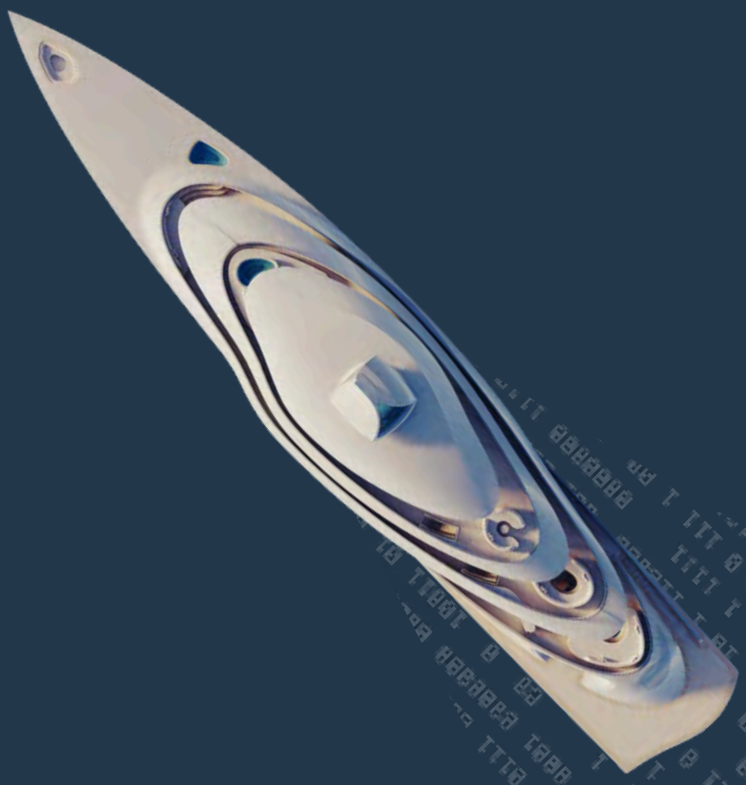
Towards Improved Resistance Predictions for Twin-Screwed Superyachts

Physical, data-driven and hybrid approaches

S. L. Duister



Delft University of Technology

TU Delft

This page has been intentionally left blank.

Towards Improved Resistance Predictions for Twin-Screwed Superyachts

Physical, data-driven and hybrid approaches

Master of Science Thesis
For the degree of MSc in Marine Technology
in the specialization of Ship Design

by

S. L. Duister

Student number:	5398746	
Report number:	MT.24/25.21.M	
Publish date:	December 13, 2024	
Committee:	Dr. ir. A. Coraddu	*Chair/responsible professor*
	Ing. G. Loeff	*Daily company supervisor*
	Dr. ir. B. Font	*Staff member*

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft FEADSHIP ROYAL DUTCH SHIPYARDS

Delft University of Technology
Faculty of Mechanical Engineering

The undersigned hereby certify that they have read and recommend to the Faculty of Mechanical Engineering for acceptance the thesis entitled **"Towards Improved Resistance Predictions for Twin-Screwed Superyachts"** by **S.L. Duister** in partial fulfilment of the requirements for the degree of **Master of Science**.

GRADUATION COMMITTEE

Dated: December 13, 2024

Committee Chair:

Dr. ir. A. Coraddu

Committee members:

Ing. G. Loeff

Dr. ir. B. Font

Preface

I've always been curious about understanding why things work the way they do, starting very young by watching and "helping" my grandfather with the endless renovation projects for the whole family. This curiosity only grew with the purchase of the most unreliable 50cc 2-stroke motorcycle, which seemed to be broken more often than fixed. Together with this curiosity, and the engineering influences of my dad and uncle, the decision to pursue mechanical engineering was not the toughest one to make. Though, the realization that naval architecture might be something for me came quite late, it feels. Not only that, in the most recent years of my studies I have also found "a drive" that was really fuelled when I started living in Delft, where being surrounded by ambitious people is the status quo. For these reasons, I feel grateful to finally present the work to obtain the Master of Science in Marine Technology, marking the end of my academic journey and the beginning of a new chapter.

I want to take this preface as an opportunity to express my gratitude to various individuals without whom this thesis would not have been possible. First and foremost, I would like to thank my company supervisor. Thank you, Giedo, for your unwavering guidance and infectious enthusiasm and energy. From day one, I was struck by your openness to exploring the relevance of the thesis project and by the freedom you granted me to carve my own path - something I deeply value and appreciate. I have learned so much from you, not only as a naval architect, but also as a person.

Next, I would like to extend my heartfelt thanks to my professor, Andrea. I clearly remember our very first meeting when you thoughtfully raised the question of whether starting a thesis in an entirely new and unfamiliar domain like machine learning was a wise choice. It was at this exact moment that I knew you would be the perfect professor for me. Andrea, thank you for taking a leap of faith and for your willingness to share your invaluable skills with me. I truly felt that you valued my learning curve, something for which I am deeply grateful.

I would also like to express my gratitude to a group of people at Feadship, who played a significant role in helping me complete my thesis. To my colleagues at the R&D department - Marc, Douwe, Martijn, Abdel-Ali and Berend - thank you for your readiness to support me in every possible way and the unforgettable laughs we shared in the office. To my fellow students - Jaap, Andrea, Leonardo, Anton, Feico, and Josha - thank you for the countless debates, coffee chats, and shared lunches that made this journey all the more enjoyable. Finally, I extend my heartfelt thanks to all my colleagues at Feadship for warmly welcoming me and offering support throughout my journey.

And from time to time, I am reminded of the importance of my family and good friends. Thank you for your support, love, and warmth, which have been my foundation throughout my academic path. A special thanks to my incredible roommates - Thomas, Tom, Mike, Arend-Jan, and Ryan - for your genuine interest, inspiration, and for keeping me grounded and sane during this project. Your presence has made all the difference.

Thank you all for being a part of this journey.

Sietse Levi Duister
Delft, December 13, 2024

Abstract

For many shipbuilders, the majority of a vessel's lifecycle greenhouse gas emissions occurs during its operational phase, commonly referred to as downstream emissions. For Feadship, this challenge is particularly pronounced, with downstream emissions accounting for 94% of a superyacht's carbon footprint. Addressing this majority requires accurate and efficient methods to predict propulsion energy use during the design stage - a task hindered by the limitations of existing prediction methods, which are either time-intensive or prone to significant errors. This is affecting the design process in two ways: an increased risk of incorrectly proportioned energy and power systems, and limited exploration of design space. Data-driven methods, based on machine learning algorithms, have been proposed in the literature. However, these methods expose two key gaps in the literature: their performance under extrapolation conditions and their limitations when applied to small datasets.

This thesis addresses these challenges by developing hybrid modeling approaches that combine physical insights through a physical model with data-driven techniques, enabling improved predictive accuracy under extrapolation scenarios and with small datasets. Three modeling approaches are tested - physical models (PMs), data-driven models (DDMs), and a combination of the two that forms hybrid models (HMs) - with a shared prediction target, namely calm-water ship resistance. Four datasets were assessed: Dataset 1 (CFD resistance), Dataset 2 (CFD power), Dataset 3 (towing tank tests), and Dataset 4 (speed-power trials), resulting in the selection of computational fluid dynamics (CFD) resistance data as the basis for training the models.

Instead of directly learning from the CFD resistance data, it appears more effective when the data-driven model learns to apply corrections to the output of a PM. Where traditionally these corrections were based on the naval architect's experience, they are now driven by data, offering fast and accurate alternative to existing methods. This philosophy is embodied in this study through a newly developed parallel HM, which achieves superior performance by learning how to apply these corrections to the PM's output automatically. During interpolation, the new HM demonstrates a mean average percentage error (MAPE) of 3.8%, outperforming the best available PM (6.7%) and the best DDM (8.9%). For extrapolation, standalone DDMs, including the best interpolator, failed dramatically, with MAPE values exceeding 180% in some cases. The new HM maintains average errors within 12% across scenarios. And with less data, the new HM consistently outperformed the best DDM, with its competitive edge most pronounced at low data availability (10% of CFD observations). By advancing these methodologies, the study not only enhances early-stage design confidence but also contributes to future steps towards automated design optimization.

Contents

List of Figures

List of Tables

Nomenclature

Symbols and Variables

ℓ_{cB}	Longitudinal center of buoyancy	(%)
ℓ_{cF}	Longitudinal center of floatation	(%)
∇	Displacement volume	(m^3)
A_t	Transom area	(m^2)
A_{bt}	Bulbous bow transverse area	(m^2)
A_{wp}	Water plane area	(m^2)
B/T	Beam-to-draft ratio	(-)
B_{mld}	Moulded beam	(m)
C_b	Block coefficient	(-)
C_m	Midship section coefficient	(-)
C_p	Prismatic coefficient	(-)
C_{wp}	Water plane area coefficient	(-)
D_t	Transom immersion at centerline	(m)
F_n	Froude number	(-)
h_{bt}	Height of bulb centroid above keel	(m)
i_E	Bow half angle of entrance	(°)
$L/\nabla^{1/3}$	Length-to-displacement ratio	(-)
L/B	Length-to-beam ratio	(-)
L/T	Length-to-draft ratio	(-)
L_{os}	Overall submerged length	(m)
L_{wl}	Waterline length	(m)
S_{app}	Wetted surface area of appendages	(m^2)
$S_{app}/(LT)$	Appendage-to-lateral cross-section ratio	(-)
S_{rudder}	Wetted surface area of rudders	(m^2)
S_{stab}	Wetted surface area of stabilizer fins	(m^2)
S_{tot}	Total wetted surface area	(m^2)
T	Draft	(m)
V_s	Ship speed	(kn)

Acronyms and Abbreviations

R^2	Coefficient of Determination
CFD	Computational Fluid Dynamics
DDM	Data-Driven Model
HM	Hybrid Model
KCV	K-Fold Cross-Validation
LOBO	Leave-One-B/T-Out
LOCO	Leave-One-C_p-Out
LOFO	Leave-One-F_n-Out
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MARIN	Maritime Research Institute Netherlands
PM	Physical Model
RMSE	Root Mean Squared Error

"Learning without thought
is labor lost;
thought without learning
is perilous."

— Confucius

Chapter 1

Introduction

Background - Shipping is increasingly governed by stringent sustainability policies, with the International Maritime Organization (IMO) leading efforts through the implementation of energy efficiency standards like the Energy Efficiency Design Index (EEDI) and the Carbon Intensity Indicator (CII). These standards have become essential for evaluating the energy efficiency of commercial vessels but are less applicable to pleasure craft, such as yachts, which have more varied usage patterns and lower annual mileage than commercial ships. Introduced in 2011, the EEDI assesses CO_2 emissions per ton shipped over a nautical mile, serving as a valuable metric for commercial shipping efficiency. Likewise, the CII, implemented in 2023, measures total CO_2 emissions based on the cargo carried and the distance traveled over a year. However, since yachts are neither designed for transporting goods nor typically used for purely extensive travel, both the EEDI and CII have limited applicability in the yachting sector.

The appetite amongst yachting clients to accelerate the technology developments is increasingly there, but encouragement from regulatory bodies remains essential. And this shift is already underway. In 2021, the IMO introduced Tier III requirements, setting strict NO_x emission limits for superyachts over 500 gross tons in Emission Control Areas (ECAs). Building on this momentum, At the 82nd session of the International Maritime Organization's Marine Environment Protection Committee (MEPC) in October 2024, superyachts' environmental impact was addressed, with SYBAss proposing revised fuel reporting and EEDI/CII calculations - advancing regulations tailored to the sector.

These developments underscore the IMO's expanding commitment to emissions reduction beyond commercial shipping, marking a pivotal shift toward sustainable yachting.

1.1 Motivation

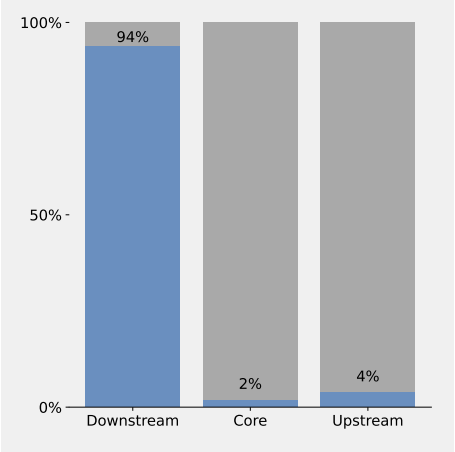


Figure 1.1: Upstream, core-process and downstream share of total company emissions, assuming a ship life cycle of 30 years (Loeff, 2024).

The greenhouse gas (GHG) emissions over the entire lifecycle of a Feadship vessel can be divided into three primary categories: upstream emissions, core emissions, and downstream emissions. Downstream emissions encompass all emissions directly resulting from the ship's operational use occurring post-sale or delivery. Internal research (Loeff, 2024) reveals that this category dominates the vessel's carbon footprint, representing a substantial 94% of total lifecycle emissions (Fig. 1.1).

For many shipbuilders, who observe similar trends (Jacquet et al., 2024) in their ship lifecycle analysis (LCAs), tackling downstream emissions is essential. This can be achieved through two routes: reducing emissions for constant energy use (e.g. by adopting alternative fuels or aftertreatment), or reducing energy use itself (through efficiency improvements). Feadship aims to pursue both pathways; however, challenges arise in accurately estimating propulsion energy use during the design process, which is critical for achieving these goals. An analysis of the discrepancies (Section 2.1) between (time-intensive) high-fidelity approaches and sea trial measurements revealed substantial average errors: over 60% for computational fluid dynamics (CFD) simulations and 26% for towing tank model tests.

This lack of confidence in estimations affects the design process in two key ways: increased risk of incorrectly proportioned energy and power systems and limited exploration of design space. A very illustrative example of the first are the energy densities of future shipping fuels like methanol and liquid hydrogen. In yacht design, methanol requires 2.3 times more storage volume than diesel, while liquid hydrogen demands 36.9 times more (Loeff, 2024). Similar issues arise with the power densities of future power systems (Van Veldhuizen et al., 2024), where solid oxide fuel cells (SOFCs) require 2.8 times

more volume than ICEs in a cruise ship scenario. These examples clearly underscore the growing risks associated with energy use predictions.

The second consequence of either inaccurate or time-intensive prediction methods is the inability to effectively explore the design space in the early stages of development, where reliable and fast predictions are crucial. At Feadship, experienced naval architects have been conducting hull and propulsion optimization studies for years but are constrained by high-fidelity CFD-based methods for these purposes, which require substantial preparation and computational time for each run. This limitation is particularly critical, as in shipbuilding, the first design choice is often the most critical, shaping everything that follows (J. Harvey Evans, 1959).

Data-driven approaches, leveraging machine learning algorithms, present a promising solution to these challenges. They offer the potential to improve propulsion energy use predictions by providing fast and highly tailored insights specifically for twin-screwed superyachts. Moreover, they enable more effective exploration of the design space, facilitating the optimization of hull and propulsion characteristics for future designs.

1.2 Problem statement

The literature review posed two gaps that are currently faced by data-driven models, including the limited performance in extrapolation scenarios and when working with small datasets.

Extrapolation - The ability to make predictions that extend beyond one's experience, is a hallmark of human intelligence. However, data-driven methods frequently struggle with extrapolation, the task of making predictions outside the boundaries of their training data. Numerous of studies (Coraddu et al., 2015; Kalikatzarakis et al., 2023; Leifsson et al., 2008; Mei et al., 2019; Skulstad et al., 2021) underscore the promising potential of hybrid models to boost extrapolation capabilities in domain-specific prediction challenges. However, despite these advancements, most studies fall short in quantifying this performance in real extrapolation scenarios. While this capability is acknowledged, few studies demonstrate substantial improvements under these conditions, even though the potential clearly exists.

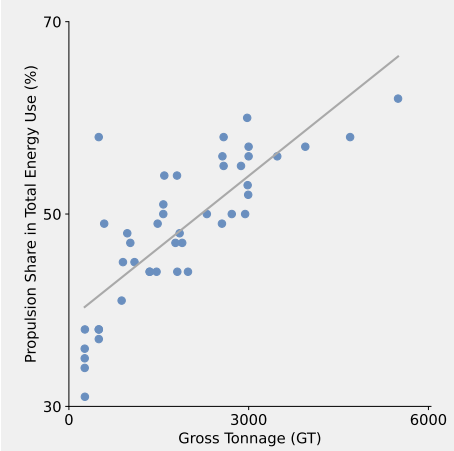


Figure 1.2: Propulsion share with respect to the total energy use of the Feadship fleet, studied by (Loeff, 2024).

Small datasets - In ship design, access to high-volume, high-velocity, and high-variety data - collectively known as Big Data - is often limited. Beyond their potential to improve extrapolation, the previously mentioned literature also emphasize the benefit of hybrid models in reducing data requirements. While this advantage is noted in the literature, quantifying its impact is often overlooked, highlighting a need for further research to address this gap.

As mentioned earlier, the critical focus for this study is propulsion energy use. Typically, energy use is categorized into (i) propulsive and (ii) auxiliary energy use, which together constitute a vessel's total energy consumption. Analysis of the Feadship fleet (Fig. 1.2) reveals that propulsion dominates energy consumption in vessels over approximately 2700 GT. Combined with the errors highlighted in Section 1.1, this underscores the critical need for improved predictions in this area.

1.3 Scope and research questions

The scope of this study is to address gaps in data-driven approaches by developing, testing, and evaluating promising hybrid modeling techniques. These approaches aim to improve propulsion energy use predictions through accurate calm-water resistance estimations, even in extrapolation scenarios with limited data. Each hybrid model combines a physical model (PM) to provide a physically grounded baseline with a data-driven model (DDM) that applies corrections to enhance predictions on unseen data. By integrating PMs and DDMs, the hybrid models leverage both physical understanding and data insights. These models are rigorously tested and compared against state-of-the-art PMs and DDMs to quantify improvements in extrapolation and data efficiency.

Thus, the master thesis centres around the following main research question:

"How can hybrid models improve early-stage predictions for calm-water ship resistance in extrapolation scenarios, while using small datasets with limited design variability?"

Answering the research question necessitates conducting several experiments, as the true model performance of the PMs, DDMs and HMs is unknown for this application. To guide this process, the study is dissected into the following sub-questions:

1. *What specific challenges persist within the Feadship design and engineering process, and what are their main causes?*
2. *What knowledge gaps exist in data-driven methods, and which modeling approaches could address these challenges?*
3. *What datasets are available relating to propulsion use, and which dataset offers the best suitability for this study?*
4. *What methodologies ensure comprehensive training, testing, and evaluation of the models?*
5. *Which of the tested exhibited superior performance in interpolation and extrapolation scenarios?*
6. *How does a reduction in dataset size impact the performance of various models?*
7. *What is the most optimal features set for twin-screwed superyachts, and what are the best methods to do this feature selection?*

1.3.1 Hypotheses

When averaged across all possible problems, no algorithm performs better than another. This idea fundamentally challenges the assumption that a "best" universal algorithm exists for all problem types. Instead, the no-free-lunch theorem from (Adam et al., 2019) suggests that an algorithm's success is context-dependent, performing well on some problems and poorly on others. This context-dependence raises essential questions about how algorithms might be tailored or adapted to achieve optimal results within specific domains, such as calm-water resistance. The following hypotheses are defined:

1. **Challenges** - Feadship aims to reduce downstream emissions but faces two challenges: disproportioned energy and power systems and limited design exploration, driven by inaccurate, time-intensive prediction methods, especially for propulsion energy use (Section 2.1.4).
2. **Modeling approaches** - Suitable approaches for this study fall into three main categories: physical models (PMs), data-driven models (DDMs), and hybrid models (HMs).
3. **Best dataset** - The sea trial dataset is considered the most suitable, as it features the largest data size and the highest design variance among all available datasets.
4. **Training, testing and evaluation** - K-fold cross-validation is appropriate for interpolation tests, whereas extrapolation tests require the exclusion of specific geometries to ensure validity.
5. **Models for interpolation and extrapolation** - For interpolation condition, data-driven models (DDMs), as they can directly learn from the data. For extrapolation condition, parallel hybrid models (HMs) as a slight advantage is presented in literature compared to DDMs and serial HMs.
6. **Less data** - Less data is required for any HM to match the performance of the best DDM, though the reduced requirement is moderate.
7. **Other features** - Add features, add complexity. It is expected that a low number of features will achieve the best model performance.

1.4 Limitations

In advanced analytics, four (or sometimes five) distinct types are commonly recognized, each representing an increasing level of complexity (Section 2.2): descriptive, diagnostic, predictive, prescriptive, and occasionally, visual analytics. While opportunities exist in *prescriptive* analytics - a category that addresses the question *"What should happen?"* - this master's thesis deliberately focuses on *predictive* analytics, which aims to determine focus on the question *"What will happen?"*. Specifically, in the context of calm-water resistance, predictive analytics involves making predictions and identifying key design factors, without explicitly optimizing for the best design choices. Although future steps toward design optimization (prescriptive analytics) are briefly explored in Chapter 6 and Chapter 7, the models developed in this thesis are limited to prediction rather than optimization.

1.5 Thesis structure

This thesis is organised in eight chapters, each focusing on different aspects of the research on developing and testing all prediction models for twin-screwed superyachts. The chapters can be clustered into different sections of this report:

- **Technical overview:** Chapters 1-3
- **Methodology and results:** Chapters 4-5
- **Reflection:** Chapter 6-8

The **Technical overview** (Chapters 1-3) introduces the study's context and data foundation. Chapter 1 outlines the motivation, research questions, and scope. Chapter 2 reviews relevant literature on data analytics and predictive modeling for ship resistance, and Chapter 3 describes the datasets, focusing on selection and enrichment for modeling suitability.

The **Methodology and results** (Chapters 4-5) details model development and testing. Chapter 4 introduces the physical, data-driven, and hybrid models, along with the evaluation and testing pipelines used to assess model performance in interpolation, extrapolation, and limited-data scenarios. Chapter 5 presents test results, comparing model strengths and limitations.

The **Reflection** section (Chapters 6-8) provides a critical analysis of findings and practical applications. Chapter 6 discusses results in light of data and model limitations. Chapter 7 addresses real-world adoption strategies for hybrid modeling, proposing dedicated teams and exploring broader applications. Chapter 8 concludes with recommendations for future research, emphasizing improvements in model accuracy and scalability.

Chapter 2

Literature Review

In this chapter, several key findings from the literature are synthesized to establish the foundation of this study. The initial problem analysis is included in Section 2.1 addresses the design and engineering process, the methods used for propulsion estimations, and the challenges in predicting ship resistance accurately. This analysis informs the exploration of different forms of analytics in Section 2.2, from descriptive to prescriptive, and their roles in ship performance assessment. The chapter then provides a detailed review of predictive modeling approaches (Section 2.3), covering three main frameworks: Physical Models (PMs), Data-Driven Models (DDMs), and Hybrid Models (HMs). A review summary is provided in Section 2.4.

2.1 Problem analysis

This section explores the company and current design and engineering process at Feadship. By examining the company's design processes, existing methods, and their limitations, this analysis lays the foundation for this study.

2.1.1 Company

Feadship's story begins in 1849, when the Akerboom family acquired a modest shipyard along the Dutch coast, focusing on boat construction and repair. This early venture laid the groundwork for future partnerships, including a notable alliance with the Van Lent family in 1927. In 1949, these two family businesses joined forces with the De Vries shipyard - another highly regarded family-run operation - to formally establish the First Export Association of Dutch Shipbuilders (Feadship). This collaboration led to the shared technical office of De Voogt Naval Architecture (DVNA), which became central to Feadship's design excellence. Founded by Henri de Voogt in 1913, DVNA quickly earned respect, with Henri himself winning races in boats he had designed and built. As De Voogt shifted its focus solely to naval architecture, it began close collaborations with the De Vries and Van Lent yards, cementing Feadship's legacy in the yacht-building industry. The launch of the Ventura in 1953 (see Fig. 2.1) is a notable example from this era.

Figure 2.1: Ventura's launch in 1953 at the C. van Lent en Zonen shipyard, which was located in Kaag, Netherlands.

Figure 2.2: Project 821, the world's first hydrogen fuel-cell superyacht (118.8 meter), launched by Feadship in May 2024, which uses hydrogen to power its auxiliary loads at anchor.

Today, Feadship operates four shipyards: two owned by Koninklijke De Vries, located in Aalsmeer and Makkum, and two under Royal Van Lent, based in Amsterdam and on the island of Kaag. Each year, Feadship produces around 5 to 7 yachts, emphasizing its commitment to highly customized and

intricate vessels. In recent years, Feadship has become renowned for building yachts of increasing size and complexity, with several recent projects exceeding 100 meters. A notable example is Project 821, depicted in Fig. 2.2.

2.1.2 Design and engineering process

In new build projects, engineering activities must be executed to support the main project activities of customer acceptance, class approval, procurement, construction, testing, final delivery and after sales of the contracted product. A considerable portion of these tasks is devoted to DVNA, where the design and engineering process is segmented into five stages: (i) concept design, (i) technical/contract design, (iii) basic design and (iv) design monitoring, shown in Fig. 2.3.

Figure 2.3: Design stages of De Voogt Naval Architects (2024)

- **Concept design** - The main purpose of the Concept Design stage is to create client interest through a concept with preliminary main parameters. Studio De Voogt works, in consultation with the Sales department, on the portfolio of Design Prospects for Feadship. A Design Number generally requires an initial assessment conducted by Studio De Voogt, which is considered as a first iteration of the engineering department to identify bottlenecks or to confirm that the associated risks are manageable. This stage investigates the preliminary vessel calculations, such as weight, stability, general arrangements (GEs), powering and additional unique features.
- **Technical / Contract design** - Upon a customer showing interest in a design prospect, the Technical or Contract Design starts and the design number transitions into a project number. This design stage ends with a signed and contracted new build project and aims to develop a feasible design according to the client and yard requirements, without any risks that may invoke major changes at a later stage. During this design stage, the design department supports the sales department in providing the necessary plans, sketches, and renderings to reflect the customer's dreams in both form and sketch. Clients can also bring forth their ideas, designers, and/or architects who have already produced plans. This stage is highly flexible and allows for various possibilities that lead to a final sale. Generally, this early phase of design takes, on average, between four and five months. However, due to the importance of the process, there are no time restrictions.
- **Basic design** - In Basic Design, the design is defined in final parameters according to the client, yard and both global and local statutory approval. The purpose of this stage is to engineer a design until it reaches a Class approval level.
- **Design monitoring** - After Basic Design, the project slowly transitions into detail engineering, where the design monitoring starts. During this stage, a reduced team will be responsible for closely overseeing the detailed engineering and refinement of the design, from the point of initial development until the yacht's delivery. The purpose is to verify and validate engineering and the finished yacht against design intent and performance requirements. Once successfully delivered, the project will be closed and the departments start to draw up their lessons learned.

Despite the appearance of linearity in the process, depicted in Fig. 2.3, a shipbuilding processes is often far from linear. DVNA uses the design spiral approach, illustrated in Fig. 2.4, in their design and engineering process. The purpose of this optimization approach (J. Harvey Evans, 1959), is to assist in organizing the thought process, enabling ship design problems to be solved most effectively. This sequential or point-based approach involves a refinement and iteration cycles until it converges to an optimal or at least single design. Some have argued the effectiveness of the design spiral approach since every iteration takes considerable effort, limiting the number of designs explored in the design space. For this reason, efforts are taken by the company to introduce concurrent design sessions (originating from NASA), which are organized in specially designed meeting rooms to promote interdisciplinary

communication. With the numerous cross-disciplinary interfaces in yacht design, these collaborative sessions facilitate improved decision-making, leading to faster development and higher quality design.

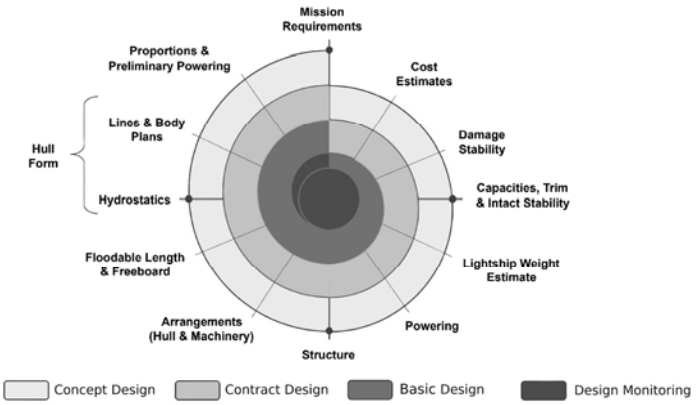


Figure 2.4: Traditional Design Spiral approach used by De Voogt Naval Architects, based on (J. Harvey Evans, 1959)

2.1.3 Methods relating to propulsion use

Typically, energy use is categorized into (i) propulsive and (ii) auxiliary energy use, which together constitute a vessel's total energy consumption. The decision to focus this study on propulsion energy use was driven by its substantial contribution to the yacht's overall energy demand - and, consequently, to downstream emissions - as outlined in Chapter 1 at Fig. 1.2, combined with the availability of relevant data.

Throughout the DVNA design process, propulsion energy use, or ship resistance predictions are typically made in four ways: statistical vessel comparisons, semi-empirical calculations, computational fluid dynamics (CFD) simulations and model tests in towing tanks by MARIN. Table 2.1 clarifies the placement of each prediction method within the design stages of DVNA, as illustrated in, shown in Fig. 2.3.

Table 2.1: Focus of every design and engineering stage, including a brief description and the fidelity level.

Design Stage	Description	Fidelity
Concept Design	Semi-empirical estimations and statistical vessel comparisons (admiralty or heickel coefficients) based on the existing fleet are used to evaluate the 1^{st} approximation of thrust power demand.	Low
Technical/Contract Design	As soon as a hull lines plan of the design prospect is available, preliminary CFD evaluations are performed to determine the initial bare-hull and appendage resistance. These evaluations consider ideal calm water condition assumptions.	Low-Medium
Basic Design	If the design prospect meets the preliminary specification, a CFD iteration process is started. These evaluations provide detailed resistance estimations, including trim and shaft line optimizations, a final wake-field analysis, and, if required, third-party model tests in conjunction with internal CFD analysis.	Medium-High
Design Monitoring	Once the basic design is considered final, parameters can be defined for validation during sea trials as input for the yard sea-trial protocol, including speed/shaft power, fuel consumption, vibration performance, and others. Sea trials are witnessed with potential troubleshooting. Validated results are reported and used for future analysis.	High

While each Feadship is highly customized, its main dimensions sometimes align with those of previously built Feadship yachts. In such cases, direct database comparisons are used for initial estimates. An internal resistance curve database is developed and is continually maintained within Studio De Voogt. The platform collects existing powering and resistance curves from all previous Feadship model tests and sea trials. This database is used to compare and analyse similar vessel trends as seen from collected data. Ultimately, it is a tool to provide designers and naval architects with an initial baseline of how similar a vessel may behave. Coupled to that, the Admiralty (AC) or Heickel coefficients (HC) are employed to provide a first estimate of the propulsion efficiency. While this is the intended use, sometimes these simple formulae, depicted in Eq. (2.1) and Eq. (2.2), are used to provide an estimate for similar-shaped vessels with differing displacement. In this case, an assumption is made for either the (AC) or (HC). Notably, the HC tends to provide better estimates for superyachts, making it particularly useful in preliminary design stages.

$$AC = \frac{D^{2/3} \cdot V^3}{P} \tag{2.1}$$

$$HC = \frac{P}{D \cdot V^3} \tag{2.2}$$

where D is the displacement (tons or m^3), V is the ship speed (knots), and P is the power required (kW).

YACHT is an internally based software that uses the Marine Research Institute Netherlands (MARIN) DESP method, commonly known as the (Holtrop & Mennen, 1984) method, correlated on Feadships. This method provides a relatively good approximation for Feadships because the fleet is primarily composed of displacement vessels to which the approach is centred. Ultimately, this tool provides a quick estimation method with a relatively moderate accuracy for new vessel which fall within the data coverage range.

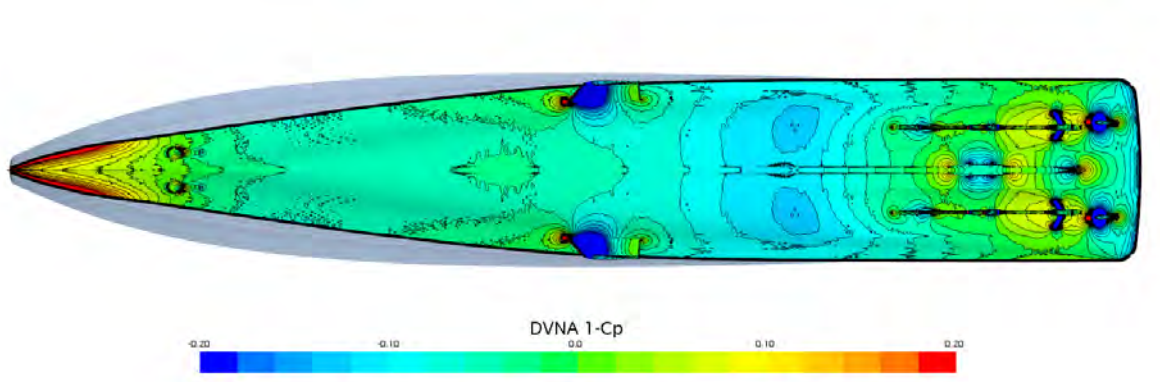


Figure 2.5: Example output from a computational fluid dynamics (CFD) simulation, illustrating the normalized pressure coefficient distribution across the yacht's entire wetted surface.

Computational fluid dynamics (CFD) or numerical simulation is employed to estimate total ship resistance, which comprises frictional and pressure resistance. Advances in computational power have made these calculations feasible even in the early stages of design. CFD allows for relatively efficient assessments, including bare-hull resistance, propeller-hull interactions, and the optimization of added appendage resistance. While the simulations, depicted in Fig. 2.5, themselves are relatively fast (e.g., approximately 2 hours for a bare-hull analysis), it is important to note that this does not account for the time required for geometry preparation, simulation setup, or post-processing of results. This process still remains time-intensive.

Finally, model testing, which is reserved for Feadship yachts with designs that push beyond existing parameter limits, involve unique parameter combinations, or introduce novel features, such as alternative propulsion configurations, appendages, or the addition of a bulbous bow.

Table 2.2 provides a detailed comparison of the existing methods, highlighting significant variations in their preparation and computational requirements. Additionally, each method has specific limitations that need to be addressed.

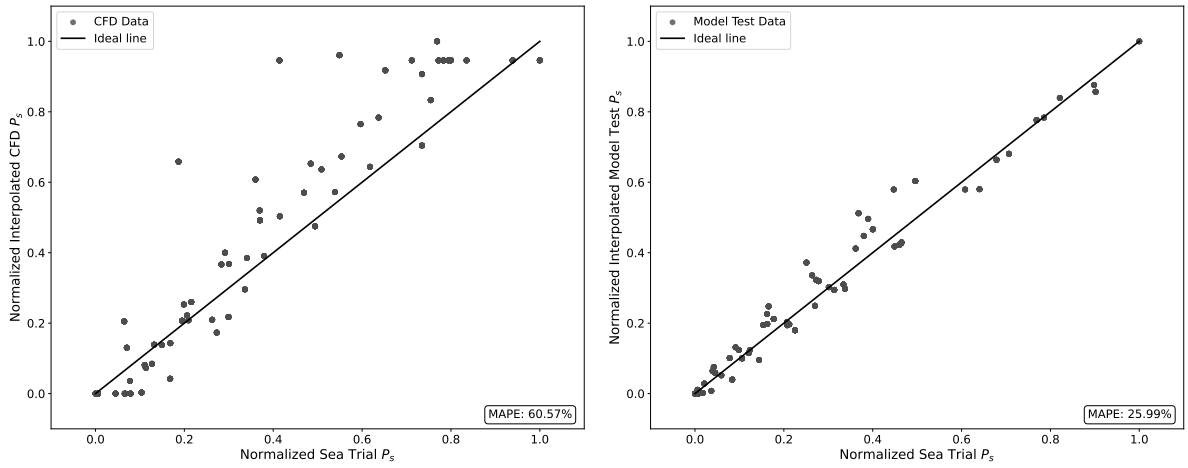
Table 2.2: Efforts and limitations associated with each existing method for estimating propulsion power.

Method	Preparation time (per geometry)	Simulation time (per run)	Limitations
Statistical vessel comparisons	Minutes	Instant	Limited to similar designs; low accuracy for hulls with novel features or unconventional proportions; allows for implicit margins by user, which compromise consistency across predictions.
YACHT simulation	Minutes	Instant	Regression method for the "average" displacement hull, with moderate accuracy on yacht geometries; estimations required for propulsion efficiencies; allows for implicit margins by user.
CFD simulation	Days	2 hours	Computationally expensive; requires high level of expertise to set up simulation; susceptible to discretization and numerical errors; estimations required for some propulsion efficiencies.
Model test	Weeks	1–2 days	Extremely time- and resource-intensive; scale effects can distort full-scale predictions; dependent on external expertise and time-planning; highly-restrict further geometry development after simulation.

2.1.4 Quantification of challenges

The scatter plots in Fig. 2.6 highlight the discrepancies between predicted and actual propulsion power values for sea trial conditions. It should be pointed out that the results for both the CFD and model tests had to be matched with the same speed at which the shaft power was measured during sea trials. The interpolation method (linear) inevitably introduces errors, and this should be taken into account when interpreting the results below. Additionally, discrepancies exist between draught estimations during design stages and actual draught at delivery. This obviously introduces errors in resistance predictions as well.

Also, the accuracy and consistency of these trials can be questioned, done in (Insel, 2008; Seo & Oh, 2021), as uncertainties are introduced by either the trial measurements itself or by post-correction methods. Industry standards like *ISO 15016: 2015* present guidelines, but these have changed significantly over the last decades, leading to inconsistency across data.



(a) Actual Sea Trial P_s (kW) vs. Predicted CFD P_s (kW)

(b) Actual Sea Trial P_s (kW) vs. Predicted Model test P_s (kW)

Figure 2.6: Current challenges with making predictions for required propulsion (shaft) power at sea trials conditions. These scatter plots both show clear discrepancies between estimations and real-world measurements.

Despite the inherent error sources, discrepancies can be analyzed by plotting the predicted propulsion power (P_s) against the actual sea trial values, as shown in Fig. 2.6(a) and Fig. 2.6(b). These plots offer a clear visual comparison of prediction accuracy by highlighting deviations from the ideal agreement line, enabling the identification of systematic biases or errors. Notably, this analysis is independent of the input parameters used, focusing exclusively on the relationship between predictions and observed values.

Figure 2.6(a) highlights substantial deviations from the ideal agreement line, accompanied by a remarkably high mean absolute percentage error (MAPE) of 60.57%, underscoring significant inaccuracies. Excluding the aforementioned sources of error, the remaining discrepancies could stem from two primary factors: (i) inaccuracies within the CFD simulation itself - modelling, discretization, iteration and programming/user errors (Ferziger & Peric, 2012) - or (ii) errors introduced during the conversion from effective power (P_e) to shaft power (P_s) via empirical propeller efficiencies. Understanding and estimating the magnitude of all errors is considered very challenging.

Figure 2.6(b) illustrates the discrepancies of model tests values and the actual sea trial values. In this case, the data points show closer alignment with the ideal agreement line, indicating better predictive performance. The reduced MAPE of 25.99% underscores the improved accuracy of the model test-based predictions compared to those derived from CFD simulations.

Overall, these results emphasize the challenges of current predictive methods for estimating propulsion power, even for high-fidelity methods like CFD-based methods and model tests, highlighting the need for improved modeling approaches.

2.2 Advanced data analytics

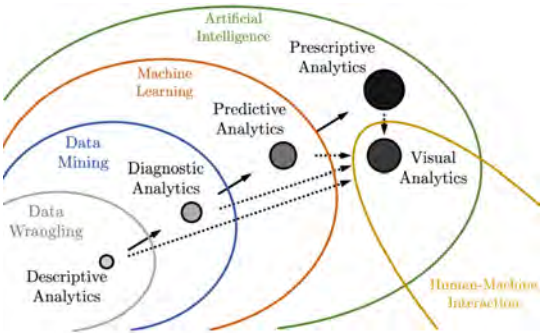


Figure 2.7: Types of data analytics in ship energy systems (Coraddu, Kalikatzarakis, et al., 2022).

Data analytics is the field of research that studies how to exploit data to derive new, meaningful, and actionable information (Coraddu, Kalikatzarakis, et al., 2022). Given the numerous sub-areas within this field, it's essential to clarify the specific scope and focus of each one. In particular, data analytics enables answers to four fundamental questions, which are illustrated in Fig. 2.7, highlighting the unique contributions and distinctions between these areas.

Descriptive Analytics - This approach addresses the question, *"What happened?"* by summarizing historical data to provide an overview of past events and performance metrics. Descriptive analytics offers valuable insights into the status of systems or processes over time. By organizing raw data into accessible information, it aids in the initial stages of analysis and understanding of system behaviour.

Diagnostic Analytics - The diagnostic approach seeks to answer *"Why did it happen?"* by examining historical data to uncover the underlying causes of specific events or patterns. This type of analytics uses statistical techniques to investigate relationships within data, helping identify root causes and contributing factors. Diagnostic insights are essential for understanding deviations from expected performance, whether due to inefficiencies, anomalies, or external influences.

Predictive Analytics - Focusing on *"What will happen?"*, predictive analytics uses statistical and machine learning models to anticipate future events and trends. By identifying patterns within historical data, predictive models estimate likely outcomes and forecast future conditions. This approach helps anticipate issues before they arise, enabling proactive planning and decision-making based on data-driven predictions.

Prescriptive Analytics - Prescriptive analytics addresses the question *"What should be done?"* by providing recommendations for future actions to achieve specific goals. Through the integration of data insights, optimization algorithms, and simulations, this approach suggests actionable steps to enhance performance, efficiency, or outcomes. Prescriptive analytics not only anticipates possible scenarios but also prescribes actions to achieve desired objectives within given constraints.

2.3 Predictive modeling approaches

State-of-the-art computational tools for predictive purposes are based on three main approaches: Physical Models (PMs), Data Driven Models (DDMs), and Hybrid Models (HMs). A brief summary is given here, but in Section 2.3.1, Section 2.3.2 and Section 2.3.3, a thorough review is presented:

- **PMs** rely on the knowledge of the phenomena and can be further subdivided into two main families:
  - Empirical and semi-empirical models (see Section 2.3.1) utilize empirical formulas to approximate with different levels of accuracy the physical phenomena, fine-tuned by means of measurement data. These models are computationally efficient, but usually not enough accurate to be exploited at design stage.
  - Computational Fluid Dynamics (CFD) models simulate fluid flow around geometry by numerically solving fluid dynamics equations, using advanced turbulence and multiphase models to predict complex interactions, such as resistance and propulsion. While able to analyse flow patterns, wave formation, and boundary layers, CFD models are computationally intensive, often limiting their use to later design stages.
- **DDMs** (see Section 2.3.2) rely on Machine Learning (ML) and historical observations to build models of the phenomena with no prior physical knowledge about them (Coraddu et al., 2017). While DDMs can be quite computationally expensive during the model creation phase, they can be highly accurate and computationally inexpensive during the prediction phase. DDMs main limitation lies on their accuracy, which is high on average but not pointwise. Therefore, in some cases, DDMs can provide physically inconsistent predictions (Coraddu, Oneto, et al., 2022).
  - Supervised DDMs require implicit or explicit handcrafting of the features to be able to achieve good recognition performance.
  - Unsupervised DDMs are able to automatically learn features directly from the data, without any labeling.
- **HMs** (see Section 2.3.3) leverage both PMs and DDMs. By combining them, they take advantage of their strengths while limiting their weaknesses (Kalikatzarakis et al., 2023). Specifically, HMs can achieve the same or higher accuracy with respect to DDMs (fully leveraging historical data), but they also leverage prior physical knowledge by exploiting computationally efficient outputs or partial computations behind PMs) to deliver physically plausible results (Coraddu et al., 2018).
  - Serial approaches include both a DDM and PM, where the prediction of the DDM is dependent on the PM's output.
  - Parallel configuration include a DDM and PM, where the PM and DDM make independent predictions, which are later aggregated by means of fusion techniques (summation, averaging or recursion).

2.3.1 Physical models

Previously, two potential approaches for utilizing physical models were outlined: (semi-) empirical models and computational fluid dynamics (CFD)-based models. This study focuses on the former, examining the application of (semi-) empirical models in detail. These methods always incorporate physical principles for key components such as viscous and wave-making resistance, while using empirical results of model tests on methodical series of hull forms, random model tests or data collected from full scale speed trials.

It will be no surprise to see that, in general, at the very first stages of the design process, statistical/empirical methods are the most appropriate tools to be used for the prediction of still or calm-water resistance. At this early stage of the design, the hull form is not yet defined and even main particulars and shape coefficients are not always at their final value. The methods concerned are usually fully computerised and able to study effects of changes of main dimensions and hull form coefficients can quickly be surveyed.

There are many of such methods available. By their nature, each method will have its balance between general applicability and accuracy. Dedicated methods for restricted class of ships may have a

somewhat higher accuracy for one specific category, but near the bounds of the parameter ranges, these methods are inclined to become progressively inaccurate.

One of these earlier methods by (De Groot, 1955) was employed more than half a century ago at the "Nederlandsch Scheepsbouwkundig Proefstation" (NSMB), which later became MARIN, netherland's biggest maritime research institute as of today. The method has been published in the periodical "Schip en Werf" of March 2 and 16 in 1951 and is a very simple procedure for the prediction of the resistance of fast ships. In De Groot's method, the distinction is made between "round-bilge" (depicted in Fig. 2.8) and "hard-chine" craft. Depending on the design speed and main dimensions, a certain preference for either of the two types can be indicated. In De Groot's method, ship resistance is determined using only three key parameters - speed, length, and displacement - which, while clearly limited in scope, represent an ingenious, simple and effective approach for the period.

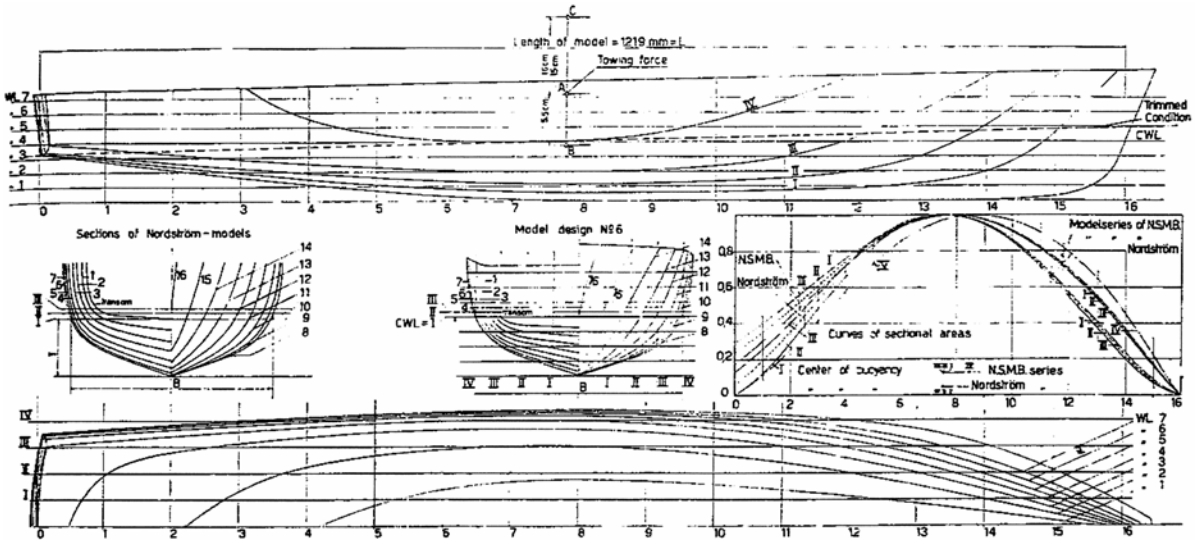


Figure 2.8: Lines drawing of the model series of U-form motor-boats of the NSMB, which are based on the body-plan of the Nordström series. Four U-forms are tested at the Delft University of Technology (De Groot, 1955).

While De Groot's method provided a practical and accessible empirical model, mainly for motorboats, the work of (Van Oortmerssen, 1971) advances this by creating a mathematically derived power prediction model that can be computerized. This model distinguishes between viscous and wave-making resistance and includes parameters such as the Froude and Reynolds numbers. Van Oortmerssen's method not only improves prediction accuracy for small ships like tugs and trawlers but also accommodates programming, by using a polynomial-based structure, allowing for efficient computational use.

The (Guldhammer & Harvald, 1974) method enhances Van Oortmerssen's polynomial-based model, which required coefficients to be manually input and tailored for each application, while Guldhammer and Harvald refined this by developing a set of regression-based formulas. Their method was applicable to a wider range of hull types and incorporating specific corrections for features like bulbous bows, appendages (e.g., rudders, bilge keels), and longitudinal center of buoyancy variations.

The Delft Systematic Yacht Hull Series (DSYHS) by (Gerritsma et al., 1981) introduces critical advancements tailored to yacht design, in particular commercial sailing yachts. The systematic model tests offered a specialized dataset for predicting resistance and stability under both upright and heeled conditions. Additionally, the method captured the impact of critical design parameters length-displacement ratio and beam-draught ratio. This unique focus enables accurate speed-power predictions and stability assessments essential for performance optimization in competitive yacht design, filling a gap in existing resistance prediction methods. Table 2.3 shows a systematic review of existing literature on PMs.

(Holtrop & Mennen, 1984; Holtrop & Mennen, 1982) further advanced the field by decomposing resistance into separate components, such as wave-making, frictional, appendage resistance and more. This detailed breakdown allows for precise adjustments based on specific hull features and configurations, such as bulbous bows and stern shapes, improving prediction accuracy across various

Table 2.3: Overview of PMs review.

Ref.	Scope(s)	Data	Inputs	Outputs	ID	Unit	Permissible range
(De Groot, 1955)	V-shaped and U-shaped motorboats	Model tests at N.S.M.B (unknown), Delft University (4 U-forms) and Stevens Institute (20 V-forms)	Appendix A	Appendix A	$V_s/\sqrt{L}$ C_p B/T	- - -	$\leq$ 2.5 for U-form 0.576 ... 0.756 1.5 ... 3.5
(Van Oortmerssen, 1971)	Small ships	Model tests at N.S.M.B of 93 tugs/trawlers for resistance analysis, 66 models for propulsion analysis	Appendix A	Appendix A	F_n L_{wl} V L/B B/T C_p C_m ℓ_{CB} i_e	- m m^3 - - - - % °	$\leq$ 0.5 8 ... 80 5 ... 3000 3 ... 6.2 1.9 ... 4.0 0.5 ... 0.73 0.70 ... 0.97 -7.0 ... +2.8 10 ... 46
(Guldhammer & Harvald, 1974)	Single and twin screw displacement vessels	Model tests and full-scale trials (unknown)	Appendix A	Appendix A	F_n C_b L/B $L/\sqrt[3]{V}$	- - - -	$\leq$ 0.36 0.55 ... 0.85 5.0 ... 8.0 4.0 ... 8.0
(Gerritsma et al., 1981)	Commercial sailing yachts	Model tests at Delft University with 22 hull variations	Appendix A	Appendix A	C_p L/B $L/\sqrt[3]{\Delta}$	- - -	0.54 ... 0.60 2.5 ... 3.5 4.5 ... 6.0
(Holtrop & Mennen, 1982)	Single and twin screw displacement vessels	Model tests at MARIN and full-scale trials	Appendix A	Appendix A	F_n C_p L/B	- - -	$\leq$ 0.45 0.55 ... 0.85 3.9 ... 9.5
(Holtrop & Mennen, 1984)	Single and twin screw displacement vessels	Model tests of 334 models at MARIN and full-scale trials	Appendix A	Appendix A	F_n C_p L/B	- - -	$\leq$ 0.55 0.55 ... 0.85 3.9 ... 9.5
(Hollenbach, 1998)	Single and twin screw displacement vessels	Model tests of 433 models at Vienna Model Basin	Appendix A	Appendix A	F_n L_{pp} L_{pp}/B B/T D/T_a C_b $L_{pp}/\sqrt[3]{V}$	- m - - - - -	$\leq$ 0.5 30.6 ... 206.8 3.96 ... 7.11 2.31 ... 6.11 0.5 ... 0.86 0.51 ... 0.83 4.41 ... 7.27

vessel types. The method's adaptability extends to high-block coefficient ships, slender naval vessels, and unconventional hull forms, offering a broader applicability than earlier models. Moreover, with refinements in Code 7, such as enhanced treatment of air resistance and hull roughness, the method achieved greater versatility and accuracy, especially for higher-speed designs, making it one of the most respected ship performance approximation methods till this day.

The (Hollenbach, 1998) method was developed from a comprehensive dataset of over 700 resistance and propulsion tests, primarily sourced from the Vienna Model Basin. Building on earlier methods like those of Holtrop and Mennen, Hollenbach's model is particularly valued for its reliability in predicting resistance for twin-screw vessels, offering both upper and lower bounds for resistance estimates. This feature provides designers with a more robust framework for accurately assessing resistance under varied operational conditions.

2.3.2 Data-driven models

While all the PMs discussed in Section 2.3.1 rely fundamentally on essential physical principles, it's important to highlight that they also incorporate data-driven techniques to a significant extent. Though, since these foundational studies (De Groot, 1955; Gerritsma et al., 1981; Guldhammer & Harvald, 1974; Hollenbach, 1998; Holtrop & Mennen, 1984; Holtrop & Mennen, 1982; Van Oortmerssen, 1971), substantial advancements have been made in the field of data-driven approaches. These developments have expanded the potential of predictive modeling by improving accuracy, enhancing computational efficiency, and broadening applicability across a wider array of domain-specific problems.

In regression problems, where one likes to forecast a numerical value for a new set of input values, data-driven approaches can be categorized in two paradigms: supervised learning and unsupervised

learning. (Shalev-Shwartz & Ben-David, 2014) (Goodfellow et al., 2016). Supervised DDMs usually require implicit or explicit handcrafting of the features to be able to achieve good recognition performance (Shalev-Shwartz & Ben-David, 2014) (Duboue, 2020). Usually, this feature set is designed based on either domain knowledge or classical signal processing techniques.

Unsupervised DDMs instead, are able to automatically learn features directly from the data (Goodfellow et al., 2016) and over-perform state-of-the-art supervised models (and in some cases also humans) in terms of recognition performance in many different applications. Unfortunately, unsupervised DDMs have also three main weaknesses. First, they require a huge number of samples to be trained on. In this thesis, the datasets have less than 500 samples in the simplest scenarios, namely interpolation, and much fewer samples in complex scenarios, namely extrapolation. The second problem is that unsupervised DDMs are very hard to interpret. It is complex to understand what they actually learned from the data, resulting in models not useful for practical applications, where insights on the problem need to be extracted (Molnar, 2020). Finally, unsupervised DDMs are seldom able to give physically plausible prediction, see for example, the well-known problem of the adversial samples (Biggio & Roli, 2018), where specially crafted inputs are designed to cause a machine learning model to make a mistake.

This study focuses on regression techniques in the supervised learning paradigm. Table 2.4 summarizes the latest applications of supervised DDMs in the field. A fairly limited amount of literature work is reviewed here as most state-of-the-art DDMs are already being utilized in the third modeling approach, hybrid models (see next Section 2.3.3).

Table 2.4 shows a systematic review of existing literature on DDMs.

Table 2.4: Overview of DDMs review.

Ref.	Scope(s)	Data	Method(s)	Parametrization	Performance	Validation	Takeovers
(Mittendorf et al., 2022)	Added-wave resistance prediction in oblique waves	Dataset derived from potential flow methods, covering 18 hull forms	RF, XGB, MLP	5 wave and geometric variables	Errors are approximately 0.75–3.1% across all scenarios	Cross-validation with synthetic data	Tree-based methods show strong performance for capturing non-linear patterns but require carefully designed datasets for generalization.
(Walker et al., 2024)	Yacht hull resistance and optimization	Delft Systematic Yacht Hull Series, 54 different yacht geometries	RF, XGB, KRR, ELM	13 hydrostatic variables	Errors approx. 0.11-0.89% for all scenarios	Leave-one-out testing (LOVO, LOGO, LOSO) and CFD simulations	Best performing surrogate is the KRR method.

2.3.3 Hybrid models

A significant challenge remains in constructing a model that effectively integrates the physical knowledge embedded in the PMs (Section 2.3.1) with hidden insights in available data, as captured by the DDMs (Section 2.3.2). This challenge is extensively highlighted in existing literature; however, many techniques aiming to fuse physical and data-driven models tend to remain relatively simplistic in their approach. Many variations exist, but most can be dissected into two buckets: either a serial configuration (Fig. 2.9a), where the prediction of the DDM is dependent on the PM's output, or a parallel configuration (Fig. 2.9b), where the PM and DDM make independent predictions, which are later aggregated by means of the following fusion techniques:

- **Option 1 - Summation**: The DDM learns to predict the residual (difference between the PM output and the actual value), which is then summed with the PM's prediction to improve accuracy.
- **Option 2 - Averaging**: The DDM learns to predict the absolute target value, and the final prediction is obtained by averaging the DDM and PM outputs (summing them and dividing by 2).
- **Option 3 - Recursion**: The DDM learns to predict the absolute target value, with the fusion process involving multiple recursive steps to minimize the disagreement between the DDM and PM, thereby reducing the prediction error.

One of the first to apply a hybrid model (Fig. 2.9) into the naval architecture domain is (Leifsson et al., 2008), with his study aiming to predict ship speed and/or fuel consumption with operational data

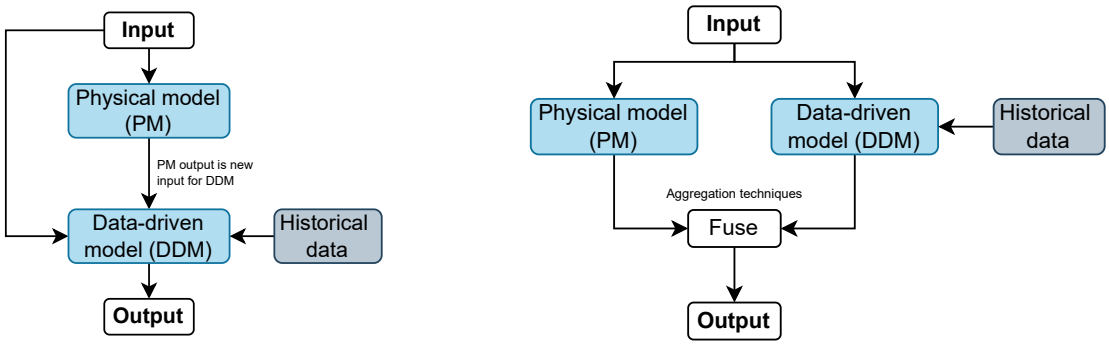
from a container vessel. This work was not extensive enough to fully assess the potential of HMs, nor was the underlying reason for the effectiveness of this approach completely understood. Though, a significant error reduction was achieved for the fuel consumption prediction of almost 65% compared to the used PM.

A novel approach was introduced by (Coraddu et al., 2015) as the PM tries to tune the regularization process of the DDM. Regularization is a technique used in machine learning and statistical modeling to prevent overfitting, which occurs when a model learns the training data too well, capturing noise or random fluctuations instead of the underlying patterns. The results showed improvements for fuel consumption and shaft power predictions, while the HM was capable of achieving similar performance with less data. Two years later, this approach was successfully applied to a different Handymax operational dataset in (Coraddu et al., 2017), alongside a serial approach (Fig. 2.9a). In the majority of experimental scenarios, the former approach demonstrated superior performance.

Other applications are found by (Mei et al., 2019) and (Skulstad et al., 2021), where hybrid models are being used for ship motion (surge, sway and yaw) and ship position predictions respectively. Again, better performance is found for both serial and parallel hybrid models compared to pure PM or DDM predictions.

The most recent studies utilizing hybrid architectures include (Odendaal et al., 2023), which focuses on predicting energy consumption under operational conditions, and (Kalikatzarakis et al., 2023), which aims to improve underwater radiated propeller noise predictions. The latter introduced an innovative fusion approach, the parallel configuration with recursions. In this setup, the PM and DDM operate in parallel, with a recursive process that minimizes the disagreement between the two outputs until convergence.

Table 2.5 shows a systematic review of existing literature on HMs.



(a) Serial hybrid model configuration

(b) Parallel hybrid model configuration

Figure 2.9: Visual representations of hybrid model (HM) configuration posed by review.

2.4 Review Summary

This chapter provided a comprehensive review of the key elements that underpin this study. The design and engineering process at De Voogt Naval Architects was detailed, highlighting the distinct stages - Concept Design, Technical/Contract Design, Basic Design, and Design Monitoring - each with increasing levels of fidelity and reliance on computational methods. Within this context, propulsion power predictions emerge as a critical challenge, as current methods (CFD, and model testing) demonstrate varying levels of accuracy and applicability, particularly in early-stage designs (Section 2.1).

A brief note on advanced data analytics (Section 2.2) was included to establish the appropriate focus for this study, resulting in the adoption of predictive analytics, with an aspiration for future integration of prescriptive analytics.

A review of predictive modeling approaches (Section 2.3) was presented, examining physical models (PMs), data-driven Models (DDMs), and hybrid models (HMs). PMs, such as semi-empirical methods and CFD, offer robustness and physical insight but often lack flexibility and accuracy in novel scenarios.

Table 2.5: Overview of HMs review.

Ref.	Scope(s)	Data	Method(s)	Configuration and fusion technique	Performance HM	Takeovers
(Leifsson et al., 2008)	Fuel consumption and vessel speed prediction	Operational data of Dettifoss container vessel	NN	Serial approach (Fig. 2.9a) and parallel approach (Fig. 2.9b) using residual summation	RMSE reduction between approx. 2.5-65% for all quantities	HMs over-performs DDM for both quantities. Greater variance in data is required for DDM.
(Coraddu et al., 2015)	Fuel consumption and shaft power prediction	Operational data of Panamax chemical/product tanker	RLS	PM tunes the regularization term during DDM's learning process	Errors approx. between 0.57-1.71% for all quantities	HM achieves good results using almost half of the data required by the DDM to reach a similar level of accuracy.
(Coraddu et al., 2017)	Fuel consumption prediction	Operational data of Handymax chemical/product tanker	RLS, LAR, RF	Serial approach (Fig. 2.9a) and novel approach with adapted regularization term	Errors approx. equal to 1.90% for best HM (RF)	Both HMs overperform the best PM and DDMs and help mitigate reliability issues with data log/processing.
(Mei et al., 2019)	Surge, sway and yaw motions prediction	Free-running (time-series) model tests	RF	Serial approach (Fig. 2.9a)	Errors approx. 0.12-0.18 (nRMSE) for all quantities	HM shows high correlation with experimental data. RF algorithm and integration with HM could be enhanced.
(Skulstad et al., 2021)	Ship's position prediction	Operational docking data of RV Gunnerus	LSTM	Parallel approach (Fig. 2.9b) with residual summation	Errors approx. 0.9-4.6 metres ship position	Addition of LSTM algorithm improved PM significantly for time-series data. Methodology can be extended to other domains.
(Odendaal et al., 2023)	Propulsion and auxiliary prediction power	Operational measurements of Feadship yachting fleet	NN	In the serial approach (Fig. 2.9a), output of the PM is used as a new feature for the DDM	Errors approx. 61.6-835.7 RMSE for all quantities	The HM is more effective for extrapolation, showing 20% improvement.
(Kalikatzarakis et al., 2023)	Underwater radiated propeller noise prediction	Cavitation tunnel tests	RF, XGB, KRR, SNN	Parallel approach (Fig. 2.9b) with recursion strategy	Errors approx. 2.8-7.2% for all quantities/extrapolation scenarios	Feature importance and test prior knowledge proofs physical plausibility of HM.

DDMs leverage machine learning to uncover complex data patterns but face limitations in extrapolation and physical plausibility. HMs aim to combine the strengths of both approaches, blending physical consistency with data-driven adaptability. Serial and parallel configurations of HMs show significant promise.

Chapter 3

Data Description

In this chapter four datasets (presented in Table 3.1) are assessed central to modeling ship resistance and propulsion performance, each offering unique attributes essential for predictive methods of current practices. Dataset 1, derived from CFD simulations, provides high-precision numerical data on hull resistance components. Dataset 2 builds on this with synthesized speed-power data, which translates CFD resistance to ship power by means of empirical propeller efficiencies. Dataset 3 includes physical modeltest data from deep-water tank (DT) tests, while Dataset 4 consists of real-world measurements from sea trials.

To identify the optimal dataset, each is evaluated based on sample size, design variance, feature richness, and data precision, as summarized in Table 3.2. Dataset 1 is selected for its balanced sample size, moderate design variance, and high data precision, making it well-suited for further analysis. This dataset will be enriched with additional features, according to Fig. 3.2, to support physical, data-driven, and hybrid modeling approaches in this study.

3.1 Current state of datasets

Table 3.1: List of four available and propulsion-related datasets in their current state, without any dataset enrichment.

Tag and speed information	ID	Unit	Set 1	Set 2	Set 3	Set 4
build number	bn	-	•	•	•	•
design revision	rev	-	•	•		
ship's speed	V_s	kn	•	•	•	•
Hydrostatic information						
length overall	L_{oa}	m	•	•	•	
length waterline	L_{wl}	m			•	
length between perpendiculars	l_{pp}	m		•	•	
beam moulded	B_{mld}	m	•	•	•	
beam waterline	B_{wl}	m				
ship draft (forward)	T_f	m	•	•	•	•
ship draft (aft)	T_a	m	•	•	•	•
displacement	V	m^3	•	•	•	•
wetted surface area bare hull	S	m^2			•	
wetted surface area appended	S	m^2	•		•	
longitudinal centre of gravity	LcG	m	•			
vertical centre of gravity	VcG	m	•			
longitudinal centre of buoyancy	LcB	m	•			
block coefficient	C_b	-			•	
midship section coefficient	C_m	-			•	
prismatic coefficient	C_p	-			•	
water plane coefficient	C_{wp}	-			•	
area exposed to wind	A_v	m^2			•	
Result information						
thrust deduction fraction	t	-			•	
wake fraction	w	-	•		•	
hull efficiency	η_h	-			•	
open-water efficiency	η_o	-			•	
relative rotative efficiency	η_r	-			•	
propulsive efficiency	η_d	-			•	
resistance bare hull	$R_{tot,bare}$	kn	•		•	
resistance appended	$R_{tot,app}$	kn	•	•	•	
thrust force	T	kn			•	
effective power	P_e	kw		•	•	
delivered power	P_d	kw		•	•	
propeller shaft power	P_s	kw		•	•	•
Data size (after pre-processing)						
ships		-	27	23	15	51
samples		-	220	594	385	350

To evaluate the proposed modeling approaches, we utilize datasets provided by Feadship, which are thoroughly detailed in this section to ensure the reproducibility of the work. The following datasets relating to resistance and propulsion power are available:

- **Set 1** (numerical) – This dataset is derived from structured computational fluid dynamics

(CFD) datasheets. These sheets report the frictional and pressure components, which, along with additional contributions (such as appendages and wind), comprise the total resistance for appended hulls in calm-water conditions. The numerical analyses are conducted using the commercial STAR-CCM+ software with a Reynolds-Averaged Navier-Stokes (RANS) solver and are typically performed for two draughts per geometry.

- **Set 2** (numerical) – This dataset is derived from structured tables in speed-power reports. These tables perform the conversion of the calm-water appended resistance data (set 1), incorporating the necessary propulsion factors and efficiencies. These factors account for the relationship between resistance and the power required to propel the vessel, offering a detailed view of vessel performance under propulsion. An in-house developed software package automates this conversion as much as possible to ensure consistency.
- **Set 3** (model tests) – This dataset is derived from the deep-water tank (DT) reports from Maritiem Research Instituut Nederland (MARIN), conducted for a significant portion of the existing Feadship fleet. In these reports, the experimental results are presented for resistance and (quasi-steady) self-propulsion tests in calm-water. For many decades, the experimental results are presented in similar format.
- **Set 4** (speed-power trials) – Typically, after a ship is built in a shipyard, various tests are conducted until the ship is delivered to the owner. The tests are broadly divided into on board tests and sea trials. The purpose of these tests is to provide a confirmation to the ship owner and classification society that the ship has been constructed in accordance with the contract and regulations.

  This dataset is derived from these trials, conducted by a third-party entity conform *ISO 15016: 2015*. These reports include tables that detail speed and fuel consumption measurements for a given total shaft power, covering the entire Feadship fleet. The majority of the speed trial reports are available.

Table 3.1 lists all datasets including common, hydrostatic and results-related information. Not all information is uniformly present in all datasets, and therefore the bullet-point indicators are highlighted when a specific variable is included in each set. At the bottom, information is included on the amount of sample size (amount of data points) and amount of design variance (unique ships).

3.2 Dataset selection

A systematic assessment can be applied to these four datasets to make a choice for the superior dataset to be used for the experimental tests. Specific criteria are applied:

- **Sample size** - Refers to the number of unique data points in each dataset. A larger sample size typically provides better statistical reliability and helps ensure that findings are representative and robust.
- **Design variance** - Measures the diversity within the dataset, in this case the amount of unique ships. Higher design variance means the dataset covers a wide range of situations, enhancing the ability to new, unseen data.
- **Feature richness** - Indicates the number and relevance of features (variables included in the dataset, giving an impression on the amount of information included in the set. If low, limited informational depth is present, often requiring feature enrichment techniques to augment the dataset, increasing efforts required in preparing the data for model application.
- **Data precision** - Provides a reflection on the expected exactness and reliability of the data, mainly based domain knowledge and observations during model development. High data precision minimizes uncertainty, making the dataset more reliable for experimental use.

Each dataset is evaluated based on scores categorized as low, moderate, or high, or through specific numerical values where applicable. The outcomes of this assessment are presented in Table 3.2.

Set 1 (Table 3.1) has a sample size of 220 data instances, resulting from two draught conditions and 4 different speeds (on average) per ship. In total, 27 different ship geometries have been counted, which represents the 2nd best design variance of all four sets. The feature richness is moderate, as a limited amount of hydrostatics and results is directly present on the CFD datasheets. The data precision is high, as the CFD experiments are conducted under controlled, relatively controlled conditions. Also, the

Table 3.2: Dataset assessment.

Criterion	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Best
Sample size	220	594	385	350	Dataset 2
Design variance (ships)	27	23	15	51	Dataset 4
Feature richness	Moderate	Moderate	High	Low	Dataset 3
Data precision	High	Moderate	High	Low	Dataset 1

hydrostatic information in set 1 is derived from the exact same 3D CAD file that is used for the CFD simulations. These characteristics make set 1 a good option for this study.

Set 2 (Table 3.1) has the largest sample size, with over 594 data points, but exhibits fairly low design variance. This is due to the software used for generating speed-power reports, which enables naval architects to create synthetic (non-existent) resistance-propulsion data at user-defined intervals, relying primarily on a minimal set of actual CFD measurements. Given that the precision of these interpolation (and potentially extrapolation) techniques is relatively uncertain, the data precision is rated as moderate. The feature richness is scored moderate, as minimal hydrostatic and result related features are included in set 2. Since most of the output results represent estimations, rather than direct measurements, the data precision is also scored moderate. Overall, this dataset does include more propulsion characteristics and for more ship speeds, compared to set 1. However, a majority of the data in this set is not stemming from direct measurements, making set 1 superior.

Set 3 (Table 3.1) has the second-largest sample size, as model tests are conducted across a wide range of speeds in the towing tank. However, model testing for Feadship yachts is reserved for designs that push beyond existing parameter limits, involve unique parameter combinations, or introduce novel features, such as alternative propulsion configurations, appendages, or the addition of a bulbous bow. Consequently, this set exhibits the lowest design variance among all four datasets but the highest feature richness and data precision. Each "diepwater tank" (DT) report provides results of highly-standardized model experiments, showing detailed tables including multiple outputs measured during the calm-water resistance and propulsion tests. A list of hydrostatics for two draught/trim conditions is included as well.

Due to its standardized and precise nature, set 3 is an excellent candidate for this study, but there is an issue as the exact design revision (rev) cannot be matched to the model tests. Each project (bn) evolves through time, and thereby new rev are introduced, causing potential misalignment between the geometrical data and resistance data (see Fig. 3.1). These differences might be minor for the majority of the set, while significant for some cases. This is an issue, as highly-precise data is of high importance in small datasets with low design variance.

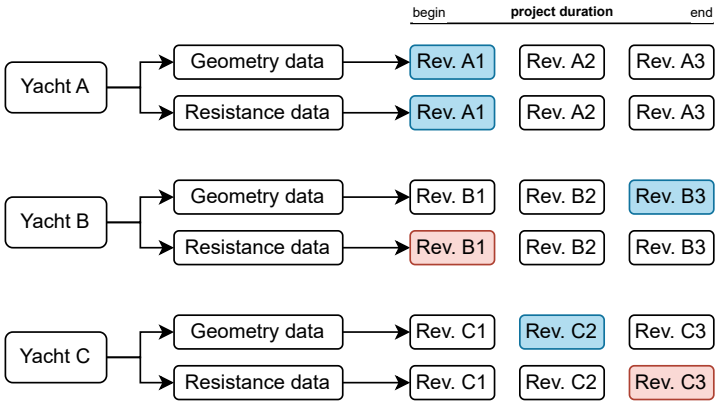


Figure 3.1: Illustrating the issue of potential misalignment between geometrical and resistance data in set 3, due to unmatched design revisions (rev).

Set 4 (Table 3.1) has by far the highest design variance and second-best sample size, as speed and power trials are mandatory for all Feadship deliveries. However, the accuracy and consistency of these trials have been questioned (Insel, 2008; Seo & Oh, 2021) as uncertainties are introduced by either the

trial measurements itself or by post-correction methods. Although the high design variance is a valuable feature of this set, the precision required for reliable data is considered lacking.

When testing different models, it is important to rule out the possibility that uncertainties or noise in the data are influencing the results. **Set 1 is chosen for the proceedings of this study**, as this set is characterized by fairly-good sample size and design variance, while including high-precision data.

3.3 Dataset enrichment

An essential part in the development of the DDMs and HMs is the selection of features to be used for training. The proposed PMs will not require any training, but they are build upon certain input variables (features), so to ensure consistent across the modeling approaches (PM, DDM, and HM), all DDMs and HMs use a feature set based on the (Holtrop & Mennen, 1984) method, detailed in Table 3.3. For this, it is necessary to add certain features, since the CFD set 1 (see Table 3.1) does not include all of them originally. This process is known as feature or data enrichment, and it will be done according to the database creation pipeline depicted in Fig. 3.2.

Table 3.3: (Holtrop & Mennen, 1982) based feature set that is used for the experimental tests (unless mentioned otherwise). For all these features, data is gathered from either the original dataset 1 (Table 3.1), or through the enrichment process (Fig. 3.2) to add missing features.

Feature name	ID	Units	From set 1	Added to set 1
Ship speed	V_s	kn	•	
Length waterline	L_{wl}	m	•	
Moulded beam	B_{mld}	m	•	
Moulded mean draft	T	m		•
Volumetric displacement	∇	m^3	•	
Longitudinal centre of buoyancy	ℓ_{cB}	m		•
Longitudinal centre of floatation	ℓ_{cF}	m		•
Wetted surface hull and appendages	S_{tot}	m^2	•	
Wetted surface of (individual) appendages	S_{a_i}	m^2		•
Transom wetted surface	A_t	m^2		•
Bulbous bow transverse area	A_{bt}	m^2		•
Height of centroid A_{bt} above keel	h_{bt}	m		•
Bow half angle of waterline entrance	i_E	deg		•
Block coefficient	C_b	-		•
Prismatic coefficient	C_p	-		•
Midship section coefficient	C_m	-		•
Water plane area coefficient	C_{wp}	-		•

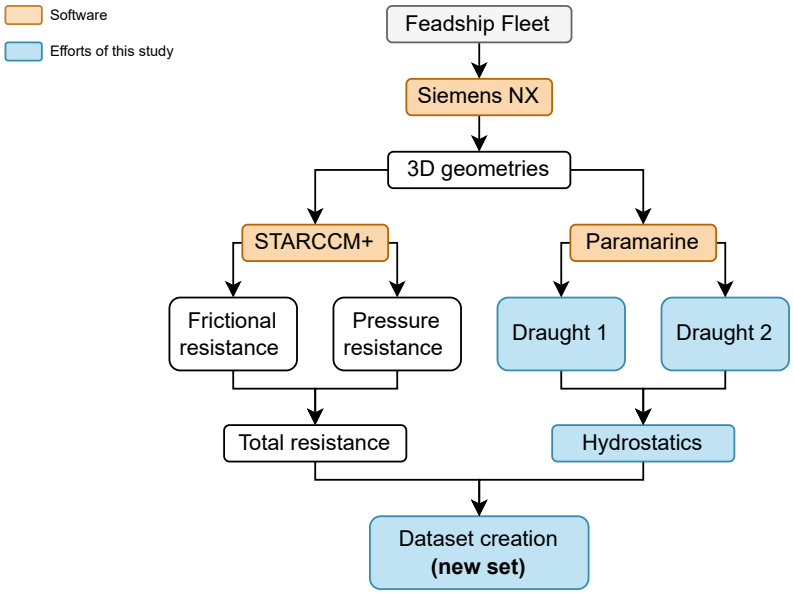


Figure 3.2: Process for dataset enrichment through 3D geometry utilization. Existing 3D geometries from the Feadship fleet serve a dual purpose: providing inputs for CFD simulations in STARCCM+ (completed by CFD engineers for resistance data) and enabling the generation of hydrostatic data (conducted in this study) using Paramarine software.

Fig. 3.3 presents a bottom view of a representative 3D geometry from the Feadship fleet, which serves as input for both STARCCM+ software (for CFD simulations) and Paramarine software (for hydrostatics generation). Since STARCCM+ typically processes only half of the geometry for CFD simulations, most 3D geometries required additional preparation. In Paramarine, these geometries were mirrored and merged to form complete solid models, ensuring the software could accurately compute hydrostatics. Additionally, since CFD simulations are performed for at least two draught conditions, these draughts were used as input to Paramarine to generate the corresponding hydrostatic data.

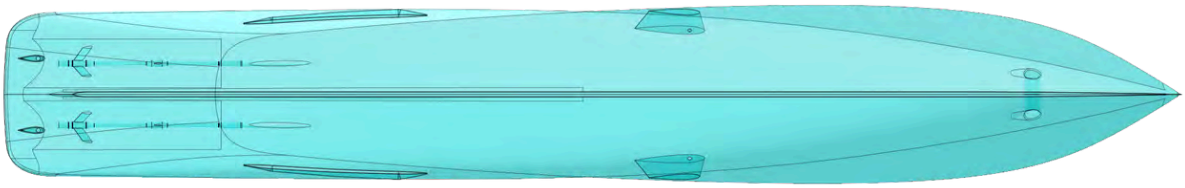


Figure 3.3: Bottom view of a representative Siemens NX geometry, showcasing the bare hull with appendages (e.g., bow thrusters, stabilizer fins, bilge keels, skeg, rudders, shaftlines, and brackets). This geometry is used for CFD simulations and dataset enrichment via Paramarine software.

Chapter 4

Methodology

In this chapter, all the proposed models will be outlined for early-stage prediction of calm-water ship resistance for twin-screwed superyachts.

The PMs section (Section 4.1.1) introduces established hydrodynamic models for capturing core resistance phenomena, providing a solid, industry-recognized baseline. The DDMs (Section 4.1.2) use machine learning algorithms, trained and validated with k-fold cross-validation, to model resistance directly from empirical data. Performance metrics such as MAE, MAPE, and R^2 evaluate each model's accuracy and robustness. The HMs (Section 4.1.3) combine PMs with DDM corrections in three configurations: serial, parallel-residual, and parallel-correction. These hybrids leverage PM predictions while integrating data-driven adjustments to improve accuracy and adaptability.

Finally, the testing framework is presented in Section 4.2 for interpolation condition, resistance curve construction for a specific ship, with less data, with other features and finally under extrapolation condition.

4.1 Proposed models

This section presents the proposed PMs, DDMs, and HMs that are going to be used for the experimental tests. All these models share the same prediction target; calm-water ship resistance.

4.1.1 Physical models

In the review (Section 2.3.1), several physical models (PMs) are proposed to computationally model the calm-water resistance for given ship speed and characteristics. In the context of the hybrid architecture, the central purpose of the PM is to provide a physically grounded baseline that captures fundamental resistance phenomena. For this reason, it is crucial that the proposed PM aligns with the problem at hand, which is the calm-water resistance for displacement yachts with a twin-screw propulsion configuration. The following models are proposed:

- **PM-a: Code 6** - (Holtrop & Mennen, 1982)
- **PM-b: Code 7** - (Holtrop & Mennen, 1984)

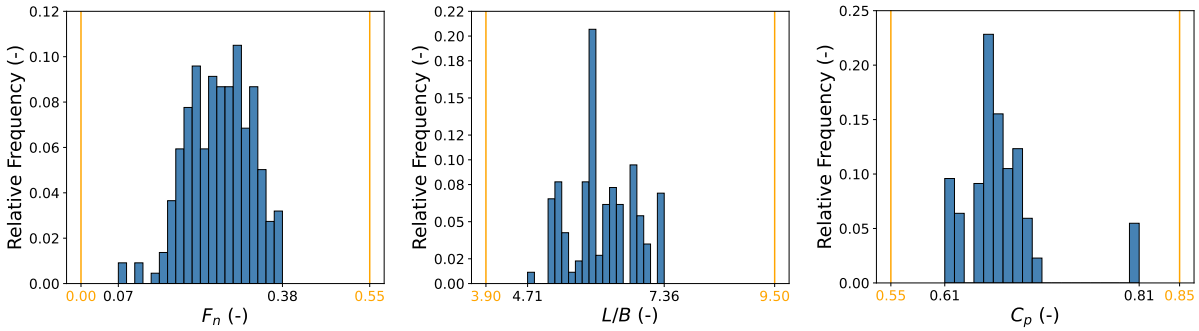
Both methods are internationally recognized methods and currently favoured by Feadship for the early-stage predictions. Thus, they are suitable PMs, but in order to improve predictions in the extrapolation scenario, the PM must have permissible ranges (see Table 4.1) that at least extend beyond the boundaries of set 1. For this purpose, a brief check is done on the Feadship numerical set 1 (Table 3.1) to see if the proposed PMs fulfil this requirement. Both models comply, as can be seen in Fig. 4.1.

Table 4.1: Proposed PMs, with corresponding permissible ranges within which the model is deemed to remain accurate.

Ref.	Inputs	Outputs	ID	Unit	Permissible range
(Holtrop & Mennen, 1982)	Appendix A	Appendix A	F_n	–	≤ 0.45
			C_p	–	$0.55-0.85$
			L/B	–	$3.9-9.5$
(Holtrop & Mennen, 1984)	Appendix A	Appendix A	F_n	–	≤ 0.55
			C_p	–	$0.55-0.85$
			L/B	–	$3.9-9.5$

4.1.2 Data-driven models

Unlike the PMs, each data-driven model has to be trained and tested. A possibility is to create a train-test split with an 80-20 ratio, meaning that the model is being trained on 80% of the data, while being tested on 20%. However, in practice this could potentially result in misleading test results as 20% of the total dataset does not represent the true characteristics of the data. Therefore, a certain randomness has to be introduced into the train and testing process, and this is achieved through k-fold cross-validation (KCV). Here, the total dataset is dissected in k folds, where k is usually equal to five or ten. In case of 5-fold cross-validation, the entire dataset is split into five folds where the first fold k represents the test set T_t, and the four remaining folds $k-1$ the training set D_n. The model is then trained on D_n and tested on T_t.



(a) Length-Froude number (F_n) **(b)** Length-over-beam ratio (L/B) **(c)** Prismatic coefficient (C_p)

Figure 4.1: Compliance check of Feadship fleet characteristics (dataset 1) against permissible ranges of proposed PMs.

It is here where all kinds of error metrics can be applied to assess the model's performance, which is better known as the error estimation (EE). Other measures of error exist, however, from a physical point of view, these three measures give a complete description of the quality of the model. Adding more measures would make the results more difficult to interpret, while not adding any new insights:

- **Mean Absolute Error (MAE):** The Mean Absolute Error quantifies the average magnitude of errors between predicted values $h(x_i^t)$ and actual values y_i^t over all m test samples $\mathcal{T}_m$. This metric is favoured for its straightforward interpretability and robustness against outliers, as it measures the average absolute deviation without emphasizing extreme errors. MAE is calculated as follows:

$$\text{MAE}(h) = \frac{1}{m}\sum_{i=1}^{m}|h(x_i^t) - y_i^t| \tag{4.1}$$

  In some cases, outliers must be penalized more heavily, and squared error metrics (e.g., MSE, RMSE) may be more appropriate. However, this is not necessary for this study, as the dataset is not expected to contain any outliers. In such scenarios, MAE offers a robust and stable measure of prediction accuracy, as noted in (Bishop, 1995) and (Hastie et al., 2009).

- **Mean Absolute Percentage Error (MAPE)**: The Mean Absolute Percentage Error expresses prediction accuracy as a percentage, by computing the average of the absolute differences between predictions and actual values, normalized by the actual values. MAPE is scale-independent, making it especially useful for data with varying scales (e.g. ship speed or physical dimensions). It is considered as the most relevant error metric for this study's context, and is calculated by:

$$\text{MAPE}(h) = \frac{100\%}{m}\sum_{i=1}^{m}\left|\frac{h(x_i^t) - y_i^t}{y_i^t}\right| \tag{4.2}$$

  However, MAPE can be sensitive to small values in y, which can lead to disproportionately large percentage errors, a point commonly noted in (Hyndman & Athanasopoulos, 2014).

- **Coefficient of Determination ($\mathbf{R^2}$)**: Also known as R^2, the coefficient of determination measures the proportion of variance in the actual values y explained by the predictions h. It ranges from 1 (perfect prediction) to negative values, where a negative R^2 indicates that the model performs worse than using the mean of the actual values ($\bar{y}$) as the predictor. It is defined as:

$$R^2(h) = 1 - \frac{\sum_{i=1}^{m}(y_i^t - h(x_i^t))^2}{\sum_{i=1}^{m}(y_i^t - \bar{y})^2} \tag{4.3}$$

  where $\bar{y}$ is the mean of y values. An R^2 value close to 1 indicates a strong fit, while values close to 0 suggest the model performs similarly to the mean prediction. Negative values imply the model performs worse than the mean prediction, as discussed in (Hastie et al., 2009).

Back to the k-fold cross-validation (KCV). The model is tested on the first fold k, the test set T_t, trained on the remaining folds $k-1$ that represent the training set D_n, and evaluated through the error metrics (MAE, MAPE, and R^2). Still, no randomness is introduced here, as this remains a regular train-test split. Therefore, the KCV method includes multiple iterations (see blue-grey folds in Fig. 4.2), where the test fold k constantly swaps position with every iteration. For example, in the second iteration, the second fold becomes the test set T_t, and the remaining folds (1, 3, 4, and 5) form the training set D_n. The model is trained and tested again on these different fold combinations, presumably performing slightly differently compared to iteration 1. This process is repeated until the number of iterations equals the number of folds, in this case five. In this way, KCV provides a more reliable and stable EE by accounting for different data partitions.

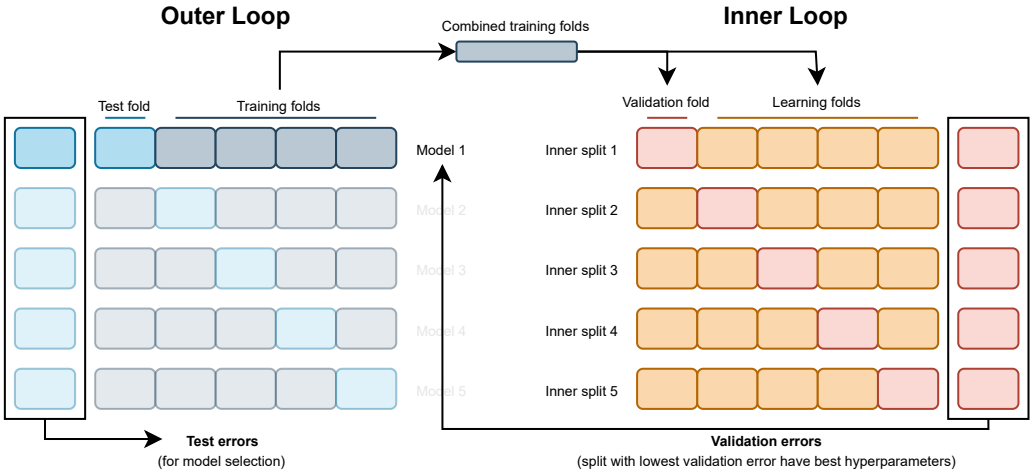


Figure 4.2: Visualization of the training and testing scheme of every data-driven model (DDM). Here, one outer loop of the nested k-fold cross validation is shown with 5-folds (k=5).

An equally important aspect of developing data-driven models is model selection (MS), which focuses on identifying the most optimal set of hyperparameters for a given application. Hyperparameters are like adjustable knobs that define the model's architecture and behaviour, such as the learning rate, regularization strength, number of leaves, or layers in a neural network. Unlike model parameters, which are learned during training, hyperparameters must be set before training begins, and their optimal values are often unknown initially. Since hyperparameters have a significant impact on model performance, the goal of model selection is to determine the configuration that minimizes the model's (validation) error. To better understand how this is achieved, it's necessary to introduce a new concept: nested k-fold cross-validation.

Nested k-fold cross-validation builds on standard k-fold cross-validation by introducing an inner loop. The **outer loop** (blue and grey folds) splits the dataset into k folds, where each fold serves as the test set T_t once, while the remaining $k-1$ folds form the training set D_n. Within each outer loop, an **inner loop** (orange and red folds) further splits D_n into training and validation sets to fine-tune hyperparameters. This nested structure ensures that hyperparameter tuning is performed independently of the test evaluation, preventing data leakage and overfitting. By iteratively training and validating across these folds, nested

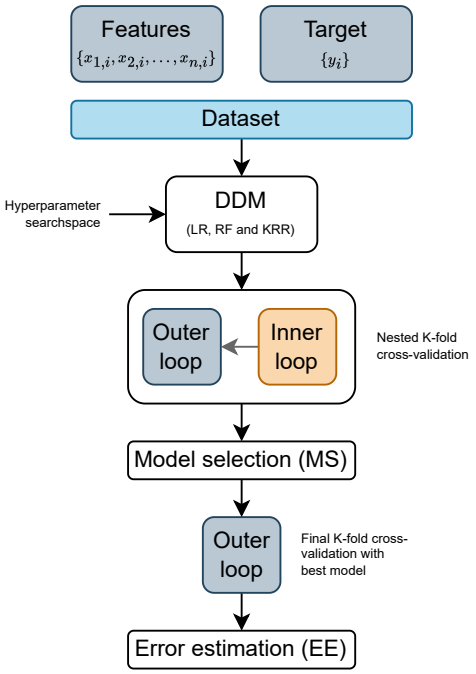


Figure 4.3: Model architecture of the data-driven models (DDM).

k-fold cross-validation provides a robust framework for both model selection and unbiased error estimation. The model architecture for each DDM is illustrated in Fig. 4.3.

In the review (Section 2.3.2), several data-driven models (DDMs) are proposed to computationally model the ship resistance for a given input. This study builds on state-of-the-art shallow DDMs by applying a series of top learning algorithms for regression. Rigorous testing is essential, as the no-free-lunch theorem (Adam et al., 2019) emphasizes that no single algorithm works best for every problem, making it necessary to evaluate multiple algorithms to identify the most suitable one for this specific application. Furthermore, each algorithm requires careful tuning of its hyperparameters (summarized in Table 4.2). More elaborate information is included in Appendix B on the learning algorithms, but the proposed DDMs are:

- **DDM-a/DDM-b: Linear Ridge (LR)** - A linear model, for example an ordinary least squares (OLS) method, is one of the simplest machine learning algorithms. During training, it tunes its coefficients by minimizing a loss function, typically the sum of squared errors, to achieve the best possible fit to the data. However, linear models can suffer from overfitting, where the model captures noise instead of the underlying patterns. To address overfitting, regularization techniques such as L_1 (lasso) and L_2 (ridge) are used to constrain the model's complexity. Ridge regression, which applies L_2 regularization, penalizes large coefficients by adding a penalty term to the loss function. This penalty helps stabilize the model, balancing fit and generalization, and is controlled by the regularization strength λ, which adjusts the trade-off between higher bias (stronger regularization) and lower variance (weaker regularization).
- **DDM-c: Random Forest (RF)** - Random forest is a robust ensemble learning method that aggregates predictions from multiple decision trees to reduce overfitting and improve accuracy. Each tree is trained on a random subset of features and samples, enhancing model diversity. Key hyperparameters include the number of trees n_t, which controls the size of the ensemble; the number of features sampled at each split n_f, which determines feature randomness; the maximum depth n_d of each tree, which limits the model complexity; and the minimum number of samples per leaf n_l, impacting leaf granularity. Carefully tuning these parameters helps in achieving a balance between computational efficiency and predictive performance.
- **DDM-d: Kernel Ridge Regression (KRR)** - Kernel ridge regression extends linear ridge regression with the Gaussian (RBF) kernel, enabling the model to capture complex, non-linear relationships in the data. By mapping input features to a high-dimensional space, KRR finds patterns that are not linearly separable in the original feature space. This model is fine-tuned by adjusting the regularization strength λ, which manages overfitting, and the kernel coefficient γ, which controls the spread of the Gaussian kernel and influences the model's sensitivity to variations in the data (Keerthi & Lin, 2003).

Table 4.2: Proposed DDMs, with corresponding hyperparameters and search space.

Model	Description	Hyperparameter	ID	Search Space
DDM-a	Linear ridge	Regularization strength	λ	$\{10^{-3}, ..., 10^{3}\}$
DDM-b	Linear ridge (with cubed speed input)	Regularization strength	λ	$\{10^{-3}, ..., 10^{3}\}$
DDM-c	Random forest	Number of trees	n_t	$\{100, 200, 1000\}$
		Number of features per split	n_f	$\{0.5, 0.75, 1.0\}$
		Maximum depth	n_d	$\{8, 10, 12, 15\}$
		Minimum samples per leaf	n_l	$\{1, 2, 4\}$
DDM-d	Kernel ridge regression	Regularization strength	λ	$\{10^{-6}, ..., 10^{2}\}$
		Kernel coefficient	γ	$\{10^{-5}, ..., 10^{-2}\}$

4.1.3 Hybrid models

The review in Section 2.3.3 presents multiple variations of hybrid models (HMs) designed to make computational predictions for various problems in the naval architecture domain. In this study, both the serial and parallel approaches (with novel contribution) are proposed to predict calm-water ship resistance. It is important to note that, similar to data-driven models (DDMs), hybrid models also require model selection (MS) and error estimation (EE). For the sake of simplicity it is not included in the visualizations, shown in Fig. 4.4, Fig. 4.5 and Fig. 4.6. The following hybrid models are being proposed:

- **HM-a: Serial model**
- **HM-b: Parallel-residual model**
- **HM-c: Parallel-correction model**

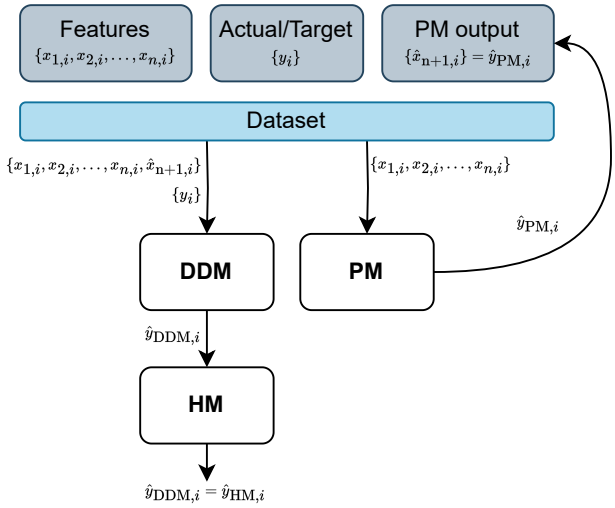


Figure 4.4: Hybrid model with serial configuration (HM-a), where the physical model (PM) output $\{\hat{y}_{PM,i}\} = \{\hat{x}_{n+1,i}\}$ is learned by the data-driven model (DDM).

The first hybrid model (HM-a) is the one using a serial approach. The dataset includes the features set X and actual (target) set Y for every single data point, denoted as i, and like any HM, the PM makes a prediction for every i. This PM output $\{\hat{y}_{PM,i}\}$ for all data points is fed back to the dataset, which is then used by the DDM as an extra input feature. For this purpose, the definition of the PM output $\{\hat{y}_{PM,i}\}$ is re-defined as $\{x_{n+1,i}\}$. And in the serial approach, this is all that is required to eventually come to a hybrid prediction. Very simplistic. The DDM will use this new input $\{x_{n+1,i}\}$ on top of the existing feature set X and the actual set Y, and makes a prediction defines as $\{y_{DDM,i}\}$, which in reality is equal to the hybrid prediction $\{y_{HM,i}\}$. For the visualization, see Fig. 4.4.

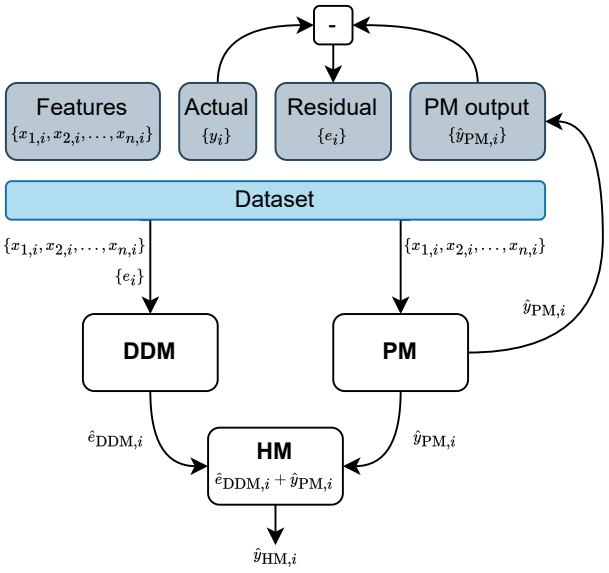


Figure 4.5: Hybrid model with parallel configuration (HM-b), where the residual $\{e_i\}$ is learned by the data driven model (DDM).

The second hybrid model (HM-b) is one using the parallel approach. Like before, the dataset included the features set X, the actual (target) set Y and the PM output $\{\hat{y}_{PM,i}\}$. Though, in all parallel configurations, an additional computation is necessary, which can be performed either within the dataset itself or in the code. For the HM-b, this computation entails a subtraction between the actual (target) $\{y_i\}$ and PM output $\{\hat{y}_{PM,i}\}$, also known as the residual $\{e_i\}$. In this HM-b configuration, it is this $\{e_i\}$ that is learned by the DDM, not the actual (target) set Y. The DDM prediction $\{\hat{e}_{DDM,i}\}$ is then simply summed with the PM prediction $\{\hat{y}_{PM,i}\}$, which together form the HM predictions $\{\hat{y}_{HM,i}\}$. For the visualization, see Fig. 4.5.

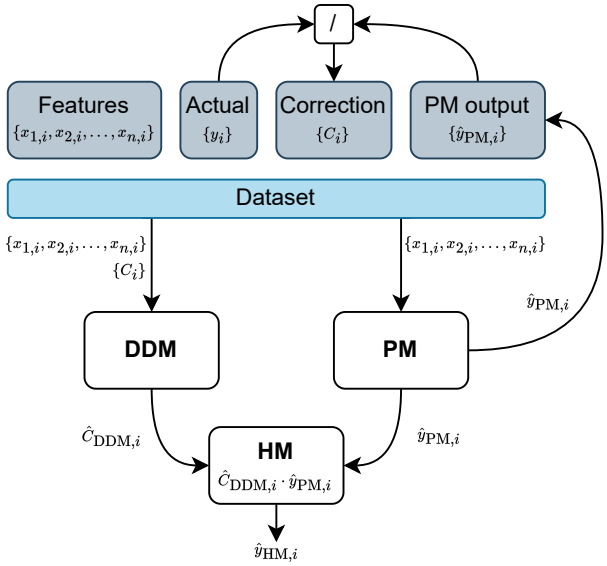


Figure 4.6: Hybrid model with parallel configuration (HM-c), where the required correction $\{C_i\}$ is learned by the data-driven model (DDM).

The third hybrid model (HM-c) also employs the parallel approach, but to the best of the author's knowledge, it introduces a novel contribution. As before, the dataset includes the feature set X, the actual (target) set Y, and the PM output $\{\hat{y}_{PM,i}\}$. The novelty lies in the additional computation: instead of computing a residual value $\{e_i\}$, a correction factor $\{C_i\}$ is derived and learned by the DDM. This correction factor is calculated through dividing the actual target by the PM output. Thus, if the PM underestimates the actual resistance $\{y_i\}$, the correction factor exceeds 1, indicating an upward adjustment needed for the PM to match the actual resistance. And when the PM overestimates $\{y_i\}$, the correction factor exceeds 1, indicating the downward adjustment required for the PM to match the actual resistance. So, the DDM predicts this correction factor $\{\hat{C}_{DDM,i}\}$, which is the multiplied by the PM prediction $\{\hat{y}_{PM,i}\}$, which together form the HM predictions $\{\hat{y}_{HM,i}\}$. For the visualization, see Fig. 4.6.

4.2 Testing framework

Building on the proposed PMs (Section 4.1.1), DDMs (Section 4.1.2), and HMs (Section 4.1.3) described earlier, rigorous testing is essential to comprehensively evaluate their predictive performance across various scenarios. Each model type is subjected to a series of structured tests, including interpolation to assess performance within the available data range, ... and extrapolation to evaluate performance beyond the data bounds.

4.2.1 Interpolation pipeline

The initial experimental tests were thoughtfully designed to serve two purposes, first, to ensure that models perform correctly within the range of the available data (interpolation) before being exposed to more complex scenarios. Second, to assess the interpolation ability of each model, establishing a robust baseline for performance across individual physical, data-driven, and hybrid models. The physical

model (PM) does not require any training, but for the data-driven models (DDMs) and hybrid models (HMs), a proper training and testing scheme is necessary (see Section 4.1.2) to provide a fair assessment of their true capabilities. Table 4.3 presents the models used for interpolation tests.

Table 4.3: Overview of models used for interpolation tests.

Type	Model	Data-driven part	Description
PM	PM-a	-	(Holtrop & Mennen, 1982)
	PM-b	-	(Holtrop & Mennen, 1984)
DDM	DDM-a	-	Linear ridge regression
	DDM-b	-	Linear ridge regression (with cubed speed input)
	DDM-c	-	Random forest
	DDM-d	-	Kernel ridge
HM	HM-a	Best DDM	Serial
	HM-b	Best DDM	Parallel-residual
	HM-c	Best DDM	Parallel-correction

The computational fluid dynamics (CFD) dataset, denoted as dataset 1 in Section 3.3, is used for the interpolation tests, with the pipeline visualized in Fig. 4.7. The working principles for each model type are detailed in Section 4.1. However, since cross-validation can be somewhat complex, additional focus is given to explaining it within the context of interpolation. Remember, as noted earlier in Section 4.1.2, there is a necessity of introducing some randomness in the training-testing scheme. Instead of a conventional 80-20 train-test split, which could be considered as "cherry-picking", the data is divided into folds, with multiple splits created by shuffling the training and test folds differently in each iteration. In nested k-fold cross-validation (KCV), this concept is combined with an additional inner loop to do the model selection (MS), a process known for selecting the most optimal set of hyperparameters. See how this is achieved for both the DDM and HM in Fig. 4.7. After the MS process, the best model needs to be retrained. However, it is essential to reintroduce randomness into the training-testing scheme using k-fold cross-validation, specifically through the outer loop. From this process, the error estimation (EE) can be obtained. It is important to note that the interpolation pipeline shown in Fig. 4.7 remains consistent throughout all interpolation tests; the only elements that are varied are the PMs, DDMs and HMs.

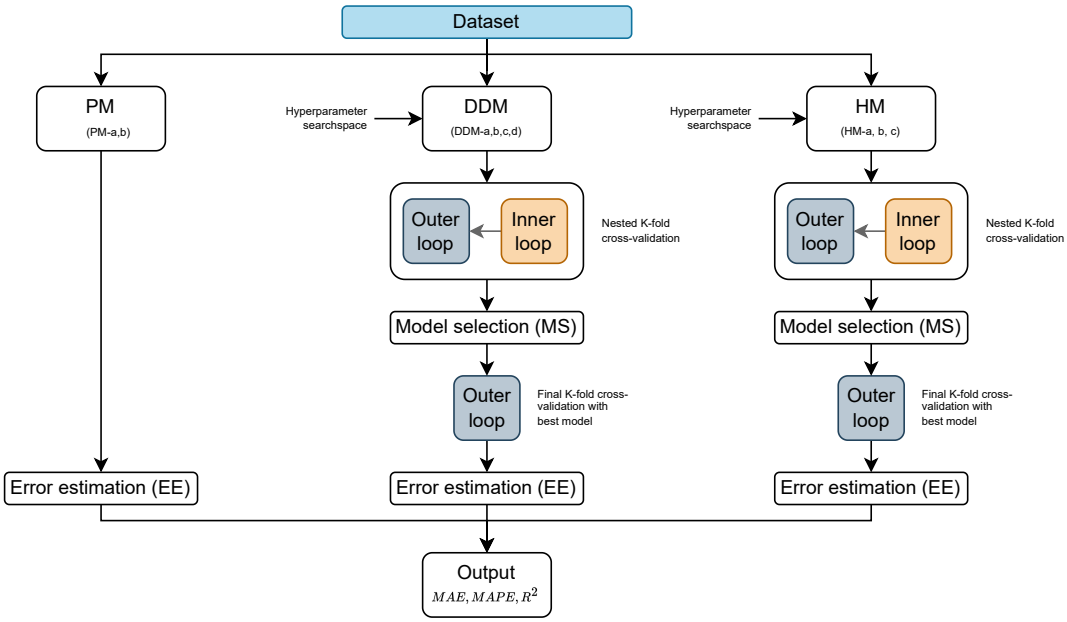


Figure 4.7: Interpolation pipeline used for all tests in Chapter 5, except during tests for extrapolation condition. This pipeline is presented in Section 4.2.5.

Three main types of results are going to be depicted in Section 5.1: test errors, scatter plots, and error distributions. First, the test error metrics (MAE, MAPE, and R^2) are presented in formatted tables, displaying two types of values: the mean (average) and the standard deviation, calculated across all

folds of the final KCV. Secondly, the scatter plots are included, which display the actual values against the predictions with a 45-degree line, known as the ideal line. When all points lie exactly on this line, it indicates perfect predictions with no errors. And thirdly, the error distributions are shown next to the scatter plots. Instead of residuals or absolute errors, relative errors (%) are used, as dataset 1 includes varying scales (e.g., ship speeds and dimensions), as noted in Section 4.1.2.

4.2.2 With specific ship

A resistance curve can be generated for a specific ship to evaluate the capabilities and limitations of each modeling approach. The methodology for this test involves applying the best-performing models from each approach to predict the calm-water resistance of a specific ship. The following steps outline the process:

1. **Model selection (MS) and error estimation (EE):** All models are trained and tested using the interpolation pipeline accounting for the model selection (MS) and error estimation (EE), as detailed in Fig. 4.7. Via this way, a performance assessment (MAE, MAPE, and R^2) can be made, to only select the best-performing PM, DDM, and HM for this test.
2. **Ship selection:** A specific ship geometry is chosen from dataset 1 (Chapter 3) for detailed testing. The selected geometry represents a practical case, ensuring relevance to real-world applications, such as early-stage propulsion system design.
3. **Resistance curve construction:** The best PM, DDM and HM are used to predict calm-water resistance at the same speeds as those available in the CFD observations. These predictions are compiled into resistance curves for direct comparison with CFD data.
4. **Analysis of results:** The deviations between the PM, DDM, and HM prediction and the actual CFD data are analysed. Special focus is placed on assessing the hybrid models' ability to integrate physical principles with data-driven corrections.

This methodology ensures a fair and consistent comparison by using an identical speed range for all predictions. Detailed results of this test are presented in Section 5.2.

4.2.3 With less data

An important knowledge gap mentioned in the introduction (Chapter 1), is the challenge of making predictions with limited data. Since the physical models (PMs), data-driven models (DDMs), and hybrid models (HMs) have already been developed at this stage, testing the best models of each type under this condition becomes a relatively straightforward next step. The methodology for this test involves the following steps:

1. **MS and EE with reduced training data:** The training process begins with only 10% of the CFD observations from dataset 1, using the interpolation pipeline described in Fig. 4.7. The MAPE error value is recorded.
2. **Incremental addition of training data:** Additional CFD observations are incorporated in 10% increments, and the MAPE is recorded after each step. This approach allows for a systematic evaluation of how increasing data availability impacts model performance.
3. **Analysis of MAPE progression:** The progression of MAPE is analyzed and visualized for the most promising PM, DDM, and HM models. This comparison identifies which model achieves the best performance at each incremental step and determines their efficiency in leveraging additional data.

This methodology provides a clear framework for evaluating the models' performance under data scarcity. The results reveal how effectively each model leverages limited data, offering valuable insights into their robustness and suitability for real-world applications. Detailed results of this test are presented in Section 5.3.

4.2.4 With other features

For any predictive model, it is essential that the features (input variables) capture sufficient information to enable accurate predictions. The Holtrop & Mennen method is widely considered as one of the most advanced semi-empirical method for ship performance and for decades a lot of effort has been put

into improving and optimizing features within this method. Therefore, the interpolation (Section 5.1), extrapolation tests (Section 5.5) and the tests with less data (Section 5.3) use a (Holtrop & Mennen, 1984) based features set, as it made sense to use existing knowledge as a starting point. Yet, is this the most optimal feature set for making calm-water predictions for twin-screw superyachts?

This raises a critical question: what methods can be used to select an optimal set of features? The most straightforward approach is to leverage domain knowledge, either to select relevant features directly or to construct meaningful combinations of them - a process commonly known as informed feature engineering. With decades of experience in yacht construction, Feadship's extensive knowledge base offers valuable insights that can guide this process effectively. For instance, it is well-known that the longitudinal centre of floatation $\ell_c F$ needs to shift slightly forward relative to the centre of gravity $\ell_c G$ for certain ship types. Similarly, the depth of the immersed transom D_t is recognized as a significant factor influencing ship resistance. Also, when estimating propulsion performance in early-stage designs, it is commonly understood that the Heickel coefficient works better for superyachts than the Admiralty coefficient. These insights, gained from decades of shipbuilding experience, establish domain knowledge as the foundation for constructing an optimal feature set for use in DDMs and HMs.

Another approach to selecting optimal features is through statistical feature selection methods, with two widely-used techniques tested: backward feature elimination (BFE) and permutation importance (PI). BFE (Coraddu, Oneto, et al., 2022) is a wrapper method that starts with all features and iteratively removes the least impactful ones, based on error metrics (e.g. MAE, MAPE). This process continues until further removals degrade performance, resulting in the smallest subset of features that maintains or improves model effectiveness. In contrast, PI is a post-hoc analysis method (Altmann et al., 2010) that evaluates feature relevance by randomly shuffling the values of each feature, one at a time, to break its relationship with the target variable. The model's performance is then measured on this altered dataset; a significant drop in performance indicates that the feature is important.

This test evaluates the impact of different feature selection methods on model performance by systematically ranking and incorporating features. The analysis focuses on the best-performing hybrid model from the interpolation tests, using three feature selection approaches: domain knowledge, backward feature elimination, and permutation importance. The methodology is outlined as follows:

1. **Feature ranking:** Features are ranked based on their estimated importance to the target variable, calm-water ship resistance. Higher ranks are assigned to features with greater influence, as determined by each feature selection method.
2. **MS and EE with the highest-ranked feature:** The training process starts with only the highest-ranked feature (rank 1), utilizing the interpolation pipeline described in Fig. 4.7. The MAPE value is recorded to evaluate model performance under this minimal feature set.
3. **Incremental feature inclusion:** Features are added iteratively based on their rank order. In each step, the model is retrained using a progressively larger feature set: starting with the top-ranked feature (rank 1), then adding the second-ranked feature (ranks 1 and 2), followed by the third-ranked feature (ranks 1, 2 and 3), and so on.
4. **MAPE progression analysis:** The progression of MAPE is analysed and visualized as features are incrementally added. This analysis identifies the feature selection method that achieves optimal performance with the fewest features. The overall impact of features on model performance can also be analysed.

This methodology provides a structured framework for assessing the effectiveness of feature selection methods. The detailed results of this evaluation are presented in Section 5.4.

4.2.5 Extrapolation pipeline

What is extrapolation? This concept is relatively straightforward to understand when applied to a ship's speed - referring to making predictions for speeds that are "outside the bag" of existing speed data. However, the notion becomes less clear when it involves extrapolation in terms of a ship's geometrical dimensions, where the relationship between features and outcomes may be more complex and less intuitive. To gain a clearer understanding of the concept of extrapolation, let us begin with an illustrative example involving a new ship design, depicted in Fig. 4.8. The design characteristics (or feature values) of this ship are represented by the red markers, while the blue markers represent the (fictional) available

data. While several features, such as the length at waterline (L_{wl}), beam moulded (B_{mld}), and draft (T), fall within the bounds of the training data, others extend beyond the range of the existing dataset. Specifically:

- **Volumetric displacement (∇)**: This feature represents the submerged volume of the hull and is larger than any value in the dataset, suggesting a heavier vessel.
- **Length-to-draft ratio (L/T)**: This ratio is lower than the dataset values, indicative of a relatively shorter and deeper hull form.
- **Water plane coefficient (C_{wp})**: The value exceeds the training range, highlighting a very full-form hull.

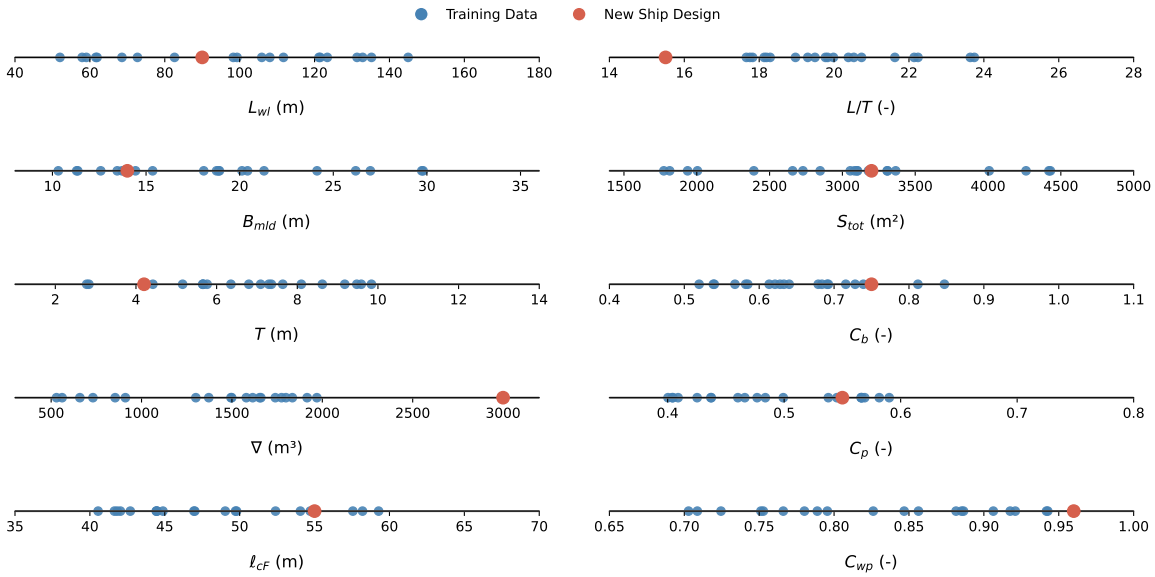


Figure 4.8: Visualization of a new ship design compared to the training data, highlighting extrapolation on certain features.

Extrapolating on such features poses challenges for predictive modeling, as relationships between inputs and outputs are less certain outside the training range. The purpose of the extrapolation tests is to evaluate the model's performance for such challenging cases. These tests use an alternative form to the final k-fold cross-validation in the interpolation pipeline (Fig. 4.7). This new form is known as leave-one-out cross-validation (LOOCV), but it requires some adaptions to be suitable for the extrapolation tests. Traditionally, in LOOCV, a single observation is removed for testing and the remaining ones are used for training. For the next iteration, another observation is used for testing and the remaining ones for training, and this process continues until all data points are touched upon.

Though, in these extrapolation tests, instead of removing a single observation, an entire subset containing multiple data points is removed, specifically on the outer regions. What are the outer regions? Well, this depends on what feature is selected for extrapolation. Once selected, all data can be dissected into three histogram bins: left extrapolation bin, middle bin and the right extrapolation bin. For this study, three testing scenarios designed and defined as the following:

- **LOFO: Leave One F_n Out:** In this scenario, the data is divided into three bins based on the Froude number F_n. Models are trained on the middle bin, excluding data from a specific F_n in the outer bins.
- **LOBO: Leave One B/T Out:** In this scenario, the data is divided into three bins based on the beam-to-draught ratio B/T. Models are trained on the middle bin, excluding data from a specific B/T in the outer bins.
- **LOCO: Leave One C_p Out:**In this scenario, the data is divided into three bins based on the prismatic coefficient C_p. Models are trained on the middle bin, excluding data from a specific C_p in the outer bins.

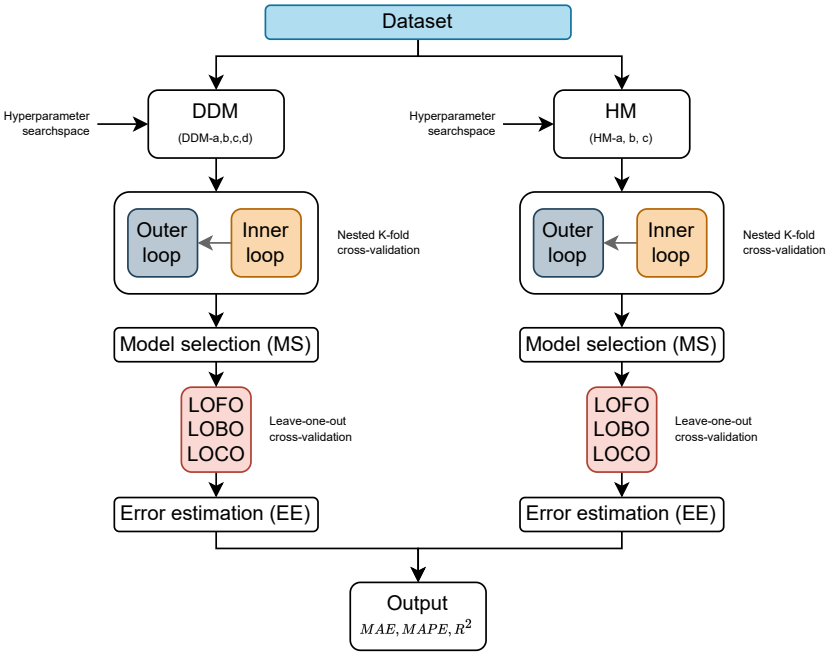


Figure 4.9: Extrapolation pipeline.

A visualization of the extrapolation pipeline, applied to CFD dataset 1, is presented in Fig. 4.9. This pipeline closely resembles the approach used for the interpolation tests, as described in Section 5.1, with two key differences: the exclusion of physical models, which inherently possess extrapolation capabilities, and the replacement of the final K-fold cross-validation (KCV) with the adapted LOOCV method.

Table 4.4: Overview of models used for extrapolation tests.

Type	Model	Data-driven part	Description
DDM	DDM-a	-	Linear ridge regression
	DDM-b	-	Linear ridge regression (with cubed speed input)
	DDM-c	-	Random forest
	DDM-d	-	Kernel ridge
HM	Best HM	DDM-a	Best HM with Linear Ridge (V_s^1)
	Best HM	DDM-b	Best HM with Linear Ridge (V_s^3)
	Best HM	DDM-c	Best HM with Random Forest
	Best HM	DDM-d	Best HM with Kernel Ridge

Table 4.4 presents all the models that are being tested in these extrapolation tests. The results in Section 5.5, are showing three main types of results: test errors, scatter plots, and error distributions, similar to the interpolation results in Section 5.1.

Chapter 5

Results

In this chapter, the performance of the PMs, DDMs, and HMs is tested in accordance with the methodology described in Chapter 4. Five tests are included here: interpolation condition, with specific ship, with less data, with other features and extrapolation condition.

In the interpolation tests (Section 5.1), three main types of results are presented: test errors, scatter plots, and error distributions. Test errors (MAE, MAPE and R^2) are summarized in tables showing the mean and standard deviation across all folds of the final KCV. Scatter plots follow, comparing actual values to predictions with a 45-degree ideal line for visualizing errors. Adjacent error distribution plots depict relative errors (%) instead of residuals or absolute errors due to varying scales in Dataset 1 (e.g., ship speeds and dimensions). For interpretating the error metrics, see explanation in Section 4.1.2.

Following this, three additional tests are conducted using the interpolation pipeline. First, the most promising models from each modeling approach are used to create a resistance curve for an arbitrary yacht (Section 5.2). Next, tests with reduced sample sizes (Section 5.3) simulate limited data scenarios, evaluating model accuracy under constrained conditions. Finally, tests with alternative input features (Section 5.4) explore the impact of feature selection on prediction accuracy, offering insights into feature importance and guiding the optimization of input configurations.

And finally, in the extrapolation tests (Section 5.5), the same three types of resuls are presented, but with a focus on extrapolation performance for different scenarios (LOFO, LOBO and LOCO). In these tests, the DDMs and most promising HM from Section 5.1 are trained on the middle bins, while sequantially being tested on the outer left and right bins.

5.1 Test for interpolation condition

The initial experimental tests were thoughtfully designed to serve two purposes, first, to ensure that models perform correctly within the range of the available data (interpolation) before being exposed to more complex scenarios, as described in Section 4.2.5. Second, to assess the interpolation ability of each model, establishing a robust baseline for performance across individual physical, data-driven, and hybrid models. In these tests, the newly enriched dataset from Section 3.3 is utilized, incorporating features derived from the methods proposed by (Holtrop & Mennen, 1984; Holtrop & Mennen, 1982), as summarized in Table 5.1.

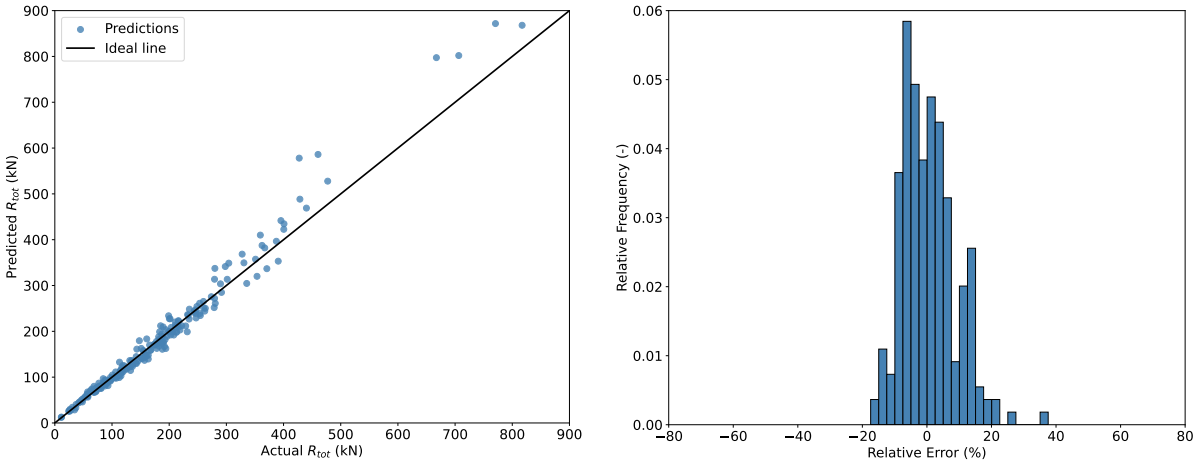
Table 5.1: Holtrop-Mennen based feature set used for the upcoming experimental tests (unless stated otherwise), derived from dataset 1 (Table 3.1) and enriched through the process detailed in Fig. 3.2.

Feature name	ID	Units
Ship speed	V_s	kn
Length waterline	L_{wl}	m
Moulded beam	B_{mld}	m
Moulded mean draft	T	m
Volumetric displacement	∇	m^3
Longitudinal centre of buoyancy	ℓ_{cB}	m
Longitudinal centre of floatation	ℓ_{cF}	m
Wetted surface hull and appendages	S_{tot}	m^2
Wetted surface of (individual) appendages	S_{a_i}	m^2
Transom wetted surface	A_t	m^2
Bulbous bow transverse area	A_{bt}	m^2
Height of centroid A_{bt} above keel	h_{bt}	m
Bow half angle of waterline entrance	i_E	deg
Block coefficient	C_b	-
Prismatic coefficient	C_p	-
Midship section coefficient	C_m	-
Water plane area coefficient	C_{wp}	-

The results in Table 5.2 show that PM-b, based on the algorithm by (Holtrop & Mennen, 1984), achieved the highest baseline performance with an R^2 of 0.96 and a Mean Absolute Percentage Error (MAPE) of 6.6%, marginally outperforming PM-a. It should be pointed out that the results for both PM-a and PM-b already highlight the effectiveness of the physical models, offering a dependable standard for the hybrid approaches.

Table 5.2: Interpolation tests - Performance assessment of all physical models (PMs).

Model	Algorithm	MAE (kN)	MAPE (%)	R^2 (-)
PM-a	(Holtrop & Mennen, 1982)	13.1	6.8	0.96
PM-b	(Holtrop & Mennen, 1984)	13.3	6.6	0.96



(a) Actual vs. Predicted values (kN)

(b) Relative errors (%)

Figure 5.1: Physical model (PM) with the best interpolation performance - (Holtrop & Mennen, 1984) (PM-b).

Figure 5.1 provides further insight into the predictive accuracy of PM-b through two complementary visualizations. The scatter plot in Fig. 5.1(a) compares the predicted resistance values with the actual values, plotting each prediction against its corresponding observed data point. This type of plot is particularly valuable because it offers a straightforward way to evaluate the overall alignment between predictions and reality; when predictions are accurate, the data points should closely follow the ideal 45-degree line. In this case, the strong alignment along this line indicates that PM-b performs well in capturing the underlying trend within the interpolation range.

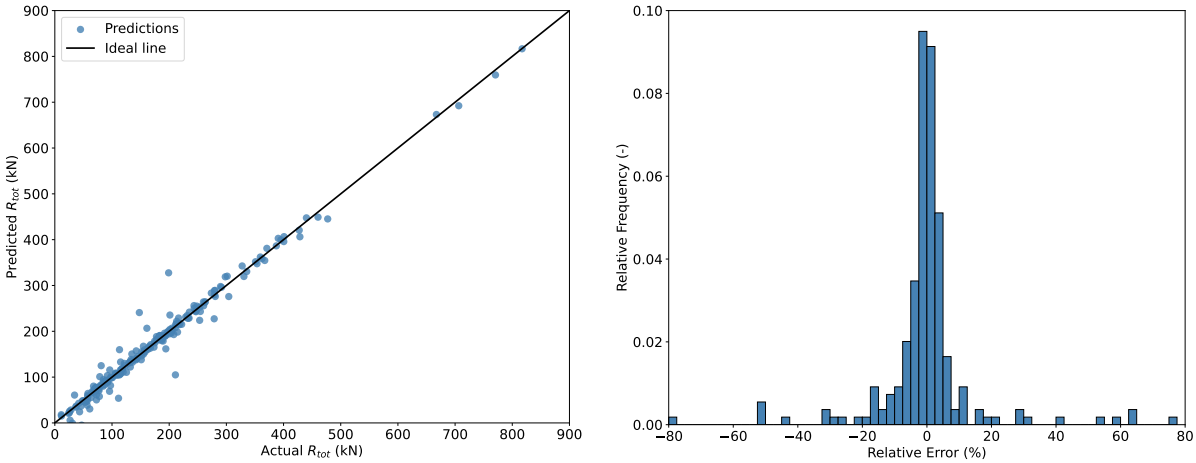
In Fig. 5.1(b), the relative error distribution provides an additional layer of understanding. Absolute error metrics, like mean average error (MAE), would not be as informative in this context because there are different scales (e.g. ship dimensions and speeds) are present in the data. The use of relative errors (expressed as a percentage) allows for a normalized view of model performance across all cases, regardless of any scale. Fig. 5.1(a) and Fig. 5.1(b) provide visual evidence of this, since the scatter plot in Fig. 5.1(a) indicates higher absolute errors in the upper resistance range, while the error distribution in Fig. 5.1(b) clearly demonstrates that these errors are not outliers compared to the majority of the data. In general, the error distribution is tightly clustered around 0%, indicating a low degree of bias and a high level of consistency in the model's predictions.

Table 5.3: Interpolation tests - Performance assessment of all data-driven models (DDMs).

Model	Algorithm	MAE (kN)	MAPE (%)	R^2 (-)
DDM-a	Linear Ridge (V_s^1)	25.8 ± 10.7	32.5 ± 31.0	0.86 ± 0.11
DDM-b	Linear Ridge (V_s^3)	13.6 ± 1.8	14.4 ± 6.0	0.96 ± 0.03
DDM-c	Random Forest	18.3 ± 3.9	12.0 ± 4.1	0.93 ± 0.06
DDM-d	Kernel Ridge	10.0 ± 5.7	8.9 ± 5.3	0.95 ± 0.09

The interpolation tests for data-driven models (DDMs), summarized in Table 5.3, show that the Kernel Ridge model (DDM-d) performed best, achieving an MAE of 10.0 kN, a MAPE of 8.9%, and an R^2 of 0.95. Among the Linear Ridge models, DDM-b, which applied a cubic transformation of ship speed (V_s^3), outperformed DDM-a, halving the MAE (13.6 kN vs. 25.8 kN) and significantly reducing the MAPE (14.4% vs. 32.5%), highlighting the value of feature engineering. The Random Forest model (DDM-c) delivered competitive performance, with an MAE of 18.3 kN and an R^2 of 0.93, demonstrating its ability to handle non-linear patterns effectively. Overall, these results highlight Kernel Ridge (DDM-d)

as the most reliable model for interpolation within this dataset, followed closely by the transformed Linear Ridge model (DDM-b).

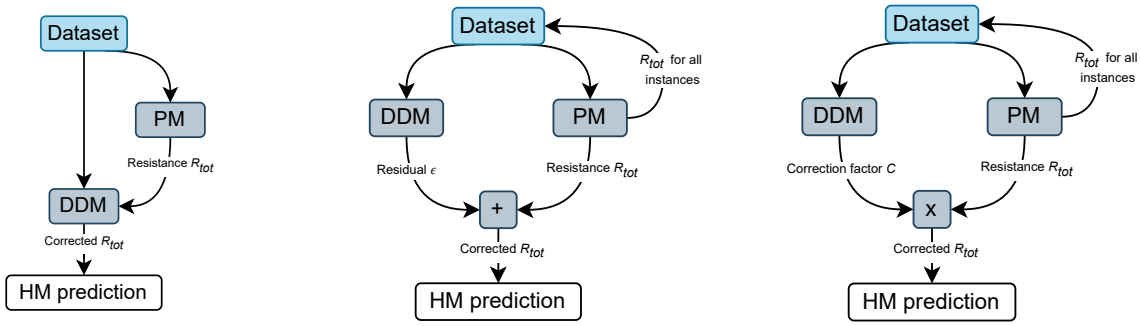


(a) Actual vs. Predicted values (kN)

(b) Relative errrors (%)

Figure 5.2: Data-driven model (DDM) with the best interpolation performance - Kernel Ridge (DDM-d).

Figure 5.2 provides a detailed visualization of the interpolation performance of the Kernel Ridge model (DDM-d). In the scatter plot (Fig. 5.2(a)), the data points cluster closely along the ideal 45-degree line, reflecting high prediction accuracy with minimal deviation from actual values. However, DDM-d exhibits occasional large deviations, visible in both the scatter plot (Fig. 5.2(a)) and the error distribution (Fig. 5.2(b)). These deviations highlight the model's sensitivity to certain data points, contrasting with PM-b's more stable performance. PM-b's error distribution (Fig. 5.1(b)) shows a broader but more uniform spread, indicating fewer extreme errors. This comparison suggests that while DDM-d's flexibility enables superior average accuracy, it sometimes compromises stability in individual predictions, whereas PM-b provides a steadier, more reliable baseline.



(a) Serial model (HM-a)

(b) Parallel-residual model (HM-b)

(c) Parallel-correction model (HM-c)

Figure 5.3: Simplified representations of the hybrid architectures used for the experimental tests.

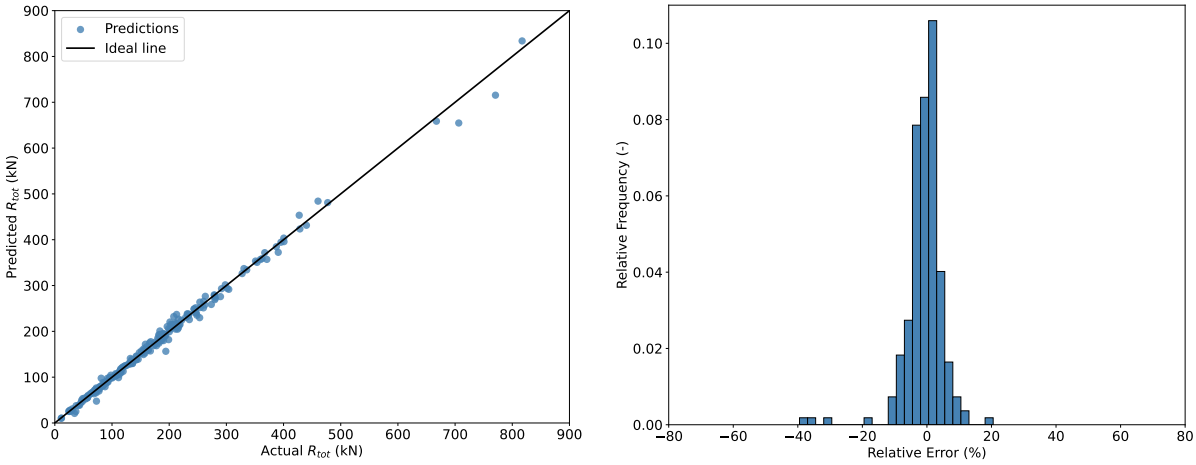
Next, the interpolation tests for hybrid models (HMs), focusing on three architectures Fig. 5.3): the Serial model (HM-a), Parallel-residual model (HM-b), and Parallel-correction model (HM-c). A more detailed visualization of all HMs can be found in Section 4.1.3.

Table 5.4 highlights the performance of each model, with HM-c (Parallel-correction) demonstrating the highest accuracy, achieving a MAPE of just 3.8% and an R^2 of 0.99, reflecting its strong consistency in predictions. HM-b (Parallel-residual) also performs well, with a MAPE of 5.6% and an R^2 of 0.97, though it is slightly less accurate than HM-c. In contrast, HM-a (Serial model) shows a higher MAPE of 12.3% and a lower R^2 of 0.88, indicating only moderate improvement over the physical models alone.

Table 5.4: Interpolation tests - Performance assessment of all hybrid models (HMs).

Model	Algorithm	MAE (kN)	MAPE (%)	R^2 (-)
HM-a	Serial model	21.5 ± 10.0	12.3 ± 4.5	0.88 ± 0.09
HM-b	Parallel-residual model	8.6 ± 6.8	5.6 ± 1.4	0.97 ± 0.06
HM-c	Parallel-correction model	5.8 ± 1.5	3.8 ± 1.1	0.99 ± 0.00

Figure 5.4 provides the scatter plot and error distribution for HM-c. The scatter plot in Fig. 5.4(a) shows a very tight clustering of data points along the ideal line, reflecting HM-c's accuracy in matching actual resistance values across the dataset. The error distribution in Fig. 5.4(b) is also narrow and centered closely around 0%, indicating minimal bias and high consistency across predictions. Compared to the best-performing DDM and PM models, HM-c demonstrates not only lower average errors but also reduced variability in individual predictions.



(a) Actual vs. Predicted values (kN)

(b) Relative errrors (%)

Figure 5.4: Hybrid model (HM) with the best interpolation performance - Parallel correction factor (HM-c).

A comparative summary is provided in Table 5.5 of all results in these tests for interpolation condition. The Parallel-correction hybrid model (HM-c) stands out with the lowest error metrics, achieving the smallest MAE and MAPE, along with the highest R^2, indicating its superior accuracy and consistency. PM-b ((Holtrop & Mennen, 1984)) and the Kernel Ridge model (DDM-d) also perform well, with moderate MAPE and R^2 values, but higher variability in individual predictions for DDM-d.

Table 5.5: Interpolation tests - Summary of all PMs, DDMs, and HMs performance assessments.

Model	Algorithm	MAE (kN)	MAPE (%)	R^2 (-)
PM-a	(Holtrop & Mennen, 1982)	13.1	6.8	0.96
PM-b	(Holtrop & Mennen, 1984)	13.3	6.6	0.96
DDM-a	Linear Ridge (V_s^1)	25.8 ± 10.7	32.5 ± 31.0	0.86 ± 0.11
DDM-b	Linear Ridge (V_s^3)	13.6 ± 1.8	14.4 ± 6.0	0.96 ± 0.03
DDM-c	Random Forest	18.3 ± 3.9	12.0 ± 4.1	0.93 ± 0.06
DDM-d	Kernel Ridge	10.0 ± 5.7	8.9 ± 5.3	0.95 ± 0.09
HM-a	Serial model	21.5 ± 10.0	12.3 ± 4.5	0.88 ± 0.09
HM-b	Parallel-residual model	8.6 ± 6.8	5.6 ± 1.4	0.97 ± 0.06
HM-c	Parallel-correction model	5.8 ± 1.5	3.8 ± 1.1	0.99 ± 0.00

5.2 Test with specific ship

At this stage, all models have been trained and tested in accordance with Chapter 4, and the interpolation results, summarized in Section 5.1 and Table 5.5, highlight the best-performing models within each modeling approach: PM-b, DDM-d, and HM-c. These results are particularly relevant in practical scenarios where designers or naval architects must proportion fuel and power systems for upcoming

projects. A key step in this process is creating a resistance curve, which estimates calm-water resistance across various speeds. This curve provides the foundation for predicting propulsion power requirements under different operating conditions. By leveraging accurate resistance estimates from high-performing models, designers can make informed early-stage decisions, optimizing the sizing of critical components like generator sets, fuel cells, electric drives and energy storage systems.

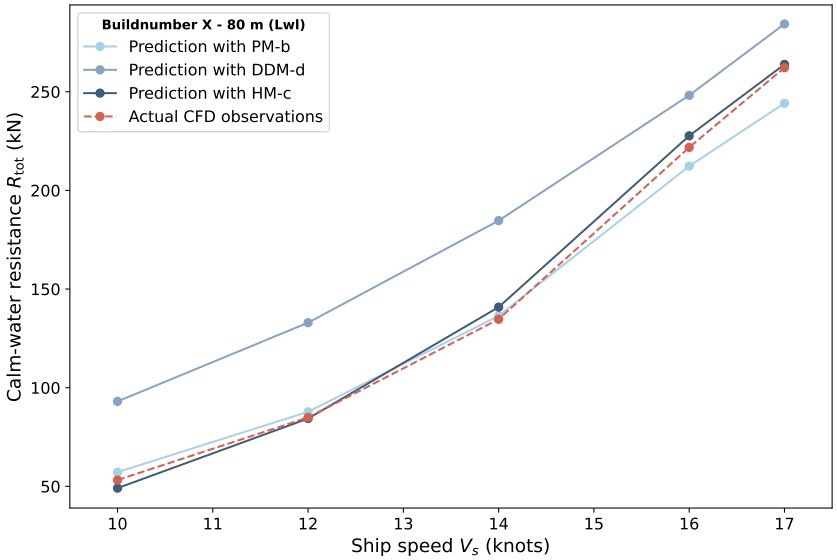


Figure 5.5: Comparison of predicted calm-water resistance curves from the best-performing models (PM-b, DDM-d, and HM-c) against actual CFD observations.

Fig. 5.5 illustrates the predictions from the best-performing models - PM-b, DDM-d, and HM-c - compared against actual CFD observations. The hybrid model (HM-c) shows the closest agreement with CFD observations across the entire speed range, showcasing its ability to combine physical principles with data-driven corrections for superior accuracy.

The physical model (PM-b) performs consistently but tends to overestimate calm-water resistance, particularly at higher speeds, reflecting limitations in its ability to capture nuanced behaviours outside its calibration range. On the other hand, the data-driven model (DDM-d), despite its strong performance in interpolation tests (Section 5.1), deviates significantly from the CFD observations in this case. This discrepancy may stem from its less stable pointwise predictions, as evidenced by the scatter plots and error distributions in Fig. 5.2, where certain cases exhibited notable variability.

These findings underscore the hybrid model's advantage in delivering reliable resistance predictions, particularly in scenarios requiring high fidelity. By integrating physical insights with data-driven adaptability, the hybrid approach proves to be the most effective for accurate and consistent resistance estimation across varying conditions.

5.3 Test with less data

This section examines model performance under limited data conditions, a challenge highlighted in the problem statement (Chapter 1). The hypothesis is that hybrid models require less data to achieve performance comparable to a state-of-the-art DDM. To test this, the most promising models - PM-b, DDM-d, and HM-c - are trained incrementally using the interpolation pipeline for consistency.

Training begins with 10% of the CFD observations from Dataset 1, with the dataset size increasing in 10% increments. At each step, the Mean Absolute Percentage Error (MAPE) is calculated to monitor performance progression as more data is added. This approach identifies the most data-efficient model, shedding light on their robustness and adaptability under data-scarce conditions. The results of this analysis are discussed in Section 5.3, and Fig. 5.6 visualizes the performance trends for the selected models: PM-b, DDM-d, and HM-c.

- **Most promising PM** - (Holtrop & Mennen, 1984) (PM-b)

- **Most promising DDM** - Kernel Ridge (DDM-d)
- **Most promising HM** - Parallel-correction architecture (HM-c)

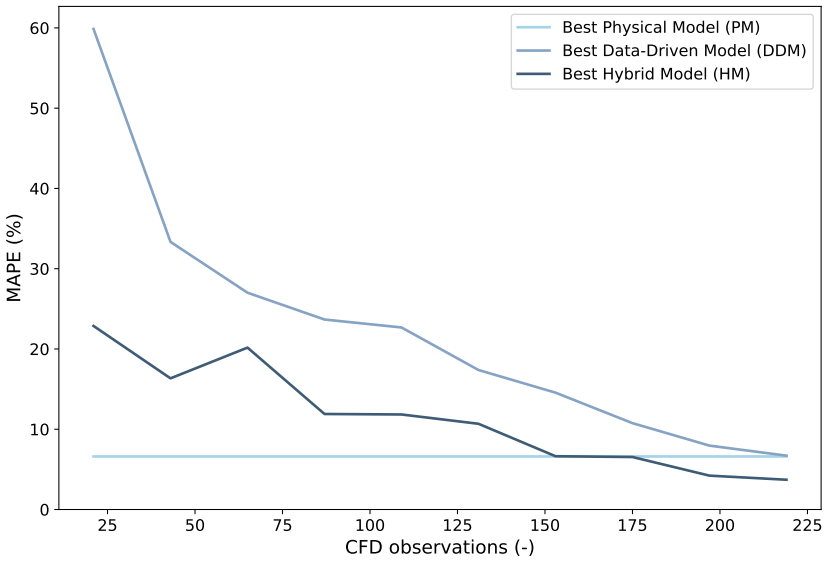


Figure 5.6: Recorded mean average percentage error (MAPE) of the most promising models (PM-b, DDM-d and HM-c) as CFD observations are added to the training dataset in incremental steps of 10%.

Figure 5.6 illustrates the progression of MAPE for the three models as the number of CFD observations increases. PM-b maintains a constant MAPE across all data sizes, reflecting its independence from the training dataset, as it relies solely on physical principles rather than data-driven learning. In contrast, DDM-d begins with a high MAPE when data availability is limited but shows significant improvement as more observations are incorporated, demonstrating its reliance on sufficient training data for accuracy. HM-c starts with an intermediate MAPE and consistently improves as the dataset grows, outperforming DDM-d across the entire range.

These results highlight distinct behaviours among the models: the DDM and HM benefit significantly from additional data, with the hybrid model offering superior accuracy and data efficiency throughout. This underscores, on top of the increase in accuracy observed in Section 5.1 and in Fig. 5.6, the hybrid model's advantage in scenarios with small datasets.

5.4 Test with other features

Accurate predictions in any model require features (input variables) that capture sufficient information about the dynamics of the prediction problem. The Holtrop & Mennen method, a leading semi-empirical method for ship performance, has undergone decades of refinement to improve feature relationships. This study leverages that established feature set as a foundation for all tests presented in Chapter 5. Yet, the unique design characteristics of twin-screw superyachts raise the question: is this feature set truly sufficient, or is there a more optimal feature set to be found for twin-screwed superyachts?

To identify the optimal feature set, a combination of one domain knowledge-based method and two statistically-driven methods for feature selection is employed. The process begins with the domain knowledge approach, leveraging insights from experienced professionals in the field. Findings from expert interviews are summarized in Table 5.6, where each feature is evaluated based on its relevance to twin-screw superyacht resistance prediction. The following experts were consulted to provide their assessments and recommendations for feature selection:

- Expert 1 - Head of Research and Development
- Expert 2 - Head of Knowledge and Innovation
- Expert 3 - Senior Specialist Design
- Expert 4 - CFD engineer

Table 5.6: Scores given by company experts on the importance of certain features, thereby using domain knowledge to construct a feature ranking, depicted in Table 5.7.

Absolute feature name	ID	Units	Expert 1	Expert 2	Expert 3	Expert 4	Total Score
Ship speed	V_s	kn	19	19	12	18	68
Length waterline	L_{wl}	m	1	17	11	20	49
Length overall submerged	L_{os}	m	18	-	-	-	18
Moulded beam	B_{mld}	m	-	14	10	17	41
Moulded mean draft	T	m	-	13	9	-	22
Volumetric displacement	∇	m^3	17	18	20	19	74
Water plane area	A_{wp}	m^2	-	11	-	-	11
Wetted surface total	S_{tot}	m^2	-	15	16	15	46
Wetted surface appendages	S_{app}	m^2	-	10	13	-	23
Wetted surface of twin-screw balance rudders	S_{rudder}	m^2	-	4	6	-	10
Wetted surface of shaft brackets	$S_{bracket}$	m^2	-	-	-	-	0
Wetted surface of skeg	S_{skeg}	m^2	-	3	-	-	3
Wetted surface of strut bossings	S_{strut}	m^2	-	-	-	-	0
Wetted surface of shafts	S_{shaft}	m^2	-	-	-	-	0
Wetted surface of stabilizer fins	S_{stab}	m^2	-	1	5	-	6
Wetted surface of bilge keels	S_{bilge}	m^2	-	2	-	-	2
Transom wetted surface	A_t	m^2	-	7	2	8	17
Transom immersion at centreline	D_t	m	-	10	7	9	26
Bulbous bow transverse area	A_{bt}	m^2	7	12	14	13	46
Height of centroid A_{bt} above keel	h_{bt}	m	6	-	-	2	8
Bow half angle of waterline entrance	i_E	deg	8	8	3	12	31
Derived/dimensionless feature name							
Froude number (length)	F_n	-	20	20	19	16	75
Block coefficient	C_b	-	16	16	17	5	54
Prismatic coefficient	C_p	-	15	-	18	6	39
Midship section coefficient	C_m	-	3	5	-	4	12
Water plane area coefficient	C_{wp}	-	2	-	8	3	13
Length-to-width ratio	L/B	-	6	-	-	10	20
Length-to-volume ratio	L/V	-	5	-	-	-	5
Length-to-draught ratio	L/T	-	4	-	-	-	4
Length-to-displacement ratio	$L/\nabla^{1/3}$	-	9	-	1	11	21
Beam-to-draught ratio	B/T	-	14	6	-	-	20
Transom-to-tranverse cross-section ratio	$A_t/(BT)$	-	13	-	-	1	14
Appendage-to-lateral cross-section ratio	$S_{app}/(LT)$	-	12	-	-	7	19
Longitudinal centre of buoyancy	ℓ_{cB}	%	11	-	15	13	39
Longitudinal centre of floatation	ℓ_{cF}	%	10	9	-	14	33

By summing the scores assigned by each individual expert, a total score is calculated for every feature in the list shown in Table 5.6. These total scores allow the features to be ranked based on their importance, reflecting the domain knowledge available within the company. This ranking is presented in Table 5.7(a). Alongside this domain-knowledge-based ranking, two additional statistical feature ranking methods are included in the same table: Backward Feature Elimination (BFE) and Permutation Importance. Their respective rankings are shown in Table 5.7(b) and Table 5.7(c).

To evaluate the effect of feature rankings on model performance, the best-performing hybrid model during interpolation tests (HM-c) was evaluated using all three ranking methods. The evaluation was conducted by iteratively training and testing the model with an increasing number of features, following the ranking order for each approach. Specifically, the model was initially trained using only the highest-ranked feature (rank 1), and the Mean Absolute Percentage Error (MAPE) was recorded. In the next iteration, the two highest-ranked features (rank 1 and rank 2) were used for training, and the corresponding MAPE was retrieved. This process was repeated incrementally, including additional features in ranking order, until all features were incorporated. The results of this iterative testing are presented in Fig. 5.7.

The results from the feature ranking evaluation (Fig. 5.7) highlight differences in the effectiveness of the three approaches - domain knowledge, Backward Feature Elimination (BFE), and Permutation Importance (PI). Across all methods, MAPE decreases rapidly as the first few top-ranked features are added, emphasizing the importance of these features. The domain knowledge approach achieves the lowest MAPE with fewer features, demonstrating the efficiency of expert-informed selection. PI shows a more uneven decline, with significant drops at specific feature thresholds, indicating sensitivity to certain features. BFE provides a smoother progression but requires more features to achieve comparable accuracy. Overall, domain knowledge delivers the most efficient feature set, while the statistical methods require more iterations to reach similar performance.

Table 5.7: Feature rankings across three approaches.

(a) Domain knowledge

Ranking	ID	Units
1	F_n	-
2	∇	m^3
3	V_s	kn
4	C_b	-
5	L_{wl}	m
6	S_{tot}	m^2
7	A_{bt}	m^2
8	B_{mld}	m
9	ℓ_{cB}	%
10	C_p	-
11	ℓ_{cF}	%
12	i_E	deg
13	D_t	m
14	S_{app}	m^2
15	T	m
16	$L/\nabla^{1/3}$	-
17	L/B	-
18	B/T	-
19	$S_{app}/(LT)$	-
20	L_{os}	m
21	A_t	m^2
22	$A_t/(BT)$	-
23	C_{wp}	-
24	C_m	-
25	A_{wp}	m^2
26	S_{rudder}	m^2
27	h_{bt}	m
28	S_{stab}	m^2
29	L/T	-
30	S_{skeg}	m^2
31	S_{bilge}	m^2
32	S_{strut}	m^2
33	S_{shaft}	m^2

(b) Backward feature elimination (BFE)

Ranking	ID	Units
1	F_n	-
2	V_s	kn
3	C_p	-
4	D_t	m
5	S_{strut}	m^2
6	C_m	-
7	$S_{app}/(LT)$	-
8	A_t	m^2
9	ℓ_{cF}	%
10	A_{bt}	m^2
11	T	m
12	L/B	-
13	S_{rudder}	m^2
14	h_{bt}	m
15	C_{wp}	-
16	C_b	-
17	L/V	-
18	∇	m^3
19	S_{skeg}	m^2
20	L_{os}	m
21	ℓ_{cB}	%
22	S_{stab}	m^2
23	B/T	-
24	L_{wl}	m
25	$A_t/(BT)$	-
26	A_{wp}	m^2
27	B_{mld}	m
28	S_{tot}	m^2
29	i_E	deg
30	$S_{bracket}$	m^2
31	L/T	-
32	S_{bilge}	m^2
33	S_{shaft}	m^2

(c) Permutation Importance (PI)

Ranking	ID	Units
1	F_n	-
2	V_s	kn
3	C_p	-
4	h_{bt}	m
5	A_{bt}	m^2
6	D_t	m
7	$S_{app}/(LT)$	-
8	C_m	-
9	A_t	m^2
10	ℓ_{cF}	%
11	∇	m^3
12	T	m
13	L/B	-
14	C_{wp}	-
15	C_b	-
16	L_{os}	m
17	A_{wp}	m^2
18	L/V	-
19	S_{tot}	m^2
20	ℓ_{cB}	%
21	B_{mld}	m
22	B/T	-
23	S_{skeg}	m^2
24	i_E	deg
25	S_{stab}	m^2
26	L_{wl}	m
27	$A_t/(BT)$	-
28	S_{bilge}	m^2
29	L/T	-
30	S_{rudder}	m^2
31	S_{strut}	m^2
32	$S_{bracket}$	m^2
33	S_{shaft}	m^2

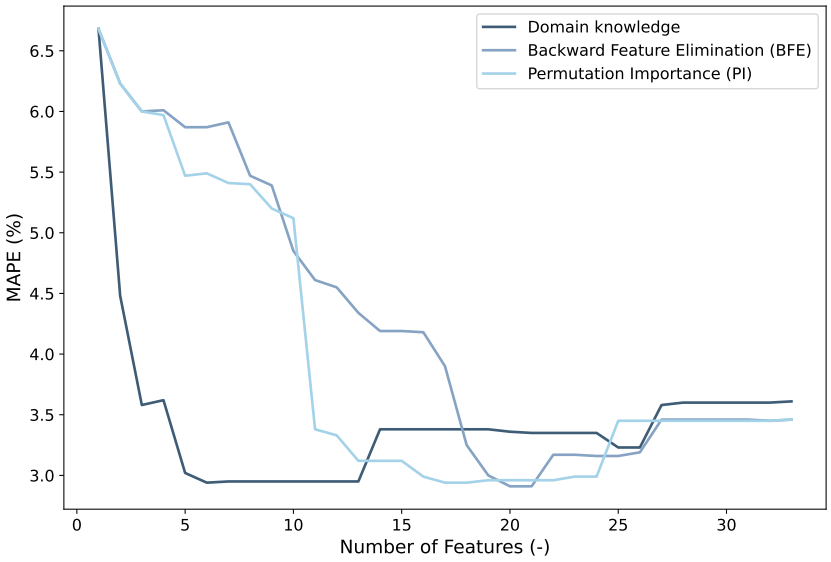
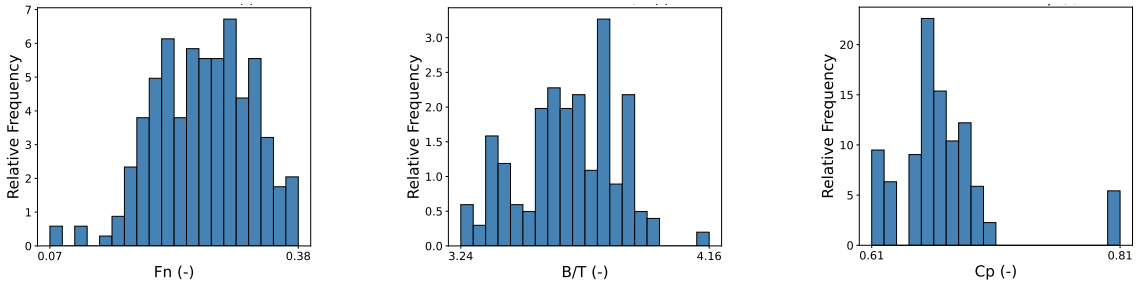


Figure 5.7: Comparison of feature selection methods (Domain Knowledge, Backward Feature Elimination, and Permutation Importance), showing the impact of only training the model on the highest ranked features, based on their importance.

5.5 Test for extrapolation condition

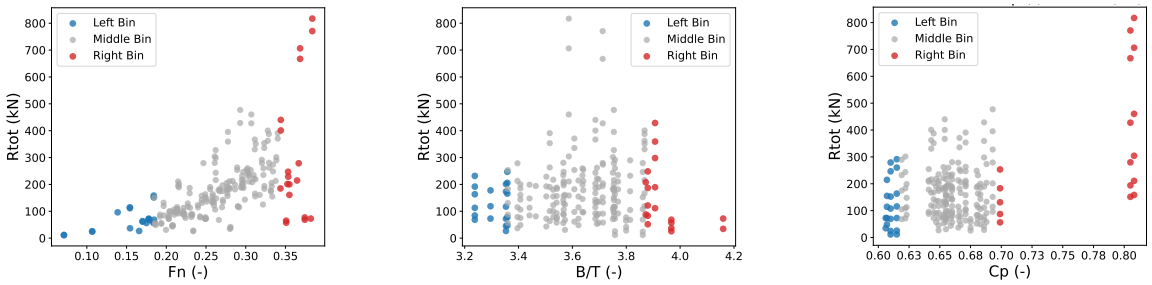
The extrapolation tests assess model performance when specific feature values are excluded from the training set, simulating conditions where predictions are required beyond the observed data range. Three scenarios are considered: Leave-one-F_n-out (LOFO), Leave-one-B/T-out (LOBO), and Leave-one-C_p-out (LOCO). Each scenario evaluates the model's ability to generalize by predicting resistance values

outside the range of a single excluded feature. Figure 5.8 presents histograms illustrating the entire range of data present for that specific feature.



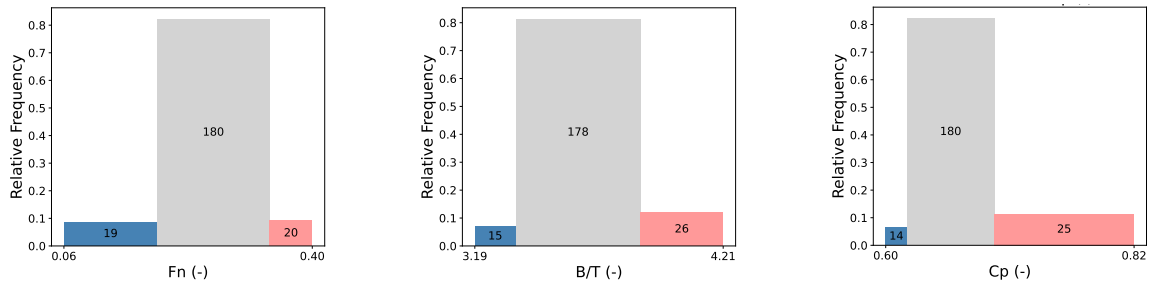
(a) Leave-one-F_n-out (LOFO) **(b)** Leave-one-B/T-out (LOBO) **(c)** Leave-one-C_p-out (LOCO)

Figure 5.8: Histogram plots visualizing data distribution for every scenario.



(a) Leave-one-F_n-out (LOFO) **(b)** Leave-one-B/T-out (LOBO) **(c)** Leave-one-C_p-out (LOCO)

Figure 5.9: Scatter plots visualizing data distribution for every scenario.



(a) Leave-one-F_n-out (LOFO) **(b)** Leave-one-B/T-out (LOBO) **(c)** Leave-one-C_p-out (LOCO)

Figure 5.10: Bin plots visualising data distribution for every scenario.

Fig. 5.9 shows scatter plots of resistance values against each excluded feature, providing a detailed view of data distribution within the bins across the full range. These plots visually clarify how the data is segmented for extrapolation, with colour-coded bins distinguishing the excluded regions from the training data. Figure 5.9(a) shows a recognizable pattern, suggesting some structure in the data, while Fig. 5.9(b) and (c) exhibit much higher variance. This variability is likely due to the wide range of ship dimensions and speeds represented in the dataset.

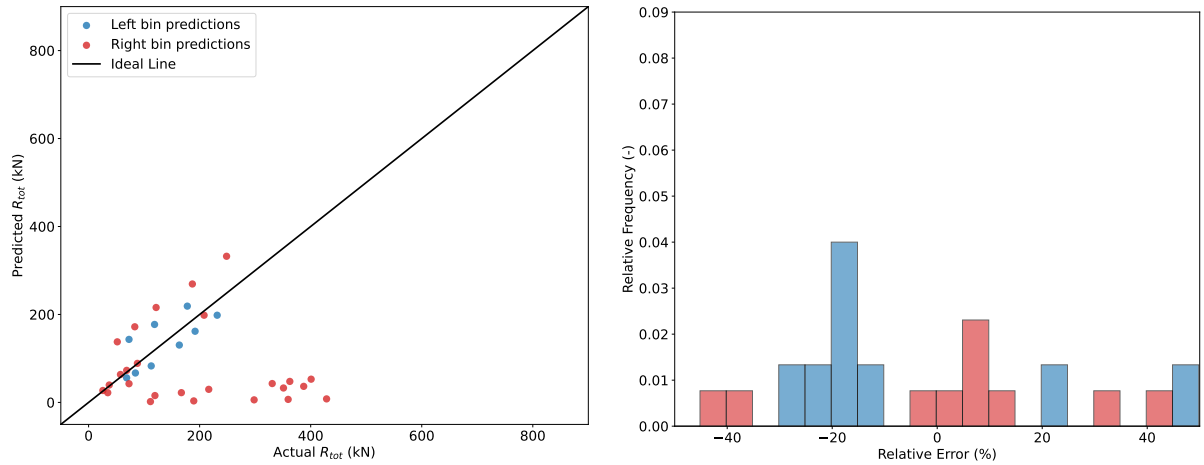
Finally, as mentioned in the extrapolation methodology Section 4.2.5, Figure 5.10 presents bin plots that categorize data into left, middle, and right bins for testing. In each scenario, the model is trained on data in the middle bin and tested on either the left or right bin, allowing for targeted evaluation of extrapolation performance. The outer bins aim to contain approximately 20 data points each. However, this is not always achievable due to the risk of data leakage. All data points associated with a specific build number (bn) are forced to remain within the same bin.

The initial extrapolation assessments are conducted using all data-driven models (DDMs). As the physical models (PMs) are already known to perform well in extrapolation, and therefore additional

testing of PMs in this context was deemed unnecessary. The results of all DDMs in extrapolation condition are presented in Table 5.8.

Table 5.8: Performance evaluation of all DDMs in the extrapolation scenarios.

		Left bin extrapolation			Middle bin			Right bin extrapolation		
		MAE (kN)	MAPE (%)	R^2 (-)	MAE (kN)	MAPE (%)	R^2 (-)	MAE (kN)	MAPE (%)	R^2 (-)
Model	**Scenario**									
DDM-a	**LOFO**	45.1	202.6	-1.04	25.8	32.5	0.86	58.6	26.1	0.88
DDM-a	**LOBO**	10.2	11.0	0.96	25.8	32.5	0.86	26.9	34.2	0.93
DDM-a	**LOCO**	26.5	98.1	0.80	25.8	32.5	0.86	56.5	20.4	0.88
DDM-b	**LOFO**	13.1	34.0	0.80	13.6	14.4	0.96	37.1	31.4	0.97
DDM-b	**LOBO**	11.1	16.6	0.96	13.6	14.4	0.96	18.9	25.2	0.97
DDM-b	**LOCO**	10.9	23.7	0.97	13.6	14.4	0.96	27.8	12.7	0.97
DDM-c	**LOFO**	21.8	86.0	0.51	18.3	12.0	0.92	109.9	54.9	0.60
DDM-c	**LOBO**	9.6	8.7	0.95	18.3	12.0	0.92	15.1	9.1	0.98
DDM-c	**LOCO**	6.4	5.3	0.99	18.3	12.0	0.92	73.0	16.8	0.64
DDM-d	**LOFO**	15.4	36.9	0.73	7.6	7.0	0.99	16.2	7.7	0.99
DDM-d	**LOBO**	139.2	182.0	-8.91	7.6	7.0	0.99	150.7	66.7	-1.43
DDM-d	**LOCO**	41.6	126.2	0.62	7.6	7.0	0.99	292.4	85.8	-1.99



(a) Actual vs. Predicted values (kN)

(b) Relative errrors (%)

Figure 5.11: Extrapolation results of DDM-d for the LOBO scenario, illustrating reduced accuracy and increased variability compared to its strong interpolation performance. Detailed metrics are in Table 5.8.

The DDM extrapolation results in Table 5.8 starkly contrast with their interpolation performance, underscoring the significant challenges of predicting beyond the training range. DDM-a fails dramatically in the left bin, with MAPE exceeding 200%, reflecting its inability to generalize effectively, and even in the right bin, its performance remains unreliable with only moderate improvements. DDM-b performs somewhat better, but its errors in the left bin and reduced reliability in extrapolation scenarios highlight its limitations compared to interpolation. DDM-c shows highly inconsistent performance, with poor results in the left bin under LOFO and only achieving strong outcomes under specific conditions like LOBO and LOCO in the right bin. Its overall reliability, however, remains questionable. DDM-d performs moderately well in the LOFO scenario, achieving reasonable accuracy, but its performance declines significantly in other scenarios, such as LOBO and LOCO, where large errors and high variability are observed.

The scatter plot and error distribution in Fig. 5.11 further emphasize the challenges faced by DDMs under extrapolation. While right-bin predictions cluster closer to the diagonal (ideal line), left-bin predictions exhibit a wide spread, leading to large relative errors, as shown in the histogram. These results confirm that none of the models maintain their interpolation-level accuracy in extrapolation

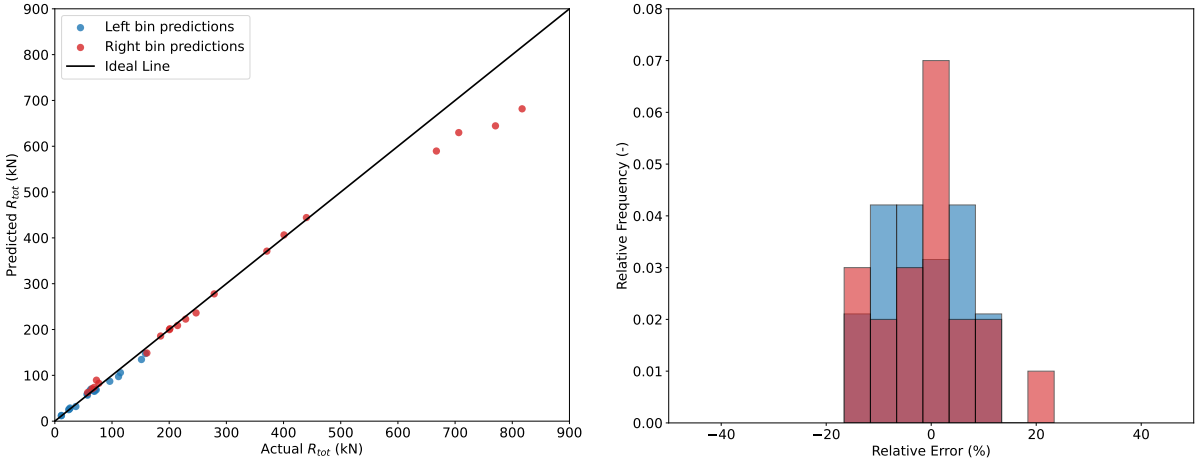
scenarios, underscoring the inherent difficulty of predicting resistance values beyond the observed training range.

Table 5.9: Performance evaluation of the optimal hybrid architecture (HM-c) with a parallel-correction configuration, tested under the LOFO scenario, with all data-driven models assessed in combination with this configuration.

		Left bin extrapolation			Middle bin			Right bin extrapolation		
		MAE (kN)	MAPE (%)	R^2 (-)	MAE (kN)	MAPE (%)	R^2 (-)	MAE (kN)	MAPE (%)	R^2 (-)
Model	**Fusion**	**LOFO**								
HM-c	**DDM-a**	7.4	10.9	0.94	9.3	5.3	0.98	21.8	11.3	0.99
HM-c	**DDM-b**	7.4	9.8	0.93	9.0	5.0	0.98	45.4	12.6	0.88
HM-c	**DDM-c**	5.1	6.8	0.97	5.4	3.3	0.99	25.1	6.6	0.96
HM-c	**DDM-d**	4.8	10.0	0.98	5.8	3.8	0.99	41.8	8.2	0.88

While the DDM results in Table 5.8 highlight significant challenges in extrapolation scenarios, hybrid models (HMs) offer an opportunity to address these limitations by combining the strengths of physical models with data-driven corrections. To explore this potential, the best-performing HM configuration (HM-c) is evaluated under the LOFO scenario, integrating physical insights with each DDM.

Table 5.9 shows that HM-c improves extrapolation accuracy across all bins compared to standalone DDMs in the Leave-one-F_n-out (LOFO) scenario, where specific Froude number ranges are excluded from the training set. The HM-c and DDM-c combination performs best in this scenario, achieving the lowest MAPE and highest R^2 in the left bin, while also maintaining strong accuracy in the middle and right bins.



(a) Actual vs. Predicted values (kN)

(b) Relative errrors (%)

Figure 5.12: Performance of the optimal configuration for the LOFO extrapolation scenario: the hybrid model (HM-c) combined with the data-driven model (DDM-c).

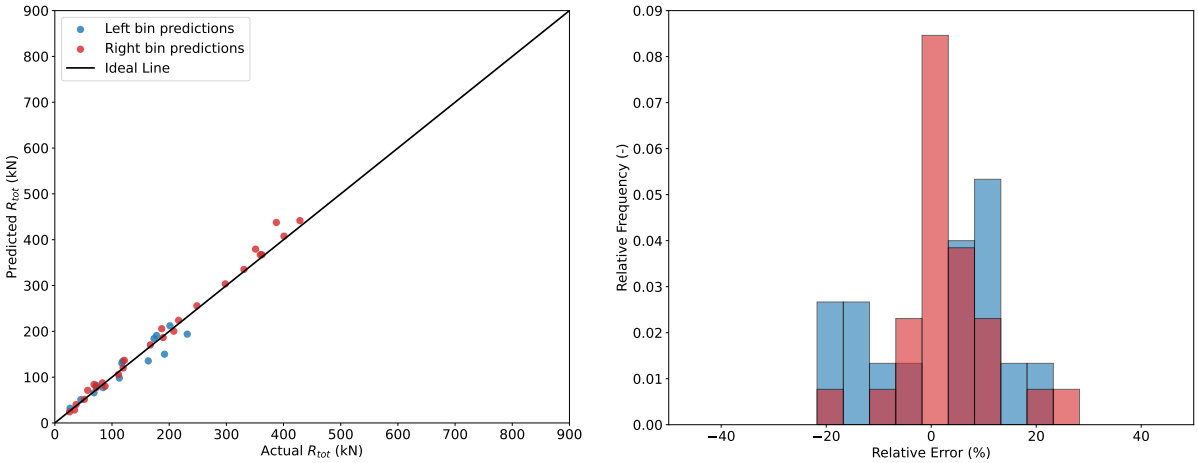
The scatter plot and error distribution in Fig. 5.12 provide further insights into these improvements. Previously, standalone DDMs struggled significantly with left bin extrapolation in the LOFO scenario, showing large variability and high error magnitudes. This is no longer the case with HM-c, as left bin predictions now align more closely with the ideal line, reflecting significantly reduced error magnitudes. The error distribution also shows narrower and more balanced relative errors across all bins, demonstrating the hybrid model's ability to address the weaknesses of DDMs.

Table 5.10 shows that HM-c significantly improves extrapolation accuracy across all bins in the LOBO scenario. The HM-c and DDM-c combination achieves the best overall performance, with low MAE (14.8 kN) and MAPE (11.9%) in the left bin and strong results in the middle and right bins, maintaining high R^2 values near 1. In contrast, the HM-c and DDM-d combination performs poorly in the left bin, with high errors (MAPE: 32.9%) and a negative R^2, indicating instability.

Figure 5.13 illustrates these results visually. The left bin predictions from HM-c paired with DDM-c align closely with the ideal line, showing reduced variability compared to standalone DDMs. The

Table 5.10: Performance evaluation of the optimal hybrid architecture (HM-c) with a parallel-correction configuration, tested under the LOBO scenario, with all data-driven models assessed in combination with this configuration.

		Left bin extrapolation			Middle bin			Right bin extrapolation		
		MAE (kN)	MAPE (%)	R^2 (-)	MAE (kN)	MAPE (%)	R^2 (-)	MAE (kN)	MAPE (%)	R^2 (-)
Model	**Fusion**	**LOBO**								
HM-c	**DDM-a**	4.3	4.4	0.99	9.3	5.3	0.98	21.9	10.9	0.91
HM-c	**DDM-b**	3.9	4.0	0.99	9.0	5.0	0.98	11.3	6.3	0.98
HM-c	**DDM-c**	14.8	11.9	0.90	5.4	3.3	0.99	9.5	6.6	0.99
HM-c	**DDM-d**	39.9	32.9	0.05	5.8	3.8	0.99	138.9	56.0	-1.16



(a) Actual vs. Predicted values (kN)

(b) Relative errrors (%)

Figure 5.13: Performance of the optimal configuration for the extrapolation LOBO scenario: the hybrid model (HM-c) combined with the data-driven model (DDM-c).

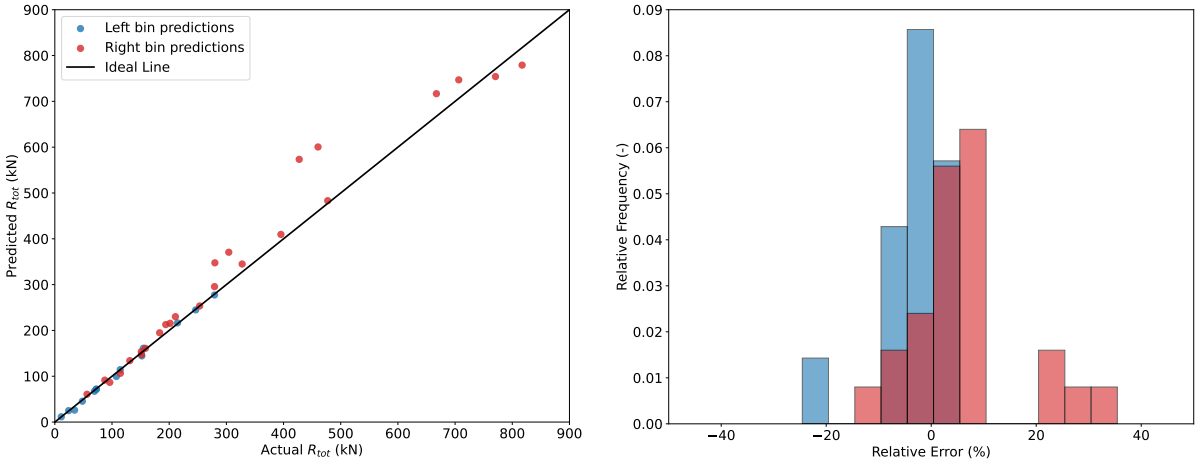
error histogram further confirms this, with a tighter distribution of relative errors centered near zero, particularly for left bin predictions, demonstrating improved consistency and accuracy across bins.

Table 5.11: Performance evaluation of the optimal hybrid architecture (HM-c) with a parallel-correction configuration, tested under the LOCO scenario, with all data-driven models assessed in combination with this configuration.

		Left bin extrapolation			Middle bin			Right bin extrapolation		
		MAE (kN)	MAPE (%)	R^2 (-)	MAE (kN)	MAPE (%)	R^2 (-)	MAE (kN)	MAPE (%)	R^2 (-)
Model	**Fusion**	**LOCO**								
HM-c	**DDM-a**	4.3	5.9	1.00	9.3	5.3	0.98	34.9	8.9	0.93
HM-c	**DDM-b**	3.1	4.2	1.00	9.0	5.0	0.98	28.8	8.7	0.95
HM-c	**DDM-c**	5.2	6.0	0.99	5.4	3.3	0.99	48.7	12.0	0.85
HM-c	**DDM-d**	15.6	21.8	0.95	5.8	3.8	0.99	287.4	80.8	-1.96

Table 5.11 highlights the performance of HM-c in the LOCO scenario, where specific prismatic coefficient (C_p) ranges are excluded from the training set. The HM-c and DDM-b combination achieves the best results, with low errors in the left bin (MAE: 3.1 kN, MAPE: 1.4%) and consistently strong performance across all bins. In contrast, the HM-c and DDM-d configuration performs poorly in the left bin (MAPE: 21.8%) and fails in the right bin, with extremely high errors (MAE: 257.4 kN, MAPE: 80.8%) and a negative R^2.

Figure 5.14 illustrates the performance of the HM-c and DDM-c combination in the LOCO scenario. The left bin predictions align closely with the ideal line, while the right bin predictions show some larger errors in the middle resistance range. The error histogram is centered around zero with limited variability, though the spread is slightly wider compared to the results for LOFO (Fig. 5.12) and LOBO (Fig. 5.13). Despite these slight inconsistencies, the HM-c and DDM-c combination maintains strong overall performance in the LOCO scenario.



(a) Actual vs. Predicted values (kN)

(b) Relative errrors (%)

Figure 5.14: Performance of the optimal configuration for the extrapolation LOCO scenario: the hybrid model (HM-c) combined with the data-driven model (DDM-c).

To summarize, the extrapolation tests highlight the poor performance of pure DDMs, which exhibited high variability and large errors across all scenarios. Surprisingly, kernel ridge-based DDMs were particularly unstable, often producing large deviations and demonstrating a lack of reliability. Even when fused with PM-b (Holtrop & Mennen, 1984) using the parallel-correction (HM-c) approach, these models remained inconsistent, yielding suboptimal results. In contrast, the hybrid approach (HM-c) demonstrated significant improvements in extrapolation accuracy by effectively integrating physical models with data-driven corrections. Among the tested configurations, the HM-c and DDM-c combination consistently delivered the best performance, achieving low errors and high R^2 values, and proving to be the most robust and reliable option for extrapolation scenarios.

Chapter 6

Discussion

In this chapter, the results of the experimental tests are discussed. The discussion begins with a detailed interpretation of the results from each experimental set, providing insights into their specific outcomes, as outlined in Section 6.1. Following this, the study's key findings are evaluated in relation to related work, highlighting areas of consistency and the validity of the findings/contributions, as discussed in Section 6.2. Finally, the discussion addresses the limitations of the study in Section 6.3 and provides recommendations for future research to build upon the insights of this study in Section 6.4.

6.1 Interpretation of results

This section provides insights into the factors influencing model performance across several tests: interpolation, with a specific ship, with less data, with other features, and extrapolation. It highlights key drivers of PM, DDM, and HM performance, offering context and guidance for future research and applications in ship resistance modeling.

6.1.1 Interpolation

The initial experimental tests were thoughtfully designed to serve two purposes: ensuring that each PM, DDM, and HM performed accurately within a widely-known approach (nested k-fold cross validation) - before being exposed to more complex scenarios like extrapolation - and to set a baseline performance to compare one model to another.

To summarize the results, the hybrid models (HMs) show clear superiority over both physical models (PMs) and data-driven models (DDMs), with the parallel-correction hybrid (HM-c) achieving the highest accuracy and stability. The PMs perform remarkably well, offering moderate accuracy and consistent results, though they show increased variance at higher resistance ranges. Among the DDMs, DDM-d (Kernel Ridge) performs best, closely followed by the linear ridge model with transformed speed input (DDM-c); both capture trends flexibly but are occasionally affected by data noise, resulting in some significantly off predictions.

The following analysis examines the underlying factors and potential drivers behind the observed variations in **PM interpolation performance**:

1. **Expected superiority of Code 7**: A slight improvement in model performance is observed for PM-b, compared to PM-a, which could be attributed to the expanded data sample in (Holtrop & Mennen, 1984) (Code 7). This code includes 334 ship models, representing a wider design variance in the data used for Code 7.
2. **Increasing absolute errors**: The scatter plot in Fig. 5.1(a) clearly demonstrates that absolute errors increase noticeably with higher target resistance values, particularly in the upper resistance range. However, the relative errors (%) remain relatively consistent, as shown in the error distribution in Fig. 5.1(b).

The following analysis examines the underlying factors and potential drivers behind the observed variations in **DDM interpolation performance**:

1. **Transforming features effective**: The only difference between DDM-a and DDM-b is the transformed speed input for the latter, resulting in $MAPE$ decrease from 32.5% to 14.4%. This demonstrates the effectiveness of domain-specific knowledge, such as the cubed speed-resistance relation, when constructing simple predictive models.
2. **Pointwise predictions can be tricky**: Some literature (Coraddu, Oneto, et al., 2022) already posed the challenge for DDMs to make pointwise predictions, meaning that on average, their accuracy is high, but not pointwise. Therefore, in some cases, DDMs can provide physically inconsistent predictions and this phenomenon is observed in Fig. 5.2.

The following analysis examines the underlying factors and potential drivers behind the observed variations in **HM interpolation performance**:

1. **PM output as DDM input too weak**: In HM-a, the PM's output is treated as an additional feature for the data-driven model (DDM), rather than directly adjusting the PM output as in HM-b and

HM-c. This means that the DDM is effectively trying to learn a mapping from both the original features and the PM output to the target, rather than explicitly correcting the PM's predictions. As a result, the PM's insights may become "diluted" when mixed with the other input features, limiting the DDM's ability to leverage the PM's baseline accuracy directly in its final prediction.

2. **Proportional adjustment favoured over additive adjustment**: In HM-b, the DDM learns an additive residual, which applies a fixed correction regardless of the magnitude of the PM's output. This can be limiting when the errors in the PM predictions scale with the target. For example, at higher values of the target, a proportional error might require a larger correction to bring the prediction close to the true value. HM-c's multiplicative correction naturally scales with the PM output, enabling it to adjust the predictions proportionally across different ranges, which may lead to greater accuracy, especially in scenarios with a wide range of values.

6.1.2 With specific ship

The purpose of the tests with a specific ship is to construct a resistance curve for the best models from each modeling approach - physical models (PM), data-driven models (DDM), and hybrid models (HM) - providing insights into their practical applicability and performance in predicting calm-water resistance.

To summarize the results, Fig. 5.5 shows that the hybrid model (HM-c) achieves the most accurate predictions, closely matching CFD observations across all speeds by combining physical principles with data-driven corrections. The physical model (PM-b) slightly overestimates the resistance at higher speeds, while the data-driven model (DDM-d) deviates significantly due to unstable pointwise predictions.

The following analysis examines the underlying factors and potential drivers behind the observed variations in **tests with specific ship**:

1. **Challenges with DDMs**: The resistance curves in Fig. 5.5 highlight the pointwise prediction challenges of data-driven models (DDMs), specifically the Kernel Ridge-based model (DDM-d), which shows significant deviations from the actual CFD observations for this ship. This underscores the need for caution when relying solely on purely data-driven approaches.
2. **Slight overshooting by HM-c**: In Fig. 5.5, HM-c slightly overcorrects the output of PM-b for lower resistance values, ending up slightly below the actual CFD values. This is likely influenced by the high errors associated with the pure data-driven model (DDM-d).

6.1.3 With less data

The purpose of the tests with less data was to evaluate the performance of the best PM, DDM, and HM models under limited data conditions by measuring the mean average percentage error (MAPE) as training data increases incrementally from 10% to 100% of CFD observations.

To summarize the results, the test depicted in Section 5.3 showed the added-value of the limited data requirement of most promising hybrid model (HM-c), compared to the most promising data-driven model (DDM-d) in interpolation scenario. As CFD observations were randomly added to the training data, HM-c constantly outperformed DDM-d and even ended up performing better than the best physical model (PM-b).

The following analysis examines the underlying factors and potential drivers behind the observed variations in **tests with less data**:

1. **Consistently lower hybrid model error with less data**: As highlighted in the introduction (Chapter 1), access to high-volume, high-velocity, and high-variety data - collectively referred to as Big Data - is often constrained. Therefore, it is encouraging to observe that the best hybrid model consistently outperforms other DDMs across the entire range (10%–100%) of CFD observations, with its advantage being most pronounced at the 10% end of the spectrum and gradually diminishing as more data becomes available for training.
2. **DDM might catch up with the HM**: A valid question to ask is whether the DDMs will eventually catch up with the HMs, as the amount of observations goes to infinity. Given the trends observed, the hypothesis is that with infinite data, the performance of the data-driven model (DDM) would continue to improve and might eventually match or slightly surpass that of the hybrid model

(HM), as purely data-driven approaches generally excel with large datasets. In a case where unlimited data is available with plenty of design variations, any arbitrary DDM is able to directly learn the variations, while an HM will always need to correct the imperfections of the physical model (PM). Still, in case of the parallel-correction hybrid model, one could debate that this HM will also learn to perfectly correct the PM in all cases. Expected is that the HM would reach a performance plateau sooner, suggesting that the HM remains more effective and stable at lower data volumes. Future research would need to investigate if the DDM would eventually catch up or even surpass the HM.

6.1.4 With other features

The purpose of the tests with other features was to determine whether a more optimal feature set exists for twin-screw superyachts compared to the traditional feature set based on (Holtrop & Mennen, 1984).

To summarize the results, the test described in Section 5.4 highlighted the advantages of feature selection guided by domain knowledge. While both statistical methods - backward feature elimination (BFE) and permutation (PI) - proved effective, this test demonstrated that domain knowledge-based feature selection remains the most reliable approach. This is largely because ship resistance is a well-studied prediction problem within the industry, with decades of research providing deep insights. However, the predictive modeling approaches presented in this study could also be applied to other problems in the naval architecture domain. In this case, a domain knowledge-based feature selection method would only be effective if the prediction problem is similarly well understood.

The following analysis examines the underlying factors and potential drivers behind the observed variations in **tests with other features**:

1. **Existence of an optimal number of features**: A notable finding from these tests is the identification of an optimal number of features for maximised model performance. As detailed in Section 5.4, all feature selection methods showed a trend where MAPE values decreased when critical features were added, indicating improved performance. However, beyond a certain threshold, adding more features led to an increase in MAPE values, suggesting that excessive features introduce noise or redundancy, ultimately reducing model efficiency.

   Why is there an optimal number of features existing? The current hypothesis is that at some point, a selection of features could explain all variance, which is sometimes already reached with the five highest-ranked features. After this, extra features will only add extra noise to the predictions.

6.1.5 Extrapolation

The purpose of the extrapolation tests in this study was to evaluate the models' performance in predicting calm-water ship resistance when exposed to data outside the training set's range.

To summarize the results, the best DDM during interpolation (DDM-d: Kernel Ridge) performed dramatically poor in basically all extrapolation scenarios (LOFO, LOBO and LOCO), for both left and right bin extrapolation. This was highly improved by the parallel-correction hybrid model (HM-c) configuration, and multiple DDMs were tested in conjunction with this hybrid configuration. These tests revealed that either the linear ridge model with transformed speed input (DDM-d) or the random forest model (DDM-d) showed the best results for extrapolation.

The following analysis examines the underlying factors and potential drivers behind the observed variations in **DDM extrapolation performance**:

1. **Poor extrapolation performance of all DDMs**: The literature review highlighted the challenge for DDMs to achieve strong performance in extrapolation scenarios—a limitation that was anticipated. However, the severity of their poor performance, as shown in Section 5.5, was unexpected, particularly for the Random Forest model (DDM-c) and the Kernel Ridge model (DDM-d). This poses a significant challenge for the hybrid modeling approach, as these DDMs require substantial corrections to outperform the PMs, which currently set the standard.

The following analysis examines the underlying factors and potential drivers behind the observed variations in **HM extrapolation performance**:

1. **Parallel-correction approach able to correct a lot**: When comparing the DDM results (Table 5.8) with the HM-c results (Table 5.9, Table 5.10 and Table 5.11) for extrapolation, there can be concluded that the parallel-correction approach is capable of applying really effective corrections. All MAPE values for the DDMs rarely fell below 10% in any scenario, with some exceeding 180%. In contrast, the HM-c model consistently demonstrated MAPE values below 12.6% across all extrapolation scenarios, except when fused with the poorly performing DDM-d. The poor performance of HM-c in combination with DDM-d can be attributed to the dramatically poor performance in extrapolation conditions, as can be seen in bTable 5.8

6.2 Comparison with previous studies

In the previous section, the interpretation of the results was discussed in detail. Here, the focus shifts to comparing these findings with existing literature.

6.2.1 Selection of hybrid model

This study focused on testing two types of hybrid models: the serial and parallel approaches. These were chosen primarily because they have consistently delivered significant performance improvements in the literature and are relatively straightforward to construct.

Hybrid models were first applied in naval architecture by Leifsson et al. (2008), achieving a 65% error reduction in fuel consumption predictions with both a serial and parallel approach. Subsequent applications by Mei et al. (2019) and Skulstad et al. (2021) demonstrated improved predictions for ship motion and positioning using both serial and parallel approaches. More recent studies, such as Odendaal et al. (2023) on energy consumption and Kalikatzarakis et al. (2023) on underwater radiated noise, showcase the versatility and effectiveness of parallel approaches, Kalikatzarakis et al. (2023) with a recursive step included in the pipeline.

Finding, computerizing and validating a physical model can be time-intensive and very challenging. Once the physical model is functioning correctly, integrating it with data-driven models becomes relatively straightforward. This simplicity is another key reason these approaches were chosen.

6.2.2 Best hybrid model

It would not make sense to compare actual error values of previous studies, as the algorithm's success is context-dependent, as described in the no-free-lunch theorem (Adam et al., 2019). Though, it is interesting to compare the relative performance of hybrid models within specific studies, as a comparative analysis between these models is conducted in this thesis as well. For this comparison, it must be noted that the exact test condition must be investigated as well, since this could highly vary from the test methodology in this study. Still, by aligning current results with past research, this section aims to highlight consistencies, discrepancies, and most of all, find further grounds for the conclusions based on the results of the experiments.

To the author's best knowledge, only (Leifsson et al., 2008) was able to construct both a serial and parallel hybrid model and actually test it. This study showed marginally small differences in both approaches, describing a slight improvement in root mean square error (RMSE) for the parallel approach. This thesis also found a superior model performance for the parallel-approach, but the improvements were significantly larger than what is found in (Leifsson et al., 2008).

6.3 Preconditions and limitations

Predictive modeling is a powerful tool for calm-water ship resistance prediction, but its effectiveness is heavily influenced by the quality of the data, the underlying assumptions, and the context in which models are applied. This section outlines the key preconditions necessary for successful model development and highlights the limitations that emerged during this study. By understanding these factors, the challenges associated with hybrid models and data-driven approaches can be better addressed. Recommendations for mitigating these challenges are provided in Section 6.4.

6.3.1 Using historical data

It should be pointed out that there is a considerable difference between the validation of a predictive model developed from a designed experiment and a model developed from data collected without the

aid of an experimental design (e.g. historical data). In a designed experiment, all features (variables) are supposedly held constant except those varied according to the design; ensuring that all relevant factors are accounted for. In this way, the true importance of a specific feature can be studied, which is highly favoured by predictive models forecasting a numerical value. However, a systematic and wide variation of all feature's numeric value is rarely present in historical data. And it should be mentioned, for complex problems like ship resistance, even with the aid of a designed experiment, achieving such systematic variation in features is not feasible in many cases.

In the context of a ship's geometry, many design parameters are interdependent due to physical constraint. For instance, changes in hull length waterline may inherently affect displacement, longitudinal centre of floatation, and wetted surface area. This interconnectedness means that varying one feature while holding others constant is often impractical or impossible. As a result, achieving systematic and independent variation of all features is unfeasible, which complicates the modeling process and can affect the predictive accuracy of models developed from such data. In regression methods, this phenomenon is known as multicollinearity, where supposedly independent variables are highly correlated with each other.

Not all prediction problems suffer from multicollinearity. Consider the classic housing price prediction problem, where features such as the square footage of the house, distance to the nearest city centre, and the number of schools in the vicinity are used. These are relatively easy prediction problems, first, because extensive housing price datasets are available, but more importantly, because these features are typically uncorrelated, making the problem free from multicollinearity. Though, in naval architecture or in engineering in general, many variables have some physics or geometry-based correlation, and care must be taken in such cases. Recommendations are included in Section 6.4.

6.3.2 Using misaligned data

A very common problem in current Feadship datasets is the presence of what seems to be misaligned data. Unless working with time-series data, any data that is used for predictive purposes must accurately capture a specific state or "snapshot" of the system in time. In shipbuilding, projects evolve rapidly, which might lead to alternative system characteristics. Common practice is to define these system changes in design revisions, but it is vital for the data to capture these changes as well. If this is not carefully monitored and misalignment between features and/or the target is present, the model's performance will decline, especially in cases where limited data is present. Recommendations are included in Section 6.4.

6.3.3 Using sparse data

Another common problem in current Feadship datasets is the presence of sparse data. In each model aiming to predict a numerical value, there are features X and one or multiple targets Y, which have to be defined in advanced of the model training. Any arbitrary selected prediction algorithm will learn X and Y for every observation (single data point), and in the most ideal case, all observations can be used. Though, a common problem in historical data is that a large proportion of the data points are zero, null, or missing, meaning it contains many inactive values compared to meaning-full entries. This is known as sparse data or a sparse dataset.

It should be mentioned that, when even a single feature or target value is missing in an observation, the complete data point becomes unusable. This could be a very small percentage of the total dataset, but when this percentage is bigger, a significant portion of the dataset needs to be removed, resulting in a loss of what could have been really valuable information. Recommendations are included in Section 6.4.

6.3.4 Using predictive analytics

The results presented in Chapter 5 basically showed that accuracy of the most-promising DDMs and HMs is very good when making prediction in interpolation condition. And while the parallel-correction hybrid model (HM-c) showed significant improvement for extrapolation condition, predictive analytics still remains challenging and tricky in these cases. Among the considerations further detailed in Chapter 7, a modeling approach can primarily be selected based on time/cost/effort constraints, the need for extrapolation, and the desired level of model accuracy. In case of calm-water ship resistance, the three modeling approaches presented in this study (PM, DDM and HM) are possible, and computational fluid dynamics (CFD). The decision framework for selecting the appropriate modeling approach is

depicted in Fig. 6.1.

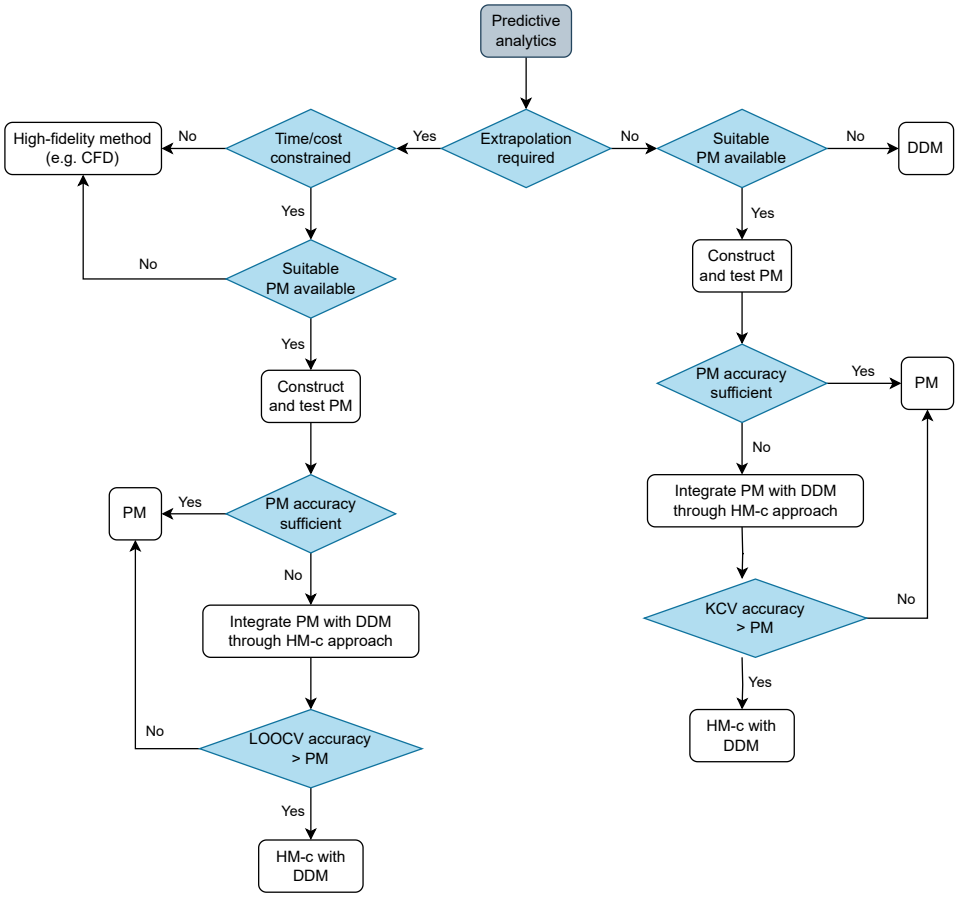


Figure 6.1: Decision framework for selecting predictive modeling approaches, incorporating physical models (PM), data-driven models (DDM), hybrid models, and CFD, based on time constraints, extrapolation requirements, and model accuracy.

In Fig. 6.1, it becomes evident that the decision-making process is highly influenced by whether extrapolation (see explanation in Section 4.2.5) is present for the prediction problem. When time, cost, and effort constraints are absent, it is recommended to conduct computational fluid dynamics (CFD) simulations, as a majority of the resistance predictions fall within interpolation conditions. Thus, when extrapolation is present and when time is contrained, CFD seems appropriate.

When time, cost, and effort are constrained, identifying an appropriate physical model (PM) becomes essential. But what defines a 'suitable' PM in context of hybrid modeling? Based on the findings of this study, the following preconditions are established:

1. **Extrapolation robustness:** In extrapolation scenarios, it makes no sense to use a PM with narrower permissible parameter ranges than the available training data. A check for these permissible ranges can be performed, as demonstrated in Section 4.1.1.
2. **Physical plausibility:** The PM's behaviour should align with the underlying physical theory relevant to the problem. For ship resistance, this requires the PM to logically decompose resistance into components such as wave-making resistance, viscous resistance, form resistance, and appendage resistance, among others. Each component must adhere to hydrodynamic principles and follow established scaling laws.
3. **Computable:** Considering future advancements toward prescriptive analytics (described in Section 2.2), it is crucial that the PM is computable to enable seamless integration with the DDM
4. **Validation:** The PM should include a predefined validation example/benchmark, enabling validation even before the integration with any DDM. For the PMs used in this study, (Holtrop & Mennen, 1984; Holtrop & Mennen, 1982) included such an example.

5. **Gaussian fit:** Certain DDMs perform optimally with a Gaussian-shaped distribution of the target data (Tax et al., 2023), among them ridge and kernel-based models, which are used in this study. Parallel hybrid models learn either the residuals or correction factors required to alter the PM's output in the right direction. For this reason, these data distribution are ideally Gaussian shaped.

The final precondition mentioned at point 5, a Gaussian fit, deserves a bit more explanation. As mentioned, some DDMs favour a Gaussian-shaped distribution of the target data. For this purpose, it helps to consult the data distribution (Fig. 6.2), presenting all correction factors required to alter the physical model's output to match the CFD values. Remember that these correction factors are learned by the DDM in the parallel-correction hybrid model (HM-c) in this study. When plotted in a histogram (Fig. 6.2), deviations from the Gaussian-fit can be detected, such as: When plotted in a histogram (Fig. 6.2), deviations from the Gaussian fit can be detected, such as:

- **Skewness:** Asymmetry in the distribution, with longer tails on one side, indicating non-uniform nature of the data.
- **Heavy Tails (Leptokurtic):** More extreme values than expected on the tails of the bell-curve, suggesting sensitivity to outliers.
- **Light Tails (Platykurtic):** Fewer extreme values than expected on the tails of the bell-curve, indicating a lack of variability in residuals.
- **Multimodality:** Multiple peaks in the distribution, indicating distinct subgroups or behaviours within the data.
- **Outliers:** Individual extreme values far from the main data cluster, suggesting inconsistency in the data.

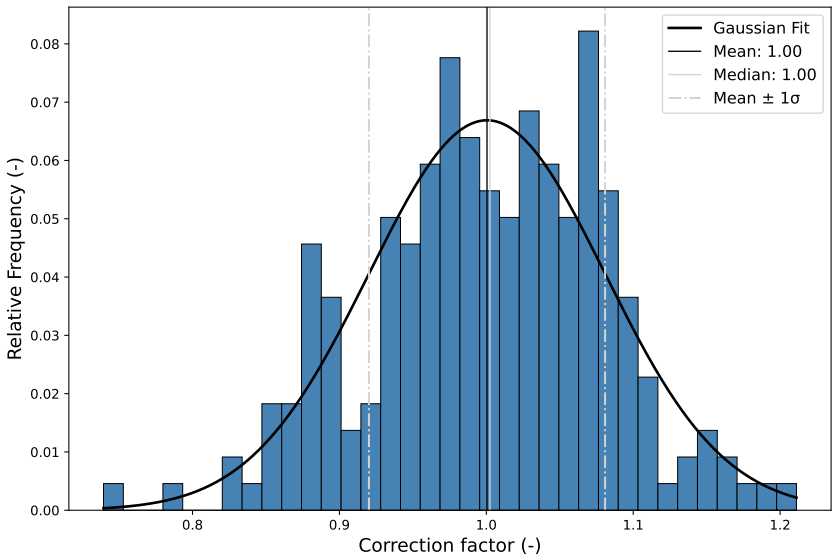


Figure 6.2: Gaussian fit for all correction factors required to alter the output of the physical model (PM-b) (Holtrop & Mennen, 1984), such that the output matches the actual CFD observation.

In addition to deviation from the Gaussian fit, this histogram plot (Fig. 6.2) can also serve more as an exploratory tool that is highly recommended once the PM is developed and coded.

6.4 Recommendations

Ultimately, this research and its findings provide only an initial glimpse into the broader potential of these methodologies. The experiments conducted were notably limited, with the models being evaluated solely on a narrowly focused dataset of computational fluid dynamics (CFD) results from ships spanning approximately the last ten years. Consequently, the insights gained may have limited applicability to other challenges within the field of naval architecture or related disciplines. A follow-up study should incorporate these lessons to enhance the reliability and scope of the results.

6.4.1 Overcoming limitations

Several limitations, primarily related to data and predictive models, are discussed in Section 6.3. This section offers recommendations for addressing these challenges.

Using historical data - In cases where correlated features are inevitable, model performance can often improve by either removing one of the correlated features or creating a combined feature. Methods such as Pearson and Spearman correlation matrices, along with Variance Inflation Factor (VIF) analysis, are valuable for identifying these correlations.

A further recommendation concerns the number of features utilized in the model. In the context of historical data, there is often a specific balance in feature selection that yields the best results. For all feature selection methods examined in Section 5.4, MAPE values decreased as important features were added, but beyond a certain point, adding more features caused MAPE values to increase again. It is strongly recommended to evaluate this during model development, as the phenomenon is straightforward to investigate, and no model should be more complex than necessary.

Using misaligned and sparse data - To prevent losing valuable information, the data can be enhanced through a process called feature or dataset enrichment. This involves carefully adding data that aligns with the system's state at a specific point in time, done in Section 3.3. This is a time-intensive task requiring thorough validation; if alignment or accuracy cannot be verified and there have been system changes over time, it is better to remove such observations, as it makes little sense for the algorithm to learn spurious relationships from incorrect data.

Using predictive models - In case of real-world adoption, it is essential to communicate the limits of the modeling approach, among them the presence of extrapolation. The developer of the predictive model is most likely aware of the model limitations, but the developer has little control over its use, when adopted in real-world context. A potential remedy is to integrate the model limits in the code itself, alarming the user when the training bounds are exceeded.

For the decision-making on the selection of modeling approaches, see Fig. 6.1.

6.4.2 Future work

This study has identified several areas for improvement that deserve further exploration in future research.

Automatic pipeline for hydrostatics retrieval - Section 5.3 demonstrated that model performance improves as the number of observations in the training data increases for both DDMs and HMs. However, this study revealed that retrieving reliable and consistent hydrostatic data is a highly time-intensive process. To facilitate future studies or real-world adoption, developing an automated solution for this retrieval process would be highly beneficial.

Further improvement of DDM performance in extrapolation - The findings in Section 5.5 clearly indicate that the extrapolation performance of the parallel-correction hybrid model is limited by the poor performance of its data-driven component. While the hybrid approach proves highly effective overall, improving the data-driven model remains a critical challenge. A follow-up study focusing specifically on enhancing the accuracy and reliability of data-driven predictions under extrapolation conditions would be highly beneficial. Emerging fields such as symbolic regression and physics-informed neural networks show promise in addressing this issue by also incorporating physical principles into data-driven models, but differently.

Optimization methods - The hybrid architecture developed in this study provides a very suitable platform for optimization (prescriptive analytics). The hypothesis is that finding optimization methods will not be the most challenging part, but rather validating if the true feature-target relation is capture by the hybrid architecture.

Categorical data - This study relies solely on numerical data for ship resistance prediction, but the current hypothesis is that incorporating categorical variables such as hull type (e.g., displacement, planing), bow types (e.g., straight, flared or bulbous bow) and driveline (e.g. conventional shaft lines or pods), immersed transom (yes/no) and trim wedge (yes/no) could significantly enhance model performance. Not all DDMs, but certain ones, such as Kernel Ridge Regression, can effectively learn from categorical data when it is appropriately encoded.

Chapter 7

Real World Adoption

Within the company, numerous of data-driven studies have been conducted for very domain-specific problems. And with success. (Odendaal et al., 2021) paved the way by improving energy consumption estimates, achieving propulsion and auxiliary power predictions for varying operational conditions within 3% and 9% of actual values. Shortly after, (De Haas et al., 2022) developed a model for marine biofouling growth prediction, demonstrating a 6.9% improvement in accuracy over previous methods. Last year, (Opstal et al., 2023) advanced HVAC energy demand estimation, reaching an accuracy of 91.3%.

All of these studies, including this thesis, used hybrid architectures for their prediction strategy and all proved to be highly effective. And although there is a diverse array of challenges within the company well-suited for this type of predictive (supervised learning) models, there still often seems to be a preference for traditional methods or even first-principles. The question, therefore, is not whether hybrid models are effective, but rather what is required to successfully adopt and scale this new wave of innovation. That is the question here, and without going into the company's internal boundaries of adoption, this chapter presents a brief study into potentional solution approaches.

7.1 A brief note on innovation and technological readiness

There's no shortage of terms to describe innovation. We hear about incremental innovations, continuous improvement initiatives, and organic growth programs. Concepts like white spaces, blue oceans, and red oceans add even more colour to the mix. And the list seems to go on forever.

Though, the definitions that resonate most with me come from The Innovator's Dilemma by (Christensen, 1997), a book that I found on Steve Jobs's favourites list, which categorizes innovation into two buckets: sustaining innovations, which enhance existing technologies for established customers, and disruptive innovations, which redefine industry standards by transforming markets. A similar concept is presented in the Harvard Business Review (HBR) article "Building an Innovation Engine in 90 Days" by (Anthony et al., 2014), which describes "core" and "new-growth innovations" and targets the same underlying idea. Both offers a straightforward framework for understanding how innovations impact both established markets and emerging ones.

A more recent HBR article from (Satell, 2017), revises this dichotomy and uses a more extensive framework with four types - basic research, sustaining innovations, breakthrough innovation, and disruptive innovation - where innovation are categorized based on *how well we can define the problem?* and *how well can we define the skill domain(s) needed to solve it?* In his framework, which is based on (Christensen, 1997)'s work, basic research focuses on poorly defined problems in unstructured domains, often exploring entirely new areas of knowledge. Sustaining innovations address well-defined problems in structured domains, refining and improving existing systems. Breakthrough innovations solve well-defined problems in unstructured domains, requiring novel approaches to achieve significant leaps forward. Disruptive innovations target poorly defined problems in structured domains, transforming markets by redefining how existing systems operate.

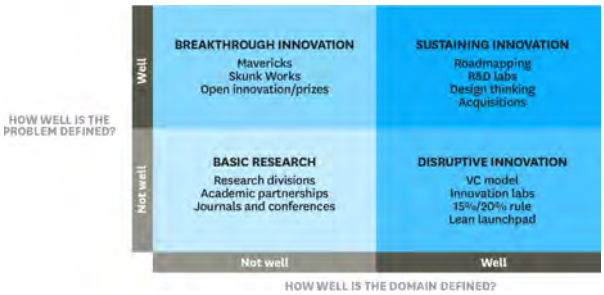


Figure 7.1: The four types of innovation and the problems they solve (Satell, 2017).

Understanding the readiness of an innovation is just as critical as understanding its type. Technology Readiness Levels (TRLs), introduced by (Mankins, 1995) at NASA, provide a structured framework for evaluating the maturity of a technology, from early research (TRL 1) to operational deployment (TRL 9). While originally developed for aerospace, TRLs are now widely used to manage innovation risk across industries. Basic research typically aligns with lower TRLs (1–3), focusing on foundational principles, while sustaining innovations progress through mid-level TRLs (4–6) as technologies become practical applications. Breakthrough innovations, representing novel solutions, operate in mid-to-high TRLs

(5–8) as they are validated and scaled. Disruptive innovations can span TRLs 3–9, evolving from early exploration to redefining markets and achieving operational deployment (Mankins, 1995; Satell, 2017).

7.2 Current state of hybrid models

This section discusses the classification, technological readiness, and value proposition of hybrid models.

7.2.1 Classification of innovation type

According to (Satell, 2017)'s framework, hybrid models are potentially best classified as *breakthrough innovations*. Why? Well, because these models require a clear problem definition with well-defined input variables and prediction target. At the same time, though, the domain in which hybrid models operate is not well-defined as their applications extend far beyond calm-water ship resistance - ship weight estimations, project planning estimations, cost estimations, and more. These two characteristics, a well-defined problem and poorly defined domain, strongly position hybrid models as breakthrough innovation.

7.2.2 Classification of TRL level

At this stage, as mentioned in this Chapter 7's introduction, several configurations of hybrid models (De Haas et al., 2022; Odendaal et al., 2021; Opstal et al., 2023) are developed, tested and deemed effective for several naval architecture use cases. However, it is important to note that the current state of these hybrid model codes remains at the proof-of-concept stage, as the thesis projects primarily focused on training and testing these model types. Actual implementation into existing engineering workflows and tools would require a different model architecture. Therefore, the technology readiness level (TRL) of hybrid models currently lies between 3 and 4 (Mankins, 1995), as they have undergone analytical validation and proof-of-concept testing in a "laboratory" environment but have not yet been demonstrated in a relevant operational setting.

7.2.3 Value proposition

The value proposition of the hybrid models:

- **Reduced computational time:** Achieving results within 2% error of high-fidelity CFD simulations under interpolation conditions is not far away, as more and more CFD observations will be added to the training set. This can be accomplished with only a fraction of the preparation and operational time, enabling faster exploration of the design space (J. Harvey Evans, 1959), compared to CFD simulation.
- **Reduced data requirement:** Considerably less data is required to equal the accuracy of modern data-driven models (DDMs), as tested in Section 5.3.
- **Enhanced extrapolation capability:** Although caution remains necessary when using hybrid models under extrapolation conditions, Section 5.5 highlights their potential to substantially outperform modern DDMs in both interpolation as extrapolation scenarios, particularly when incorporating the recommendations outlined in Chapter 6.
- **Enhanced quality and consistency:** During literature review, it was found that naval architects seek to tailor several existing estimation approaches (Section 2.1.3) to yacht design, forcing them to apply explicit or implicit corrections to these methods. This creates inconsistencies in, for example, early-stage ship resistance forecasts, which hybrid models address by autonomously applying the necessary corrections.
- **Enhanced integration:** The computationally accessible nature of hybrid models allows for seamless integration with tools like optimization algorithms (Section 2.2), aligning with the company's envisioned future developments.

7.3 Strategy for adoption

Organizations aiming to achieve transformative change through internally developed innovations can gain valuable insights by adopting strategies inspired by the venture capital and startup ecosystem. Much of the literature referenced in this section is grounded in these approaches.

7.3.1 Opportunity to start data lake ecosystem

Hybrid models provide an excellent first use case for developing a data lake ecosystem. Configuring such frameworks (Fig. 7.2) can be challenging without clear use cases, but hybrid models come with well-defined data requirements, simplifying the development process. Once the data lake structure is in place, integrating additional applications, such as other machine learning (ML), artificial intelligence (AI), or even business intelligence (BI) tools, becomes relatively straightforward, in essence only requiring the creation of alternative data pipelines. As Feadship strives to enhance its data literacy and governance in the coming years, this represents a significant step in the right direction.

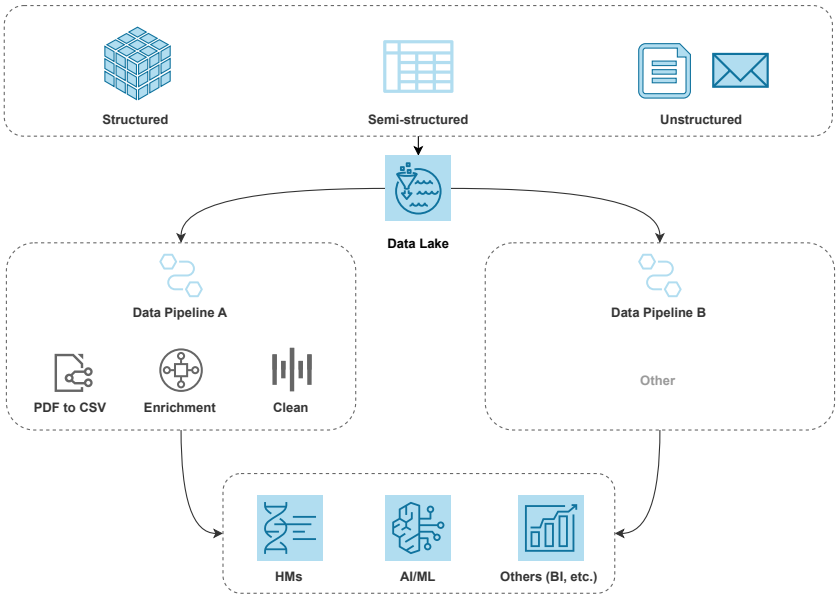


Figure 7.2: Simplified visualization of a data lake ecosystem, showcasing raw data (structured, semi-structured, and unstructured), its ingestion into the data lake, and the creation of data products through pipelines for hybrid models (HMs) and other applications such as machine learning (ML), artificial intelligence (AI), and business intelligence (BI).

7.3.2 Succeeding in a niche

One of the key principles for successfully adoption of innovations is to first demonstrate their value within a well-defined niche before attempting to scale their application across broader domains. Drawing inspiration from concepts outlined in Crossing the Chasm by (Moore, 1991) and The Lean Startup (Ries, 2011), a targeted, focused approach can mitigate risks and provide the foundation for widespread adoption.

In the context of hybrid models for ship resistance prediction, the niche can be defined as early-stage design within the Feadship fleet, where traditional methods struggle with a balance between accuracy and efficiency. This niche is ideal because it aligns with a clear pain point: early-stage designs require rapid, reliable predictions, yet resource-intensive CFD simulations or error-prone empirical methods are the current norm. Though, success in the niche is yet to be achieved, as integration in real-world engineering workflows would require a different model architecture (see Section 7.2.2).

7.3.3 Developing trust with pilot-projects

Once success is achieved in a niche, the next step is to expand into adjacent markets (or use cases), building on the credibility and insights gained from the initial focus. Geoffrey Moore refers to this initial focus as the *beachhead* (initial market) in Crossing the Chasm (Moore, 1991), where securing dominance in the beachhead facilitates a gradual and strategic expansion into larger or related markets. In the context of the hybrid models, this entails an expansion to other pilot-projects.

This raises the question of other areas where hybrid models could provide added value. It is crucial to emphasize that the availability of a suitable physical model (PM) is one of the most important preconditions, as outlined in Section 6.3.4. With this precondition in place, potential pilot-projects could include:

- **Bio-fouling growth:** (De Haas et al., 2022) has already contributed here, but the work can be expanded upon by applying the new hybrid model (HM-c) from this study and incorporating new ship data. Actual adoption of a hybrid approach could yield immediate benefits, given the pressing and contemporary nature of this challenge.
- **Wave bending moments:** A potential new use case could be forecasting the wave bending moments for a new ship. A task that currently requires time and domain-knowledge to convert a wave scatter diagram (from operational area) into wave bending moments for several longitudinal nodes of the ship.

Not having a suitable physical model does not have to limit the data-driven opportunities. The pure data-driven models (DDMs) Chapter 5 showed remarkable performance for interpolation condition, which entails the majority of the predictions. In this case, examples of potential pilot-projects could include:

- **Predictive maintenance**: Forecasting maintenance needs for yachts, which could provide immediate benefits as limited building/maintenance slots are available yearly. A more streamlined planning process could enhance operational efficiency in this case.
- **Ship weight:** Forecasting (sea) trial displacement or draughts for new ship designs, which remains a challenging task.
- **Deck high-stacking** Forecasting the required height for each deck to optimize design and usability.
- **Project planning:** Forecasting construction timelines to improve planning accuracy.
- **Cost estimations:** Forecasting overall project costs, including materials, labour, and operational expenses.

7.3.4 Refinement

At this stage, significant success has been achieved for multiple pilot-projects with either data-driven models or hybrid models, ideally integrated within the data lake ecosystem. This presents an opportunity to further refine model architectures as they are tested against real-world problems, diverse datasets, and user feedback. Additionally, the process from raw data to actionable data products within the data lake ecosystem can be streamlined.

7.4 People and skills

The innovation matrix (Fig. 7.1) by (Satell, 2017) highlights various approaches to adopting breakthrough innovations (like hybrid models), including skunkworks, which are small, autonomous teams within organizations that focus on innovative projects with minimal bureaucracy. This concept originated from Lockheed Martin's "Skunk Works," famed for pioneering aircraft designs during World War II. This centralized approach to innovation is also found in recent literature (Anthony et al., 2014) on innovation broadly, as well as in two HBR articles focusing specifically on building an AI-powered organization (Fountaine et al., 2019) and on competing with data analytics (Davenport, 2006), even talking about the need for an "überanalytics" group.

The need for a dedicated group to support analytics innovation still feels distant for the company, as there are currently few analytic innovation projects that are scalable - even at the basic research phase. Consequently, hiring specialists with distinct skills feels not yet practical. Think of skills like *data engineering* (building and optimizing data pipelines), *data science* (extracting insights from data), *software development* (creating scalable applications), *UX design* (ensuring user-friendly interfaces), and *AI innovation management* (overseeing the strategic adoption of AI technologies).

Until that day arrives - having a pipeline of multiple advanced and scalable analytic innovations - it seems more practical to focus on hiring well-rounded individuals with strong analytical skills and a problem-solving mindset. And I believe this approach can go a long way in these modern times. At Delft University of Technology, an entirely new generation of engineers is about to graduate, that has already embraced modern learning tools like large language models (LLMs), enabling everyone to quickly acquire expertise, even in unfamiliar domains. And this is happening, having seen fellow students pushing their master's theses beyond their programs. In my view, it is perfectly valid for individuals to specialize in a single skill set, but it has never been easier to acquire additional ones.

Chapter 8

Conclusion

For many shipbuilders, the majority of a vessel's lifecycle greenhouse gas emissions occurs during its operational phase, commonly referred to as downstream emissions. For Feadship, this challenge is particularly pronounced, with downstream emissions accounting for 94% of a superyacht's carbon footprint. Addressing this majority requires accurate and efficient methods to predict propulsion energy use during the design stage - a task hindered by the limitations of existing prediction methods, which are either time-intensive or prone to significant errors. This is affecting the design process in two ways: an increased risk of incorrectly proportioned energy and power systems, and limited exploration of design space. Data-driven methods, based on machine learning algorithms, have been proposed in the literature. However, these methods expose two key gaps in the literature: their performance under extrapolation conditions and their limitations when applied to small datasets. This thesis responds to these challenges by developing hybrid modeling approaches that combine physical insights with data-driven techniques, addressing both these gaps.

8.1 Sub-questions

In the introduction of this thesis, the research was divided into several sub-questions. These sub-questions address different aspects of the research question. The sub-questions are answered individually below:

What specific challenges persist within the Feadship design and engineering process, and what are their main causes?

Feadship aims to reduce downstream emissions, facing challenges that are two-fold: an increased risk of improperly proportioned energy and power systems and limited exploration of the design space. A primary contributing factor is the reliance on either inaccurate and/or time-intensive prediction methods, particularly for propulsion energy use in twin-screwed superyachts, as outlined in Section 2.1.4.

What knowledge gaps exist in data-driven methods, and which modeling approaches could address these challenges?

Data-driven (machine learning) models struggle with extrapolation and limited datasets (Chapter 2), posing challenges in ship design. Hybrid models, combining physical and data-driven approaches, show promise in addressing these gaps by enhancing extrapolation capabilities and reducing data requirements. This study utilized serial and parallel hybrid models for their proven effectiveness in the literature, with parallel models showing superior performance and significantly improving prediction accuracy.

What datasets are available relating to propulsion use, and which dataset offers the best suitability for this study?

Four datasets were assessed (Chapter 3): Dataset 1 (CFD resistance), Dataset 2 (CFD power), Dataset 3 (towing tank tests), and Dataset 4 (speed-power trials). At the outset of this study, the hypothesis was that sea trial data might serve as the best source; however, significant uncertainties were identified, stemming from the trial measurements and post-correction methods. Instead, Dataset 1 was selected as the most suitable due to its high-precision data from a controlled CFD environment and the ability to exactly match ship geometries with resistance results, alongside its balanced design variance and moderate sample size.

What methodologies ensure comprehensive training, testing, and evaluation of the models?

Two distinct pipelines are required: for interpolation and extrapolation condition. Both pipelines achieve robust training and testing through a process called nested k-fold cross validation, responsible for selecting the best hyperparameters (model selection). After model selection, re-training is required with the best model through either ordinary k-fold cross-validation (for interpolation) or through leave-one-out cross validation (for interpolation). In total, three model types are tested - physical models (PMs), data-driven models (DDMs) and hybrid models (HMs) - PMs are tested directly, while DDMs and HMs follow training and testing methodologies outlined in Chapter 4.

Which of the tested exhibited superior performance in interpolation and extrapolation scenarios?

The tested models (Chapter 5) consist of two PMs (Holtrop & Mennen, 1984; Holtrop & Mennen, 1982), four DDMs (ridge regression, cubed-speed ridge regression, random forest, and kernel ridge), and three HMs (serial, parallel-residual, and parallel-correction), which integrate the strengths of the best-performing DDM and PM.

The parallel-correction hybrid model (HM-c), a novel configuration developed in this study, demonstrated the highest accuracy in interpolation, achieving a mean average percentage error (MAPE) of 3.8% and a determination coefficient (R^2) of 0.99. The physical models (PM-a and PM-b) provided a reliable baseline, with PM-b performing best (MAPE: 6.6%, R^2: 0.96), though it was less precise than HM-c. Among the data-driven models (DDMs), Kernel Ridge (DDM-d) performed best, achieving moderate accuracy (MAPE: 8.9%, R^2: 0.95), but it was still outperformed by both HM-c and the physical models. See results in Section 5.1.

In extrapolation, HM-c again proved superior, achieving robust accuracy across all scenarios (e.g., MAPE below 12.6% in LOFO, LOBO, and LOCO conditions). Standalone DDMs, including the best interpolator (DDM-d), failed dramatically, with MAPE values exceeding 180% in some cases. However, robust performance was observed with Random Forest (DDM-c), and when combined with PM-b, the HM-c/DDM-c configuration achieved the best results in extrapolation, effectively integrating PM robustness with data-driven flexibility for improved generalization beyond the training range. See results in Section 5.5.

How does a reduction in dataset size impact the performance of various models?

HM-c consistently outperformed DDM-d under limited data conditions, with its competitive edge most pronounced at low data availability (10% of CFD observations). Notably, there is a crossover point where HM-c surpasses the performance of the physical model (PM-b) as more CFD observations are added. While DDM-d demonstrated steady improvements with increasing data, potentially surpassing HM-c in scenarios with unlimited CFD observations, both models are expected to plateau in performance at some point. Based on the observed trends, the HM-c is likely to reach this plateau sooner, but future research is required to determine whether the DDM could ultimately exceed the HM's performance. See results in Section 5.3.

What is the most optimal features set for twin-screwed superyachts, and what are the best methods to do this feature selection?

Feature selection methods, including Backward Feature Elimination (BFE), Permutation Importance (PI), and a domain knowledge-based approach, enhance model performance by identifying and prioritizing relevant features (Chapter 5). While all methods improved accuracy and reduced computational complexity by removing less-informative variables, the domain knowledge-based approach achieved slightly better results, underscoring the value of expert insights in optimizing feature sets. See results in Section 5.4.

8.2 Research question

Using the answers to the sub-questions, the main research question can now be addressed:

"How can hybrid models improve early-stage predictions for calm-water ship resistance in extrapolation scenarios, while using small datasets with limited design variability?"

Instead of directly learning from the CFD resistance data, it appears more effective when the data-driven model learns to apply corrections to the output of a PM. Where traditionally these corrections were based on the naval architect's experience, they are now driven by data, offering fast and accurate alternative to existing methods. This philosophy is embodied in this study through a newly developed parallel HM, which achieves superior performance by learning how to apply these corrections to the PM's output automatically. During interpolation, the new HM demonstrates a mean average percentage error (MAPE) of 3.8%, outperforming PM (6.7%) and DDM (8.9%). For extrapolation, the new HM maintains average errors within 12% across scenarios. And with less data, the new HM consistently outperformed the best DDM, with its competitive edge most pronounced at low data availability (10% of CFD observations). By advancing these methodologies, the study not only enhances early-stage design confidence but also contributes to future steps towards automated design optimization.

References

Adam, S., Alexandropoulos, S., Pardalos, P., & Vrahatis, M. (2019). *No Free Lunch Theorem: A Review* (I. C. Demetriou & P. M. Pardalos, Eds.; Vol. 145). Springer International Publishing. https://doi.org/10.1007/978-3-030-12767-1

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

Anthony, S., Duncan, D., & Siren, M. (2014, December). *Build an Innovation Engine in 90 Days* (tech. rep.). Harvard Business Review. Boston, MA.

Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, *84*, 317–331. https://doi.org/10.1016/j.patcog.2018.07.023

Bishop, C. (1995). *Pattern Recognition and Machine Learning* (tech. rep.).

Christensen, C. M. (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business School Press.

Coraddu, A., Kalikatzarakis, M., Oneto, L., Meijn, G. J., Godjevac, M., & Geertsmad, R. D. (2018). Ship diesel engine performance modelling with combined physical and machine learning approach. *Proceedings of the International Ship Control Systems Symposium*, *1*. https://doi.org/10.24868/issn.2631-8741.2018.011

Coraddu, A., Kalikatzarakis, M., Walker, J., Ilardi, D., & Oneto, L. (2022, January). Data science and advanced analytics for shipping energy systems. In *Sustainable energy systems on ships: Novel technologies for low carbon shipping* (pp. 303–349). Elsevier. https://doi.org/10.1016/B978-0-12-824471-5.00014-1

Coraddu, A., Oneto, L., Baldi, F., & Anguita, D. (2017, January). Vessels Fuel Consumption Forecast and Trim Optimisation: A Data Analytics Perspective. https://doi.org/10.1016/j.oceaneng.2016.11.058

Coraddu, A., Oneto, L., Baldi, F., Anguita, D., & Member, S. (2015). *Ship Efficiency Forecast based on Sensors Data Collection: Improving Numerical Models through Data Analytics* (tech. rep.).

Coraddu, A., Oneto, L., Cipollini, F., Kalikatzarakis, M., Meijn, G., & Geertsma, R. (2022). Physical, data-driven and hybrid approaches to model engine exhaust gas temperatures in operational conditions. *Ships and Offshore Structures*, *17*(6), 1360–1381. https://doi.org/10.1080/17445302.2021.1920095

Davenport, T. H. (2006, January). Competing on Analytics. www.hbrreprints.org

De Groot, D. (1955). *Resistance and Propulsion of Motor-Boats* (tech. rep.). Member of the Research Department, Netherlands Ship Model Basin (N.S.M.B). Wageningen, Holland.

De Haas, M., Coraddu, A., & Baptista, M. L. (2022). *Early Stage Fouling Effects Prediction for Yacht Design A grey-box model approach using operational voyage data MT54035: MT MSc Thesis* (tech. rep.).

Duboue, P. (2020). *The art of feature engineering : essentials for machine learning*. Cambridge University Press.

Ferziger, J. H., & Peric, M. (2012). *Computational Methods for Fluid Dynamics*.

Fountaine, T., Mccarthy, B., & Saleh, T. (2019, August). Building the AI-Powered Organization.

Gerritsma, J., Onnink, R., & Versluis, A. (1981, December). *Geometry, Resistance and Stability of the Delft Systematic Yacht Hull Series* (tech. rep. No. 328). Delft University of Technology. Delft.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning* (tech. rep.). MIT press.

Guldhammer, H., & Harvald, S. (1974). Ship Resistance: Effect of Form and Principal Dimensions. *Akademisk Forlag, Copenhagen*.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning Data Mining, Inference, and Prediction* (tech. rep.). Springer Science \& Business Media. New York, NY.

Hollenbach, K. (1998). Estimating Resistance and Propulsion for Single-Screw and Twin-Screw Ships. *Schiffstechnik/Ship Technology Research*, 72–76.

Holtrop, J., & Mennen, G. G. (1984). *An Approximate Power Prediction Method* (tech. rep.). MARIN. Wageningen.

Holtrop, J., & Mennen, G. (1982). *An Approximate Power Prediction Method* (tech. rep.). Maritime Research Institute Netherlands (MARIN). Wageningen.

Hyndman, R., & Athanasopoulos, G. (2014). Forecasting: Principles and Practice. *International Journal of Forecasting*.

Insel, M. (2008). Uncertainty in the analysis of speed and powering trials. *Ocean Engineering*, *35*(11-12), 1183–1193. https://doi.org/10.1016/j.oceaneng.2008.04.009

J. Harvey Evans. (1959). Basic Design Concepts. *American Society of Naval Engineers (ASNE) Journal*, 671–678.

Jacquet, L., le Duigou, A., & Kerbrat, O. (2024, April). A systematic literature review on holistic lifecycle assessments as a basis to create a standard in maritime industry. https://doi.org/10.1007/s11367-023-02269-4

Kalikatzarakis, M., Coraddu, A., Atlar, M., Gaggero, S., Tani, G., & Oneto, L. (2023). Physically plausible propeller noise prediction via recursive corrections leveraging prior knowledge and experimental data. *Engineering Applications of Artificial Intelligence*, *118*. https://doi.org/10.1016/j.engappai.2022.105660

Keerthi, S. S., & Lin, C.-J. (2003). *Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel* (tech. rep.). Neural Computation.

Leifsson, L. T., Sævarsdóttir, H., Sigurdsson, S. T., & Vésteinsson, A. (2008). Grey-box modeling of an ocean vessel for operational optimization. *Simulation Modelling Practice and Theory*, *16*(8), 923–932. https://doi.org/10.1016/j.simpat.2008.03.006

Loeff, G. (2024). *Bright internal report* (tech. rep.). Feadship. Hoofddorp.

Mankins, J. C. (1995). *Technology Readiness Levels: A White Paper* (tech. rep.). NNASA Office of Space Access and Technology. Washington, D.C.

Mei, B., Sun, L., & Shi, G. (2019). White-Black-Box Hybrid Model Identification Based on RM-RF for Ship Maneuvering. *IEEE Access*, *7*, 57691–57705. https://doi.org/10.1109/ACCESS.2019.2914120

Mittendorf, M., Nielsen, U. D., & Bingham, H. B. (2022). Data-driven prediction of added-wave resistance on ships in oblique waves—A comparison between tree-based ensemble methods and artificial neural networks. *Applied Ocean Research*, *118*. https://doi.org/10.1016/j.apor.2021.102964

Molnar, C. (2020). *Interpretable Machine Learning A Guide for Making Black Box Models Explainable* (tech. rep.). http://leanpub.com/interpretable-machine-learning

Moore, G. A. (1991). *Crossing the Chasm: Marketing and Selling Technology Products to Mainstream Customers* (tech. rep.). Harper Business. New York, NY.

Odendaal, K., Pruyn, J., Alkemade, A., De Vos, P., & Baptista, M. (2021, June). *Enhancing early-stage energy consumption predictions using dynamic operational voyage data: A greybox modelling investigation* (tech. rep.). Delft University of Technology. http://repository.tudelft.nl/.

Odendaal, K., Alkemade, A., & Kana, A. A. (2023). Enhancing early-stage energy consumption predictions using dynamic operational voyage data: A grey-box modelling investigation. *International Journal of Naval Architecture and Ocean Engineering*, *15*. https://doi.org/10.1016/j.ijnaoe.2022.100484

Opstal, A., El Mouhandiz, A., & Van Biert, L. (2023, May). *A Grey Box HVAC Energy Demand Estimation For Yachts* (tech. rep.). Delft University of Technology.

Ries, E. (2011). *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses* (tech. rep.). The Crown Publishing Group. New York, NY.

Satell, G. (2017). The 4 Types of Innovation and the Problems They Solve. *Harvard Business Review Press*.

Seo, D. W., & Oh, J. (2021). Uncertainty analysis of speed–power performance based on measured raw data in sea trials. *International Journal of Naval Architecture and Ocean Engineering*, *13*, 396–404. https://doi.org/10.1016/j.ijnaoe.2021.04.001

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms* (tech. rep.). Cambridge University Press. http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning

Skulstad, R., Li, G., Fossen, T., Vik, B., & Zhang, H. (2021). A Hybrid Approach to Motion Prediction for Ship Docking - Integration of a Neural Network Model into the Ship Dynamic Model. *IEEE Transactions on Instrumentation and Measurement*, *70*. https://doi.org/10.1109/TIM.2020.3018568

Tax, D. M. J., Loog, M., & Krijthe, J. (2023). *CS4220 Machine Learning 1* (tech. rep.). Delft University of Technology.

Van Oortmerssen, G. (1971). *A Power Prediction Method and Its Application to Small Ships* (tech. rep.). Netherlands Ship Model Basin (NSMB). Wageningen.

Van Veldhuizen, B. N., Van Biert, L., Ünlübayir, C., Visser, K., Hopman, J. J., & Aravind, P. V. (2024). Preprint: Component Sizing and Dynamic Simulation of a Low-Emission Power Plant for Cruise Ships with Solid Oxide Fuel Cells. https://ssrn.com/abstract=4943022

Walker, J. M., Coraddu, A., & Oneto, L. (2024). Data-Driven Models for Yacht Hull Resistance Optimization: Exploring Geometric Parameters Beyond the Boundaries of the Delft Systematic Yacht Hull Series. *IEEE Access*, *12*, 76102–76120. https://doi.org/10.1109/ACCESS.2024.3404495

Appendices

A

Review physical models

In the literature review, various physical models are examined, and to assess the complexity of implementing such models, a systematic overview of their input and output parameters is developed. This comprehensive analysis is presented in the appendix.

Table A.1: Required and optional input parameters for (De Groot, 1955) method.

Parameter	ID	Remarks
	required parameters	
ship speed	V_s	
length waterline	L_{wl}	
volumetric displacement	V	
	optional parameters	
wetted surface	S	approx. $2.75\sqrt{VL_{wl}}$

Table A.2: Required and optional input parameters for (Van Oortmerssen, 1971) method.

Parameter	ID	Remarks
	required parameters	
froude number	F_n	is $V_s/\sqrt{gL_D}$
displacement length	L_D	rather than L_{wl}, calculated with $\frac{1}{2}(L_{pp} + L_{wl})$
moulded beam	B	
moulded mean draft	T	
volumetric displacement (moulded)	V	
longitudinal centre of buoyancy	FB	from forward perpendicular
midship coefficient	C_m	
prismatic coefficient	C_p	
half angle of entrance	i_E	at load (design) waterline
	optional parameters	
longitudinal centre of buoyancy	ℓ_{CB}	approx. $(\frac{1}{2}L_D - FB)/L_D * 100\%$
entrance load waterline coefficient	C_{wl}	

Table A.3: Required and optional input parameters for Guldhammer and Harvald, 1974 method.

Parameter	ID	Remarks
	required parameters	
ship speed	V_s	
length between perpendiculars	L_{pp}	usually $L_{WL} = L_{pp} + L_{aft}$
length of aft overhang in waterline	L_{aft}	
extension of S beyond fore perpend.	L_{fore}	
computation length	L	usually equal to L_{OS}
maximum moulded beam in waterline	B	
moulded draft	T	
block coefficient	C_B	or volumetric displacement V
prismatic coefficient	C_P	or midship section area A_M
transverse cross-section area of bulb	A_{BT}	at forward perpendicular FP
	optional parameters	
propeller diameter	D_P	for propulsion analysis
longitudinal centre of buoyancy	ℓ_{CB}	or assume optimum position
wetted surface (hull + rudder)	S	
wetted surface of appendages	S_{APP}	bilge keels, stabilizer fins, bossings, etc.
form factors for fore and aft body	F_F, F_A	$-3 \leq F_A, F_F \leq +3$

Table A.4: Required and optional input parameters for Delft Systematic Yacht Hull Series (Gerritsma et al., 1981), also known as the DSYHS method.

Parameter	ID	Remarks
	required parameters	
ship speed	V_s	
length waterline	L_{wl}	
beam waterline	B_{wl}	
draught of canoe body	T_c	
volumetric displacement of canoe body	∇_c	
prismatic coefficient	C_p	
longitudinal centre of buoyancy	ℓ_{cB}	
	optional parameters	
wetted surface of canoe body	S_c	with $(1.97 + 0.171 \frac{B_{wl}}{T_c}) \sqrt{\nabla_c \cdot L_{wl}}$
wetted surface of keel	S_k	not mentioned
wetted surface of rudder	S_r	not mentioned

Table A.5: Required and optional input parameters for the (Holtrop & Mennen, 1982) and (Holtrop & Mennen, 1984) method.

Feature name	ID	Notes
	Hull parameters	
Ship speed	V_s	
Length waterline	L_{wl}	
Moulded beam	B	
Moulded mean draft	T	Typically $T = \frac{1}{2}(T_A + T_F)$
Moulded draft at aft perpendicular	T_a	
Moulded draft at forward perpendicular	T_f	
Volumetric displacement (moulded)	∇	
Prismatic coefficient (based on L_{wl})	C_p	
Midship section coefficient	C_m	or use $C_m = C_b/C_p$
Waterplane area coefficient	C_{wp}	
Longitudinal centre of buoyancy	ℓ_{Cb}	Positive forward; with respect to $L_{wl}/2$
Immersed transom area	A_t	Measured at rest
Stern shape parameter	C_{stern}	Differs for every stern type
	Propulsion parameters	
Propeller diameter	D	
Number of propeller blades	Z	
Propeller chord length	$c_{0.75}$	At a radius of 75 percent
Propeller blade thickness-chord length ratio	t/c	
Propeller blade surface roughness	k_p	Standard figure for new propeller is 0.00003 m
Thrust coefficient	$K_{T,B-series}$	From B-series polynomials
Torque coefficient	$K_{Q,B-series}$	From B-series polynomials
Open-water efficiency	η_O	From B-series polynomials
Shaft efficiency	η_S	$P_d/P_s = 0.99$
	Optional parameters	
Wetted surface (hull)	S	
Wetted surface of appendages	S_{app_i}	Bilge keels, stabilizer fins, etc.
Area of ship and cargo above waterline	A_{air}	Projected in direction of V_s
Transverse area of bulbous bow	A_{bt}	Measured at forward perpendicular
Height of centroid of A_{bt} above keel	h_b	Has to be smaller than $0.6T_f$
Half angle of waterline entrance	i_E	
Diameter of bow thruster tunnel	d_{th}	

Table A.6: Required and optional input parameters for (Hollenbach, 1998) method.

Parameter	ID	Remarks
	required parameters	
ship speed	V_s	
length between perpendiculars	L_{PP}	
length in waterline	L_{WL}	check definition for ballast condition
length over wetted surface	L_{OS}	check definition for ballast condition
moulded beam	B	
moulded draft at aft perpendicular	T_A	
moulded draft at forward perpendicular	T_F	
propeller diameter	D	
block coefficient (based on L_{PP})	C_B	
transverse vertical area above waterline	A_V	for air resistance
number of rudders	$N_{rudders}$	1 or 2 (for twin screw vessels)
number of shaft brackets	$N_{brackets}$	0, 1 or 2 (for twin screw vessels)
number of shaft bossings	$N_{bossings}$	0, 1 or 2 (for twin screw vessels)
number of side thrusters	$N_{thrusters}$	between 0 and 4
	optional parameters	
wetted surface (hull)	S	
wetted surface of appendages	S_{APP_i}	bilge keels, stabilizer fins, etc.
diameter(s) of thruster tunnels	d_{TH}	for appendage resistance fins, etc.

B
Review data-driven models

In the literature review, various data-driven models are examined, and to assess the complexity of implementing such models, a systematic overview of working principles is developed. This comprehensive analysis is presented in the appendix.

B.1 Linear model

A linear regression model aims to describe the relationship between a dependent variable (target) $\mathbf{y}$ and independent variables (features) $\mathbf{X}$ through a linear equation:

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \varepsilon, \tag{B.1}$$

where:

- $\mathbf{y} \in \mathbb{R}^n$: Target variable vector.
- $\mathbf{X} \in \mathbb{R}^{n \times p}$: Feature matrix with n samples and p features.
- $\mathbf{w} \in \mathbb{R}^p$: Vector of regression coefficients.
- $\varepsilon \in \mathbb{R}^n$: Vector of errors or residuals.

The goal of linear regression is to estimate the coefficient vector $\mathbf{w}$ by minimizing the residual sum of squares (RSS):

$$\mathcal{L}(\mathbf{w}) = \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2. \tag{B.2}$$

Ridge regression, also known as Tikhonov regularization, extends the linear model by adding an L2 regularization term to mitigate overfitting and improve generalization, particularly when multicollinearity exists among the features. The Ridge regression objective function is defined as:

$$\mathcal{L}(\mathbf{w}) = \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \alpha\|\mathbf{w}\|_2^2, \tag{B.3}$$

where:

- $\alpha \geq 0$: Regularization parameter controlling the penalty term.

The first term represents the RSS, measuring the error in predicting $\mathbf{y}$ from $\mathbf{X}$. The second term penalizes large coefficients, thereby reducing overfitting. As α increases, the coefficients are shrunk more aggressively toward zero, balancing the trade-off between bias and variance.

The Ridge regression solution can be computed analytically using the closed-form expression:

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X} + \alpha\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}, \tag{B.4}$$

where $\mathbf{I} \in \mathbb{R}^{p \times p}$ is the identity matrix. The regularization term $\alpha\mathbf{I}$ ensures numerical stability by preventing singularities in the matrix inversion, particularly when $\mathbf{X}^T\mathbf{X}$ is ill-conditioned. A matrix is considered ill-conditioned if it has a high condition number, meaning small changes in the input data can cause large changes in the solution, making computations numerically unstable.

B.2 Random Forest

To understand a Random Forest, it is helpful to start with its building block: a decision tree. A decision tree is a flowchart-like structure used for making predictions. At each internal node, the tree evaluates a feature and partitions the data based on its value, creating branches that represent possible outcomes of the test. This process continues recursively until reaching the leaf nodes, which provide the predicted values. For regression tasks, these predicted values are typically the mean of the target variable in that partition. Decision trees excel at modeling complex relationships but can suffer from overfitting, especially with deep trees trained on small datasets.

A Random Forest builds on the idea of decision trees by combining many of them into an ensemble model. The core idea is analogous to consulting multiple experts to make a decision: each decision tree provides its own prediction based on slightly different subsets of the data and features. By aggregating these predictions, the Random Forest delivers a more accurate and robust output than any single tree could achieve. In regression tasks, the Random Forest predicts the output $\hat{y}$ by averaging the predictions of all the individual trees:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} \hat{y}_t, \tag{B.5}$$

where:

- T: Total number of trees in the forest.
- $\hat{y}_t$: Prediction from the t-th tree.

The ensemble nature of the Random Forest mitigates the high variance of individual decision trees by leveraging two key strategies:

- **Bootstrap Aggregating (Bagging):** Each tree is trained on a randomly sampled subset of the data, drawn with replacement. This introduces variability among the trees, reducing the overall model variance.
- **Feature Randomness:** At each split, a random subset of features is considered when determining the best partition. This decorrelates the trees, enhancing the robustness of the ensemble.

These techniques ensure that the trees in the forest are diverse and independent, making the model less prone to overfitting and improving its generalization performance. The key hyperparameters of a Random Forest include:

- **Number of Trees:** Specifies the size of the ensemble. A larger number generally leads to better performance at the cost of increased computational effort.
- **Maximum Depth:** Limits the depth of individual trees. Controlling the depth helps balance the trade-off between underfitting and overfitting.
- **Minimum Samples per Leaf:** Defines the smallest number of samples required to create a leaf node, influencing the granularity of the tree's partitions.
- **Number of Features:** Determines how many features to consider for splitting at each node, promoting diversity among trees.

By combining the simplicity of decision trees with ensemble techniques, Random Forests provide a powerful and versatile tool for regression tasks, offering strong predictive accuracy and resistance to overfitting.

B.3 Kernel Ridge

Kernel Ridge Regression (KRR) extends the linear regression model by incorporating the "kernel trick" to capture non-linear relationships between the features and the target variable. It combines ridge regression with kernel methods, enabling it to operate in a high-dimensional feature space without explicitly computing the transformation. KRR begins with the same foundation as the linear model:

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \varepsilon, \tag{B.6}$$

where $\mathbf{X}$ is the feature matrix, $\mathbf{w}$ is the vector of coefficients, and ε represents residual errors. While the linear model fits $\mathbf{w}$ directly in the original feature space, KRR maps the features into a

high-dimensional space through a kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$ and solves the ridge regression problem in this transformed space.

The transformation to a higher-dimensional feature space is achieved through a mapping function $\phi(\mathbf{x})$, which projects the original features into a space where complex, non-linear relationships can be represented as linear combinations. However, explicitly computing $\phi(\mathbf{x})$ is often computationally expensive or infeasible, especially when the dimensionality of the transformed space is very high or infinite. For example, the Radial Basis Function (RBF) kernel implicitly maps the data into an infinite-dimensional space. Instead of explicitly calculating $\phi(\mathbf{x})$, the kernel trick computes the dot product between two transformed feature vectors directly in the high-dimensional space:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j). \tag{B.7}$$

This approach allows KRR to operate efficiently without needing to construct or manipulate the transformed feature space explicitly. Common kernel functions include:

- **Linear Kernel:** $k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$, equivalent to standard ridge regression.
- **Polynomial Kernel:** $k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + c)^d$, capturing polynomial relationships of degree d.
- **Radial Basis Function (RBF) Kernel:** $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$, capturing complex relationships based on distance.

KRR solves the ridge regression problem in the kernel space:

$$\mathcal{L}(\mathbf{w}) = \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \alpha \|\mathbf{w}\|_2^2, \tag{B.8}$$

where $\alpha \geq 0$ is the regularization parameter. Using the kernel trick, the solution is expressed in its dual form:

$$\alpha = (\mathbf{K} + \alpha \mathbf{I})^{-1} \mathbf{y}, \tag{B.9}$$

where:

- $\mathbf{K}$ is the kernel matrix, with $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$.
- α represents the dual coefficients.

The prediction for a new input $\mathbf{x}$ is:

$$\hat{y} = \sum_{i=1}^{n} \alpha_i k(\mathbf{x}_i, \mathbf{x}). \tag{B.10}$$

KRR generalizes ridge regression by using kernel functions to introduce non-linearity. When the linear kernel is used, KRR reduces to standard ridge regression. However, with non-linear kernels such as the RBF kernel, KRR can model highly complex relationships, making it suitable for data with intricate patterns that a linear model cannot capture. Key hyperparameters include:

- **Regularization Parameter (α):** Controls the trade-off between model complexity and fit to the data.
- **Kernel Type:** Determines the nature of the non-linear transformations.
- **Kernel Parameters (e.g., γ for RBF kernel):** Controls the flexibility of the kernel function.

By leveraging the kernel trick, KRR achieves the ability to model non-linear patterns while maintaining computational efficiency, making it a powerful extension of ridge regression.

This is the final page