

Table 8.28: Comparison for segmented method between Dataset 2 and Dataset 4 for HTVs

HTV	Dataset 2	Dataset 4
Segments [#]	5858	5826
Total Distance [nm]	2,330,783	2,137,907

For the semi-submersible fleet, this difference in segmentation is even smaller (Table 8.29). This shows that both datasets were segmented almost identically. Therefore, it is expected that the resulting operational profiles will not differ significantly when using Dataset 2 or Dataset 4 at fleet level.

Table 8.29: Comparison for segmented method between Dataset 2 and Dataset 4 for Semi-Submersibles

SS	Dataset 2	Dataset 4
Segments [#]	1911	1902
Total Distance [nm]	994,653	947,891

Although most of the data was segmented in a similar way, a different segmentation occurred for the endurance of the HTV fleet, resulting in a different maximum endurance. Table 8.30 shows that the difference in maximum endurance is relatively large. However, when comparing an endurance value that represents 98% of all trips, either by reducing the duration threshold or by considering only the shortest 98% of the segments, the difference decreases significantly.

The maximum endurance derived from Dataset 2 was visually checked using Plotter and SOG data. This inspection showed that a period of approximately one hour, during which the vessel was not sailing, was not included in the segmented result. The maximum endurance of Dataset 4 was also visually verified and was found to be correct.

Table 8.30: Endurance comparison for HTVs between Dataset 2 and Dataset 4

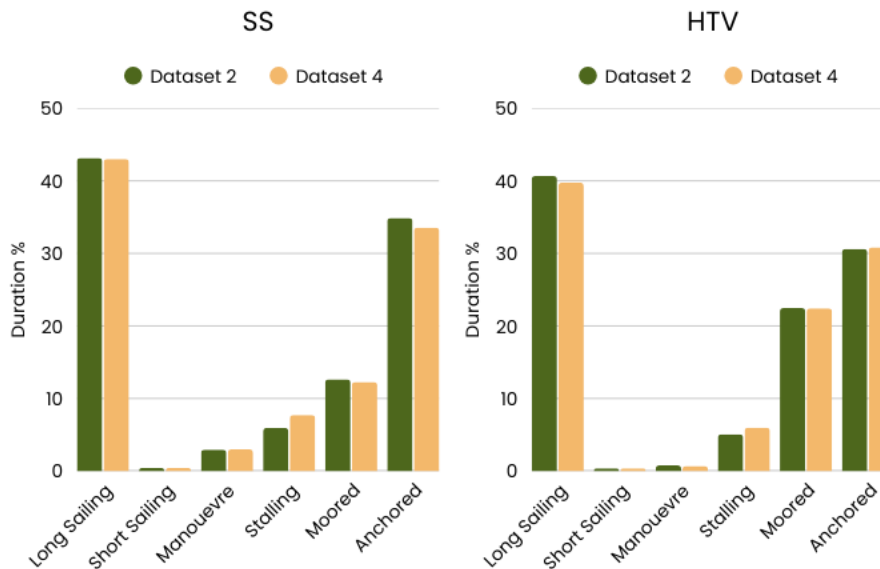
HTV	Dataset 2	Dataset 4	Difference [%]
Endurance max [h]	2,218	1,398	37.06
Endurance 98% (duration) [h]	666	699	4.72
Endurance 98% (segments) [h]	655	648	1.07

For the semi-submersibles, no difference in segmentation was observed when analysing the endurance (Table 8.31). The maximum endurance for the semi-submersibles was also visually verified using Plotter. This indicates that when using the segmented method, the influence of cleaning and repairing is only significant when analysing the absolute maximum endurance. When considering an endurance value that represents 98% of the trips, the difference becomes small.

Table 8.31: Endurance comparison for Semi-Submersibles between Dataset 2 and Dataset 4

SS	Dataset 2	Dataset 4	Difference [%]
Endurance max [h]	1,513	1,513	0.00
Endurance 98% (duration) [h]	907	907	0.00
Endurance 98% (segments) [h]	878	878	0.00

To summarize how the vessels are operated on average per year, Figure 8.6 shows the distribution per operational mode. The figure again confirms that, due to the very similar segmentation in Dataset 2 and Dataset 4, there is no significant difference between the datasets for both the HTV fleet and the semi-submersibles.

**Figure 8.6:** Distribution of operational modes for HTVs and SS

As expected due to the difference in total sailed distance, a small difference can be observed when analysing the mean distance sailed per long sailing trip. The most influential difference between Dataset 2 and Dataset 4 occurs in the travelled distance. The same trend can be seen in Table 8.32.

Table 8.32: Mean sailing distance comparison for HTVs between Dataset 2 and Dataset 4

HTV Mean Distance [nm]	Dataset 2	Dataset 4	Difference [%]
Long Sailing	1,583	1,435	9.35
Short Sailing	15	15	0.00
Manoeuvre	14	21	33.33
Long + Short	947	859	10.24
Long + Short + Manoeuvre	781	727	6.91

For the semi-submersibles, the same trend can be observed, with a decrease in distance for Dataset

4 compared to Dataset 2. Table 8.33 shows smaller differences between the datasets, indicating that fewer inconsistencies were present in these datasets.

Table 8.33: Mean sailing distance comparison for Semi-Submersibles between Dataset 2 and Dataset 4

SS Mean Distance [nm]	Dataset 2	Dataset 4	Difference [%]
Long Sailing	2,428	2,321	4.41
Short Sailing	20	20	0.00
Manoeuvre	62	69	10.14
Long + Short	1,294	1,230	4.95
Long + Short + Manoeuvre	1,018	983	3.44

The mean duration for all non-sailing segments was calculated and can be seen in Table 8.34. This again shows only a small influence of Dataset 4. This is expected because the total duration between Dataset 2 and Dataset 4 is similar, and the segmentation results are also comparable.

Table 8.34: Mean duration comparison for HTVs between Dataset 2 and Dataset 4

HTV Mean Duration [h]	Dataset 2	Dataset 4	Difference [%]
Moored	123	129	4.65
Anchored	117	121	3.31
Stalling	60	59	1.67
Manoeuvre	8	8	0.00
Anchored + Stalling	103	104	0.96

For the semi-submersibles, the mean duration of the non-sailing segments is shown in Table 8.35. This table again shows only limited differences between the datasets.

Table 8.35: Mean duration comparison for Semi-Submersibles between Dataset 2 and Dataset 4

SS Mean Duration [h]	Dataset 2	Dataset 4	Difference [%]
Moored	115	114	0.87
Anchored	138	134	2.90
Stalling	67	81	17.28
Manoeuvre	28	30	6.67
Anchored + Stalling	120	120	0.00

8.4.4. Comparison between HTVs and Semi-Submersibles operational profiles with the segmented method

To evaluate the operational profiles of the HTV fleet and the semi-submersibles, only the results from Dataset 4 are considered. Although the results for Dataset 2 are very similar, Dataset 4 is used as the preferred input because it contains fewer errors and is therefore considered more accurate.

When comparing the maximum endurance of HTVs and semi-submersibles, it can be observed that semi-submersibles are designed for slightly higher endurance. Table 8.36 shows that, in order to complete 98% of all trips without stopping, semi-submersibles require approximately 9 days more sailing endurance. This indicates that semi-submersibles generally perform longer maximum sailing segments.

Table 8.36: Comparison between HTVs and SS for endurance using the segmented method

Endurance 98%	HTV	SS	Difference
Duration [h]	648	878	9 days
Distance [nm]	5,018	8,555	41.34 %

Figure 8.7 also shows that, on average, semi-submersibles spend slightly more time sailing during a year. In addition, semi-submersibles spend less time moored in ports compared to HTVs. The figure also indicates that semi-submersibles operate more frequently in manoeuvring conditions. This may suggest that semi-submersibles more often transfer cargo offshore or between vessels rather than directly between ports.

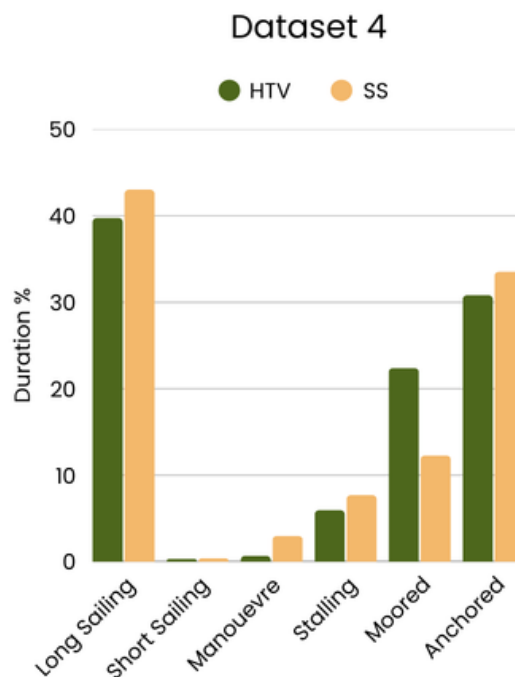


Figure 8.7: Distribution of operational modes for HTVs and Semi-Submersibles

Consequently, HTVs spend more time in ports than semi-submersibles. Table 8.37 shows that the mean duration a HTV spends moored in port is approximately 15 hours longer than for a semi-submersible. It was expected that this value might be higher for semi-submersibles, as they can also operate using the FLO/FLO loading method, which is typically more time-intensive. However, it may be that the semi-submersibles included in this research did not frequently perform FLO/FLO cargo operations. This remains speculative, as it cannot be verified based solely on AIS data.

Table 8.37: Comparison between HTVs and SS for moored duration using the segmented method

Moored	HTV	SS	Difference
Mean Duration	5 days 9 hours	4 days 18 hours	15 hours

The most noticeable difference between HTVs and semi-submersibles is the mean distance sailed per

long sailing segment. On average, HTVs sail 38.17% less distance per long sailing trip than semi-submersibles. However, the total time spent sailing during a year is relatively similar for both vessel types. This indicates that HTVs tend to perform more trips per year, but with shorter sailing distances, whereas semi-submersibles perform fewer trips but over longer distances.

Table 8.38: Comparison between HTVs and SS for mean sailing distance using the segmented method

Long Sailing	HTV	SS	Difference
Mean Distance [nm]	1,435	2,321	38.17 %

Overall, the largest difference between the operational profiles of HTVs and semi-submersibles can be observed in the distribution of non-sailing time. HTVs spend more time moored in ports, while semi-submersibles spend relatively more time at anchorage. This may have several explanations. For example, HTVs may make more use of just-in-time operations, while semi-submersibles may wait longer at anchorage areas. Another explanation could be that HTVs mainly transport cargo between ports, whereas semi-submersibles may more frequently operate offshore. It might also be that semi-submersibles need more preparation time in anchorage. However, these remain speculations and cannot be verified using AIS data alone.

Another clear difference is the distance sailed per long sailing segment. The average SOG during long sailing trips is similar for both vessel types, indicating that semi-submersibles do not sail faster. Instead, the longer distances per trip are most likely caused by longer voyage legs. This suggests that semi-submersibles generally perform fewer trips per year, but over longer distances, whereas HTVs perform more trips but over shorter distances.

If this trend is representative for the broader HTV and semi-submersible fleets, it may indicate that semi-submersibles benefit more from larger vessel sizes, allowing them to transport larger cargo volumes per trip.

8.4.5. Trip distance analyses

After all data is divided into segments, it becomes possible to analyse specific operational modes in more detail. To understand the operations of HTVs, both the speed and the trip distribution are important. For designers, this provides valuable insights into whether a vessel type performs many short trips or fewer but longer trips over the course of a year.

This analysis is performed for both Dataset 2 and Dataset 4, as well as for the HTV fleet and the semi-submersibles. First, all sailing segments are considered. Two types of distributions are constructed: the first shows the number of segments within a certain distance range, while the second represents the distribution of the duration spent within these distance ranges. In Figure 8.8, both the segment-based and duration-based distributions are shown.

It can be observed that short segments occur significantly more frequently than longer segments. However, when analysing the duration spent within each distance range, a more evenly distributed pattern is observed.

8.4. Operational Profile Results

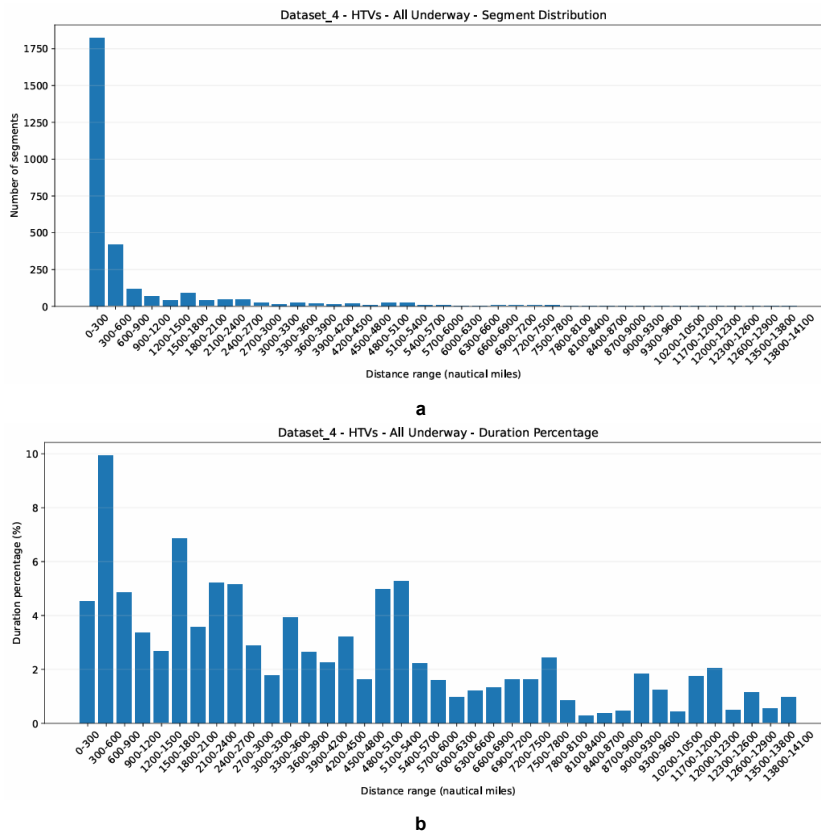


Figure 8.8: Distribution based on segments and duration for HTVs using Dataset 4

Overall, the results show that HTVs operate across a wide range of trip distances and spend a considerable amount of time sailing trips of varying lengths. Therefore, it cannot be concluded that HTVs primarily perform short trips, as longer trips, although less frequent, still account for a significant portion of the total sailing time.

When comparing the results from Figure 8.8b with those based on Dataset 2 (Figure 8.9), it can be observed that the shape of the distribution is almost identical. However, the distance range for Dataset 2 extends further, indicating that some segments include inconsistencies in the data, leading to unrealistically large travelled distances. These distances reach values up to 24,600–24,900 nm, which is not consistent with the known endurance of the vessels included in this study.

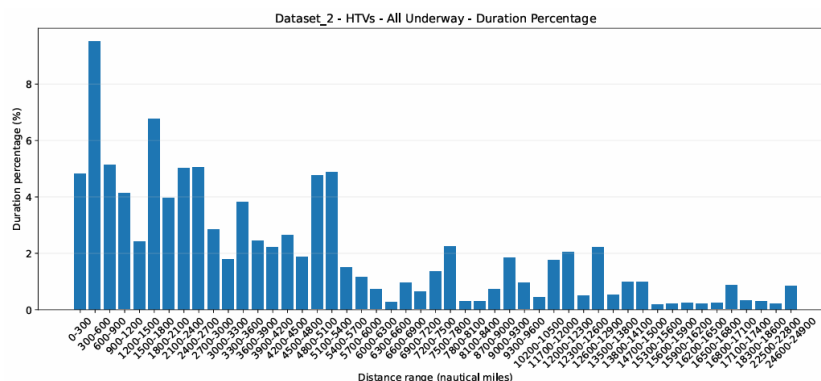


Figure 8.9: Duration spent within distance bins per trip based on Dataset 2 for HTVs

For the semi-submersibles, a similar difference between Dataset 4 and Dataset 2 can be observed:

the shape of the distribution remains comparable, but the range extends to higher maximum values in Dataset 2. Figure 8.10 shows that semi-submersibles exhibit a more evenly distributed pattern across trip distances.

Overall, the same conclusion as for the HTVs can be drawn: these vessel types spend time across a wide range of trip distances, with comparable durations spent in both shorter and longer trips. To draw more robust conclusions, it may be beneficial to include additional vessels in this analysis to determine whether the distribution becomes more peaked or remains relatively flat.

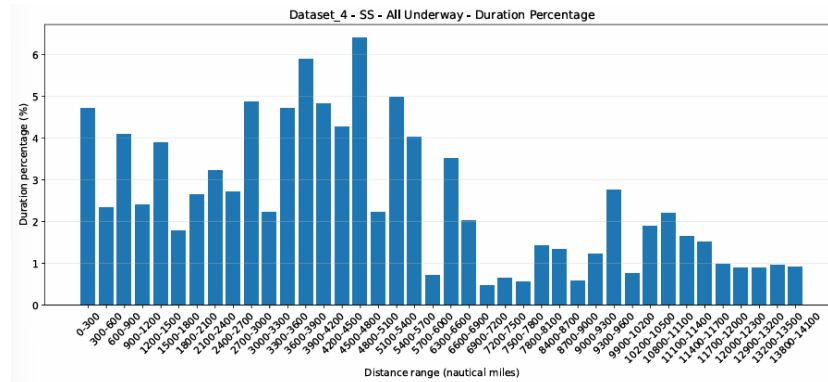


Figure 8.10: Duration spent within distance bins per trip based on Dataset 4 for Semi-Submersibles

8.4.6. Comparison between methods for data-driven operational profiles

When comparing the different datasets at operational profile level, it becomes clear that the influence of cleaning and repairing is minimal when considering a fleet-level operational profile. However, when analysing maximum values, the influence of cleaning appears more significant, as these are based on individual vessels and specific years. The difference between Dataset 2 and Dataset 4 is therefore more pronounced when considering maximum values compared to mean values.

Thus, when comparing the impact of Dataset 2 and Dataset 4 at fleet level, the influence is very limited. In contrast, when analysing the datasets at a more detailed level, such as per vessel and per year, more significant differences can be observed. An example of this can be found for the vessel *Biglift Baffin*, for which specific errors were identified (see Appendix X).

When comparing the unsegmented method with the segmented method, the main difference between the two methods lies in the level of detail that can be obtained in the operational profile. For the unsegmented method, it becomes clear how the vessel has been operated based on speed bins. This provides an overall picture of how much of the time the vessel has been sailing and not sailing. When comparing only the percentages of sailing and non-sailing, both methods give roughly the same result: approximately 60% of the time the vessel is not sailing and 40% of the time it is sailing. This can also be observed in Figure 8.11.

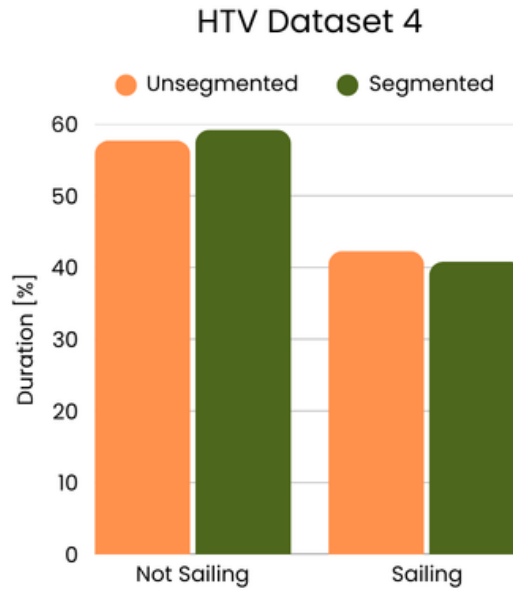


Figure 8.11: Unsegmented compared to segmented on sailing and non-sailing distribution

For the unsegmented method, the total sailed distance is slightly higher. To determine the total sailed distance for the unsegmented method, the distance travelled within the bin 0–0.5 knots has been excluded. The result can be seen in Table 8.39. With the unsegmented method, drift and jitter data points with a SOG > 0.5 knots are still included, which increases the total sailed distance. Although this difference is not very large, an advantage of the segmented method is that the total sailed distance can be determined using only the sailing segments. In this way, the travelled distance without the influence of drift and jitter can be identified, provided that the segmentation is performed correctly.

Table 8.39: Comparison of sailed distance between the unsegmented and segmented methods

HTV Dataset 4	Unsegmented	Segmented
Total sailed distance [nm]	2,125,676	2,098,563

Another limitation of the unsegmented method is that only the maximum endurance can be determined. Lower endurance values can also be estimated, but in that case it remains unknown how many trips can still be completed with that endurance. Therefore, reducing the maximum endurance becomes more of an assumption. Table 8.40 shows that the maximum endurance determined with both methods is nearly the same. However, as mentioned above, it may not be desirable to design a vessel for the absolute maximum endurance for several reasons. With the segmented method, a reduced endurance can easily be identified, together with its effect on the percentage of historical trips that can still be completed.

Table 8.40: Comparison of endurance between the unsegmented and segmented methods

HTV Dataset 4	Unsegmented	Segmented
Max endurance [h]	1399	1398
98% Endurance [h]	-	878

Overall, the unsegmented and segmented methods provide similar results at fleet level, without significant differences. The main distinction lies in the level of detail that can be obtained from each method. To determine in which speed bins the vessel has operated most and how much of the time it has been

sailing or not sailing, the unsegmented method provides sufficient insight. However, if more detailed operational information is required, such as per-trip characteristics, the data needs to be divided into segments.

Table 8.41 shows the difference in the possibilities offered by each method. The overall conclusion is that both methods provide similar general results, but the segmented method offers a higher level of detail.

Table 8.41: Comparison between the possibilities of the unsegmented and segmented methods

Unsegmented	Segmented
Max endurance	Endurance based on the number of segments
Total sailed distance including distance due to drift and jitter	Total sailed distance including only sailing segments
Percentage spent in speed bins	Percentage spent in operational states
Cruising speed	Speed per operation
	Average distance per trip
	Average duration per operation
	Speed distribution per trip

8.5. Verification and Validation of the Segmentation Method

To determine whether the methods used in this research are implemented correctly, several verification and validation checks are performed on key components of the methodology.

8.5.1. Verification Results

The haversine equation is used throughout this research to determine the distance between consecutive AIS data points. To verify the correct implementation of this function, several sanity checks are performed:

- No calculated distances have negative values.
- If the coordinates of two consecutive data points are identical, the calculated distance is zero.
- If the coordinates of two consecutive data points are reversed, the calculated distance remains the same.

Another important component of this research is the segmentation method used to divide the AIS data into operational segments. This method is verified by comparing the segmentation results with the visual AIS data in Plotter. In Plotter, the speed distribution can be visualised together with the vessel location over time. The background map is set to OpenStreetMap to analyse whether the vessel is moored, anchored, or sailing.

This comparison is performed for several datasets. The results are shown in Table 8.42. The table shows that the segmentation method may still produce small errors, for example by ignoring a short sailing period of approximately 2 knots lasting about 30 minutes. However, overall it can be concluded that the segmentation method identifies the operational segments accurately.

Table 8.42: Verification of the segmentation method

	Wrong Segments	Wrong Labels
Audax 2023	1	11
Sun Shine 2024	0	0
Sun Shine 2023	1	1
Target 2018	0	2
Target 2019	1	0
Target 2020	2	0
Dongbang Giant No.6 2024	0	0

8.5.2. Validation Results

To validate the haversine distance calculation used in this research, the calculated distances are compared with the results obtained from an online distance calculation tool. This comparison is used to confirm that the implemented distance calculation in Python produces accurate results.

Table 8.43 shows an example of an interpolation error in Dataset 2. For each data point, the calculated distance using the method described in this research is shown, together with the distance calculated using an online tool [33]. The comparison shows that the difference between both methods is negligible, confirming the correctness of the implemented haversine distance calculation.

Table 8.43: Snapshot of dataset 2 (Interpolated) from Biglift Baffin 2019 containing interpolation mistake

Date and UTC Time	Latitude [degree]	Longitude [degree]	SOG [kn]	Distance [nm]	Duration [s]	SOG calc [kn]	Distance calc [nm]
4/27/2019 6:45	21.9273	-179.957308	8.3	0.694	300	8.328	0.694
4/27/2019 6:50	21.9266	-179.969585	8.3	0.685	300	8.22	0.685
4/27/2019 6:55	21.925995	-175.615337	8.3	242.512	300	2910.144	242.496
4/27/2019 7:00	21.92528	-30.074947	8.3	7490.614	300	89887.368	7489.20
4/27/2019 7:05	21.924567	115.465443	8.2	7490.680	300	89888.16	7489.20
4/27/2019 8:55	21.909247	179.726913	8.1	3550.218	6600	1936.48	3550.216
4/27/2019 9:00	21.908415	179.714782	8.2	0.677	300	8.124	0.677

8.6. Conclusion Results

To conclude, this research used 11 different HTV vessels and 6 Semi-Submersibles, resulting in 91 datasets (one dataset per year per vessel) to derive data-driven operational profiles. These datasets were obtained through MadeSmart [4] and showed very high coverage, where 97% of all data points had durations smaller than 30 minutes.

After obtaining Dataset 2, it was cleaned and repaired by implementing the methods described in this research. The cleaning procedure showed that interpolation errors do not occur frequently; however, when they do occur, they can have a significant impact on the total sailed distance. The repairing procedure reduced the missing data from 17.27% to 3.84%. Overall, the cleaning and repairing showed that if errors exist in the dataset, their influence can be significant when analysed per vessel, but smaller when analysed at fleet level due to their low occurrence.

When using both datasets to obtain operational profiles without segmentation, the largest difference can be observed in the maximum endurance, where Dataset 2 is less reliable than Dataset 4. This also showed a deviation in the sailed distance, as expected. However, when analysed at fleet level, the significance is reduced due to the low occurrence of errors across the fleet as a whole.

Next, the operational profiles obtained with the segmentation method were compared. Due to the multi-level segmentation method, the difference in the number of segments based on Dataset 2 and Dataset 4 was minimal. A significant difference was still observed when determining the maximum endurance of the whole fleet; however, when considering the endurance required to cover 98% of all segments, the difference between Dataset 2 and Dataset 4 was much smaller. Overall, the influence of cleaning and repairing the dataset on operational profiles at fleet level is limited. However, at a more detailed level, such as per vessel, the importance of cleaning and repairing increases.

The segmentation method was also validated using visual trajectories in Plotter [4]. This validation was performed on 7 randomly selected datasets and showed high accuracy for the segmentation method.

9

Blended Design as early stage design tool

This research provides a framework for data-derived operational profiles and demonstrates its application to the HTV fleet. In this chapter the impact of data-derived operational profiles in the early-stage vessel design will be discussed. The data-derived operational profiles are compared with the assumed operational profile for an HTV by UDSBV in section 9.1. In section 9.2, the design results from Blended Design using different operational profiles as input are discussed to assess the impact of data-derived operational profiles. This chapter is briefly concluded in section 9.3.

9.1. Operational profiles in Blended Design

Blended Design is briefly introduced in chapter 5. To align Blended Design with the data-driven operational profiles obtained from AIS data, several adjustments to the standard input parameters are required.

The operational profiles contain a large amount of information. Therefore, specific parameters must be selected to serve as input for Blended Design. The parameters derived from the operational profiles are listed below:

1. Distance per single trip (nm)
2. DP days
3. Endurance (days)
4. Port time (hours)
5. Waiting time (hours)
6. Operational days per year

The mean distance of the long sailing segments is used to represent the distance per single trip in nautical miles.

The number of DP days is set to zero, as the use of dynamic positioning (DP) is uncommon for most vessels in the dataset. In addition, distinguishing between DP operation and anchoring solely based on AIS data proved difficult and was not reliably achieved in this research. However, the percentage of time spent in manoeuvring operations is known. Based on this percentage, a rough estimation of DP days could be made to investigate the potential influence on the results. The accuracy of such an estimation remains uncertain and has therefore not been used.

For the endurance parameter, multiple values are evaluated to assess their influence on the resulting ship design. The endurance durations corresponding to 100% and 98% segment coverage are considered. In addition, the endurance values derived from the duration-based analysis (100% and approximately 98%) are also included.

The port time is determined as the mean duration of all moored segments, while the waiting time is defined as the mean duration of anchored and stalling segments combined. Stalling segments are included because they also represent waiting behaviour.

The number of operational days per year is taken as a full year of 365 days. The analyses are based on yearly averages; therefore, non-operational days are already implicitly included in the waiting times. Using AIS data alone, it is not possible to determine whether a vessel is not sailing due to a lack of contract or temporary inactivity. Therefore, the operational days are set to 365 days.

In the original Blended Design method, port time is estimated using Equation 9.1, where the port time depends on the number of cargo items handled.

$$t_{port} = 5 + 4 \cdot N_{cargo} \tag{9.1}$$

In this research, this relation is replaced by the mean port time derived from the AIS-based segmentation. The same approach is applied to the waiting time: Equation 9.2 in Blended Design is replaced by the mean anchored and stalling time derived from the AIS-based segmentation. This original equation was based on waiting time as a function of whether the vessel could operate in DP or not.

$$\frac{t_{DP}}{DP_{capability}/100} - t_{DP} \tag{9.2}$$

When the operational profiles are adapted to fit the above described format these can be compared based on their Blended Design input this comparison can be seen in Table 9.1. The first two are the assumed operational profiles made by UDSBV.

It was determined that OP 2 D4 HTV represents the most suitable data-driven operational profile. For this profile the cleaned and repaired Dataset 4 is used together with the segmented method. The endurance is not based on the maximum observed value, but on the endurance required to sail 98% of all segments.

OP 1 D4 HTV is also based on the segmented operational profile derived from Dataset 4, but in this case the maximum endurance is used. This profile was included to determine the influence of designing the vessel to be able to sail all segments instead of 98% of the segments.

For OP 3 D2 HTV the segmented operational profile based on Dataset 2 is used. In this case the endurance is also based on 98% of the trips. Finally, OP 4 D4 SS uses the same approach as OP 2 D4 HTV but is based on the operational profile derived for the semi-submersible vessels.

Table 9.1: Input Blended Design

	UDSBV 1	UDSBV 2	OP 1 D4 HTV	OP 2 D4 HTV	OP 3 D2 HTV	OP 4 D4 SS
Distance single trip [nm]	300	3300	1435	1435	1583	2321
DP days	5	0	0	0	0	0
Endurance [days]	40	40	58	27	27	36
Port time [h]	$5 + 4 \cdot N_{cargo}$	$5 + 4 \cdot N_{cargo}$	129	129	123	114
Wait time [h]	*	*	104	104	103	120
Operational days	300	300	365	365	365	365

* $\frac{t_{DP}}{DP_{capability}/100} - t_{DP}$

It can be observed that the distance per single trip estimated by UDSBV lies roughly in the same range as the data-derived input values. The endurance also appears to be a reasonable estimation. The largest difference can be observed in the no-sailing time.

UDSBV expected the HTVs to spend less time in port and less time waiting, which would allow the vessels to perform more trips per year. To reach the same port time as derived from the AIS-based operational profile, approximately 31 monopiles would need to be transported per trip. This is not realistic, as the analysed vessels typically transport around 10 monopiles per trip.

The waiting time was also adjusted compared to the previous design iteration. In the earlier estimation this waiting time only included delays due to weather conditions that prevented DP operations. Waiting time in anchorage or during stalling could be partly represented by the standard 5 hours included in the port time estimate. However, even with this adjustment there remains a large difference compared to the waiting times derived from AIS data.

Overall, the initial estimations made by UDSBV appear to be somewhat optimistic regarding the amount of time the vessel is able to spend in sailing operations. But UDSBV assumed that the vessel are not in operation 65 days per year. Where as the data-derived operational profiles are analyzed throughout a whole year. Table 9.1 shows that there are some deviations between the assumed operational profile and the data-derived operational profile to see the impact on the vessel design there needs to be looked at the results from Blended Design.

9.2. Ship design from Blended Design

The Blended Design tool results in the same optimal hull design for all operational profile inputs. The only parameter that differs between the scenarios is the optimal design speed. Table 9.2 shows the vessel design that performs best over the vessel’s lifetime according to the Blended Design method.

Table 9.2: Ship design output from Blended Design

	UDSBV 1	UDSBV 2	OP 1 D4 HTV	OP 2 D4 HTV	OP 3 D2 HTV	OP 4 D4 SS
Length [m]	184.1	184.1	184.1	184.1	184.1	184.1
Breadth [m]	50.4	50.4	50.4	50.4	50.4	50.4
Depth [m]	11.9	11.9	11.9	11.9	11.9	11.9
Design Speed [kn]	11–12	15	12–14	12–14	12–14	12–15
Block Coefficient [-]	0.64	0.64	0.64	0.64	0.64	0.64

When analysing the design results, it can be observed that an increase in trip distance results in a higher optimal design speed. The hull form remains unchanged because the vessel operates in long sailing conditions for a significant portion of its lifetime. In order to sail as efficiently as possible and maximize the number of trips per year, the block coefficient is optimized at the lower boundary of the design space.

OP 2 D4 HTV and OP 3 D2 HTV were compared to analyse the influence of the repaired dataset (Dataset 4) versus the interpolated dataset (Dataset 2). Figure 9.1 shows the comparison between the operational profiles for the design speed. The bold section of the curves represents values where the Return on Investment (ROI) is nearly identical.

The results show that the optimal design speed is almost the same for both operational profiles. However, the repaired dataset results in a slightly higher ROI. This difference can be linked to the input values shown in Table 9.1, where the interpolated dataset results in slightly longer trips which means the vessel can sail less trips per year leading to a smaller ROI.

It should be noted that the absolute value of the ROI is not the primary focus of this analysis, as it is

based on estimated income values provided by UDSBV. Instead, the relative differences between the ROIs are used to assess the influence of the different operational profiles. The other Blended Design outputs show similar minimal differences between the repaired and interpolated datasets; therefore, the remaining figures are provided in Appendix J.

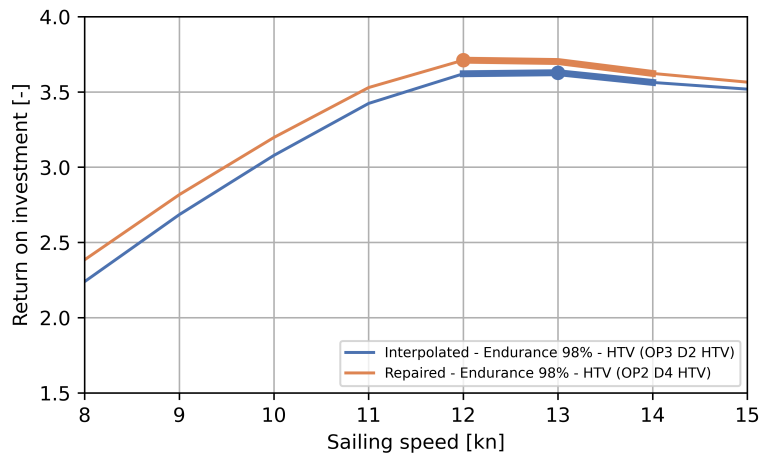


Figure 9.1: OP 2 D4 HTV compared with OP 3 D2 HTV for optimal design speed

The influence of reducing the endurance from 58 days to 27 days between OP 1 D4 HTV and OP 2 D4 HTV is also minimal and not clearly visible in the Blended Design output. These results are presented in Appendix J.

A larger difference can be observed when comparing OP 2 D4 HTV with OP 4 D4 SS. Semi-submersibles typically sail longer distances per trip compared to HTVs. When both vessel types sail at the same speed, the semi-submersible will complete fewer trips per year, which results in a lower lifetime ROI. This effect can be observed in Figure 9.2, where the optimal ROI for the semi-submersible is lower than for the HTV.

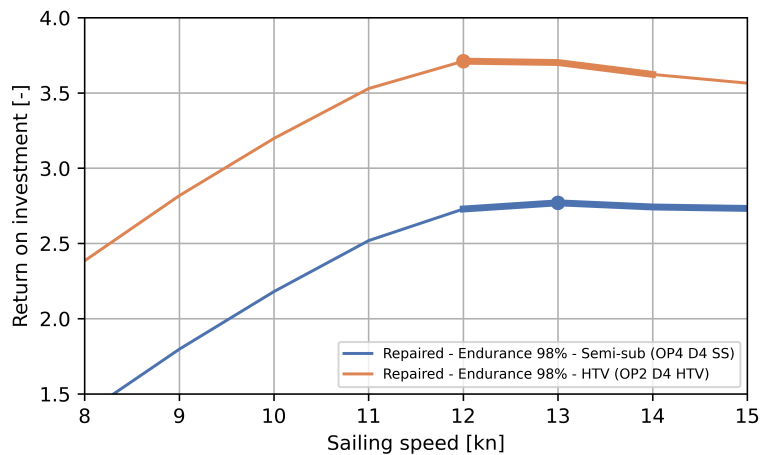


Figure 9.2: OP 2 D4 HTV compared with OP 4 D4 SS for optimal design speed

The next figures compare the data-driven operational profile (OP 2 D4 HTV) with the estimated operational profiles provided by UDSBV. The other operational profiles follow a similar trend to OP 2 D4 HTV but with slightly lower ROI values.

When comparing the data-driven operational profile with the estimated profiles, the main difference can be attributed to the trip distance. In Figure 9.3 it shows that when the vessels are expected to sail

longer distance per trip the ROI will be slightly lower. The blue line is the assumed operational profile with the short distance whereas the orange line represents the longer mission.

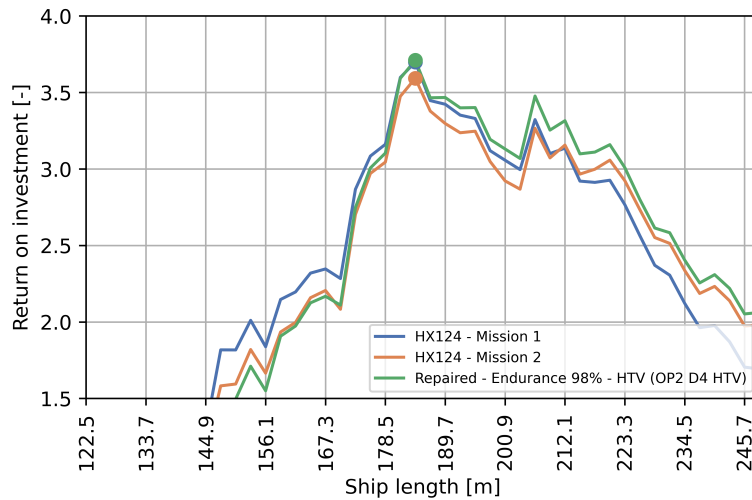


Figure 9.3: OP 2 D4 HTV compared with UDSBV 1 and 2 for vessel length

Figure 9.4 shows that the optimal breadth would likely be larger for all operational profiles. However, the breadth is constrained to 50.4 meters to allow the vessel to pass through the Panama Canal. Without this constraint, the optimal design would likely have a larger breadth.

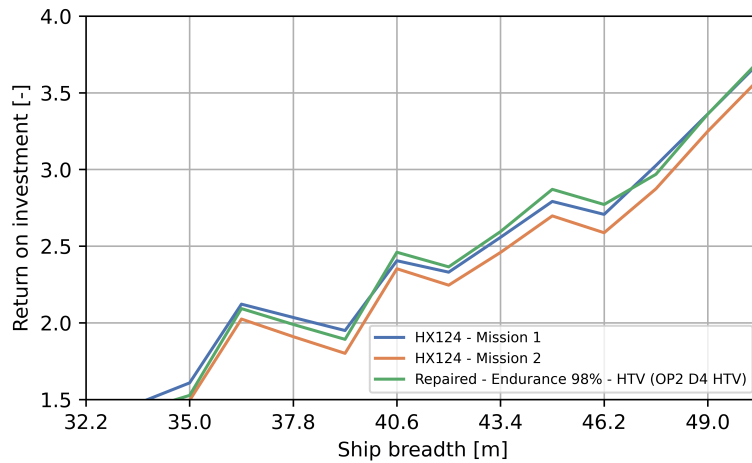


Figure 9.4: OP 2 D4 HTV compared with UDSBV 1 and 2 for vessel breadth

The largest difference between the operational profiles can be observed in the optimal design speed, as shown in Figure 9.5. Operational profiles with shorter trip distances result in a lower optimal design speed. Therefore, trip distance has the strongest influence on the optimal design speed, while port time and waiting time mainly influence the ROI.

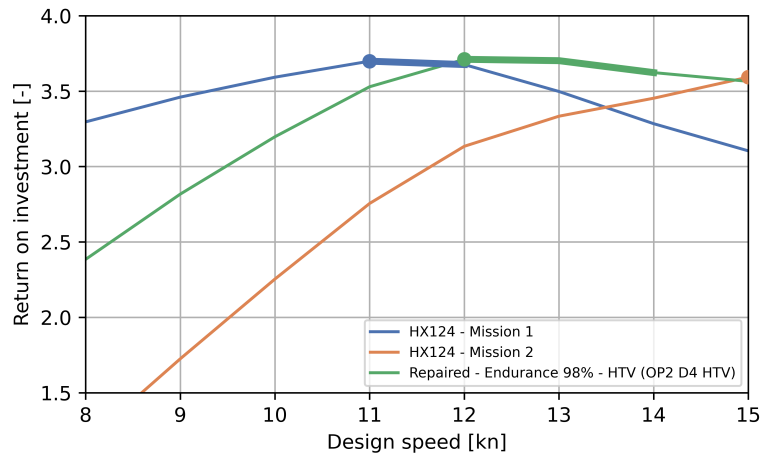


Figure 9.5: OP 2 D4 HTV compared with UDSBV 1 and 2 for design speed

Overall it can be concluded that the impact of data-derived operational profiles in this case study is minimal. UDSBV made different assumptions in the operational profiles than the historical AIS data showed but the impact on the design and the ROI is minimal.

9.3. Conclusion Blended Design as early stage design tool

In conclusion it can be seen that the assumptions made by UDSBV do differ from the data derived operational profiles. The assumptions mostly differ in the port and waiting time where the assumptions are significantly lower compared to what the historical AIS data shows. Another difference is the trip distance UDSBV assumed two types of operations one with short trips and the other with long trips whereas the AIS data showed that HTVs operate in both short and long trips. Lastly UDSBV assumes that 65 per year the vessel is not in operation sailing beneficial trips whereas for data-derived operational profile this is taken into account in the time spent in port and waiting.

When looking at the ROI and design impact it can be seen that the ROI for the assumed operational profiles and the data-derived operational profiles is almost similar this shows that although the not sailing time was underestimated by UDSBV the 65 not operational days compensated this.

Overall, the trip distance has the largest influence on the optimal design speed, while waiting time and port time primarily influence the ROI. For HTVs, the vessel design is largely constrained by the cargo requirements. In the Blended Design analysis the vessels are designed to transport monopiles; therefore, the length is limited by the expected future monopile length. The breadth is constrained by the requirement to pass through the Panama Canal. If this constraint were removed, vessels operating on longer routes would likely benefit from a larger breadth in order to transport more monopiles per trip.

The influence of the different datasets used for the data-driven operational profiles is minimal in the Blended Design results.

10

Additional ship insight through AIS data

The impact on the ship design of data-derived operational profiles seems to be limited for the HTV fleet analysed in this research. But AIS data can also provide additional ship insights useful for early-stage vessel design. In this research, several AIS-based insights were obtained to support key research decisions. In section 10.1, the vessels in this research are analysed to determine whether they visited known monopile fabrication sites. section 10.2 discusses whether yearly trends can be identified that may need to be incorporated into the operational profiles. It is also insightful to assess whether the predefined design criteria are aligned with actual operations; this is discussed in section 10.3. Lastly, this chapter is briefly concluded in section 10.4.

10.1. Monopile fabrication site visits by the vessels in this research

In this section, the activity of vessels around known monopile fabrication sites is analysed to estimate how frequently the vessels operate within this specific market.

To determine the activity at known monopile fabrication sites, a list of fabrication locations was compiled using publicly available information obtained through Google. For each of these monopile fabrication sites, a polygon was drawn in the Cartographer tool provided by MSG [4]. Cartographer uses OpenStreetMap as a background layer, and for some sites the monopiles themselves are even visible in the satellite imagery. The polygons drawn in Cartographer were exported to an Excel file containing the coordinates of each corner of the polygons. The full list of monopile fabrication sites and their corresponding polygons can be found in Appendix K.

After defining the locations of the monopile fabrication sites, a Python script was developed to analyse all no-sailing segments in the AIS dataset and determine whether such a segment occurs within the boundaries of one of the fabrication site polygons. When at least 80% of the AIS data points within a segment fall inside the polygon boundaries, the segment is classified as a visit to a monopile fabrication site.

This analysis showed that 13 out of the 17 vessels visited at least one of the identified fabrication sites, of which 8 were HTVs and 5 were semi-submersibles. Table 10.1 shows consecutive visits to known monopile fabrication sites made by HTVs.

Table 10.1: AIS-derived monopile transport performed by HTVs

	Bigroll Bering	Dongbang Giant No.6	Dongbang Giant No.7	Zhong Yu 188	Fan Zhou 8
Start Date	03/13/2020	01/02/2023	11/29/2022	03/22/2025	03/05/2025
End Date	06/29/2020	06/28/2023	02/07/2023	04/30/2025	10/21/2025
Fabrication Site	Steelwind Nordenham	EEW Rostock	SDMS	Century Wind Power	Dajin Penglai
Country	Germany	Germany	Germany		China
Number of Visits	2	11	5	3	4

Table 10.2 shows consecutive visits to monopile fabrication sites made by the semi-submersible fleet. Both vessel types operate in the monopile transportation market, although it is clear that this is not their only market segment.

Table 10.2: AIS-derived monopile transport performed by Semi-Submersibles

	Target (SS)	Sun Shine (SS)	GPO Grace (SS)	Blue Marlin (SS)
Start Date	05/24/2023	01/22/2024	03/26/2023	03/26/2024
End Date	02/25/2024	08/22/2025	02/24/2024	08/27/2024
Fabrication Site	Steelwind Nordenham	EEW Rostock	EEW Rostock	Dajin Penglai
Country	Germany	Germany	Germany	China
Number of Visits	22	12	7	4

For the operational profile, an estimate was made for the average time vessels spend in port. When analysing the time spent at monopile fabrication sites, a large variation can be observed. The duration varies from a few hours to multiple days. This indicates that loading and unloading operations for these vessels can vary significantly. Even within the same port and for the same vessel, the duration can differ by several days. This variation can be seen in Figure 10.1.

Start	End	Duration (dd hh:mm:ss)	Port	Mean SOG	Pts in port	% in port	Majority label
2023-03-26 18:45:00	2023-03-27 14:20:00	00 19:35:00	EEWRostock	0.04	232	98.3	Moored
2023-03-30 08:20:00	2023-04-04 17:25:00	05 09:05:00	EEWRostock	0.01	1550	100.0	Moored
2023-06-10 07:55:00	2023-06-13 17:30:00	03 09:35:00	EEWRostock	0.02	977	99.7	Moored
2023-10-16 01:10:00	2023-10-18 04:30:00	02 03:20:00	EEWRostock	0.02	613	99.4	Moored
2023-11-22 16:10:00	2023-11-25 14:25:00	02 22:15:00	EEWRostock	0.10	844	100.0	Moored
2024-01-26 11:20:00	2024-01-28 18:20:00	02 07:00:00	EEWRostock	0.02	660	99.8	Moored
2024-02-24 10:20:00	2024-02-29 07:20:00	04 21:00:00	EEWRostock	0.01	1403	99.9	Moored

Figure 10.1: Duration spent in EEW Rostock port by GPO Grace

The vessels have also visited known monopile fabrication sites without performing consecutive visits. Overall, it appears difficult and unrealistic to estimate a fixed loading or unloading time per monopile, as the duration varies significantly in practice.

Therefore, it appears more realistic to represent the total operational time spent in mooring operations by analysing the full dataset and using the mean duration. Looking only at visits to monopile fabrication sites would not be representative, as the vessels analysed in this research do not operate exclusively in the monopile transportation market.

One advantage of the Blended Design approach is that the vessel design is based on the full operational lifetime of the vessel. By using the mean operational behaviour derived from the AIS data, a more realistic estimation of operational conditions can be obtained.

10.2. Trends over the years for the HTV fleet

When using historical AIS data to design a new vessel, it is important to analyse trends in the historical data in order to estimate whether certain operational parameters may change in the future. For example, if a significant increase in the distance sailed per trip is observed over the years, this could indicate

that offshore wind farms are being constructed further offshore. Such a development could influence the endurance requirements of future vessel designs.

To analyse possible trends, the results from Dataset 4 for the HTV fleet have been used. First, the vessel speeds over the years were analysed. There could potentially be a trend in vessel speed due to the increasing adoption of slow steaming, which is often applied as a short-term measure to reduce greenhouse gas emissions [38]. Figure 10.2 shows the median SOG for four different speed ranges for all HTVs combined. No clear trend can be observed in the SOG over the analysed period. Therefore, it cannot be concluded whether vessels are expected to sail faster or slower in the future. When analysing the data per vessel, no consistent trend in the SOG could be identified either.

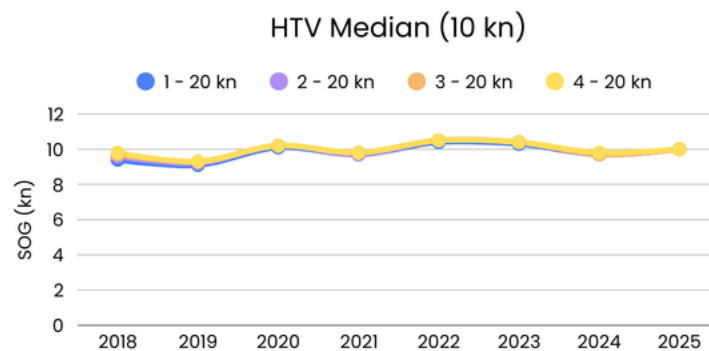


Figure 10.2: Median SOG for HTVs per year

Another analysed trend is the average sailed distance per year. Figure 10.3 shows the average distance sailed per vessel for each year. No overall trend can be identified from this figure. Some vessels show a decrease in activity during the COVID-19 period; however, this effect is not consistent across the entire fleet.

The COVID-19 years were not excluded from the analysis. The Blended Design methodology considers the operational behavior of vessels over their entire lifetime. During a vessel’s operational lifetime, external influences such as economic fluctuations or global disruptions are expected to occur. By using long-term averages, the realistic influence of such external factors is inherently incorporated into the analysis.

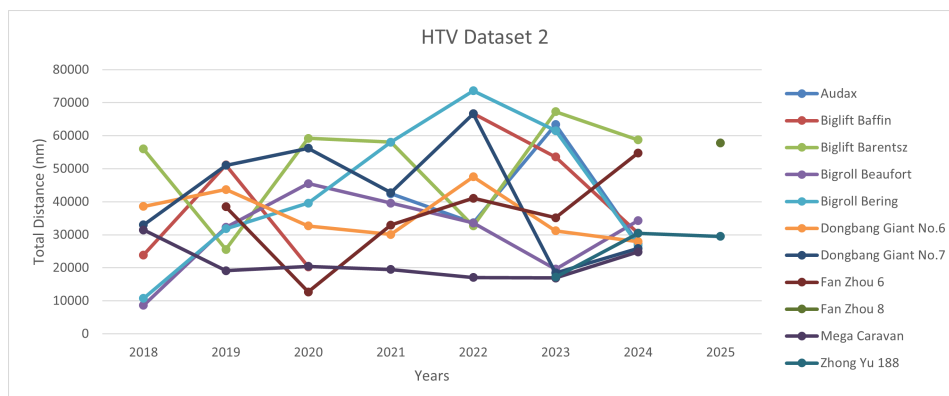


Figure 10.3: Average distance sailed per year for the HTV fleet

Overall, no clear trends were identified within the analysed HTV fleet. The vessels operate very diversely: in some years a vessel performs many short trips, while in other years it operates globally with longer voyages. The yearly variation per vessel in the time spent sailing versus non-sailing operations can reach differences of up to approximately 30%.

Due to this operational variability, using average values to derive the operational profiles appears to provide the most representative description of the operational behaviour of the HTV fleet.

10.3. Vessel characteristics compared to operations

In Table 8.1 all vessels used in this research are listed together with several of their main characteristics. For most vessels the design speed is known and for some vessels the endurance is provided in nautical miles. These vessels are designed to operate at their design speed and to be able to sail at least their specified endurance without stopping.

Using AIS data, it is possible to analyse whether the vessels actually operate close to their design speed. For all vessels, the percentage of time spent sailing within a range of ± 2 knots around the design speed has been analysed. From the previous results it became clear that the average SOG for the HTV fleet is approximately 10 knots, which differs considerably from the design speeds listed in Table 8.1.

Figure 10.4 shows the percentage of time vessels operate within a ± 2 knot range around their design speed and around the AIS-derived cruising speed for the HTV fleet as a whole. The figure shows that for most vessels, except for Fan Zhou 8 and Hua Hai Long, the vessels operate more frequently around 10 knots than around their design speed as stated in the specification sheets. This indicates that the vessels generally operate below their design speed. This behaviour may be influenced by several factors such as weather conditions, operational requirements from clients, fuel efficiency considerations, or scheduling constraints.

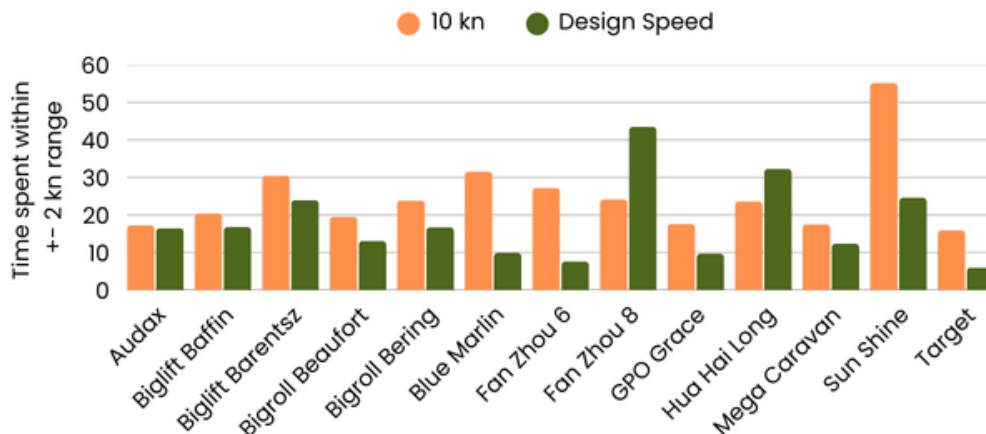


Figure 10.4: Comparison between time spent sailing near AIS-derived cruising speed and design speed

For several vessels, namely Fan Zhou 8, Fan Zhou 6, Audax, and GPO Grace, the endurance is specified in nautical miles in the specification sheets. The long sailing segments derived from Dataset 4 were analysed to determine whether the vessels needed to sail distances close to their design endurance during the period covered by this research.

Table 10.3 shows the endurance listed in the specification sheets compared with the longest sailing segment identified in the AIS data for each vessel. The longest sailing distance observed in the AIS data is 14,072 nautical miles. This suggests that, for example, the endurance of GPO Grace may be somewhat oversized relative to the distances observed during the analysed operational period.

Table 10.3: Comparison between endurance based on specification sheets and AIS data

	Endurance Spec Sheet [nm]	Endurance AIS [nm]
Fan Zhou 8	16,000	7,753
Fan Zhou 6	13,000	11,817
Audax	13,000	12,065
GPO Grace	25,000	13,321

Overall, the results indicate that the design speed and endurance for which these vessels are designed are not fully aligned with their observed operational behaviour. Most vessels operate significantly below their design speed and rarely utilize their full design endurance during the analysed operational period.

10.4. Ship insights conclusion

Overall AIS also showed that both the HTVs and the Semi-Submersibles operate in the offshore wind market therefore it is important to also take the Semi-Submersibles into account. The data also shows that the vessel do not only operate in this market as not all port visits are to known monopile fabrication sites. Furthermore the data showed that the assumed port time should not be based on the amount of monopiles as first of all the vessels do not solemnly operate in this market and secondly this duration is highly dependent on the vessel and port itself as well.

Furthermore the data showed no significant trends over the years it seems that the HTVs can operate very differently per year. The advantage of using Blended Design as an early stage vessel design tool is that it finds the optimized vessel design over its full lifetime and therefore the mean already takes into account some uncertainties which can occur during its lifetime.

Lastly the AIS data showed that most of the vessels actually operate most of the time below the design speed. This shows that the assumptions for these vessels are not in line with the reality. When looking at the Blended Design results, the chance of a significant change of vessel dimensions with a lower design speed is not expected due to the bounded dimensions within this market but the ROI might be lower then assumed.

IV. Conclusion

11

Conclusion

This research investigated how AIS data can be used to derive data-driven operational profiles for HTVs and how these profiles can support early-stage vessel design. The motivation for this research was based on the observation that most vessel designs rely on assumed operational profiles, rather than on actual operational data. As a result, these assumptions may not accurately reflect real vessel behaviour, introducing uncertainty in early-stage design decisions.

To address this gap, this research aimed to establish a structured, end-to-end approach linking AIS data to early-stage vessel design. To answer the main research question, several sub-questions were formulated and are addressed below.

How can AIS data be processed and repaired to enable reliable operational analysis?

AIS data inherently contains noise, inconsistencies, and missing values, which introduce uncertainty in the derived operational profiles. This research demonstrated that a structured preprocessing approach, consisting of multi-level cleaning and trajectory repair, is required to improve data reliability.

The results show that preprocessing significantly improves the quality of individual vessel datasets, as inconsistencies due to interpolation can strongly effect the operational profile. However, at fleet level, the impact of cleaning and repairing is limited due to aggregation effects. As more vessels or years are included, the influence of individual data errors decreases.

This demonstrates that AIS data will never be fully error-free, but that its reliability increases when analysed at fleet level. For vessel-level analysis, and particularly when evaluating extreme values, preprocessing remains essential. It should be noted that this is based on already processed data obtained from MSG [4] and thus not raw AIS data.

How can AIS trajectory data be segmented to identify operational states of HTVs?

Even after preprocessing, AIS trajectories contain noise such as positional drift and speed fluctuations, making it difficult to directly interpret operational behaviour. A multi-level segmentation method was therefore developed, incorporating speed, spatial, and temporal constraints.

The results show that segmentation methods from literature cannot be directly applied without adaptation, likely due to differences in sampling rate, vessel type, and operational region. A tailored segmentation approach was therefore developed.

The developed segmentation method proved to be more accurate than non-segmented approaches for operational profiles and enables more detailed insight into vessel operations. In particular, segmentation allows for the identification of operational states such as anchoring, which cannot be reliably captured using aggregated operational profiles. This additional level of detail provides valuable insight into operational behaviour.

How do AIS-derived operational profiles compare to the assumed operational profiles used by UDSBV?

The comparison showed that the assumed operational profiles used by UDSBV are generally reasonable, but do not fully represent actual vessel behaviour.

UDSBV assumed two distinct operational scenarios (short and long trips), whereas AIS data shows that both behaviours occur in practice and often within the same year. The data-driven operational profiles therefore represent a combination of these scenarios, indicating that the assumptions were not incorrect, but incorrectly distributed.

A key finding is that vessels spend significantly more time in anchorage than assumed. This insight was unexpected and has led UDSBV to reconsider the importance of this operational condition in future vessel designs.

Although UDSBV assumed fewer operational days, this difference is largely compensated by waiting times observed in AIS data. As a result, the overall yearly operational balance is comparable, but the distribution between sailing and non-sailing activities differs.

In addition, Semi-Submersible vessels were found to operate in similar offshore wind markets and exhibit comparable operational behaviour to HTVs. However, due to longer sailing distances per trip, their Return on Investment (ROI) is lower, which aligns with expectations.

What is the impact of incorporating AIS-derived operational profiles on HTV design within early-stage vessel design?

The impact of incorporating AIS-derived operational profiles on the main vessel design parameters was found to be limited. Key characteristics such as vessel dimensions remained largely unchanged, even when comparing assumed and data-driven operational profiles, as well as cleaned and non-cleaned datasets.

This limited sensitivity can be explained by the nature of HTVs, where vessel dimensions are strongly constrained by cargo characteristics, such as monopile dimensions, and external limitations such as the Panama Canal width. As a result, the design space is restricted and less sensitive to variations in operational input.

However, AIS-derived operational profiles do influence operational performance and economic indicators, particularly the Return on Investment and optimal vessel speed. This indicates that the primary value of data-driven operational profiles lies in performance evaluation rather than in determining vessel geometry for the HTV fleet.

How can AIS-derived operational profiles support early-stage vessel design?

AIS-derived operational profiles support early-stage vessel design by providing a realistic and data-driven representation of vessel operations. This enables designers to move beyond purely assumed operational scenarios.

A key contribution of this research for UDSBV is the identification of operational conditions that were previously underestimated, such as anchorage. These insights directly influence design considerations and improve the realism of design input.

Furthermore, no clear long-term trends were identified within the AIS data that require explicit adjustments for future operations. Instead, the variability in operations across vessels and years suggests that using averaged operational profiles over the vessel lifetime is an appropriate approach for early-stage design.

How can data-driven operational profiles be derived from AIS data for early-stage vessel design?

This research demonstrates that data-driven operational profiles for HTVs can be derived from AIS data through structured preprocessing and tailored segmentation methods, and can be effectively integrated within early-stage vessel design frameworks such as Blended Design.

Within this research, a structured link has been established between AIS data and early-stage vessel design, covering the full chain from data preprocessing and trajectory segmentation to operational profile derivation and design application. In this process, both existing tools and newly developed methods were combined to bridge the gap between raw operational data and design input.

While the influence of these profiles on vessel geometry is limited within the applied design context, their main added value lies in improving the realism and reliability of operational assumptions. By grounding design input in real-world AIS data, uncertainty in early-stage vessel design is reduced and operational performance can be evaluated more accurately.

Furthermore, this research shows that for fleet-level operational analysis, the impact of additional data cleaning and repairing is limited when a sufficiently large and diverse dataset is used. By including multiple vessels of the same type, across different years and operational regions, the influence of individual data errors is reduced through aggregation effects. As a result, inconsistencies within specific vessels, years, or regions become less dominant in the overall operational profile.

However, for vessel-level analysis or when evaluating extreme values, data cleaning and repairing remain essential, as local errors can still have a significant impact on the results. This research contributes to a more transparent, data-driven, and robust approach to early-stage vessel design. Table 11.1 shows the overall influence found in this research for the HTV fleet.

Table 11.1: Summary of impact of cleaning and repairing and data-derived operational profiles

	Influence of cleaning and repairing	Influence of data-derived operational profiles
Vessel Level	Large	-
Fleet Level	Small	New insights
Design impact	None	Limited

12

Discussion

This research demonstrates the potential of data-driven vessel design using AIS data as input for design tools. The results show that AIS-derived operational profiles can provide valuable insight into how HTVs are actually operated in practice. By analysing AIS data for multiple vessels over several years and applying a multi-level segmentation method, representative operational profiles were derived. These profiles reveal operational patterns that are not captured by the assumed operational profiles traditionally used during vessel design. Although the impact on the final design was limited, valuable insights were obtained. However, several limitations should be considered when interpreting the results of this research.

Limitations of the segmentation method

The segmentation method developed in this study was specifically designed for HTV operations. As a result, the thresholds and operational states identified in the segmentation process are closely linked to the operational characteristics of this vessel type. Applying the same segmentation method to other vessel types would maybe require adjustments to both the thresholds and the segmentation logic. Furthermore, some HTVs are capable of performing DP operations. The segmentation method used in this research was not able to explicitly identify DP operations, which may therefore be included within other operational categories.

Limitations of the AIS dataset

Another limitation of the dataset is the uneven usage of AIS data across the vessels. For some vessels, up to eight years of AIS data were analysed, while for others only two or three years were used. Although the analysis showed that operational profiles can vary significantly between years for the same vessel, suggesting that the influence of uneven datasets is limited in this case, this effect may be more significant for other vessel types. If a vessel with a specific operational profile is analysed over a much longer period than others, it may exert a disproportionate influence on the fleet-level operational profile.

Representation of operational behaviour

The operational profiles derived in this research were based on mean values per trip. This approach was chosen because it provides suitable input for the Blended Design methodology used in this study. However, using mean values results in a simplified representation of operational behaviour. Alternative approaches could include identifying the most common operational patterns using clustering techniques, such as k-means clustering. These methods could potentially identify representative operational profiles rather than averaging behaviour across multiple vessels and years.

Uncertainty in data processing

Uncertainty may also be introduced through the trajectory repair method used to reconstruct missing AIS data. In this research, missing data was repaired using linear interpolation on a sphere when consistent vessel behaviour was observed before and after the gap. Although this method proved effective for many trajectory gaps, more advanced interpolation techniques could further improve the

reconstruction of missing vessel movements. However, given the high data coverage in this dataset, the influence of improving trajectory reconstruction is expected to be minimal in terms of its impact on the final design.

Another limitation concerns the use of AIS vessel status information for identifying operational states. In this research, the majority status of non-sailing segments was used to distinguish between different stationary behaviours. However, AIS status information is not always reliable. In some cases, delays were observed in updating the vessel status, or the correct status was not transmitted at all. Although these inconsistencies appeared to occur relatively infrequently in the analysed dataset, this remains an assumption. In datasets where AIS status information is less reliable, alternative methods based on movement behaviour may be required to identify operational states.

Lack of contextual information

Furthermore, AIS data does not provide contextual information about vessel operations. The segmentation method developed in this research identifies whether a vessel is sailing or not, but it does not reveal the underlying reasons for these operational states. Non-sailing periods may occur due to cargo loading or unloading, waiting for weather windows, operational decisions, or port availability constraints. Without additional contextual information, it is therefore not always possible to determine the exact cause of operational behaviour. The inclusion of weather data, such as that provided by MSG [4], could offer additional context. While this may not directly influence the operational profile, it could enhance understanding and support more informed decision-making during early-stage vessel design.

Influence of design methodology

Finally, the influence of AIS-derived operational profiles on vessel design is also dependent on the characteristics of the design tool used. In this research, the Blended Design methodology was applied to evaluate the impact of operational profiles on HTV design. This approach is primarily focused on vessels transporting monopiles and evaluates performance over the vessel's lifetime. As a result, strong design constraints, such as cargo dimensions and infrastructure limitations, can restrict the influence of operational profile variations on the final vessel geometry. In markets where cargo types differ significantly from monopiles, the influence of operational profiles on vessel design may become more pronounced.

Overall reflection

Despite these limitations, the results of this research demonstrate that AIS-derived operational profiles provide valuable insight into real vessel operations and can support more data-driven decision-making during early-stage vessel design. Incorporating AIS-based operational analysis into the design process can contribute to a better understanding of vessel operational profile, and economic performance. Furthermore, this research shows that highly detailed cleaning and repairing procedures are primarily beneficial at the individual vessel level rather than at the fleet level. Therefore, when applying AIS data in early-stage ship design, it is important to clearly define the scope of the analysis in order to determine which preprocessing steps are necessary and justified.

13

Recommendations

Based on the results, conclusions, and limitations identified in this research, several recommendations for future research can be made.

One recommendation is to test the segmentation method for other vessel types and analyse how difficult it is to obtain reliable segments for vessels with different operational behaviour. The segmentation approach developed in this research was specifically tuned to the operational characteristics of HTVs, and therefore adjustments may be required when applying it to other vessel classes.

A second recommendation is to develop a reliable method to determine whether a vessel is moored or not. Existing research proposes several methods for identifying mooring behaviour, but these approaches were not feasible within this research due to limitations in the available AIS data and time constraints. A preliminary attempt was made by clustering the no-sailing segments based on behavioural characteristics such as very low average SOG and minimal heading variation. The elbow method was used to determine the number of clusters, after which k-means clustering was applied to both the speed distribution and heading change. Although this approach showed promising results, challenges remained because certain mooring segments and DP segments have very similar movement characteristics and were therefore sometimes clustered together. Future research could improve this classification by combining movement-based clustering with additional contextual information such as distance to shore or port locations, although such datasets were not available during this research.

In addition, it is recommended to investigate the influence of different market scenarios within the Blended Design framework. HTVs operate in multiple markets, and incorporating different operational scenarios into the design process could provide further insight into how operational variability influences vessel performance and design outcomes.

Finally, it would be valuable to apply the developed methodology to other vessel types. Different vessel classes may exhibit significantly different operational patterns, and the influence of data-driven operational profiles on vessel design may therefore vary depending on the vessel type.

References

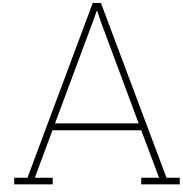
- [1] J. J. Zwaginga. *Exploring market uncertainty in early ship design*. 2020.
- [2] *AIS transponders*. URL: <https://www.imo.org/en/OurWork/Safety/Pages/AIS.aspx>.
- [3] Fugro. *Fugro Offshore Wind Farm Site Appraisal and Concept Design - 9.43MB - 1m 36s*. Apr. 2023. URL: <https://www.fugro.com/news/long-reads/2023/how-offshore-wind-is-evolving-to-hit-2050-net-zero-targets>.
- [4] Made Smart Group. *Made Smart Group*. May 2025. URL: <https://help.madesmart.nl/m/prospector/1/736495-introduction>.
- [5] Robert Keane, Laurent Deschamps, and Steve Maguire. "Reducing Detail Design and Construction Work Content by Cost-Effective Decisions in Early-Stage Naval Ship Design". In: *Journal of Ship Production and Design* 32 (May 1, 2016), pp. 110–123. DOI: 10.5957/JSPD.32.2.140028.
- [6] Henrique Murilo Gaspar. *Data-Driven Ship Design*. Accepted: 2021-03-23T10:01:43Z Publication Title: 426-439. 2018. ISBN: 978-3-89220-707-8. URL: <https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2735012> (visited on 06/20/2025).
- [7] Apostolos Papanikolaou. *Ship design: methodologies of preliminary design*. Springer, 2014.
- [8] Julie Chalfant. "Early-Stage Design for Electric Ship". In: *Proceedings of the IEEE* 103.12 (2015), pp. 2252–2266. DOI: 10.1109/JPROC.2015.2459672.
- [9] Taewoo Nam and Dimitri Mavris. "Multi-stage reliability-based design optimization for aerospace system conceptual design". In: *49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 16th AIAA/ASME/AHS Adaptive Structures Conference, 10th AIAA Non-Deterministic Approaches Conference, 9th AIAA Gossamer Spacecraft Forum, 4th AIAA Multidisciplinary Design Optimization Specialists Conference*. 2008, p. 2073.
- [10] Francesco Baldi et al. "Energy and exergy analysis of a cruise ship. NOTE: PLEASE REFER TO THE JOURNAL ARTICLE https://www.researchgate.net/publication/327796672_Energy_and_Exergy_Analysis_o In: June 2015.
- [11] N.L. Trivyza, A. Rentizelas, and G. Theotokatos. "The influence of ship operational profile in the sustainability of ship energy systems". English. In: *International Conference of Maritime Safety and Operations 2016, MSO 2016*; Conference date: 13-10-2016 Through 14-10-2016. Oct. 2016. URL: <http://www.incass.eu/mso-2016/welcome/>.
- [12] Anna Swider and Eilif Pedersen. "Data-Driven Methodology for the Analysis of Operational Profile and the Quantification of Electrical Power Variability on Marine Vessels". In: *IEEE Transactions on Power Systems* 34.2 (Mar. 2019), pp. 1598–1609. ISSN: 1558-0679. DOI: 10.1109/TPWRS.2018.2876252. URL: <https://ieeexplore.ieee.org/abstract/document/8492442> (visited on 06/20/2025).
- [13] Sepideh Jafarzadeh and Ingrid Schjøberg. "Operational profiles of ships in Norwegian waters: An activity-based approach to assess the benefits of hybrid and electric propulsion". In: *Transportation Research Part D: Transport and Environment* 65 (Dec. 2018), pp. 500–523. ISSN: 13619209. DOI: 10.1016/j.trd.2018.09.021. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1361920918300580> (visited on 06/20/2025).
- [14] Celine Mysen. "Understanding Offshore Support Vessel Activity by the Use of AIS Data". Accepted: 2019-11-01T15:00:33Z. Master thesis. NTNU, 2019. URL: <https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2626177> (visited on 06/19/2025).
- [15] Michael Smith. "Improvements to the design process of the energy distribution system". In: ().
- [16] Roy de Winter and B Veen. *The Added Value of AIS Data Analysis for Naval Architects*. Sept. 12, 2024.

- [17] Jongseo Park and Minjoo Choi. "A K-Means Clustering Algorithm to Determine Representative Operational Profiles of a Ship Using AIS Data". In: *Journal of Marine Science and Engineering* 10.9 (Sept. 2022). Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, p. 1245. ISSN: 2077-1312. DOI: 10.3390/jmse10091245. URL: <https://www.mdpi.com/2077-1312/10/9/1245> (visited on 05/19/2025).
- [18] International Electrotechnical Commission. *IEC 61993-2:2018 – Maritime Navigation and Radio-communication Equipment and Systems – Automatic Identification Systems (AIS) – Part 1: AIS Class A Shipborne Equipment*. International Standard. Geneva: IEC. 2018.
- [19] Planevision Systems. *What is AIS? | Planevision Systems*. Oct. 17, 2025. URL: <https://planevision.systems/what-is-ais/>.
- [20] Comar Systems. *What is AIS? - Comar Systems*. Oct. 15, 2018. URL: <https://comarsystems.com/what-is-ais/#:~:text=How%20does%20AIS%20work?,communicate%20with%20other%20nearby%20ships.&text=information%20is%20combined%20with%20other,RADAR%20picture%20when%20used%20together..>
- [21] *How AIS works | Navigation Center*. URL: <https://www.navcen.uscg.gov/how-ais-works?pageName=AboutDSC>.
- [22] Xiaoguang Wang et al. "Vessel route anomaly detection with Hadoop MapReduce". In: *2014 IEEE International Conference on Big Data (Big Data)*. 2014, pp. 25–30. DOI: 10.1109/BigData.2014.7004464.
- [23] International Electrotechnical Commission. *IEC 61993-2:2018 – Maritime Navigation and Radio-communication Equipment and Systems – Automatic Identification Systems (AIS) – Part 1: AIS Class A Shipborne Equipment*. International Standard. Geneva: IEC. 2018.
- [24] International Electrotechnical Commission. *IEC 62287-1:2017 – Maritime Navigation and Radio-communication Equipment and Systems – Automatic Identification Systems (AIS) – Part 2: AIS Class B Shipborne Equipment*. International Standard. Geneva: IEC. 2017.
- [25] Duong Nguyen et al. "Multi-task Learning for Maritime Traffic Surveillance from AIS Data Streams". In: *CoRR* abs/1806.03972 (2018). arXiv: 1806.03972. URL: <http://arxiv.org/abs/1806.03972>.
- [26] Yiheng Chen et al. "Optimizing vessel trajectories: Advanced denoising and interpolation techniques for AIS data". In: *Ocean Engineering* 327 (May 30, 2025), p. 120988. ISSN: 0029-8018. DOI: 10.1016/j.oceaneng.2025.120988. URL: <https://www.sciencedirect.com/science/article/pii/S0029801825007012> (visited on 05/21/2025).
- [27] *Satellite-based AIS*. July 11, 2013. URL: <https://www.barentswatch.no/en/articles/satellite-based-ais/>.
- [28] J.J. de Ridder. *A Moonshot for Future Offshore Wind Turbine Installation Vessels for an Uncertain Market*. 2023.
- [29] Jorick Tjaberings, Stefano Fazi, and Evrim Ursavas. "Evaluating operational strategies for the installation of offshore wind turbine substructures". In: *Renewable and Sustainable Energy Reviews* 170 (Dec. 2022), p. 112951. ISSN: 13640321. DOI: 10.1016/j.rser.2022.112951. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1364032122008322> (visited on 07/01/2025).
- [30] Clarksons. *Heavy Transport Vessels Market Analysis & Business Case HX121*. June 19, 2023.
- [31] Ulstein Design & Solutions BV. *TBLSHIPQ32025*. 2025.
- [32] Jianwen Ma, Yue Zhou, and Zhaoxin Zhu. "Identification and analysis of ship waiting behavior outside the port based on AIS data". In: *Scientific Reports* 13.1 (2023), p. 11267.
- [33] Chris Veness, www.movable-type.co.uk. *Calculate distance and bearing between two Latitude/Longitude points using haversine formula in JavaScript*. URL: <https://www.movable-type.co.uk/scripts/latlong.html>.
- [34] Wayan Mahardhika Wijaya and Yasuhiro Nakamura. "Port performance indicators construction based on the AIS-generated trajectory segmentation and classification". In: *International Journal of Data Science and Analytics* 20.3 (2025), pp. 2473–2492.

- [35] Zhaojin Yan et al. "Extracting ship stopping information from AIS data". In: *Ocean Engineering* 250 (2022), p. 111004.
- [36] Roll Group. *MC-Class – Roll Group*. Apr. 15, 2025. URL: <https://www.roll-group.com/fleet-equipment/mc-class/>.
- [37] Xuan Guo et al. "Ship trajectory segmentation by movement states while addressing uncertainty and sparsity". In: *Ocean Engineering* 312 (2024), p. 119218. ISSN: 0029-8018. DOI: <https://doi.org/10.1016/j.oceaneng.2024.119218>. URL: <https://www.sciencedirect.com/science/article/pii/S0029801824025563>.
- [38] Nestor Goicoechea and Luis María Abadie. "Optimal Slow Steaming Speed for Container Ships under the EU Emission Trading System". In: *Energies* 14.22 (2021). ISSN: 1996-1073. DOI: 10.3390/en14227487. URL: <https://www.mdpi.com/1996-1073/14/22/7487>.
- [39] V.P.M. Peeten. *Evaluating emission of offshore transportation vessels*. 2022.
- [40] J.F. Schuitemaker. *Impacts of lifecycle perspectives in early stage ship design*. 2024.
- [41] *Spinergie*. URL: https://market-intelligence.spinergie.com/sso/email-identification?_target_path=https%3A%2F%2Fmarket-intelligence.spinergie.com%2Fdashboard%2Fpage%2Fvessel%3Fmap%3Aactive%3Dwindfarms%252CvesselsLatestPositions%26map%3AweatherLayerShown%3Dnone%26map%3Apos%3D56.6799%252C7.0803%252C6%26blocks%3AisNotOpenLicense%3Dtrue%26vesselId%3D9962%26componentsState%3Dvessel-schedule_groupby%3Dvessel_title%24vessel-schedule__splitby%3DprojectActivityType%24vessel-schedule__excludedEntities%3D%24vessel-schedule__period%3D1718480713000%252C1859223600000%24vessel-schedule__layers%3Dannotations%252CvesselEvents%252Ccharters%252CcustomContracts%252Cworkpackages%252Cvoyages%252Cactivities%24vessel-schedule__chart_analysis%3DwindAtVessel.
- [42] Dick Hill. *Clarksons Report Fan Zhou 8 Bound for Eemshaven with 10 monopiles for RWE Offshore Wind Farm*. Apr. 28, 2025. URL: <https://www.heavyliftnews.com/clarksons-report-fan-zhou-8-bound-for-eemshaven-with-10-monopiles-for-rwe-offshore-wind-farm/>.
- [43] *MV BigLift Baffin delivers 38 jacket foundations at Saint-Brieuc offshore wind farm*. URL: <https://www.bigliftshipping.com/en/projects/mv-biglift-baffin-delivers-38-jacket-foundations-at-saint-brieuc-offshore-wind-farm>.
- [44] *Dongbang Transport Logistics Co., Ltd.* URL: <https://www.dongbang.co.kr/eng/service/ser04.php>.
- [45] Han Heilig. *CWHI completes delivery monopiles for Inch Cape OWF*. Feb. 27, 2026. URL: <https://ocean-energyresources.com/2026/02/27/cwhi-completes-delivery-monopiles-for-inch-cape-owf/>.
- [46] *Initial Loadouts for the Thor OWF Project*. Apr. 22, 2025. URL: <https://eew-group.com/projects-references/success-stories/news-detail/initial-loadouts-for-the-thor-owf-project/>.
- [47] *Terrein Buss Terminal gevuld met onderdelen voor windpark Kaskasi*. Jan. 23, 2022. URL: <https://eemshavenonline.nl/havenonline/terrein-buss-terminal-gevuld-met-onderdelen-voor-windpark-kaskasi/>.
- [48] Dick Hill. *DEME signs with Port of Thyboron for Marshalling Vesterhav Nord and Syd Foundations*. Feb. 10, 2023. URL: https://www.heavyliftnews.com/deme-signs-with-port-of-thyboron-for-marshalling-vesterhav-nord-and-syd-foundations/?utm_source=chatgpt.com.
- [49] *Last monopiles for Hornsea two left the port of Rostock*. Aug. 9, 2021. URL: <https://eew-group.com/projects-references/success-stories/news-detail/last-monopiles-for-hornsea-two-left-the-port-of-rostock/>.
- [50] *MANUFACTURING - CWHI (CNOOD-WENCHONG HEAVY INDUSTRIES)*. URL: <https://www.cwhi.com/manufacturing/>.
- [51] *Dajin Offshore Heavy industry*. URL: <https://www.dajin.cn/index/index/ports?lang=EN>.

- [52] *Taisheng Landau held the delivery ceremony of the first batch of monopile foundations for the 310MW offshore wind power project in Pingda, Vietnam*. *Nantong Taisheng Blue Island Offshore Co., Ltd. self-elevating platform Fixed platform Maritime wind power equipment*. URL: <http://en.biooffshore.com/news/3.html>.
- [53] *Facilities - CS Wind Offshore*. URL: <https://www.cswoffshore.com/facilities/>.
- [54] *EEW Group | Represented worldwide with 6 production sites*. URL: <https://eew-group.com/company/our-facilities/>.
- [55] *Haizea Wind Group. Home Haizea Wind Group - English - Haizea Wind Group*. Feb. 24, 2026. URL: <https://haizeawindgroup.com/en/inicio-haizea-wind-group-english/>.
- [56] *Navantia UK. Methil*. Mar. 3, 2026. URL: <https://navantia.co.uk/methil/>.
- [57] *Morgan Roy, Century Wind Power HomeMesh*. URL: <https://www.cwptw.com/en/>.
- [58] *Monopiles | Marktleder met 75+ jaar ervaring*. Jan. 21, 2026. URL: <https://sif-group.com/nl/producten-diensten/monopiles>.
- [59] *Steelwind Nordenham. Monopiles - Steelwind Nordenham*. Nov. 2024. URL: <https://en.steelwind-nordenham.de/products/monopiles/>.
- [60] *Home - English - Titan Wind Energy (Germany) GmbH*. Feb. 9, 2026. URL: <https://titan-wind.de/en/titan-wind-cuxhaven-monopile-manufacturing-for-offshore-wind/>.
- [61] *Navantia Seanergies-Windar Renovables will build 21 monopiles for Iberdrola for its Windanker offshore wind farm in the Baltic Sea - Navantia*. URL: <https://www.navantia.es/en/news/press-releases/navantia-seanergies-windar-renovables-will-build-21-monopiles-for-iberdrola-for-its-windanker-offshore-wind-farm-in-the-baltic-sea/>.

V. Appendix



Blended Design additional information

In this appendix, additional information about the design tool Blended Design is presented. As mentioned previously, Blended Design was developed by UDSBV in collaboration with several MSc students and was provided as a tool for this research. An in-depth understanding of the tool was beyond the scope of this study. For further details, the theses of Zwaginga [1], Peeten [39], de Ridder [28], and Schuitemaker [40] are recommended. It should be noted that the tool has been further developed internally since these studies.

Blended Design requires multiple input files some of which are changed each time the tool is used, while others are modified less frequently. Currently, Blended Design is configured for four different vessel types: Subsea Rock Installation (SRI), HTV, Foundation Installation Vessel (FIV), and Wind Turbine Installation Vessel (WTIV). For each vessel type, there is a corresponding CSV file containing a reference vessel with its particulars and design assumptions. Additionally, Blended Design includes a CSV file with constants. Market data for different cargo types, sourced from Spinergie [41], is also incorporated through CSV files. While this market data should be updated regularly, it does not need to be refreshed every time Blended Design is used this is the same for fuel price data.

Figure A.1 illustrates the input for Blended Design it is divided into categories for example all the green colored input is used to create the ship configurations where the orange colored input is used to determine the cost of the vessel and the pink is related to the income, purple to the markets, grey is related to fuel type and the salmon color is related to the operation the vessel will fulfill during its lifetime. The first thing in the input file is selecting the ship type. Once the ship type is chosen, Blended Design will automatically load the relevant market data for the associated cargo. For example, selecting an HTV will load data on monopiles and jacket dimensions. The next step involves defining the ranges and increments for the ship configurations, as previously explained. The mission input is provided via a CSV file, which currently contains various operational profiles based on parameters such as distance per trip, number of days in dynamic positioning, endurance days and waiting times in port. Additional inputs must also be specified, such as contract type, fuel type, the number of persons on board, and the number of operational days per year. These inputs are essential for evaluating the performance of each configuration over its lifetime, ensuring that it can meet the predefined operations and objectives.

	Input Name	Meaning	Input Format	Example	Note
	Vessel_type	Vessel type	Selection	SRI / HTV / FIV / WTIV	
	Assumptions	-	CSV	Table	CSV for each vessel type
Ship Configuration	L	Length	Range	[start, stop, end]	
	B	Breadth	Range	[start, stop, end]	
Costs	D	Depth	Range	[start, stop, end]	
	Vs	Design Speed	Range	[start, stop, end]	
Income	CC	Crane Capacity	Range	[start, stop, end] / 0 (no crane)	
	CB	Block Coefficient	Range	[start, stop, end]	
Market	POB	Persons on Board	Number	#	
	Rock_cap	Rock Capacity	Range	[start, stop, end]	Only applicable for SRI
Fuel	CO2taxvalue	CO2 Tax	Range	[start, stop, end]	Euro / ton CO2
	Motioncompcost	Motion Compensation Cost	Number	#	Euro / (t * m) (Specific for WTIV)
Operation	Shipcost	Ship Costs	CSV	Table	Reference values for CAPEX and OPEX, based on ship parameters
	Shipfinancing	Ship Financing	CSV	Table	Financial parameters
	Unit_income	-	Number	#	FIV/HTV : Income per ton cargo SRI : Income per ton rock WTIV: Income per MW of turbine
	Distance_rate	-	Number	#	Bonus reward based on distance in Euro / ton / NM
	Charter_rate	-	Number	#	Fixed day rate
	Contract_type	-	Selection	Performance / Charter / Both	Unit Income and Distance income are used for performance-based contracts and charter rate is used for charter contracts
	Projection Input	-	CSV	Table	Input for market module
	Market_	-	CSV	Table	For both jackets and monopiles
	Pmatrix_	-	CSV	Table	For both jackets and monopiles
	Sizes_	-	CSV	Table	For both jackets and monopiles
	Rock_data	Rock Data	CSV	Table	Market specifics for SRI
	Fuel_type	Type of Fuel	Selection	MDO / Methanol / Ammonia / Hydrogen	
	Fuel_data	Fuel Data	CSV	Table	CSV for each fuel type containing fuel constants
	Fuelprices	Fuel Prices	CSV	Table	Current fuel prices and projections
	Mission	-	CSV	Table	CSV includes different missions
	DP_capability	Dynamic Positioning Capability	Number	Percentage	Shows how much of the standard environment by DNV can be handled
	Oper_days	Operational Days	Number	#	

Mission	Distance	Total Distance	DP	Self Supportive
1	nm	nm	days	days
2	nm	nm	days	days
3	nm	nm	days	days

An arrow points from the 'Mission' column of the second table to the 'Mission' input row of the main table.

Figure A.1: Inputs for Blended Design

Figure A.2 illustrates how Blended Design generates multiple ship configurations based on the input ranges for main dimensions. Users define start and end values and a step size for each parameter, allowing control over the level of detail. This approach limits the design space to relevant configurations while reducing computational effort.

Length	Beam	Depth	Speed
L 1	B 1	D 1	Vs 1
			Vs 2
	D 2		Vs 1
			Vs 2
	B 2	D 1	Vs 1
			Vs 2
D 2		Vs 1	
		Vs 2	
L 2	B 1	D 1	Vs 1
			Vs 2
	D 2		Vs 1
			Vs 2
	B 2	D 1	Vs 1
			Vs 2
D 2		Vs 1	
		Vs 2	

Figure A.2: Example of how ship configurations are generated in Blended Design

In the first step of the ship module, all parameter combinations are generated. However, not all represent realistic designs. Evaluating all configurations would be computationally inefficient, and many combinations are known beforehand to be infeasible or incompatible with the applied methods. For example, the Holtrop & Mennen method for resistance is only valid within specific L/B ratios.

To address this, Blended Design computes key design coefficients (e.g., Froude number and block coefficient) using scaling from a reference vessel. Configurations outside valid method ranges or inconsistent with physical and naval architectural principles are filtered out.

The remaining configurations are evaluated through a sequence of ship modules, which determine characteristics such as resistance, propulsion performance, installed power, light ship weight, deadweight, and loading conditions. An overview of these modules is shown in Figure A.3.

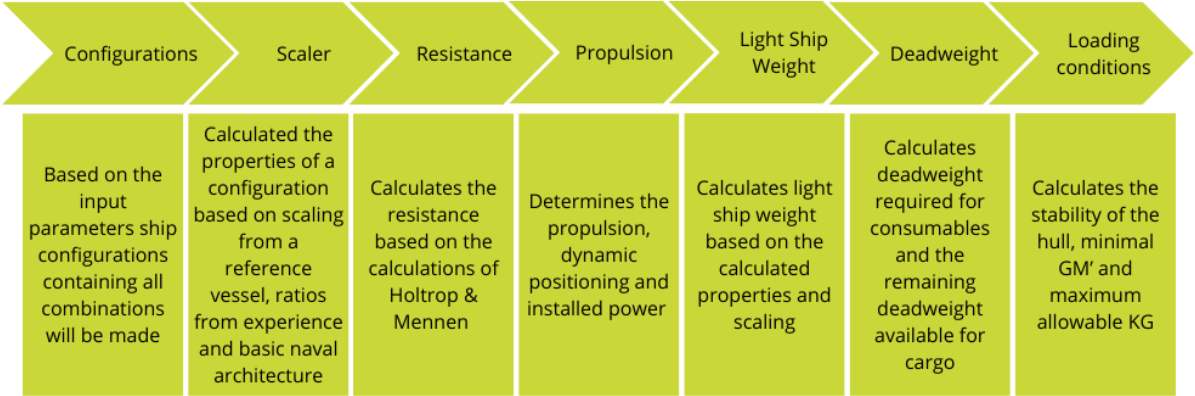


Figure A.3: Overview of the ship modules in Blended Design

The market model simulates the current offshore wind market and projects future trends over the economic lifetime of a vessel. It is based on data from the Spinergie [41] database, which includes information on both existing wind farms and planned projects. Figure A.4 illustrates how the market forecast is generated for monopile diameters. Cargo dimensions, such as length and diameter, are derived from known parameters, for example, the rated power of a wind turbine. Alternatively, the user can provide a custom table that cross-references turbine power with other parameters. From the available data, a trend line is extrapolated to estimate future market developments. At the same time, the designer retains the flexibility to generate forecasts using experimental data or professional experience, enabling the exploration of both broad and narrow market scenarios.

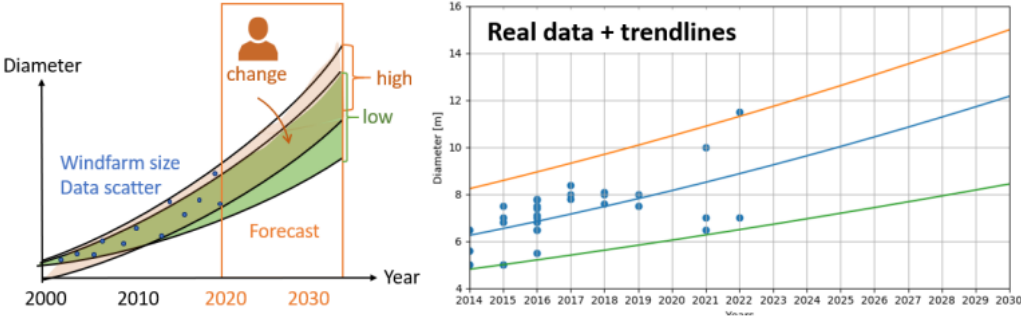
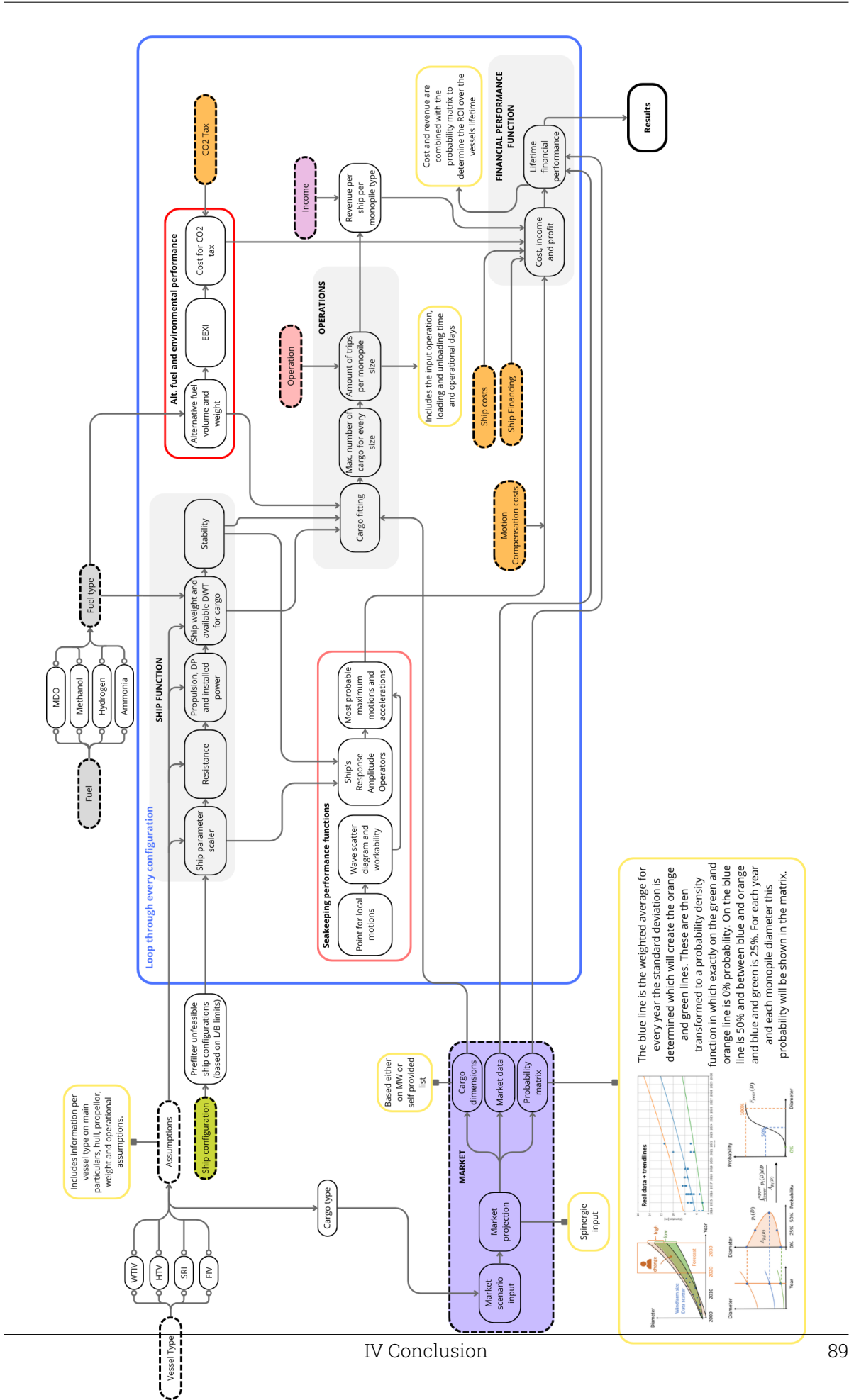


Figure A.4: Market module forecast by Zwaginga [1]

Based on these trend lines explained above the probability function explained in chapter 5 are made. To eventually obtain a probability matrix. When combining all the information for all the configurations the champion with the highest return on investment over its whole lifetime will be chosen first design step.

Figure A.5 shows an overview of the loops within Blended Design. The red-circled functions represent features available in Blended Design but are unlikely to be used in this research. The seakeeping performance function, for example, is only applied when the vessel type is WTIV. The yellow blocks provide additional information, and the blue box highlights the steps that loop through each configuration. The dotted aligned blocks represent the inputs shown in Figure A.1.



IV Conclusion

Figure A.5: A loop trough Blended Design

B

Settings used in Prospector by MSG to collect the AIS data

In this appendix, the download settings used to obtain the datasets for this research are presented. Figure B.1 shows the first menu encountered when downloading data from MSG [4]. For both Dataset 1 and Dataset 2, the source was set to *All Data* and the type to *Positions and Statics*.

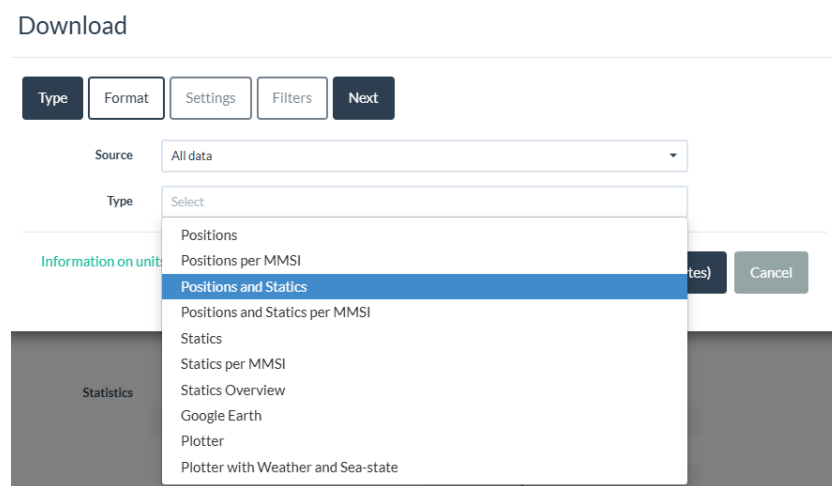


Figure B.1: Download type settings

Figure B.2 shows the format options. In this research, the *Tab Separated* format was used.

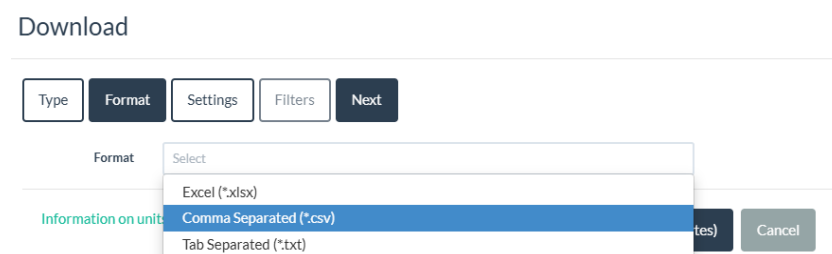


Figure B.2: Download format settings

For both Dataset 1 and Dataset 2, the setting *Remove Position Jumps* was enabled. For Dataset 2,

the *Interpolate* option was additionally enabled, using an interval of 300 seconds and a maximum gap of 30 minutes. The option to preserve original positions was not enabled, as this would have resulted in duplicate values.

Download

Type Format Settings Filters Next

General

Adjust time for AIS time stamps No

Remove position jumps Yes

Interpolate

Interval 300 seconds

Maximum gap 30 minutes

Preserve original positions No

Decimate

[Information on units and conventions used](#) Download (0 bytes) Cancel

Figure B.3: Download settings

Finally, Figure B.4 shows that no filters were applied during this research.

Download

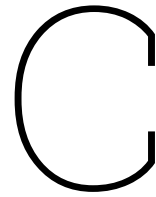
Type Format Settings Filters

+ -

Name	Description
No filters added	

[Information on units and conventions used](#) Download (0 bytes) Cancel

Figure B.4: Download filter settings



Overview of Python-Based Workflows Supporting the Methodology

This appendix provides an overview of the Python-based workflows used throughout the methodology for creating Dataset 4. The workflow transforms raw and interpolated AIS data into a cleaned and repaired dataset that is used in this research for further analysis. The overall process consists of preprocessing, metric computation, data cleaning, gap identification, and gap repair.

The process starts with converting the txt files from MSG [4] into structured csv files. To allow flexibility in the selection of vessels, an input matrix is used in which the vessels included in the process can be switched on or off. Files are standardized, timestamps are formatted consistently, and Dataset 2 is merged into a single chronological dataset per vessel instead of separate datasets per vessel per year. This process is shown in Figure C.1.

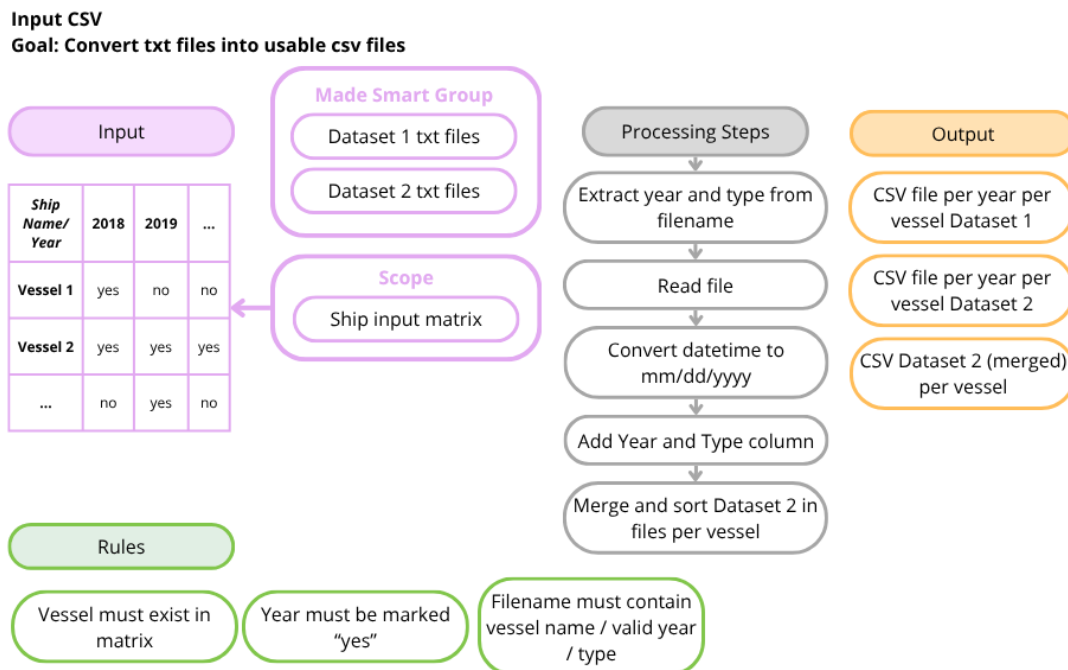


Figure C.1: Workflow for converting raw AIS text files into standardized CSV files and merging vessel data based on the selected years from the input matrix.

Next, additional metrics are computed for each AIS point. These include the time difference between

consecutive points, the travelled distance, and a calculated SOG based on these values. Special care is taken to avoid unrealistic calculations across large time gaps, such as between non-consecutive years in the merged Dataset 2. These steps are illustrated in Figure C.2.

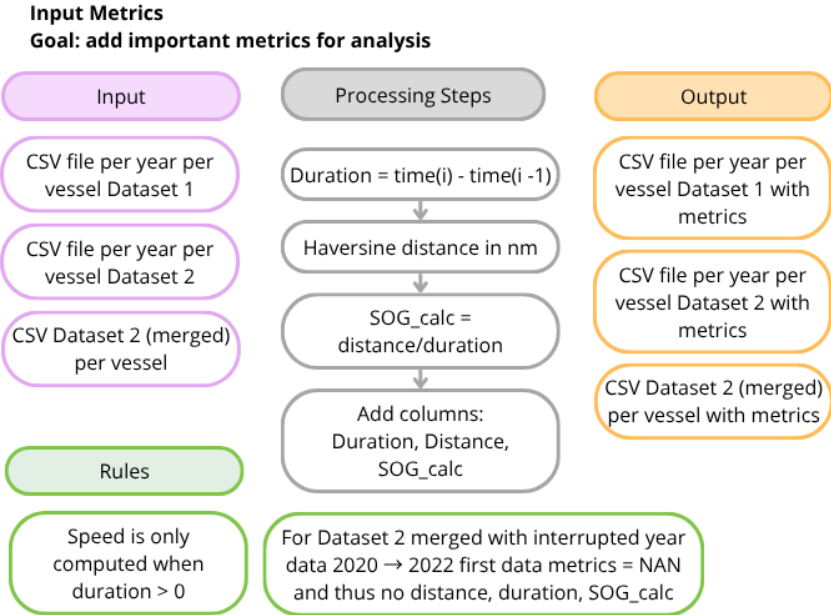


Figure C.2: Computation of AIS-based metrics, including time differences, travelled distance, and calculated speed between consecutive points.

The merged dataset is then cleaned in three consecutive steps. First, invalid or missing values are removed. Second, unrealistic jumps are filtered using a last valid point approach. Third, inconsistencies between reported speed and travelled distance are removed for fixed time intervals. After each step, the metrics are recalculated. This cleaning process is shown in Figure C.3.

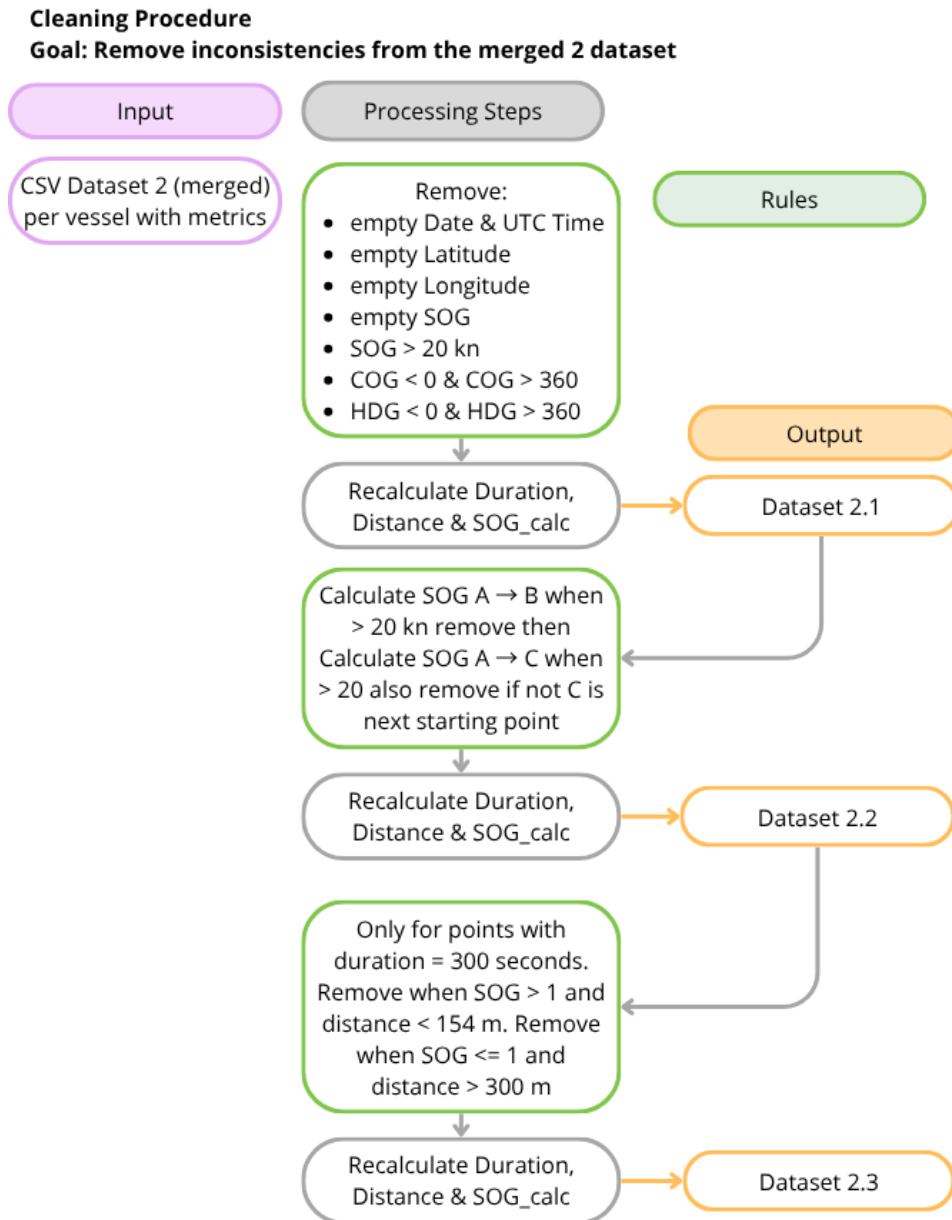


Figure C.3: Three-step cleaning process applied to the merged AIS dataset, including removal of invalid data, filtering based on last valid points, and consistency checks between speed and distance.

After cleaning, temporal gaps in the data are identified based on the time difference between consecutive points. These gaps are classified as stationary, moving, or unknown depending on the vessel behaviour before and after the gap. This classification is shown in Figure C.4.

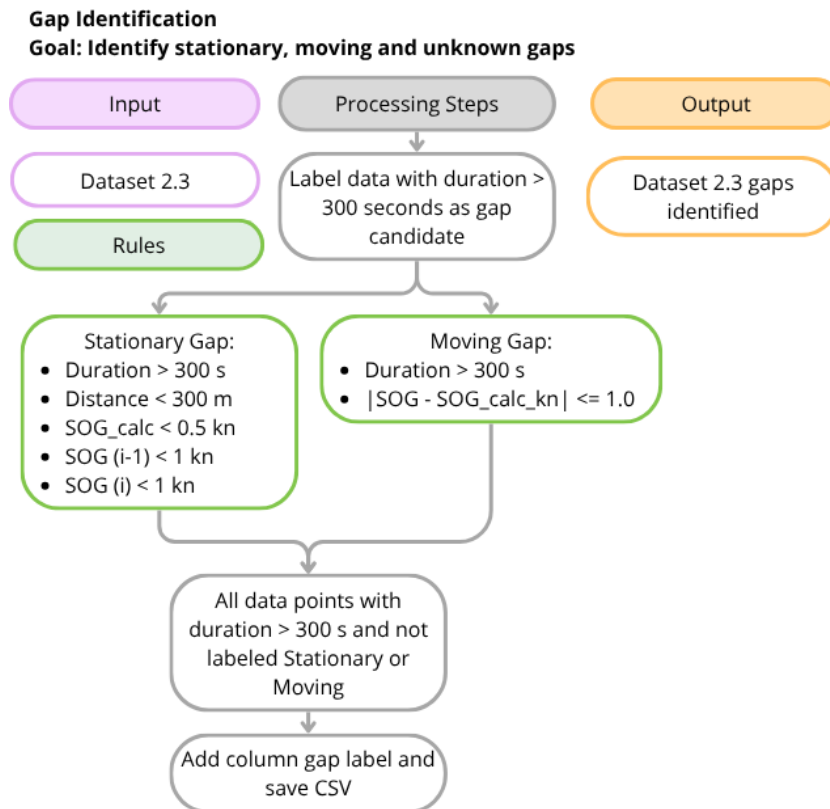


Figure C.4: Classification of temporal gaps into stationary, moving, and unknown categories based on vessel behaviour before and after the gap.

Finally, selected gaps are repaired by inserting interpolated points. Stationary gaps are filled with points at a fixed location, while moving gaps are filled using interpolation between known positions. Unknown gaps are left unchanged. The repair process is illustrated in Figure C.5.

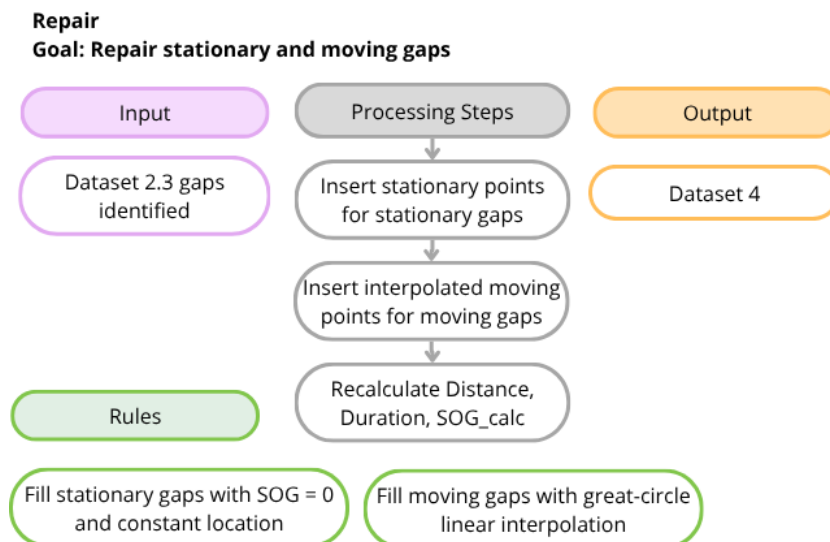


Figure C.5: Gap repair process where stationary gaps are filled with fixed-position points and moving gaps are interpolated between known positions.

D

Operations within the offshore wind market

In this appendix, several operations within the offshore wind market performed by some of the vessels used in this research are presented. In Table D.1, several operations involving the transportation of offshore wind foundations carried out by vessels in this research are shown. It was difficult to determine how the transportation of offshore wind foundations was performed and by which vessel, as this information was often not explicitly reported.

Table D.1: Operations within the offshore wind market

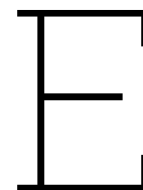
Project	Vessel	Type	Cargo	Year	From - To
Windpark Nordseecluster [42]	Fan Zhou 8	SS	Monopiles	2025	Penglai (CN) - Eemshaven (NL)
Saint-Brieuc [43]	Biglift Baffin	HTV	Monopiles	2023	El Ferrol (ES) - Saint Brieuc
Yunlin [44]	Dongbang Giant No.6	HTV	Monopiles	2020	Nordenham (DE)- Mai Liao (TW)
Formosa II [44]	Dongbang Giant No.6	HTV	Jackets	2021 - 2022	Batam (ID) - Taichung (TW)
Calvados [44]	Dongbang Giant No.6	HTV	Monopiles & TP	2023	Rostock (DE) - Flushing (NL)
Zhongneng [44]	Dongbang Giant No.7	HTV	Jackets	2022 - 2023	Goseong (KR) - Kaohsiung (TW)
Changfang [44]	Dongbang Giant No.8	HTV	TP & Jackets	2023	Goseong (KR) - Taipei (TW)
Moray West [44]	Dongbang Giant No.3	HTV	TP	2023	Penglai (CN) - Hamriyah (UAE)

When identifying which vessels were involved in the transportation of monopiles, additional projects were found. This showed that not only semi-submersibles and HTVs operated within this market, but also self-gearred vessels and barges were frequently used. Some of these projects are presented in Table D.2.

Table D.2: Operations in the offshore wind market with vessels outside of this research

Project	Type	Cargo	Year	From - To
Inch Cape [45]	SS	Monopiles	2026	Qinzhou (CN) - Port of Leith (UK)
Thor [46]	SS	Monopiles	2025	Rostock (DE) - Denmark
Kaskasi [47]	Barges	Monopiles	2022	Aalborg (DK) - Eemshaven (NL)
Vesterhav [48]	Barges	Monopiles	2023	Rostock (DE) - Port of Thyboron (DK)
Hornsea 2 [49]	Self-gearred	Monopiles	2021	Rostock (DE) - Eemshaven (NL)

To accurately identify an operational profile solely representing operations within this market, it would be relevant to determine which vessels visited known monopile fabrication sites and to focus exclusively on these operations. For this purpose AIS data should not be queried on vessel but on location.

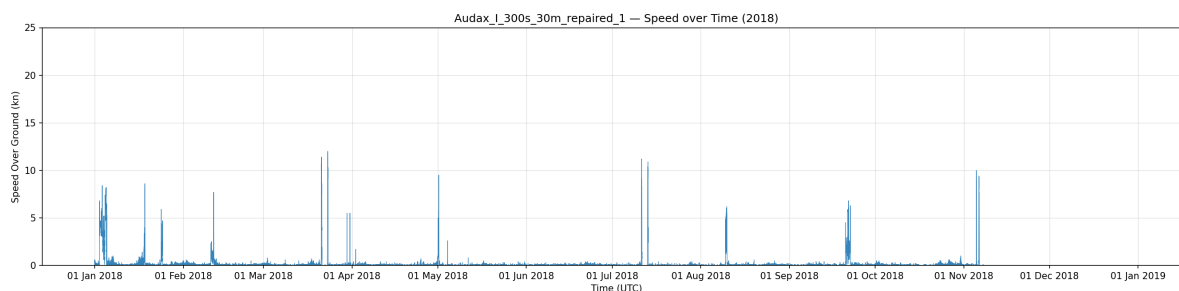


Speed Distribution Plots

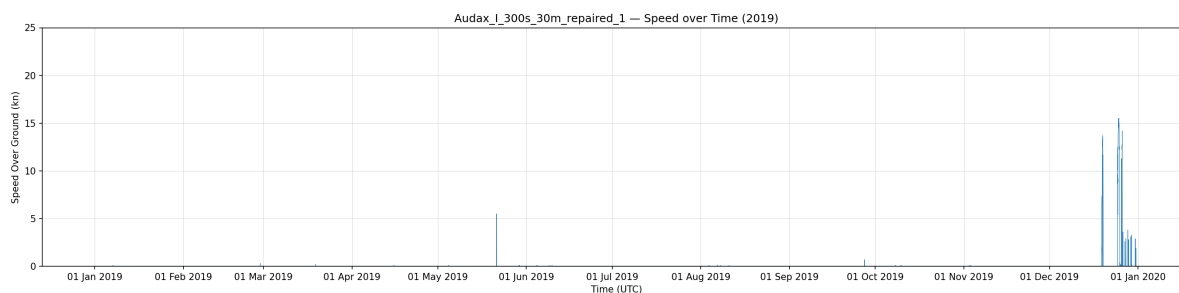
In this appendix, all speed distributions are presented. The speed over ground (SOG) is plotted over time for each vessel and for each year included in this research.

For several years, clear differences in operational behaviour can be observed, as discussed in the main text. These speed distributions provide a clear visual representation of whether a vessel predominantly operates on long-distance trips, short-distance trips, or a combination of both.

E.1. Speed Distribution Audax

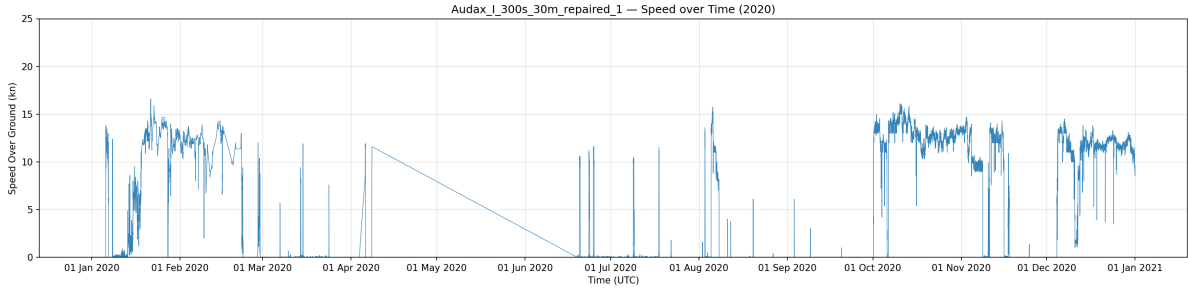


Audax 2018 Speed Distribution

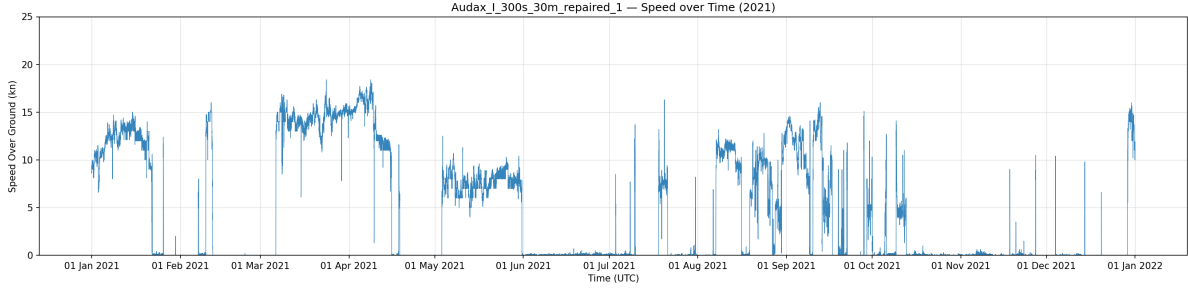


Audax 2019 Speed Distribution

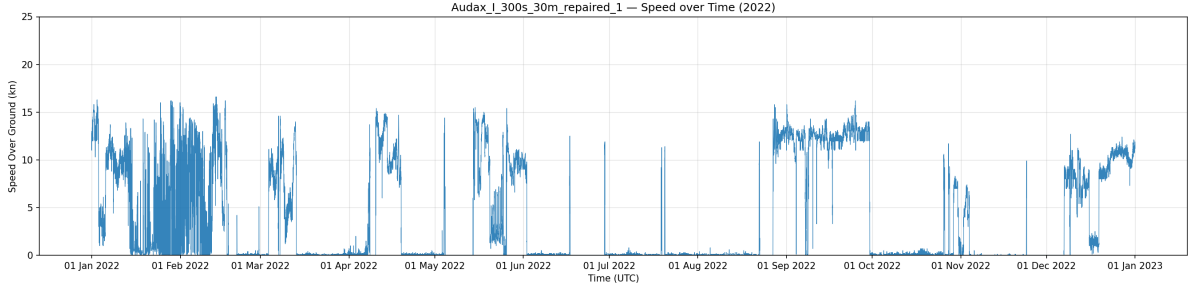
E.1. Speed Distribution Audax



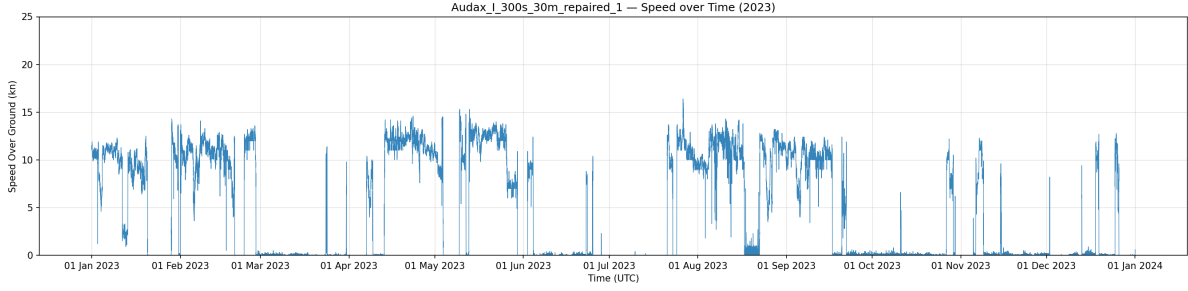
Audax 2020 Speed Distribution



Audax 2021 Speed Distribution

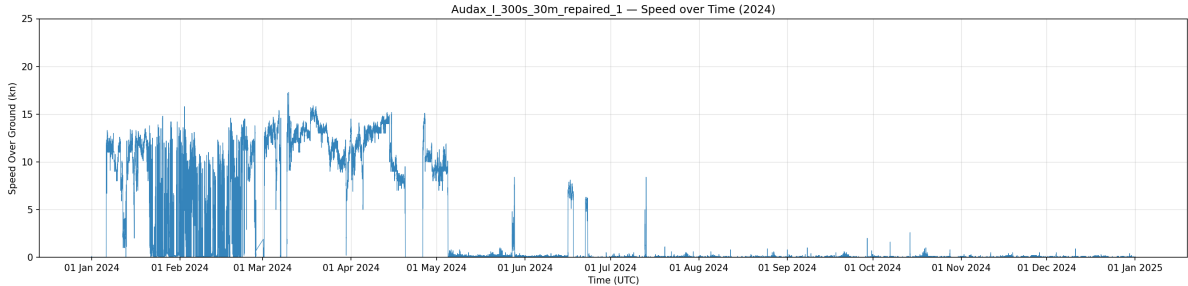


Audax 2022 Speed Distribution



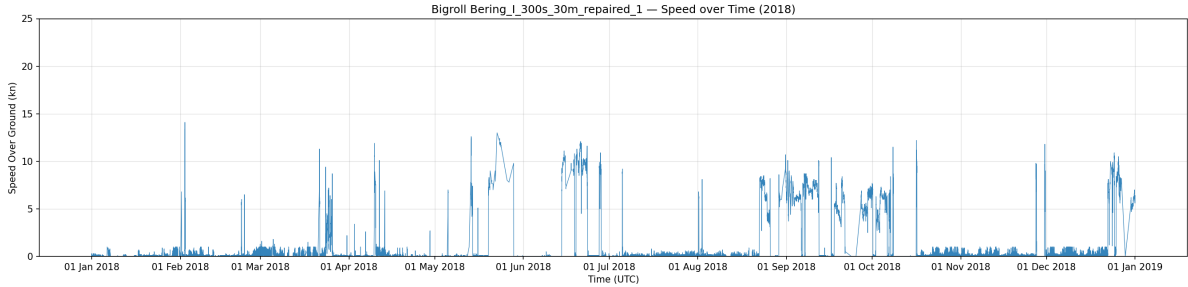
Audax 2023 Speed Distribution

E.1. Speed Distribution Audax

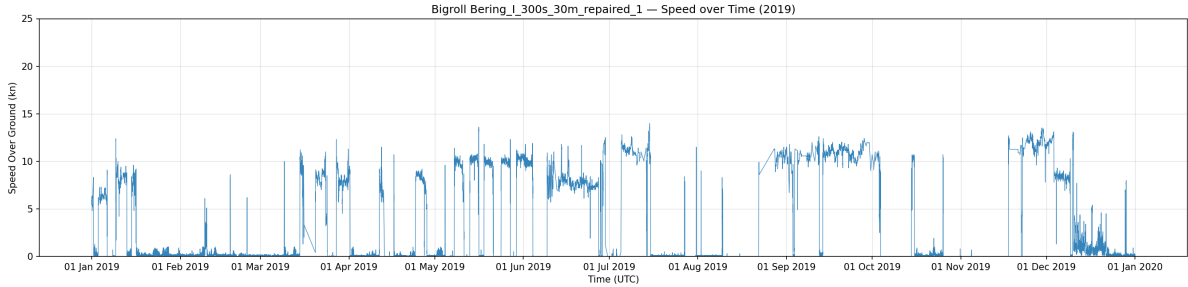


Audax 2024 Speed Distribution

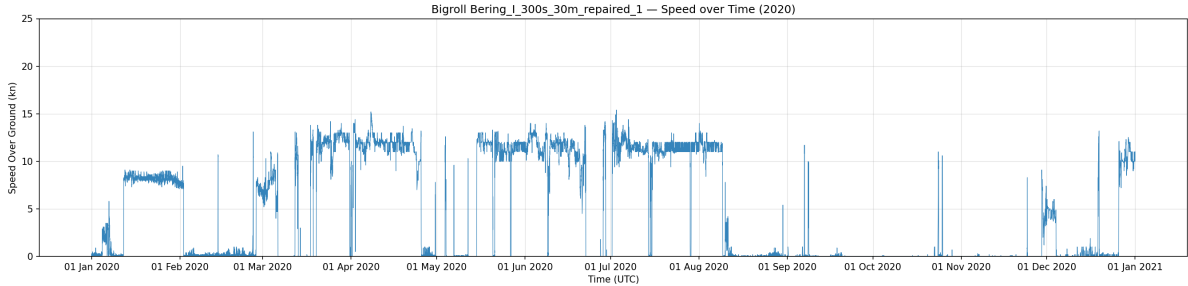
E.2. Speed Distribution Bigroll Bering



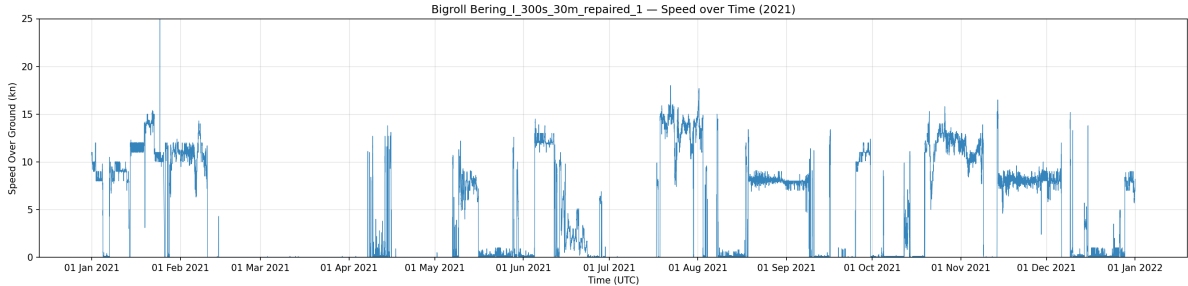
Bigroll Bering 2018 Speed Distribution



Bigroll Bering 2019 Speed Distribution

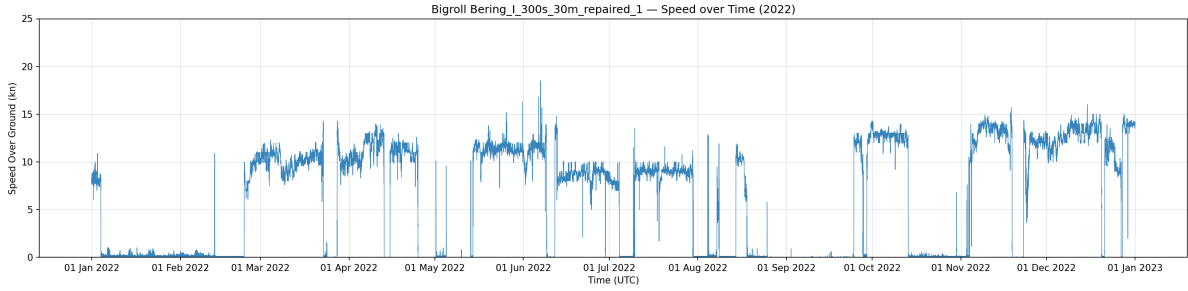


Bigroll Bering 2020 Speed Distribution

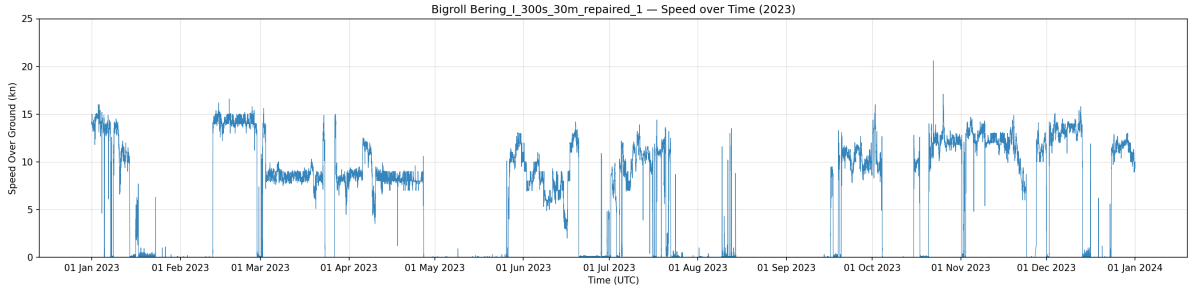


Bigroll Bering 2021 Speed Distribution

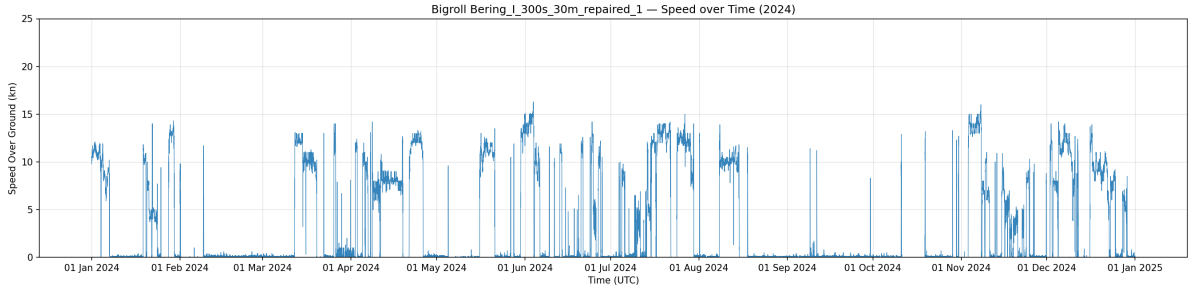
E.2. Speed Distribution Bigroll Bering



Bigroll Bering 2022 Speed Distribution

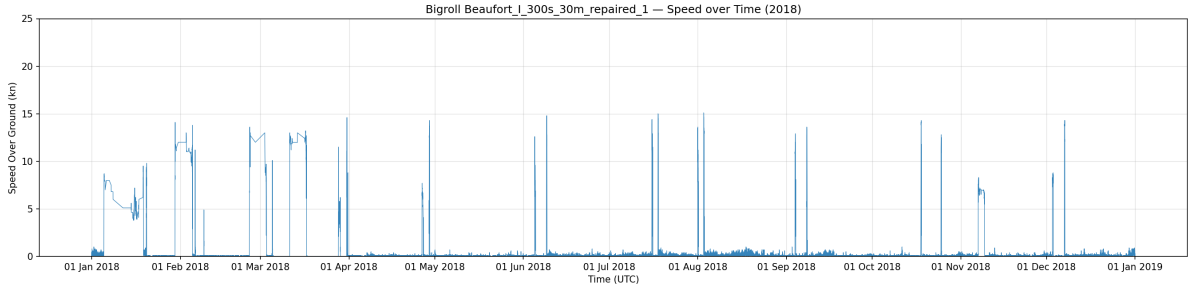


Bigroll Bering 2023 Speed Distribution

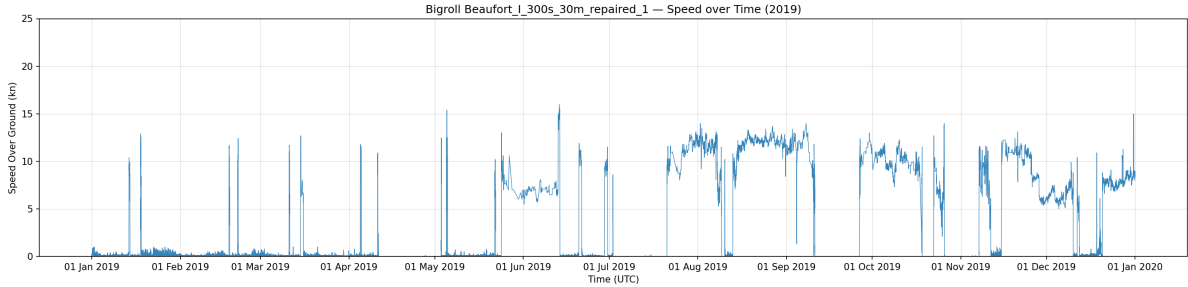


Bigroll Bering 2024 Speed Distribution

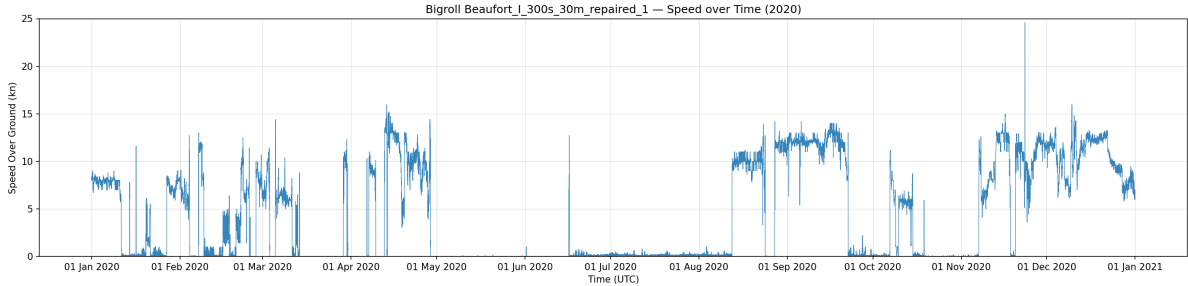
E.3. Speed Distribution Bigroll Beaufort



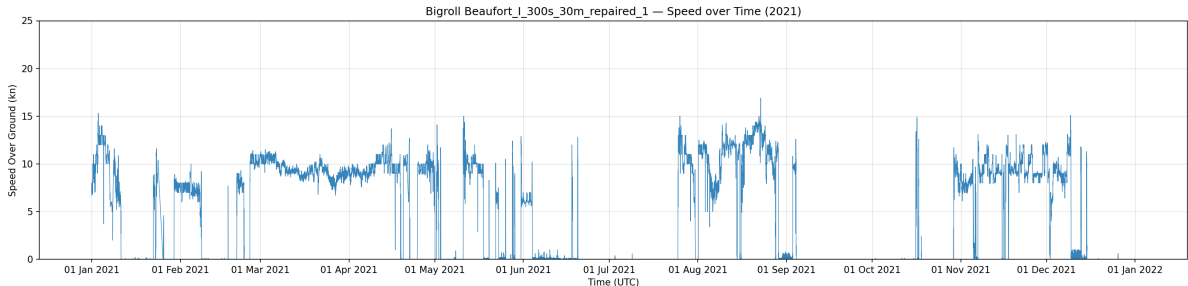
Bigroll Beaufort 2018 Speed Distribution



Bigroll Beaufort 2019 Speed Distribution

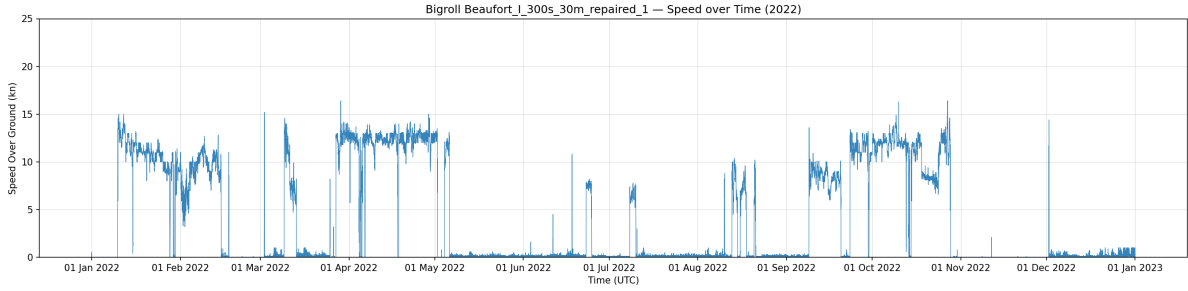


Bigroll Beaufort 2020 Speed Distribution

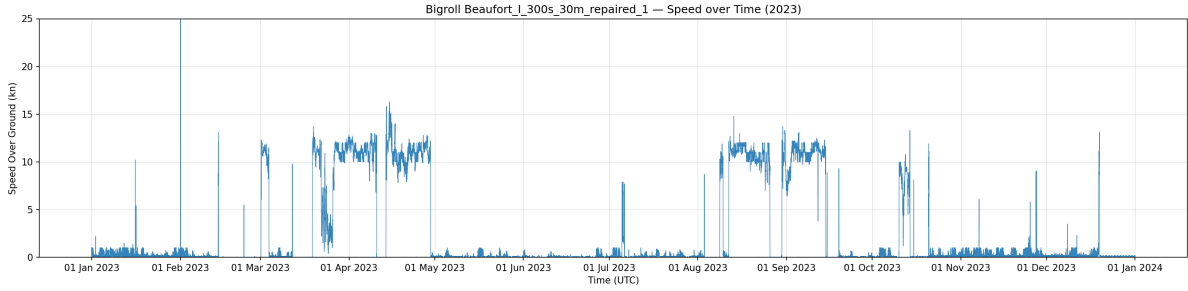


Bigroll Beaufort 2021 Speed Distribution

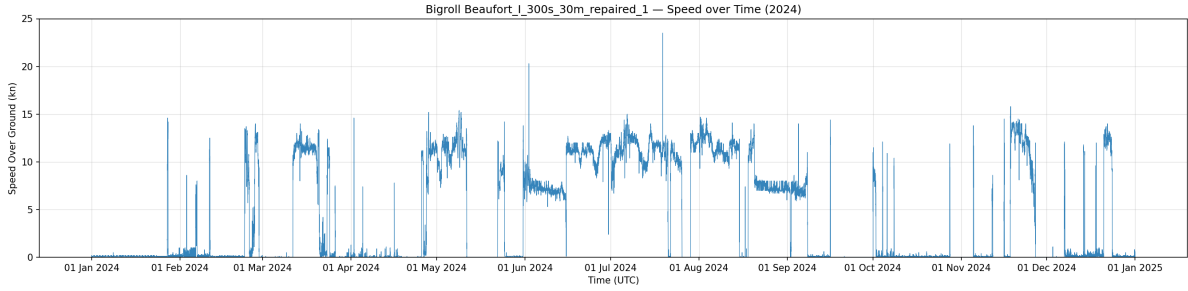
E.3. Speed Distribution Bigroll Beaufort



Bigroll Beaufort 2022 Speed Distribution

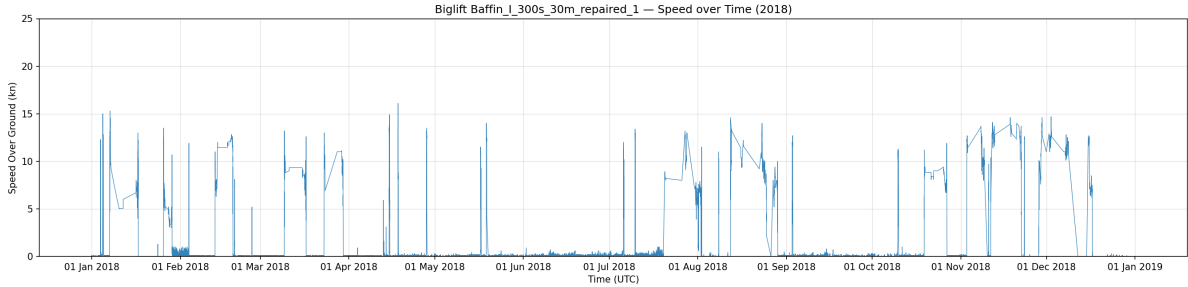


Bigroll Beaufort 2023 Speed Distribution

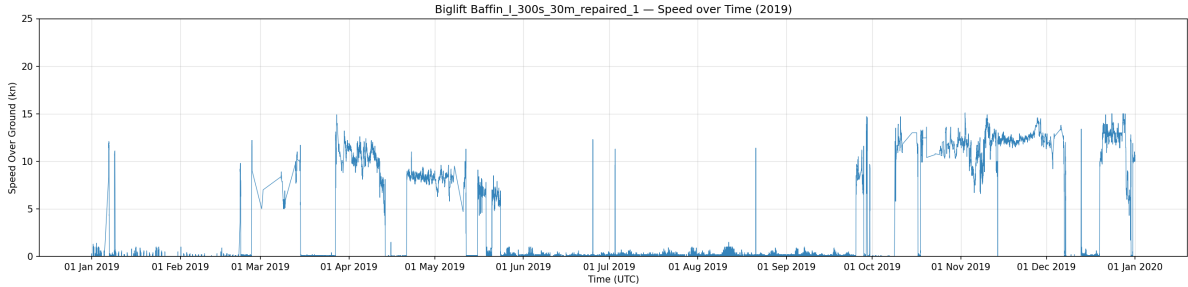


Bigroll Beaufort 2024 Speed Distribution

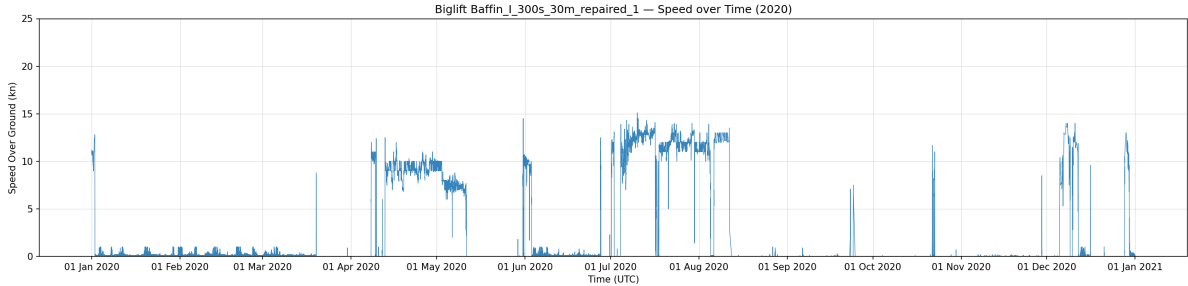
E.4. Speed Distribution Biglift Baffin



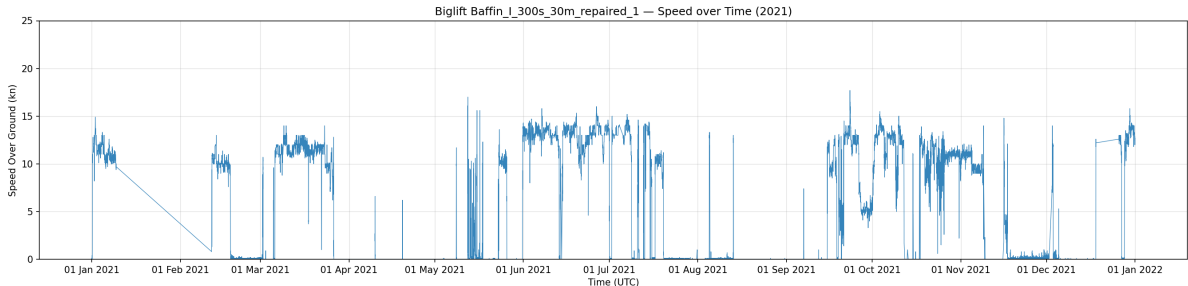
Biglift Baffin 2018 Speed Distribution



Biglift Baffin 2019 Speed Distribution

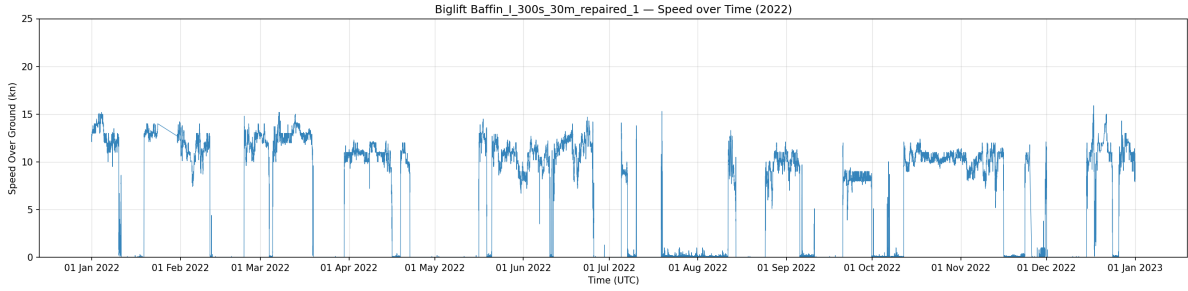


Biglift Baffin 2020 Speed Distribution

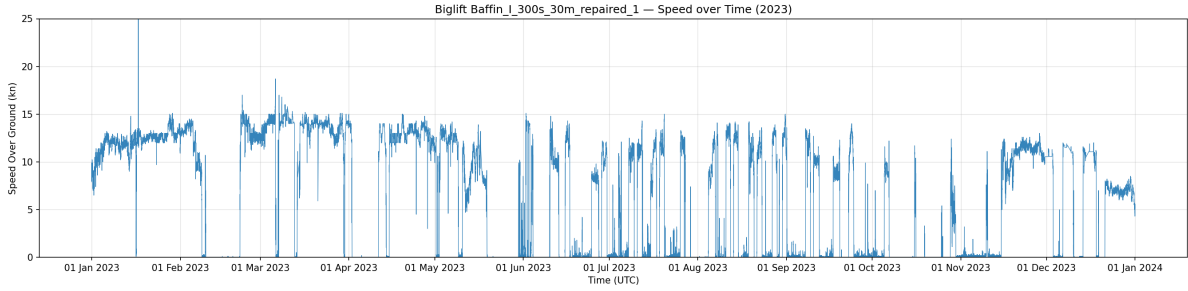


Biglift Baffin 2021 Speed Distribution

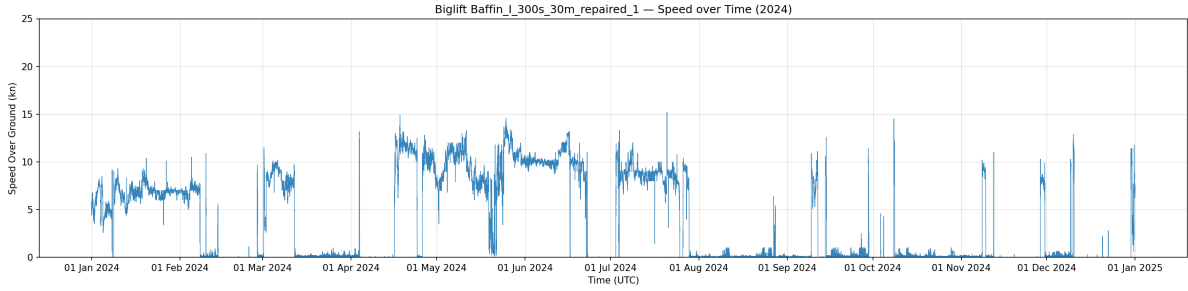
E.4. Speed Distribution Biglift Baffin



Biglift Baffin 2022 Speed Distribution

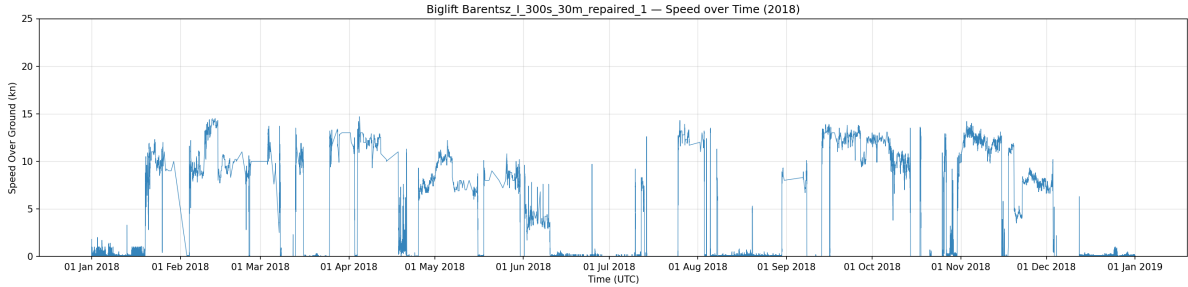


Biglift Baffin 2023 Speed Distribution

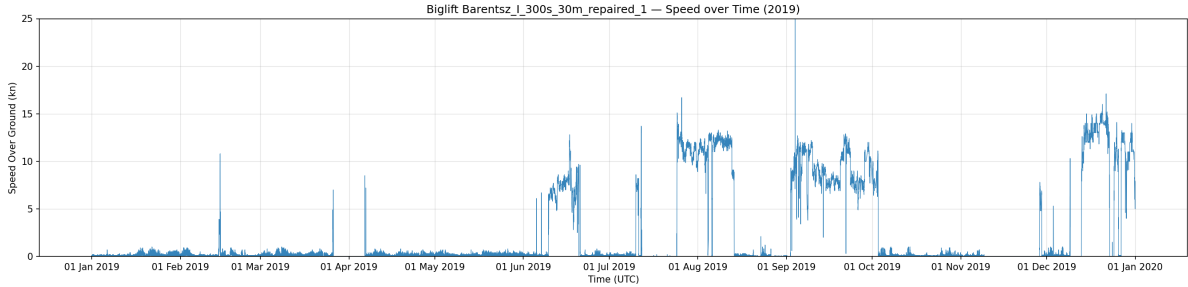


Biglift Baffin 2024 Speed Distribution

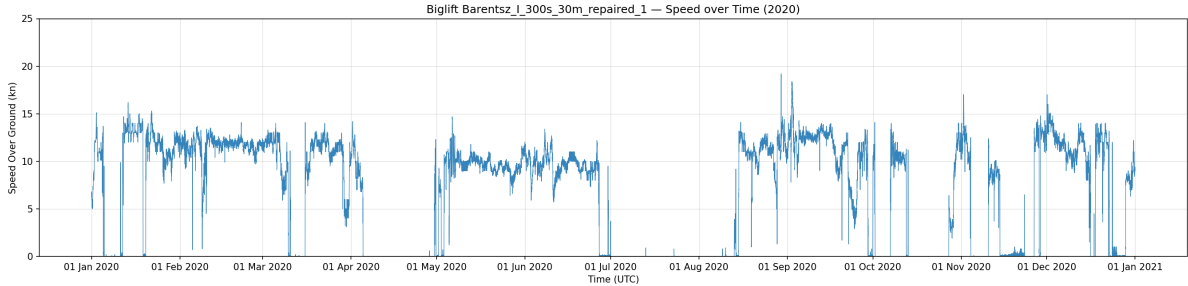
E.5. Speed Distribution Biglift Barentsz



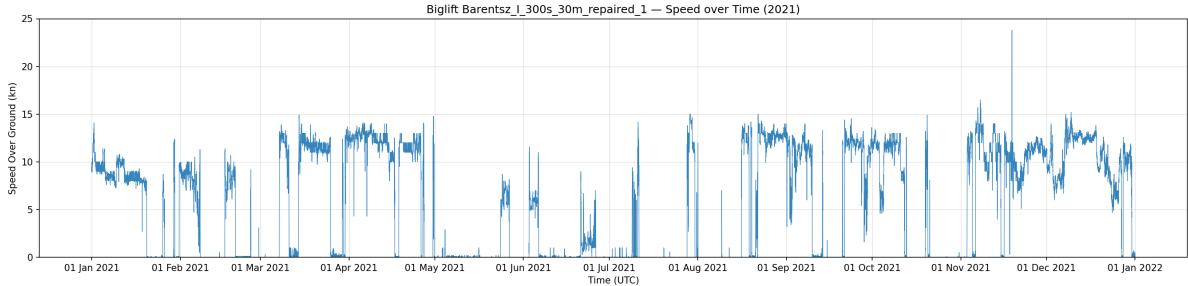
Biglift Barentsz 2018 Speed Distribution



Biglift Barentsz 2019 Speed Distribution

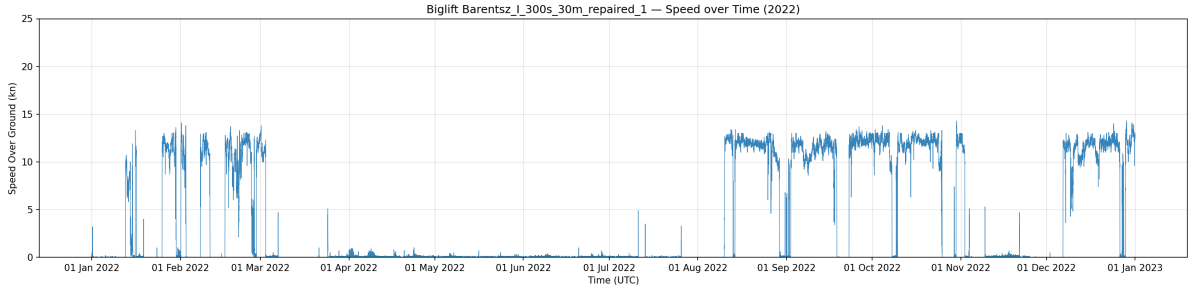


Biglift Barentsz 2020 Speed Distribution

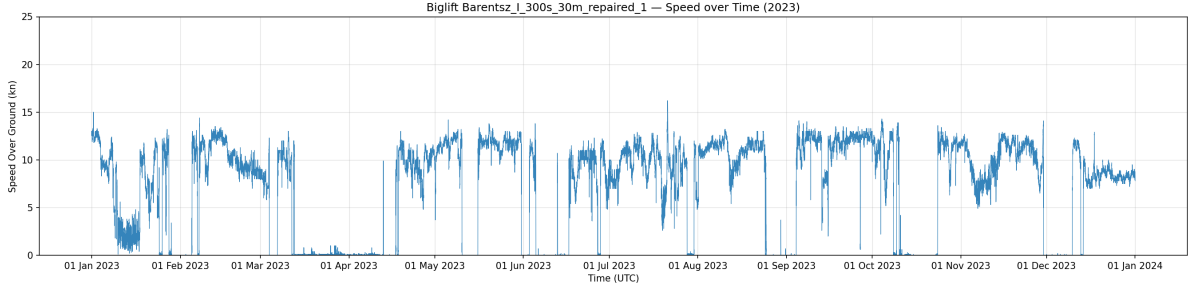


Biglift Barentsz 2021 Speed Distribution

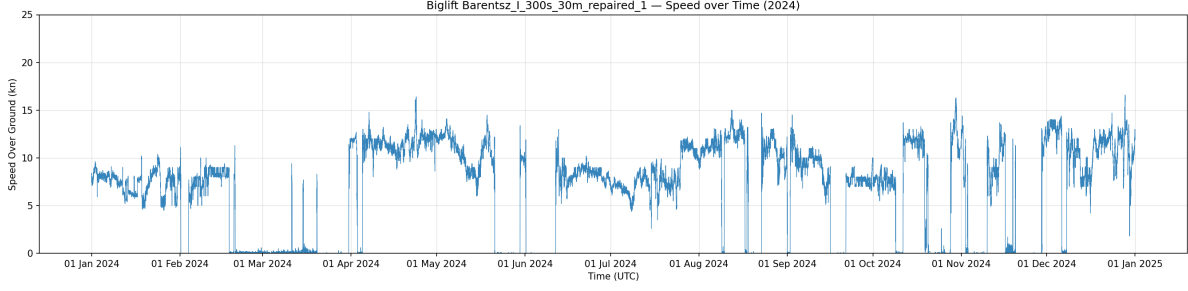
E.5. Speed Distribution Biglift Barentsz



Biglift Barentsz 2022 Speed Distribution

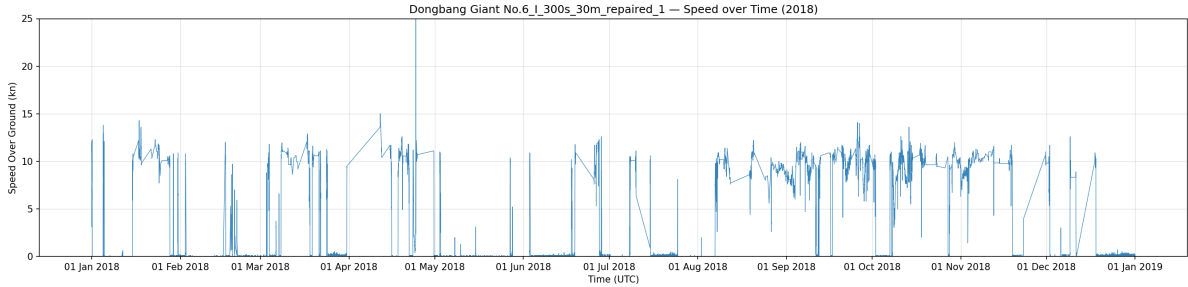


Biglift Barentsz 2023 Speed Distribution

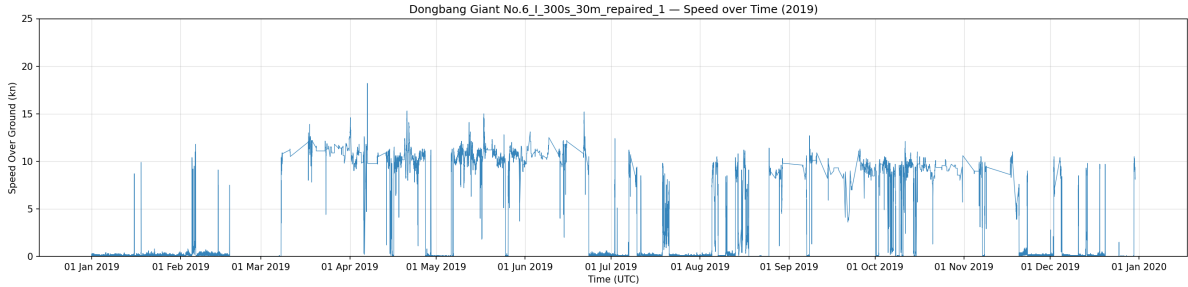


Biglift Barentsz 2024 Speed Distribution

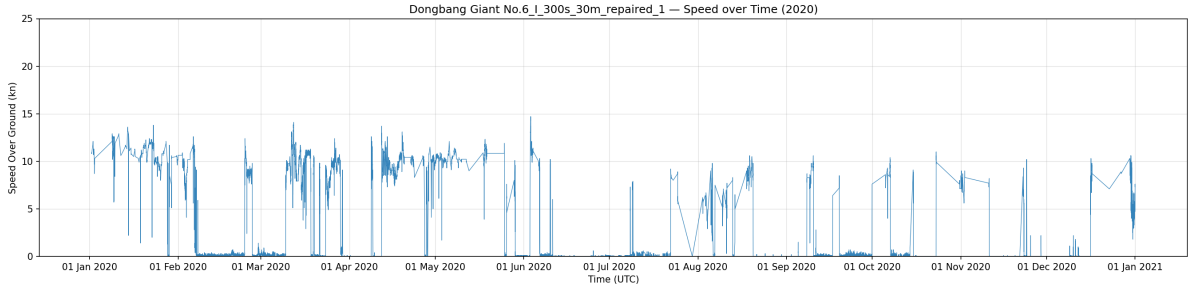
E.6. Speed Distribution Dongbang Giant No.6



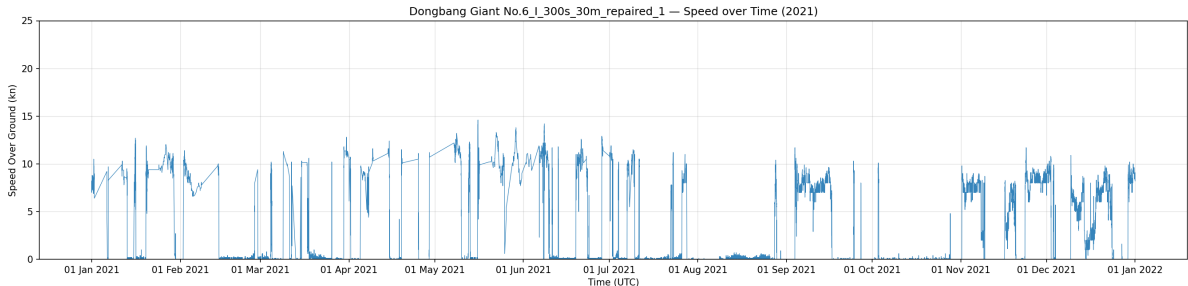
Dongbang Giant No.6 2018 Speed Distribution



Dongbang Giant No.6 2019 Speed Distribution

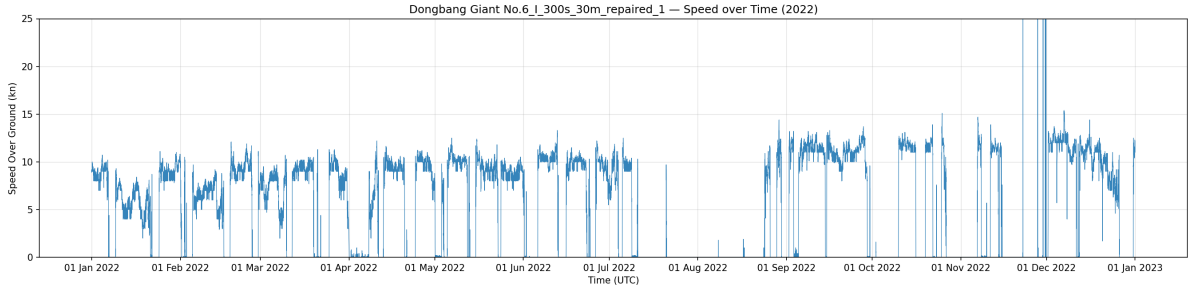


Dongbang Giant No.6 2020 Speed Distribution

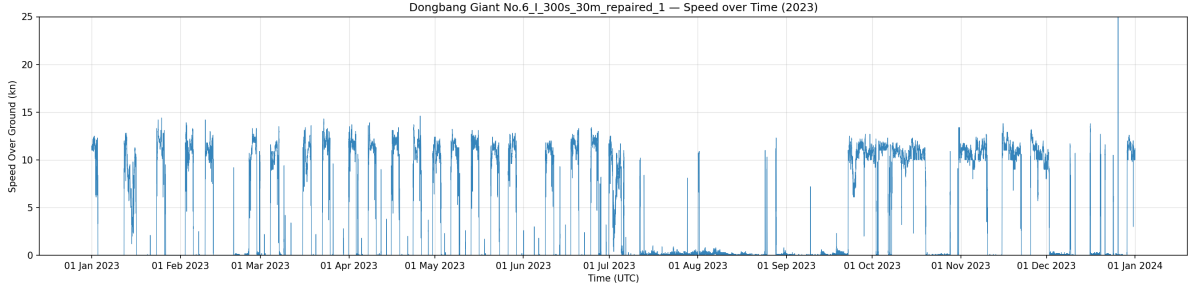


Dongbang Giant No.6 2021 Speed Distribution

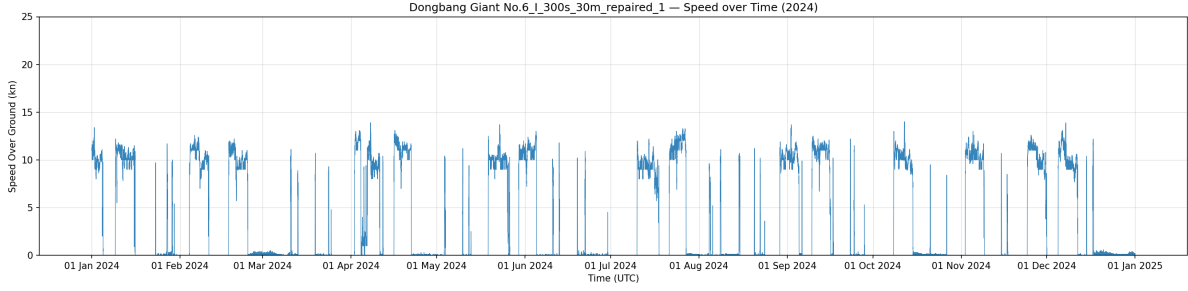
E.6. Speed Distribution Dongbang Giant No.6



Dongbang Giant No.6 2022 Speed Distribution

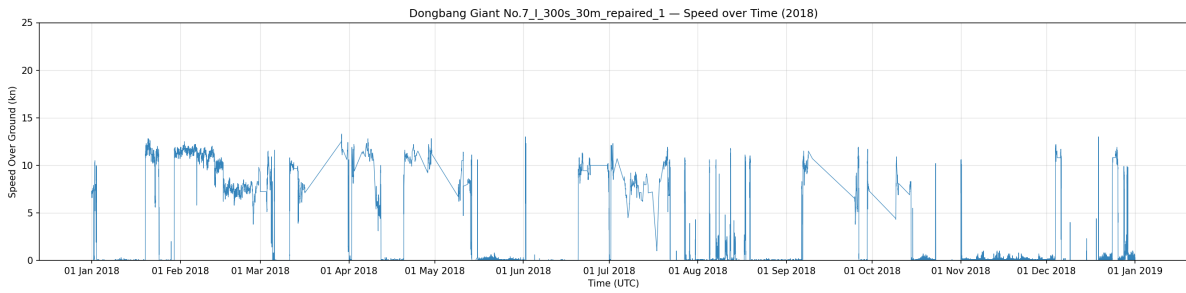


Dongbang Giant No.6 2023 Speed Distribution

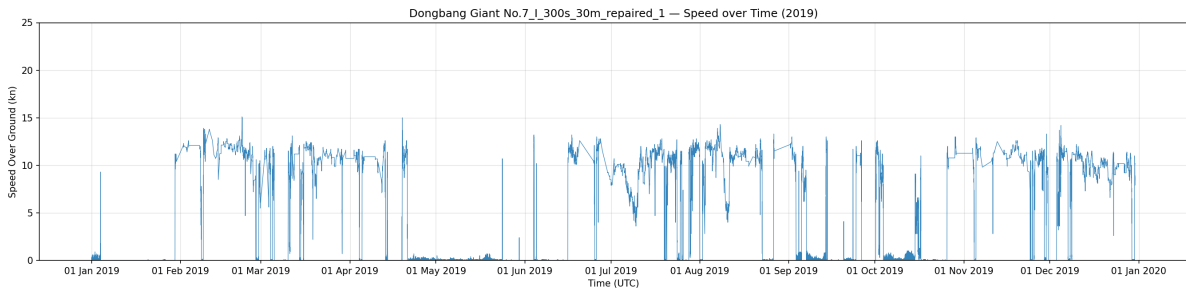


Dongbang Giant No.6 2024 Speed Distribution

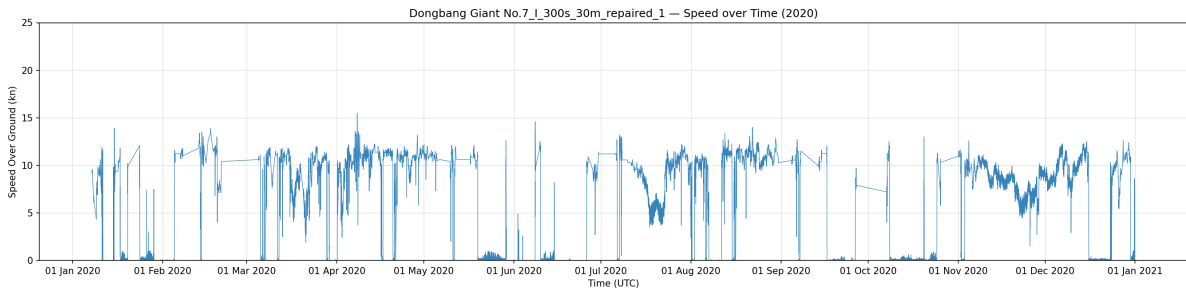
E.7. Speed Distribution Dongbang Giant No.7



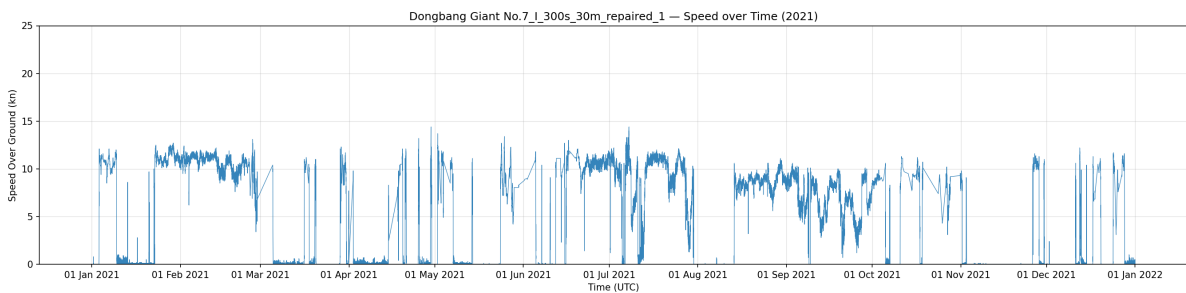
Dongbang Giant No.7 2018 Speed Distribution



Dongbang Giant No.7 2019 Speed Distribution

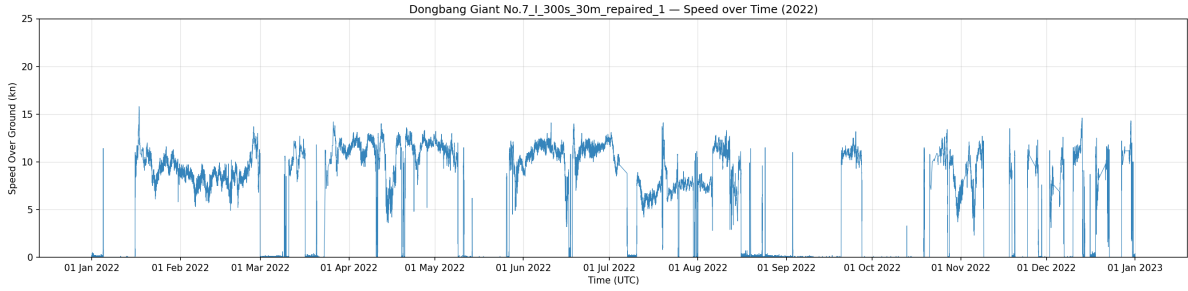


Dongbang Giant No.7 2020 Speed Distribution

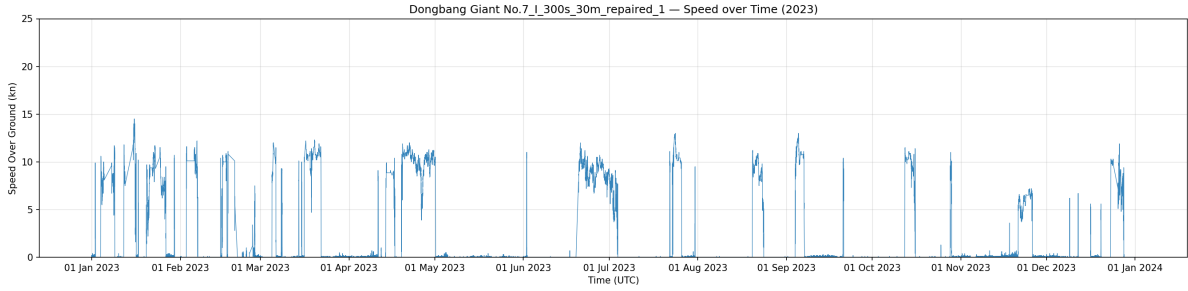


Dongbang Giant No.7 2021 Speed Distribution

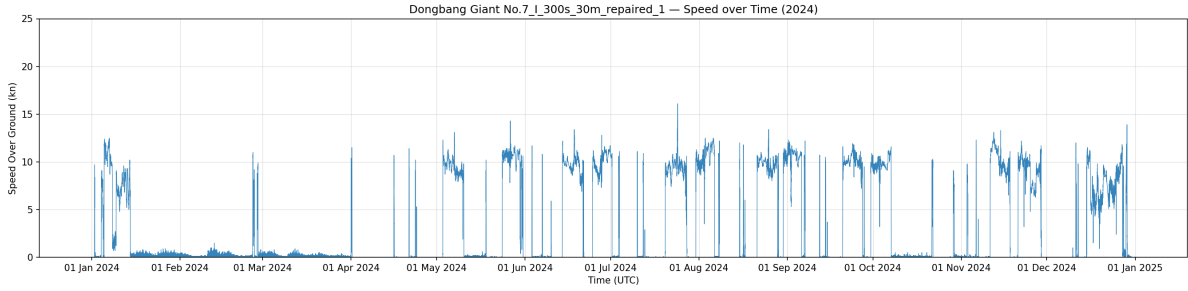
E.7. Speed Distribution Dongbang Giant No.7



Dongbang Giant No.7 2022 Speed Distribution

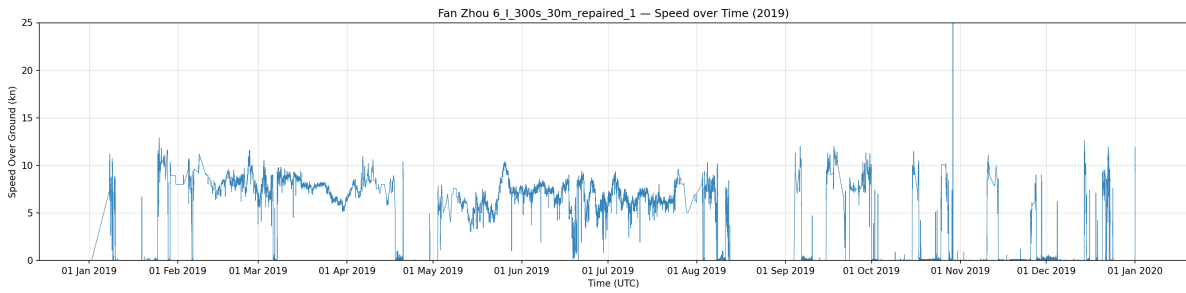


Dongbang Giant No.7 2023 Speed Distribution

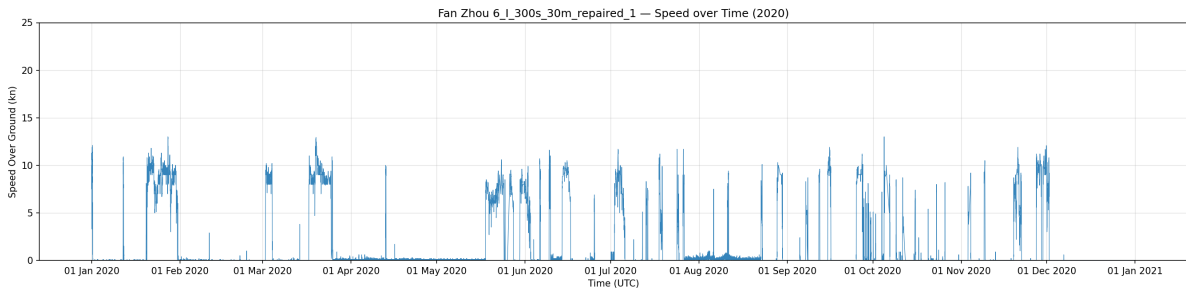


Dongbang Giant No.7 2024 Speed Distribution

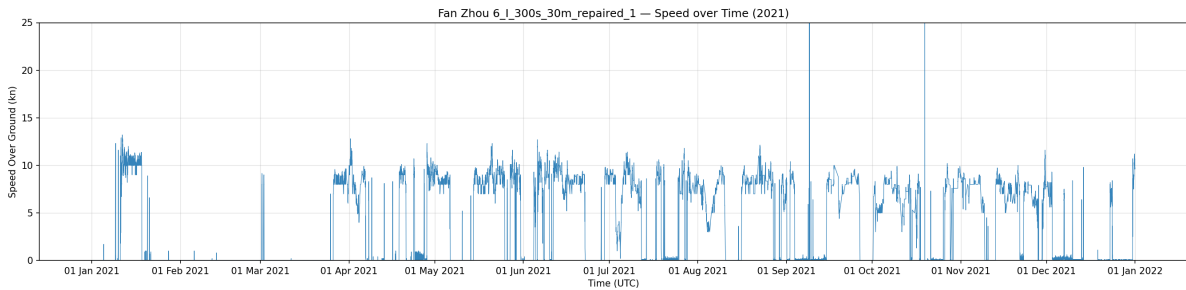
E.8. Speed Distribution Fan Zhou 6



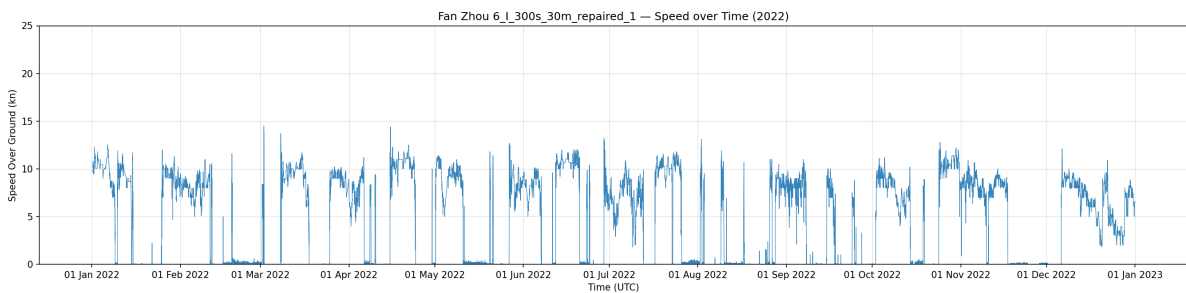
Fan Zhou 6 2019 Speed Distribution



Fan Zhou 6 2020 Speed Distribution

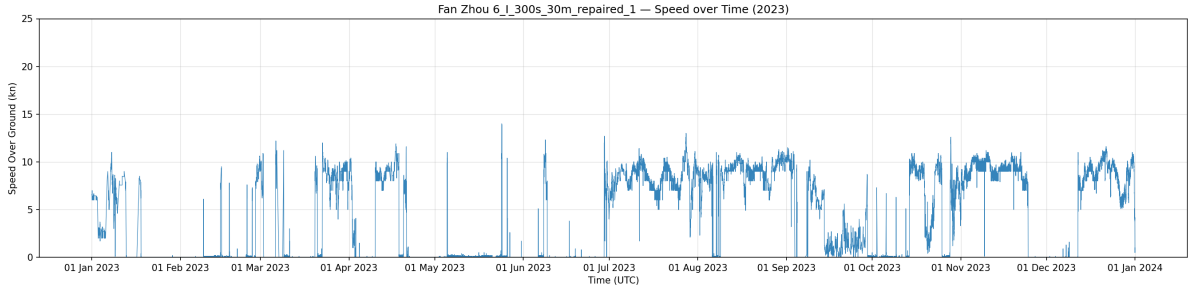


Fan Zhou 6 2021 Speed Distribution

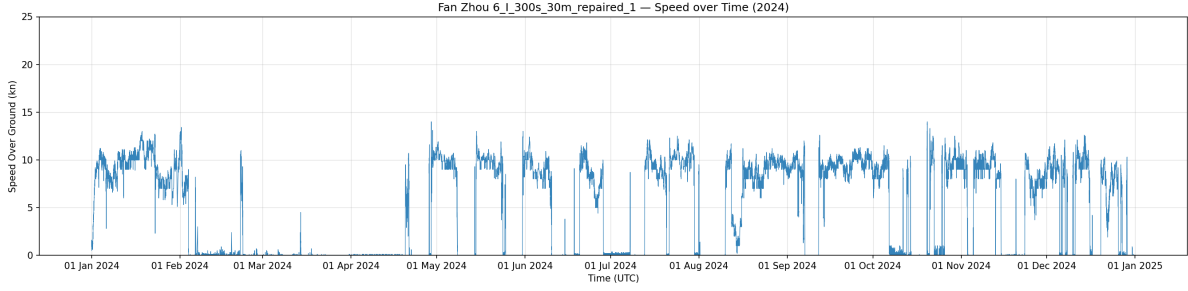


Fan Zhou 6 2022 Speed Distribution

E.8. Speed Distribution Fan Zhou 6

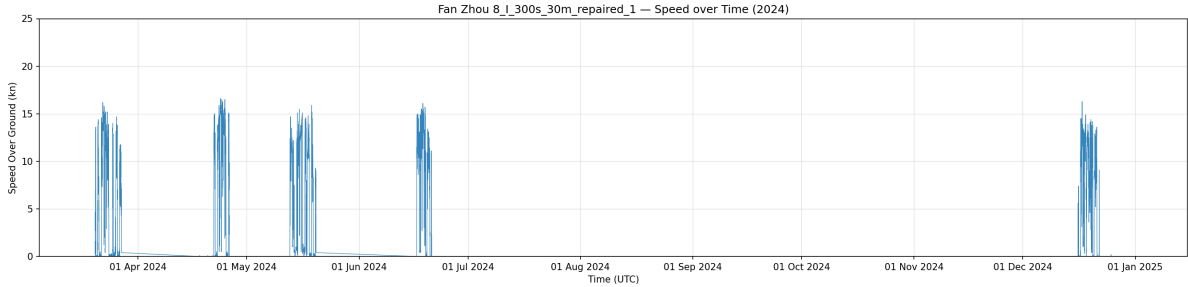


Fan Zhou 6 2023 Speed Distribution

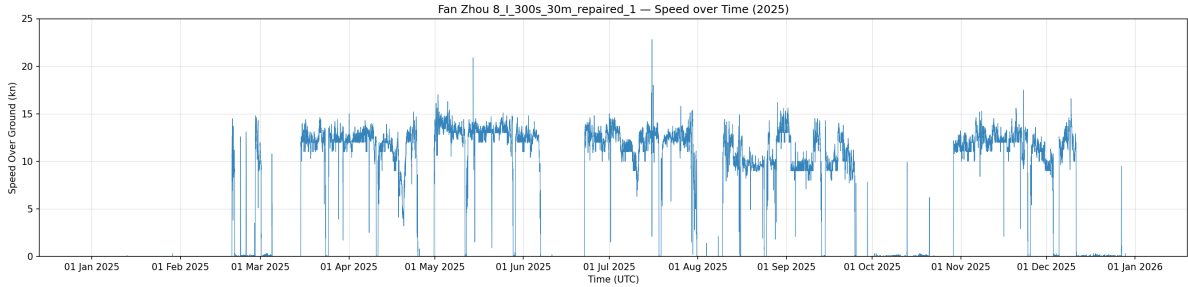


Fan Zhou 6 2024 Speed Distribution

E.9. Speed Distribution Fan Zhou 8

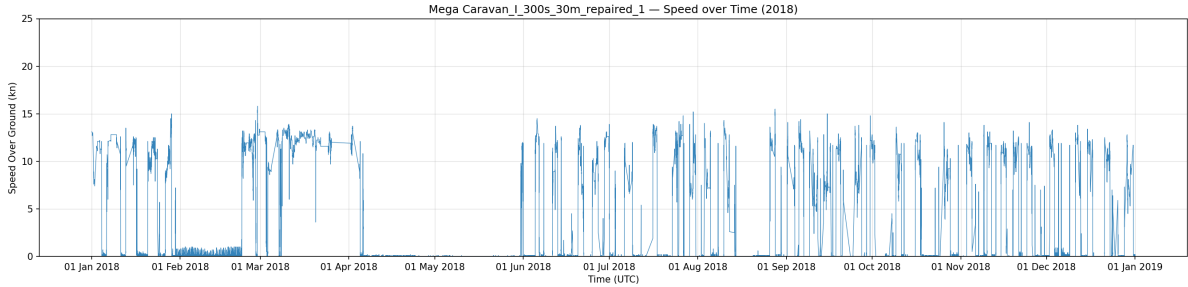


Fan Zhou 8 2024 Speed Distribution

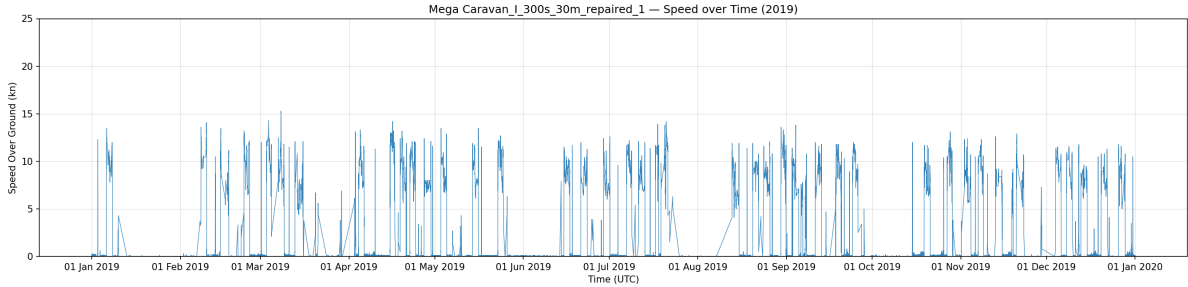


Fan Zhou 8 2025 Speed Distribution

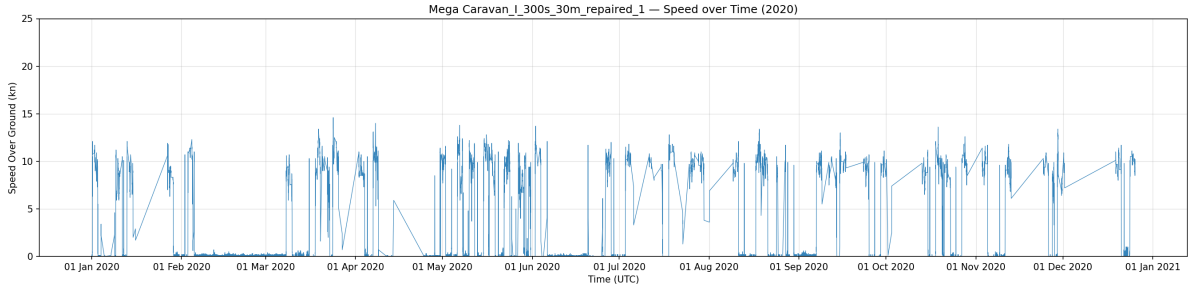
E.10. Speed Distribution Mega Caravan



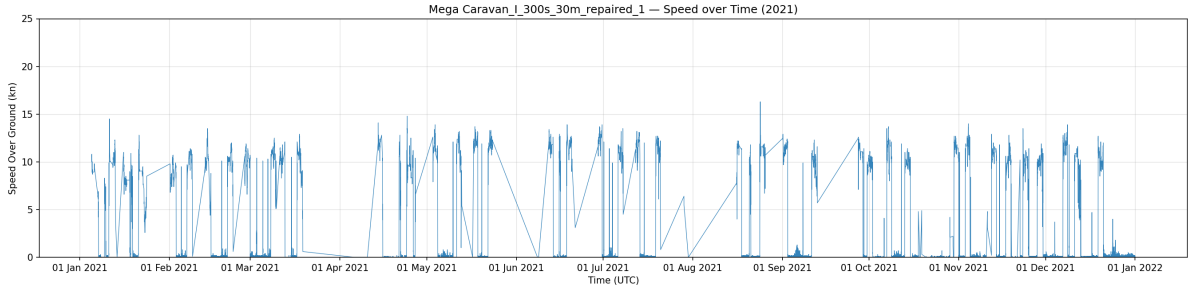
Mega Caravan 2018 Speed Distribution



Mega Caravan 2019 Speed Distribution

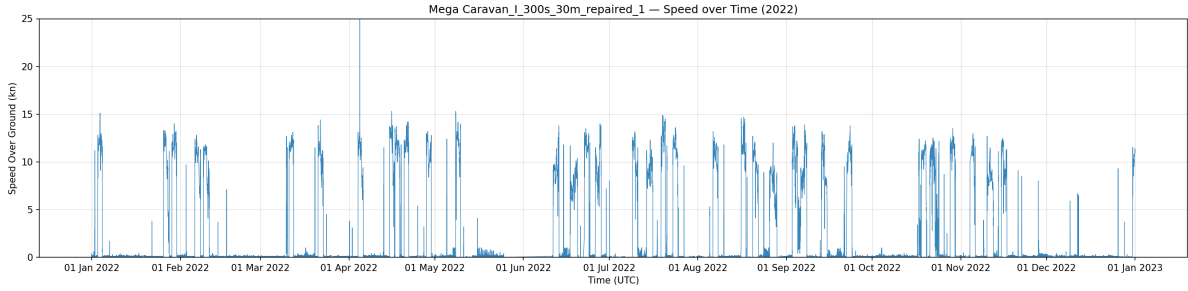


Mega Caravan 2020 Speed Distribution

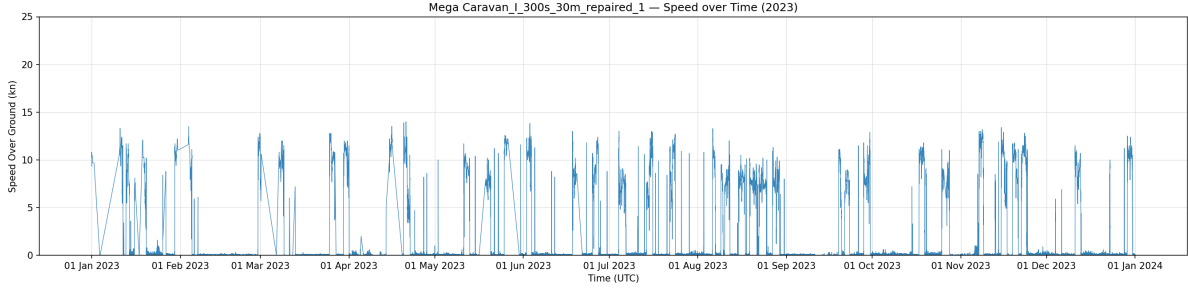


Mega Caravan 2021 Speed Distribution

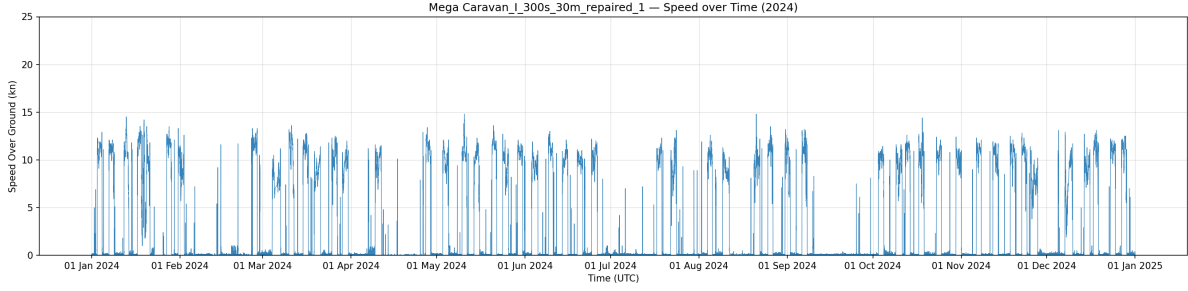
E.10. Speed Distribution Mega Caravan



Mega Caravan 2022 Speed Distribution

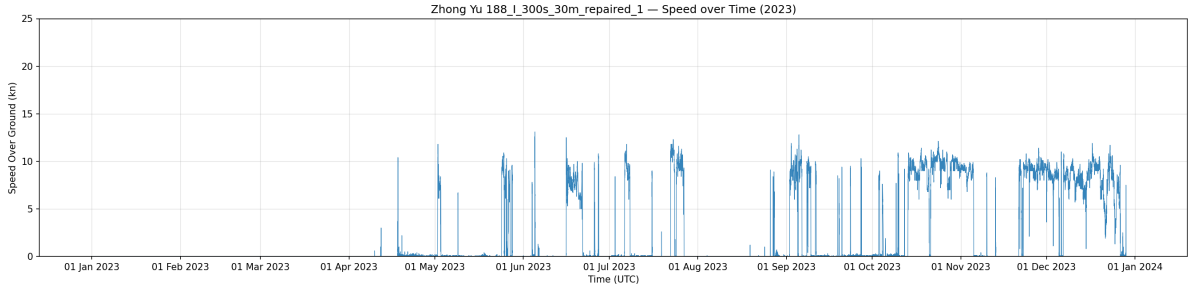


Mega Caravan 2023 Speed Distribution

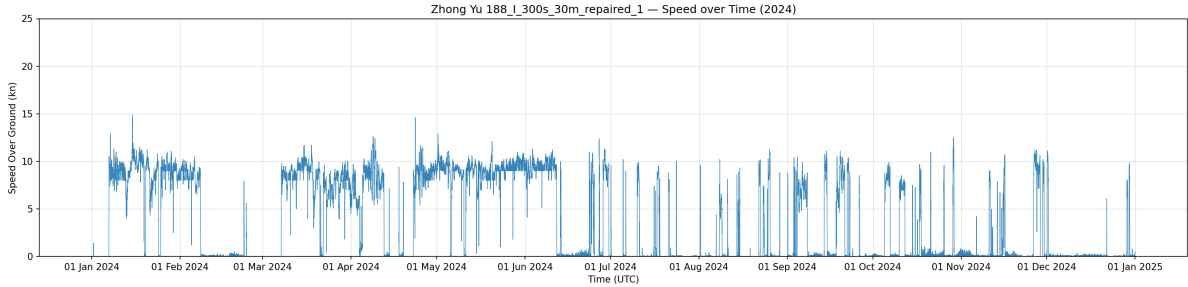


Mega Caravan 2024 Speed Distribution

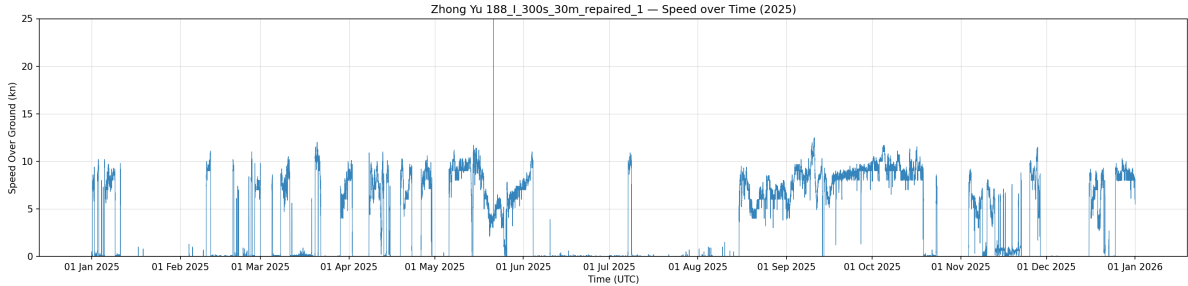
E.11. Speed Distribution Zhong Yu 188



Zhong Yu 188 2023 Speed Distribution

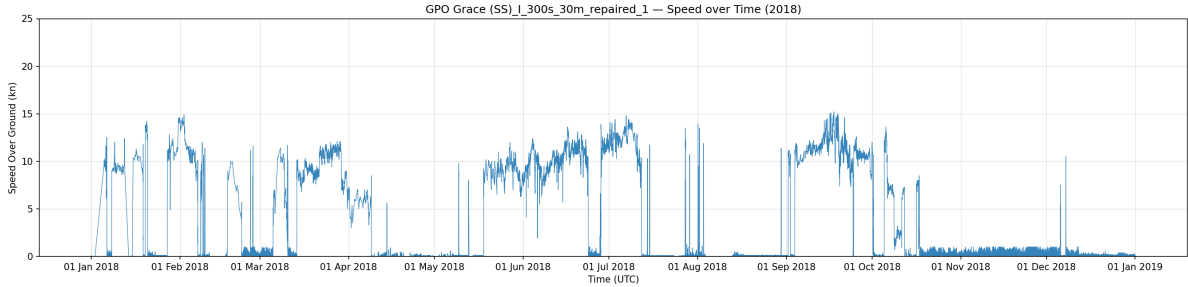


Zhong Yu 188 2024 Speed Distribution

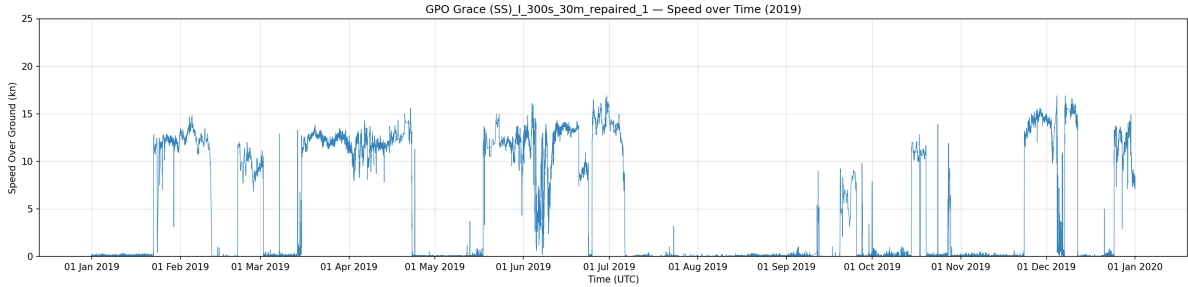


Zhong Yu 188 2025 Speed Distribution

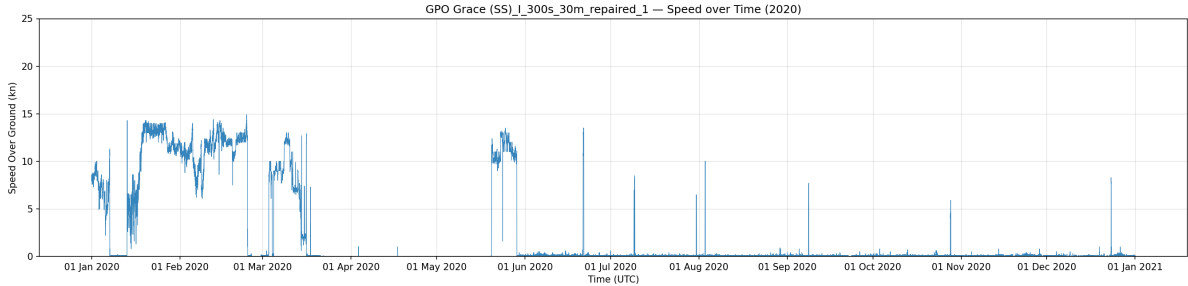
E.12. Speed Distribution GPO Grace



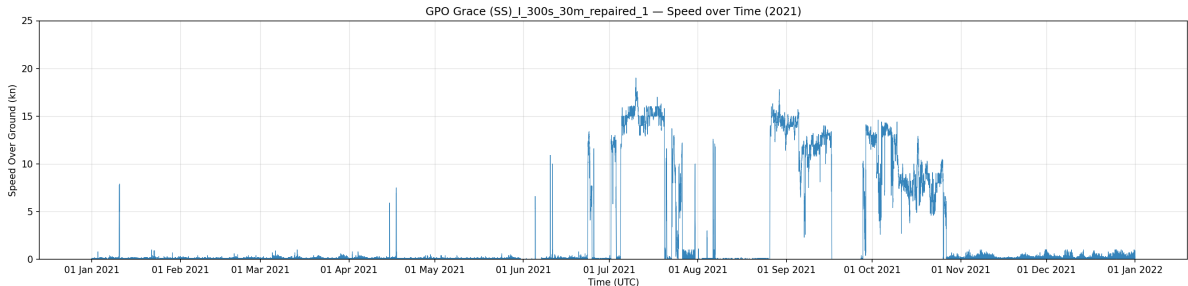
GPO Grace 2018 Speed Distribution



GPO Grace 2019 Speed Distribution

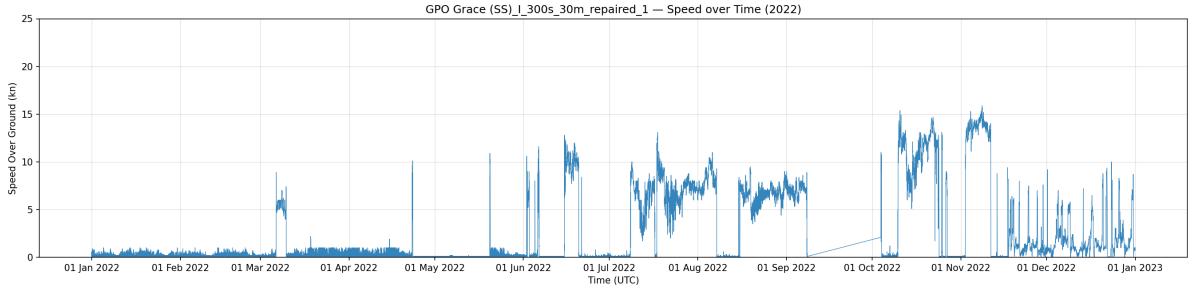


GPO Grace 2020 Speed Distribution

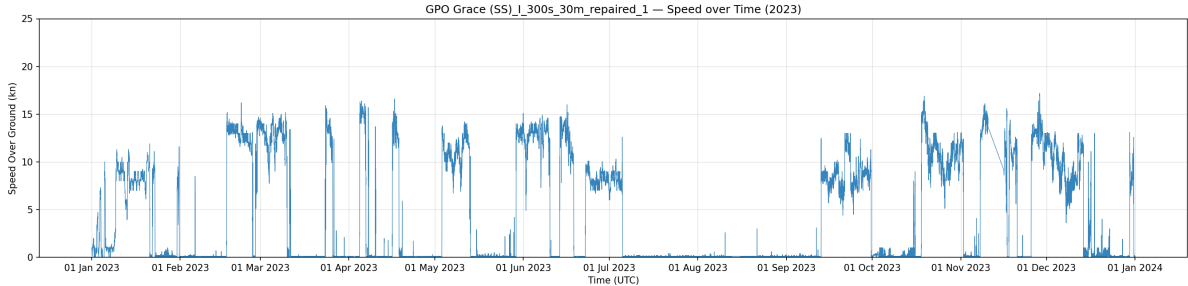


GPO Grace 2021 Speed Distribution

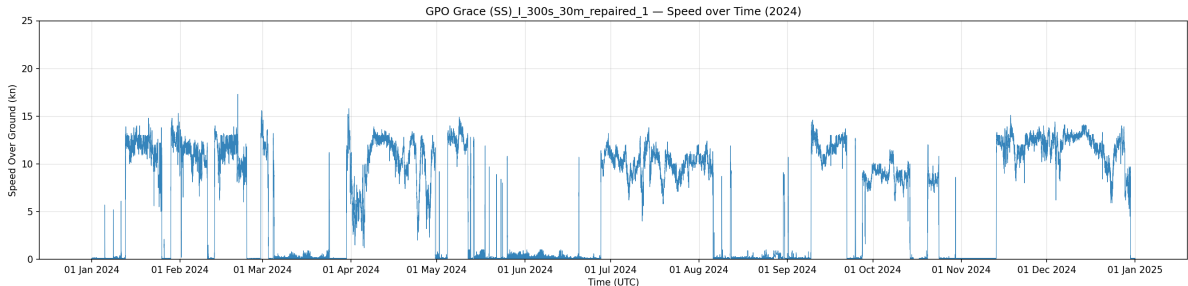
E.12. Speed Distribution GPO Grace



GPO Grace 2022 Speed Distribution

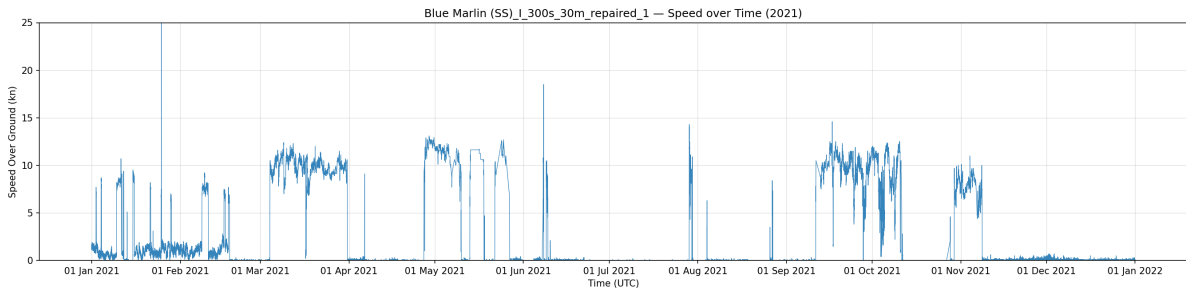


GPO Grace 2023 Speed Distribution

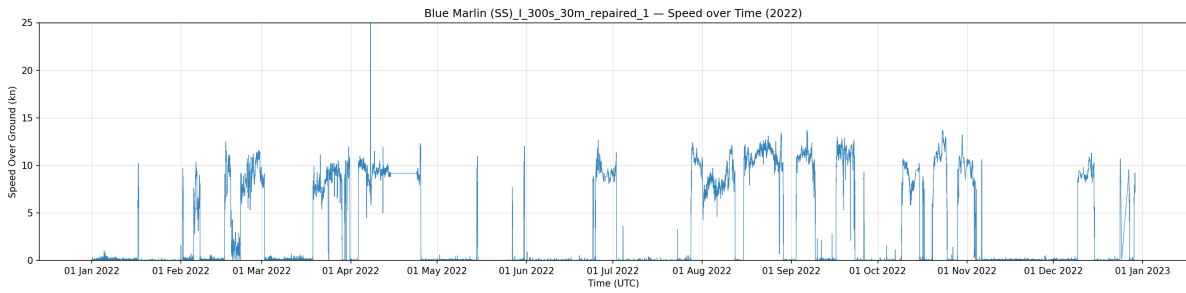


GPO Grace 2024 Speed Distribution

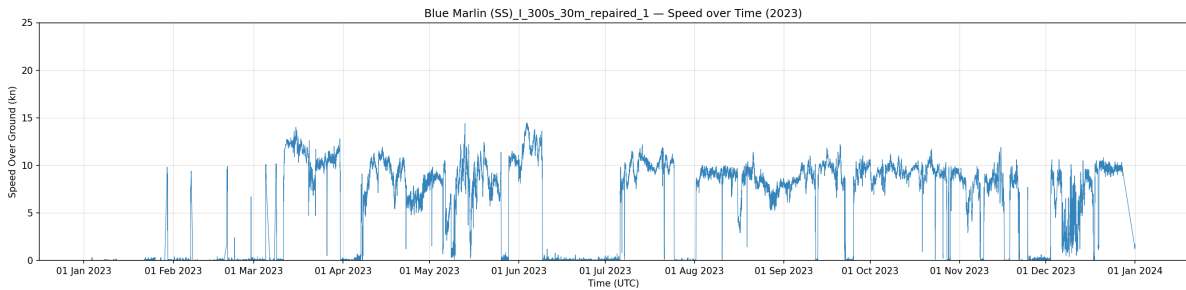
E.13. Speed Distribution Blue Marlin



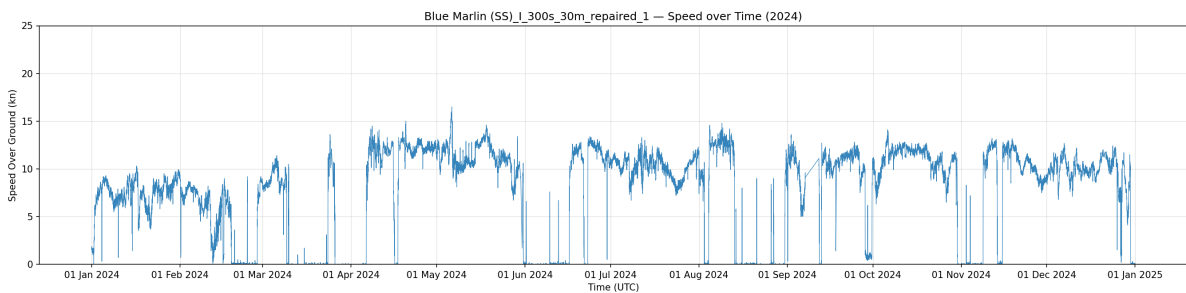
Blue Marlin 2021 Speed Distribution



Blue Marlin 2022 Speed Distribution

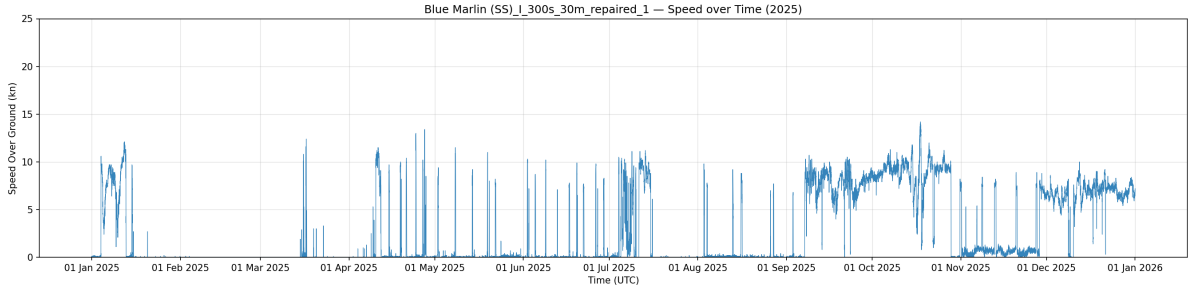


Blue Marlin 2023 Speed Distribution



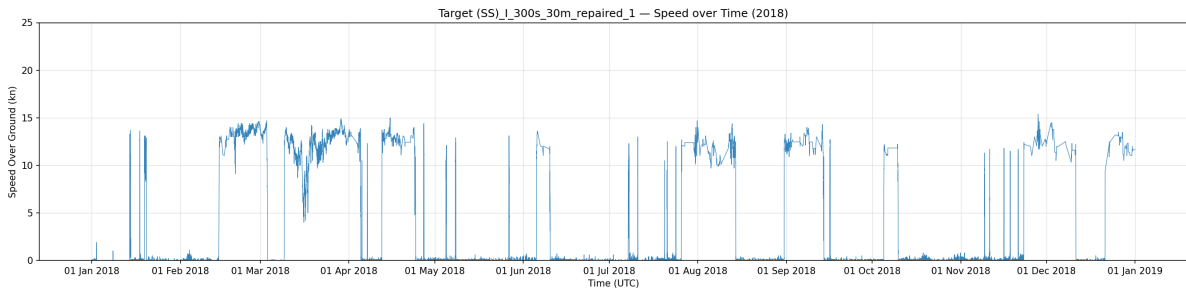
Blue Marlin 2024 Speed Distribution

E.13. Speed Distribution Blue Marlin

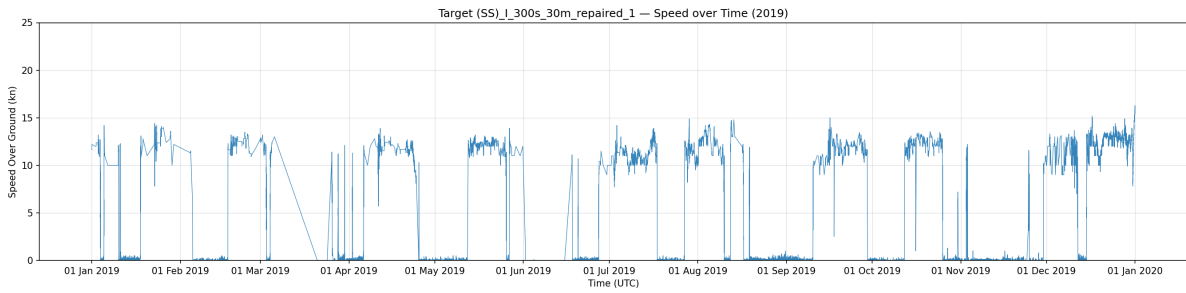


Blue Marlin 2025 Speed Distribution

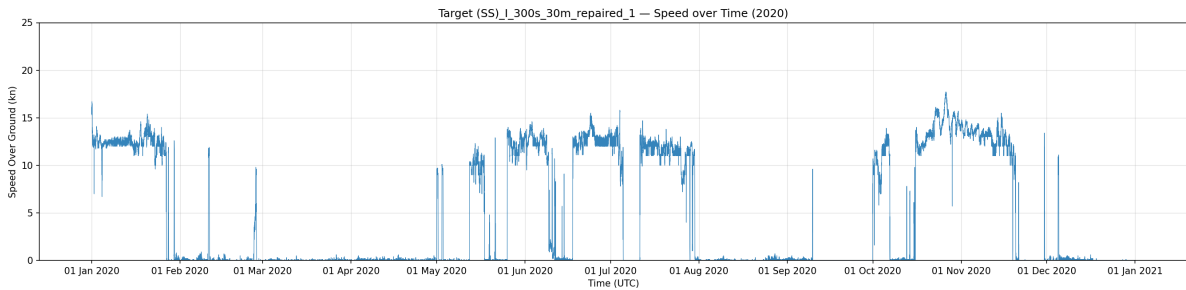
E.14. Speed Distribution Target



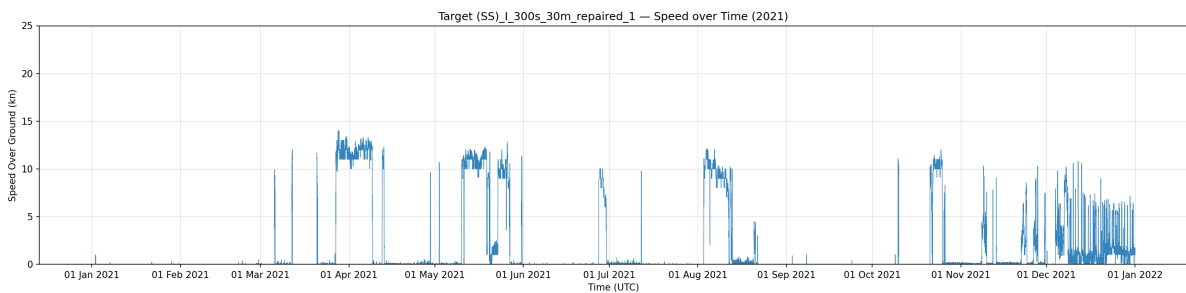
Target 2018 Speed Distribution



Target 2019 Speed Distribution

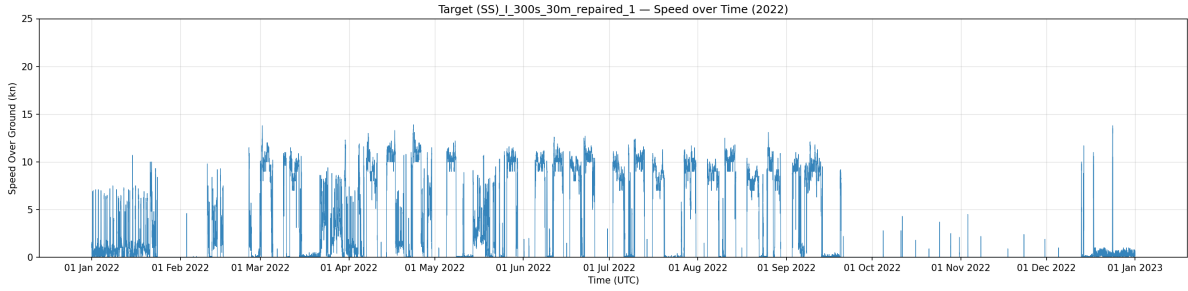


Target 2020 Speed Distribution

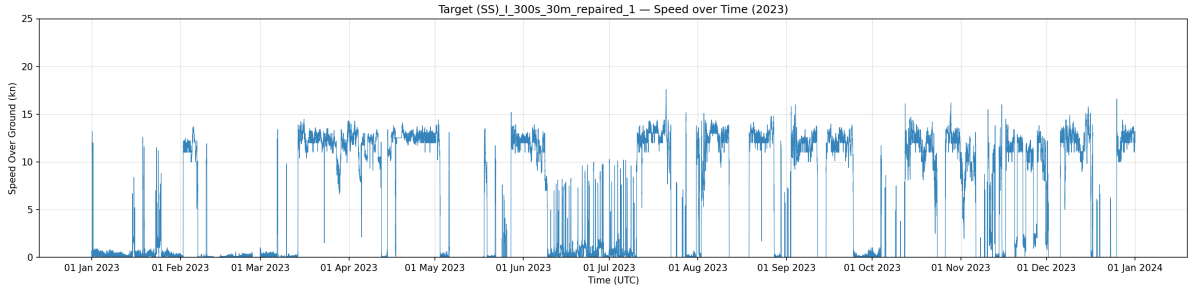


Target 2021 Speed Distribution

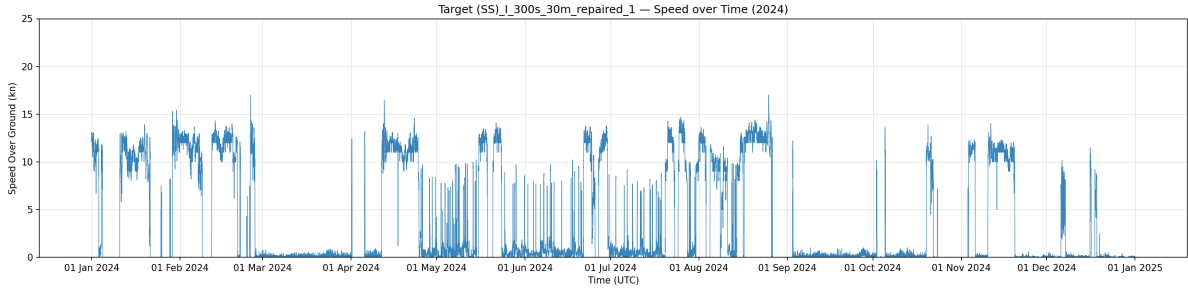
E.14. Speed Distribution Target



Target 2022 Speed Distribution

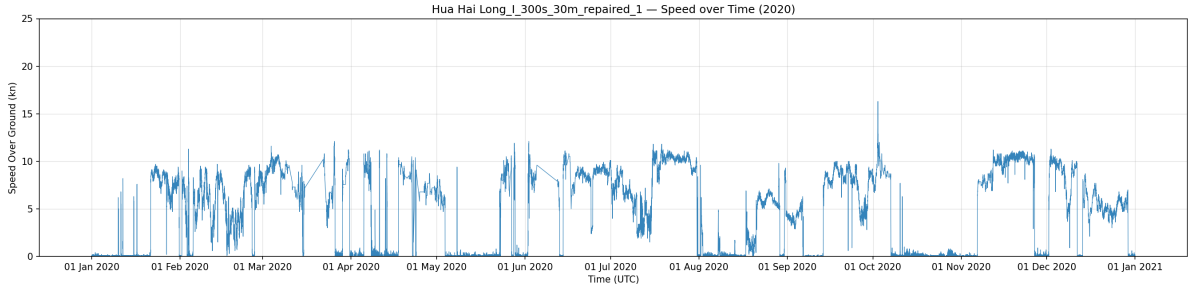


Target 2023 Speed Distribution

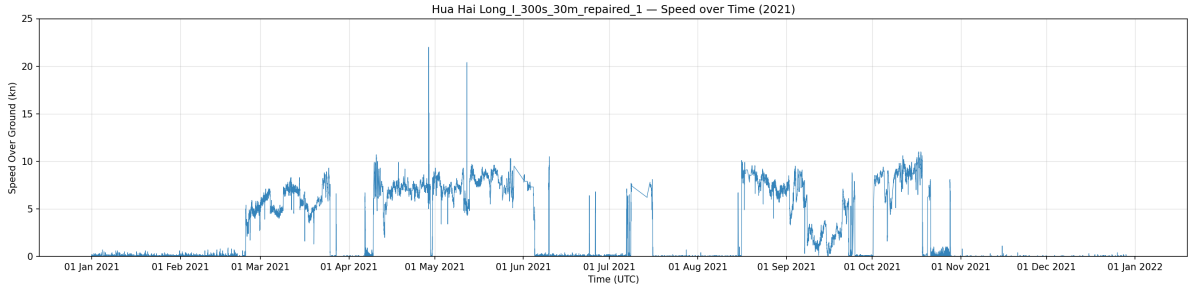


Target 2024 Speed Distribution

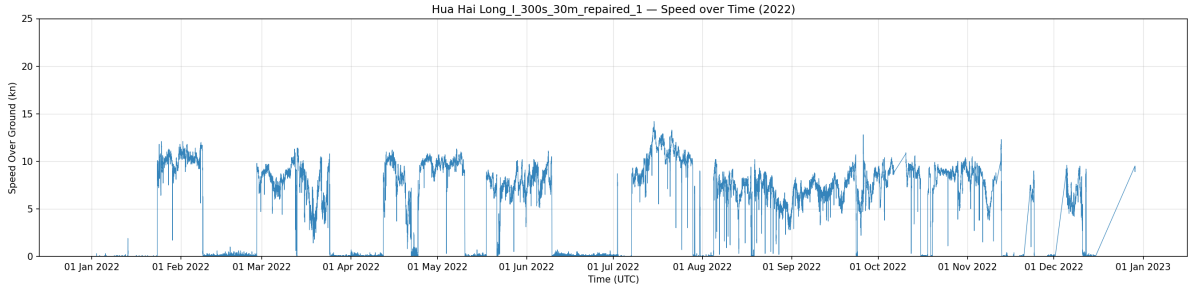
E.15. Speed Distribution Hua Hai Long



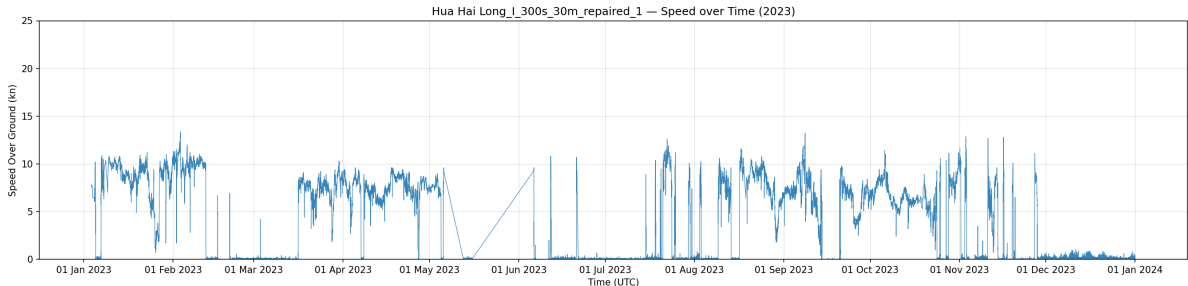
Hua Hai Long 2020 Speed Distribution



Hua Hai Long 2021 Speed Distribution

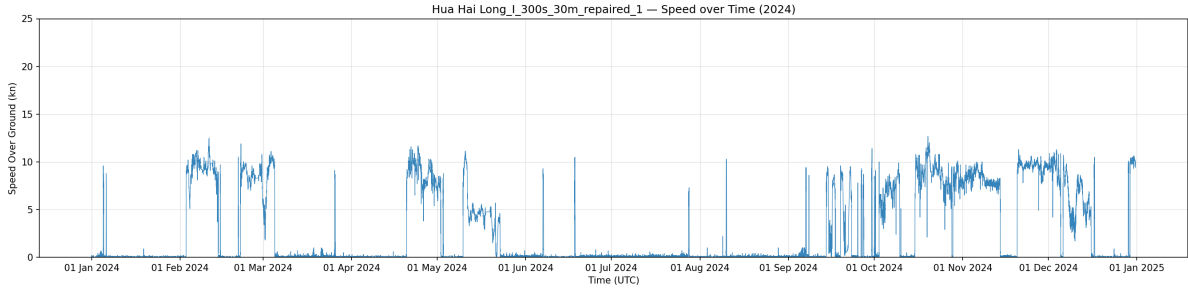


Hua Hai Long 2022 Speed Distribution



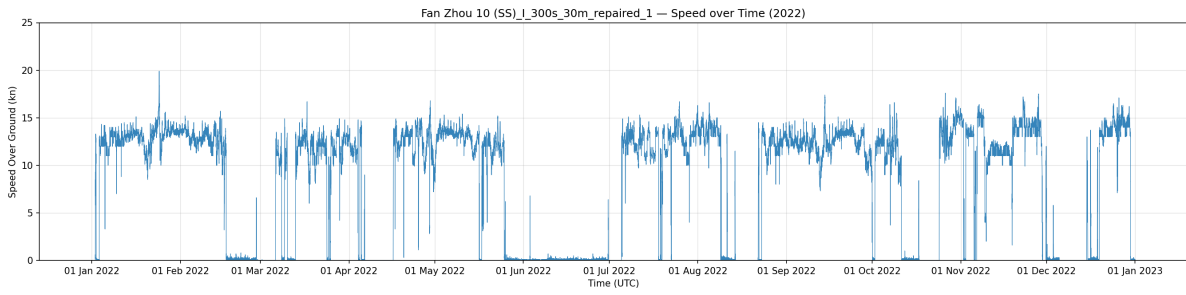
Hua Hai Long 2023 Speed Distribution

E.15. Speed Distribution Hua Hai Long

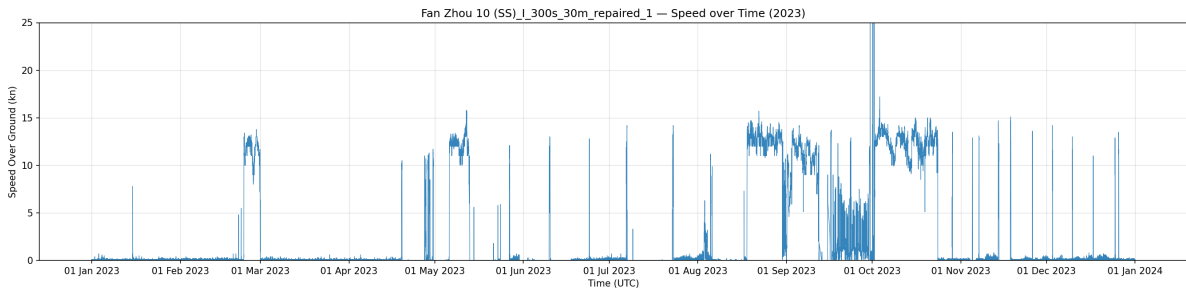


Hua Hai Long 2024 Speed Distribution

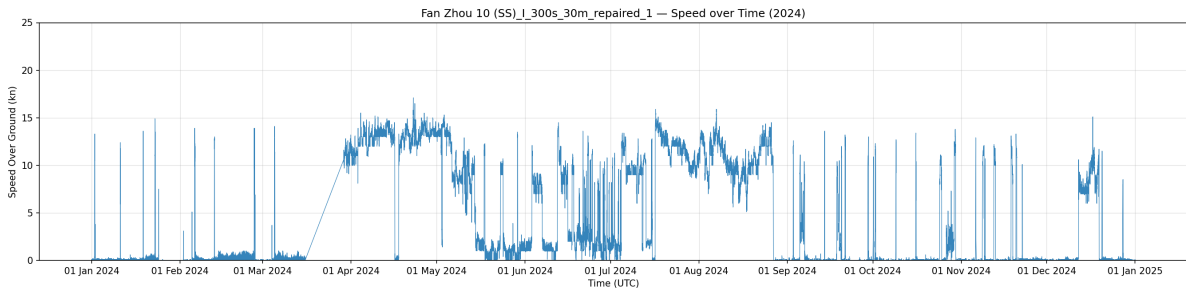
E.16. Speed Distribution Fan Zhou 10



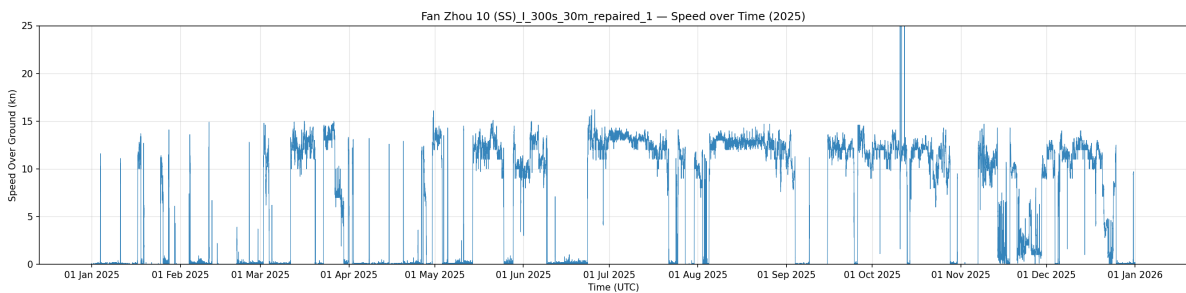
Fan Zhou 10 2022 Speed Distribution



Fan Zhou 10 2023 Speed Distribution

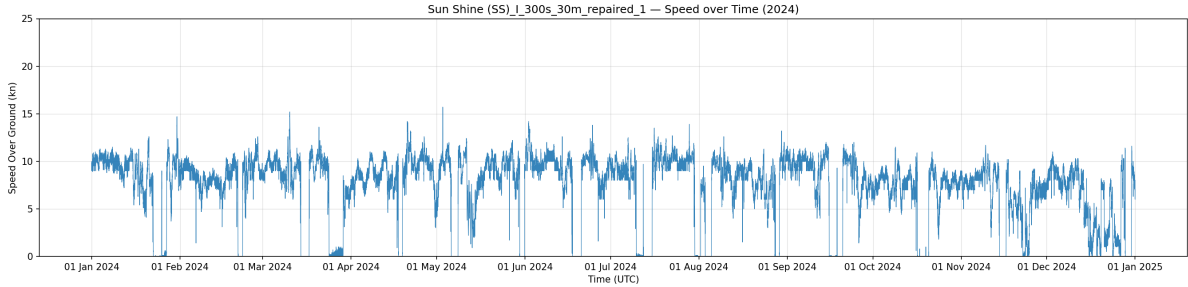


Fan Zhou 10 2024 Speed Distribution

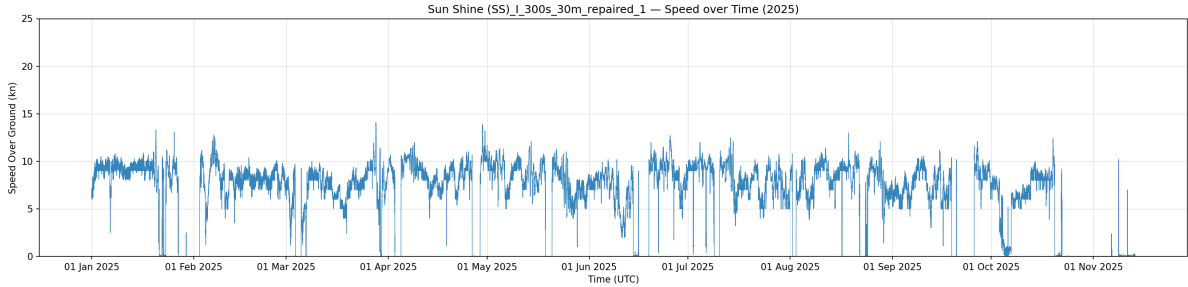


Fan Zhou 10 2025 Speed Distribution

E.17. Speed Distribution Sun Shine



Sun Shine 2024 Speed Distribution



Sun Shine 2025 Speed Distribution

F

Interpolation inconsistency example of Biglift Baffin 2019

In this appendix, additional data illustrating an interpolation inconsistency are presented. For the Biglift Baffin in 2019, a small section of the AIS data from Dataset 2 (interpolated) is shown in Table F.1. This section shows a large jump in position and therefore in travelled distance, while the time interval is only 5 minutes.

Table F.1: Snapshot of Dataset 2 (Interpolated) from Biglift Baffin 2019 containing an interpolation mistake

Date and UTC Time	Latitude [degree]	Longitude [degree]	SOG [kn]	Distance [nm]	Duration [s]	SOG calc [kn]	Distance calc [nm]
4/27/2019 6:45	21.9273	-179.957308	8.3	0.694	300	8.328	0.694
4/27/2019 6:50	21.9266	-179.969585	8.3	0.685	300	8.22	0.685
4/27/2019 6:55	21.925995	-175.615337	8.3	242.512	300	2910.144	242.496
4/27/2019 7:00	21.92528	-30.074947	8.3	7490.614	300	89887.368	7489.20
4/27/2019 7:05	21.924567	115.465443	8.2	7490.680	300	89888.16	7489.20
4/27/2019 8:55	21.909247	179.726913	8.1	3550.218	6600	1936.48	3550.216
4/27/2019 9:00	21.908415	179.714782	8.2	0.677	300	8.124	0.677

When investigating that same section in Dataset 1 (Table F.2), it can be observed that the travelled distance remained consistent and did not show large jumps. This means that the mistake occurred due to interpolation. Furthermore, it can be seen that in Dataset 1 the vessel moved from -179 degrees to 179 degrees, which is actually a very small distance, but during interpolation this was interpreted as travel over a large distance across the globe.

Table F.2: Snapshot of Dataset 1 from Biglift Baffin 2019 at the same timestamp as the interpolation mistake in Dataset 2

Date and UTC Time	Latitude [degree]	Longitude [degree]	SOG [kn]	Distance [nm]	Duration [s]	SOG calc [kn]
4/27/2019 6:49:30	21.92665	-179.96835	8.3	0.041	19	7.76
4/27/2019 6:52:11	21.926383	-179.974983	8.2	0.370	161	8.27
4/27/2019 6:52:31	21.926333	-179.9758	8.3	0.046	20	8.28
4/27/2019 6:52:51	21.9263	-179.976617	8.5	0.046	20	8.28
4/27/2019 6:53:11	21.92625	-179.977433	8.2	0.046	20	8.28
4/27/2019 6:54:11	21.9261	-179.979917	8.4	0.139	60	8.34
4/27/2019 6:54:51	21.926017	-179.98155	8.3	0.091	40	8.19
4/27/2019 7:07:13	21.92425	179.98835	8.2	1.680	742	8.15
4/27/2019 7:41:32	21.919167	179.90335	8.2	4.744	2059	8.29
4/27/2019 8:18:51	21.9147	179.81365	7.9	5.004	2239	8.05
4/27/2019 8:51:34	21.90985	179.735183	8	4.381	1963	8.03

In Table F.3, Dataset 1, Dataset 2, and Dataset 2.2 are compared. Dataset 2.2 is the dataset in which the interpolation inconsistencies were identified and removed. The comparison between Dataset 2 and Dataset 2.2 shows that only 4 data points were identified as interpolation inconsistencies, but this resulted in a very large difference in the total distance. Thus, a small number of mistakes can have a significant effect on the total distance when considering the data of one vessel in a specific year.

Table F.3: Comparison between Dataset 1, Dataset 2, and Dataset 2.2 cleaned only on SOG calculation for Biglift Baffin 2019

	Data points [#]	Total Distance [nm]	Max SOG_calc [kn]	SOG_calc >17 kn [#]
Dataset 1	89882	32,822	32	430
Dataset 2 (Interpolated)	68463	51,395	89,888	4
Dataset 2.2 (Cleaned interpolation)	68459	32,638	16	0

G

Speed plot comparison between Dataset 2 and 4

This appendix shows the CDF and ranked speed plots for each vessel. In Figure G.1, the CDF plots show the influence of cleaning and repairing on the median SOG. They also show that for some datasets cleaning and repairing led to a change in the overall yearly speed distribution, whereas for others the change was minimal.

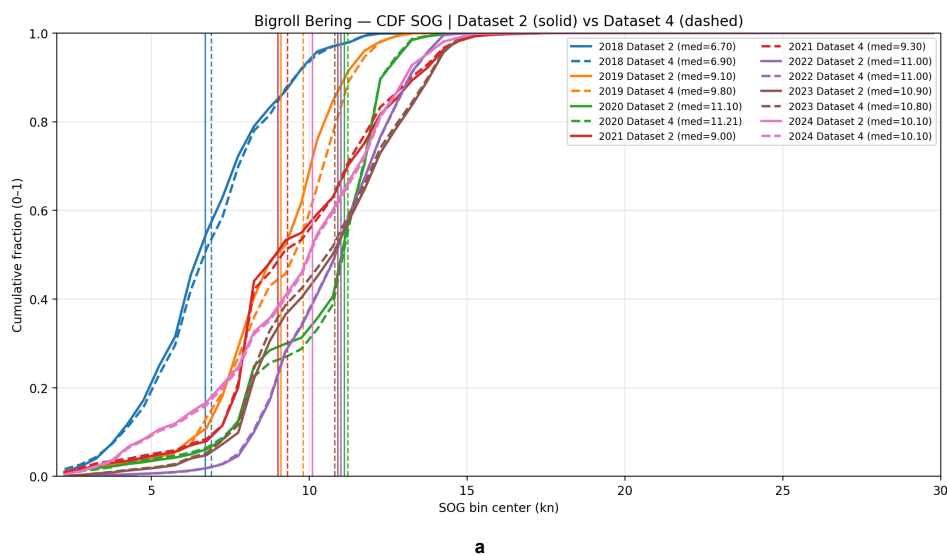
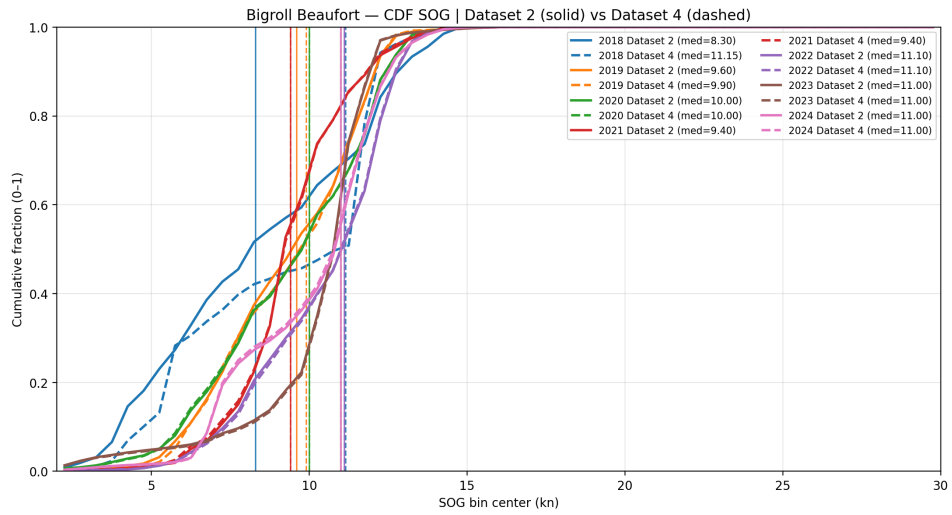
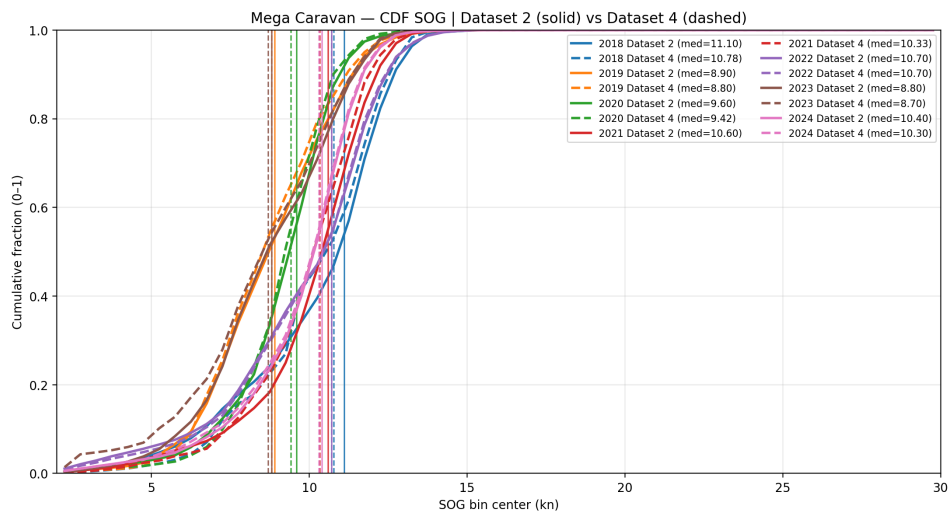


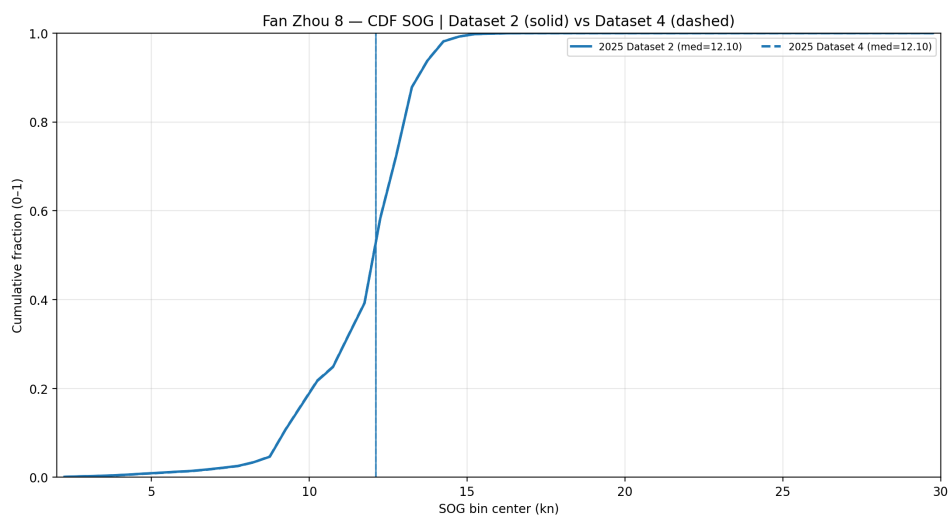
Figure G.1: CDF comparison between Dataset 2 and 4 for all vessels



b

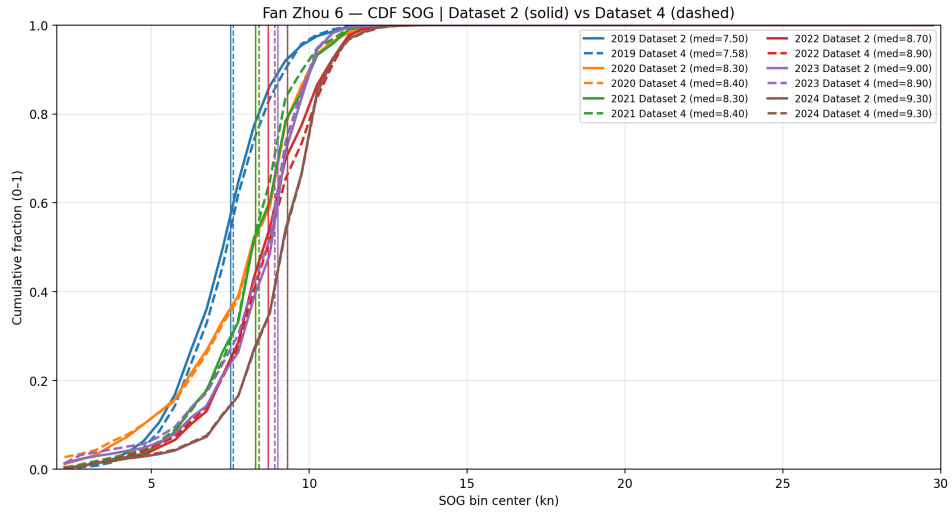


c

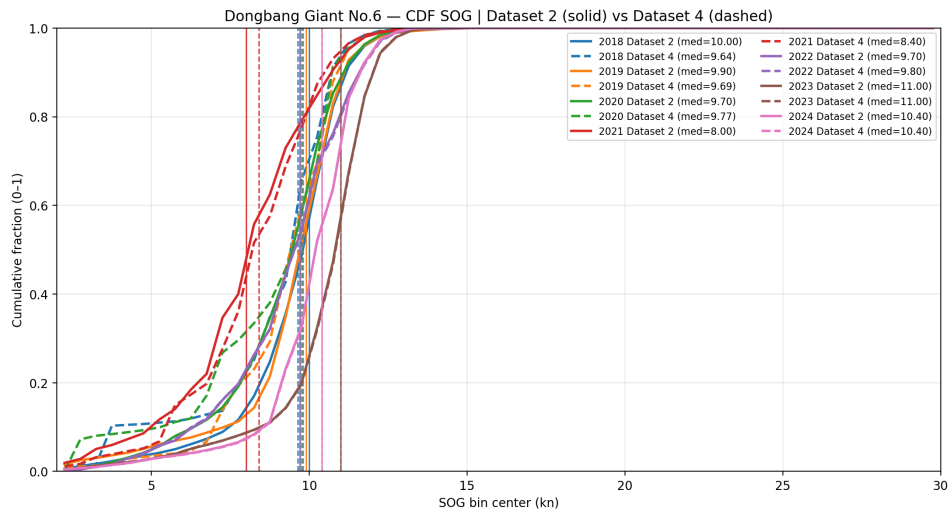


d

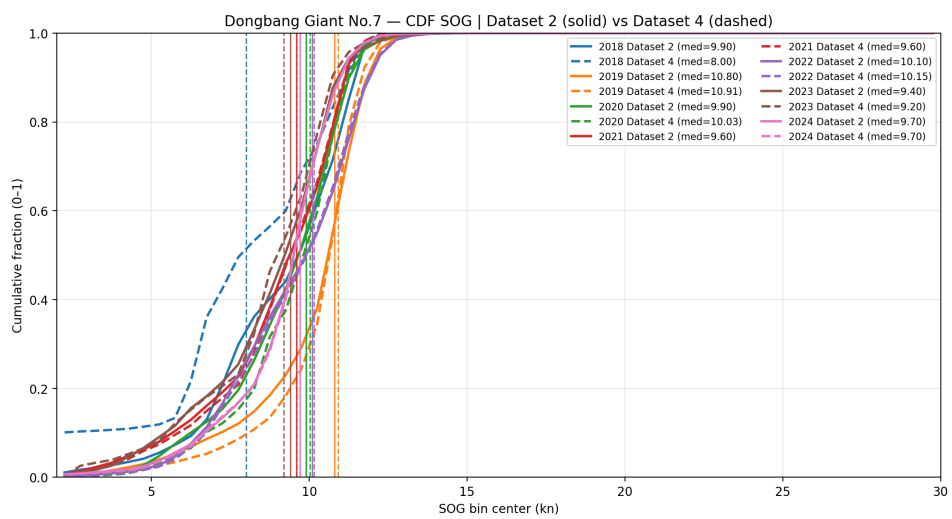
Figure G.1: CDF comparison between Dataset 2 and 4 for all vessels (continued)



e



f



g

Figure G.1: CDF comparison between Dataset 2 and 4 for all vessels (continued)

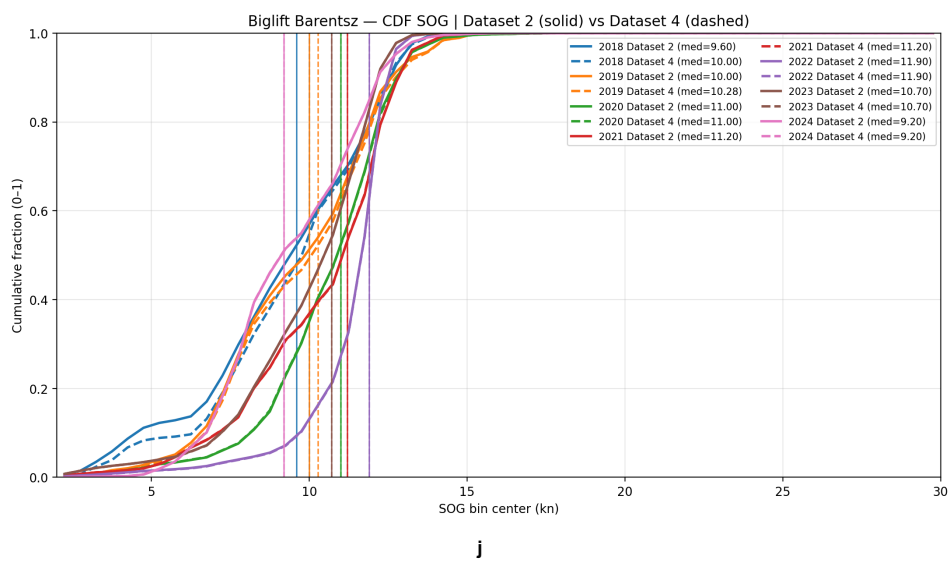
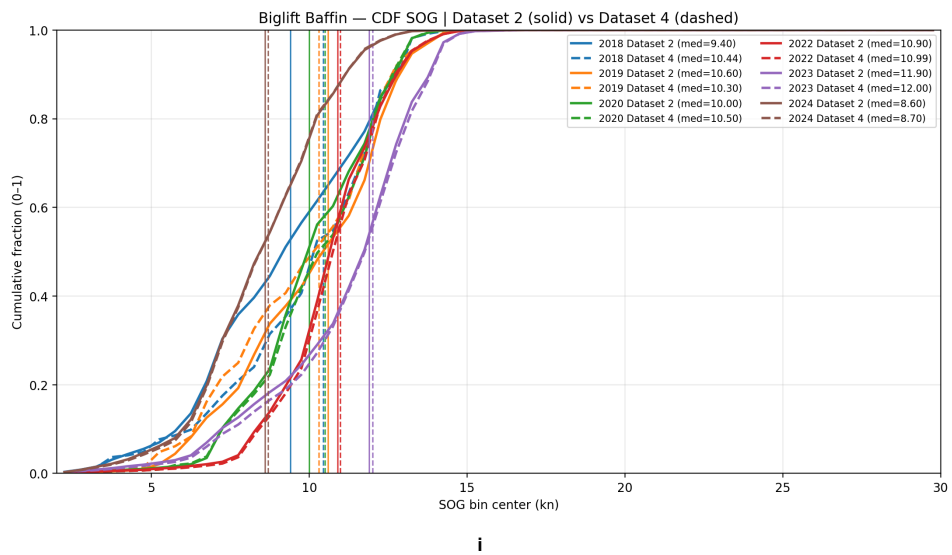
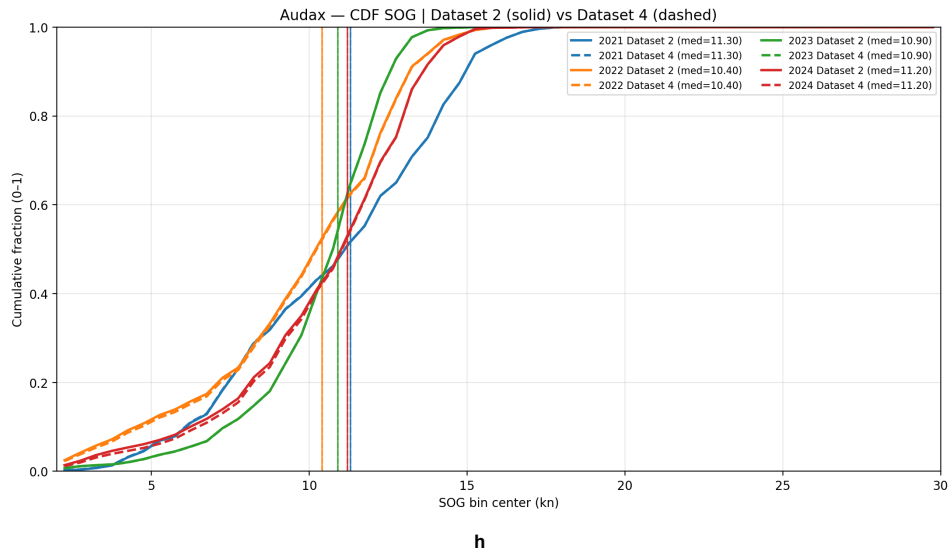


Figure G.1: CDF comparison between Dataset 2 and 4 for all vessels (continued)

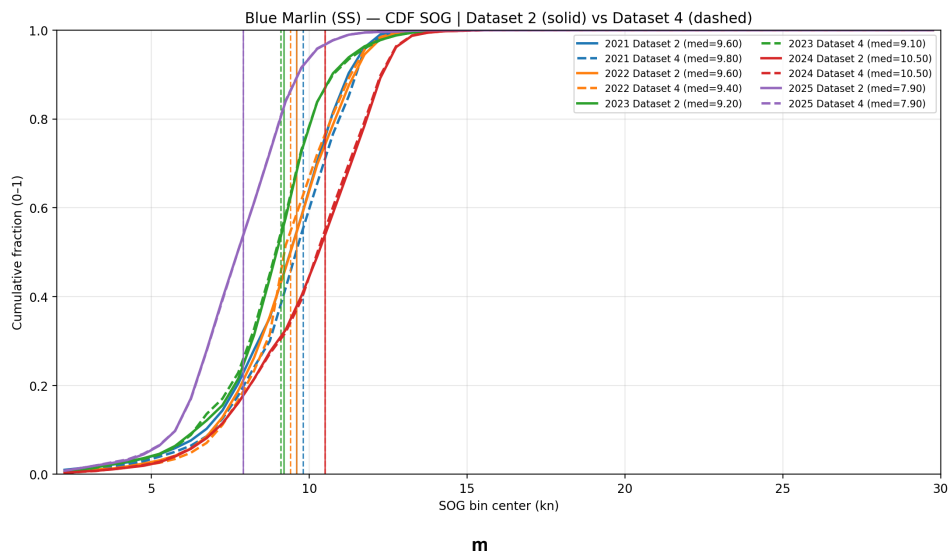
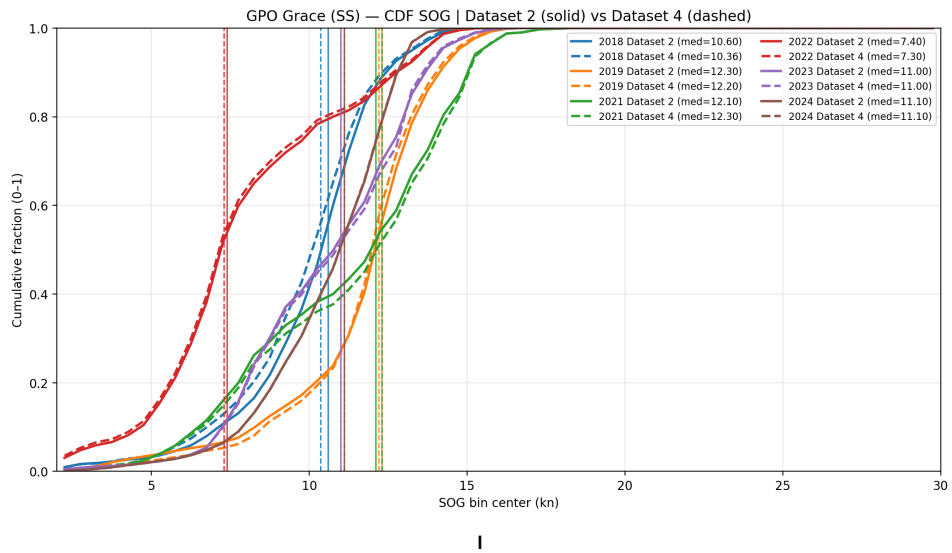
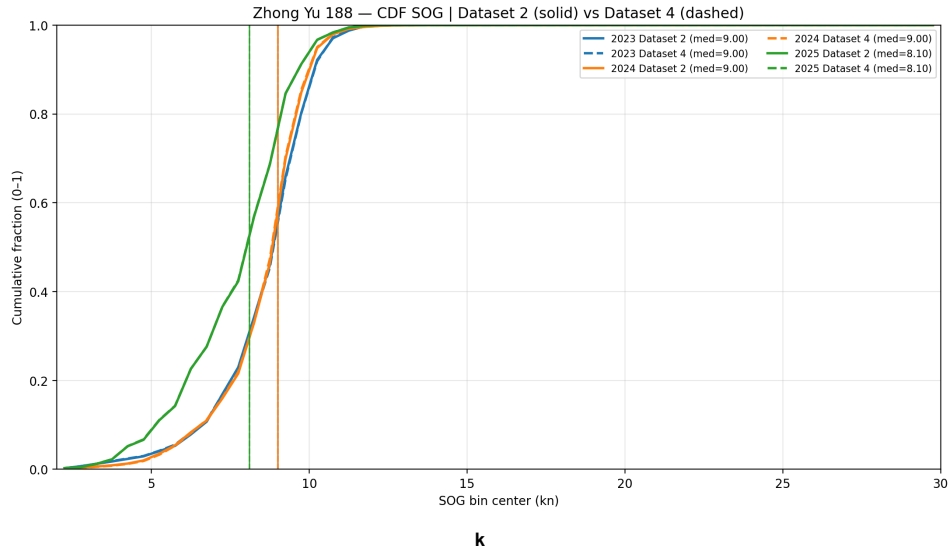


Figure G.1: CDF comparison between Dataset 2 and 4 for all vessels (continued)

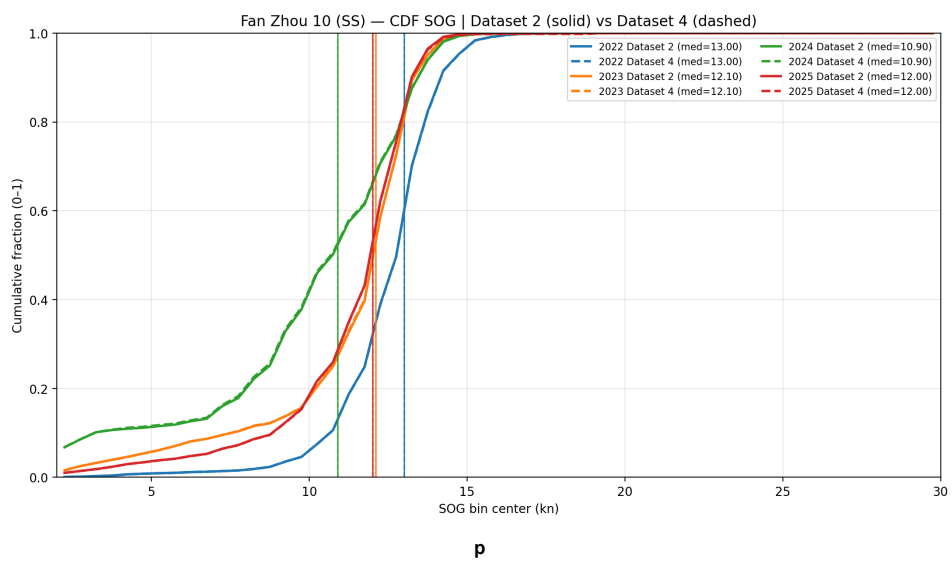
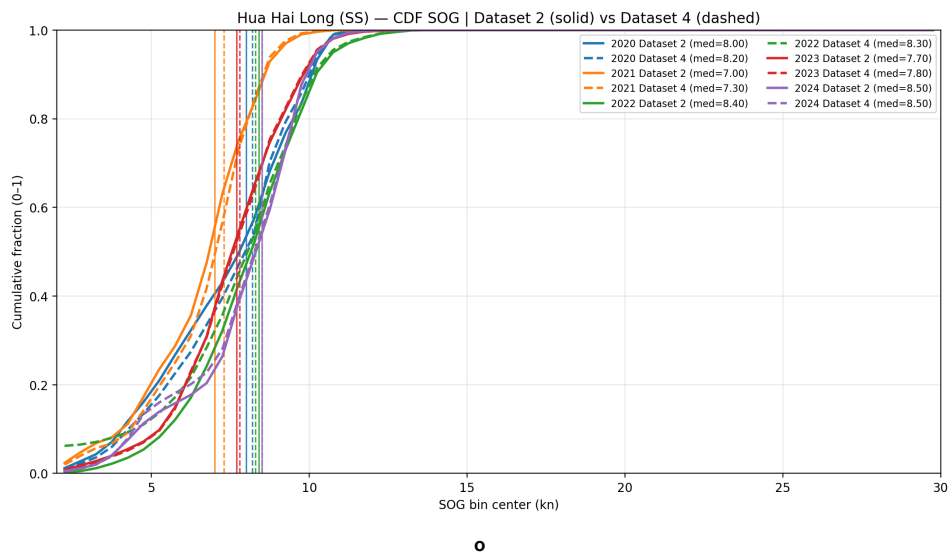
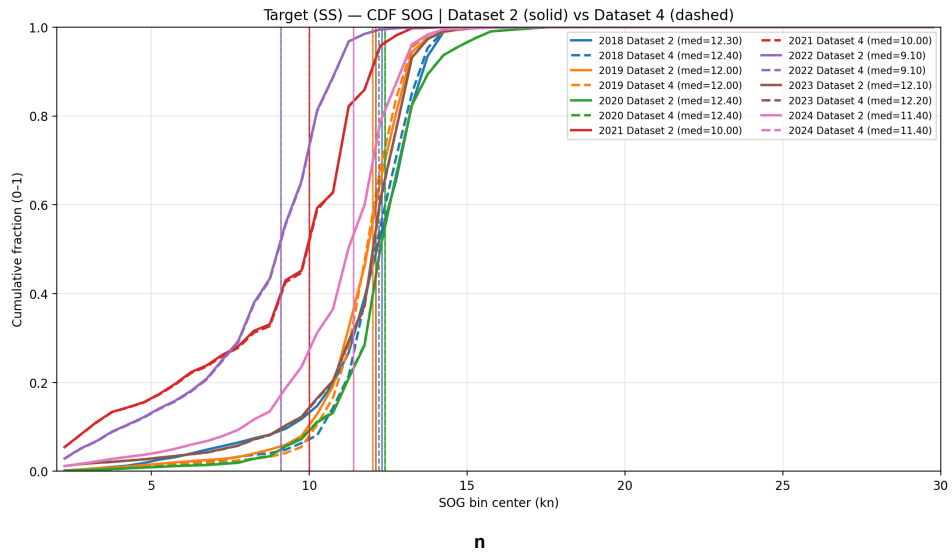


Figure G.1: CDF comparison between Dataset 2 and 4 for all vessels (continued)

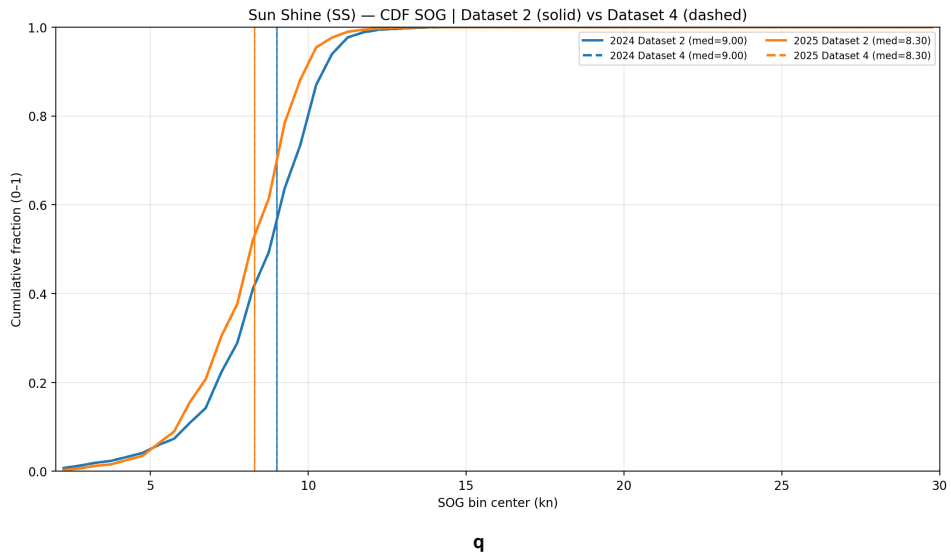


Figure G.1: CDF comparison between Dataset 2 and 4 for all vessels (continued)

To also show the impact of cleaning and repairing on speeds from 0 to 20 knots, a ranked speed plot was made for each vessel separately. This can be seen in Figure G.2. It shows in which speed range the repair had an impact for each dataset.

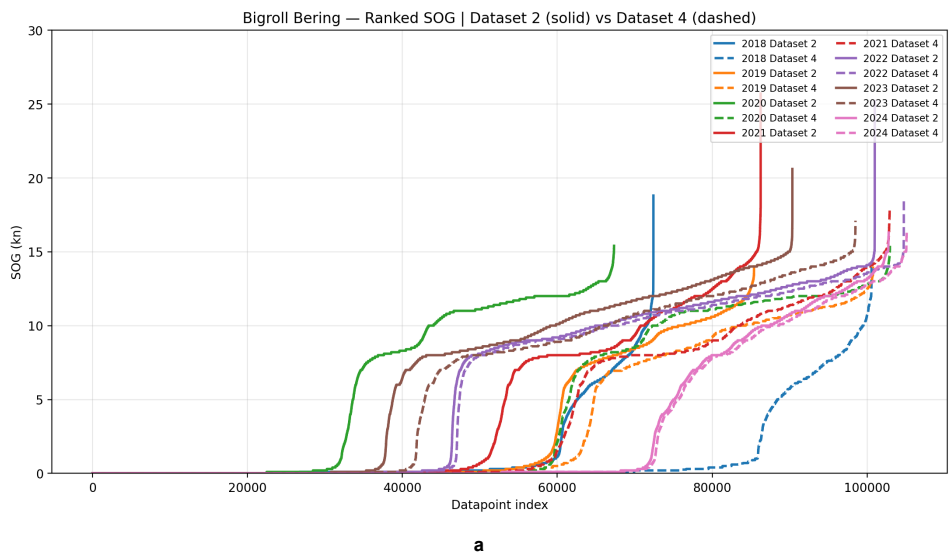
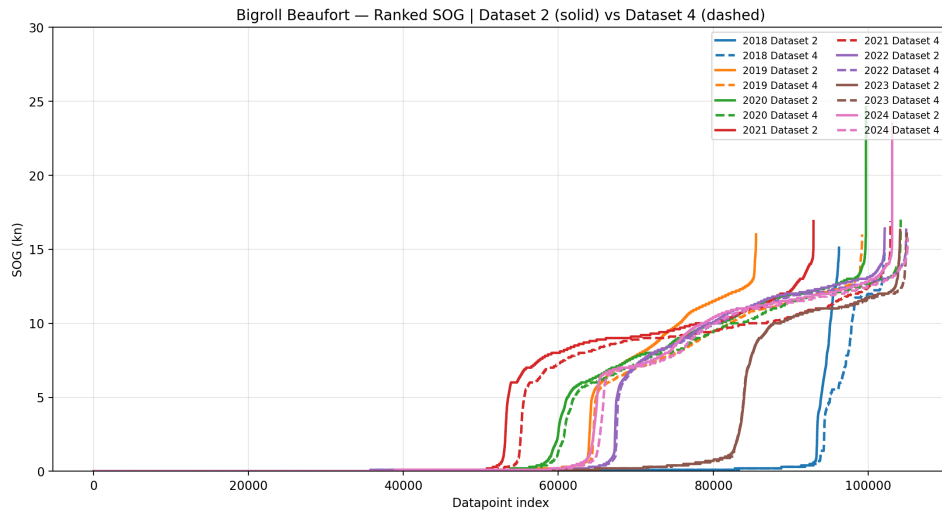
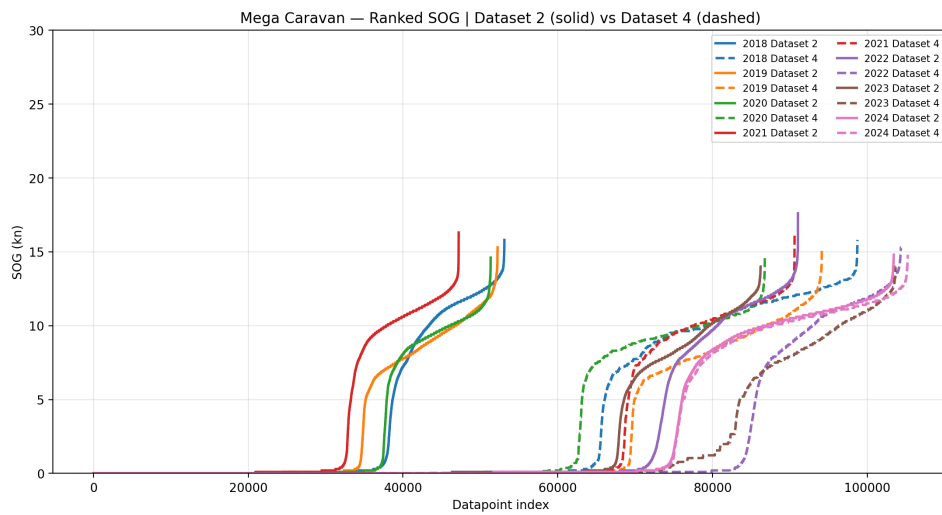


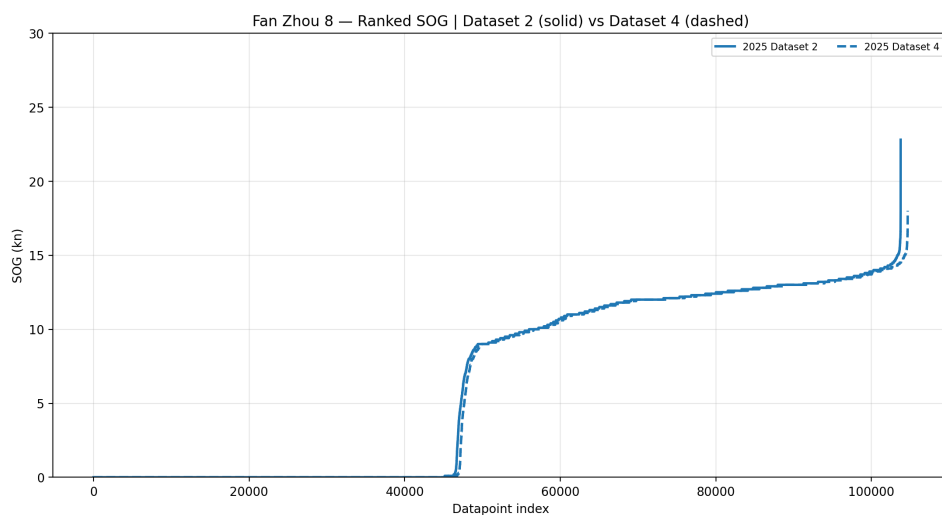
Figure G.2: Speed ranked plot comparison between Dataset 2 and 4 for all vessels



b

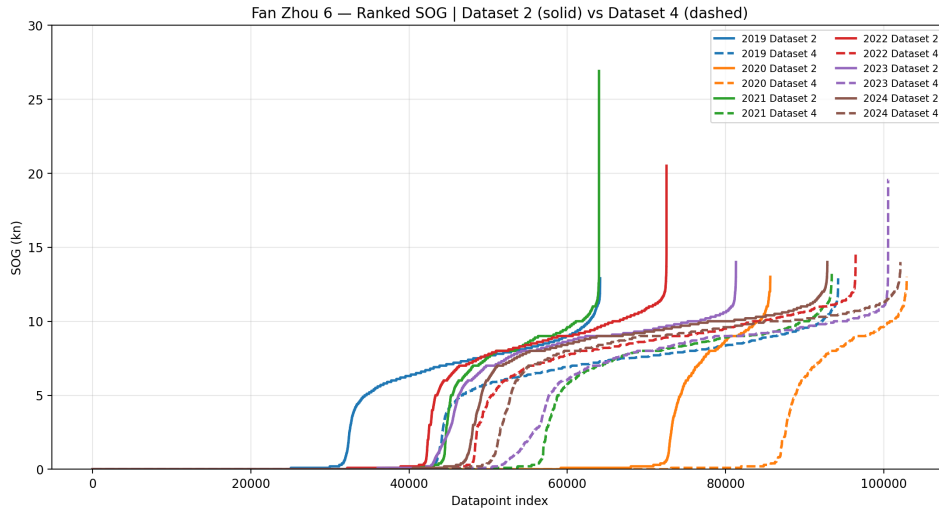


c

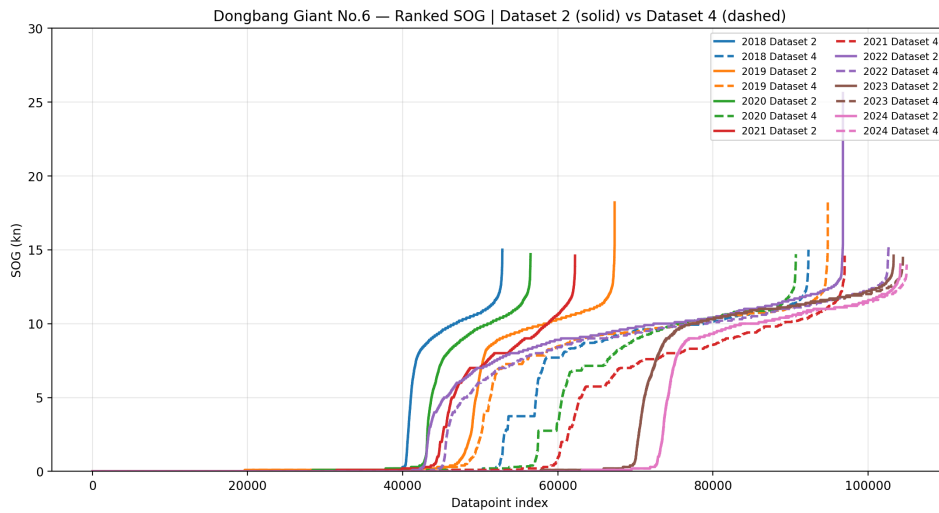


d

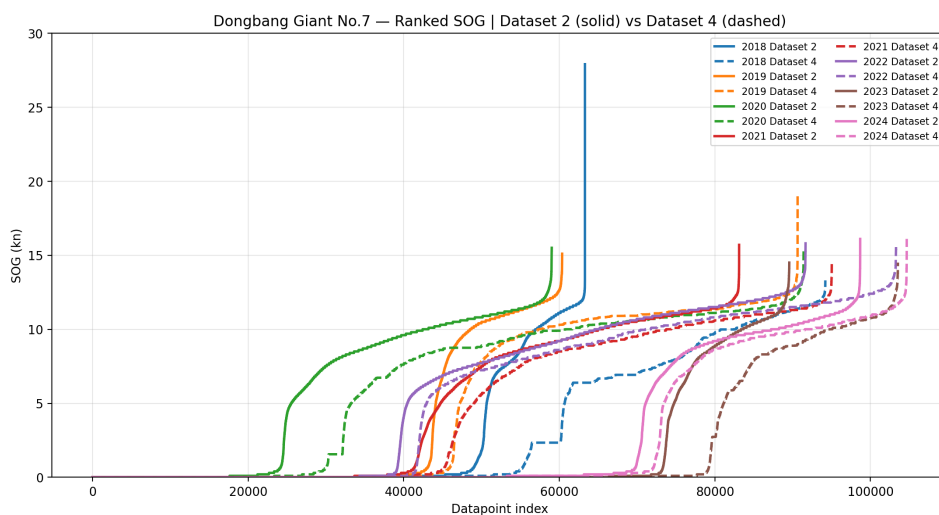
Figure G.2: Speed ranked plot comparison between Dataset 2 and 4 for all vessels (continued)



e

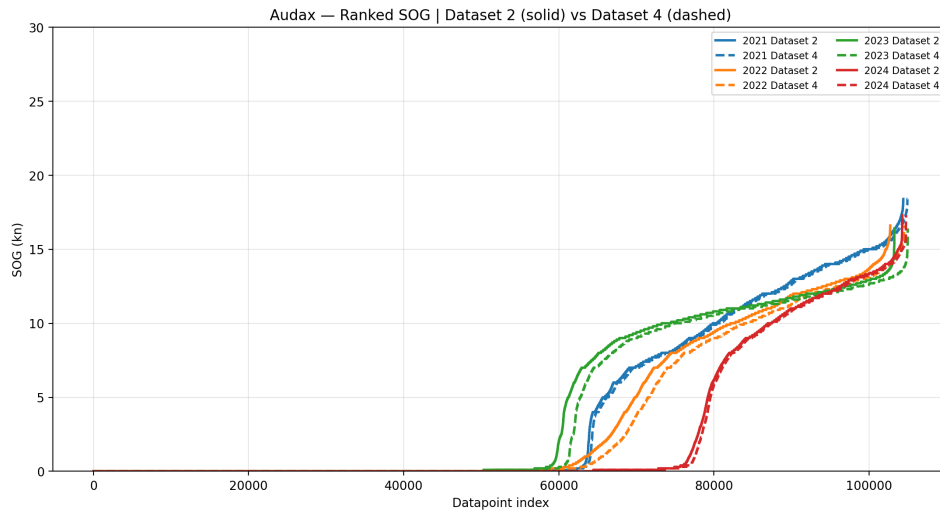


f

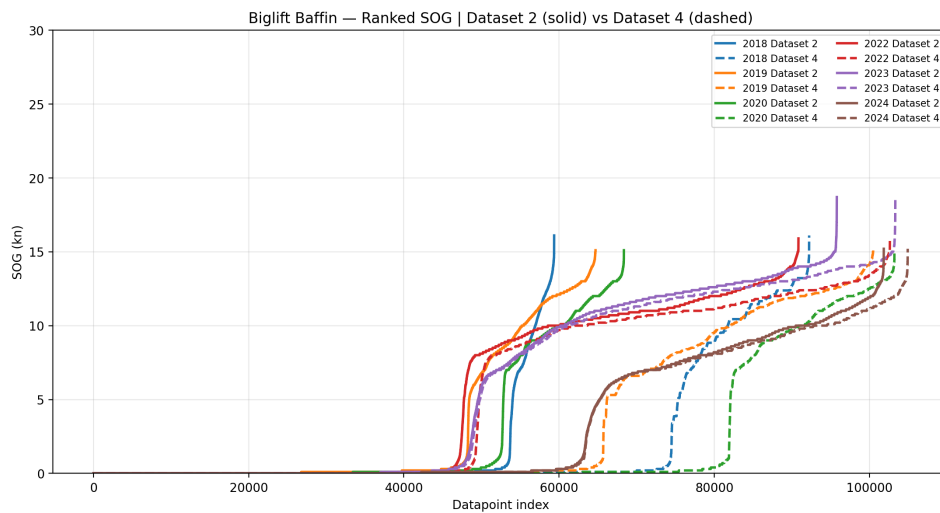


g

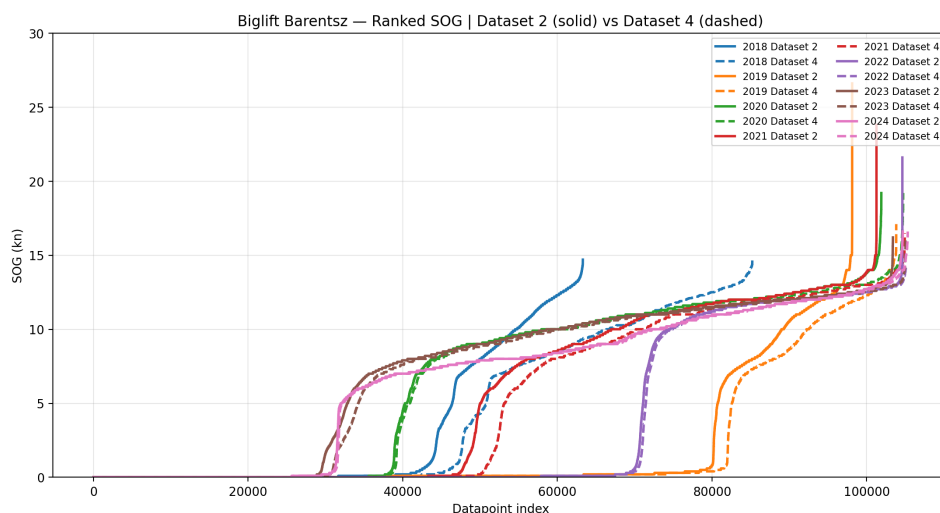
Figure G.2: Speed ranked plot comparison between Dataset 2 and 4 for all vessels (continued)



h



i



j

Figure G.2: Speed ranked plot comparison between Dataset 2 and 4 for all vessels (continued)

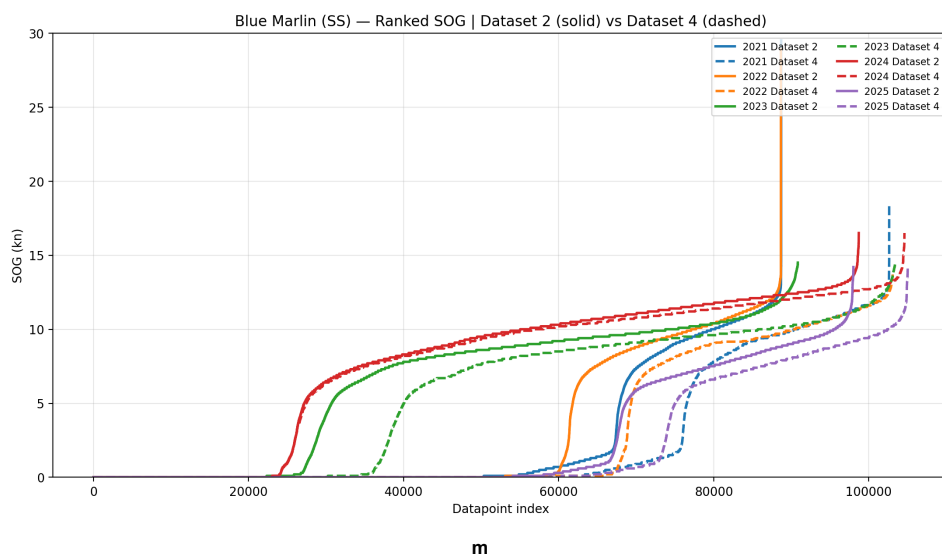
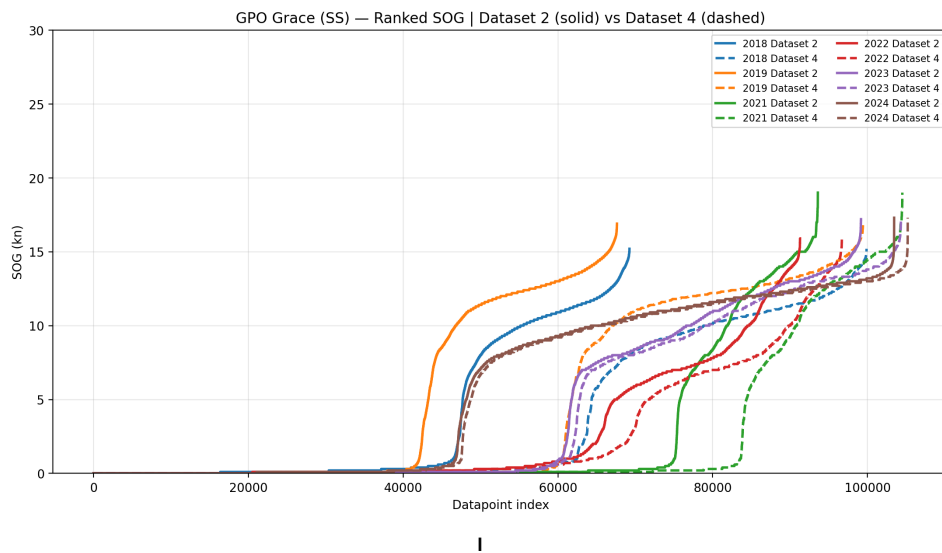
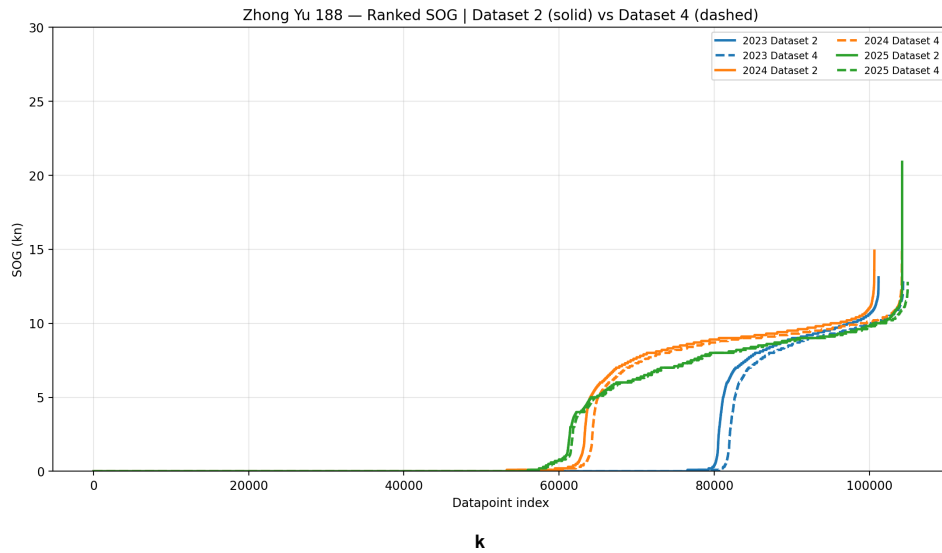


Figure G.2: Speed ranked plot comparison between Dataset 2 and 4 for all vessels (continued)

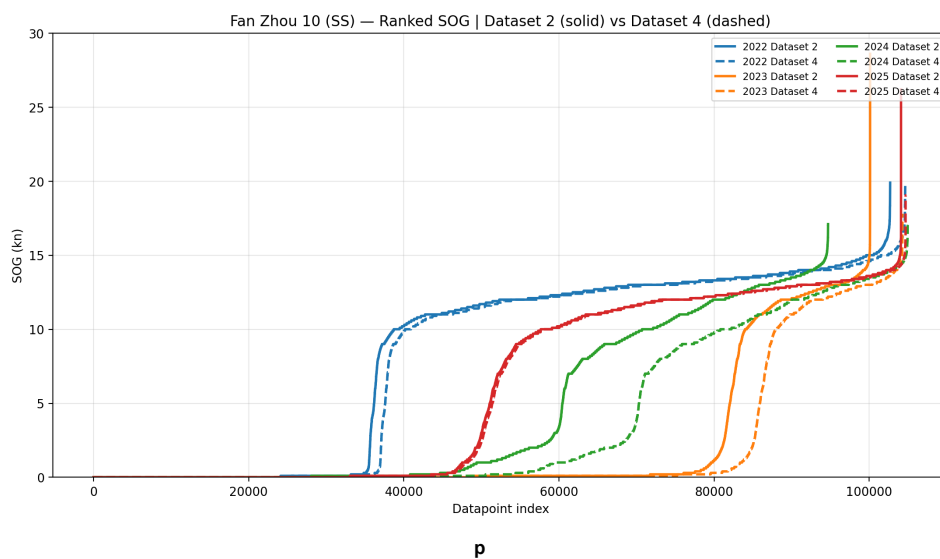
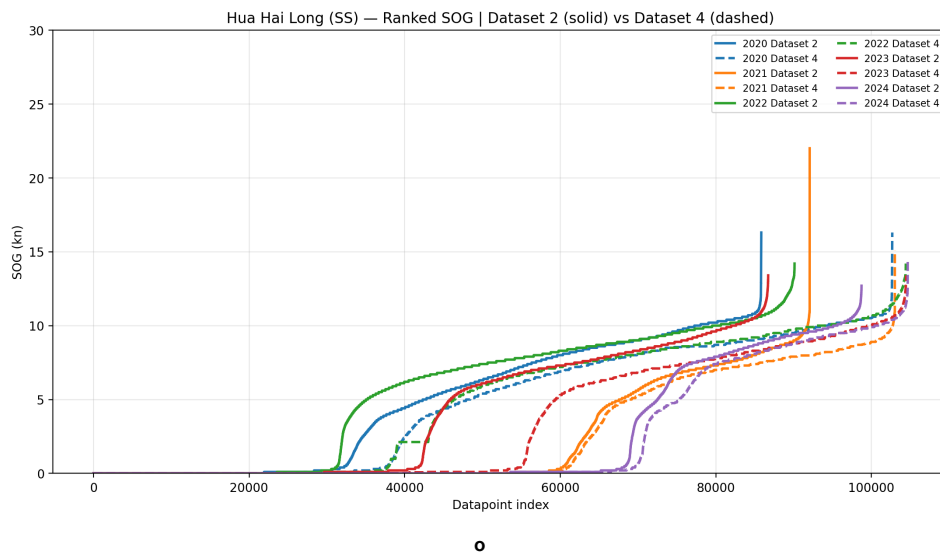
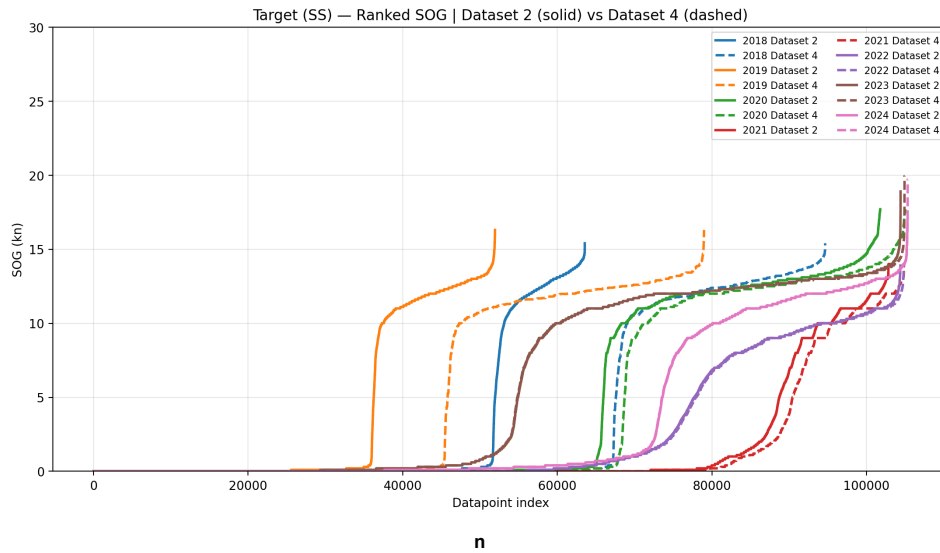
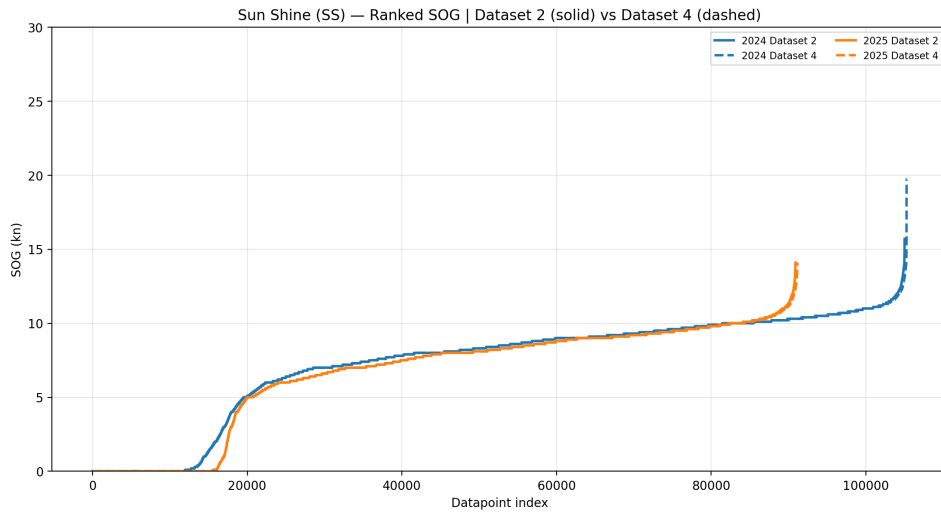


Figure G.2: Speed ranked plot comparison between Dataset 2 and 4 for all vessels (continued)



q

Figure G.2: Speed ranked plot comparison between Dataset 2 and 4 for all vessels (continued)

H

Results for Semi-Submersibles with Dataset 2 as input for the speed bins

This appendix shows the results derived from the unsegmented Semi-Submersible operational profile using Dataset 2 as input. Table H.1 shows the exact results, but these were very similar to the results obtained using Dataset 4 (cleaned and repaired) and were therefore not shown and discussed in the main research.

Table H.1: Semi-Submersibles Dataset 2 speed distribution

	Data points [#]	Duration [h]	Duration [%]	Distance [nm]	Distance [%]	Mean SOG [kn]	Median SOG [kn]
0 - 0.5 kn	1,426,286	130,919	52.73	8,349	0.72	0.044	0
0.5 - 2 kn	94,993	8,509	3.36	8,786	0.76	0.981	0.9
2 - 4 kn	35,179	3,846	1.52	9,188	0.79	2.855	2.8
4 - 6 kn	64,159	5,984	2.37	30,961	2.67	5.09	5.1
6 - 20 kn	1,053,988	103,793	41.01	1,103,929	95.07	10.218	10.1
Total for SS	2,674,605	253,053	100	1,161,214	100	-	-
Median 1 +- 2 kn	595,036	58,708	23.20	572,428	49.30	9.632	9.6

Operational Profile Using Segmented Method: Comparison Between Dataset 2 and Dataset 4 for Biglift Baffin

This appendix presents an example of the impact of the data cleaning and repair steps applied in this research at vessel level. Instead of comparing all vessels, a single representative case is analysed: the *Biglift Baffin* in 2019.

Table I.1 shows the overall differences between Dataset 2 and Dataset 4. It can be observed that Dataset 2 results in a higher number of segments compared to Dataset 4. However, the most significant difference is found in the total sailed distance. This substantial discrepancy is caused by interpolation inconsistencies present in Dataset 2. As a result, the vessel-level operational profile is strongly affected, particularly in terms of total distance travelled.

Table I.1: Comparison of vessel-level operational profile between Dataset 2 and Dataset 4 for the Biglift Baffin (2019)

Biglift Baffin 2019	Unit	Dataset 2	Dataset 4	Difference
Segments	#	45	39	15.4%
Distance	nm	51433.28	32672.41	36.5%

When analysing the distribution of time across operational modes, it can be seen that the relative differences between Dataset 2 and Dataset 4 are minimal. As shown in Table I.2, all differences remain below 1%. This indicates that, when considering only the percentage of time spent in each operational mode, the influence of data cleaning and repair is limited.

Table I.2: Time distribution comparison between Dataset 2 and Dataset 4 for the Biglift Baffin (2019)

Biglift Baffin 2019	Unit	Dataset 2	Dataset 4	Difference
Long Sailing	%	37.12	37.74	0.58
Short Sailing	%	0.32	0.29	0.03
Manoeuvre	%	0.82	0.00	0.82
Stalling	%	0.00	0.00	0.00
Moored	%	10.35	10.50	0.15
Anchored	%	51.39	51.48	0.09

The largest differences at vessel level are observed in the mean values per operational segment, as presented in Table I.3. As expected, the most significant discrepancy occurs in the long sailing distances,

where a difference of 72.2% is observed. This again highlights the impact of interpolation errors in Dataset 2, which artificially increase sailing distances.

Table I.3: Mean value comparison between Dataset 2 and Dataset 4 for the Biglift Baffin (2019)

Biglift Baffin 2019	Unit	Dataset 2	Dataset 4	Difference
Long Sailing	nm	4641.32	2695.33	72.2%
Short Sailing	nm	28.00	26.82	4.39%
Moored	h	151.16	181.40	16.67%
Anchored + Stalling	h	281.55	317.65	11.36%

Overall, it can be concluded that data cleaning and repair, particularly addressing unrealistic values, interpolation errors, and mismatches, are essential when deriving vessel-level or fleet-level operational profiles that rely on absolute values such as distance or segment characteristics. However, when only the relative time distribution across operational modes is of interest, the impact of these data improvements is limited.

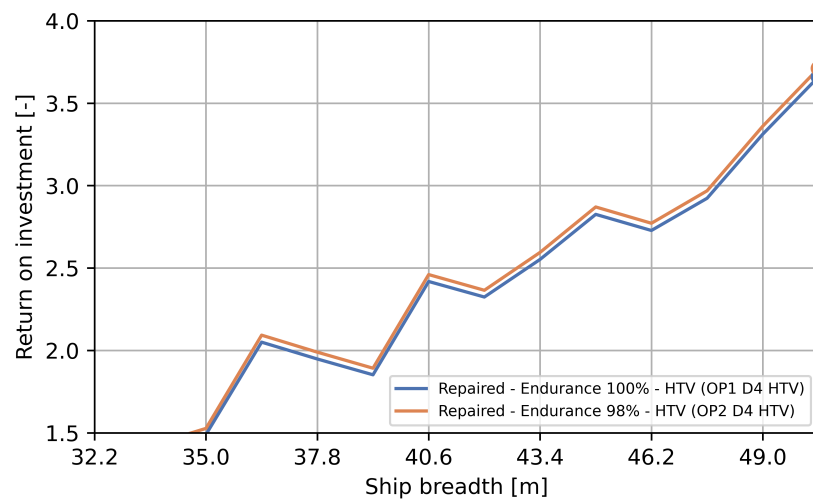
J

Blended Design Additional Results

This appendix presents additional results from the Blended Design analysis. These results were not discussed in detail in the main body of the report, as the observed differences are minimal. However, they are included here for completeness.

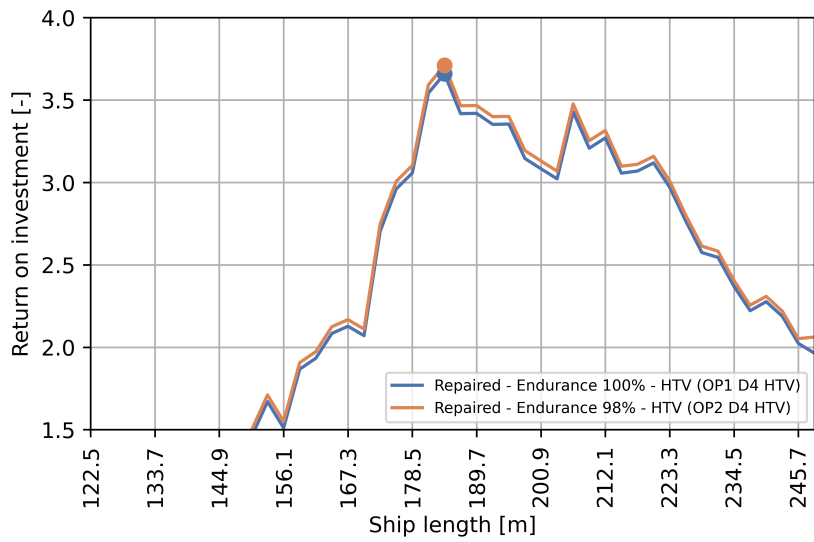
In Figure J.1, the results of the Blended Design using OP 1 D4 HTV and OP 2 D4 HTV as input are presented. These operational profiles are described in section 9.1. OP 1 D4 HTV is based on the maximum identified endurance, whereas OP 2 D4 HTV uses the endurance required to complete 98% of the segments. The difference in endurance between these profiles is 31 days, which is expected to influence onboard requirements such as fuel capacity.

Despite this difference, the impact on the overall design results is negligible. As shown in Figure J.1, both operational profiles yield nearly identical trends, following the same relationships between ROI and the main design parameters. A slightly higher ROI is observed for the reduced endurance case; however, no significant influence on the vessel dimensions can be identified.

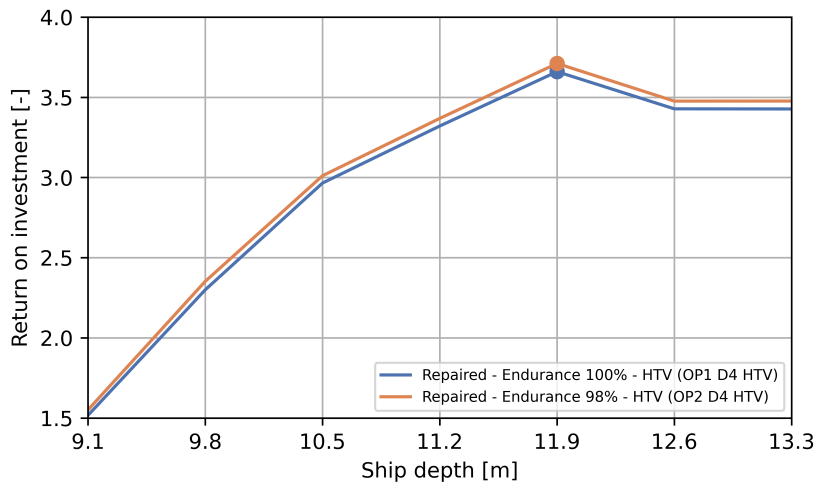


a

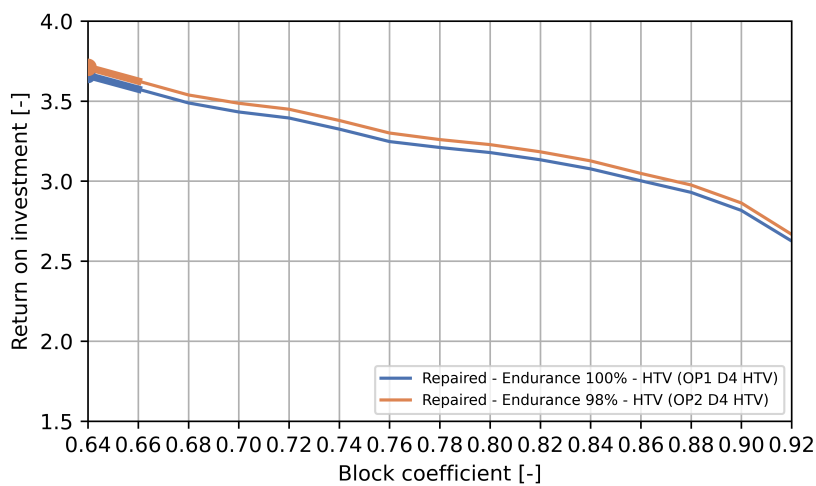
Figure J.1: Comparison between Blended Design results with different endurance



b

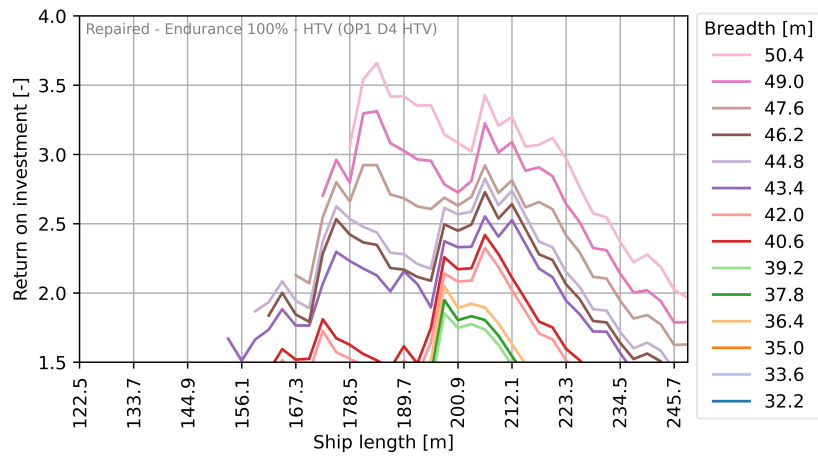


c

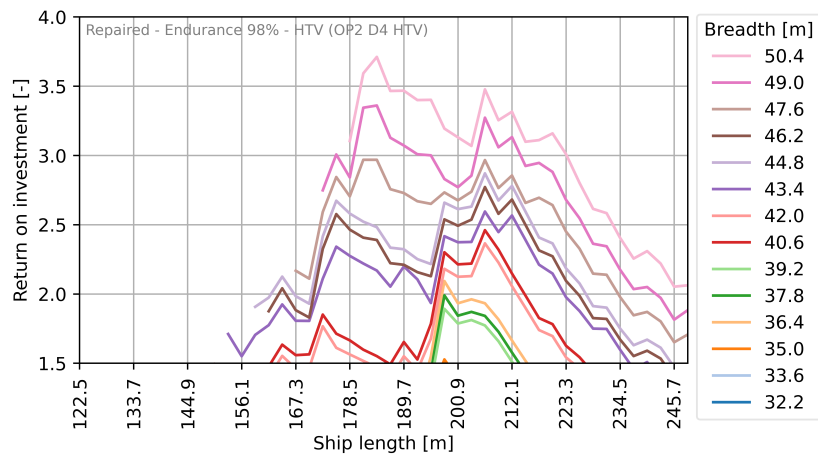


d

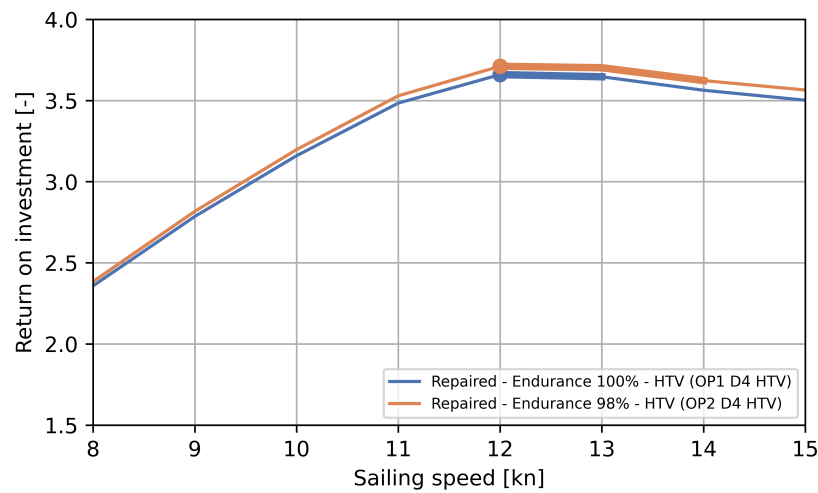
Figure J.1: Comparison between Blended Design results with different endurance (continued)



e



f



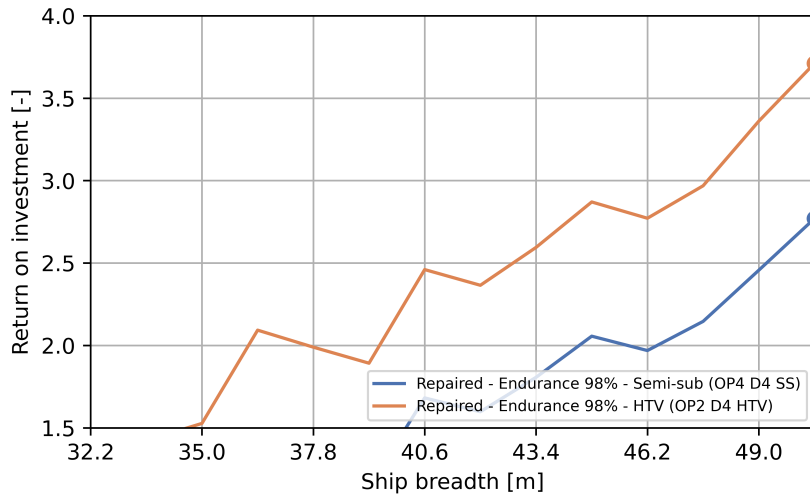
g

Figure J.1: Comparison between Blended Design results with different endurance (continued)

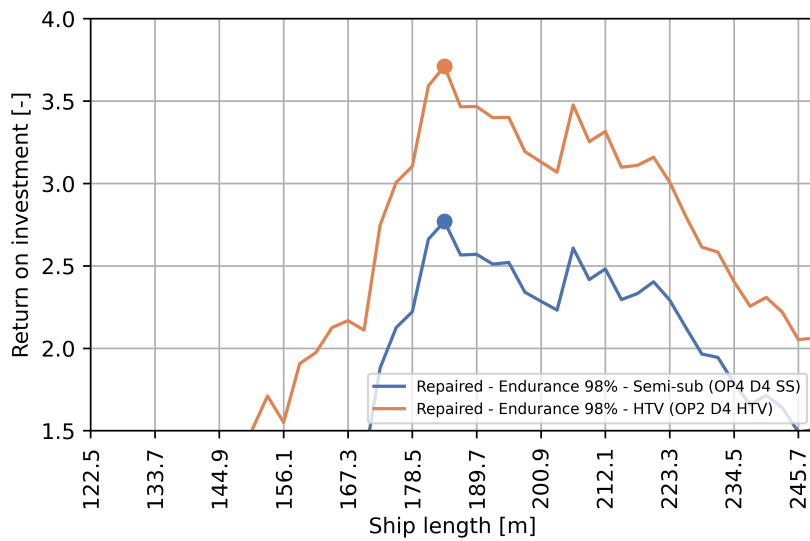
In Figure J.2, additional results comparing operational profiles of HTVs and Semi-Submersibles are presented. In contrast to the endurance comparison, these results show more pronounced differences,

as already discussed in chapter 9.

Although the overall trends and relationships between design variables and ROI remain similar, the Semi-Submersible configurations consistently result in lower ROI values. This can be attributed to their longer trip distances and different operational characteristics. Consequently, while the relative influence of design parameters remains unchanged, HTVs are expected to achieve higher ROI values overall.

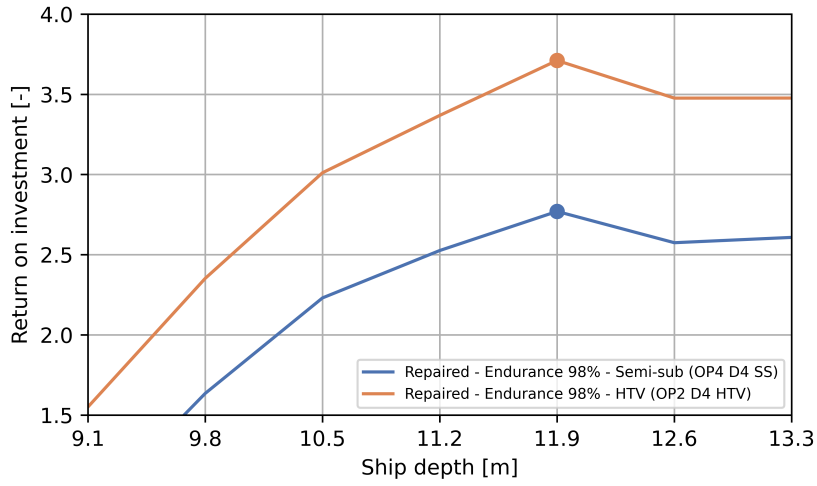


a

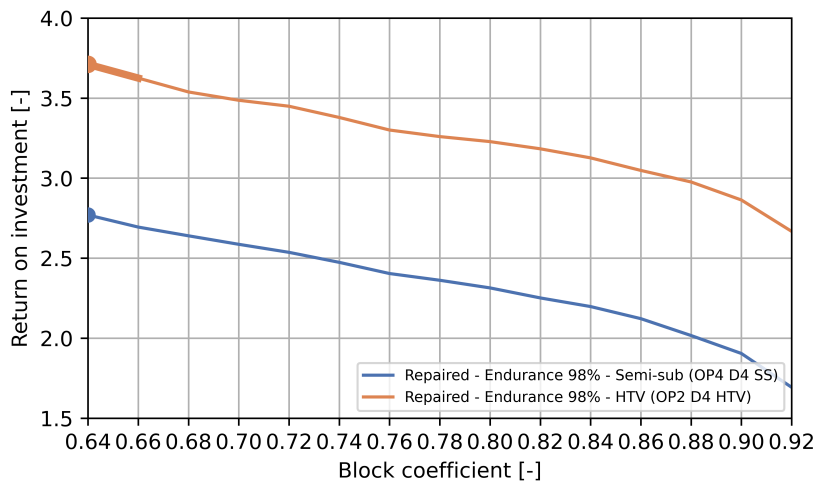


b

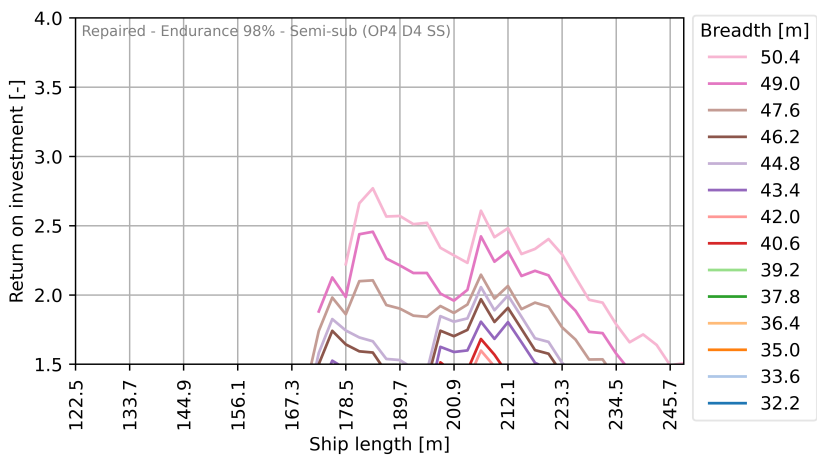
Figure J.2: Comparison between Blended Design results of HTV and Semi-Submersible



c



d



e

Figure J.2: Comparison between Blended Design results of HTV and Semi-Submersible (continued)

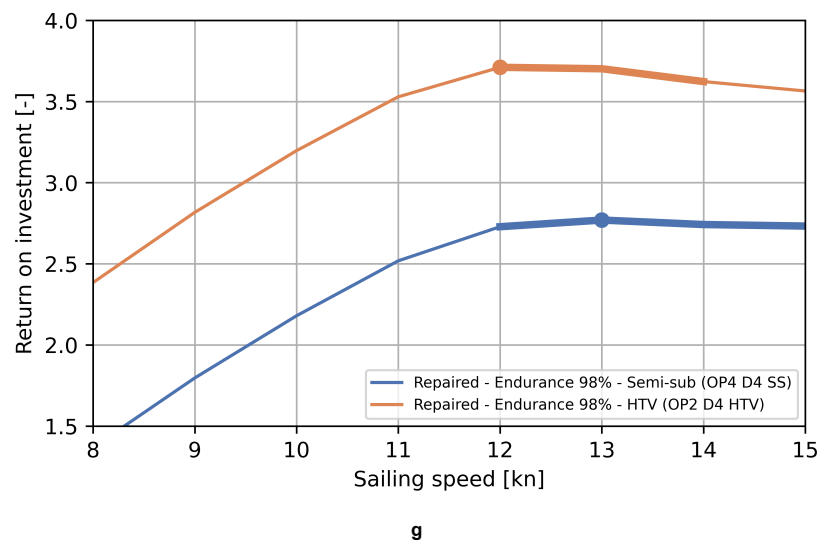
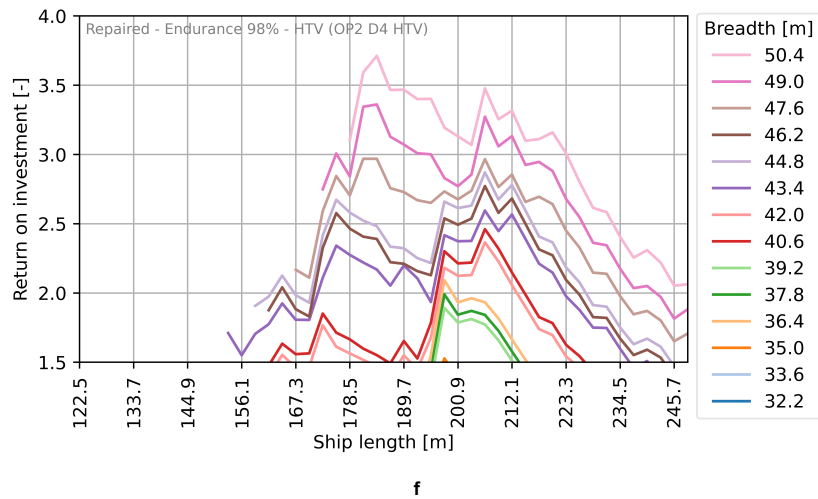
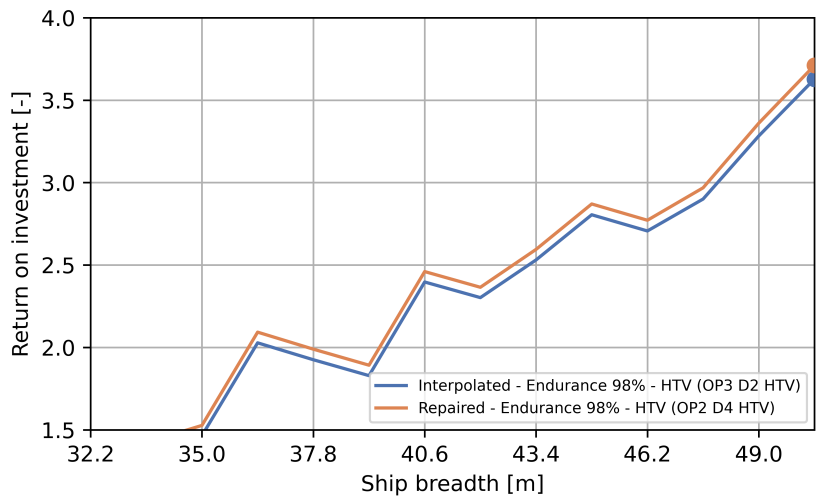


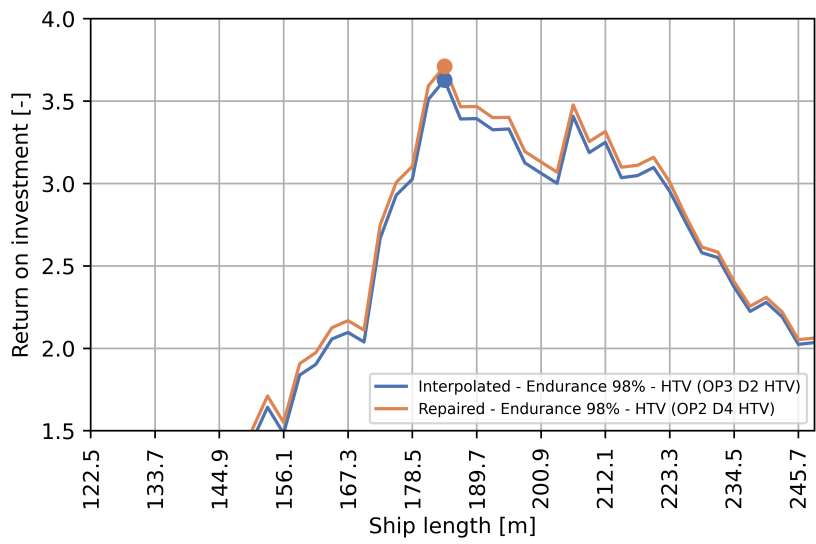
Figure J.2: Comparison between Blended Design results of HTV and Semi-Submersible (continued)

Finally, the impact of the data cleaning and repairing procedure on the Blended Design results is shown in Figure J.3. As previously discussed, the effect of cleaning and repairing on the operational profiles themselves is limited. This is also reflected in the design results, where only minor differences can be observed.

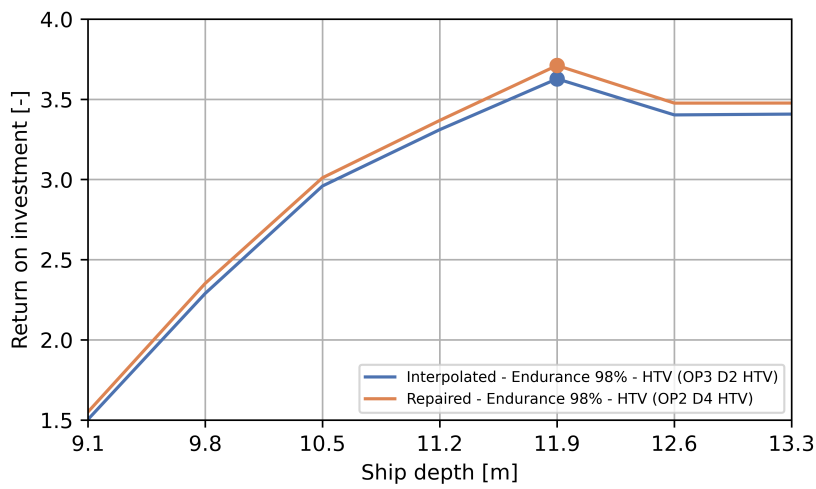
The repaired dataset results in a slightly higher ROI compared to the interpolated dataset. This suggests that neglecting data cleaning and repairing may lead to a modest underestimation of future performance. However, the overall relationships between vessel dimensions and ROI remain unaffected.



a

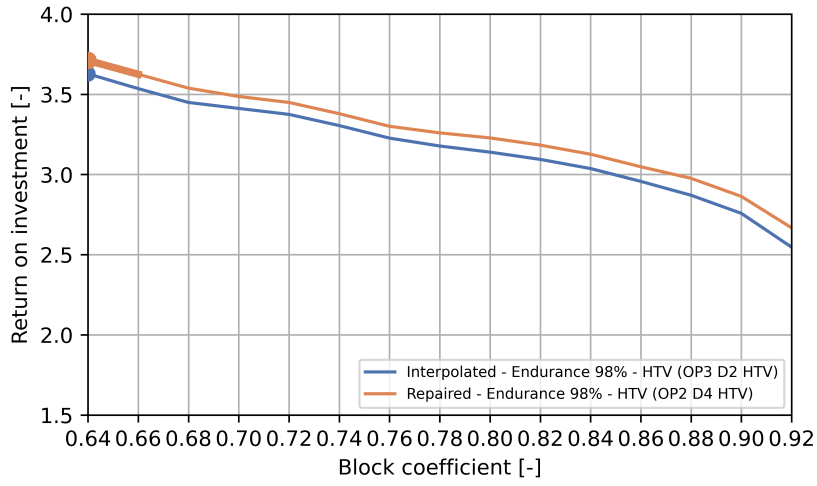


b

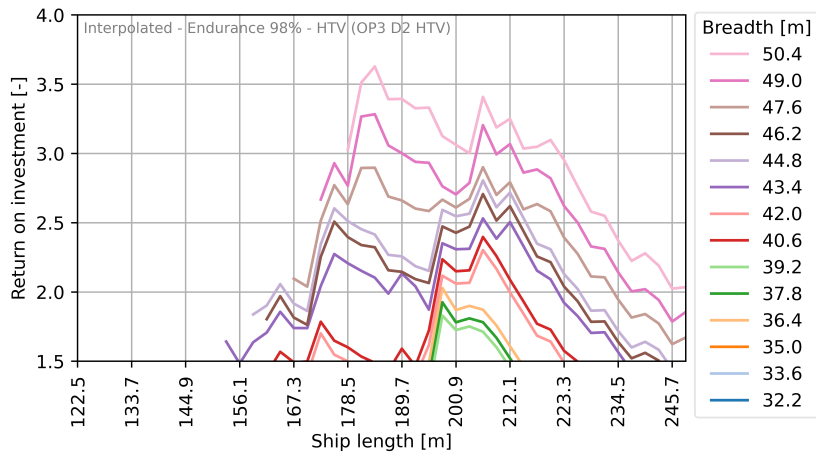


c

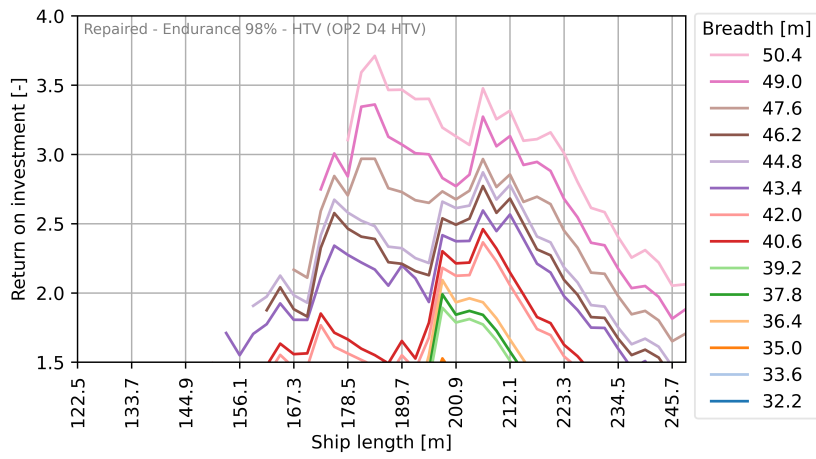
Figure J.3: Comparison between Blended Design results of HTV using Dataset 2 and 4



d



e



f

Figure J.3: Comparison between Blended Design results of HTV using Dataset 2 and 4 (continued)

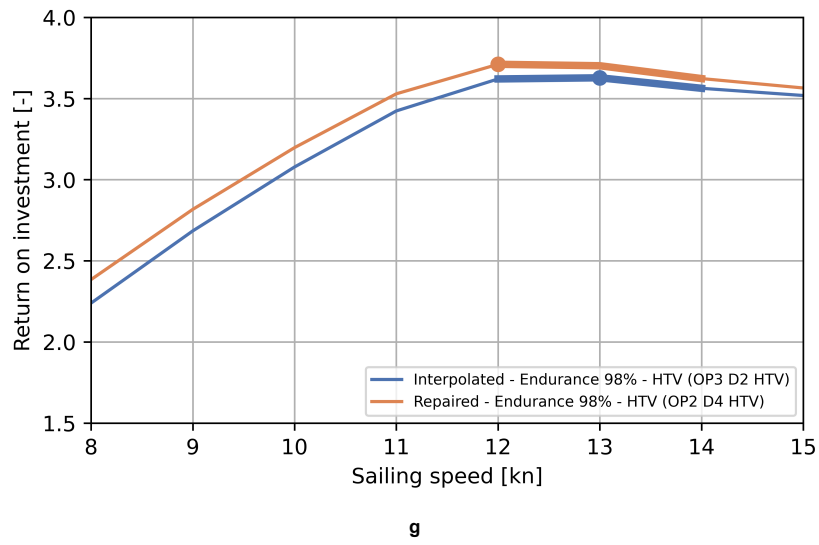


Figure J.3: Comparison between Blended Design results of HTV using Dataset 2 and 4 (continued)

K

Monopile Fabrication Sites

In Table K.1, an overview of the monopile fabrication sites considered in this research is provided. For each site, polygonal boundaries were defined using Cartographer by MSG [4]. These polygons were subsequently used to determine whether a vessel visited a fabrication site.

Table K.1: Monopile Fabrication Site

Port	Coordinates	Port	Coordinates
CWHI Qinzhou Yard [50]	21.724935° N, 108.644228° E 21.726530° N, 108.637808° E 21.726912° N, 108.635989° E 21.721884° N, 108.633350° E 21.711294° N, 108.632316° E 21.708955° N, 108.647662° E 21.723436° N, 108.650923° E	Dajin Penglai [51]	37.832899° N, 120.894854° E 37.815887° N, 120.895285° E 37.816077° N, 120.905413° E 37.830981° N, 120.907932° E
TSP Blue Island [52]	31.690262° N, 121.846285° E 31.690058° N, 121.854181° E 31.680154° N, 121.854902° E 31.680738° N, 121.841135° E	Dajin Yangjian [51]	21.703775° N, 111.821310° E 21.692514° N, 111.829103° E 21.698830° N, 111.845411° E 21.711972° N, 111.837824° E
CS Wind Aalborg [53]	57.056795° N, 10.031122° E 57.048691° N, 10.047052° E 57.056944° N, 10.058313° E 57.064859° N, 10.042795° E	EEW New Jersey [54]	39.855132° N, 75.243948° W 39.840793° N, 75.236944° W 39.844800° N, 75.224172° W 39.859770° N, 75.232138° W
CS Wind Esjberg [53]	55.462409° N, 8.454354° E 55.443329° N, 8.446938° E 55.434214° N, 8.483331° E 55.456102° N, 8.489236° E	EEW Rostock [54]	54.158938° N, 12.116359° E 54.140680° N, 12.127071° E 54.142370° N, 12.139980° E 54.160787° N, 12.126522° E
CS Wind Odense [53]	55.467256° N, 10.515241° E 55.456979° N, 10.536939° E 55.482045° N, 10.571271° E 55.491071° N, 10.546003° E	Haizea Bilbao [55]	43.346269° N, 3.057235° W 43.360448° N, 3.041717° W 43.365341° N, 3.058951° W 43.364941° N, 3.074332° W 43.354657° N, 3.076392° W
CWHI Xiangshui [50]	34.435522° N, 119.817227° E 34.447358° N, 119.822274° E 34.450925° N, 119.802567° E 34.437872° N, 119.803803° E	Haizea Brest [55]	48.388792° N, 4.442779° W 48.379216° N, 4.451911° W 48.374427° N, 4.442024° W 48.381952° N, 4.435981° W
CWHI Zhongshan [50]	22.553542° N, 113.615428° E 22.555127° N, 113.626963° E 22.567043° N, 113.621117° E 22.572810° N, 113.618098° E 22.568760° N, 113.608149° E	Navantia Methill [56]	56.180341° N, 3.011179° W 56.181469° N, 3.015367° W 56.179711° N, 3.019521° W 56.176290° N, 3.022954° W 56.173538° N, 3.018182° W
Century Wind Power [57]	25.153432° N, 121.369264° E 25.151262° N, 121.363991° E 25.150128° N, 121.360867° E 25.139810° N, 121.366377° E 25.143306° N, 121.375149° E 25.149180° N, 121.371561° E	SIF Maasvlakte [58]	51.961761° N, 4.014833° E 51.973622° N, 4.013994° E 51.974341° N, 3.997377° E 51.961311° N, 3.999643° E
Steelwind Nordenham [59]	53.520332° N, 8.546888° E 53.524210° N, 8.555540° E 53.527877° N, 8.551789° E 53.522863° N, 8.544176° E	SIF Roermond [58]	51.200979° N, 5.984854° E 51.204528° N, 5.990724° E 51.201796° N, 5.996080° E 51.203624° N, 5.999720° E 51.206378° N, 5.994295° E 51.206313° N, 5.990553° E 51.201710° N, 5.983755° E
Titan Wind Cuxhaven [60]	53.842815° N, 8.755471° E 53.847615° N, 8.759969° E 53.855595° N, 8.737515° E 53.849843° N, 8.734460° E	Windar Navantia Fene [61]	43.475891° N, 8.173981° W 43.480675° N, 8.181878° W 43.475842° N, 8.193688° W 43.470909° N, 8.193825° W 43.468068° N, 8.186547° W

