Enhancing early-stage energy consumption predictions using dynamic operational voyage data
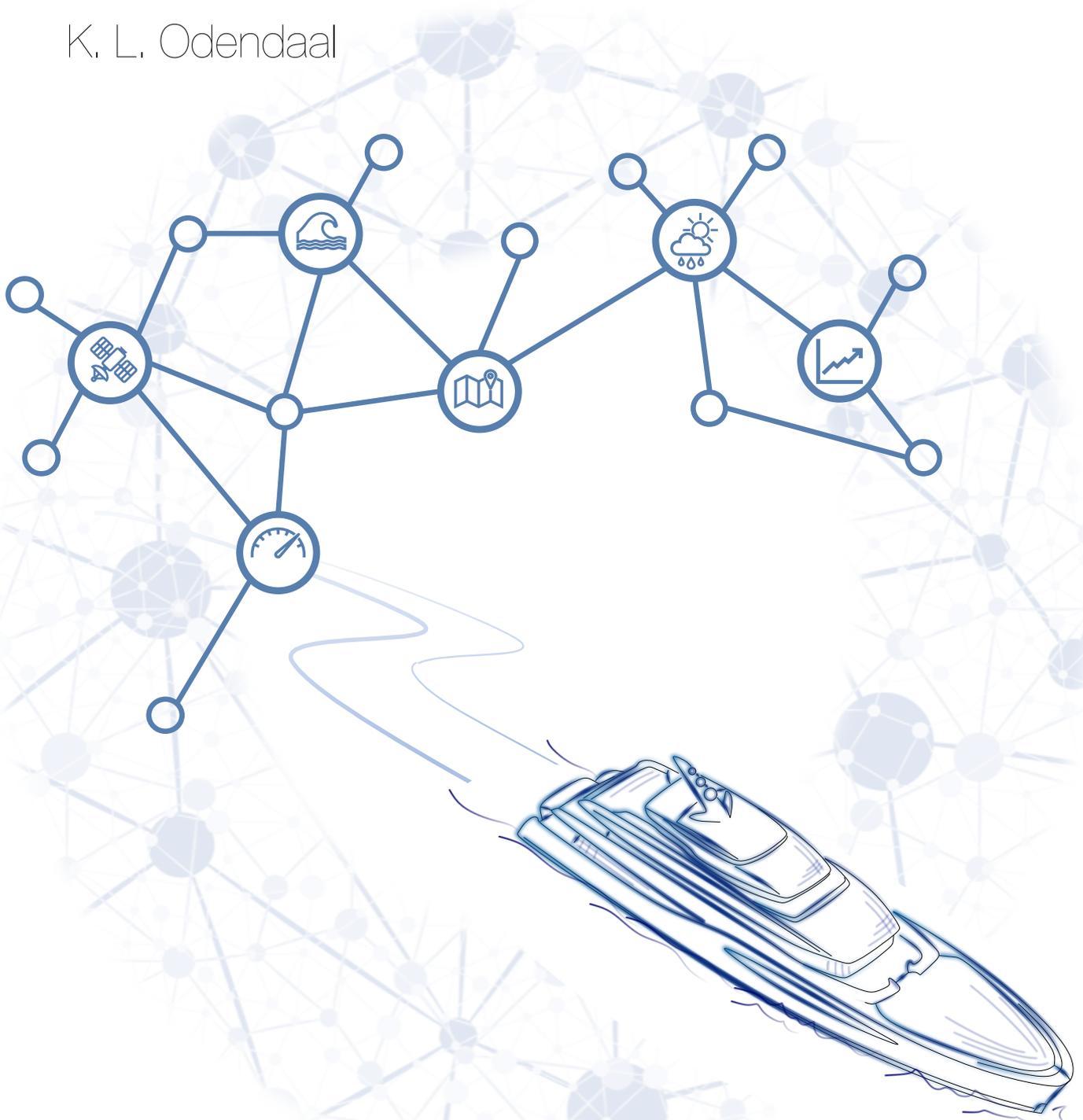
A grey-box modelling investigation

K. L. Odendaal

TUDelft   De Voogt
NAVAL ARCHITECTS

Enhancing early-stage energy consumption predictions using dynamic operational voyage data

A grey-box modelling investigation

by

K. L. Odendaal

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday June 22, 2021 at 14:00.

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

TUDelft   **DE VOOGT** NAVAL ARCHITECTS

Summary

The adverse human contribution to global climate change has been recognized as a significant risk to future generations. Therefore, the yachting industry has acknowledged the need to reduce its environmental impact due to consumer's increasing pressure and potential future regulations to limit the environmental effects. Unfortunately, current real-world data presents a large disparity between predicted and actual gathered energy consumption results. Therefore, this research aims to develop an approach to accurately predict total dynamic Energy Consumption (EC) using real operation voyage data for the improved early-stage design of new future yachts.

Therefore, three modelling approaches within the maritime industry are investigated: White-box, Black-box, and state-of-the-art grey-box modelling. White-box models are considered 100% deterministic, where the physics are easily interpretable. In contrast, Black-box models are based on observed data and require no prior physical system information to function. Grey box modelling attempts to combine the advantages of both techniques while minimizing the disadvantages.

Upon a thorough investigation of relevant literature sources and identifying a clear literature gap within the yachting industry, a precise assessment of the method requirements was conducted to determine the most suitable modelling solution to evaluate propulsion and auxiliary power consumption. It was determined that the approach which will most likely satisfy all method requirements is an Artificial Neural Network Grey-Box modelling (ANN-GBM) approach using a serial configuration.

Further technical descriptions of each white box model and black models are then presented. These overviews give technical details, limitations, and assumptions which must be adhered to during application. Furthermore, the artificial neural network's general working principles and technological foundations for optimal performance are detailed. Ultimately, a secondary literature review is conducted to provide a baseline solution for propulsion and auxiliary solutions, respectively. Here, it is concluded that the available data input features and data quantity closely align with successful literature studies; thus, providing initial confidence in the method potential.

Using a novel preparation and uncertainty evaluation methodology, a 10-month period dataset is applied and orientated to three operations: Sailing, Anchor and Combined situations. A series of studies are then conducted to determine the GBMs interpolation, extrapolation, and exploitation performance potential by comparing each modelling category for each operational dataset. Within the data training ranges, the GBM consistently was a top-performer, managing to make propulsion and auxiliary power predictions within 3% and 9% of actual operational conditions. When making estimations beyond the training ranges, the GBM shows the capability of improving extrapolation capacity. However, improvement limitations were found directly to be related to the strength between dynamic WBM input-output correlations. Finally, in a study utilizing both the GBMs interpolation and extrapolation capabilities, internal relationships are isolated and extracted to estimated the fouling and daylight cycle effects on powering demand.

It is ultimately found via a verification and validation (V&V) analysis that the GBM model solution performs better on average than similar literature models for auxiliary and propulsion power estimations, respectively. The improvements are likely related to the data quality (continuously monitored), input-output feature relations, and data preparation/evaluation steps. Unfortunately, a V&V analysis of the extrapolation capacity could not be conducted as no outright literature comparisons are available for usage. Nevertheless, GBM improvements over the pure-BBM are qualitatively observed and similarly reinforced in alternative literature studies, thus indicating the immense potential of GBM solutions over the conventional BBM approaches.

Finally, while the study provided technical insight into *How* the GBM can be applied, a general understanding of *When* such solution approaches can be applied within design processes is not considered in most literature. Thus, a general consideration decision structure is detailed to provide naval architects with practical knowledge and confidence in future applications of the GBM modelling approach.

Preface

I have always had a strong passion for boats and yachts. Unfortunately, being from mainland Canada, not many opportunities were present to learn about such topics. As such, my passion and love for ships have led me to explore and learn from some of the best within the industry overseas. After two full years of rigorous education, I am excited to admit there is still much to learn! Therefore, I am looking forward to the continuing journey, including the challenges, knowledge, and growth that come along with it. However, the road is long, and therefore each accomplishment should be enjoyed and reflected. While I am incredibly proud of how far I have come, I would not be here today without the help and support of many people. While I wish I could thank everyone involved, completing the project would not have been possible without the following individuals.

I am incredibly thankful for the chance given by *De Voogt Naval Architects* to work on such a relevant and practical problem. **Giedo Loeff**, thank you for the opportunity and for presenting me with the initial problem faced within *DVNA*. Your knowledge and insight into all things ships were a tremendous asset throughout the entire process! Mainly thank you for seeing the potential in my proposed solution! **Aaron Alkemade**, thank you for the constant and continuous support throughout the entirety of the graduate study. You were always readily available for questions, and when you did not know them, you went out of your way to find them for me! My smooth transitions in the early stages are mainly attributed to you quickly finding and addressing my, often challenging, questions. Specifically, thank you for being a cheerful and excited partner throughout this journey! **De Voogt Graduate Interns** (you know who you are), thank you for being friendly, accepting, and a source of distraction when needed. Our weekly coffee meetings have been a highlight and gentle reminder that working together is always more fun than apart!

Next, I am incredibly fortunate and thankful for the opportunities completing my studies at the *TU Delft* have presented me! The quality of education, world-class instructors, and overall support provided throughout the process have directly lead to my academic success. **Austin Kana**, thank you for your knowledge, insight, and enthusiasm, which have been a consistent guiding force throughout my independent research project and final graduate study. This has been a long road; however, I genuinely have enjoyed the process and feel my work has found another level under your supervision.

Ultimately, I would not be here today without the love and support of my parents – **Morné & Soretha Odendaal**. Your endless love and support have instrumental in following my dreams. Thank you for your teachings in kindness, compassion, and dedication which I have found ever-important throughout this time of global uncertainty and fear. Finally, **Breanna Fisher**, thank you for taking a leap of faith and joining me in an extraordinary journey across the world. You inspire me every day, and your love and optimism are an infectious reminder that no matter how bad or difficult things seem, there is always a light at the end of the tunnel! Words will never be able to capture my gratitude and thankfulness for having you in my life.

Kirsten Lombard Odendaal
Delft, June 2021

Contents

List of Figures

List of Tables

All models are wrong, but some are useful.

-George E. P. Box

1

Yachting and the Environment

The adverse human contribution to global climate change has been recognized as a significant risk to future generations. As such, the shifting perspective towards sustainability is currently being driven by both public image and social responsibility. Therefore, to meet these worldwide demands, the yachting industry has acknowledged the need to reduce its environmental impact due to the increasing pressure of consumers and future regulations to limit the effects on the environment.

Currently, the yacht fleet is rising rapidly worldwide. As investigated by Lindstad et al. [70], the yachting industry expects the total number to rise to 7,200 by 2030, as highlighted in figure 1.1. Without any immediate interventions, the yachting industry's environmental impacts are expected to grow proportionally with the increasing fleet size. As a result, discussions regarding the ecological consequences of global maritime fleet growth have become necessary within the current climate debate.
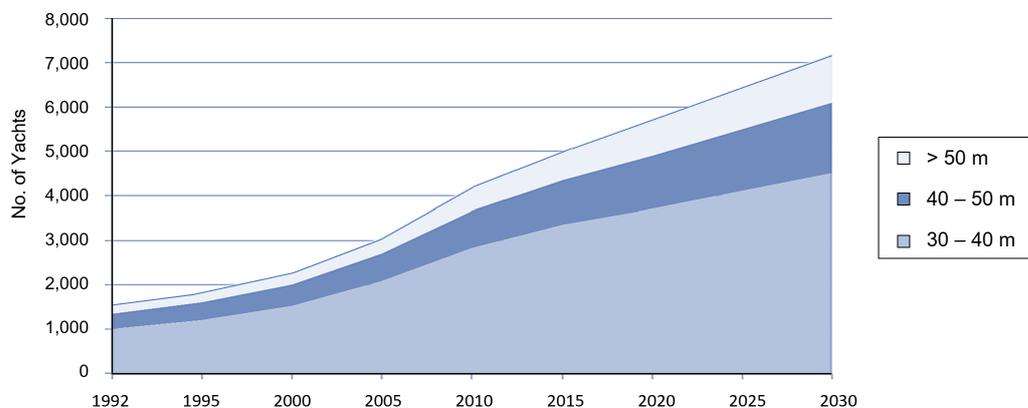
Figure 1.1: Yacht fleet growth predictions from 1992 to 2030, (Lindstad et al. [70])

These implications are well documented by the Intergovernmental Panel on Climate Change (IPCC). Currently, the IPCC is the main contributor to the United Nations Framework Convention on Climate Change (UNFCCC) for objective and scientific information relevant to understanding the scientific risk of human-induced climate change (IPCC 2013 [57]). From the report IPCC 2007 [56], it is estimated that greenhouse gas emissions need to be reduced by around 50% to 85% in 2050, compared with current levels, to achieve a stabilization temperature at 2 °C above pre-industrial levels. To that end, The Paris Agreement was established to strengthen the global response to the threat of climate change by keeping a global temperature rise well below 2 °C and pursue efforts to limit the temperature increase even further to 1.5 °C (United Nations [102]). However, as outlined in CE Delft [21], The Paris Agreement does not set specific targets for countries or sectors but instead relies on 'nationally determined contributions' (NDCs). International maritime emissions are currently not covered by NDCs. Still, given their share in the total emissions, it presents a significant risk towards reaching The Paris Agreement's

goals if unchanged.

To avoid such a disaster, the Marine Environment Protection Committee (MEPC) of the International Maritime Organization (IMO) adopted the Initial IMO Strategy to reduce greenhouse gas (GHG) emissions from ships. However, due to the immense global contribution of GHG emissions within the maritime transport industry, the public eye has mainly fallen on shipping instead of yachting. To date, three IMO Greenhouse Gas Studies have been published (IMO [55]). The first IMO GHG study, published in 2000, estimated that international shipping in 1996 contributed approximately 1.8% of the global anthropogenic CO_2 emissions. In comparison, the second IMO GHG study estimated that international shipping emissions in 2007 contributed approximately 2.7% of the global anthropogenic CO_2 emissions. Finally, the third IMO GHG study, published in 2014, estimated international shipping emissions in 2012 contributed approximately 2.2% of the global anthropogenic CO_2 emissions. The study also updated the CO_2 estimates for 2007 to 2.8%. Based on these findings, the IMO remains committed to reducing GHG emissions from international shipping. As a matter of urgency, the IMO aims to phase emissions out as soon as possible in this century. To achieve these ambitious goals, three strategic levels, as seen in figure 1.2, have been set by the IMO [55].

1. The reduction of vessel carbon intensity through the implementation of the energy efficiency design index (EEDI) for new ships with the aim to strengthen energy efficiency design requirements.

2. International shipping to reduce CO_2 emissions per transport work, by at least 40% by 2030, pursuing efforts towards 70% by 2050, as compared to 2008; and

3. To reduce the total annual GHG emissions by at least 50% by 2050 compared to 2008 while pursuing efforts towards complete emission phase-out to ensure a pathway consistent with the Paris Agreement temperature goals.
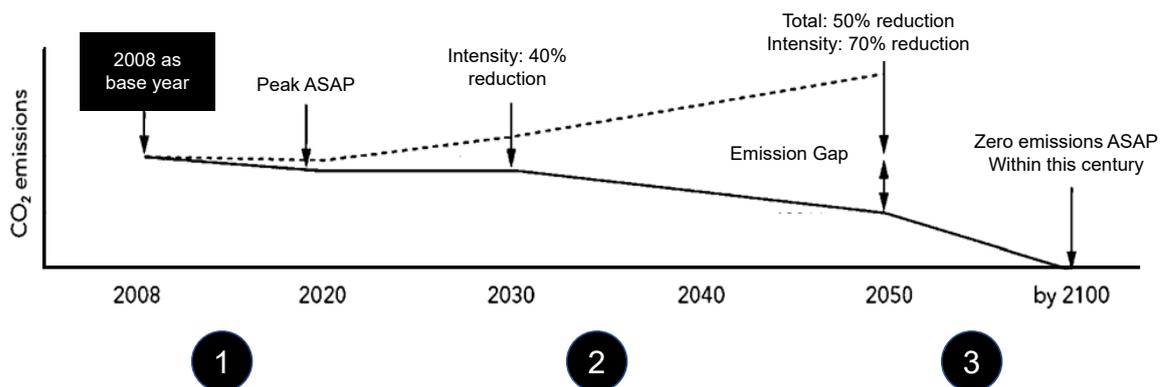
Figure 1.2: Initial IMO strategy on reduction of GHG emissions: visions and ambitions, (IMO [55])

At present, the expression of the EEDI is in CO_2 emissions per ton shipped over a nautical mile. Unfortunately, this metric does not translate well to pleasure crafts such as yachts. Although yachts are vessels, they are not associated with the shipping industry as they do not carry goods for the sake of commerce. One primary purpose of a yacht is to provide luxury mobile accommodations to experience the world's natural beauty. As highlighted by Cozijnsen [28], due to the difference in operational profiles, the expression of emissions from yachts in shipped tons over nautical miles becomes complicated and inaccurate. Currently, the EEDI does not include yachts, as the influence of these vessel types typically shows a lesser impact than the global shipping fleet. This is evidently shown and analyzed by Olmer et al. [81], where varying ship types are decomposed into three categories: *International, Domestic*, and *Fishing*. In this grouping scheme, yachting falls within *Domestic*, alongside six other vessel classes. The detailed CO_2 emissions per ship class can be seen highlighted in figure 1.3.

Nonetheless, if the Paris Agreement's goals are to be reached, it is expected that regulations affecting yachts will most likely be established in the future (Letschert [69]). As such, *De Voogt Naval Architects*, a leading luxury yacht design and engineering company, is currently developing the notion of an environmental impact index for yachts. The Yacht Environmental Transparency Index (YETI) has
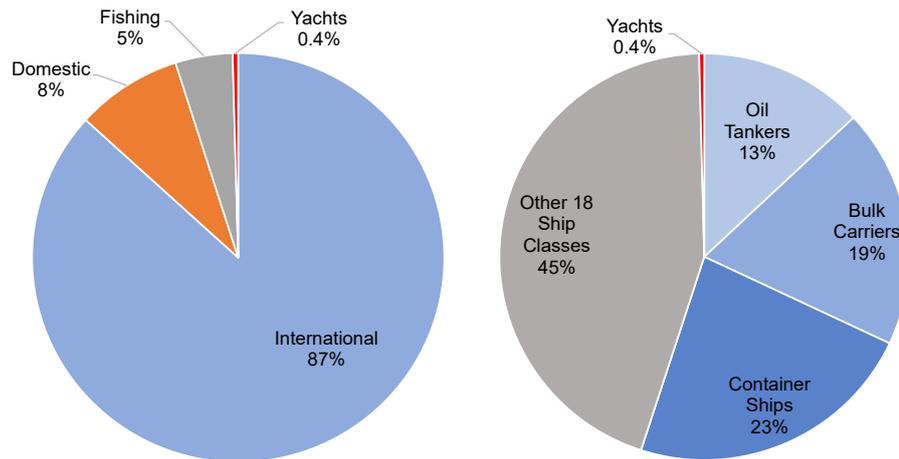
Figure 1.3: Average share of global CO_2 emissions (2.6%, 2015) by category (left) and ship class (right), (Olmer et al. [81])

emerged as a result by evaluating the gap between existing indices and the willingness to make yachts more sustainable. As detailed by Cozijnsen [28], the index's ultimate goal is to push the industry to become more sustainable through a clear and unbiased method, completely capturing the full yacht life-cycle (build, operation, and refit) environmental footprint.

While the YETI project is a big step in the right direction, it is still a very rigorous life–cycle analysis. Therefore, to comply with this strict framework, the understanding and prediction of propulsion and auxiliary demand must be met with a high degree of confidence within the early design stages when developing the yachts of the future.

1.1. Introduction to De Voogt Naval Architects

The yachting industry is regarded as an advanced technology sector of the maritime industry, and it benefits from the social awareness involved in emissions reduction and sustainability. One such company leading the charge in cutting-edge sustainable developments and research is *De Voogt Naval Architects (DVNA)*. *De Voogt* is an internationally renowned and leading yacht design and engineering firm located in Haarlem, Netherlands. The company has currently designed and developed a diverse portfolio of luxury custom and semi-custom motor yachts up to a length of approximately 110 meters. All yachts that *De Voogt* creates are tailor-made for the very exclusive export market and currently is situated as an industry leader in the 70 to 100-meter market (de Vries [30]).
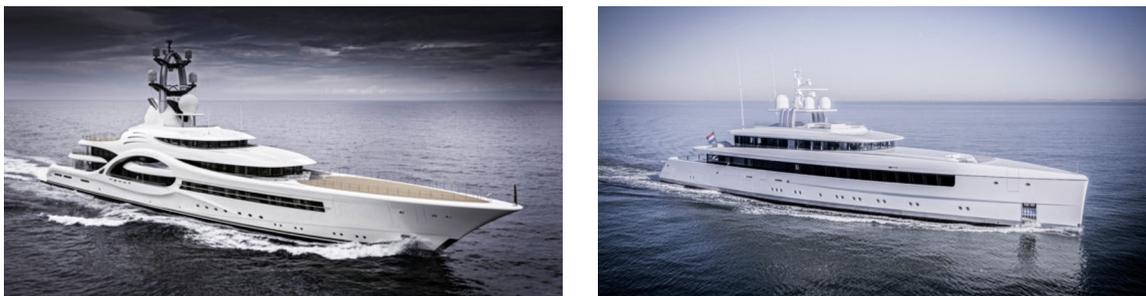
Figure 1.4: De Voogt Naval Architect designed and built Anna (left) and Najiba (right)

De Voogt Naval Architects works - under the brand name *Feadship* - closely with the yards of Koninklijke De Vries Shipbuilding and Royal Van Lent. In addition to designing yachts in the classic *Feadship* tradition, *De Voogt* is increasingly asked to develop revolutionary new concepts aligned with current green market movements. Being an industry leader, and having the market push towards sustainability, a clear focus on innovation and advancement has been adopted. This market trend has to-date presented revolutionary yachts such as the 110m *Anna* and the 58m *Najiba*, as seen in figure

1.4.

As outlined in Jachtbouw Actueel [60], the former *Anna* discarded the traditional engine-crankshaft-prop setup favouring a more sustainable diesel-electric system, where an electric engine drives the propeller instead. Additionally, *Anna* has a comprehensive, integrated feedback system, enabling real-time data collection and vessel management seamlessly. Even more impressive is the latter, *Najiba*, using only 11.4 litres of fuel per nautical mile at a 12 knot cruising speed. This achievement is upwards of 25% less than any equivalent yacht on the market. *De Voogt's* success was achieved through shifting demand from devices – washing machines, lighting, etc.– to nighttime duties. Ultimately, this 'peak shaving' has the potential to cut power use by an estimated 10% to 15%.

In seeking a greener future, ground-breaking tools and initiatives have pushed *De Voogt* towards adopting more green technology for further sustainable applications within the yachting industry. However, challenges and limitations regarding the accuracy of total energy consumption estimations at an early stage are currently being faced.

1.1.1. Goals, Challenges, and Known Uncertainties

Currently, a hurdle is being faced within the maritime sector regarding observed discrepancies between real-operation voyage data and predicted energy consumption results. These gaps have led to much discussion on the implications of future sustainable yacht design. As such, *De Voogt's* current goals, challenges, and uncertainties regarding energy demand estimations are extracted and paraphrased from one-on-one internal interviews. These discussions and associated questioning can be seen in the appendix A.1. It should be noted that, while the following study focuses specifically on the yachting outlook within *De Voogt Naval Architects*, such challenges and discrepancies are also occurring within other maritime industries such as the shipping and public transport sector. Therefore, the relevance is widely applicable.

Currently, yachts are designed based on sea-trial conditions and ideal assumptions. Ultimately, many of the calculations and optimizations are evaluated under standard academic conditions and are not realistic in practice. For instance, all bare-hull and appendage optimizations are currently done for sailing purely ahead. Additionally, all propulsion analysis is based on propeller loading cases in the clean hull, clean propeller, and calm water environmental situations. These considerations are an incorrect reflection of real-world phenomena and pose significant consequences in consumption and emission evaluation discrepancies. As such, to improve on these current practices, a series of goals, as outlined by engineers of *DVNA*, can be identified.

1. First, improved operational estimations and feedback are required by collecting, filtering, and analyzing all relevant vessel data. The lack of data evaluation and integration disconnects accurate system consumption and proportioning between propulsion and auxiliary power estimations and actual observed operational demand.

2. Secondly, much related to the first point, *De Voogt* needs accurate predictions of the yacht's actual impact: fuel consumption and emissions. While the parameters are related, they are not the same, especially when considering polluting emissions and associated regulation impacts.

A clear future vision of *De Voogt* and *Feadship* is the introduction of new hybrid systems onto the market. This challenge requires a new, more accurate prediction of overall energy usage and impacts under dynamic conditions. Therefore, a primary goal and challenge of *De Voogt* are to take strides towards an applied dynamic design approach, in which real operational data and the use of the vessel is considered. Fundamentally, adapting and evolving towards a functional design approach as opposed to the current estimation practice. While the ambitions are relatively straightforward, known uncertainties within both the propulsion and auxiliary power components challenge meeting these objectives.

Auxiliary Power: As observed by measured operational fleet data, the auxiliary power fluctuates a great deal in different operational situations. The current *De Voogt* engineering tool for estimating peak auxiliary power demand is focused on equipment load lists. These lists are extensive as all

consumers, peak load demands, and load factors are presented. Ultimately, these load factors, which indicate operational usage, are chosen based on prior experience. Systems known to be continuously operating or never operating when held in redundancy can be easily evaluated. However, systems that work intermittently are more challenging to assess. For instance, a ship stabilizer consumes and functions differently when sailing than at anchor. Ultimately, selecting how and when these factors are applied is mainly dependent on prior engineering experience, which lacks a sound scientific basis for optimal use.
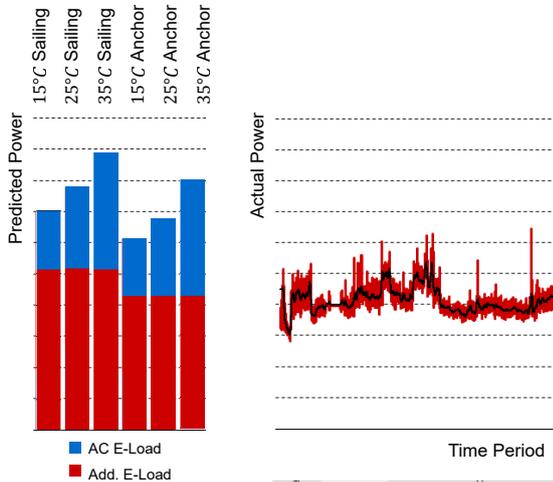
Figure 1.5: Measured differences between predicted and actual operational auxiliary load
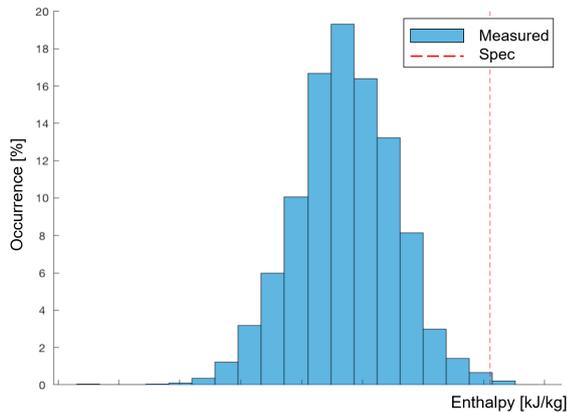
Figure 1.6: Hindcasted weather data probability of occurrence enthalpy difference comparison

Inspecting one such *Feadship* case, figure 1.5, there is a clear difference as the total e-load predictions are almost 2x more than what is happening in operation. Thus, the current estimation method is not an accurate representation of real-world functions.

There are various reasons identified for the discrepancies. However, one known situation is the determination of the design points. When an HVAC system is designed, it is usually for the worst-case environmental loads: the hottest potential point on the day. When considering the probability of occurrence, these loads for the specific humidity and temperature conditions are incredibly infrequent. Hence the likelihood of actual occurrence is extremely low. For example, when considering a particular area on board and incorporating hindcast weather conditions, actual enthalpy load variance can be determined. This technique can be illustrated in figure 1.6. Based on the collected data, the observed load spectra are much lower than the predicted design point. Therefore, the design point worst-case estimation methodology contains a large margin of uncertainty and conservatism.

Ultimately, it boils down to a difference between estimated load profiles as opposed to data-driven load profiles. The former is what naval architects have always been doing, designing for sea-trial, for things previously thought relevant. However, with access to real-operational data, new methods and techniques can be utilized for functional design to further enhance future efficiency and effectiveness.

Propulsion - Shaft Power and Fouling Effects: Additionally, uncertainties within the propulsion predictions are also observed. In the early design stages, it has been noticed by *De Voogt* that the power prediction estimations can be lower than the operation data results by upwards of approximately 50%. Once again, this has to do with the ideal calm-water assumptions; however, things are never perfect in reality. Dynamic phenomena such as waves and wind are always present; thus, deviations between predictions and operational results are consistently produced.

Additional, dynamic effects such as bio-fouling also provide a large source of uncertainty. Ultimately, yachts are stationary for a large portion of their operation. As such, high calcium fouling on the entire hull and propeller is a common occurrence. However, the fouling amount depends on the external conditions and the state of the anti-fouling system in place, [103]. Nonetheless, this biological growth has the potential to increase the expected vessel resistance significantly. Recently, due to the unforeseen

resistance increase, catastrophic exhaust system failure occurred. When analyzing the data, it was observed that the power demand on the main engines was well beyond the safe consumption ranges. The collected results can be seen highlighted in figure (1.7).
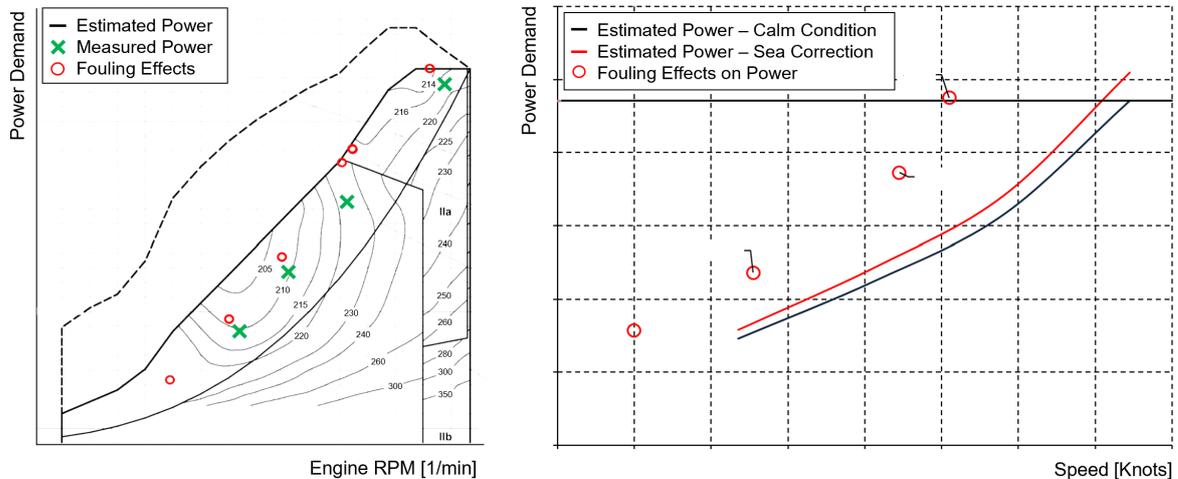
Figure 1.7: Measured differences between predicted and actual operational shaft load

The green points show initial sea-trials under ideal conditions: clean hull, propeller, ideal weather, and appropriate corrections. In comparison, red dots are the extracted shaft power in the failure scenario where dynamic conditions such as fouling were present. On the right-hand side, we see a speed power graph. In this case, the black line is the initial sea trial, and the red line is corrected for the sea-trail conditions. However, the actual shaft powers can be seen to exceed these estimations significantly, producing roughly a 1.5x load increase.

This phenomenon has been observed in many other situations as well, thus garnering well-deserved attention. Currently, *De Voogt* has determined, re-engineered and re-calculated the expected power increase due to fouling using numerical flow simulations. This internal study significantly increased the understanding and accuracy of power predictions during such occurrences. Unfortunately, such simulations are not practical in the early design stages as the methodology is too computationally demanding. As such, *De Voogt* is faced with the challenge of accurately predicting the expected propulsion power under dynamic operational conditions and estimating the influence and state of fouling within the earlier design stages.

1.2. Research Goal and Objectives

As outlined and elaborated within section 1.1.1, current prediction methods underestimate the propulsion power and overestimate the auxiliary power within dynamic operating conditions, respectively. Therefore, the purpose of the investigation is to,

Develop an approach to accurately predict total dynamic Energy Consumption (EC) using real operation voyage data for the improved early-stage design of new future yachts

To reach this research goal, a few key research questions will be addressed and answered throughout the course of the study. These include,

1. *How do the DVNA design process, calculation methods, and data availability influence the overall modelling requirements when estimating the EC?*

2. *What methods currently exist to predict EC for both propulsive and auxiliary loading accurately, and which approaches are most suited to achieve the modelling requirements?*

3. *How do the proposed technical solutions operate, and which modelling conditions, limitations, and assumptions are necessary for optimal performance?*

4. *What current data preparation methodologies exist to incorporate raw operational information within the modelling approach?*

5. *How is the general performance influenced by varying modelling categories, data-preparation procedures, and vessel-specific operational usage?*

6. *What estimation capabilities does the proposed solution exhibit outside the design domain, and how can the total performance be leveraged to isolate and extract hidden relationships such as,*

   (a) *Fouling effects overtime on total propulsion power?*
   (b) *Daylight cycle effects on total auxiliary power?*

7. *How can the proposed solution approach be integrated within conventional ship design processes, and what criteria must be considered for successful implementation?*

1.3. Report Outline

The complete report structure can be seen outlined in figure 1.8.
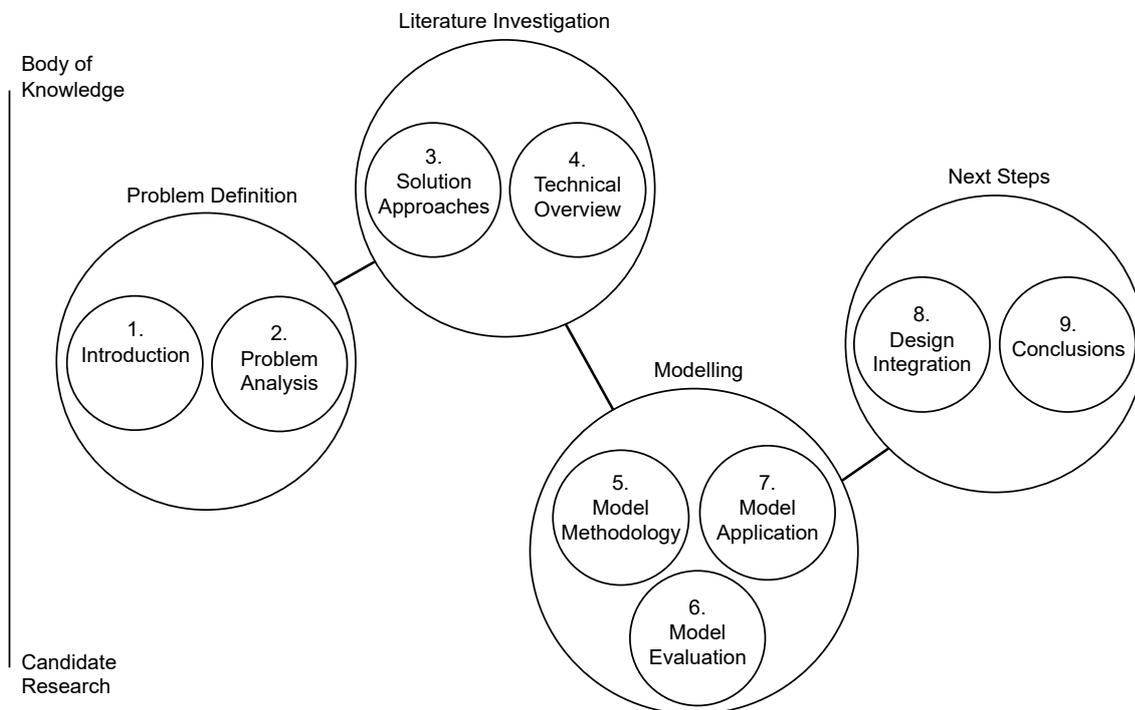
Figure 1.8: General report structure and outline

Chapter 1 introduces the general problem of prediction discrepancies amongst actual vessel energy consumption. This problem identification leads to the proposed research goals and objectives necessary to overcome such challenges. Chapter 2 further defines the problem by investigating the associated design phase focus, internal calculation methods, data availability and uncertainties, and modelling requirements essential to fulfilling the research objectives.

Since the investigation primarily focuses on model development and analysis, new modelling terminologies such as White-box, Black-back, and Grey-box modelling are be introduced within Chapter 3. While these terms are standard within computer science, the maritime industry has yet to adopt these descriptors. Therefore, a detailed introduction of these concepts and their relation to the marine industry is further addressed. Once all possible modelling categories and literature are reviewed, a proposed solution, potential risks, and work scope is addressed. Chapter 4 provides a technical overview of the proposed solutions while addressing each model's critical assumptions, considerations, and limitations. The literature review can be considered complete at this stage.

Next, the modelling stage presents the bulk of the candidate research. Chapter 5 outlines a novel methodology to implement the proposed solution approach. Here, each evaluation stage is detailed and outlined to provide a clear and transparent procedure for universal application. Once all critical steps are detailed, Chapter 6 applies the presented approach to estimate the main powering components of a single vessel. The developed model's critical assumptions, limitations, differences, and performance are considered and addressed throughout each preparation and evaluation step. This chapter ultimately demonstrates both the methodology and solution approach accuracy capabilities within a data design range. Chapter 7 continues to explore the proposed solution approach. Critical performance details and reflections on the approach's capacity to extrapolate beyond its data design domain are addressed, and inherent exploitation input-output relationships are demonstrated. These studies enforce the broad solution approach applications while acknowledging potential pitfalls and limitations.

Upon conclusion of the case modelling evaluation and application studies, general design considerations, findings, and recommendations are addressed. Chapter 8 presents an overview of how the proposed solution can be integrated within conventional maritime design processes. A clear road map is provided to ultimately provide Naval Architects with knowledge and confidence when such methods are suitable from a practical perspective. Finally, Chapter 9 presents the main research conclusions. First, the research questions are addressed individually, leading to the final research goal fulfillment and modelling requirement assessment. Upon closing, practical industry and academic contributions are put forth, and recommendations for future works are presented.

2

Problem Analysis

This chapter aims to expand on critical topics that are beneficial in further understanding and defining the proposed problem by answering the first research question,

'How do the DVNA design process, calculation methods, and data availability influence the overall modelling requirements when estimating the EC?'

First, an introduction and highlight of the current *De Voogt Naval Architects* design and engineering procedure are presented. This section further elaborates on the associated level of detail required for power prediction metrics during each associated phase. Upon summarizing each design stage's relative accuracy and objectives, a targeted design phase focus is determined. Secondly, the current calculation procedures for thrust loading and auxiliary loading is investigated and outlined. This study allows for a check on what tools are currently implemented (or not) for each calculation method. Next, an analysis of data availability is addressed. This study includes an investigation into the various data sources, including new initiatives such as the *7SEAS* and *Project HOTEL*. This section also outlines current data limitations and associated uncertainties. Finally, when the problem and parameters are fully explored and analyzed, the chapter concludes with a set of design requirements necessary to address accurate energy consumption predictions.

2.1. The Design and Engineering Process

Generally, in a new build program, engineering activities must be executed to support the main project activities of customer acceptance, class approval, procurement, construction, testing and final delivery of the contracted product. The engineering packages must provide the required technical information, which increases in extent and detailing during the project's progress, from the concept design to completion and delivery.

The following sections elaborate on the design and engineering processes of *De Voogt*. All acquired information about the internal *DVNA* design processes and general project requirements are obtained via one-on-one interviews and internal company documentation (Feadship [38, 39, 40]). The interviews and associated questioning can be seen in the appendix A.1. As outlined in section 1.1, *De Voogt* is a part of an integrated group known as *Feadship*. Therefore, the design processes between *De Voogt* and the various yards can deviate slightly. Thus, the investigation focuses on the current *DVNA* design and engineering process perspective, as this particular structure is most in line with the associated study.

The *DVNA* design process can be subdivided into five critical stages: Concept Design, Contract Design, Development Design, Definitive Design, and Concurrent Design. Figure 2.1 highlights these main stages and how these initial phases fit into the general yard end-stage production processes.

Initially, the design process begins from a blank sheet of paper, every piece of metal, panelling, ducting, wiring, and outfitting is designed just for that particular yacht. As outlined by van Wijngaarden
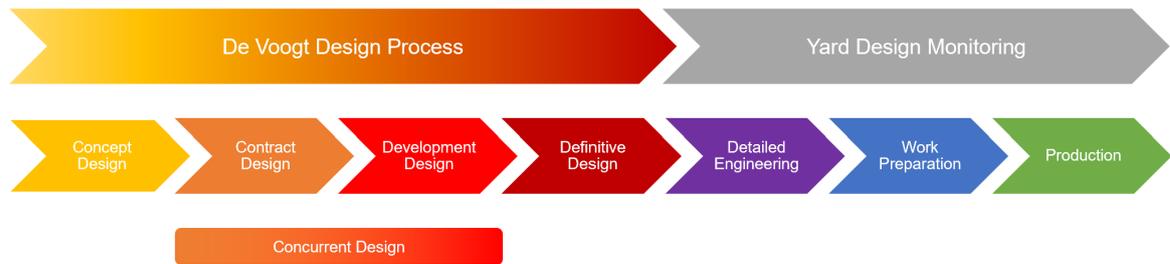
Figure 2.1: Operational Excellence (OE) process as used by De Voogt Naval Architects

et al. [104], from the first sketch to the smallest detail, a traditional design spiral approach, figure 2.2, is followed during the vessel's formation.
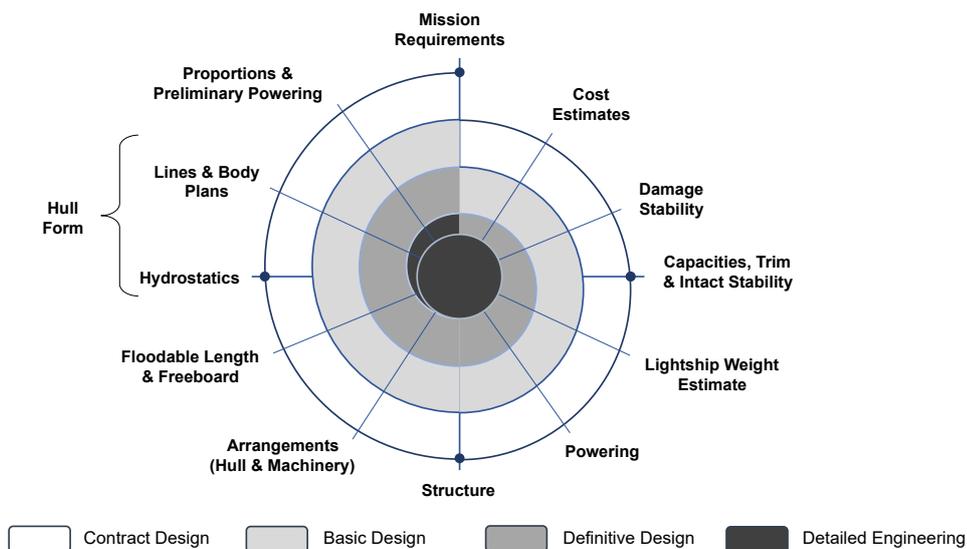
Figure 2.2: Traditional Design Spiral approach used by De Voogt Naval Architects, (based on Evans [37])

Design requirements are given by dedicated stakeholders and/or *Feadship* specifications, which are then interpreted by designers and engineers, resulting in an overall design satisfying the brief. These proposals are then thoroughly reviewed by all relevant stakeholders. Within all review-iterations, the spiral convergence approach is implemented and followed. This optimization strategy of design is known as 'Point-Based Design' and involves slow refinement and iteration until convergence upon an optimal solution. As experienced by van Wijngaarden et al. [104], this method requires many calculations to be redone with every iteration. Unfortunately, in most cases, time and budgets are limited, which ultimately limits the number of iterations cycles a design product can complete. However, to overcome this traditional design approach's deficiencies, a concurrent design approach is incorporated to reduce such iteration and cost bottlenecks.

It should be noted that the concurrent design sessions are situated to invoke and involve interdisciplinary communication, which leads to a higher quality of decisions, faster development of the design, and better understanding throughout the design process. As such, direct theoretical ship calculations are not made during this stage. Therefore, when considering energy consumption components such as propulsion and auxiliary power demand, direct calculations are hardly required. Thus, the concurrent design stage will not be considered a relevant addition to the design phase focus investigation.

2.1.1. Concept Design

At the end of the initial Concept Design Stage, theoretical shipbuilding is incorporated through an internal engineering review process conducted by the *De Voogt* department. This initial assessment

process is considered a first iteration of the engineering department to identify bottlenecks or confirm the designer and/or manufacturing that the associated risks are manageable. This phase investigates the preliminary vessel calculations, such as weight, stability, general arrangements (layouts), powering, and unique features. Since detailed information is highly lacking at this stage, it is usually accomplished through comparative studies and basic empirical and statistical formulations.

The prediction of propulsion load in the Concept Design phase is generally of low accuracy due to the high design uncertainty. During this stage, 2D empirical estimations and statistical vessel comparisons determine the expected resistance and thrust demand. Additionally, auxiliary load demand is roughly estimated within the Concept Design stage. Much like the propulsion load estimations, vessel comparisons are used to determine an expected initial load balance, also known as load lists. Firstly, these lists contain all large and necessary equipment such as; engines, generators, HVAC units, and stabilizers. Ultimately, these calculations are primarily used as an initial baseline starting point and are highly subject to change.

2.1.2. Contract Design

Upon customer interest in a Conceptual Design, the Contract Design begins. This design stage ends with a signed and contracted new build project. 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.

Once the Contract Design stage begins, CFD is generally applied to determine hull resistance. However, these results are based on preliminary hull lines plans. As such, they are highly subject to change and iteration. As such, while the method is of high fidelity, the subsequent inputs are not. Thus, the outcome is still far from complete. Nonetheless, bare-hull resistance and appendage optimizations are performed for a more detailed estimation. A preliminary speed-power curve is developed and further investigated in terms of speed, range, and power requirements using these initial results. At this point in the design process, the auxiliary power considerations are no longer done by *De Voogt.* These estimations are outsourced to contractors or associated shipyards. From a technical perspective, initial load balances and equipment lists are developed to give a rough total energy prediction. These load balances and associated equipment activity rates are based on past industry experiences and reference vessels. As such, no transparent or standardized methods are applied when selecting such load factors, resulting in large variance and uncertainty within the estimations.

2.1.3. Development Design

The Development Design phase usually begins immediately after the signing of the sales contract. During this design stage, a dedicated team is formed within *DVNA* and the associated shipyard. The teams generally consist of Technical Advisors (TA), a Project Manager (PM), as well as Project Engineers (PE) of all disciplines. Ultimately, this phase aims to optimize the design by indicating possible risks and providing interdisciplinary solutions for the detailed engineering phase. This check strategy enhances the engineering process's efficiency by limiting the possibility of any significant changes in the design later. As such, best practices are universally implemented to achieve this objective,

- Focus on risk identification, not on solving problems
- Clearly outline the scope of the Design Development
- Take up lessons learned in the Design Development checklist for forthcoming projects

Hydrodynamics plays an essential role in the overall design. As soon as the initial Contract Design deliveries are achieved, the Design Development CFD iteration process continues. If the preliminary optimized hull is compliant with the specifications, a bare-hull trim wedge analysis is performed. Otherwise, the hull shape is rejected, then the initial Contract Design process is wholly repeated until

compliance is achieved. In addition to the general hull and appendage optimization, an initial nominal wake-field interaction procedure is completed during this stage. This procedure allows for estimating an initial wake-field and optimization of the shaft line positioning. The deliverable is once again a speed-power curve. However, this stage demands compliance in terms of specifications for both speed and range. While the detail level can be accurately determined using CFD platforms, the certainty level in the calculation results is not assured due to simplifications made within the design stage. These include propeller momentum sources and assumed thrust deductions.

During the Design Development stage, all auxiliary power estimations and load lists are not handled in-house. These estimations are conducted directly by contractors and associated yards. As such, the detailing behind the process is not well understood. Nonetheless, HVAC system power estimations are generally performed using first-principles heat load balances on supplied general arrangements. These considerations are evaluated for varying operational temperature points to define multiple load cases clearly. Completed HVAC evaluations and estimations are then integrated within the electrical load balance sheets. These estimations are continually updated as more detailed information about supplier equipment, layout spaces, and specialized machinery specifications become available.

2.1.4. Definitive Design

Once the design has been de-risked and the dedicated build teams are formed, the Definitive Design stage begins. During this stage, the construction's main characteristics, the systems, and the interior are matched per rules and regulations and yard standard requirements. Additionally, approval of classification societies and flag states is obtained within this phase. Finally, the primary drawings are packaged and presented to the authorized representative of the customer.

At this later point in the design process, a final lines plan with aligned appendages, and shaft lines is used to determine the final wake-field analysis. Once the final wake parameters are estimated, including hull efficiency and thrust deduction, the final appended hull is evaluated for varying drafts, which results in the final speed power relation. It should be noted that these CFD calculations are applied under a calm water assumption and do not consider environmental effects such as waves or fouling. During this design stage, the associated level of detail is refined to the highest level of accuracy. These final resistance results are used and compared with sea-trail protocol results to evaluate the actual vessel range and emissions.

During the Definitive Design, auxiliary power load balances are fully defined and detailed. Only minor equipment changes may influence the mean power demands. As such, the auxiliary load balance is considered a high fidelity estimation since all significant and most (if not all) minor machineries, such as HVAC systems, lighting, stabilizers, and engines, are incorporated. Additionally, these auxiliary power estimations are available for multiple operational conditions. These considerations and procedures in evaluating and the breakdowns are highly influential on associated equipment activity rate selections. Unfortunately, assumptions behind selections are hardly transparent as these procedures generally lack a standardized method other than a company or user experience.

2.2. Design Phase Focus

Based on the *De Voogt* design process evaluation (section 2.1), a detailed summary of each associated phase and connection with energy demand predictions can be seen in table 2.1.

From the evaluation, the *Contract Design* phase is identified as the phase where an improved operational prediction method for both auxiliary power and propulsion loading can add the most value to the engineering process of *De Voogt Naval Architects*. There are multiple reasons and benefits for focusing on this particular stage within the design process.

First, the ability to speed up the acquisition of problem knowledge early in the engineering process allows for mitigation of risk instead of costly redesigns later in the project. An essential aspect of this problem is accurately estimating the required power proportions from both loading components: auxiliary power and propulsive power. Underestimating or overestimating these results can lead to

Table 2.1: De Voogt design process summary

Design Stage	Propulsion Demand	Auxiliary Power Demand	Fidelity
Concept Design	2D empirical estimations and statistical vessel comparisons based on the existing fleet is used to evaluate the 1st approximation of thrust power demand.	All essential machinery loads are compiled into a preliminary load balance. Additionally, statistical vessel comparisons are made to give a 1st approximation of auxiliary power demand.	Low
Contract Design	As soon as a hull lines plan is available, preliminary CFD evaluations are performed to determine the initial bare-hull and appendage resistance. These evaluations consider ideal calm water condition assumptions	Auxiliary power estimations are not under the control of DVNA but instead out-sourced to contractors and associated shipyards. Received deliverables include detailed load balance estimations of all essential machinery, which are broken into multiple operational conditions. For each condition, varying equipment running activity rates are selected through a company or personnel experience. These balances also consider and include HVAC estimations based on first-principles heat load balances for varying environment temperature conditions and design set-points. Ultimately, assumptions and procedures are not explicitly transparent as no standardized process is available from outsourced estimations	Low - Medium
Development Design	A CFD iteration process is closely followed. If the initial hull designs meet the preliminary specifications, a detailed trim and shaft line optimization process is completed. If not, the Contract Design process is redone until the specifications are achieved		Medium - High
Definitive Design	Model tests, in conjunction with CFD, is performed both externally and internally. These evaluations provide detailed resistance estimations. Additionally, a final wake-field analysis is performed to evaluate thrust performance metrics such as hull efficiency and thrust deduction. All evaluations still consider calm water conditions. The final estimations for thrust demand		High

incorrect total operational energy consumption. These uncertainties can influence essential metrics such as emission predictions and energy indices, but they can also impact future yacht designs, such as hybrid vessels, which require a high degree of prediction accuracy during the early-stage feasibility studies. Secondly, this phase in the design generally has the most considerable leap in available data and total input parameters required to accurately determine critical theoretical shipbuilding calculations. Therefore, developing a simple, robust, and accurate model to save time while reducing the required staff-hours of the experienced naval architects are seen as incredibly important. This can ultimately speed up the Contract Design process and allow for more time allocation to other critical areas of the design processes. Thirdly, as outlined in the table 2.1, during the existing Contract Design stage, no in-house capabilities exist to accurately predict the estimated auxiliary power demand under varying operational conditions. As such, a tool that can accurately estimate such metrics would be seen as a huge benefit. This enhanced capability will not only allow for enhanced designer knowledge and

understanding, it will also reduce informational bottlenecks as usually experienced when dealing with external sources or parties. This closing of the informational loop means internal processes can move forward with fewer time delays.

2.3. Current Methods for Energy Demand Estimations

To further understand the current problem and establish meaningful modelling requirements, an investigation into the current calculation processes and available design tools must first be completed and understood. Therefore, the critical methods: propulsion loading predictions, and auxiliary power prediction are highlighted. It should be noted that the current calculation method discussions are based upon internally held interviews with esteemed *DVNA* designers and engineers. The interviews and associated questioning can be seen in the appendix A.1.

2.3.1. Resistance and Propulsive Power Estimation

Throughout the *DVNA* design process, resistance and propulsion power predictions are typically made in three ways: semi-empirical calculations, direct database comparison, and CFD analysis.

YACHT is an internally based software that uses the Marine Research Institute Netherlands (MARIN) DESP method, commonly known as the Holtrop and Mennen [52] method, correlated on *Feadships*. This method is a relatively good semi-empirical method for *Feadships* because the fleet is primarily composed of displacement yachts to which the approach is centred. Ultimately, this tool provides a quick estimation method with a relatively moderate accuracy level of new vessels which fall within the data coverage range. In addition to the semi-empirical approach, an internal resistance curve database has been developed and is continually maintained within *Studio De Voogt*. The platform collects existing powering and resistance curves from all previous *Feadship* model tests. This database is used to compare and analyze similar vessel trends as seen from collected model and sea-trail results. Ultimately, it is a tool to provide designers and naval architects with an initial baseline idea of how a similar vessel may behave.

Finally, CFD is also used in the estimation of propulsion power. These calculations can be applied very early on in the design stages due to increased available computational power. The method allows for relatively quick bare-hull resistance estimations, propeller-hull interactions, and added appendage resistance optimization results with a relatively high degree of accuracy. However, applying these tools too early within the design stage adds additional challenges since vessel and hull designs are continually changing. Designers should have the freedom to explore the design space to determine an optimal solution. If CFD and refinement are implemented too early, designs become more rigid and lack the necessary flexibility in the early stages to explore the full model space. Nonetheless, due to the relatively cheap computational demand and the relatively high accuracy of the results, it is continually being implemented earlier within the design processes.

Added Resistance due to Fouling   Currently, the process of predicting fouling consists of estimating the roughness of the surface for the sea-trial conditions. To-date all *Feadships* are designed for sea-trail conditions. Ultimately, this means that the whole design process is focused on one-day results and does not consider the operational use. To account for these conditions, stochastic-based roughness values are implemented to predict the increase of resistance. However, the prediction of fouling is complex.

Fouling mainly causes frictional resistance to increase. However, pressure resistance is also affected due to the added turbulence imposed by the surface roughness. Ultimately, estimating fouling is not a trivial solution. Recently, an internal study by Alberts and Jacoby [5] was done to investigate, model, and predict the change in resistance coefficient factor, ΔC_f. This estimation was achieved by implementing a novel CFD procedure to estimate the effects of surface roughness. At which point, a surrogate modelling approach using a Gaussian Process Regression model was trained on numerous output results. The development of this tool has allowed for a much deeper understanding of how to estimate surface roughness. It has also allowed for validation of current estimation procedures, which

has never been considered before. Currently, design specifications do not require the consideration of hull fouling. As such, there is no current place for the implementation of such tools within the design process.

Added Resistance due to Waves and Wind   Multiple studies in collaboration with MARIN have been completed to determine the added thrust in waves. These investigations have resulted in a semi-empirically based quadratic transfer function *taw* (added thrust in waves) model based on a range of 18 *Feadship* yachts. In total, 258 tests in irregular seas and 18 tests in regular waves were conducted. The experiments consisted of varying seagoing conditions such as wave heights, wave periods, and relative headings while monitoring the vessel's required thrust. This new model has been designated as VoogtWAVE and requires, next to heading and speed, only four ship parameters to predict the *taw*; length overall submerged (L_{os}), beam at the waterline (B_{wl}), the displacement (∇) and the waterline coefficient of the foreship ($C_{wp,fore}$).

Ultimately, VoogtWAVE gives a robust prediction of *taw* and, as outlined by Grin [48], performs equally well as industry-standard, SPAWAVE (Grin [47]) when considering all other headings. However, it performs significantly better in head and bow-quartering seas. Unfortunately, some discrepancies exist between the empirical relations and the observed loads. From the results, there is much more added thrust in stern quartering seas than estimated. Grin [48] elaborates that the causes could be due to many factors (besides uncertainties in the prediction method and/or measurements). However, it has likely to do with a combination of a relatively limited number of wave encounters, roll resonance, active fin stabilizers, speed variations, yaw motions and steering actions.

In addition to waves, the added resistance due to wind can also be determined. Currently, there exists an extensive database of wind resistance coefficients, C_x. This database contains results from both direct wind tunnel model testing as well as CFD wind tunnel simulations. Wind tunnels have been completed for approximately half of the *Feadship* fleet. Due to the determined wind force coefficients' availability, the International Towing Tank Conference (ITTC) calculation procedure [94], can be closely followed. When inspecting the wind force coefficient of multiple *Feadships*, the coefficient determinations' general trends follow a very similar shape overall. Additionally, the deviation is relatively narrow banded. However, the main takeaway of these results is that the wind force's overall influence is relatively small compared to that of the wave added resistance component.

Unfortunately, at the moment, all things related to dynamic effects such as fouling, wind, and waves are currently not accounted for in the design process at any point. Nonetheless, a more detailed breakdown and elaboration of relevant existing propulsive and added resistance models are outlined in chapter 3.2.

2.3.2. Auxiliary Power Estimation

Auxiliary power estimations are determined empirically in the form of load balance sheets. This extensive sheet lists all components on board. A manually assigned running activity percentage is assigned next to each component to determine the average load for various operating conditions. When the load balance is developed, it is a highly subjective process to determine the equipment activity rates. From a practical perspective, the method has been routinely implemented as estimations generally provide a conservative evaluation of actual operation targets. Unfortunately, no explicit knowledge of what the systems need to do in practice is available. As such, empirical-based load balances are still commonly applied since they continue to "work fine." However, if this knowledge gap can be filled, clear improvements in operational use and efficiency can be achieved.

In addition to the contractor supplied load balance, a first-estimate empirical tool is the process of development. The approach provides a low-fidelity annual mean auxiliary power prediction. This method is still in the early stages of development and only considers a handful of vessels. Nonetheless, it can evaluate the average power consumption depending on the empirical vessel gross-tonnage comparisons. Much more data and considerations are required to have a robust estimation; however, detailed predictions are possible as the data set continually expands.

It should be noted that external contractors or shipyards currently perform nearly all auxiliary power

estimations. As such, expertise in estimating or evaluating this demand is not available. Neverthe-less, the knowledge and resources to perform such metrics are well within the realm of *De Voogt's* capabilities. Therefore, great strides and new initiatives are currently proposed to account for these considerations. Section 3.2.3 provides a more detailed breakdown and elaboration of relevant and currently applied auxiliary power estimation techniques.

2.3.3. Available Tools

Based on the above calculation method investigation for both propulsive and auxiliary power, an inventory of available tools and processes can be composed. These tools and whether they are currently being applied within the *DNVA* design process is seen summarized in table 2.2.

Table 2.2: Available and implemented De Voogt design tools

Criteria	Type	Design/Engineering Tool	Applied?
Operational Profiles	Database	7SEAS Portal	Yes
Propulsion Power Estimation	Semi-Emp.	H&M Propulsion Resistance	Yes
	Empirical	Resistance Curve Database	Yes
	CFD	Bare-Hull/Appendage Resistance	Yes
	CFD/S.Emp.	Added Resistance/Thrust in Waves	No
	CFD	Fouling Coefficient Estimation	No
	CFD/S.Emp.	Wind Resistance/Force Estimation	No
Auxiliary Power Estimation	Empirical	Load List Spreadsheet	Yes
	Empirical	Low Fidelity Prediction Model	No

2.4. Available Data

The amount of collected data used for operational monitoring from maintenance to voyages is currently substantial. As such, a further investigation into the availability and uncertainties associated with each source is crucial when implementing any future modelling applications. The types of data, their corresponding sources, and a description of each can be seen in table 2.3.

In-house design department benefits allow for easy access to critical vessel parameters such as; hull shape information and design specifications. Additionally, the close connections with the dedicated *Feadship* shipyards allow for easy access to general maintenance data. These databases are comprehensive; however, critical diagnostic features such as hull and propeller cleaning data and general engine diagnostics are accessible. Additionally, many new vessels are fitted with complex sensors and monitoring equipment. These devices allow for continuous data collection on parameters such as engine and generator power demand and experienced ship motions, respectively. Hindcasted weather databases are also accessed. These external databases are utilized to ensure accurate, up-to-date recorded climate phenomena for vessel operation evaluations. Finally, internal data collection and analysis initiatives related to voyage reporting and auxiliary power data are also available for use. While these are newly formed within *De Voogt Naval Architects* and still in their early stages, real-time system feedback to centralize, filter and evaluate existing operations of the vast *Feadship* fleet is currently growing at a rapid pace.

2.4.1. Data Limitations and Uncertainties

Technological improvements in sensors and sensor networks have opened many opportunities to use and combine geospatial data from sensors (Rodríguez and Servigne [90]). However, the data readings that are retrieved from these devices are often imprecise due to certain uncertainties. These can range from measurement errors due to the sensor itself and/or a discrete sampling rate of the associated

Table 2.3: Available dataset types, locations, and descriptions

Data Type	Data Source	Description
Ship Design Specifications	Inhouse design department databases	Ship design parameters: hull shape information (Length, Width, C_b, C_p) design loading conditions (draft, speed, range), LCG, LCB, propulsion systems, engine specifications, propeller selection, general arrangements
Maintenance Data	Inhouse maintenance and shipyard databases	State of hull and propeller fouling, date of last hull cleaning
Engine and Motion Data	Sensor monitoring	Main engine and generator power (shaft power), tanks levels (consumption), ship motions (pitch and roll)
Hindcast Weather Data	Copernicus Climate Change Service (CCS) database	All weather related parameters associated to past, present, and future climate conditions within Europe and the world
Voyage Report Data	*7SEAS Portal Initiative*	Onboard feedback monitoring: ship speed and heading, wave conditions (height, period, and directions), wind conditions (speed and direction) and corresponding measured operational profiles for *Feadship* fleet
Auxiliary Power Data	*Project HOTEL Initiative*	Additional hotel load system feedback monitoring: recorded total auxiliary loads (incl. sampling rates), AC power (voltage, amperage, and fan speed), air and sea temperature, and exterior relative humidity

measurements. As outlined in appendix A.1, such uncertainties and limitations exist within the retrieved datasets, which can impose enormous challenges on the associated analytics.

One such noticed challenge within *De Voogt* datasets, which is a common occurrence in all sensor systems, is related to incurred noise. As noted by Rodríguez and Servigne [90], sensor noise is a problem in all data collection systems due to interactions between various signal propagation. However, onboard yachts, this problem can be magnified since the allotted spaces are much more confined. This arrangement means that all wiring and cabling are packed tightly together. Due to this extreme packing, it has been seen that external cable signals have in the past influenced the retrieved generator datasets through additional noise within the system. However, it should be noted that this noise influence is generally marginal but has the ability to fluctuate the actual energy readings and provide unreliable readings.

Another source of uncertainty within the retrieved data can be related to the sensor systems' actual sampling rates. There is currently vessel information available for approximately 50+ ships; however, nearly all sampling rates are based on 3-minute averages. This discrete averaging sampling process smears out peak consumption values within the *De Voogt* datasets. As such, the recorded observations have the potential to underpredict energy usage in some instances.

The third source of uncertainty can be directly related to data management and measurement validity within *De Voogt* datasets. In some instances, total consumption readings do not directly align with the summation of individual consumption readings. For instance, during some operational cases, the HVAC system is recorded as providing zero power output. However, when considering each individual component readings, it can be seen that consumption is measured and provided. The total consumption readings in such cases are not a true reflection of the actual consumption. In these cases, engineering sense must be applied, and fact-checking should be completed when analyzing the data sets. Unfortunately, with massive datasets, it is challenging to filter where correct consumption is considered. Luckily, data is provided for individual generator sets. This means that instead of relying on uncertain combined data, a summation across the individual sources can be considered instead.

2.5. Method Requirements

Based on the problem introduction (chapter 1), and associated problem analysis (chapter 2), a clear series of problems can be identified. These problems can then be converted into a solution requirement based on the need to fulfill such challenges. Table 2.4 highlights the critical problems and associated requirements.

Ultimately, the main problem is that all energy consumption predictions are focused on highly conservative system design points and ideal operational conditions within the early stages. As such, *De Voogt* vessels are experiencing a large degree in uncertainty in the total energy consumption estimations upon operational use of the vessel. This deviation between operational conditions and prediction results concerns the future development of hybrid vessels and optimal design practices. Therefore, a solution must be found to provide a high degree of accuracy and certainty in both auxiliary power and propulsion loading by incorporating dynamic operational voyage data. The method should also highlight and elaborate upon its limitations and uncertainties.

After numerous discussions with both the *De Voogt* Design and Engineering departments (appendix A.1), it can be understood that no design accuracy requirement other than the fulfillment of the project design condition specifications are necessary. As such, a self-established accuracy margin has to be set to evaluate the proposed method's success. Therefore, a baseline accuracy for both auxiliary power and propulsion loading predictions should be no larger than a 15% deviation within the Contract Design stage. As such, summarizing the tabulated solution requirement criteria, the following method requirements which the methodology must adhere to can be proposed as,

1. *Estimate power for propulsion and auxiliary systems under dynamic conditions within $\pm 15\%$ with 95% Confidence Intervals (C.I.)*
2. *Ability to proportion both auxiliary and propulsion power consumption independently*
3. *Be based on available data within De Voogt Naval Architects databases*
4. *Be based on a modular methodology to easily incorporate various estimation tools and results*
5. *Be able to deal with discrepancies and errors in voyage report data*
6. *Be able to incorporate a range of ship sizes within the De Voogt fleet*

Table 2.4: Summarized problem definitions and solution requirements

Source	Summarized Problem Definitions	Solution Requirements
Section 1.1.1	Uncertainty exists in the prediction of auxiliary power loading. Currently predicted results are overpredicting operational consumption by an upwards factor of two (100% difference). This uncertainty is mainly attributed to the extreme probability of occurrence design points when considering the dynamic weather and sea conditions established in the early design stages.	A **low fidelity method** must be constructed that can accurately **predict auxiliary power demand within a degree of 15% using actual operational data** through the incorporation of available and measured weather and sea conditions metrics. Additionally, this model should highlight and exhibit a **95% confidence interval** region in which the model can effectively predict.
Section 1.1.1	Uncertainty exists in the prediction of propulsion demand within the early design stages. Estimations are currently under predicting the required shaft power by an upwards difference ranging 40% to 60% compared to operational results. This uncertainty is mainly attributed to conservative sea-trail design scenarios and neglect of environmental conditions such as wind, waves, and bio fouling.	A modelling technique capable of accurately **predicting the propulsion loading within a degree of 15% under varying operational conditions** must be constructed in addition to the current sea-trial correction methods. This model should be able to consider and incorporate available data metrics such as weather and sea conditions. The model should also consider the influence of bio fouling on the propulsion loading. Additionally, this model should provide a **95% interval region of confidence** that the model can predict.
Section 2.3	Multiple design tools are currently used to estimate total energy consumption and its various components: thrust and auxiliary power demand. However, these tools vary in both fidelity and where they are incorporated within the design process.	A modelling method needs to be developed, which **supports a modular framework.** The model should be able to incorporate empirical, semi-empirical or CFD based powering estimations tools to enhance prediction with the use of operational voyage data.
Section 2.4	Currently, only two vessels have both auxiliary and propulsion data available. However, soon this data set will grow to include more vessels. To-date *Feadship* consists of 50+ vessels with approximately 271 years of voyage specific data collected.	A modelling method needs to be developed to **easily incorporate a growing set of vessel operational data** to predict the energy consumption of **multiple vessel sizes** to further enhance and support feasibility studies within an early design stages.
Section 2.4	An extensive database of collected operational metrics is still in its infancy and is not currently being utilized within the *De Voogt* design framework to transition towards real-operational design.	The method should be **based on the existing and readily available internal databases** within the *Feadship*, *7SEAS Portal* and *Project HOTEL* platforms to assess energy consumption and loading proportions. It should also be able to handle, filter, and flag any discrepancies or errors within the data sets.

3

Solution Approaches

This chapter aims to expand and understand critical modelling approaches commonly applied within the maritime industry to evaluate full ship power demand predictions. As such, this section will seek to answer the following research question,

'What methods currently exist to predict EC for both propulsive and auxiliary loading accurately, and which approaches are most suited to achieve the modelling requirements?'

First, section 3.1 presents an introduction to what dynamic modelling methods are currently available. This section introduces three primary modelling categories found within the maritime field: the White-Box, Black-Box, and Grey-Box Modelling. Next, each approach is further elaborated, explored and detailed, respectively. Section 3.2 begins with an investigation into what White-Box modelling techniques exist for both propulsion and auxiliary power predictions. Black-Box modelling and associated frameworks will then be outlined and analyzed within section 3.3. Finally, within section 3.4, Grey-box modelling approaches are explored and analyzed. Upon conclusion of the various investigated solution approaches, a detailed literature summary is outlined within section 3.5. Here, the relevant literature gaps to justify the study objectives are highlighted. Next, section 3.6 relates the observed literature gaps with the proposed solutions. Each option is then compared with the model requirements for a final selection of the appropriate method. Finally, upon the literature investigation's conclusion, the study's completed scope is detailed and outlined within section 3.8.

3.1. Modelling Approaches within the Maritime Industry

Modelling of physical systems is usually applied by implementing two mathematical approaches: White-box models (WBM) or Black-box models (BBM). As outlined by Coraddu et al. [26], Huotari et al. [54], Leifsson et al. [68], Simian and Stoica [96], Zwart [108], the White-box approach models a physical system entirely using physical laws and deterministic first-principle relations, which is based on prior knowledge. The black-box approach models a system entirely based on observed data, such as input-output measurements, and requires no prior knowledge of the overall system. These methods usually focus on a range of statistical approaches, such as auto-regressive models or machine learning methods. As presented by Baldi et al. [12], Yang et al. [106], well trained BBMs can be more accurate than WBMs. However, BBMs require large amounts of high-quality data for model training, and more importantly, lack interpretability and extrapolating ability in contrast to WBMs. As further highlighted by Simian and Stoica [96], it is also possible to have models that deeply integrate both the White- and Black-box approaches, generating what is known as a Grey-box model (GBM). Ultimately, GBM attempts to combine both WBM and BBM advantages to overcome both individual solution's apparent drawbacks.

These three modelling approaches have been applied within the maritime industry when considering and evaluating dynamic ship power predictions. As opposed to building design, ships can be regarded

as self-sufficient islands. However, additional influences related to sea effects such as waves, currents, and biofouling are significant contributors to the required energy demands. Therefore, modelling the major power consumers is typically decomposed into two central portions: propulsion and auxiliary power. In the following sections (3.2, 3.3, 3.4), an investigation into each of the relevant modelling techniques, WBM, BBM, GBM, and their corresponding applications concerning power demand estimation are further investigated.

3.2. White Box Modelling

As noted within section 3.1, the White-box modelling approach has long been used to predict ship performance and estimate vessel energy consumption within ship design. These methods are often applied early within the design stages to evaluate and predict initial performance metrics. Generally, ship performance evaluations can be decomposed into the following individual components: total propulsive demand and auxiliary power demand. As outlined in Huotari et al. [54], a ship's power demand profile is most profoundly dominated by propulsion demand, especially in merchant ships. Nonetheless, mean auxiliary power predictions can significantly impact total energy predictions, where cruise vessels and yachts can sometimes be upwards of 50% the total demand, (Akershoek [4], Boertz [15]). As such, these models are typically separated and determined independently from the propulsive power demand. The total propulsive load, which can be considered in terms of resistance, can be further divided into two main elements: the calm-water resistance and the added resistance components.

$$R_T = R_{CW} + R_{AR} \tag{3.1}$$

Where R_T, R_{CW}, and R_{AR} are the total, calm-water, and added resistance, respectively. The latter component consists of predicting the added resistance of external forces such as wind, waves, and fouling. Until now, no distinction between thrust demand and resistance has been made; however, they are not the same. Stapersma and Klein Houd [98] highlights that the thrust–resistance relation must consider a commonly termed thrust deduction factor, t.

$$R = (1 - t)\, k_p\, T \tag{3.2}$$

Where k_p is the number of propellers on the ship. The deduction factor is used to account for the effect of the propulsor adding frictional resistance on the hull surface due to the phenomenon of drawing water along the hull, rudders and appendages. Further details about the propulsive chain processing and integration can be found in section 4.1, as the process is universal for all vessels, including yachts.

As detailed in section 3.1, the definition of a pure WBM depends on 100% deterministic metrics. However, most WBMs incorporate a degree of empirical results. These results generally consist of statistical analysis evaluations, which are used to enhance existing physical first-principle relations. The maritime industry consists of complex considerations such as fluid mechanics and fluid-structure interactions. As such, many implemented solutions can be considered semi-empirical and not true a WBM. Nonetheless, due to the relative practicality and generally low-fidelity of such models, they are commonly viewed as WBMs, (Yang et al. [106]).

3.2.1. Calm Water Resistance

The first component that is typically modelled is the calm-water resistance component. As outlined by Stapersma and Klein Houd [98], it is the force that is required to tow the ship at a specified speed (without the propulsor). The International Towing Tank Conference (ITTC) provides a list of recommended estimation methods to evaluate a ship's calm-water resistance. As reported by the Seakeeping Committee of the 29th ITTC [94], these approaches include direct towing tank tests, CFD numerical computations, and empirical formulations, which can be seen outlined in figure 3.1. Furthermore, the ITTC connects the relative fidelity and practicality of each method.

By nature, each method has its balance between general applicability and accuracy. As detailed by MARIN [73], dedicated methods for a restricted class of ships may have somewhat higher accuracy for one specific category. However, when nearing the generally narrow bounds of the parameter ranges,
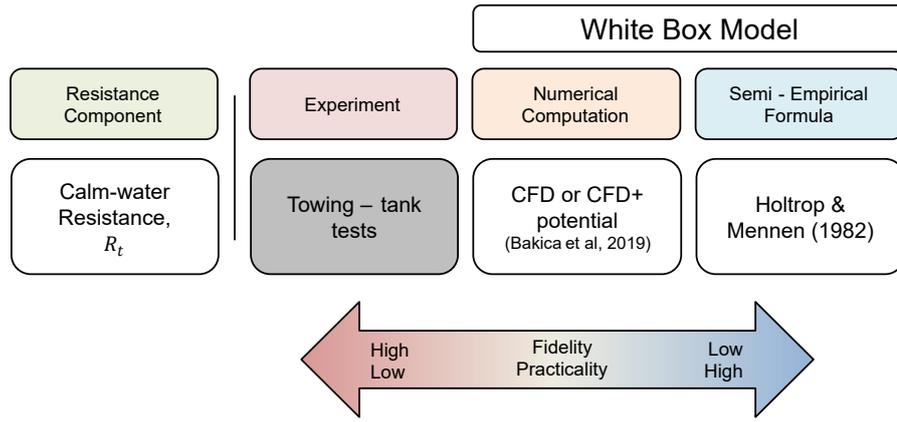
Figure 3.1: ITTC recommended Calm-water procedures (Seakeeping Commitee of the 29th ITTC [94])

it becomes ever more inaccurate. In contrast, generalized vessel methods are typically not very accurate for specific combinations of dimensions, speeds, and hull form coefficients. Nonetheless, The parameter ranges are broad; thus, the quality of the prediction is naturally sacrificed. As such, great care and understanding must always be applied when implementing such modelling techniques.

Fundamentally, direct towing tank model tests cannot be considered a WBM due to their empirical nature. Nonetheless, these techniques have been the foundation of many semi-empirically driven WBMs. These experimentally driven results generally produce accurate results; however, they are subject to experimental facility settings, procedural influences, and model-scaling effects, (Burger [19]). While these uncertainties are generally small, the cost in time and resources make these methods impractical for early-stage design.

The most commonly applied semi-empirical WBM is the Holtrop-Mennen method [51, 52]. As outlined by Yang et al. [106], this approach has been highly influential during the initial design stage to estimate calm-water resistance for displacement type vessels due to its ease in application and moderate accuracy. The method is based on regression analysis of a vast range of model tests and trial data, which gives it broad applicability. Ultimately, the approach divides total calm-water resistance into multiple independent resistance components,

$$R_{T,CW} = (1 + k)R_F + R_{APP} + R_W + R_A + R_B + R_{TR} + R_{AA} \tag{3.3}$$

These components include frictional resistance, R_F with form factor k for the hull variations, the resistance of appendages R_{APP}, a wave-making and wave-breaking resistance R_W, a model-ship correlation resistance R_A, pressure resistance due to bulbous bow R_B, an additional pressure resistance of the immersed transom R_{TR} and an air resistance component R_{AA}. Additional reviews of other WBMs can be found in Molland [76], Larrson and Raven [67], and Carlton [20].

In addition to empirical-based modelling, CFD can also be considered a WBM due to its strong first-principal roots. Furthermore, due to the decreased computational demand, CFD is routinely performed to estimate calm-water conditions. Nonetheless, for increased accuracy, more computational power is required. This approach has been extensively studied and validated with both model and full-scale results. As such, CFD can provide a much more robust and high-fidelity estimation of the induced resistance parameters. Unfortunately, as Zwart [108] outlines, CFD can suffer from both scale effects and simplifications, such as actuator disk propellers, free-surface effects, or disregard of trim and sinkage to reduce computing demand further.

Bakica et al. [9], investigated a novel early-stage numerical approach that compared critical resistance and propulsion parameters between model-scale and CFD. The results showed that an accuracy between 1% to 2.5% could be achieved for calm-water resistance. This analysis included actuator disk simplifications, which successfully managed to estimate hull-propeller nominal wake distribution within 3% of the model test estimates. Additionally, state-of-the-art CFD simulations can perform full-scale simulations directly to avoid simplifications by using more computational power. Niklas and Pruszko [80] compared the results determined by Holtrop-Mennen, towing tank experiments, full-scale CFD

simulations and sea trial measurements of a research vessel. As a result, the calm-water predictions using the empirical Holtrop-Mennen approach varied by approximately 18% compared to the sea–trial data. Furthermore, the model-scale predicted calm water resistance varied from −6% to 11% relatively to sea trials data. In comparison, the results calculated by full-scale CFD varied from −10% to 4% as compared to sea trial data. Niklas and Pruszko [80] stresses that for innovative hull forms, full-scale CFD simulations should support the towing tank method as the results can provide high accuracy at little cost.

These deviations in sea trial results depend on wind and waves, which are uncontrollable and not considered within calm-water conditions. As such, added resistance phenomenon such as waves and wind is exceptionally costly to predict well. Nonetheless, Niklas and Pruszko [80] investigation proves that full-scale CFD simulations can estimate the calm-water power requirement more accurately than the empirical-based WBM methods and equally (if not better) to model scale testing. In summary, accurately estimating the calm-water power demand is highly feasible with both model testing and in-depth CFD analysis. However, model-scale to full-scale scale effects is seen in both CFD and towing tank tests. These scaling effects can be overcome using full-scale CFD estimations. Nevertheless, this approach requires high demand for computational power. Unfortunately, such methods have reduced practicality for early-stage design due to the exceedingly high cost of time and resources, (ITTC [94]). As such, semi-empirical WBMs within the maritime sector, including the yachting industry, can still give a 1st powering estimation with relatively good accuracy for fast concept explorations and early-stage design.

3.2.2. Added Resistance

In addition to calm–water resistance, the added resistance due to wind, waves, and fouling can significantly impact the overall energy consumption. A detailed breakdown of ITTC recommended estimation methods for added resistance can be seen within figure 3.2. Based on a detailed analysis by Yang et al. [106], among wind, waves and current, the influence of the first two is considered much more significant than currents. As such, most relations focus on quantifying these complex component's impacts, among which the semi-empirical method is most widely used.

Resistance Component	Experiment	Numerical Computation	Semi - Empirical Formula
Added resistance due to wind, ΔR_{wind}	Wind tunnel test	CFD – Virtual Wind Tunnel (Janssen et al 2017)	Blendermann (1993), Fujiwara (2005)
Added resistance due to waves, ΔR_{wave}	Seakeeping test	CFD+ potential or Far/Near-field (Deng et al, 2010)	STAWAVE -I,II, SPAWAVE (Grin, 2014, 2020)
Added friction resistance due to Fouling, ΔR_{foul}	Statistical data collection and measurement	CFD+ potential (Demirel et al, 2014)	Townsin (2003)

White Box Model

High Low ← Fidelity Practicality → Low High

Figure 3.2: ITTC recommended added resistance procedures (Seakeeping Commitee of the 29th ITTC [94])

Most wave-generated thrust methods focus either on ship motion-induced added resistance or added resistance resulting from wave reflection. Considering the former, Jinkine and Ferdinande [62] developed a wholly empirical method that can provide the wave-added resistance in regular head seas for any wavelength. Whereas, Fujii and Takahashi [42], developed a semi-empirical approach that considers only the wave reflection component for blunted bow shapes. As detailed by Grin [47], the limitations of such methods are that they only consider head and short waves, respectively. Additionally, most formulations require detailed information on hull lines; thus, they are susceptible to change. Due to such restrictions, new initiatives requiring only the vessel's main particulars have been proposed.

The first new method is the STAWAVE2, which predicts added resistance in head seas only. The approach combines both the Jinkine and Ferdinande [62] and Fujii and Takahashi [42] methods with an updated estimation of the peak and tail height based on block coefficients. The second new method is SPAWAVE, which handles all wave directions. As outlined in section 2.3.1, SPAWAVE has further been adapted specially for yachts, namely *Feadships*. Both the STAWAVE2 and SPAWAVE method outputs consist of quadratic transfer functions (QTF), which generally exhibit similar trends, as seen in figure 3.3. In addition to the mentioned, several additional semi-empirical methods have been further investigated by Lu et al. [71].
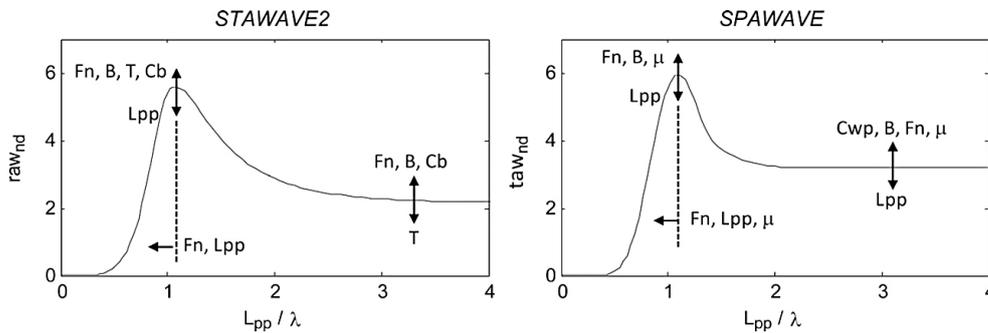
Figure 3.3: Parameters that influence STAWAVE2 (left) and SPAWAVE (right), (Grin [47])

Generally, wind resistance does not play a critical role in estimating total added resistance compared to the influence of waves. Nonetheless, Haddara and Guedes Soares [49] stresses that without proper considerations of the peak mean static forces and induced moments, effects can become critical in terms of efficient ship operation. Furthermore, Haddara and Guedes Soares [49] suggests that while wind tunnel tests are the most accurate procedure to estimate the wind-induced forces on ships, they are very time-consuming and expensive to complete. Therefore, when wind force coefficients, C_x, are available, the ITTC recommended procedure, [94], provides a novel formulation to evaluate the added wind resistance component for all ship types, including yachts.

$$\Delta R_{AA} = \frac{1}{2}\,\rho_{air}\,A_v\,C_x(\beta_{w,ref})\,V_{w,ref}^2 - \frac{1}{2}\,\rho_{air}\,A_v\,C_x(0°)\,V_s^2 \tag{3.4}$$

Where A_v is the area of the maximum transverse exposed section, V_s is ship speed over ground, $\beta_{w,ref}$ denotes the apparent wind direction, and $V_{w,ref}$ is the apparent wind speed. Unfortunately, due to such coefficient's general unavailability, semi-empirical white-box models have become popular for predicting wind resistance effects on all vessels. Both Ueno et al. [101] and Haddara and Guedes Soares [49] investigated commonly applied WBM approaches, namely, Isherwood [58], Blendermann [13, 14], and Fujiwara et al. [43, 44]. Generally, all method approaches are similar, where differences are mainly attributed to the vessel datasets used to perform the necessary regressions. Currently, ITTC recommends the former two approaches within their recommended procedures to account for added wind resistance if coefficient data is unavailable.

Fouling can also significantly contribute to total energy consumption. Unlike the wave and wind resistance components, this element is a component that is directly linked to the calm-water frictional resistance. As emphasized by Bressy and Lejars [18], heavy calcareous fouling may result in powering penalties of more than 85%. Even slime films can lead to significant increases in resistance and power of up to 20%, which may be particularly problematic for yachts that spend a large portion of their time

stationary. Unfortunately, this dynamic phenomenon is extremely complex and difficult to predict. As such, only low-fidelity analytical expressions exist for the first estimate of the roughness allowance. To-date the Propulsion Committee of the 28th ITTC [89] recommends using Townsin [100] to estimate the roughness allowance coefficient, ΔC_F, directly,

$$\Delta C_F = 0.044 \left[\left(\frac{k_s}{L_{wl}} \right)^{1/3} - 10 \cdot Re^{-1/3} \right] + 0.125 \cdot 10^{-3} = \frac{\Delta R_F}{1/2 \, \rho_w \, S_{BH} \, V_G^2} \tag{3.5}$$

The equivalent sand grain roughness height, k_s is in micrometres, Re is the Reynolds number based on vessel length, L_{wl} is the waterline length, and S_{BH} is the wetted surface of the hull. As emphasized by the Propulsion Committee of the 28th ITTC [89], there is currently no accurate and universal method to predict ship fouling. Suggesting that only by studying a large variety of ships can statistically reliable data be obtained over an extended period. As such, Uzun et al. [103] offers a practical time-dependent growth model to predict fouling of service vessel-based data analysis. The model's outcomes presented a solution that shows a linear increase in power demand of approximately 10% in the effective power per year for service vessels.

As mentioned above, CFD is an excellent alternative to semi-empirical formulations, as they are generally more robust and accurate. In addition, they have been widely verified and validated for calm-water conditions; however, added resistance due to waves, wind, and fouling has proven to be more challenging.

Sadat-Hosseini et al. [92] investigated and evaluated the effects of added resistance in short and long head waves by use of RANS CFD. The investigation was verified using experiment benchmark data from various model tests and analytical approaches. Ultimately, the added resistance trend follows quite closely; however, the simulation under predicts the resistance for wavelength ratios, $\lambda/L > 1.2$ and over predicts for $\lambda/L < 1.2$, where an average error of 20% was found. Additionally, the added resistance of a container vessel in head waves was studied during a Gothenburg 2010 workshop. Here, Deng et al. [33] determined added resistance and motions estimations within an average error deviation of 17.7% using a RANS CFD solver. While general trends were accomplished well, exact replications proved difficult even with the additional computational power.

Estimation of wind force utilizing CFD is also a possibility to improve accuracy. Koop et al. [66] investigated the effects of wind on typical offshore vessels in open-water conditions. When comparing the force coefficients obtained from CFD simulations with wind tunnel data, deviations ranged from 1% to 20% relative difference depending on wind heading. However, the average deviation, while not explicitly stated, is well below 5%. Janssen et al. [61] performed a CFD sensitivity analysis on cruise vessels, where the impact of geometrical simplifications on the loading coefficients was investigated. From the results, the most detailed model had an average difference in measured and CFD wind loads of only 5.9%, whereas the most simple box solution average deviation was 37.8%. Much like CFD for wave predictions, wind estimations also suffer from high computational costs due to airflow complexity. However, as mentioned previously, and enforced by MAN Diesel & Turbo [72], the influence of wind compared to wave effects is minor as it only contributes around 2% of total calm-water ship resistance.

Fouling remains a challenge to predict accurately; however, CFD has shown promising results. Demirel et al. [31] demonstrated and validated a novel approach to accurately predicting surface roughness influences on container vessels. The validation study showed a good correlation between the experimental data with the CFD experiment, where the friction coefficient deviation ranged from 0.14% to 2.54%. As noted in section 2.3.1, Alberts and Jacoby [5] adapted this methodology towards *Feadships*, where similar results were achieved. However, it was stressed that the major issue with predicting fouling with CFD is the high computational cost. Unfortunately, this becomes detrimental and impractical within the early design stages, where the ship's vessel geometry still changes considerably.

Ultimately, it can be concluded that a high degree of accuracy can be achieved for added resistance components when implementing CFD. However, the computational cost of such simulations is extensive and time-consuming to perform. Thus, for early-stage predictions, these methods are generally impractical to perform. As such, the benefits and insight gained by determining the dynamic influences are mainly attributed to semi-empirical WBM practices. While these methods allow for a high degree of practicality, their accuracy is only suitable for early estimations and has proven to show a significant degree of variation.

3.2.3. Auxiliary Power

As mentioned in section 3.2, the propulsion and auxiliary are generally discussed as separate entities. The auxiliary power and its associated systems function to ensure the vessel is habitable in terms of hotel systems, survivability functions, and all-electric requirements to satisfy entertainment and comfort needs. Ultimately, many methods have been developed to estimate total auxiliary power demand, including empirical formulations, electrical load analysis, and simulations as highlighted in figure 3.4.
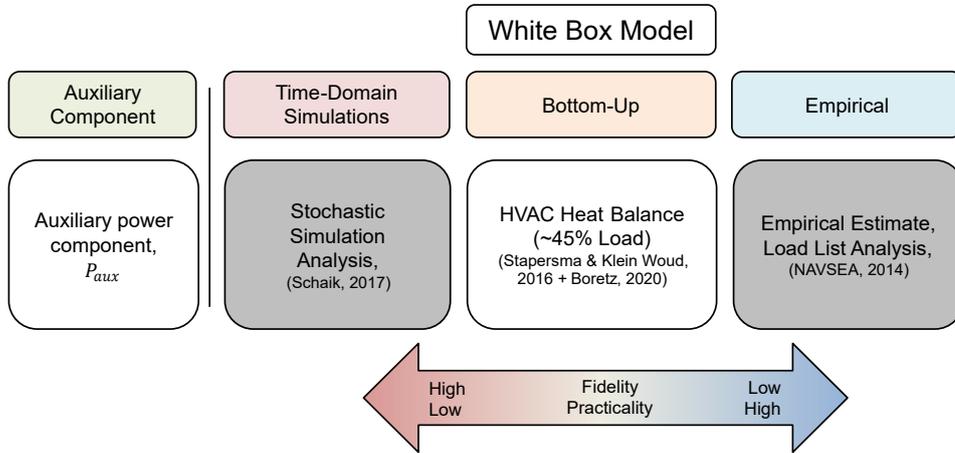
Figure 3.4: Various recommended auxiliary power estimation procedures (based on Stapersma and Klein Houd [98])

As outlined by Stapersma and Klein Houd [98], empirical formulae can be used successfully to obtain a first estimate of the electric power demand in the pre-design stage. However, this is only possible if the determined relations are based on sufficient ship data and similar operations. From a purely WBM perspective, wholly empirical estimations do not fall within this category. Besides the general lack of accuracy, no physical insight into the system is known a priori. As such, auxiliary power demand is generally determined either through a load balancing analysis or simulations, which usually consist of multiple aggregated WBM components.

The Electric Load Analysis (ELA) approach is a commonly applied method for determining electrical power demand. Historically, the algorithm combines individual loads to determine the total power generation. Thus, the actual power demand is calculated as a function of absorbed consumer power, P_a, and multiple empirical factors for different operating conditions, as seen in the equation.

$$P_{total} = \sum P_a \cdot L_f \cdot k_u \cdot k_s \tag{3.6}$$

L_f is the load factor and is defined as the relative percent load of the maximum electric power the equipment absorbs in operation. The load factor typically varies between 0 and 1, signifying no-load and fully-loaded, respectively. The number in service, k_n, is defined as the number of running equipment in the operational condition. The third and final factor is the simultaneity factor, k_s, which varies between 0 and 1. This factor accounts for each item's relative mean operational use, such as intermittent operation instead of continuous. Subsequently, the product results in the average absorbed power, which can be totaled to obtain the complete auxiliary power.

Stapersma and Klein Houd [98] emphasize that the load and simultaneity factor's estimation is the most challenging part of ELA. These load factors are ultimately based on experience and are often estimated too high to reduce the possibility of designing an undersized generator. This conservatism overestimates the electric power demand and leads to high procurement costs and low average loading efficiencies for the generators. While the ELA approach determines the maximum operational demand, the relative accuracy is limited due to the uncertainties associated with the ambiguous load factor selections (Naval Sea Systems Command [77]). Thus, to overcome these accuracy deficiencies, Stochastic Load Simulations (SLS) can be performed under various operational conditions. Instead of providing fixed demand factors, this approach incorporates the uncertainty margins using probability distribution functions (PDF) to evaluate an expected full operating range. Unfortunately, both these modelling

techniques cannot be considered WBMs since the determination of load factors is still based on pure experience and empirical vessel comparisons. Thus, to overcome these ambiguous challenges, Boertz [15] investigated a bottom-up approach.

A bottom-up approach is a foundational method that considers multiple independent model contributions. This procedure allows for energy consumption to be calculated for every known activity individually and aggregate each component for a total estimation. Therefore, a high degree of operational insight is required, which often leads to the consensus that the approach is too complicated within the early design stages. Nonetheless, Boertz [15] managed to distinguish, predict and validate many individual auxiliary power sub-models and their interactions concerning average 24-hour loading under various cruise ship operational conditions. The primary consumer components can be seen in figure 3.5.
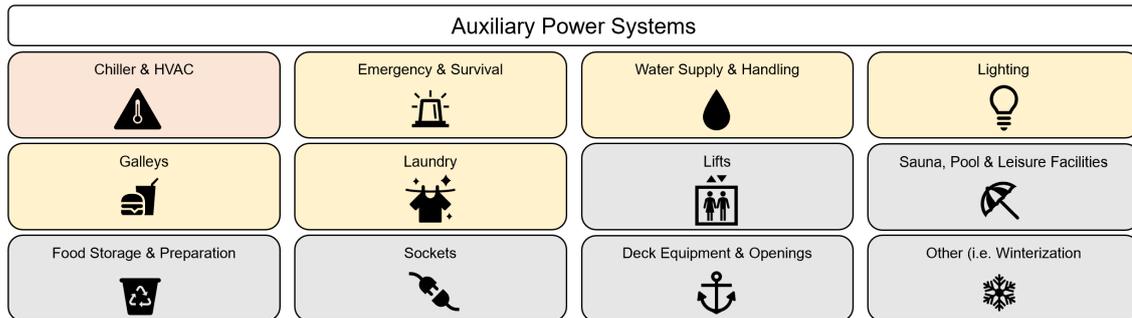
Figure 3.5: Auxiliary power demand breakdown, (Boertz [15])

Of these models, it was observed that the HVAC systems are dominant, providing a mean contribution of 30% to 50% demand within the total auxiliary load due to high comfort requirements, passenger and crew behaviours, and external environmental conditions.

Typically, the required heating and cooling loads onboard vessels are determined through the first-principles heat balancing approaches. This WBM estimation serves as the basis for determining the required power to maintain a state of equilibrium when considering dynamic factors such as heat transmission through envelopes (walls, floors, and ceilings), solar radiation, people, infiltration, lighting and any additional equipment. The ISO 7547 [59] and Stapersma and Klein Houd [99], provides guidance in the design of ventilation and air-conditioning systems in ships. Additionally, ASHRAE [7] provides general design guidance for a multitude of interior spaces and their associated space requirements in terms of heating loads, occupancy and minimal air changes required. Typically, for each space, estimations are done independently as the usage of onboard spaces can vary dramatically. The commonly suggested procedures are typically the basis for more detailed modelling and simulation tools. Thus, the provided frameworks give an excellent estimate and are typically used to evaluate the air handling system's selection.

In summary, many techniques exist to evaluate the complete auxiliary power demand of vessels. These include empirical approaches as well as Electric Load Analysis and Stochastic Load Simulations. However, due to their inherent empirical foundations and lack of physical interpretability, the methods cannot be considered WBMs. However, while more complicated, a bottom-up approach can incorporate multiple sub-models to evaluate and aggregate individual power demand. Many of these models can be considered as WBMs as they principally rely on deterministic and physical foundations. As suggested by Boertz [15], HVAC systems are the most significant contributors to the total auxiliary energy consumption. Therefore, detailed WBMs to accurately estimate required heat load balancing are generally applied to capture dynamic features such as ambient environmental conditions. It should be noted that these methods are universally used in not one, but all sectors of the maritime field, including the yachting industry.

3.3. Black Box Modelling

In addition to WBMs, the Black Box Model (BBM) approach can be applied to estimate the maritime industry's powering requirements. Ultimately, a BBM is a mathematical method that directly describes the relations between system inputs and associated outputs. As mentioned in section 3.1, these models do not rely on any prior knowledge or theoretical considerations. Leifsson et al. [68] highlights that these modelling techniques are beneficial when a system's behaviours are not fully understood or when WBM lacks either predictability or accuracy.

While such methods have clear advantages over WBMs, many disadvantages exist as well. As outlined by Haranen et al. [50], all BBM must be trained on parameters estimated from data collection. Therefore, models are highly dependent on the quality of the associated data. If these measured results are not reliable, the modelling uncertainty can be high. Additionally, Yang et al. [106] and Leifsson et al. [68] emphasize that due to the high dependency on data, BBMs' extrapolation characteristics are limited to the data sets from which they are derived. Uncertainty within BBMs can also be associated with the fundamental intrinsic model parameter selections. These parameter uncertainties can cause detrimental effects such as underfitting or overfitting of an associated dataset. This phenomenon hinders the model's capability to generalize itself to new datasets without the influence of data bias. As outlined by Zwart [108], it is expected that the model ideally learns genuine relationships between the input and output variables. If care is taken to effectively tune the model's parameters, predictions and interpolations between data points are possible in new unobserved circumstances where no data is accessible.

Nonetheless, state-of-the-art literature has demonstrated BBM's use to predict ship performance characteristics, such as speed, propulsion power, fuel consumption, fouling influences, and early-stage wave effects. Whereby, a multitude of different machine learning algorithms have been tested, such as Artificial Neural Networks (ANN) and Gaussian Process Regressions (GPR), with remarkable results for an array of vessel types. The general schematics for comparison of each BBM structure can be seen in figure 3.6.
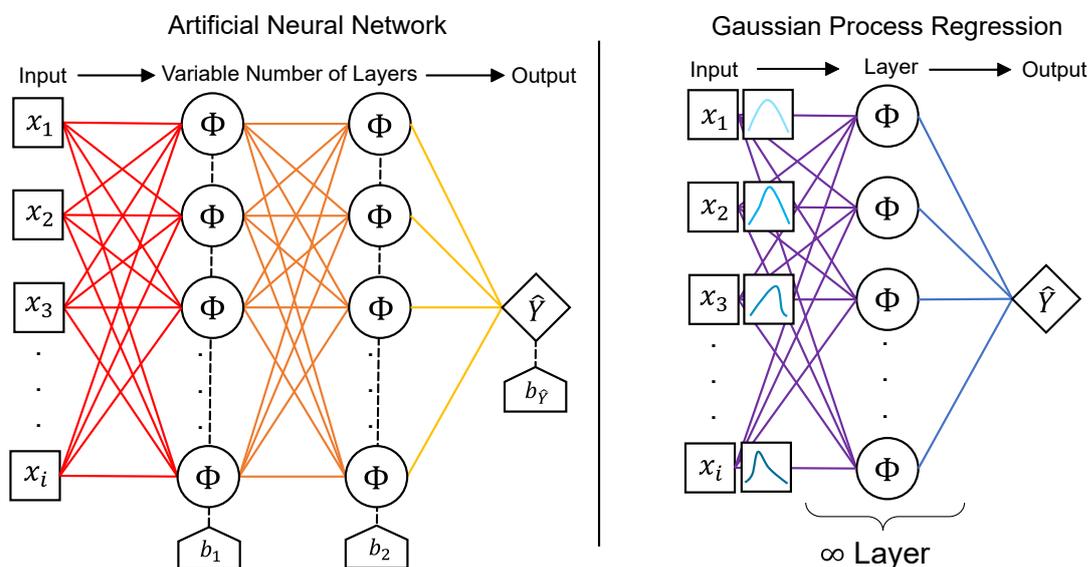
Figure 3.6: Conventional schematic of an ANN (left) and GPR equivalent (right), (based on Zeni et al. [107])

Arguably one of the most applied methods within the maritime industry is the ANN. Pedersen and Larsen [84] was one of the first to train an ANN to predict propulsion power for container vessels. Full-scale propulsion power measurements were used as target prediction data; whereas, ship speed, wind speed and direction, sea, and air temperature, from four different loading conditions were used as input features. It was determined that the network could predict propulsion power with a mean relative error of less than 2.7%. These estimations were directly compared to traditional semi-empirical WBM resistance methods, which showed an associated relative error ranging from 17.9% to 28.1%.

Furthermore, Pedersen and Larsen [85] enhanced the previous ANN model by incorporating hindcasted weather data. The ANN estimation improved significantly, reducing from 2.7% to 1.6% relative error. A novel methodology by Parkes et al. [83] demonstrated that a simple ANN structure could predict, with a high degree of repeatability, the propulsive power with an accuracy of 8% for shipping vessels. It was further enforced that such a model can be used in weather-routing optimization and establish new power margins for newbuild projects. In addition to propulsive power, Bal Beşikçi et al. [11] developed an ANN to estimate fuel consumption. The model showed promising results compared to a multiple regression analysis, where a relative mean error estimation of 6% was observed.

Gaussian Process Regression (GPR)is also a popular BBM approach applied within the maritime industry. Petersen et al. [88], analyzed and compared both ANN and GPR approaches for modelling fuel efficiency and propulsion power. The investigation incorporated a novel and publicly available high-frequency dataset for a ferries operational window of two months. In all cases, the ANN predicted slightly more accurate results. In contrast, the GPR models inherent probabilistic properties allowed for quantification of uncertainty directly. Similarly, Pedersen and Larsen [86] evaluated and compared the ANN with the GPR. Both results demonstrated comparable accuracy when using noon report and hindcasted weather data. It was noted that an advantage of the GPR was the extraction of known 'length scales,' which provided informed features analysis. However, for large datasets, the ANN computationally scaled much better, allowing for much faster responses.

Due to BBM's success in predicting ship performance metrics, the application has expanded to related areas. While most literature focused on fuel consumption and propulsive power estimations, added resistance characteristics such as waves, winds, and fouling estimation also show a high degree of potential.

Cepowski [23] developed an ANN model to predict the non-dimensionalized added wave resistance coefficient, C_{AW}, and compared the results with experimental and popular semi-empirical WBM models for 14 different vessels. It was shown that the BBM was able to accurately predict within $\pm 1.2\%$ of the real-world figures, thus enforcing a practical method for preliminary design stages. However, Cepowski [23] emphasizes that the estimations outside of the parameter ranges are unreliable, therefore hindering the use cases. Haddara and Guedes Soares [49] used a conventional ANN methodology to estimate the wind force coefficients using the Blendermann [13, 14] datasets as training data. The investigation determined that the numerical estimations agreed well with the experimental results while coping with various vessel draft conditions. Unfortunately, no prediction results were provided; but, it was concluded that such techniques could deliver better results than WBM approaches typically outlined in the literature. Most recently, Coraddu et al. [27] performed an investigation into estimating the speed loss caused by the effect of fouling on a ship's propeller and hull. The investigation determined that BBMs proved to be more accurate and consistent in predicting performance loss over time. However, it should be noted that the process involved a detailed and expansive dataset dedicated to the estimation of fouling for new future regulations analysis. Nonetheless, the proposed method could effectively determine the intervals between propeller and hull cleaning actions while accurately estimating ship efficiency.

Unfortunately, the use of ANN or GPR, to the best of the author's knowledge, has not been applied to estimate auxiliary power estimations within the maritime industry. As such, a noticeable gap in the literature can be observed. However, within the building engineering field, this technique has been widely applied for estimating energy demand. Runge and Zmeureanu [91] conducted an intensive literature investigation composed of 91 sources. It showed that BBM ANNs were used for 81% of all the studies. Of these studies, the results showed a high degree of prediction accuracy, where the mean average deviations ranged from 0.001% to 19.1%. Additionally, Kalogirou and Bojic [63], Karatasou et al. [64] and Neto and Fiorelli [79] developed conventional ANNs for estimating energy demand within commercial buildings. The results enforce the BBM's capability by estimating energy demand with a relative mean average of 8.7%, 9.0%, and 16.5%, respectively. As such, it can be expected that the auxiliary power estimation using BBM should likewise show similar promise within the yachting industry since the foundational habitability principles remain the same.

3.3.1. Black-Box Complexity and Uncertainty

In summary of the BBM literature investigation, mention of scalability and uncertainty characteristics between ANN and GPR was consistently observed. However, none of the literature quantified the degree difference between complexity or delved into each approach's probabilistic aspects.

In computational science, time-complexity is the computational metric that describes the time it takes to run an algorithm. The time complexity is typically estimated by tallying the number of basic operations performed within the algorithm, supposing that each elementary function takes a fixed amount of time to perform. This process is generally described as the big O notation, where the operation relations n describes the relationship between time and input size. While many forms of complexity notation exist, figure 3.7 outlines the commonly determined dependencies.
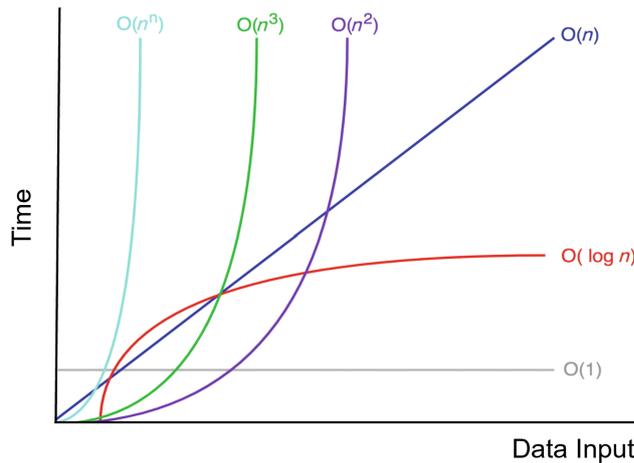
Figure 3.7: General model time-complexity comparison, (Bae [8])

Pang et al. [82], investigated and compared neural networks, the Gaussian process regression, and a variant mixture of the two. The study outlined that no matrix inversion is required for the ANN within the training process. Therefore, the complexity is generally represented as a quadratic $O(n^2)$, where n is the number of neurons and weights within the ANN. For the GPR, the training results in a non-convex loss function. This shape implies that multiple starting points are necessary, resulting in a matrix inversion technique. Inversion requires $O(N^3)$ complexity, where N is the number of training data points. Ultimately, Pang et al. [82], concluded that for large datasets (greater than 1000 data points), GPR is a much less attractive option than the ANN due to the significant disparity in complexity.

While complexity is unattractive, GPR has the inherent ability to quantify modelling uncertainty within estimations. This is done using a Bayesian inference approach that considers the probabilistic distributions over all possible functions consistent with the observed data. Thus, the evaluation of the variance within estimations allows for the determination of the total uncertainty. Unfortunately, ANNs are not built upon this principle and cannot quantify the uncertainty directly. However, techniques do exist to evaluate such metrics artificially. The most commonly applied method is bootstrap aggregation, otherwise known as bagging (Ferrario et al. [41], Mazloumi et al. [75]). This technique implements a random sampling with replacement approach to develop an empirical probability distribution of each parameter. These distributions are then used to evaluate variance independently for an ensemble of ANNs. This method allows for a synthetic development of modelling uncertainty, which can be expanded into a total prediction uncertainty by further incorporating the target variances.

The modelling comparison between ANN and GPR yielded essential considerations of each approach. The first is computational time; for large datasets, the ANN is much more suitable in all instances. However, GPR does have inherent uncertainty characteristics as opposed to ANNs. Nonetheless, uncertainty can be obtained for ANNs if statistical ensembling techniques are leveraged. Unfortunately, this involves developing random distributions using a series of models, thus increasing the computational cost. Therefore, a trade-off must be considered: uncertainty for time-cost.

3.4. Grey Box Modelling

The term grey-box modelling (GBM) refers to the combination of both a WBM and BBM. These modelling approaches aim to extract each model technique's advantages to overcome each model's consequences. Ultimately, such a method seeks to retain critical physical relations present from a WBM, while the BBM integrates complex dynamic behaviours obtained via operational data. As outlined by Yang et al. [106], such models have successfully increased extrapolation capacity, interpretability, modelling accuracy, and reduced the amount of required data.
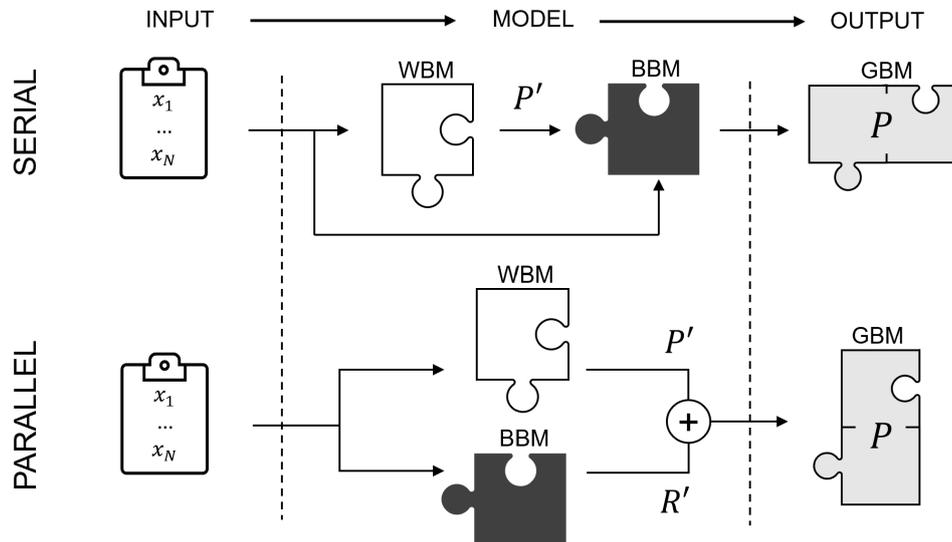
Figure 3.8: Serial grey-box modelling (top) and parallel grey-box modelling (bottom), (Leifsson et al. [68])

As detailed by Leifsson et al. [68], GBMs can be typically distinguished into two main categories depending on the application: serial modelling and parallel modelling. These framework's different compositions can be seen in figure 3.8. A serial modelling approach involves a WBM and BBM set in series. The inputs are fed to the WBM and BBM; however, the initial prediction (P') of the WBM is integrated directly into the BBM. In this situation, a mapping between the applied physics and the operational data can be internally developed. The parallel modelling approach involves a WBM estimation, where a BBM is used in parallel to minimize the residual (R') between prediction and target data. These are then combined to determine the final prediction.

Using these two categories, Leifsson et al. [68] initiated a comparative investigation between WBM, BBM, and GBM. The results determined that GBM modelling significantly increased the vessel fuel consumption predictability over the WBM, where marginal differences were observed between the serial and parallel approaches. Ultimately, Leifsson et al. [68] concluded two primary advantages of the GBM. The first is an indication of the improved extrapolation ability of the GBM. The second indicates how a GBM can integrate influences of physical phenomena ignored in the WBM. Coraddu et al. [26] also investigated the effects of GBM on the estimation of shaft power, shaft torque, and fuel consumption for a varying number of data samples. Here, two different serial approaches were adopted. A general serial approach and a more advanced serial- approach where internal parameter optimization was performed. The investigation showed that the GBM performed much better than both WBM and BBM, where a mean average percent error range of 0.8% to 1.5% for 1000 data points was recorded. Additionally, comparing the two serial GBM approaches showed minimal deviations of approximately 0.1%. Ultimately indicating that a simple approach can be both practical and effective.

Much like the above section 3.3, the GBM has not been applied to evaluate the auxiliary power consumption within the maritime sector, let alone within the yachting industry. However, within the field of buildings engineering, this approach has been lightly explored. For example, Braun and Chaturvedi [17] adopted a thermal network grey-box modelling approach to model office buildings. The grey-box model achieved an 8.6% root mean squared error (RMSE) in the predicted sensible load with only two weeks of training data. Furthermore, Siemann [95] investigated two different Grey-box models for

residential building energy demand. These GBMs led to improvements of 39.6% and 28.3% over the conventional models in estimating the hourly load. Thus, while the industries are different, parallels between buildings and maritime engineering can be drawn, suggesting similar results can be obtained.

3.5. Literature Summary and Gap Analysis

To successfully address the proposed research question,

'What methods currently exist to predict EC for both propulsive and auxiliary loading accurately, and which approaches are most suited to achieve the modelling requirements?'

The literature investigation examined three main modelling techniques within dynamic energy consumption predictions, WBM, BBM, and GBM. The comparison and associated literature source summary can be seen in table 3.1.

Table 3.1: Literature summary and comparison of different modelling techniques

Model	Type	Author Sources	Advantage	Disadvantage
White Box Model	Calm Water	CW [1] [51, 52, 94] CFD CW [1] [9, 80]	• Can interpret prediction results and system behaviour • Can extrapolate beyond the given data range • Does not require any historical data to function	• Accuracy of predictions depends on assumptions and uncertainties implicit in the models • Requires a lot of priori knowledge
	Added Resistance	Wave [1] [42, 47, 48, 62] Wind [1] [13, 14, 43, 58, 94] Foul [89, 100, 103] CFD Wave [1] [33, 92] CFD Wind [1] [61, 66] CFD Foul [1] [5, 31, 32]		
	Auxiliary Power	Aux. Power [1] [15, 77, 98, 99] HVAC [1, 2, 7, 59]		
Black Box Model	ANN	Prop. Power [83–85] FO Con. [11] Wave [23] Wind [49] Foul [27] Aux. Power [63, 64, 79, 91]	• Does not require any a priori system knowledge • More accurate compared to WBM	• Requires a large amount of historical data • Poor model interpretability • Poor extrapolation capacity • May result in unreasonable results (overfitting, under-fitting, scale-effects) • No universal solution as different models are better suited for different applications
	GPR	FO Con. [53, 86, 88]		
	Model Compare	Complexity [82] Uncertainty [41, 75]		
Grey Box Model	Parallel (P) and Serial (S)	P - FO Cons. [68] S - FO Cons. [26, 68, 106, 108] S - Pro. Power [26, 108] Aux. Power [17, 95]	• Accuracy > WBM • Historical data required < BBM • Model interpretability > BBM • Extrapolation capacity > BBM	• Accuracy < BBM • Historical data required > WBM • Model interpretability < WBM • Extrapolation capacity < WBM

WBM techniques are universally used within the early design stages of all marine sectors. Currently, a great deal of overlap exists between the investigated WBMs outlined in section 3.2 and the methods outlined in section 2.3; however, these approaches have been adapted explicitly towards *Feadship's* through the use of empirically obtained results. General WBMs can decompose complex problems, such as propulsive and auxiliary demand, into smaller sub-models for increased physical insight. Unfortunately, this creates scenarios where many individual parameters are required to suitably solve such problems. Additionally, due to the complex nature of dynamic interactions, such models, while offering a high degree of practicality, lack the necessary accuracy required for later design stage calculations. While CFD, model tests, and electric load simulations can be used to offer a much more

---

[1] Calculation tools within *DNVA* are available as seen in table (2.3.3)

robust and accurate solution, these techniques require a high degree of system information generally not available within the initial design stages of yacht design.

On the other hand, BBM techniques have been successfully applied within the maritime industry to accurately evaluate dynamic effects. While these approaches have been applied and verified within numerous areas related to energy predictions, many challenges still exist. For example, BBM is highly dependent on input data quality, which directly influences the estimation capacity. Additionally, these models are mapped as input-output relations. As such, the interpretability and physical understanding of the internal interactions are exceedingly difficult to evaluate. Ultimately, it can be concluded that, while these methods are powerful, data quality, data amount, and the correct parameter tuning are driving factors in regards to accuracy. Unfortunately, no literature has explicitly studied yachts; however, extrapolation of the method is likely as success has been documented for multiple vessel classes, including shipping vessels, passenger ferries, cruise ships, and workboats.

GBM is a state-of-the-art method that combines both the WBM and BBM to overcome each individual's deficiency. Success has been observed in improving accuracy, interpretability, and extrapolation capacity. However, only a few literature sources are available for study due to the method's relatively new induction within the maritime field. Nonetheless, the approach shows a great deal of promise, where a detailed literature comparison of the various modelling approaches has been developed and detailed by Yang et al. [106].

A clear literature gap can be formed when considering all sources relevant to energy consumption predictions under dynamic loading conditions. Figure 3.9 outlines this evident break within the investigated literature.

Figure 3.9: Dynamic energy prediction literature gap analysis

It can be observed that the prediction of propulsive demand or the associated fuel consumption has been investigated for all modelling categories. However, the studies have primarily been orientated towards shipping vessels, where little literature for other vessel classes has been discovered. This divide leaves a large gap specifically orientated to yachts, which are designed for a completely different purpose and operational profile in comparison.

Furthermore, while BBM and GBM dynamic auxiliary power estimations have been well documented in building engineering, the maritime industry has yet to apply these techniques to such areas. While GBMs seemingly overcomes many WBM and BBM deficiencies, not many areas other than propulsive power and fuel consumption estimations have utilized these modelling approaches. As such, considerable academic and industry contribution can be gained depending on the proposed solutions (section 3.6) and associated scope of work (section 3.8).

3.6. Proposed Solution

Having outlined the various solution approaches in table 3.1, and identifying a literature gap in figure 3.9, a precise evaluation of the method requirements (table 2.4) can be conducted to determine the most suitable modelling solution to evaluate both propulsion and auxiliary power consumption under dynamic conditions. The evaluation and comparison breakdown of each modelling approach with solution requirements can be seen in table 3.2.

Table 3.2: Method requirements and solution approaches evaluation

Requirements	White Box Model	Black Box Model	Grey Box Model
Estimate power for propulsion and auxiliary systems under dynamic conditions within $\pm 15\%$ with 95% C.I.	**Unlikely** ⊖	**Likely** ⊕	**Likely** ⊕
Ability to proportion both auxiliary and propulsion power consumption independently	**Yes** - Many Models ⊕	**Yes** - Single Model ⊕	**Yes** - Few Models ⊕
Be based on available data within *De Voogt Naval Architects* databases	**Limited** ⓘ	**Yes** - All data ⊕	**Yes** - All data ⊕
Be based on a modular methodology to easily incorporate various estimation tools and results	**Yes** - Multiple WBM ⊕	**No** - Singluar Model ⊖	**Yes** - WBM + BBM ⊕
Be able to deal with discrepancies and errors in voyage report data	**Limited** ⊖	**Yes** - Pre-processing ⓘ	**Yes** - Pre-processing ⓘ
Be able to incorporate a range of ship sizes within the *De Voogt* fleet	**Yes** - Physics dependent ⊕	**Yes** - Data dependent ⓘ	**Yes** - Data + Physics ⓘ

Based on the detailed above comparison, the approach that will most likely satisfy all method requirements is an Artificial Neural Network Grey-Box modelling (ANN-GBM) approach using a serial configuration. By nature, the GBM is a modular solution that can incorporate all data types, be it empirical, semi-empirical, CFD, or data-sensor features. The serial framework will be used, where a WBM is integrated as a direct input within the BBM. This framework will offer a relatively simple implementation process compared to the parallel or advanced serial techniques at little to no cost towards modelling accuracy (section 3.4).

A total of four WBMs will be incorporated to initially evaluate both propulsive (three models combined) and auxiliary (one model) energy demand as highlighted in figure 3.10. The first propulsion-focused WBM will be based on the Holtrop and Mennen [51, 52] method for calm-water resistance. Additionally, two WBMs to account for wind and wave effects will be implemented. The Wind model will be based on the ITTC method [89], where the wind-force coefficients are extracted from previous *DNVA* simulations and wind-tunnel tests. The VoogtWAVE model [48] will be used to account for wave added thrust from all directions. All three models are well established and readily available for use (section 2.3.3) within the GBM framework. While WBMs to capture dynamic environmental effects are not necessarily needed since the BBM can account and scale for this portion, an improved mapping between the physics and the operational data can be internally developed. This interaction will hopefully allow for enhanced interpolation and extrapolation capacities. Unfortunately, due to the difficulty of accurately accounting for fouling contributions, this portion will not be explicitly modelled. Instead, an annual power increase of 10% will be incorporated as per Uzun et al. [103].
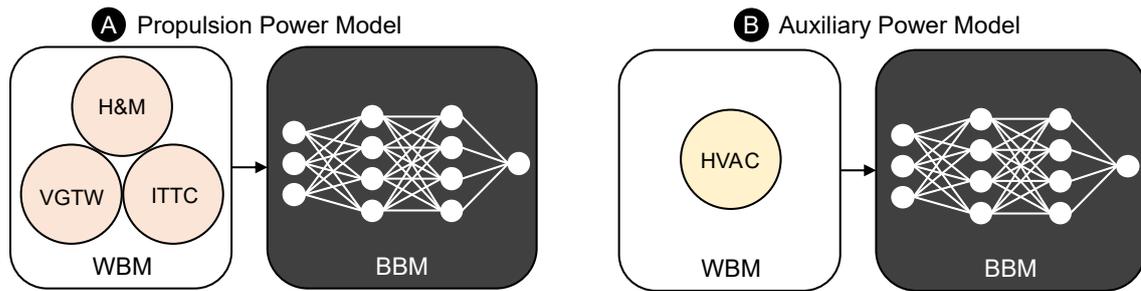
Figure 3.10: ANN-GBM proposed solutions for propulsion (left) and auxiliary power estimation (right)

A WBM considering the HVAC consumer demand will be developed as no current model exists. This model will be based on the work of Stapersma and Klein Houd [98, 99]. Since HVAC systems are generally the largest consumer influenced by dynamic environmental conditions, such a model will be exceedingly essential to capture the physical properties within the GBM (section 3.2.3). While the HVAC is only a portion of the total auxiliary load, it is expected that the BBM can sufficiently scale the additional power system effects as documented by relevant literature (table 3.1). An additional benefit to having multiple WBMs can lead to increased insight regarding model accuracy and model dynamics. This internal study can provide insight into the modular capabilities and effects of the GBM approach on prediction precision and modelling uncertainty.

Due to the successful use within the maritime industry and the general ease of implementation, the ANN algorithm will be used as the BBM. This method scales well with large datasets; thus, ANNs can account for growing amounts relatively easily. Additionally, uncertainty techniques such as boot-strapping (section 3.3.1) are widespread and have been incorporated within other industries numerous times to evaluate modelling uncertainty using well-defined procedures. Ultimately, two independent GBM models will be developed and aggregated to assess individual energy proportioning and total demand. This solution will allow for a robust method that can incorporate physical and observed data phenomena to improve the predictions of both propulsive and auxiliary power under dynamic environmental loading while not restricting the integration of external methods.

3.7. Critical Reflection of Potential Risks and Mitigation Strategies

While the ANN – GBM provides a solution that neatly fits all defined method requirements, some potential risks to the solution's success exist. The first and most transparent risk is related to the quality of the datasets. Sensor noise, signal rattle, signal dropping, and time-lagging are intrinsic sources of uncertainty that may skew and provide undesirable physical mappings within the model, (Rodríguez and Servigne [90]). Therefore, a robust data pre-processing framework is a straightforward mitigation strategy to eliminate any outlier points. However, while such an approach can add tremendous value to the overall modelling performance, the model can only be as accurate as the collected data itself. As such, sensor information must be monitored and validated to ensure the operating results are reliable.

Potential secondary risk can be related to the quantity of collected data. While it is expected that the GBM requires a lesser data amount than the BBM, a sufficiently broad spectrum is still needed, (Pedersen and Larsen [86]). The narrower the data width, the less reliable the modelling space when considering unseen data results. This, for example, can influence the model's ability to extrapolate or interpolate to new data solutions when investigating multiple vessel sizes, fouling contributions, or varying speed consumption. Mitigation of this risk can be achieved through increased data selection and internal parameter tuning for prediction generalization. Unfortunately, this strategy relies on data collection and time. Alternative methods such as synthetic over-sampling to balance datasets can also be utilized but may provide bias to prediction estimations, (Du and Swamy [34]).

A third recognized risk is directly presented due to the WBMs. Since many maritime specific WBMs include a portion of empirical data, the models may inherently lean towards outdated solutions. Ulti-mately this implies that future vessel estimations will be sources of current data trends and not tech-nological advancements. One approach to mitigate such a risk is to eliminate outdated data sources

and train the models on data that truly reflects the current yachting industry's state regarding energy efficiency. Additionally, the implementation of modern WBMs must also be considered as they also generally reflect the current state of modern design solutions.

A final source of error can also be related to the WBM accuracy and completeness. The more precise the solution to the target, the more closely the internal modelling parameters adapt to the dynamic physical processes. However, if the WBM does not fully reflect the target outcomes, extrapolation capacity and estimation accuracy may be at risk. To mitigate this risk, WBMs must accurately reflect the most dynamic variation of the target prediction. For example, as detailed in section 3.2.3 the HVAC demand provides the most significant and dynamic contribution to the total auxiliary consumption. Whereas the other portions remain relatively constant for each operational condition, as enforced in figure (1.5). In such situations, the WBM still aids in the dynamic mapping, whereas the BBM can account for the constant portions with relative ease, (Zwart [108]). A further example can exist due to uncertainty within the annual fouling increase contribution outlined in section 3.2.2. The applicability of yachts is questionable as the study is based on service vessels that are rarely anchored. Therefore, this factor may present an inaccurate bias within the GBM. Nonetheless, while it may not be the most accurate estimation, the dynamic increase plays a critical role in the GBMs mapping the fouling contribution.

3.8. Scope of Work
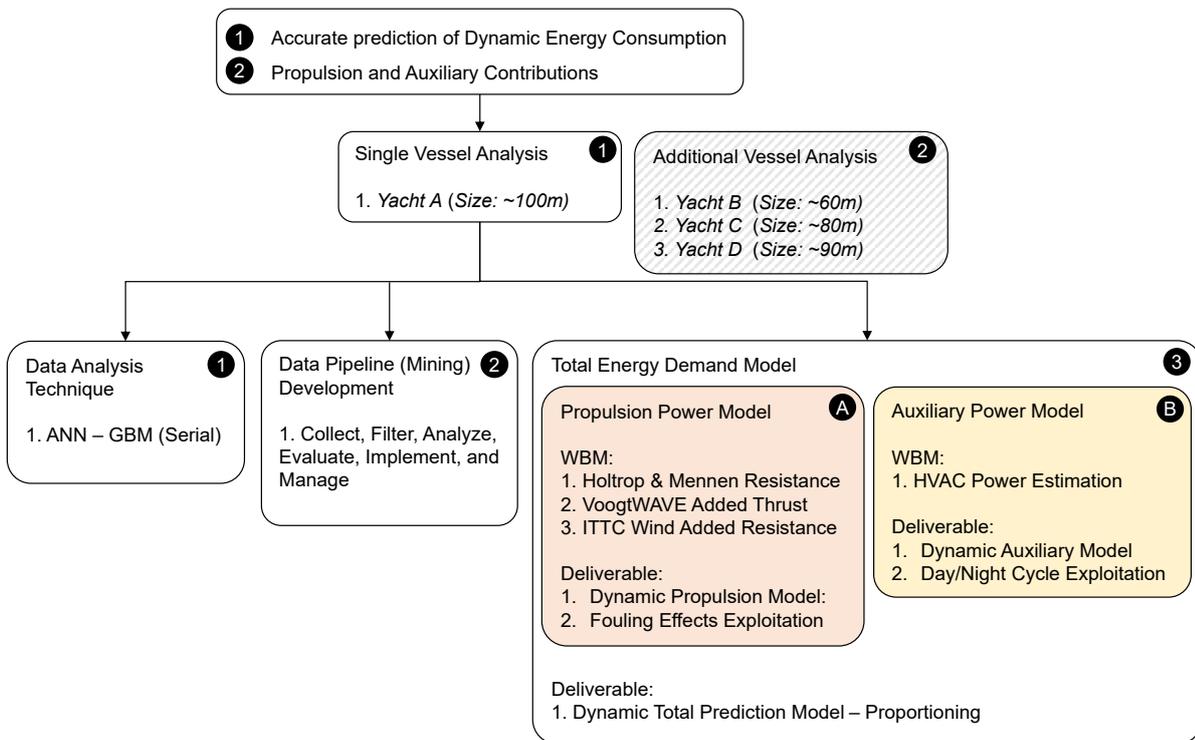
Figure 3.11: General project scope overview

Dynamic energy analysis is a broad field with many avenues available to evaluate consumption predictions. As such, having established the method requirements and solution approaches, it is crucial to define the investigation's focus. Scope refinement ensures the main objective and associated conditions are prioritized within the allotted time constraints. The ultimate goal is for a fully robust study; however, the focus on estimation accuracy is essential. As such, the investigation's quality shall not be compromised in favour of quantity within the time frame. The general scope outline can be seen presented in figure 3.11.

The research's primary focus is the accurate prediction of dynamic energy consumption for propul-

sion and auxiliary power demand. Therefore, these objectives take priority. Due to time and data availability limitations, the investigation will first focus on a single vessel. However, if the results show a high degree of promise and adequate time is available, other vessels will be included within the scope. Three main components will be covered in the study.

The first is a detailed description of the proposed solution, ANN-GBM. This contribution is essential in understanding the capabilities and overall limitations and assumptions of the solution approach. While many BBM exists, only the ANN algorithm will be considered. As such, the qualitative comparison conducted in the literature investigation is the driving measure. The building of WBM is not a focus but instead incorporating them within a GBM framework. Therefore, all models will be based on available methods within *De Voogt*. While CFD and simulations offer much better estimations, only commonly applied semi-empirical models will be implemented due to the relative ease of integration and quick computational times. For the auxiliary power component, no such WBMs exist. As such, a low-fidelity heat load balance method will be developed based on commonly applied methods.

The second scope element involves the development of a detailed process pipeline. The study should be completely reproducible. As such, clarity and transparency in all aspects, including data collection, preparation, management, and implementation, must be provided in addition to the modelling procedures. This contribution will allow for a clear generalized framework that can be followed for any future grey-box implementation.

The final scope component will focus directly on the modelling capacity and analysis of the GBM. Three models will be developed, each considering propulsion, auxiliary, and total power demand, respectively. The models will demonstrate the accuracy of predictions under dynamic environment effects, including a 95% modelling confidence band inside and outside the data training ranges. Once all models have been developed, a proportioning comparison between the aggregated sub-models and the total energy model will be conducted. Finally, each sub-model will be individually investigated to deduce both fouling effects on propulsion estimations and the influence of night/day equipment operations on system consumption.

<div style="text-align: right; font-size: 3em;">4</div>
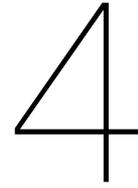
Technical Overview

In chapter 3, a singular solution approach was proposed as the ANN GBM. This modelling approach consists of four white-box models: three for propulsion and one for auxiliary demand, and one machine learning black-box algorithm for each group. This chapter aims to further explore and expand upon the working principles, critical assumptions, and inputs required of each WBM. Additionally, the artificial neural network BBM's general foundations and considerations are also investigated and outlined. As such, this section seeks to answer the following research question,

'How do the proposed technical solutions operate, and which modelling conditions, limitations, and assumptions are necessary for optimal performance?'

Section 4.1 analyzes the separate prediction models, beginning with the Holtrop and Mennen calm-water resistance method. Furthermore, propulsion considerations and efficiencies are addressed to ensure the model parameters reflect the obtained operational data. Next, the ITTC wind resistance and VoogtWAVE added thrust models are detailed and outlined. Finally, the hotel heat balancing approach is further explored along with its critical considerations. Section 4.2 consists of an introduction to artificial neural networks. This section further elaborates on the training processes, hyperparameter considerations, and ensembling methodology. Upon conclusion of the black-box modelling overview, section 4.3 investigates existing literature parameters to evaluate successful and optimal ANN structures. Finally, a small-scale BBM proof-of-concept is developed to confirm the modelling feasibility further.

4.1. White Box Model Overview

4.1.1. Total Calm-water Resistance - Holtrop & Mennen Method

As outlined in section 3.2, the Holtrop and Mennen method provides a quick, practical approach to estimate the calm-water resistance. The technique offers a parametric procedure, which results in a moderate degree of accuracy. The full calculation procedure and all relevant formulations, form coefficients, and empirical regression relations can be seen in appendix B.1.1. Additionally, a complete list of required and optional input parameters can be seen summarized in table B.1.

While the approach allows for broad application, a range of applicability must still be satisfied. Holtrop and Mennen [52] highlighted ranges for prismatic coefficient, length to beam ratio, and Froude number, which are considered in the initial regression analysis. As such, care must be taken to ensure these limits are satisfied; however, reasonable estimates can be expected for cases that fit the

following conditions.

$$Fr \leq 0.45$$
$$0.55 \leq C_p \leq 0.85$$
$$3.9 \leq {}^L/_B \leq 9.5$$

Fortunately, most *Feadships* fall within this range. However, it becomes not applicable when considering alternative hull concepts such as fast yachts or multibody vessels. The general evaluation procedure can be broken into multiple components, which each consider a portion of resistance. When accumulated, the method computes a dimensionalized total resistance, R_T (equation 3.3). Upon evaluation of the total resistance, effective power can be determined as the multiplication of both ship speed and resistance,

$$P_E = R_T \cdot V_s \tag{4.1}$$

The effective towing power (kW) is the required force to drive a ship to a necessary speed while considering the opposing resistance within calm-water conditions. The propulsor's corresponding required thrust power can be determined using associated hull-propeller interaction parameters as outlined in section B.1.2.

$$P_T = T \cdot V_A \tag{4.2}$$

The power delivered by the propeller in water moving at the velocity of advance V_A with useful output T is the thrust power P_T per propeller. Where the thrust – resistance relation can be seen in equation 3.2. The advance velocity is a function of the wake factor. As highlighted by Stapersma and Klein Houd [98], the wake factor represents the difference between the ship's speed and the water's velocity at the propeller location due to the boundary layer influence.

$$V_A = (1 - w) \cdot V_s \tag{4.3}$$

Ultimately, any form of propulsion power can be determined using what is known as the propulsion chain. This power sequence, as presented by Stapersma and Klein Houd [98], links the ship (demand) to the prime movers (supply) through various efficiencies. The detailed propulsion chain can be seen in figure 4.1.
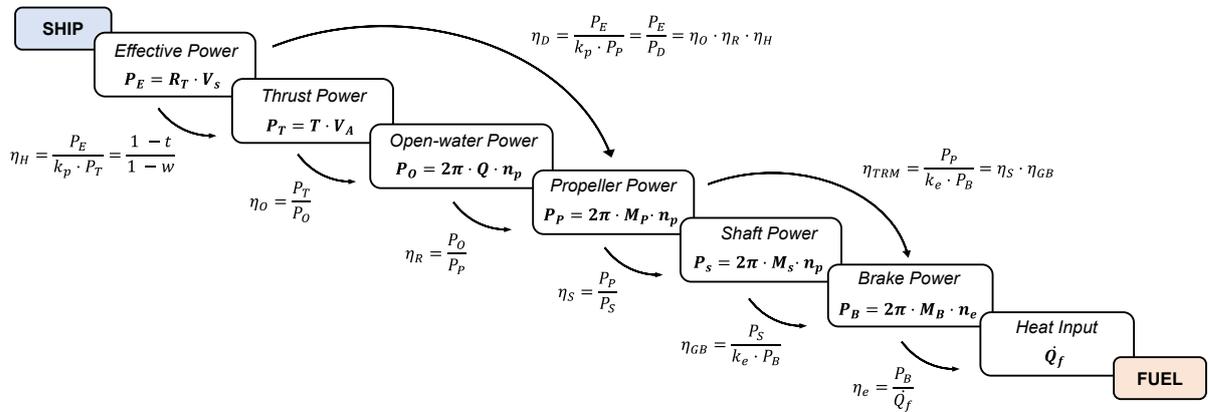
Figure 4.1: Propulsion chain breakdown for efficiencies and corresponding powers, (Stapersma and Klein Houd [98])

Therefore, to determine the power experienced directly by the engine (brake power), the propulsion efficiency (η_D), which consists of the hull efficiency (η_H) open water efficiency (η_O), relative rotative efficiency (η_R), as well as the transmission efficiency (η_{TRM}) must all be known.

$$P_B = \frac{P_E}{\eta_S \eta_{GB} \eta_O \eta_R \eta_H} = \frac{P_E}{\eta_{TRM} \eta_D} = \frac{P_D}{\eta_{TRM}} \tag{4.4}$$

The rotative efficiency can be determined using semi-empirical relations based on the Holtrop and Mennen method [51, 52], equation B.55. However, open-water efficiency is dependent on many factors, including propeller loading and speed. For the sake of the study, this parameter is extracted from MARIN model scale tests. The remaining efficiencies are based on literature sources and discussions with experienced engineers at the cooperating shipyard and displayed in table B.2 in appendix B.1.3.

4.1.2. Wind Added Resistance - ITTC Method

Section 3.2.2 outlines the many wind resistance models available when the corresponding wind force coefficients are not available. However, when the parameters have been determined, the ITTC method of determining wind added resistance is suggested. The full calculation procedure and all relative equations and parameter descriptions can be found in appendix B.1.4. Additionally, table B.3 highlights the required input parameters for the associated WBM.

As outlined by the Seakeeping Commitee of the 29th ITTC [94], the total resistance under the respective wind conditions can be determined by adding the calculated added resistance component, R_{AA}.

$$R_{T,WI} = R_{T,CW} + \Delta R_{AA} \tag{4.5}$$

This resistance component is expressed as the relative change between wind resistance effects and the calm-water air resistance, as highlighted in equation 3.4. The general ITTC wind sign conventions can be seen highlighted in figure 4.2. In this practice, $0°$ on the bow represents pure headwinds, whereas $180°$ represents tailwinds. Additionally, counter-clockwise is considered positively orientated. It should be noted that a great deal of care must be taken when analyzing wind data. In many datasets, wind directions must first be orientated from the global axis to the local ship axis, which can be achieved using the global ship heading. While the conversion is straightforward, understanding which orientation system is used is crucial to accurate modelling evaluations.
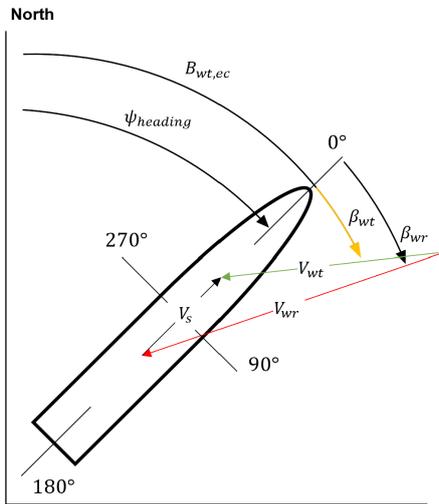
Figure 4.2: Sign conventions for incoming wind directions
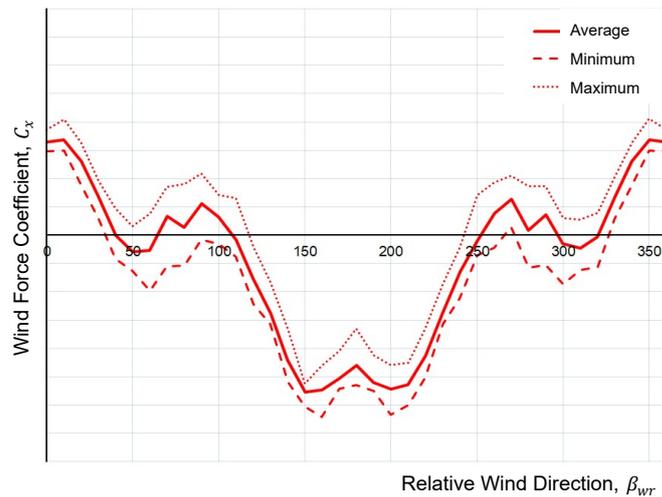
Figure 4.3: Wind force coefficient for incoming relative wind orientation for a typical *Feadship*

As outlined in section 2.3.1, a *Feadship* database consisting of multiple vessel wind coefficients exists. The corresponding coefficient means and respective minimum and maximum deviations can be seen in figure 4.3. While a vast range of vessels was analyzed, the range of applicability must still be considered. The associated collected vessel frontal area, longitudinal area, waterline length spans can be seen below,

$$158\,m^2 \leq A_{(x,v)} \leq 238\,m^2$$
$$625\,m^2 \leq A_{(y,v)} \leq 1065\,m^2$$
$$66.5\,m \leq L_{WL} \leq 100\,m$$

Upon direct inspection, the overall trends of all vessels are similar. Thus, indicating that *Feadships* generally fit a standard shape profile as the associated mean-variance of the coefficients appears marginal. It should be noted that due to confidentiality, actual values cannot be stated. Nonetheless, it is observed that the general magnitudes are respectively small in comparison to other resistance coefficients. Within pure headwinds, the most significant contributing resistance is obtained. However, as the relative heading changes, a quick drop is observed. The region between $105°$ to $250°$ allows for a

negative coefficient contribution as the wind aids the sailing vessel. Within this region, the coefficient's absolute magnitude is nearly double compared to the worst-case scenario.

4.1.3. Waves Added Thrust - VoogtWAVE Method

As further elaborated in section 3.2.2, various models can estimate a ship's added resistance component in waves. Unfortunately, this phenomenon is quite a complicated process, and as such, relies on model scale empirical results. Thus, the VoogtWAVE method is introduced as a *Feadship* orientated semi-empirical methodology. Ultimately, the model presents a relatively accurate and straightforward procedure needing only six input parameters, as highlighted in table B.1.5. The full WBM calculation method can be seen in appendix B.1.5. In addition, parameter influences on the added thrust WBM model can be seen highlighted in figure B.1.5. The added thrust, ΔT_w, can be considered an independent component of the propulsor's total thrust. As such, the total thrust can be determined through means of direct addition.

$$T_{T,WA} = T_T + \Delta T_w \tag{4.6}$$

Within the VoogtWAVE method, thrust deduction factors were integrated to convert wave added resistance to wave added thrust. Therefore, total resistance must likewise be converted before the Voogt-WAVE model can be applied. The approach provides a quadratic transfer function (QTF) to the incoming regular wave height, ζ_a. As such, the added thrust component can be determined from the direct multiplication of the shape function, *taw*.

$$\Delta T_w = taw \cdot \zeta_a^2 \tag{4.7}$$

The general sign conventions differ slightly from that of the typically ITTC wave and wind conventions. These orientations can be seen in figure 4.4.
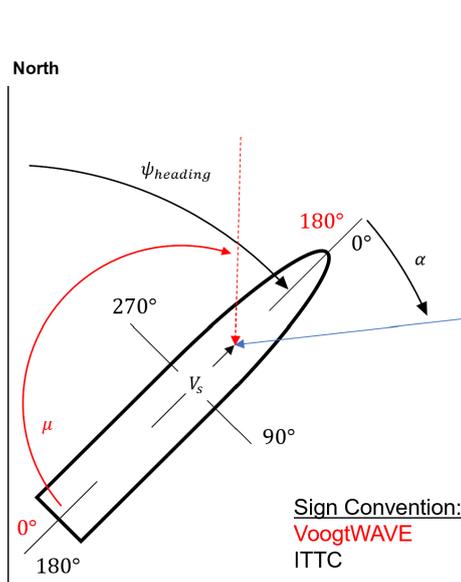
Figure 4.4: Sign conventions for incoming wave directions
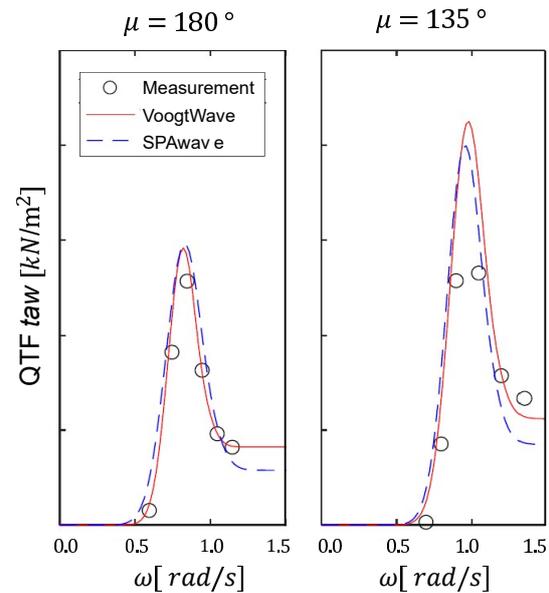
Figure 4.5: Wave Model and experimental comparison for varying incoming wave directions, (Grin [48])

The incoming wave direction, μ, considers the stern, whereas the ITTC incoming direction, α, considers the bow. Since the model applies a mid-plane symmetry wrapping, a direct relationship between the two conventions can be found as,

$$\mu = 180° - \alpha \tag{4.8}$$

As initially outlined in section 2.3.1, the results allow for a robust prediction of added thrust in waves, as it performs equally if not better in all wave directions. As highlighted in figure 4.5, the general trends

fit the experimentally measured data very well for both head and bow-quartering waves. However, beam-quartering waves have provided a wide uncertainty interval range, which, as suggested by Grin [48], requires further study to properly assess the discrepancies.

The modelling methodology is limited to four different *Feadship* bow shapes, split into the categories: flare, reversed, straight, bulbous, where only full load vessel conditions were investigated. Furthermore, the associated applicability range of the total analyzed vessel roll radius of gyration (k_{xx}), pitch radius of gyration (k_{yy}), and waterline length (L_{WL}) spans can be seen below,

$$36.3\% \cdot B \leq k_{xx} \leq 41.4\% \cdot B$$
$$22.9\% \cdot L_{pp} \leq k_{yy} \leq 28.3\% \cdot L_{pp}$$
$$51\,m \leq L_{WL} \leq 108\,m$$

Ultimately, the investigation determined that the roll radius k_{xx} does not influence the results much; however, k_{yy} does. Additionally, the VoogtWAVE approach only considers regular wave conditions directly. Nonetheless, irregular waves can be accounted for indirectly by applying the full-wave spectra and the general shape of the *taw*.

4.1.4. Hotel Heat Gain - HVAC Balance Method

Section 3.2.3 delves into the different considerations of estimating total auxiliary power. As highlighted, these models can be evaluated using a bottom-up approach to ensure estimations are based on first principles instead of experience-driven load factors. Ultimately, the heating, ventilation, and air-conditioning (HVAC) systems are the largest dynamic auxiliary consumers onboard a yacht. These systems must provide comfortable ambient surroundings under various external environmental factors while adhering to regulatory requirements. The full HVAC power estimation calculation procedure and all relative equations, considerations, and parameter descriptions can be found in appendix B.1.6. The industry required ([7, 15, 59]) inner temperatures, area heating loads, required air exchanges, and fresh air proportions are defined in table B.6. Additionally, table B.4 highlights the required input parameters for the associated WBM.

Ultimately, the prediction model consists of a single HVAC evaluation for each associated cabin, stateroom, and area. Fan coil units (FCU) are implemented for all spaces to ensure the ambient conditions are met with a high degree of control. Each system's architecture consists of a heater, cooler, and humidifier, as seen in figure 4.6. Air recirculation is used for each area to pre-heat the incoming air using conventional mixing boxes. This device reduces the required power to maintain the set conditions. The amount of air recirculation is based on the maximum fresh air demand (F) as provided by either *Feadship* design specifications (70% F), ISO 7547 [59] regulations, or calculated CO_2 content per person.
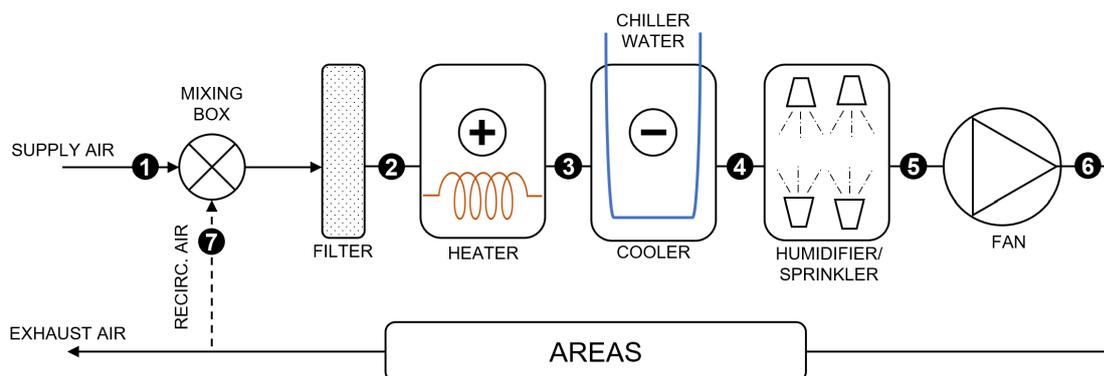
Figure 4.6: Fan coil unit (FCU) schematic for cabin HVAC systems, (Stapersma and Klein Houd [99])

Ultimately a person's overall comfort level is dependent on these factors. Therefore, a region known as the comfort zone is well established (ASHRAE [7] and Stapersma and Klein Houd [99]) to ensure

pleasant conditions are maintained within a degree of equipment design point flexibility. Therefore, the predictive model allows for an achievable relative indoor humidity of 50%, under both the worst-case ambient summer and winter conditions of $35°C$ & $70\%RH$ and $-15°C$ & $30\%RH$, respectively. Additionally, as outlined in the guidance, the maximum inside supply air temperature difference should not exceed $10°C$ when cooling and should not exceed $8°C$ when heat spaces. Tighter interior temperature ranges lead to increased power consumption to maintain the design set-points. Furthermore, the corresponding interior air temperatures and minimum required air exchanges per space are defined in table B.6.

It should also be noted that equipment efficiencies within the HVAC units must also be considered. Electrical heaters, humidification pumping mechanisms, heat exchangers, and fans are all subject to losses. Therefore, general intermediate equipment efficiencies and estimated coefficients of performance for heating, cooling, fan supply, and humidification can be seen collected in table B.7. These estimations are based on previous literature implementations for early-stage design; thus, these metrics are considered constant throughout the prediction model for increased simplicity.

It is acknowledged that the following prediction process does not necessarily provide a one-to-one accuracy estimation of what is occurring in practice. First and foremost, the main areas are individually considered by separate FCU's within the prediction model. This function is different from actual installations, where singular units generally cover multiple areas. Additionally, ducting pressure drops, duct connection losses, smart recirculation, and smart reheating systems are not considered within the prediction model. Furthermore, considerations into personnel and guest movements are neglected, which can significantly impact HVAC unit sizing. Nonetheless, the methodology is based on the practical procedure outlined by Stapersma and Klein Houd [99], where the focus is on the system's demand and not on unit integration. Ultimately, the purpose of the WBM is to provide an estimate of the dynamic variation on the auxiliary load demand; as such, the approach is deemed acceptable for use within the GBM framework to link the dynamic environmental relationships.

4.2. Black Box Model Overview

4.2.1. Basic Principles of Neural Networks

Artificial neural networks are a computing system that is vaguely inspired by the biological networks found within the brain. These systems are composed of multiple individual synaptic components known as perceptrons. The perceptron is the earliest and simplest form of the neural network model (Du and Swamy [34]). The general topology can be seen in figure 4.7 where the general input-output relationships can be described as,

$$\hat{Y}_j = g \cdot \left[\sum_{n_j=1}^{j} \left(\overline{X^T} \cdot \overline{W} \right) - b \right] \tag{4.9}$$

Where the associated j^{th} neuron output $\hat{Y}_j$ is a function of multiple i^{th} feature input parameters, $\overline{X^T} = (x_0, \dots, x_i)^T$. The corresponding weight input from the ith input and for the jth neuron is $\overline{W} = (W_{j,0}, \dots, W_{j,i})$. The b parameter represents the associative layer bias or threshold. The activation function, $g(\cdot)$, is some continuous or discontinuous function that maps the real numbers to an interval between [-1,1] or [0,1]. Therefore, a single neuron computes two functions within the node; the summation of each link's weights and bias and a mapping conversion. Additional details about activation functions can be found in appendix B.2.1.

When multiple perceptrons are placed in connection with one another, a multi-layer perceptron (MLP) network is formed, commonly known as a Feed-forward neural network (FFNN). As detailed by da Silva et al. [29], these networks feature at least one intermediate (hidden) layer, which is placed between the input and output layers. The general architecture of an FFNN can be seen in figure 4.8. As further outlined by da Silva et al. [29], these networks are known for a wide range of applications such as pattern recognition and function approximations. Ultimately, the popularity of such methods stems from the FFNN's universal approximation capability. Du and Swamy [34] enforces that the Universal Approximation Theorem is a mathematically proven phenomenon where, no matter what function exists, a network can approximately approach the results for any number of inputs and outputs. Unfortunately,
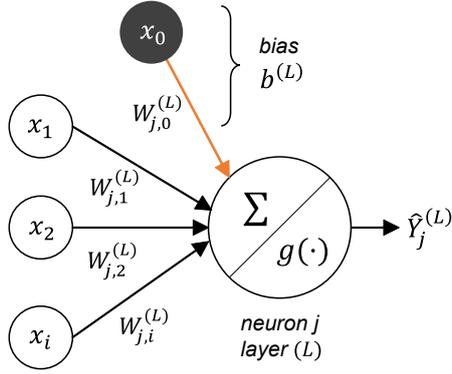
Figure 4.7: A single perceptron architecture, (da Silva et al. [29])
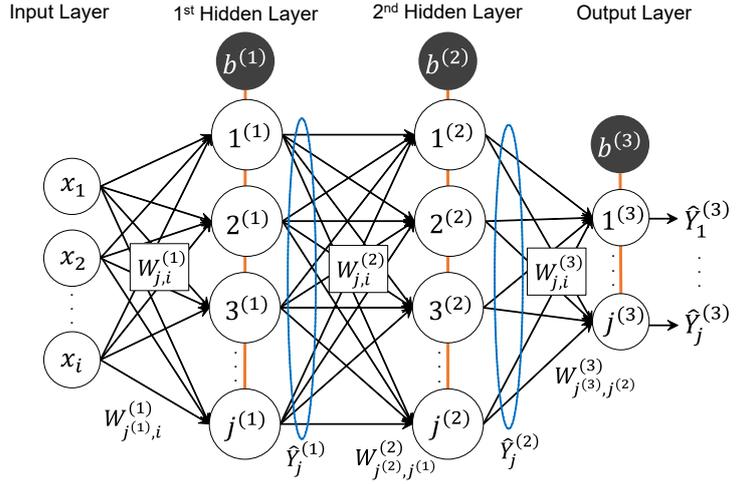
Figure 4.8: A multi-layer (3-layer) perceptron network with weights, biases, and outputs, (da Silva et al. [29])

while the FFNN is very efficient for function approximations, the necessary network architecture for approximating a target function depends only on the target function's basic geometrical shape. As such, optimal network features can vary drastically for each situation and application.

4.2.2. Training Process of Neural Networks

The general training process of an FFNN is accomplished using the backpropagation algorithm. The process is split into three specific stages: forward propagation, loss function calculation, and backward propagation. Ultimately, the training algorithm is well documented (Aggarwal [3], da Silva et al. [29], Du and Swamy [34], Nelles [78], Zwart [108]) and falls within the supervised learning category. This classification means the neural network finds patterns that explain the relationship between the input and target features, where iterations during the learning process aim to minimize the difference between the ground truth targets and the predicted values.

Stage 1: Forward Propagation   In this phase, the inputs, x_i, for a training dataset, D_{TR}, are fed into the neural network. Each layer is sequentially considered resulting in a forward cascade of computations, using the current set of weights and bias, to arrive at an initial prediction, $\hat{Y}_j$. As further detailed by Saleh [93], the calculations performed within each neuron include a linear function that multiplies the input data by some weight plus a bias, which is then passed through an activation function to allow for non-linearity and complex mappings. The mathematical forward propagation process, based on the topology as seen in figure 4.8, can be defined and decomposed as,

$$\hat{Y}_j^{(3)} = g^{(3)} \left[\Sigma(\hat{Y}_j^{(2)} \cdot W_{j^{(3)},j^{(2)}}^{(3)}) - b^{(3)} \right] \tag{4.10}$$

$$\hat{Y}_j^{(3)} = g^{(3)} \left[\Sigma(g^{(2)} \left[\Sigma(\hat{Y}_j^{(1)} \cdot W_{j^{(2)},j^{(1)}}^{(2)}) - b^{(2)} \right] \cdot W_{j^{(3)},j^{(2)}}^{(3)}) - b^{(3)} \right] \tag{4.11}$$

$$\hat{Y}_j^{(3)} = g^{(3)} \left[\Sigma(g^{(2)} \left[\Sigma(g^{(1)} \left[\Sigma(x_i \cdot W_{j^{(1)},i}^{(1)}) - b^{(1)} \right] \cdot W_{j^{(2)},j^{(1)}}^{(2)}) - b^{(2)} \right] \cdot W_{j^{(3)},j^{(2)}}^{(3)}) - b^{(3)} \right] \tag{4.12}$$

It should be noted that synaptic weights and bias are randomly initialized within a neural network and remain unmodified during the execution of this stage. Additionally, the network bias's primary purpose is to help shift each neuron activation function to avoid zero values, hindering the training process.

Stage 2: Loss Function and Performance Evaluation   As detailed by da Silva et al. [29], the next stage consists of defining a function representing the approximation error, whose purpose is to measure the deviation of the network's output responses ($\hat{Y}_j$) to the corresponding desired target values (Y_j).

Thus, considering the $x_{TR,i}$ training sample for the topology illustrated in figure 4.8, the mean squared error (MSE) function is employed to measure the local performance associated with the results by the output neurons for the given sample.

$$E_{MSE}(x_{TR,i}) = \frac{1}{N} \Sigma_{x_{TR}=1}^{N} \left(\Sigma_{n_j=1}^{j^{(3)}} \left(Y_j - \hat{Y}_j^{(3)}(x_{TR,i}) \right)^2 \right) \tag{4.13}$$

The MSE loss function is standard in regression and function approximation analysis, (da Silva et al. [29], Du and Swamy [34]). However, multiple other loss functions can be used. For instance, the mean average error (MAE) is an alternative within regression analysis. Whereas, when considering classification based problems, binary-based loss functions are implemented instead. Ultimately, each iteration of the training process aims to minimize the loss function by changing the parameters (weights and biases) that are used to perform the calculations during the forward pass.

Stage 3: Backward Propagation   The backward propagation stage is an efficient method of computing the gradients (partial derivatives) of the loss function to each subsequent layer's weights and biases. The approach is a smart implementation of the chain rule of derivatives computations, which gives you the ability to compute all required partial derivatives linearly.

$$\nabla E_{MSE}^{(L)} = \frac{\partial E}{\partial W_{j,i}^{(L)}} = \frac{\partial E}{\partial \hat{Y}_j^{(L)}} \cdot \frac{\partial \hat{Y}_j^{(L)}}{\partial I_j^{(L)}} \cdot \frac{\partial I_j^{(L)}}{\partial W_{j,i}^{(L)}} \tag{4.14}$$

Where each component can be decomposed as,

$$I_j^{(L+1)} = \sum_{i=0} \hat{Y}_j^{(L)} \cdot W_{j,i}^{(L+1)}; \quad \frac{\partial I_j^{(L)}}{\partial W_{j,i}^{(L)}} = \hat{Y}_j^{(L)}; \quad \frac{\partial \hat{Y}_j^{(L)}}{\partial I_j^{(L)}} = g'^{(L)} \cdot (I_j^{(L)}); \quad \frac{\partial E}{\partial \hat{Y}_j^{(L)}} = \sum_{j=1} \frac{\partial E}{\partial I_j^{(L+1)}} \cdot \frac{\partial I_j^{(L+1)}}{\partial \hat{Y}_j^{(L)}} \tag{4.15}$$

Upon determining the gradients, adjusting the output and intermediate layers' synaptic weights can be completed sequentially. The full backward propagation calculation procedure and component descriptions can be found in appendix B.2.2 as outlined by da Silva et al. [29]. While not always explicitly stated, the backpropagation technique only determines the associated gradients of each layer. An optimization search algorithm must then be employed to reduce the loss function in the next iteration step by changing the associated weights and biases. The optimization algorithm's final objective is to find the global minimum where the loss function has reached the least possible value.

Stochastic Gradient Descent (SGD) is a first-order optimization (meaning the algorithm focuses on first-derivatives) method commonly used to reduce the loss functions. This optimizer is often applied together with backpropagation to make efficient updates in terms of neural networks. As such, backward propagation is not limited to singular optimizers; instead, efficient alternatives can be used in conjunction with the backward propagation routine for enhancing training operations. Other commonly applied 1st-order optimizers are SGD with Momentum, Root Mean Squared Propagation (RMSprop), and Adaptive Moment Estimation (Adam). According to Kingma and Ba [65], the adam optimizer is considered a robust and modern solution since the approach is computationally efficient, little memory requirement, invariant to diagonal re-scaling of gradients, and is well suited for problems that are extensive in terms of data/parameters. It should be noted that while 1st-order methods are the norm, alternative optimization techniques such as 2nd-order Gauss-Newton Methods, Levenberg-Marquardt, Scaled Conjugate Descent, Bayesian Optimization, and Evolutionary Algorithms has also been successfully implemented (da Silva et al. [29], Nelles [78]).

4.2.3. Hyperparameter Tuning for Generalization

One of the most relevant features concerning an artificial neural network is its ability to generalize the acquired knowledge, enabling the estimation of solutions using inputs that have never been encountered before. However, to achieve this state, appropriate model hyperparameters (factors specific to the network architecture) must be selected to avoid the practical problems of over and under-fitting.
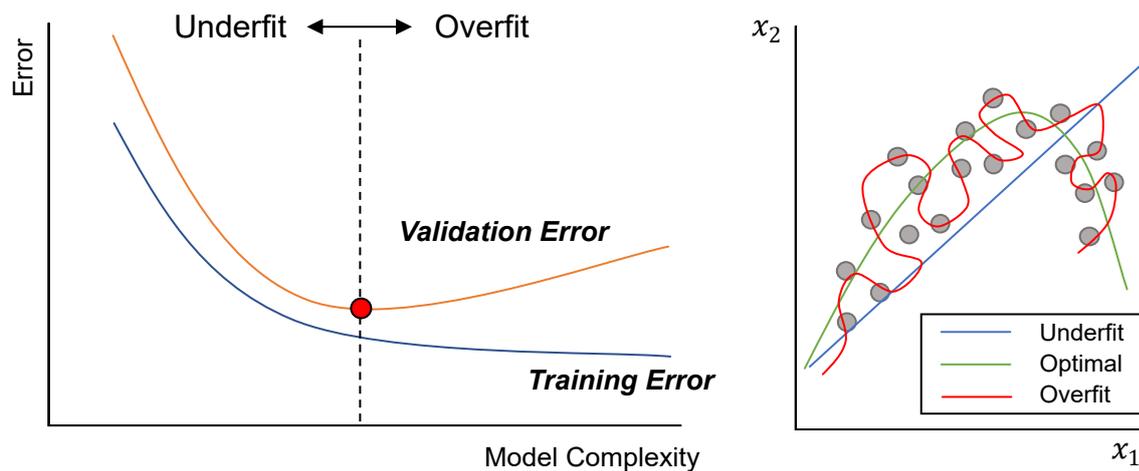
Figure 4.9: Model error comparison (left) and corresponding fitting relations (right), (based on da Silva et al. [29])

As outlined by Aggarwal [3], the problem of overfitting refers to the situation where fitting a model to a training data set does not guarantee good prediction performance on unseen test data. In other words, there is always a gap between the training and validation data performance. This situation can be highlighted in the left of figure 4.9, where the deviations in training and validation sets represent the onset of over-fitting. The opposite phenomenon for under-fitting can be seen as well. Here, the error deviations are slight, but the model's global error has not reached the minimum position. The optimal point, where a model can be considered generalized, can be visualized in the right of figure 4.9.

The most well-known approach to determine the optimal hyperparameters is known as grid searching. All combinations of selected values of the hyperparameters are tested to determine the optimal choice. However, as suggested by Aggarwal [3], one issue with the approach is that the number of points in the grid increases exponentially with the number of hyperparameters. Due to these limitations, random gird searches are sometimes employed instead. The critical parameters that influence these phenomena are the model topology, early-stopping criteria, dropout regularization, and modelling ensembling.

Optimal Neural Network Topology   As enforced by da Silva et al. [29], the topology of a neural network depends on many aspects such as the adopted learning algorithm, how the weight matrices were initialized, the complexity of the problem, the spatial distribution of the training samples, and the quality of the training set available. As such, the most suitable configuration is usually based on empirical means.

The most commonly applied empirical method is known as random sub-sampling cross-validation. This approach initially divides the entire data set into a training and validation set, where then each candidate topology is tested incrementally. This separating must be repeated multiple times during each topology learning process to provide different samples in both subsets. The global performance of each candidate topology is then obtained from the average individual performance. As suggested by da Silva et al. [29], 60% to 90% of the entire dataset, D, are selected at random for the training subset, D_{TR}. Both Zwart [108] and Parkes et al. [83] stresses that the general accuracy is mainly determined by the number of hidden layers and neurons. As the number of units increases, more complex relations can be modelled by a network; however, having too many can overfit the dataset and significantly reduce the model generalization characteristics.

Early-Stopping and Dropout Regularization   Generalization qualities can also be improved by incorporating early-stopping and dropout regularization techniques within the learning processes.

Early-stopping is a simple procedure in which the optimization process is ended after only a few iterations to avoid overfitting, (Aggarwal [3], da Silva et al. [29], Zwart [108]). During the procedure, the learning process is continuously monitored, where the performance is evaluated after each repetition.

The training process is stopped when the mean squared error increases between successive epochs (all subset data is presented to the model). This process requires separating the dataset into three subsets; training, validation (used for early-stop), and testing. In summary, early stopping essentially constrains the optimization process by restricting the number of steps within the optimization process to reduce both over-fitting and, in effect, training time. Early stopping is generally incorporated by either setting the number of successive increases of error (validation failure), setting the maximum number of epochs, or specifying a computation time limit.

Dropout is a form of model regularization, meaning the technique discourages learning a more complex or flexible model to prevent overfitting. The dropout technique randomly drops units (along with their connections) from the neural network during training batch iteration. This probabilistic process forces the network to become robust against these random disruptions and prevents the risk of over-fitting to the supplied training set. As detailed by Du and Swamy [34], dropout can be interpreted to regularize a neural network by adding noise to its hidden units for improved model robustness to unseen data. Additionally, both Du and Swamy [34] and Nelles [78] found that dropping out 20% of the input units and 50% of the hidden units were often optimal.

4.2.4. Data Scaling for Improved Performance

Saleh [93] emphasizes that it is essential to re-scale the input and output data variables to achieve higher modelling accuracy. This concept roots from the fact that having different data scales for each feature may skew the importance of higher valued features. However, if all the components are equally scaled, the model can give higher weights to the more significant features towards the target feature. Furthermore, by removing the model's need to learn from the data invariance, training times can decrease. Currently, two primary re-scaling methodologies are applied: Normalization and Standardization.

Normalization consists of re-scaling the feature value, x_i so that all the features' values are between [0,1] or, using Thales' conversion, [-1,1] as seen below,

$$x_{norm} = 2 \cdot \left(\frac{x_i - x_{min}}{x_{max} - x_{min}} \right) - 1 \tag{4.16}$$

da Silva et al. [29] further details that all input and output variables need to be individually normalized to their maximum and minimum values, considering all the available data. Additionally, the bounding values should be included within the training set. Otherwise, the data extents are reduced as the values are contained within the out-of-sample testing sets. In contrast, standardization converts the values to a mean of 0 and standard deviations equal to 1, as seen below,

$$x_{std} = \frac{x_i - \bar{x}}{\sigma} \tag{4.17}$$

Where $\bar{x}$ is the sample set mean, and σ is the sample set standard deviation. Between the two data scaling methodologies, no rule exists between selecting one over another, (Saleh [93]). However, it should be noted that it is essential that they are used individually and not in conjunction with one another.

4.2.5. Model Ensembling for Uncertainty Assessment

As detailed by Mazloumi et al. [75], it is common to consider two sources for the uncertainty associated with neural networks: epistemic uncertainty and aleatoric uncertainty.

Epistemic uncertainty results from unknowns within the model structure, such as topology, limited data, hyperparameter selection, and convergence uncertainty (measured by the variance and denoted by σ_m^2). Aleatoric uncertainty is the uncertainty arising from the natural stochasticity of observations. This inherent uncertainty (measured by the variance and denoted by σ_e^2) may occur from intrinsic noise in randomly selecting a training dataset from a population or measurement errors; thus, it cannot be reduced even when more data is presented.

One commonly applied method to evaluate these components is known as Bootstrapping. The technique can quantify the model (σ_m^2) and inherent (σ_e^2) uncertainty by considering an ensemble of ANNs built on different datasets sampled with replacement. The ANN bootstrapping procedure, as detailed by Ferrario et al. [41], to identify the modelling and variance uncertainty intervals can be seen in figure 4.10, and detailed as the following,

1. Divide the entire available dataset (D) of x_i input/output parameters into training, validation, and test datasets, as D_{TR}, D_{VAL}, and D_{TEST}, respectively.

2. Generate B bootstrap training samples datasets ($D_{TR,b}$, $b = 1, ..., B$) by sampling with replacement from the original training dataset D_{TR}. Each set $D_{TR,b}$ is composed of the same number of sample points, N_{TR} of the original training dataset. However, due to the sampling with replacement, some of the D_{TR} input/output patterns will appear more than once in the $D_{TR,b}$ whereas some will not. As detailed by Efron and Tibshirani [35], for estimating a probability distribution, the number B will ordinarily be in the range of 25 to 200.

3. Build the bootstrapped ANN models ($f_b(x_{TR,i})$, $b = 1, ..., B$), based on the datasets $D_{TR,b}$ ($b = 1, ..., B$) generated at the previous *step 2*, and the validation set D_{VAL}. The training sets are used to train and tune individual models; whereas, the validation sets are used to monitor network performance accuracy.

4. Use the ANN models of *step 3* to compute the target estimates ($\hat{Y}_{boot,b}(x_{TEST,i})$, $b = 1, ..., B$), on the new data set D_{TEST}. This allows for the development of N_{TEST} bootstrap-based empirical probability distribution functions (PDF) for the quantity of the target points. In correspondence of a new input, x_{TEST}, the bootstrap mean estimate, $\overline{Y}$ is given by the average of the B functions.

$$\overline{Y}_{boot}(x_{TEST,i}) = \frac{\sum_{b=1}^{B} \hat{Y}_{boot,b}(x_{TEST,i})}{B} \tag{4.18}$$

And the bootstrap estimate of the modelling variance, σ_m^2, can be determined by,

$$\sigma_m^2 = \frac{\sum_{b=1}^{B}(\hat{Y}_{boot,b}(x_{TEST,i}) - \overline{Y}_{boot}(x_{TEST,i}))^2}{B - 1}, \quad x_i(n) \in \{1, ..., N_{TEST}\} \tag{4.19}$$

The process can be similarly repeated to determine the inherent uncertainty. However, in this instance, the residual error between the target and bootstrap estimate is considered as,

$$\hat{E}_{boot,b}(x_{TR,i}) = Y_j - \hat{Y}_{boot,b}(x_{TR,i}) \tag{4.20}$$

Therefore, the error spectrum's mean and variance can be likewise evaluated as $\overline{E}$ and σ_e^2, respectively. It should be noted that since prediction intervals rely on target data to evaluate residual errors, prediction intervals can only be evaluated during the training processes, where ground-truth values and estimated bootstrap targets are both known. This implies that for out-of-sample data, only the modelling uncertainty can be evaluated successfully. Nonetheless, the worst-likelihood uncertainty can be extrapolated from the prediction uncertainty parameter.

5. Assuming that the target values follow a normal distribution based on the bootstrapping sampling principles (Ferrario et al. [41], Mazloumi et al. [75]), the confidence intervals (with confidence level, $100 \cdot (1 - \gamma)\% = 95\%$) of the total uncertainty intervals can be constructed using,

$$\overline{Y} \pm Z \cdot \sigma_t, \quad \text{where } Z = 1.96 \text{ (95\%)} \tag{4.21}$$

Where both σ_m^2 and σ_e^2 have been quantified, the total prediction uncertainty can be obtained from general propagation principles as $\sigma_t^2 = \sigma_m^2 + \sigma_e^2$. This means for each input vector x_i, a range where the mean of a dependent population occurs within a 95% probability can be identified. Once identified, the confidence interval of the mean bootstrap prediction can be expressed as, $\pm[\underline{\delta}, \overline{\delta}]$. While the sampled features generally form a normal distribution (Mazloumi et al. [75]), the risk of slight underestimation of the confidence range can exist otherwise.
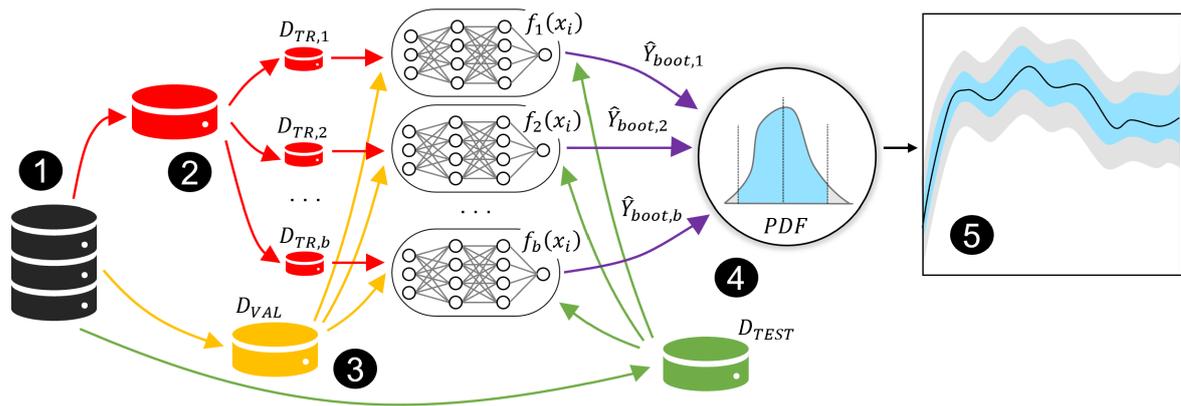
Figure 4.10: General construction of bootstrap-empirical PDF for uncertainty estimations, (based on Ferrario et al. [41])

4.3. Literature Comparisons for Optimal Network Structures

As outlined in section 4.2.3, the optimal artificial neural network structure is generally found via empirical means or grid search evaluations. As such, a comparison of similar network architectures when evaluating both propulsion and auxiliary demand can be conducted to find an initial search reference (based on a similar comparison by Zwart [108]). The results for both propulsion and auxiliary power demand can be seen in tables 4.1 and 4.2, respectively. While many references exist, the listed literature conforms most closely to the associative thesis study, where all relevant parameters were clearly outlined and detailed. Unfortunately, there is currently no literature for yachts when evaluating ANNs for both propulsion and auxiliary power. Nonetheless, the extracted sources have a high degree of transparency, allowing for a clear ANN modelling basis. Therefore, upon review of the literature, the following modelling considerations can be concluded as follows,

The number of input features: The maximum number of input features used in both conditions, propulsion and auxiliary, is thirteen (Zwart [108]). It can be seen that for propulsion, the average feature count (~ 9) is larger as compared to the auxiliary portion (~ 5). While a higher feature count does not necessarily mean the best modelling accuracy, it does provide a good indication for more modelling precision. While some features are related to equipment specific measurements, most of the features are due to environmental forces. Intuitively, this makes sense as the dynamic variations greatly influence each component's energy consumption (chapter 3). As such, these features must be accurately captured to ensure the model can reflect such effects. The full feature variables for each study are additionally outlined in tables 4.1 and 4.2.

The number of data samples: Filtered data samples seemingly vary between 223 to 1988 for propulsion demand. These samples have either been based on noon reports or sensor data measurements. Additionally, continuous onboard measurements have also been incorporated. These sets generally consist of extensive dataset samples. However, as seen in the tables, the number of samples' influence on accuracy is not as substantial as the data quality. Typically the reliability of noon-reported data is lower than that of data sensors as the sampling frequency is much lower (24h). However, the technical reliability of onboard sensors and associated data noise and signal losses may also hinder the accurate representation of the actual operating conditions. For auxiliary power estimations within the building's field, data samples range between 286 to 4402 data sample points, consisting of both 3-minute and 10-minute sampling rates. Ultimately, collecting more data is the fastest way to improve modelling accuracy; however, collecting high-quality data allows for the most extensive prediction improvements.

The training percent split: The data training range falls between 70% and 85%, with an average of 75% for both the propulsion and auxiliary demand components. These ranges closely align with the ANN literature suggestions of Aggarwal [3], da Silva et al. [29], Du and Swamy [34], Nelles [78]. The remainder of the datasets is generally equally split between a validation set to apply early-stopping and dropout regularization techniques (section 4.2.3) and an independent test set to verify the modelling results.

The neural network topology: The results show that the number of hidden layers is generally limited to 1 for both propulsion and auxiliary demand. Most investigated studies performed a systematic empirical trial-and-error method to evaluate the layer effects on modelling accuracy. Layers are incrementally considered for a varying range of neurons per hidden layer and then directly compared on an accuracy basis. Similarly, the number of neurons per layer is determined via a trial-and-error methodology to evaluate the optimal topology. As described in section 3, the more neurons and layers, the more considerable the risk of overfitting the modelling. As such, empirical approaches are generally coupled with validation datasets to aid in model generalization. Based on the review, neurons typically fall within a similar average range of 24 and 19, respectively.

The optimization routine: From the investigation, it can be seen that a broad range of optimizers has been applied in the previous studies. However, based on the results, no discernable advantages from one to another have been noticed. Nonetheless, many studies suggest the comparison of the multiple optimizers to develop an optimal ANN model.

Table 4.1: Propulsion ANN Literature Summary

ANN Parameters		Propulsion				
Reference	Average	Zwart [108]	Parkes et al. [83]	Bal Beşikçi et al. [11]	Pedersen and Larsen [84]
Input Features	9	13	6	9	9
Data Amount	816 (exluding CM[1])	1988	45983 (CM[1])	223	238
Training Data (%)	73%	70%	70%	70%	80%
Hidden Layers	2	1	3	1	1
Neurons	24	15	50	12	20
Optimizer	Varies	Bayesian	Conjugate Gradient	Levenberg-Marquardt	Bayesian
Error	5.78%	6.63%	7.80%	6.00%	2.70%
		Input Features				
Vessel	Ship Speed	✓	✓	✓	✓
	Trim	✓	✓	✓	✓
	Mean Draft		✓	✓	✓
	Load (DWT)	✓		✓	
Environmental	Sea (Wind Wave) Height	✓	✓	✓	✓
	Sea (Wind Wave) Direction	✓		✓	✓
	Swell Height	✓			
	Swell Direction	✓			
	Wind Speed	✓	✓	✓	✓
	Wind Direction	✓	✓	✓	✓
	Sea Temperature	✓	✓		✓
	Air Temperature				✓
Engine	Engine Speed			✓	
Fouling	Days Hull Clean	✓			
	Days Propeller Clean	✓			
White Box	Propulsion Power	✓			
Output Target	Shaft Power (kW)	✓	✓		✓
	Fuel Consumption (MT/h)			✓	

Table 4.2: Auxiliary ANN Literature Summary

ANN Parameters		Auxiliary - Buildings Engineering			
Reference	Average	Karatasou et al. [64]	Neto and Fiorelli [79]	Kalogirou and Bojic [63]
Input Features	5	6	5	5
Data Amount	1715	4402	286	456
Training Data (%)	77%	70%	75%	85%
Hidden Layers	1	1	1	1
Neurons	19	12	21	23
Optimizer	Varies	Levenberg-Marquardt	Stochastic Gradient Descent	Stochastic Gradient Descent
Error	9.00%	1.50%	16.50%	9.00%
		Input Features		
Environmental	Air Temperature	✓	✓	✓ (Seasons)
	Relative Humidity	✓	✓	
	Global Solar Radiation	✓	✓	
	Diffuse Solar Radiation		✓	
	Wind Speed	✓		
Building Properties	Insulation Type			✓
	Wall Thickness			✓
	Heat Transfer Function			✓
Time Classification	Time of Day	✓		✓
	Day-Type	✓	✓	
Output Target	Energy Demand (kWh)	✓	✓	✓

[1] CM - Continuously monitored datasets

4.4. Small-scale Black Box Model

A small-scale model is developed to confirm the feasibility further and enhance the confidence of black-box modelling capabilities. The proof-of-concept focuses on estimating the maximum installed main engine power as a function of vessel speed and displacement. The target relation can be seen in equation 4.22,

$$P_{ME} = \frac{\Delta^{(2/3)} \cdot V_s^3}{C_{ADM}} + F_{Noise} \cdot Random(-1 \dots 1) \tag{4.22}$$

This relation is typically known as the admiralty equation, where P_{ME} is the main engine power, Δ is the displacement of the vessel, V_s is the ship's speed, and C_{ADM} is the admiralty constant based on the selected vessels. The admiralty coefficient's determination is based on investigating 47 Feadship's and can be seen outlined in appendix C.1.1. In addition to the admiralty relation, a random noise term, F_{Noise}, is included to introduce non-linearity within the estimation. A total of 100 random data points are used to evaluate the function with limiting ranges of,

$$200\,t \leq \Delta \leq 3600\,t \quad \& \quad 5\,knots \leq V_s \leq 18\,knots$$

The neural network structure can be seen highlighted in figure 4.11. A conventional feed-forward neural network structure is applied with two inputs, five neurons using one hidden layer, and a single output. The network is decomposed into a 70%-15%-15% training, validation, and testing split scheme, which can be further visualized in appendix C.1.2. Additionally, the corresponding network is trained using default *Python Keras* modelling parameters (Chollet [24]),

1. Training Split: 70% (Validation: 15% & Test: 15%)
2. Max Epochs: 500 & Max Patience: 10
3. Activation = *'tanh'*
4. Optimizer = *'adam'*
5. Initializer = *'uniform'*
6. Dropout Rate = 0.0
7. Layers = 1 & Neurons = 5
8. Number of Boots = 10
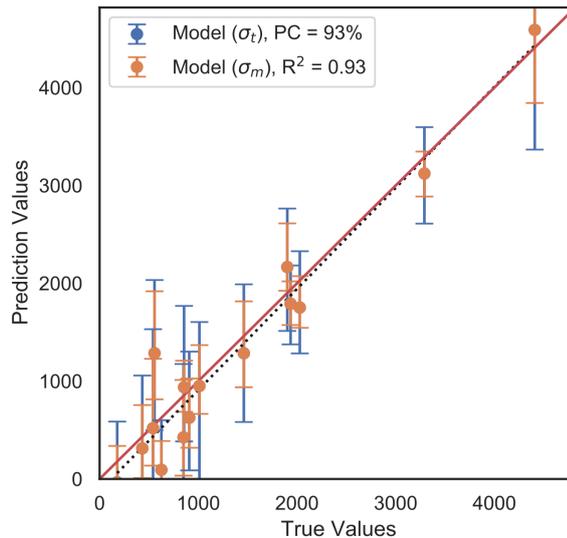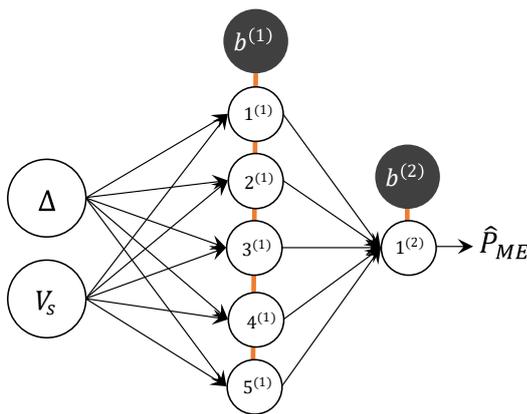
Figure 4.11: General small-scale artificial neural network model structure (In:2, 1HL:5, Out:1)

Figure 4.12: Small-scale artificial neural network target and prediction performance

Table 4.3: Small-scale model results summary

Speed, V_s (knots)	Displacement, Δ (t)	True, P_{ME} (kW)	Prediction, $\hat{P}_{ME}$ (kW)	Error (kW)	Percent Error (%)
6.1	3049	432.6	312.8	119.7	27.7
6.4	907	178.5	-33.5[1]	211.9	118.7
6.7	1336	625.8	93.8	532.0	85.0
6.8	3377	541.4	519.1	22.2	4.1
9.1	3543	556.4	1285.6	-729.2	-131.1
9.5	2343	855.8	937.9	-82.1	-9.6
10.8	1434	1011.2	951.1	60.1	5.9
10.9	3036	1934.9	1790.7	144.3	7.5
11.1	746	908.1	625.5	282.6	31.1
11.2	2697	2028.3	1751.9	276.3	13.6
11.6	331	850.4	425.3	425.1	50.0
12.1	1367	1461.4	1284.9	176.5	12.1
12.2	2523	1901.7	2166.3	-264.5	-13.9
13.1	3178	3290.7	3121.7	169.0	5.1
17.2	2188	4408.8	4589.6	-180.8	-4.1

Performance		
Coefficient of Determination (R^2)	0.926	
Mean Error (kW)	77.5	
Mean Percent Error (%)	13.5	
Average 95% Uncertainty, $\sigma_{t,avg}$ (kW)	Lower: 716.7	Upper: 664.1

It should be noted that the following architecture is not optimized and that an arbitrary structure is selected to highlight the general modelling robustness.

Based on the described network architecture, default settings, and target function, the corresponding results of 15 unseen samples are highlighted in figure 4.12 and further detailed in Table 4.3. It can be seen that the BBM model does an excellent job in estimating the target main engine power, where the statistical fitting relation (R^2) and mean error (ME) are 0.93 and 77.5 kW, respectively. Additional modelling results can be seen in appendix C.1.4. In addition to the general performance criterion, confidence intervals using the previously described methodology (section 4.2.5) can be seen determined for both modelling (σ_m) and total (σ_t) uncertainty. The full development of the empirical probability distributions can be seen highlighted in appendix C.1.3. Ultimately, these intervals allow for further understanding and interpretation of the developed model's capabilities. It can be seen that the last data point has a relatively large 95% confidence bound. This is most probably since the sampling point is nearing the training limits. As such, the modelling estimations have likely not been trained to the degree of the other points, which showcase narrow uncertainty bands.

One interesting point of note is related to the second prediction data point, where the prediction falls into a negative region. This data point thus represents a non-physical estimation as the main engine power can not be below zero. However, when inspecting figure C.1 within appendix C.1.2, it can be seen that the added noise values have allowed for negative inputs. As such, the inherent dependency on the input parameters becomes apparent. Therefore, not only does a model need to be optimized, the data must be prepared such that non-physical outliers are removed from the learning sets to ensure reliable results.

Nonetheless, the corresponding small-scale model evaluation highlights the immense practicality and general working ability in non-linear function approximations. Therefore, the confidence of achieving successful application within the study is further attained in the modelling capabilities.

---

[1]Unrealistic target estimation output

5

Modelling Methodology

Having outlined the technical solution approaches in chapter 4, along with each modelling approach's main assumptions and limitations, further elaboration on essential processes and data preparation steps are detailed and described. Ultimately, this chapter hopes to provide a clear, efficient, and universal methodology to successfully develop and implement data-driven modelling solutions. In addition to modelling methodologies, critical assumptions and decision processes are addressed throughout each step. As such, this chapter will seek to answer the following research question,

'What current data preparation methodologies exist to incorporate raw operational information within the modelling approach?'

First, section 5.1 presents and outlines a general modelling methodology. This section introduces the key steps to efficiently and successfully implement digital modelling solutions. Each primary step is then further decomposed and outlined. Section 5.2 highlights the critical processes within the data preparation stages. This section elaborates on all relevant considerations to convert raw data into cleaned modelling features effectively. Next, section 5.3 outlines the general GBM methodology in more detail. Here, the required and available modelling input parameters are presented. Section 5.4 further expands on the black-box modelling process and the necessary sequence for optimal modelling performance and evaluation. Finally, section 5.4.3 outlines the general verification and validation considerations in the context of data-driven modelling.

5.1. General Modelling Methodology

A universal modelling methodology can be seen highlighted in figure 5.1. As visualized, seven steps are necessary to fully understand, evaluate, and implement an appropriate data-orientated modelling solution. These include; *Problem Definition, Data Collection, Data Preparation, Modelling, Model Evaluation, Deployment, and Performance Monitoring*. This sequential modelling pipeline is extracted and adapted from a multitude of novel literature works, namely, Pedersen and Larsen [85], Petersen et al. [88], and Parkes et al. [83].

The first stage deals with identifying the problem and understanding the inherent difficulties that indicate a data-orientated solution is appropriate. Such an evaluation has been conducted in chapters 2 and 3 of the report. The second step, *Data Collection*, identifies which data sources are required, which sources are available, and when they are available within various design stages (see chapter 2). At which point assemblage of these sets is completed. *Data Preparation* is a critical step that converts raw information into usable modelling inputs, which allow for efficient data-orientated modelling operations. The next step, *Modelling*, implements a technical modelling solution, in this case, a GBM scheme (see chapter 3). After the models are developed, the *Performance Evaluation* step assesses each model using essential metrics to gain a deeper insight into independent modelling functionality. Upon successful modelling assessments, the *Deployment* stage relates to applying the optimal model to the
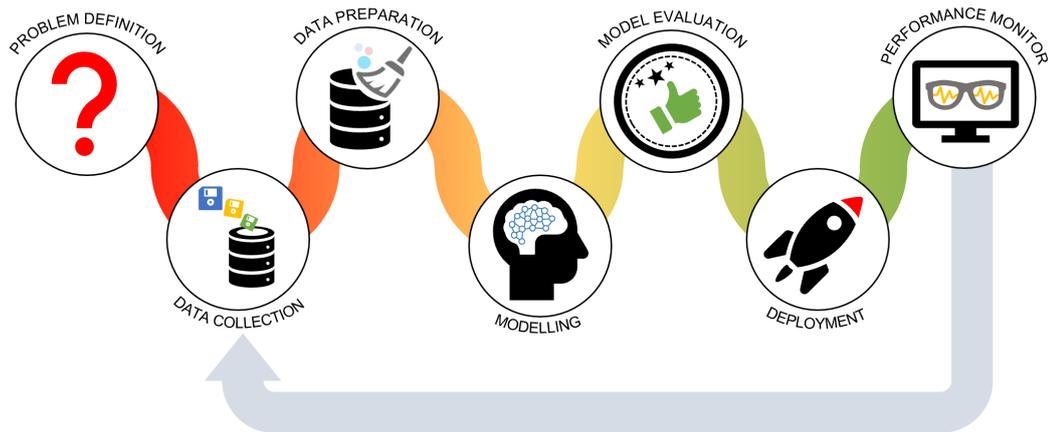
Figure 5.1: General data-orientated modelling methodology

experienced problem. Finally, *Performance Monitoring* is a post modelling process where application and performance are continually monitored. Here, new data information is retained, and modelling solutions are continually updated. This study focuses on the first six stages, as the remaining stage is a continual evolutionary process that requires an appropriate time period of application. As such, the focus is orientated only on the development and analysis of the GBM solution in regards to the fulfillment of the main research objective and associated research questions of section 1.2.

5.2. Data Preparation Methodology

The preparation of raw data is arguably one of the most critical steps in any GBM modelling approach. This influence is due to the inherent data-driven nature of the BBM. If the data quality is flawed and not an accurate reflection of physical results, the corresponding output estimations will be equally poor. As such, an appropriate data preparation methodology is presented to ensure that the highest possible quality of data is retained. Zwart [108] adopted a novel cleaning approach developed by García et al. [45] which is commonly applied within computer science applications. This approach has been further adapted to increase clarity and ease by enhancing the process's sequential nature to apply grey-box modelling. While the general preparation approach has been applied in other studies, alternative techniques and method repeatability are further explored by considering new continuous datasets and auxiliary powering applications. Figure 5.2 highlights the proposed sequential data preparation methodology.

Figure 5.2: Data preparation procedure (based on García et al. [45] and Zwart [108])

The pipeline consists of 8 critical steps: *Data Integration, Data Transformation, Missing Value Imputation, Noise Identification, Data Cleansing, WBM Evaluation, Feature Selection, and Data Scaling*. At the end of the process, the data can be considered *'cleaned'* and ready for the next step within the

GBM general methodology outlined in section 5.3. A detailed breakdown of each preparation step's critical objectives and processes can be viewed in the following sections.

5.2.1. Data Integration

Generally, data originates from multiple external sources with varying lengths and parameter dimensions. Therefore, it is necessary to determine both the data's location and the general structural shape. From table 2.3 within section 2.4, the available data sources are detailed and outlined. However, the integration of these sets is not discussed. To ensure unequal data lengths are capable of merging, mutual data features must be compared and aligned. Fortunately, many characteristics such as timestamps, latitudes, longitudes, and vessel speeds are shared within the available datasets. Once alignment is verified, higher frequency datasets can be used as an interpolation foundation for the remaining sources. This process allows the raw in-between data structures to be additionally aligned.

Unfortunately, this process is not without its flaws. For example, interpolating between datasets creates artificially introduced data points that may not truly exist in the operational environment. Therefore, it is recommended to adopt as few varying datasets as possible to reduce the inherent interpolation errors. Nonetheless, when overlapping parameters display a high degree of correlation between various sets, such influences are generally minor or are naturally eliminated through the cleaning process. The Data Integration procedure is constructed using a *Python* environment.

5.2.2. Data Transformation

The data transformation process is a general procedural transformation that converts string or character data points into numerical inputs. The process is necessary for internal ANN computations. Generally, these inputs are classification-based data inputs that describe a condition or scenario. For instance, operational labels such as 'Sailing' or 'Anchor' must be encoded. This transformation procedure is typically done using either ordinal encoding or one-hot-encoding.

Ordinal encoding is a conventional procedure that assigns each unique character string a finite discrete numerical value of ranked ordering ranging from 0 to the number of entries. This conversion allows for a distinct designation within the ANN structure. However, with pure classification-type problems, it imposes a ranked relationship where no such relationship may exist. To overcome these weaknesses, one-hot-encoding can be used instead. In this instance, each unique entry becomes its own data feature with a binary activation of either 0 or 1. The benefits gained by having independent relationships are opposed to adding more input features. Fortunately, in the available datasets, Data Transformation is not required as all parameters are either continuous or binary numeric.

5.2.3. Input Missing Data

Missing data can cause significant errors in the ANN training process. As such, data points containing any missing information must usually be removed to avoid such problems. However, removing all data points is typically unfavourable as information is expensive to obtain. Therefore, interpolation between nearest neighbouring data entries is usually applied. If available entries encompass missing entries, interpolation methods can be applied to retain value data points without discarding the entire entry. Additional techniques such as imputation, where missing points are filled with the mean feature value, can also be valuable to retain certain features.

Unfortunately, much like the Data Integration method, interpolation relies on the assumption that the neighbouring points are narrowly spaced to provide reliable estimations. However, the act of discarding an entire sample due to a few missing inputs is not typically suggested with limited samples. Nevertheless, most machine learning models can not handle any missing input features. Therefore, if they are incapable of being filled, they must ultimately be removed.

5.2.4. Noise Identification

Outliers within datasets can present significant prediction discrepancies if found within the ANN training stages. Thus, detecting such irregularities during the data preparation stages is highly important to ensure efficient operation. There are multiple methods used to flag or eliminate outliers; however, the most common procedures are the parametric-based Standard Deviation (SD) and Inter-Quartile Range (IQR) evaluation methods. Generally, data extracted from real-world operations are stochastic in nature, as such Gaussian-like distributions can be expected.

When implementing the Standard Deviation approach, an outlier can be considered beyond the 3rd SD of the distribution. The IQR approach differs slightly; however, typically, outliers are considered 1.5 times the IQR. Thus, while both methods are commonly applied, the IQR approach is typically more robust to skewed data sources. Ultimately, both approaches can be visualized using both histograms and boxplots, respectively.

However, it should be reiterated that both these features assume a parametric form (normally distributed). Nonetheless, distribution-free data points also exist. These more advanced non-parametric methods, such as clustering or density-based grouping, exist to handle such data abnormalities in these situations. While these methods can be compelling, they are also indiscriminate. In other words, a high understanding of the features is required as the process has the potential to reduce data ranges without intention drastically.

5.2.5. Data Cleaning

Data Cleaning is a general procedure where engineering insight is applied to remove irregular data points further. Such examples can include eliminating unrealistic directional values which fall beyond the 360-degree spectrum or impossible recorded ship behaviour such as attaining values greater than the maximum speed or installed power. In such cases, detailed insight into the data must be fully grasped to assess whether the remaining entries are sensible. Ultimately, Data Cleaning aims to reduce irregular or noisy entries in a physical application sense for entry within the ANN, which can be highly susceptible to both outliers and noise.

5.2.6. White Box Model Evaluation

All data entries have been cleaned at this position within the preparation framework, and as such, no missing values shall be present within the set. Therefore, each WBM can be evaluated for each corresponding data entry which can be extracted for use within the GBM framework. Ultimately, WBM solutions can be used as a new input feature within the ANN-BBM structure so long as the feature relevance is high, which can be compared directly using scatter comparison and residual error plots.

WBM modelling assumptions and general technical considerations can be seen in section 4.1, whereas the required WBM data inputs and technical details are presented in appendix B.1.

5.2.7. Feature Selection

Once all features are cleaned and processed, selecting the most relevant features through a Feature Selection Method is crucial. As detailed by Parkes et al. [83], introducing variables that are poorly correlated to the target variables can add unnecessary complexity, negatively impacting model performance and generalization capacity.

For example, both sea and air temperature are usually highly correlated with one another. Including both parameters introduces a feature that provides no new information but consequently increases the modelling complexity instead. Typically, many methods exist to evaluate which features are optimal; however, many become complex in their own right. However, as shown by both Zwart [108], and Parkes et al. [83], using a Spearman's Rank correlation method has proven effective in determining critical inputs for modelling energy consumption.

It should be noted that this method, while simple and effective, only indicates non-linear relations,

which are either monotonically increasing or decreasing. Therefore, parabolic or oscillatory relations will effectively show little to no correlation between the parameters. As such, this method should only be used as a mere guide in flagging highly correlated features.

5.2.8. Data Scaling

Once all features are integrated, cleaned, and selected, re-scaling the input and target data using the most appropriate scaling method is required. This approach means either applying normalization or standardization to the input and output features of the completed dataset. Each method's importance and working principles can be seen in detail within section 4.2.4.

This process is completed within the *Python* environment using the *sklearning.preprocessing* toolbox *MinMaxScaler* and *StandardScalar* features, (Pedregosa et al. [87]). It should be noted that scaled modelling results are not one-to-one reflections of actual-world results until the final estimations are re-scaled to the input data ranges using the inverse of the described procedure.

5.3. Grey Box Modelling Methodology

A grey-box methodology can be further detailed between the Data Preparation, Modelling, Modelling Evaluation, and Deployment stages of the overall general modelling methodology. The overall structure and GBM methodology pipeline can be seen highlighted in figure 5.3 which ultimately follows the *Serial* approach as introduced by Leifsson et al. [68].
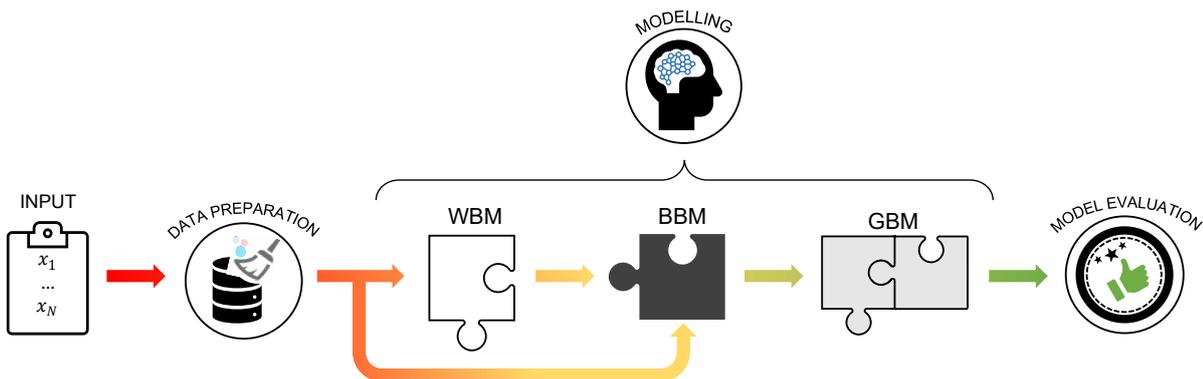
Figure 5.3: Serial grey-box modelling procedure (based on Leifsson et al. [68])

Data inputs are prepared and passed through the WBM features towards the BBM solution. At this point, the BBM is trained on the associated input features, which inherently alters into a GBM solution. As proposed in section 3.6, the BBM will be the Multi-Layer Perceptron Artificial Neural Network.

As two models focused on propulsion and auxiliary power will be developed, the target features, main engine shaft power and dedicated generator auxiliary powers, are extracted from the initial data sets and used to train both regression models, respectively. Inputs range from currently available ship specifications, hindcasted weather sets, onboard sensors, and WBM evaluations (see section 5.3.1). Upon completing the model training with respect to the target data features, a GBM function is developed. This function requires the same input features as it was trained, allowing for targeted power estimations. Additional information on the BBM technical developments can be found in sections 4.2 and 5.4. At this point, the created models are evaluated using appropriate performance metrics, as seen in section 5.4.2, to select the best solution.

It should be noted that each model is subject to feature upper and lower limits. These are the parameter intervals in which the model can confidently function. Extending beyond these limits requires the model to extrapolate beyond its modelling space and, as such, can fall victim to the limitations of the BBM component. Therefore, the Serial approach hopes to introduce WBM features with high input-output dependencies to influence the ANN to behave more in line with the underlying physics,

thus aiding in estimation performance within limited training data regions. However, suppose marginal relations between the WBM estimations and the target parameters exist. In that case, the model's behaviour ultimately reverts itself to a pure-BBM approximation, including all associated modelling existing strengths and limitations. Based on feature selection principles (see section 5.2.7), if the dependencies are too poor, engineering decisions regarding the inclusion of the WBM are ultimately required.

5.3.1. WBM and BBM Input Parameters

The available input data for both WBMs and BBMs have been listed in table 5.1. Based on the optimal ANN literature summary tables 4.1 and 4.2, it can be seen that the available data almost wholly aligns with the novel literature ANN input features.

Table 5.1: Available input parameters for both the WBM and BBM

BBM Parameters		
Ship Characteristics	Atmospheric	Environmental
Speed Over Ground, V_s (knots)	Air Temperature, T_{air} (°C)	Wind Speed, V_{wi} (knots)
Heading, ψ (°)	Sea Temperature, T_{sea} (°C)	Wind Direction, β (°)
Gross Tonnage, GT (t)	Relative Humidity, RH (%)	Wave Height, H_s (m)
Displacement, ∇ (t)	Surface Radiation, E_e (W/m^2)	Wave Period, T_s (s)
Froude Number, Fn (-)		Wave Direction, α (°)
Hull Clean Interval, HCI (days)		
Classification	WBM Contribution	Target
Sailing Factor, S_{oper} (-)	Calm-water Power, $P_{s,cw}$ (kW)	Propulsion Power Load, P_s (kW)
Sunlight Factor, α_{sun} (-)	Wind Power, $P_{s,wi}$ (kW)	Auxiliary Power Load, P_a (kW)
	Wave Power, $P_{s,wa}$ (kW)	Total Power Load, P_{total} (kW)
	Total Shaft Power,[1] $P_{s,t}$ (kW)	
	HVAC Power, P_{hvac} (kW)	
WBM Parameters		
Calm-Water Contribution	Wind Contribution	Wave Contribution
Table B.1	Table B.3	Table B.1.5
HVAC Contribution		
Table B.4		

From the propulsion power perspective, only propeller cleaning information and operational trim/draft are unavailable. While the latter two features are very influential to the overall energy demand, the trim and draft are not expected to vary drastically from the yacht design draft. This parameter is mainly attributed to shipping vessels, where the load and overall trim can fluctuate drastically throughout its continuous operation. In comparison, yachts fluctuations are minor, as they are mainly attributed to tank levels instead of added payload and cargo placement. However, while not directly influential to the overall research objective, propeller cleaning information directly impacts the models' capability to inherently learn any fouling relationship to the overall propulsion demand. As such, this missing parameter can be a detriment in the building of internal physical modelling relationships.

From an auxiliary power perspective, all environmental information is seemingly available. The only unknown parameters can be related to the vessel's structural properties that may influence the overall heat transfer capabilities. It should be noted that while most of these parameters are known, they are purely based on buildings engineering evaluations. As such, essential missing system factors such as active stabilizers and rudders can play a very influential role in the auxiliary load's overall dynamic relationships in addition to the HVAC systems. Therefore, the neglect of such parameters has the potential to disrupt the model's overall learning capabilities.

---

[1]Includes a 10% yearly fouling contribution

5.4. Black Box Modelling Methodology

The general methodology can be further magnified to consider the critical modelling functions of the BBM component within the GBM, as seen in figure 5.4.
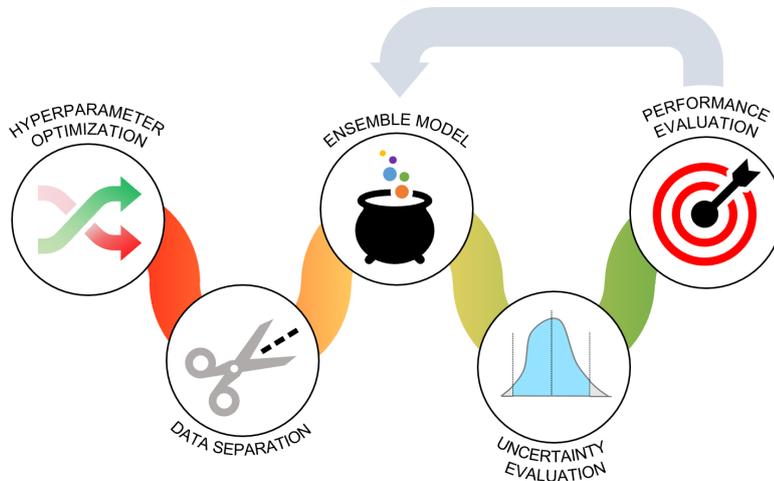
Figure 5.4: Black-box modelling procedure

This sequential pipeline considers the necessary steps outlined and detailed in section 4.2 to ensure optimal performance and adequate modelling evaluations. The main steps are *Hyperparameter Optimization, Data Separation, Ensemble Modelling, Uncertainty Evaluation, and Performance Evaluation*. To further make sure of modelling uniformity, each ANN-BBM modelling step is developed using the open-source *Python* package *Keras* as a foundation (Chollet [24]).

5.4.1. Neural Network Critical Selections

The first stage, *Hyperparameter Optimization*, incorporates a grid searching technique using a cross-validation scheme to evaluate the optimal modelling parameters and structure for the given dataset. Ultimately, as listed in table 5.2, eight common hyperparameters are varied for each application to ensure maximum modelling generalization and efficiency.

Table 5.2: Critical ANN hyperparameters

Dropout	Max. Epochs	Batch Size	Initializer[1]	Activation	Optimizer	Layers	Neurons
0	10	1	Glorot Uniform	ReLu	Adam	Varies	Varies
0.5	100	32		Sigmoid	SGD		
		64		Tanh			

A detailed breakdown of the individual parameters and technical considerations can be seen in section 4.2.3. Additionally, tables 4.1 and 4.2 are used as starting foundations for the grid searching application. It should be noted that for each model tuning, a two-tier magnification was conducted. The 0th Zoom's focus is on the first six hyperparameters, whereas the last two parameters relating to the network shape (layers and neurons) are left broad and general. Once the best solution is determined, the 1st Zoom is conducted. This phase focuses on enhancing the layer and neuron ranges to consider the neighbouring regions to determine the optimal network topography. Ultimately, this two-tier analysis allows for a more efficient refinement process as computational demand can quickly become astronomical due to the infinitely variable parameter permutations.

The second stage, *Data Separation*, as the name suggests, is the process of separating the cor-

---

[1]Default *Python Keras* hyperparameter

responding dataset into three varying datasets: Training Set, Validation Set, and Testing Set. The associated data split follows a proportioning scheme of 70%, 15%, and 15%, respectively. This data splitting ratio is in line with both da Silva et al. [29] suggestions and the ANN literature investigations' findings, as highlighted in sections 4.3.

The third and fourth stages, *Ensemble Modelling and Uncertainty Evaluations*, incorporates the Bootstrap aggregation methodology. The complete technical procedure of ensembling and evaluating the associated empirical confidence intervals is outlined in section 4.2.5. The study will ensemble 50 bootstraps per modelling evaluation to develop the associated uncertainty components as per the recommendations. However, it should be reiterated that while both modelling and inherent error uncertainties can be obtained during the training stages, only modelling uncertainty can be confidently obtained during the deployment stage due to the unknown target parameters. Nonetheless, if sufficient samples are considered, the average uncertainty band ratios can give a relatively good indication of the generally expected confidence intervals within the deployment phase.

Finally, the last stage deals with the evaluation metrics necessary to quantify the modelling performance capabilities accurately. These critical parameters are further detailed in section 5.4.2. It should be noted that no single metric can reflect the total modelling performance. As such, multiple metrics are typically applied, each with individual characteristics to fully assess the model's performance.

5.4.2. Performance Evaluation

The performance of the models will be evaluated by analyzing four commonly applied metrics as outlined by Botchkarev [16]. These include the *Coefficient of Determination (R^2), Mean Error (ME), Mean Absolute Error (MAE)*, and the *Root Mean Squared Error (RMSE)*. The associated formulations for each and their respective percent representation can be found in table 5.3.

Table 5.3: General Performance Indicators

Metric	Formulation	Percent Formulation
Coefficient of Determination	$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2}$	
Mean Error	$\bar{E} = \frac{1}{N} \cdot \sum_{i=1}^{N}(y_i - \hat{y})$	$\bar{E}_\% = \frac{100}{N} \cdot \sum_{i=1}^{N} \frac{(y_i - \hat{y})}{y_i}$
Mean Absolute Error	$|\bar{E}| = \frac{1}{N} \cdot \sum_{i=1}^{N}|y_i - \hat{y}|$	$|\bar{E}_\%| = \frac{100}{N} \cdot \sum_{i=1}^{N}\left|\frac{y_i - \hat{y}}{y_i}\right|$
Root Mean Square Error	$\sqrt{\bar{E}^2} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N}(y_i - \hat{y})^2}$	$\sqrt{\bar{E}_\%^2} = \sqrt{\frac{100^2}{N} \cdot \sum_{i=1}^{N}\frac{(y_i - \hat{y})^2}{(y_i)^2}}$

The coefficient of determination (R-squared) is a classical statistical measure that indicates the overall goodness of fit. The metric ranges from 0 to 1, where the former is a fit through the global population mean, and the latter is a perfect fit. It should be noted that the parameter can potentially produce a negative result. Ultimately, this indicates that the chosen model fits the data poorly and does not follow the associated data trend. Unfortunately, while the R-squared approach shows general model fitting, it does not provide any actual prediction error indications.

The ME and its counterpart MPE indicate the residual difference between the estimated and the target values. While this metric provides a relative indication of performance, it does not show the variance amongst the residual results. Nonetheless, it provides information on whether the developed model systematically underestimates (more negative error) or overestimates (positive error). As such, this metric is usually considered only an indication of the relative modelling bias and is sometimes known as the bias error.

The MAE and MAPE focus on the absolute residual error difference. Ultimately, this performance indicator describes the typical magnitude of the estimation errors. This metric is one of the most commonly applied performance indicators for its simplicity. Each residual contributes proportionally to the total error, meaning that more significant errors contribute linearly to the overall error. In other words, a lower MAE represents a more excellent estimation, while a larger MAE may suggest that all or a portion

of the estimation regions behave poorly. Unfortunately, due to the MAPE structure, a critical weakness results from the division operation. As such, the MAPE can grow unexpectedly large if the actual target values near zero.

Another commonly applied performance metric is the RMSE. This performance metric applies the conventional mean squared average, thus imposing a high penalty on outlying data points. Ultimately, the RMSE can be highly skewed by outliers; therefore, it provides a robust result to punish significant modelling errors. This metric can be used in conjunction with the MAE to understand the relative modelling variance. The MAE will always be less than the RMSE; therefore, the closer the two metrics, the better the general model is in deviations between estimations and the targets. The RMSPE gives a relative percent indication of the RMSE performance metric. However, much like the MAPE, it is subject to artificial numerical increases when the target parameters approach zero.

Ultimately, each modelling performance indicator has its advantages and disadvantages. As such, Botchkarev [16] suggests that a single metric cannot thoroughly diagnose the global modelling performance. However, by using multiple performance metrics, a general understanding of the modelling behaviour can be successfully captured, investigated, and assessed.

5.4.3. Performance Verification and Validation

Upon completion of the model development, a global modelling verification and validation must be considered. These two criteria are critical in the development of any model and, as such, must be judged carefully.

Verification of the model is an attempt to evaluate whether the model has been built and functioning correctly. Ultimately, this criterion is an investigation into whether there are any errors within the development phase. In simpler terms, verification is confirmation that the model has been created correctly. On the other hand, validation is a process in which the modelling outputs are evaluated and judged according to the established requirements. This criterion considers conditions such as usefulness and whether the approach addresses real-world needs. In the case of the following investigation, the verification and validation are to be considered as follows,

Verification: Since the methodology has no indicators of whether or not errors exist within the development process, comparisons between existing case investigations are the best gauges of successful development. As such, a direct comparison between the highlighted propulsion and auxiliary literature summaries seen within tables 4.1 and 4.2 are to be used as a comparative baseline. Since these literature sources provide detailed and transparent performance indicators, the successful development of the model should provide results within similar error ranges. Conversely, if the corresponding literature and actual results deviate considerably, an internal scripting error is likely high.

Validation: This criterion is much more challenging to evaluate as this metric can be highly subjective. Nonetheless, within the Computer Science community, model validation is often referred to as the process where a trained model is evaluated with an independent testing dataset (Wang and Zheng [105]). This procedure provides the unbiased performance of the model on data entries it has yet to encounter; thus, indicating true performance and generalization ability. As such, the model validation process implements the above-established performance metrics and directly compares the determined outcomes with the established modelling requirements seen in table 3.2.

Ultimately, the verification and validation process is used to ensure that all modelling intended functions are operating sufficiently and that the model does not produce any adverse and unintended results.

6

Model Evaluation

The following chapter demonstrates the applicability of the modelling methodology outlined in chapter 5. The approach is demonstrated and proven through a vessel case analysis where both the propulsion, auxiliary, and total dynamical loads are estimated and evaluated. Ultimately, this chapter aims to answer the following research question,

'How is the general performance influenced by varying modelling categories, data-preparation procedures, and vessel-specific operational usage?'

Section 6.1 presents a general introduction and overview of the case vessel. Next, applying the modelling methodology critical steps, the Data Preparation procedure will be implemented within Section 6.2. Here, case-specific details on dataset integration, data quantity and quality, white-box model evaluations, optimal feature selection, and BBM parameter optimizations are presented and compared. At which point, Section 6.3 presents the different modelling results for the various dynamical loading components determined within the proposed GBM training stage. In addition, this section explores the ability to aggregated various models to verify the feasibility of multi-model proportioning. Finally, section 6.4 compares the corresponding modelling results amongst the various modelling categories amongst relevant literature results based on similar investigations.

6.1. General Case Introduction

The case in question focuses on a single modern *Feadship* vessel, *Yacht A*. This yacht is an ideal candidate for the following case investigations for two overarching reasons.

The first is that this yacht is one of the newest amongst the *Feadship* Fleet. As such, the implementation of onboard continuous monitoring systems allows for direct measurement of most power-consuming systems with a high-frequency 3-minute sampling rate. These measurements allow for a robust and varied selection of potential feature inputs to further enhance the capabilities of the GBM. Furthermore, since the ANN uses a supervised learning approach, accurate target parameters such as individual power demand found on both main engines and generator sets are critical.

The second reason is that both the data quantity and quality are expected to be high. While the ship is a newer vessel, 2+ years of high-frequency onboard sensor information and voyage AIS location data are readily available since launch. In addition to these high quantity datasets, large amounts of resources have been invested in ensuring the obtained information's quality is accurate via multi-set validations and comparisons. The main vessel parameters can be found in table 6.1

Yacht A is a large (~100m) semi-displacement styled vessel with no attached bulbous bow. Where, the apportionment between propulsion and auxiliary power demand can be obtained directly by dedicated generator sets due to a Diesel-Electric configuration. These generators are linked with two electric engines and coupled via shafts towards two 5-bladed fixed-pitch propellers with a diameter of approximately 2.5m each.

Table 6.1: Vessel design specifications

Yacht A Main Particulars			
Parameter	Symbol	Value [1]	Unit
Waterline Length	L_{WL}	100.0	m
Waterline Breadth	B_{WL}	15.0	m
Draft	T	4.0	m
Gross Tonnage	GT	4500	gt
Maximum Speed	V_{max}	18.0	knots
Range Speed	V_{range}	15.0	knots
Range	R	5000.0	nm
Configuration	DE	Diesel-Electric	-

As noted in section 3.8, the investigation will first focus on a single vessel due to time and data availability limitations from other vessels. Nonetheless, by first investigating a single ship, the methodology's success and limitations can be deeply explored to give detailed insight into the modelling approach's feasibility. At which point, if the results show promise, the research's continuation via application to the remaining *Feadship* fleet can be conducted.

6.2. Data Preparation

The Data Preparation stage implements the procedure outlined and highlighted within figure 5.2. As detailed within section 5.2, this process is critical to converting raw operational information from multiple datasets into one processed set. The following sub-sections highlight the corresponding case application of the preparation methodology and ultimately demonstrates how the associated approach can be successfully applied.

6.2.1. Data Integration and Transformation

As identified in section 5.2.1, many sources are available and must be integrated to allow for singular data entries within the GBM. In the associated case, three primary datasets are obtained between the periods of 01-10-2018 and 01-07-2020. These include onboard sensor data (AMCS), hindcasted weather data (ECMWF), and additional voyage data features (Valid). Upon collection, the individual sources can then be aligned, cross-validated, and merged using a general interpolation method. The alignment process is based on associated data entry time stamps, latitudes, longitudes, or any data feature with overlapping presence in any two collections. The corresponding aligned sets can be seen highlighted in figure 6.1.

Unfortunately, while the full data sensor ranges nearly cover a 2-year period, the associated weather and voyage dataset is a limiting factor. As such, the actual data range quickly drops to an overlapping 10-month period. In addition to reducing the usable data ranges, the associated fidelity of each resulting collection dramatically varies as well. As emphasized in section 5.2.1, the degree difference in the dataset frequency can, unfortunately, provide considerable uncertainty between the datasets once interpolation is applied. While the AMCS dataset is collected using a 3-minute sampling frequency, the corresponding voyage and weather data have hourly intervals. While voyage information does not radically change over time, weather conditions behave much more dynamically. As such, a degree of uncertainty amongst the actual operational conditions can exist as some interpolated features may not be suitably represented.

It should be noted that, on brief occasions, irregularities within the onboard sensor datasets are noticed. Such errors can be directly visualized within the latitude and longitude sub-figures. Seemingly, when the signals are lost, a corresponding null placeholder data entry is obtained instead of valid measurements. Unfortunately, such recording errors are tough to locate and can significantly skew the modelling analysis. As such, only by carefully applying the data preparation procedure throughout the

---

[1]Approximate parameters due to *Feadship* required confidentiality
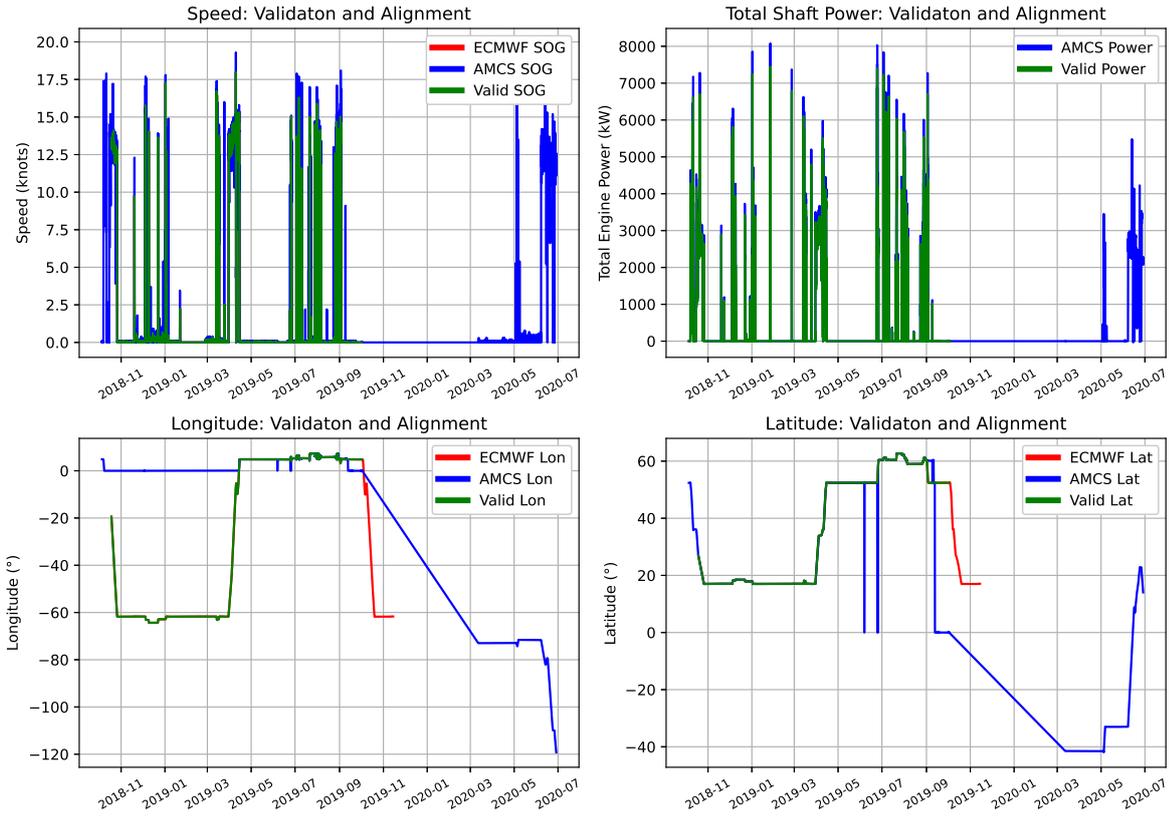
Figure 6.1: General dataset alignment and validation

remainder of the study can true data points be ascertained as opposed to extreme interpolation and measurement errors.

In addition, a few critical input features can be manipulated and transformed into more GBM appropriate data entries upon completing the data alignment and integration. These include vessel operation classification, sunlight classification, hull clean interval quantification, and wind or wave relative direction orientations.

The sailing operation factor can be quantified as a binary classification between *Sailing* and *Anchor* as follows,

$$S_{oper} = \begin{cases} 0 & V_s < 1 \ knot \\ 1 & V_s \geq 1 \ knot \end{cases} \tag{6.1}$$

When the vessel has a speed over ground greater than 1 knot, the vessel operation can be classified as *Sailing*. Anything less is quantified as at *Anchor*. This threshold was deemed adequate for the sake of distinguishing between pure sailing as well as low-speed manoeuvring operations. In the study's context and the general design process focus of the early-stage design, manoeuvring power demand is eliminated as this application currently falls outside the bounding scope.

Like the sailing factor, a binary daylight factor can be established. This factor quantifies the radiation component throughout a typical day. The selected threshold considers the interior time between 06:00 to 18:00 as *Day* and the outer time as *Night*.

$$\alpha_{sun} = \begin{cases} 0 & t_{localhours} \leq 6 \quad or \quad t_{localhours} > 18 \\ 1 & 6 < t_{localhours} \leq 18 \end{cases} \tag{6.2}$$

It should be noted that all voyage and onboard sensor timestamps are provided in Coordinated Universal Time (UTC). Therefore, a conversion using local timezones to align the vessel's location with the local time must be completed before the classification process.

Additionally, the hull cleaning intervals can be determined as,

$$HCI = \text{Current Date} - \text{Date Hull Clean} \tag{6.3}$$

This factor quantifies the linear time increases since the last experienced hull cleaning. While it should be noted, as of yet, no hull clean for *Yacht A* has been recorded. Nonetheless, this factor plays a critical role and establishing a connection between power and fouling.

Finally, both directional wave and wind features can be converted into relative orientations as opposed to the earth-centred coordinate systems,

$$\beta_{360} = \beta_{t,ec} - \Psi_{heading} \tag{6.4}$$

$$\alpha_{360} = \alpha_{t,ec} - \Psi_{heading} \tag{6.5}$$

The function is a basic conversion between the global ship heading and the incoming wave or wind direction. Nonetheless, the determined parameter can be further manipulated into a new, more concise data input feature. It is assumed that the effects of wind and waves from the portside are equal to that of the starboard side due to yacht hull forms' symmetrical nature. Therefore, any directions with an angle greater than 180° are mirrored to an equivalent direction,

$$\alpha_{180}, \beta_{180} = x_{360} = \begin{cases} x_{360} & \text{if } 0° < x_{360} \leq 180° \\ 360° - x_{360} & \text{if } 180° < x_{360} \leq 360° \end{cases} \tag{6.6}$$

If the corresponding angles lie in any other region, a complete 360° unit is either added or subtracted depending on the corresponding sign direction.

6.2.2. General Data Cleaning

Upon the data integration and transformation conclusion, a complete dataset containing $61,944$ potential data entries is developed. At this stage, a general data cleaning process can now be conducted to ensure all entries meet the necessary GBM requirements and irregular data is eliminated. This process considers the following three critical stages of the preparation methodology: *Input Missing Data, Noise Identification*, and *Data Cleaning*.

However, before continuing the data preparation steps, the combined dataset must be orientated towards each individual powering focus. Since the proposed solution is to develop models considering propulsion shaft power and auxiliary power, the associated datasets must reflect these different outcomes. As outlined in section 6.2.1, a classification is developed to distinguish between both *Sailing* and *Anchor* operations. Thus, the corresponding datasets can be likewise divided using this data feature. Ultimately, this means that three distinct datasets can be used in the subsequent investigation, *Sailing, Anchor,* and a *Combined* set. The associated datasets and each corresponding data feature for the collection can be visualized using polar charts, histograms, and boxplots in appendix D.1.1.

Upon initial inspection of the collected dataset, no standout irregularities exist. However, on further examination, the vessel heading parameter within the *Anchor* condition may contain a source of error. As remarked prior, once a signal is lost, the onboard sensors inputs a null measurement. This measurement error is reflected by the overall dominance of a zero heading seen within the polar chart. Unfortunately, this parameter dramatically distorts the subsequent relative wind and wave directions, which ultimately can influence the GBM's internal physical dependencies. As such, a fourth dataset is created, which removes any vessel heading entry of precisely zero. Unfortunately, it is nearly impossible to distinguish between actual zero headings and measurement errors; nonetheless, it is expected that perfectly zero-degree orientations are exceedingly rare in actual operation. Nevertheless, this elimination strategy reduces the associated *Anchor* condition dataset by 57%.

A general summary of each corresponding dataset operation and the associated data cleaning processes and associated total decrease can be seen in table 6.2. A detailed summary of each individual operation can be further found in appendix E.1.1.

Each obtained data point is a valuable commodity. Therefore, it is always the goal to retain as much information as possible. However, when many of the data features are missing inputs, this is

Table 6.2: Data preparation results summary

	Outlier Detection: None		Outlier Detection: IQR	
Initial Datapoints	61944		61944	
Pre - Processing Step	Amount Dropped	Data Remaining	Amount Dropped	Data Remaining
Sailing Operation				
Sailing Only	55999	5945	55999	5945
Missing Data	993	4952	993	4952
Data Specifications	185	4767	111	4841
Outlier Drop	0	4767	1245	3596
Final Datapoints	**57177 (⇓19.8%)**	**4767**	**58348 (⇓39.5%)**	**3596**
Anchor Operation				
Anchor Only	5945	55999	5945	55999
Missing Data	14506	41493	14506	41493
Data Specifications	8076	33417	552	40941
Outlier Drop	0	33417	11491	29450
Final Datapoints	**28527 (⇓40.3%)**	**33417**	**32494 (⇓47.4%)**	**29450**
Anchor Operation ($\Psi \neq 0$)				
Anchor Only	5945	55999	5945	55999
Missing Data	14506	41493	14506	41493
Data Specifications	34158	7335	32177	9316
Outlier Drop	0	7335	2431	6885
Final Datapoints	**54613 (⇓86.9%)**	**7335**	**55063 (⇓87.7%)**	**6885**
Combined Operation ($\Psi \neq 0$)				
Anchor + Sail	0	61944	0	61944
Missing Data	15512	46432	15512	46432
Data Specifications	34341	12091	33976	12456
Outlier Drop	0	12091	1964	10492
Final Datapoints	**49853 (⇓80.5%)**	**12091**	**51452 (⇓83.1%)**	**10492**

not always possible. Nonetheless, the nearest neighbour interpolation is possible if the bounding cells are known. By implemented this strategy, an average of 6% of the missing data points across all the operations is retained. Then, unfortunately, the remaining entries must be dropped, of which the total average drop is proportionate to 23% across all operations. A further investigation into the subsequent datasets reveals that most data features contain most, if not all, data entries and that only a few data features are the root cause,

- Auxiliary Load: 16% missing data entries

- Wave Height, Period, and Direction: 7% missing data entries each

- Sea Temperature: 5% missing data entries

The next most significant feature with missing inputs is the wind parameters with a minuscule 0.02% of missing entries.

As detailed in section 5.2.4, extreme outlier parameters can significantly hinder the general GBM training process. Therefore, a method to reduce the general noise within each corresponding dataset is critical. As such, a parametric *IQR* methodology is adopted. This method is highly effective with skewed datasets and is routinely applied. Unfortunately, one of the main fallacies of the approach is that, while most extreme outliers are removed, potentially valid points can be caught within the elimination process. Thus, the technique is generally considered over-conservative in its approach to detection and removal. Therefore, two additional cases are applied for each corresponding operation; *IQR* or *non-IQR* removal.

The resulting *IQR* data bounds can be visualized within appendix D.1.1 whereas, the direct effects can be seen highlighted within table 6.2. Based on the figures, the features most influenced by the *IQR*

method are then ones that exhibit a high degree of concentration and low standard deviation around a specific region. In this instance, representative features such as SOG, Surface Radiation, and Auxiliary Load are typically targeted.

- Speed over Ground (V_s) under the *Sailing* condition shows a large concentration of values in the cruising speed region of approximately 10 to 16 knots. While other speeds are present, the counts of these instances are drastically lower in comparison. However, under the *Anchor* Operation, the degree magnitude of zero-speeds dominates the datasets. As such, any speeds above 0 knots are ultimately considered outliers and eliminated.

- The Surface Radiation (E_e) data feature presents a highly skewed data set in all operational datasets. Thus, the outlier detection method is used to eliminate most of the data feature's high variance tail region.

- Auxiliary Load Component (P_a) similarly presents outliers in all operational datasets. This feature illustrates an ideal distribution for the parameter approach. As such, the extreme upper and lower entries are typical targets. Since this data feature is essential to the GBM training processes, extreme outliers can play a prominent role in the model development's overall success.

Ultimately, it is seen that the application of the outlier detection method results in an average 12% data elimination as compared to the total operation amount.

The general cleaning procedure's final step is applying self-established specifications to ensure all data features contain physically possible values. This step requires a high degree of insight and application knowledge to verify feasible specifications. The following specifications have been incorporated to ensure most if not all input features are sensible,

- $Data < 0$ (excl. T_{air}): 30 entries removed - All data that enters a negative region are not physical, excluding exterior air temperatures, which may fall below 0 °C are to be eliminated.

- $\Psi = 0$: 32,000 entries removed - Known sensor errors resulting in a precisely 0 heading are to be eliminated (Alternative *Anchor* Operation).

- $RH > 95\%$: 430 entries removed - A relative humidity greater than 95% is to be eliminated. This situation is a highly extreme case; while physically possible, the probability of occurrence is slight and should not be used within the GBM training procedure.

- $V_s \leq 0$ & $P_s > 0$: 280 entries removed - If the speed over ground is zero and some propulsion load exists, the entry is eliminated. Since the propulsion load is proportionate to the speed, zero speed should incidentally reflect a zero propulsion load.

- $V_s > 0$ & $P_s \leq 0$: 1,350 entries removed - Opposite to the above condition, if the vessel records a noticeable speed, and the propulsion power demand is zero, the entry is likewise eliminated.

The resulting specification drops result in an average of 32% data elimination compared to the total operation amount. However, it should be noted if the heading specification is not considered, the self-specifications represent a mere 5% drop overall to the collected datasets. A more detailed overview of each individual specification for the various operational datasets can be seen in appendix E.1.1.

6.2.3. White-Box Model Evaluation

At this stage in the preparation methodology, the developed data collections can be considered *cleaned*. Thus, the WBM evaluations can be conducted for each operational situation. Both the *non-IQR Sailing* Propulsion and *non-IQR Anchor* Auxiliary WBM results can be seen in figures 6.2 and 6.4, respectively. Further details can be found within appendix D.1.3 where all case results are plotted for all considered operational variations.

When inspecting the propulsion comparisons, it can be seen that there is a relatively good correlation between the collected data results and the estimated predictions. The calm-water model generally presents a lower bound to the operational points. This relation is to be expected as this model only considers idealistic environmental conditions. When the additional resistance components are included,
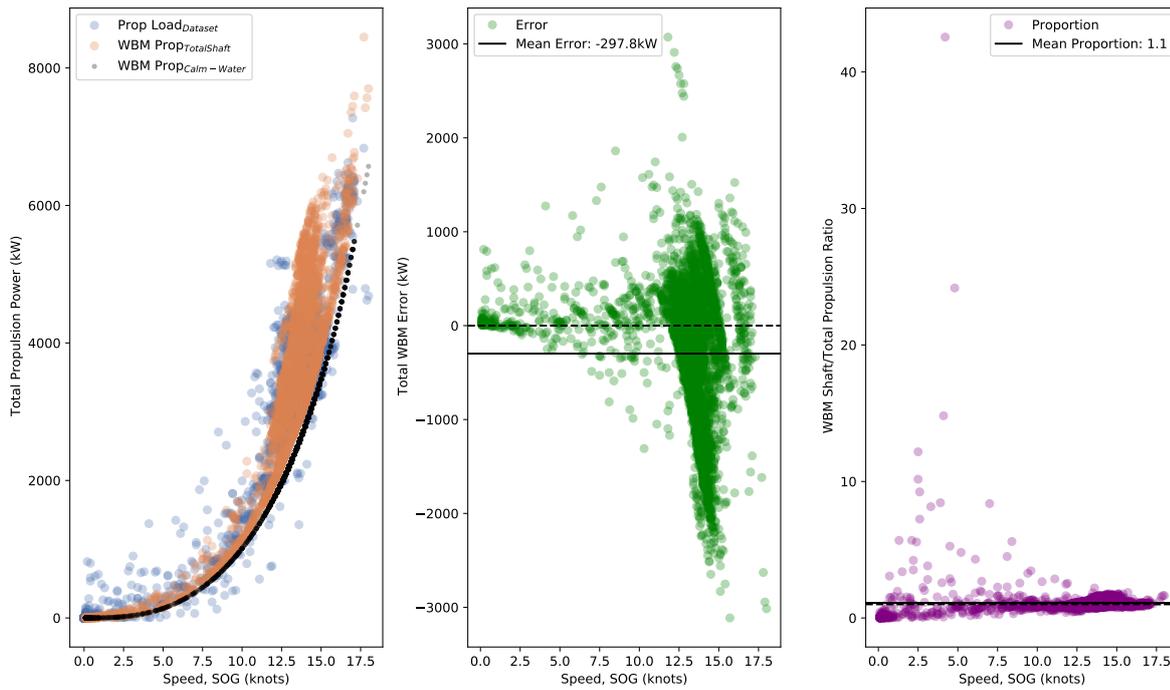
Figure 6.2: Comparisons between total Propulsion power (non-IQR Sailing), WBM total shaft, and WBM calm-water

the total shaft WBM presents a much more realistic representation. However, there still exists a degree difference between the actual gather results and the estimations. As seen within the residual error plot, a significant deviation is shown between the ranges of 12.5 and 15 knots. This discrepancy indicates that the added powering models cannot successfully capture the dynamic environmental interactions undergone in operation during this range. Figure 6.3 highlights each individual WBM modelling component contribution to the total WBM shaft propulsion power and their corresponding proportions to the actual operational results, respectively.
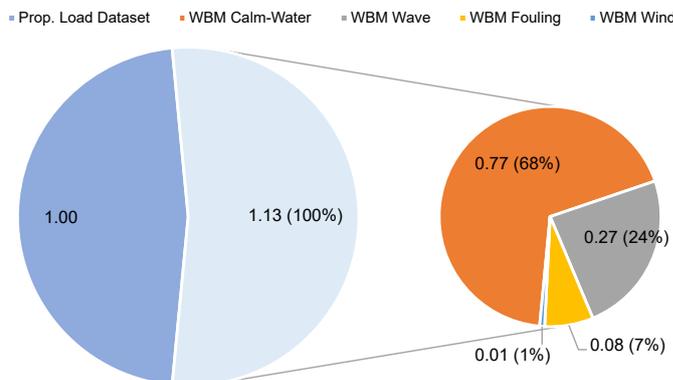
Figure 6.3: WBM Propulsion power modelling contributions and proportion comparisons

The calm-water proportion makes up the most considerable portion at 77%. The VoogtWave component likewise makes up a substantial proportion at 27%. However, it can be seen that the overall influence of the wind model on propulsion power is at a mere 1%. While this proportion is low, it does fall in line with literature investigation suggestions that wind usually encompasses about 2% of the calm-water component. Ultimately, it can be seen that the total shaft power WBM overestimates the average operational load by approximately 13%, where the majority of the dynamic influence exists due to the added thrust in waves.
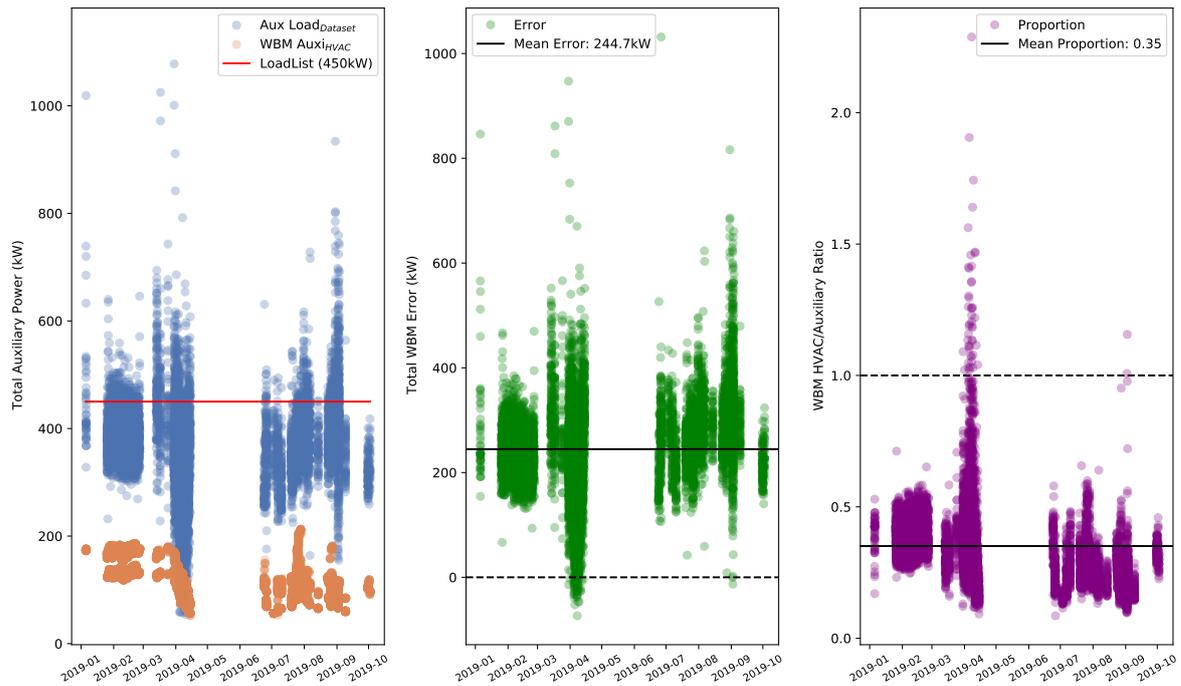
Figure 6.4: Comparisons between total Auxiliary power (non-IQR Anchor), WBM HVAC power, and calculated Load Lists

The Auxiliary power comparison deviates quite drastically from the propulsion WBM evaluations. It can be seen that the collected results have substantial variance. As noted in section 3.6, the developed WBM only considers the HVAC portion of the auxiliary power component. Nevertheless, based on qualitative observation, the WBM does seemingly follow the general trending of the recorded auxiliary powers, thus capturing, to a degree, some dynamic effects.

Additionally, the WBM falls within the 30% to 50% contribution range as per the initial literature findings. However, the model is not an exact replication, and in some instances, the actual recorded auxiliary power falls below the WBM evaluations. In addition to the WBM, a developed operational load list is used as a comparison. Unlike the WBM, this developed estimation represents the total expected load. Unfortunately, these calculations only provide a singular point-based solution, wherein the case vessel predicts a constant required 450 kW for successful operation. It should be noted that on closer inspection of the WBM, a slight gap between estimations is present. This difference is due to the daylight factor, which influences the inherent radiation contribution. During the Night, radiation is not present; thus, the additional required power is neglected in such situations. Nevertheless, while not perfect and having no alternative other than the constant operational load list estimate, the WBM HVAC power model provides a large dynamic portion of the total demand.

6.2.4. Feature Selection

Now that all parameters are known, a feature selection method can be conducted (see section 5.2.7). A detailed look into the full Spearman Correlation matrices can be seen in appendix D.1.2, where the mean absolute (MASCo) summary of the top 5 correlators for each dataset variation can be seen in table 6.3. Each of these dataset variations can be further decomposed into its specific function, i.e. whether the dataset evaluates auxiliary, propulsion, or total power demand. The complete dependency lists with all input features can be seen within appendix E.1.2.

When comparing the various operational conditions, it is clear that the auxiliary power dependencies are much lower than that of the other cases. Unfortunately, this is likely to provide initial insight into the overall prediction accuracy of the future GBM. There are many potential reasons that the correlations are lower in comparison to the other operations. One likely explanation is that the auxiliary power

systems are composed of many different complex components and, once combined, the Spearman Correlation is not suitable to provide an isolated univariate analysis. For instance, a cubic relation between vessel speed and shaft propulsion power may be expected due to the underlying physics. Therefore, both ship speed as well as the corresponding WBM's show significant correlations in all cases. However, in the case of the auxiliary power, no dominating feature is known due to the complex cumulative nature of the total auxiliary power. Amongst the collected data features, the HVAC WBM and atmospheric conditions show the most dynamic dependencies; however, these only make up a portion (approximately 35%) of the total loading. As such, the input-output target dependencies may not be universally related to one or two features but instead composed of multiple low-order contributions from numerous parts.

Table 6.3: Spearman correlation results summary for all operational conditions

	Outlier Detection: None				Outlier Detection: IQR			
Sailing:	Propulsion		Auxiliary		Propulsion		Auxiliary	
Rank	Input	SCo	Input	SCo	Input	SCo	Input	SCo
1	V_s	+0.71	H_s	-0.43	V_s	+0.64	H_s	-0.41
2	$P_{s,cw}$[1]	+0.71	T_0	-0.34	$P_{s,cw}$[1]	+0.63	T_0	-0.33
3	$P_{s,t}$[1]	+0.64	α_{360}	-0.26	T_{air}	-0.63	α_{360}	-0.30
4	T_0	+0.48	β_{360}	-0.23	$P_{s,t}$[1]	+0.53	β_{360}	-0.26
5	T_{air}	-0.42	RH	-0.16	T_0	+0.49	RH	-0.18
Target	P_s	1.00	P_a	1.00	P_s	1.00	P_a	1.00
MASCo (T5)	0.53		0.29		0.58		0.30	
Anchor:	Auxiliary		Auxiliary ($\Psi \neq 0$)		Auxiliary		Auxiliary ($\Psi \neq 0$)	
Rank	Input	SCo	Input	SCo	Input	SCo	Input	SCo
1	P_{hvac}[1]	+0.51	P_{hvac}[1]	+0.32	P_{hvac}[1]	+0.50	P_{hvac}[1]	+0.33
2	T_{air}	+0.46	E_e	+0.31	T_{air}	+0.46	E_e	+0.31
3	T_0	+0.40	α_{sun}	+0.27	T_0	+0.39	α_{sun}	+0.24
4	H_s	+0.36	T_{air}	+0.15	H_s	+0.35	T_{air}	+0.15
5	β_{360}	-0.33	β_{360}[2]	-0.06	β_{360}	-0.33	β_{360}[2]	-0.07
Target	P_a	1.00	P_a	1.00	P_a	1.00	P_a	1.00
MASCo (T5)	0.41		0.22		0.41		0.22	
Combined:	Auxiliary ($\Psi \neq 0$)		Total ($\Psi \neq 0$)		Auxiliary ($\Psi \neq 0$)		Total ($\Psi \neq 0$)	
Rank	Input	SCo	Input	SCo	Input	SCo	Input	SCo
1	P_{hvac}[1]	+0.34	V_s	+0.86	P_{hvac}[1]	+0.30	V_s	+0.84
2	T_{air}	+0.27	$P_{s,cw}$[1]	+0.86	T_{air}	+0.23	$P_{s,cw}$[1]	+0.84
3	H_s	-0.25	$P_{s,t}$[1]	+0.86	E_e	+0.21	$P_{s,t}$[1]	+0.84
4	V_s	-0.23	S_{oper}	+0.84	α_{sun}	+0.17	S_{oper}	+0.82
5	S_{oper}	-0.22	T_{air}	-0.43	H_s	-0.17	T_{air}	-0.39
Target	P_a	1.00	P_{Total}	1.00	P_a	1.00	P_{Total}	1.00
MASCo (T5)	0.26		0.67		0.22		0.64	

It should be noted that a considerable degree of difference between the *Anchor* operational cases exists. It was expected that the elimination of the irregular heading characteristics would improve the consistency of the dataset. While the consistency may have been improved, the dependencies have considerably diminished in the process. One plausible explanation for such an occurrence is not the data quality but the data quantity influence. Elimination of the 0° headings resulted in a drastic drop in data amount for all data features. As such, many influential data entries are potentially eliminated as a consequence.

One note of interest is highlighted by the *Sailing* Auxiliary situation. This operational situation is

---

[1]White-Box Model (WBM) evaluated input feature
[2]Low Spearman Correlation: manually added data feature

the only one that does not significantly depend on the developed HVAC power model. Additionally, in all the auxiliary power operations, this is the only case where all top features indicate a negative correlation to the wave environmental parameters. One possible explanation for such phenomenon can be related directly to the ambient environmental correlations. More significant wave heights indicate a greater propulsion power demand – as expected. However, an increase in wave heights also relates to increasing speeds. Intuitively this is difficult to interpret; however, throughout the 10-month dataset, it is known that two complete transatlantic crossings occur. As such, the vessel transverses beyond the usual cruising speed range into much harsher oceanic conditions. A deeper look into the dependencies indicates that increased vessel speed relates to an inversely correlated ambient air temperature. This dependency can be either related to increasing ambient convection or the external transatlantic environments. Nevertheless, since the exterior ambient conditions are positively correlated with the auxiliary power, a decrease in temperature presents a reduction in required auxiliary power. It can also indicate having guests onboard. With increased speeds and wave heights, we can expect that, most likely, only crews are onboard. As such, stabilization gear is probably retracted, thus lowering overall energy usage during operation. However, when speeds are lower and hence wave heights are lower, the guest is onboard; therefore, more systems are activated for improved comfortability. Generally, it can be expected that more significant wave heights should increase auxiliary power relations as active stabilizers and rudders will play a more critical role. However, since these are data-driven correlations, the human behavioural aspects can play a substantial role in each operational condition.

Looking into the effects of the outlier detection methods can also provide potential insight into the modelling performance. It is expected that if the *IQR* method is applied, extreme outliers are eliminated within the dataset. Therefore, the technique should inherently support strong internal correlations as noise that may skew the modelling results is removed. While this seemingly influences the *Sailing* operation by improving the overall correlations, the *Anchor* operation does not reflect much difference. A more interesting note is that the *Combined* dataset shows that the *IQR* slightly reduces the highest correlations. A possible explanation for such an occurrence is due to the shape of the combined data features. Since the *IQR* method is a parametric approach, it relies on traits that exhibit Gaussian-like distributions. However, as highlighted in figure D.4, many of the highest correlated data features do not present this form, potentially providing distortion within the cleaning process.

Another worthy point of mention is related to the directional components. While the directional features are numeric and indicate a degree of correlation to the target parameters, the corresponding results are highly unreliable. These parameters are orientated using a polar coordinate axis, thus are a radial indication of direction. As such, the corresponding results may or may not be monotonic. Nonetheless, these parameters are critical in operation and thus must be included within the GBM training stages. However, it should be noted that the wind speed component is typically correlated very low. As this component generally has a marginal contribution to the overall power demands (see section 6.2.3), thus the corresponding wind direction may incidentally contribute very little.

While the dependencies indicate some numerical relation, not all features with low correlations should be disregarded. As such, a minimum threshold is established to ensure some degree of correlation. This established threshold is set at ± 0.10. Anything below the limiting bound is individually evaluated and judge for suitability based on the specific operation and the corresponding functions. For instance, Relative Humidity, Daylight Factor, and Surface Radiation commonly indicate low dependencies. However, all of these parameters generally exhibit oscillatory behaviour. As such, the Spearman Correlation evaluation can not suitably determine whether a strong relationship between the input and output features exists. However, such factors have the potential to influence the HVAC component of the auxiliary power demand significantly. Therefore, any correlator below the established threshold is evaluated based on operational suitability and engineering sense to ensure all essential features are retained.

Ultimately, based on the Spearman correlations and all available features (see table 2.3), an average of 10 data features are removed across all operational conditions. It should be noted that the *Combined* operational datasets naturally have more input features as both auxiliary and propulsion orientated data features are to be retained.

6.2.5. Hyperparameter Optimization

At this stage within the GBM methodology breakdown, the black-box modelling phase starts as each operational dataset has been fully prepared. This part of the training process begins with grid search hyperparameter optimization. The general technical process and decision-making details can be found within sections 4.2.3 and 5.4.1, respectively. A summary of each operational dataset's optimal parameters and structure can be found in table 6.4.

Table 6.4: Hyperparameter optimization results summary

Type	Outlier Detect	Dropout	Epoch	Batch Size	Initializer	Activation	Optimzer	Layers	Neurons $(Z_0 \rightarrow Z_1)$
				Sailing Condition					
Propulsion	IQR	0	100	32	G.U	ReLu	Adam	2	$30 \rightarrow 38$
Propulsion	None	0	100	32	G.U	ReLu	Adam	2	$30 \rightarrow 38$
Auxiliary	IQR	0	100	32	G.U	ReLu	Adam	2	$20 \rightarrow 14$
Auxiliary	None	0	100	64	G.U	ReLu	Adam	2	$30 \rightarrow 38$
				Anchor Condition					
Auxiliary	IQR	0	100	64	G.U	ReLu	Adam	2	$30 \rightarrow 40$[1]
Auxiliary	None	0	100	32	G.U	ReLu	Adam	2	$30 \rightarrow 36$
Auxi. ($\Psi \neq 0$)	IQR	0	100	32	G.U	ReLu	Adam	2	$30 \rightarrow 40$[1]
Auxi. ($\Psi \neq 0$)	None	0	100	32	G.U	ReLu	Adam	2	$30 \rightarrow 38$
				Combined Condition					
Auxi. ($\Psi \neq 0$)	IQR	0	100	32	G.U	ReLu	Adam	2	$30 \rightarrow 40$[1]
Auxi. ($\Psi \neq 0$)	None	0	100	64	G.U	ReLu	Adam	2	$30 \rightarrow 40$[1]
Total ($\Psi \neq 0$)	IQR	0	100	32	G.U	ReLu	Adam	2	$30 \rightarrow 40$[1]
Total ($\Psi \neq 0$)	None	0	100	32	G.U	ReLu	Adam	2	$30 \rightarrow 40$[1]

Based on the results summary, many of the BBM parameters and structures converge towards a similar solution space. This conformity is understandable, as many of the datasets exhibit an overlap of data features. Additionally, since the data sources are only distinguished between self-established operational conditions, much of the data entries contain a degree of resemblance and similarity. Therefore, it is natural to expect a similar range of hyperparameters. Nonetheless, a slight degree of difference amongst each operation can be observed due to the associated design space variations.

A keynote of interest can be related to the investigated literature optimal parameters seen within section 4.3. While the propulsion power structure is comparable, the auxiliary modelling structure is noticeably different. Both the topographical layers and neurons are increased. This rise can be due to the overall complexity difference between buildings and ships. While many of the influential data features reveal a degree of overlap, additional dynamic considerations such as vessel speed, waves, and fouling are unique to ships. An additional note of interest can be related to the grid search refinement methodology. While a two-tier magnification was conducted, a few operational datasets reached the bounding limits of the grid search evaluation, as indicated within table 6.4. Therefore, while the overall structure is improved, there is still room for further enhancement. However, while keeping in mind time limitations versus the overall expected modelling gains, further optimization of the network structure has been neglected. Ultimately, the grid refinement process can be conducted for an infinite number of hyperparameter selections and permutations to determine the optimal global structure; however, the respective computational demand increase must be judged carefully.

6.3. Modelling Evaluation and Comparison

The models can be thoroughly evaluated for each corresponding condition upon determining the optimal BBM structure for each operational dataset. As noted in section 5.4.1, a conventional 70%-15%-

---

[1]Grid search optimization exterior bound has been reached, indicating further optimization is possible

15% data proportioning scheme is to be implemented to ensure adequate training and validation is achieved. The associated performance evaluations, model comparisons, and visualizations for each condition, namely propulsion, auxiliary, and total energy consumption, can be found in sections 6.3.1, 6.3.2, and 6.3.3, respectively. Additional case results, either not meeting the general requirements ($< 15\% \pm 95\%CI$ as outlined in section 2.5) or not outperforming an equivalent operation, can be found within appendix E.1.3.

6.3.1. Propulsion Power Evaluation

The complete *Sailing* Propulsion power analysis can be conducted using the outlined GBM procedure, established comparative performance metrics, and baseline requirements. The associated case output results for both the *IQR* and the *non-IQR* operational datasets can be seen summarized in tables 6.5 and E.8, respectively. Additionally, the best performing model can be visualized with and without developed uncertainty intervals in figures 6.5 and D.20, respectively.

Table 6.5: IQR Sailing Propulsion performance summary comparisons between GBM, BBM and WBM

Sailing Propulsion - IQR					
Test Values	539				
Perf. Metric	$GBM + P_{s,t}$	$GBM + P_{s,cw}$	BBM	$P_{s,t}$	$P_{s,cw}$
R²	0.947	0.949	0.947	-1.219	-2.485
ME (kW)	-3.348 (O)	-2.963 (O)	-1.541 (O)	-343.183 (O)	907.515 (U)
MAE (kW)	73.164	72.765	73.542	600.625	910.225
RMSE (kW)	121.145	120.02	122.413	237.96	996.826
Percent Errors					
MPE	-0.311%	-0.280%	-0.249%	-10.141%	25.836%
MAPE	2.226%	2.213%	2.231%	17.339%	25.938%
RMSPE	3.517%	3.485%	3.554%	22.982%	28.801%
CI95% Lower	258.74	253.86	260.77	-	-
CI95% Upper	255.07	250.38	264.31	-	-
Cover%	96.66%	95.73%	95.92%	-	-
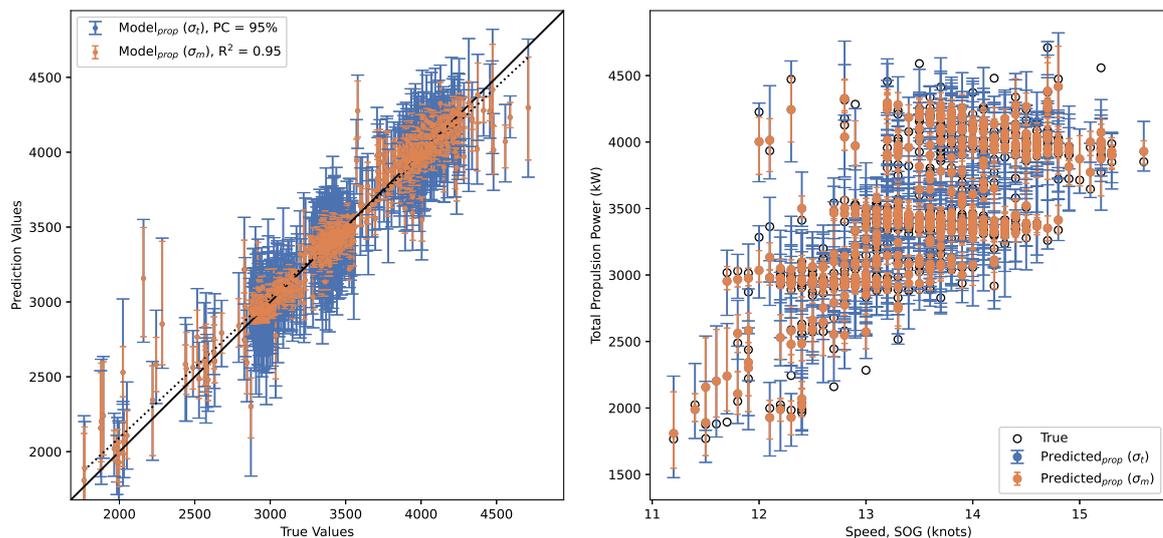
Figure 6.5: IQR Sailing Propulsion $GBM + P_{s,t}$ prediction performance (left) and speed-power relationship (right) with 95%CI

From the results, the best-developed model is determined to be the $GBM + P_{s,cw}$ with a fitting coefficient of 0.946. The solution also exhibits phenomenal estimation characteristics well below the established 15% early-stage design threshold. The developed model effectively makes mean estimations

within 3.5% of the actual operational energy demand while exhibiting a slight overestimation.

In comparison to the other modelling results, both the GBM and BBM heavily outperform the WBM. Two WBM's are used in the comparative study in the associated operation: total shaft power and calm-water power. While the total shaft WBM is the best performing approach, the performance is nearly 20% larger than the best performing GBM. On the other hand, both the GBM and BBM can be seen to behave very similarly, all of which boast impressive prediction performance within a degree of 0.5% of each other.

While all three cases, $GBM + P_{s,t}$, $GBM + P_{s,cw}$, and BBM, meet the established modelling requirements for the *IQR* case, an apparent performance reduction is noticed when an outlier detection method is not applied. In this situation, while the order of performance is still maintained, the performance metrics begin to fall outside the established threshold. Interestingly, in this case, the MAPE is shown to have a higher percent error than the RMSPE. Due to the nature of the performance metric, the lower numerical evaluations are causing significant error discrepancies due to the division operator. As such, the percent indication is not an ideal metric for evaluation.

Nonetheless, the RMSPE, which is more robust to lower-order discrepancies, indicates worse modelling performance for all established models. Thus, the *IQR* approach to eliminate extreme outliers has effectively improved the overall modelling performances. It should be noted that while performance is enhanced, the corresponding data bounds have been substantially reduced as a consequence which inherently reduces modelling functionality and applicability. This limitation becomes apparent via the bootstrap statistical confidence intervals. These indicators allow for detailed interpretation of where the model performs best relative to the testing data. Based on figure 6.5, it can be seen that within the 2800 kW to 4500 kW powering region, both the modelling and inherent uncertainty confidence interval bands are narrowed. Thus, indicating quite good performance. However, when nearing the outer bounds, the associated bars show a relative increase in length. This rise is because the outer data regions are typically limited to the amount of training data available within these regions compared to the interior data domain. Thus, the overall modelling performance decreases when approaching these bounds as well. The practical modelling bounds for each associated data feature can be seen in table E.16.

While the results show promise, a few irregularities have been noticed within the evaluation. From the results, the $WBM + P_{s,t}$ provides an overall better estimation than the pure $WBM + P_{s,cw}$. This performance discrepancy makes a degree of sense as the dynamic contributions are considered instead of only considering calm-water conditions. However, the best functioning model implements the calm-water WBM as opposed to the total operational WBM estimation. Intuitively this isn't easy to interpret. However, when analyzing the determined dependencies established in table 6.3, additional insight into the modelling behaviour can be gained. It can be seen that in all cases, the calm-water correlation ranks above the total shaft estimations. Thus, while the estimation may not necessarily provide a better immediate solution, the internal mappings between the target propulsion and WBM are more closely represented. Unfortunately, this adds a degree of complexity for future case evaluations in that WBM accuracy is less important than the overall dynamic dependencies.

6.3.2. Auxiliary Power Evaluation

Similarly, the complete Auxiliary power analysis can be conducted using the outlined GBM procedure, established comparative performance metrics, and baseline requirements. However, in this case, two different operational datasets are evaluated: *Anchor* and *Combined*. Auxiliary power usage is not limited to *Anchor* or *Sailing* conditions and must be considered in all operations. As such, a comparative analysis between *Anchor-only* and *Combined* conditions is explored. It should be noted that *Sailing-only* Auxiliary power does not meet the established requirements and thus is neglected. For each case result, the associated outputs for both the *IQR* and *non-IQR* operational datasets can be seen summarized in tables 6.6, 6.7 and E.11, E.14, respectively. Additionally, the best performing models can be visualized with and without developed uncertainty intervals in figures 6.6, 6.7 and D.21, D.22.

From the results, the best-developed *Anchor* model is determined to be the BBM with a fitting coefficient of 0.300. The solution, while overall fitting is moderately lower than the propulsion model, indicates performance metrics below the established 15% threshold. It can be seen that estimations are

Table 6.6: IQR Anchor Auxiliary (ΨØ) performance summary comparisons between GBM, BBM and WBM

Anchor Auxiliary - IQR ($\Psi \neq 0$)				
Test Values	1033			
Perf. Metric	$GBM + P_{hvac}$	BBM	$Load\ List$	P_{hvac}
R^2	0.299	0.300	-2.78	-39.18
ME (kW)	1.185 (U)	0.559 (U)	-68.227 (O)	256.073 (U)
MAE (kW)	28.189	28.13	70.062	256.073
RMSE (kW)	35.094	35.069	79.556	259.36
Percent Errors				
MPE	-0.548%	-0.722%	-19.260%	67.004%
MAPE	7.401%	7.399%	19.644%	67.004%
RMSPE	9.166%	9.159%	20.839%	67.936%
CI95% Lower	61.03	60.37	-	-
CI95% Upper	69.71	69.09	-	-
Cover%	92.84%	92.84%	-	-
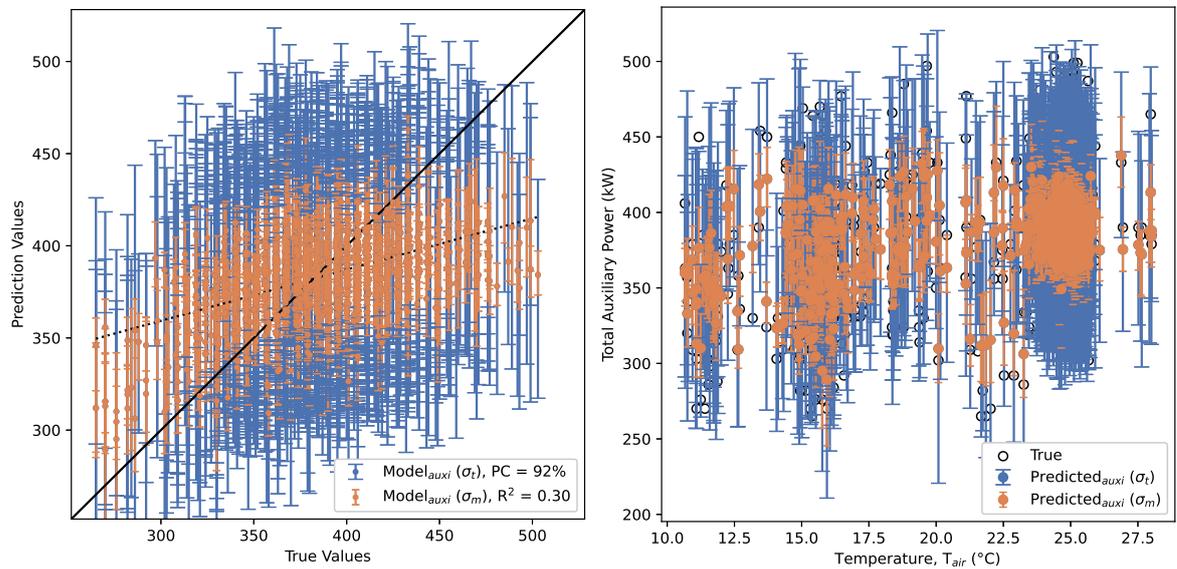
Figure 6.6: IQR Anchor Auxiliary (ΨØ) BBM prediction performance (left) and temperature-power relationship (right) with 95%CI

made with a general error within 35 kW (9%) of the actual operational energy demand while exhibiting a marginal underestimation bias.

In comparison, the best BBM *Combined* model has a slightly increased fitting correlation at 0.308. However, in this case, all performance metrics are somewhat worse than the *Anchor-only* condition, with an average increase of 3.4%. Nevertheless, this case can consider both *Anchor* and *Sailing* operations, thus allowing for much more enhanced implementation and usage.

In both operations, the BBM is the superior model; however, the $GBM + P_{hvac}$ modelling results are nearly identical, with an extremely low prediction error deviation at almost 0.10% on average. In comparison to the other existing modelling results, both approaches heavily outperform the alternative methods. The HVAC model is an inadequate comparison as this only considers a portion of the total auxiliary power. However, the established Load List estimation, which estimates the whole designed auxiliary load, can be compared directly. Based on the results, the singular point Load List estimation captures neither the mean nor the dynamic auxiliary power loading as the average error is approximately 12% greater than the best-developed models.

While the general requirements are satisfied for the zero heading and *IQR* cases in both operations, an apparent reduction in performance is observed for the alternative considerations. When comparing

Table 6.7: IQR Combined Auxiliary ($\Psi\emptyset$) performance summary comparisons between GBM, BBM and WBM

Combined Auxiliary - IQR ($\Psi \neq 0$)				
Test Values	1574			
Perf. Metric	$GBM + P_{hvac}$	BBM	$Load\ List$	P_{hvac}
R^2	0.307	0.308	-1.736	-18.415
ME (kW)	-1.647 (O)	-1.858 (O)	-76.771 (O)	250.129 (U)
MAE (kW)	36.824	36.719	81.661	250.129
RMSE (kW)	47.959	47.921	96.379	256.735
Percent Errors				
MPE	-2.181%	-2.248%	-23.843%	66.645%
MAPE	10.351%	10.339%	24.830%	66.645%
RMSPE	12.947%	12.936%	25.823%	68.788%
CI95% Lower	86.47	85.66	-	-
CI95% Upper	98.04	98.38	-	-
Cover%	93.01%	93.39%	-	-
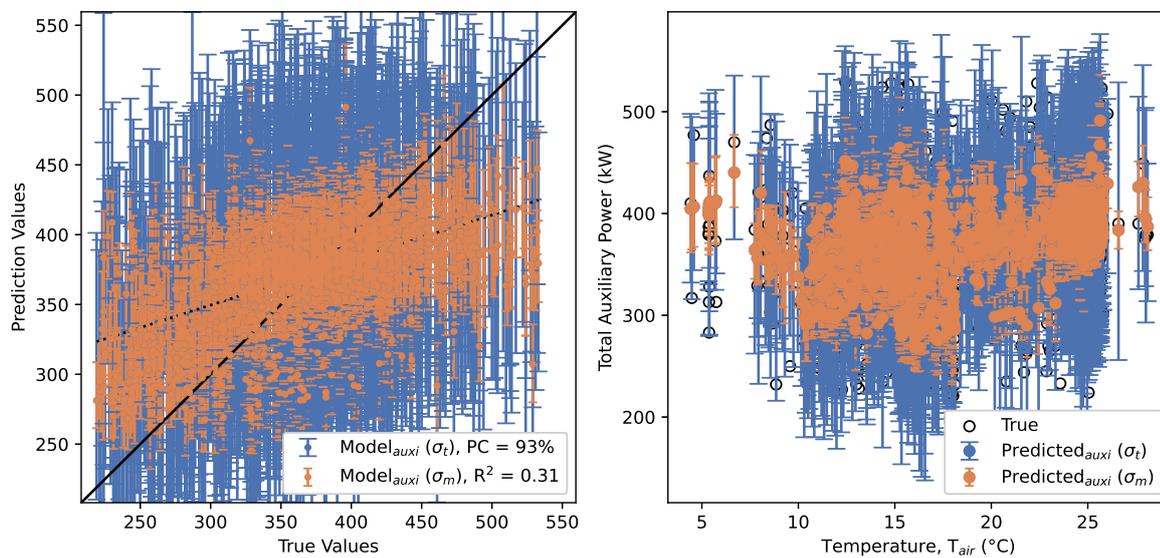
Figure 6.7: IQR Combined Auxiliary ($\Psi\emptyset$) BBM prediction performance (left) and temperature-power relationship (right) with 95%CI

the non-outlier detection approaches, a noticeable performance reduction is once again observed. However, in the *Combined* operation, this drop is sufficient enough to fall above the general 15% requirements. This situation indicates sufficient need and success of the data preparation methodology.

A further comparison between the zero heading specification criteria can additionally provide modelling insight. It can be seen that similar to the outlier detection approach, removing the uncertain sensors entries improves the overall performance. However, this improvement is not quite as significant as expected. Curiously, without removing zero headings, the overall fitting function for the *Anchor IQR* condition improves to an impressive 0.525 (75% improvement). However, it must be stressed that this is not an indication of a better model. The main reason for the drastic improvement of the fitting coefficient is the drastic difference in training data amounts. As seen in table 6.2, the complete *Anchor* condition dataset has nearly 22,500 more data entries. Thus the singular fitting coefficient improvement, while showing a general reduction in performance metrics, strongly indicates over-fitting. This situation provides insight into the complexity of data quantity versus data quality. While more information is always favourable, if the input data entries are incorrect, then faulty outputs will always be produced.

The question remains why the relative performance between the auxiliary and propulsion models

is so much lower in comparison. While there are no singular solutions, many aspects can contribute to the overall limited dynamic success.

First, there is a clear difference in Spearman Correlation magnitudes. It has been observed that the overall dependencies within the auxiliary cases are generally much lower than the operational propulsion equivalents. Therefore, as reflected within section 6.2.4, the overall correlations prove to be significant initial gauges as to the expected performance. Lower dependencies mean that the BBM component of the GBM can not adequately map internal physical connections, thus resulting in a lowered goodness of fit.

Another reason, much related to the previous point, can be due to the HVAC WBM. It was hoped that this model would capture the most critical dynamic operational effects. Unfortunately, while this model generally maintains the most significant correlations of all relevant data features, it is nowhere near the propulsion WBM. This can mainly be attributed to the fact that overall auxiliary power is composed of multiple highly complex systems that continuously interact. Thus, while the HVAC demand is generally considered the most prominent load proportion, the remaining system effects and influences are not retained as the degree of scattering is substantial. These critical missing components can include; active stabilizers and rudders, which are known to contribute significantly to both auxiliary power demand and its dynamical operational contributions.

While the model does not behave as the propulsion models, the developed solutions are still below the established requirement thresholds. In a global sense, prediction accuracy within 35 kW of the total auxiliary load is still significantly improved over the existing Load List estimation at a relative 80kW difference.

6.3.3. Total Power Evaluation

Finally, the completed total power analysis can be similarly evaluated using the outlined GBM procedure, established comparative performance metrics, and baseline requirements. In this case, both the Sailing and Anchor operational are merged to consider a Combined operation. The associated case output results for both the IQR and the non-IQR operational datasets can be seen summarized in tables 6.8 and E.19, respectively. Additionally, the best performing model can be visualized with and without developed uncertainty intervals in figures 6.8 and D.23.

Table 6.8: non-IQR Combined Total ($\Psi\emptyset$) performance summary comparisons between GBM, BBM and WBM

Combined Total Power - None ($\Psi \neq 0$)					
Test Values	1814				
Perf. Metric	$GBM + P_{s,t+hvac}$	$GBMP_{s,cw+hvac}$	BBM	$P_{s,t} + LL$	$P_{s,cw} + LL$
R²	0.995	0.995	0.994	0.902	0.893
ME (kW)	-11.778 (O)	-3.91 (O)	-1.612 (O)	-201.782 (O)	259.662 (U)
MAE (kW)	61.363	60.609	63.199	293.613	348.164
RMSE (kW)	120.343	119.786	126.972	541.9	564.938
Percent Errors					
MPE	-2.156%	-0.960%	-0.197%	-15.869%	-3.560%
MAPE	6.563%	6.322%	6.366%	19.049%	20.162%
RMSPE	7.166%	7.133%	7.560%	32.063%	33.476%
CI95% Lower	276.880	266.810	287.19	-	-
CI95% Upper	252.040	245.180	259	-	-
Cover%	97.74%	97.85%	97.79%	-	-

Based on the results, the best-developed model is the $GBM + P_{s,cw+hvac}$ with a near-unity fitting coefficient of 0.995. Not only does the model have a near-perfect fit, but the estimation performance metrics are also well below the established 15% threshold. The mean absolute error that the model effectively estimates is 6.3% while exhibiting a slight overestimation bias of approximately 1%.

Much like the previous investigations, both the GBM and BBM vastly outperform the WBM. In this case, two WBM's are used in the comparative study in the associated operation: total shaft power and
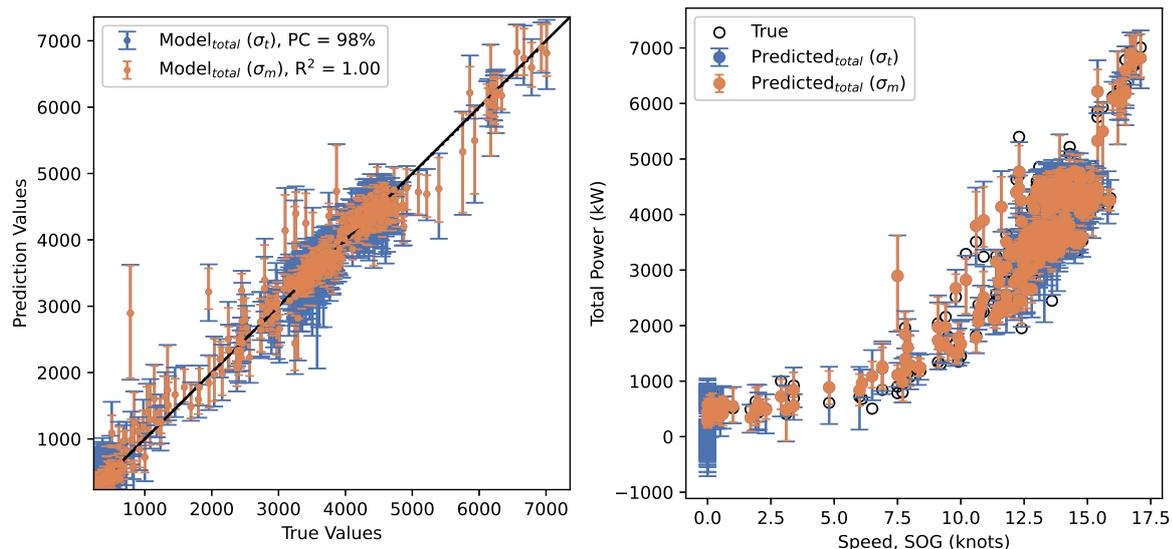
Figure 6.8: None Combined Total $GBM + P_{s,cw+hvac}$ prediction performance (left) and speed-power relationship (right) with 95%CI

calm-water power combined with the Load List estimation. While $P_{t,s} + LL$ WBM is the best performing approach, the recorded error is nearly 25% worse than the best performing GBM.

Interestingly, the total demand model behaves much like an average between the best-developed auxiliary and propulsion models. This behaviour is reflected in both the performance metrics and the overall uncertainty the model exhibits. The best-averaged auxiliary and propulsion MAPE and RMSPE are 6.3% and 8.2%, respectively. When comparing the two metrics to the actual Combined operational performance, a slight 1.0% global improvement is experienced. While this is favourable, it indicates that the model does not inherently improve each contribution but instead manages to retain much of the same limitations each part demonstrated. This comparison can be enforced with figure 6.8, where the uncertainty bounds of the pure auxiliary portion (zero speed) are much larger than the remaining regions. Ultimately, this signifies that although the Sailing operation dominates the modelling behaviour, the worse auxiliary performance is ultimately retained and reflected within increased confidence bands.

While this model seems to reflect the combination of behaviours from each individual component, some additional inconsistencies exist. This operation is the only case where the non-IQR outlier detection approach performed best. However, this can be related to the feature selection correlation study seen within section 6.2.4. Applying the outlier detection method reduces the Spearman Correlation between the two cases since many of the data feature distributions no longer exhibited a parametric shape. Thus the approach is not sufficient in that radically merging different operational datasets can impose problems on data orientation and structure. Nevertheless, the developed model provides total energy demand estimations well below the performance threshold of 15% while allowing a broad application range and usage.

6.3.4. Modelling Aggregation Comparison

Section 6.3 has demonstrated the model approach's ability to estimate energy consumption for various operational conditions accurately. However, one critical requirement is to proportion total energy demand into its individual components: auxiliary and propulsion power. While multiple models have been developed for various operational situations, estimations cannot be decomposed once fully developed. Unfortunately, this is a limitation of the ANN process, where a singular dependent target parameter is used to supervise the training process.

However, just because models cannot be decomposed does not mean that models can not be aggregated using a traditional bottom-up approach. As such, a comparison between the aggregated

propulsion and auxiliary models as well as the total demand model is investigated. Figure 6.9 highlights 12 training test samples under varying conditions that have been used unilaterally between the three models. Here, the actual target, aggregated model predictions, and total power model predictions are directly compared. Based on the results, it can be confirmed that the propulsion component ultimately
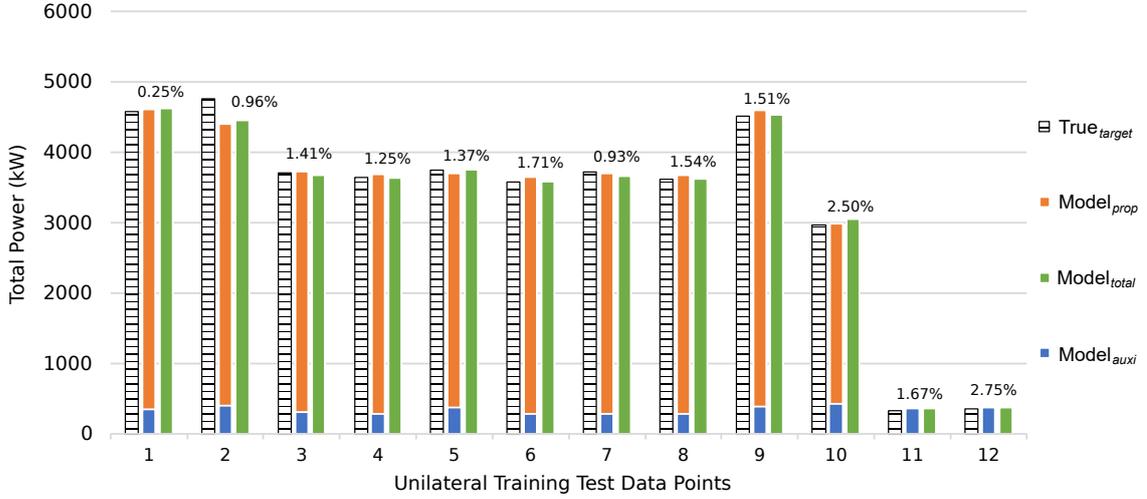
Figure 6.9: Comparison between Total ($GBM + P_{s,cw+hvac}$) and Auxiliary ($\Psi\emptyset\ GBM + P_{hvac}$) + Propulsion ($GBM + P_{s,cw}$) power demand

dominates the total load. Nevertheless, when both models are aggregated, the mean degree difference between each sample is marginal 1.5%. This deviation falls well below the mean error of each corresponding model and, as such, provides a suitable indication that a bottom-up GBM aggregation for system proportioning is a viable solution to capture each propulsion, auxiliary, and total powering demand. However, it should be noted that the corresponding training data ranges should always be considered when aggregating multiple models. In this case, the Combined Auxiliary ($\Psi\emptyset$) model is applied in conjunction with the Sailing Propulsion model. Thus, the limiting model is the propulsion model, as the bounds are significantly more narrow.

A second comparison between the evaluated modelling uncertainty can also be studied. Since each developed model comprises an ensemble of individual models, the uncertainty varies amongst the solutions. Thus, when aggregating models, the associated mean uncertainty must be combined using a standard propagation method.

$$\text{Average Aggregated Lower Bound: } \sqrt{85.66^2 + 253.86^2} = 267.92 \text{ kW}$$

$$\text{Average Aggregated Upper Bound: } \sqrt{98.38^2 + 250.38^2} = 269.01 \text{ kW}$$

Therefore, each developed modelling uncertainty contribution can be seen as follows,

$$\hat{P}_{prop}(+250.38, -253.86) + \hat{P}_{auxi}(+98.38, -85.66) = \hat{P}_{total,agg}(+269.01, -267.92) \tag{6.7}$$

When comparing these new bounds to the best-developed total energy consumption model, results are relatively closely aligned with a degree difference between lower and upper of 0.41% and 9.71%, respectively. It should be noted that while the upper bound presents a moderate degree difference, the propagated uncertainty only considers the mean outcomes and is only compared to ensure the orders of magnitudes are relatively aligned. Furthermore, the individual uncertainty regions within each model present a much deeper level of interpretability and understanding; thus, uncertainty application is highly dependent on the individual modelling capabilities. Ultimately, the uncertainty behaviour exhibited in the total and aggregated model indicates a high degree of similarity as both indicate that unique model characteristics are preserved. Therefore, it can be determined that successful proportioning by aggregating GBM models and their associated uncertainty bounds can be achieved. Thus, allowing for a broad application spectrum where the complete vessel energy demand can be inspected, decomposed, and compared under various operational conditions.

6.4. Modelling Verification and Validation

Section 6.3 has evaluated each operational condition via a direct comparison between all modelling categories: GBM, BBM, and WBM. In addition, each developed model was compared using the multiple performance metrics necessary to fully understand each model's potential and limitations. By having identified the best-developed models, a Validation and Verification analysis can be further conducted. As outlined and detailed in section 5.4.3, this procedure ensures that the acquired results are realistic and in line with the established modelling requirements. A detailed summary of the results and corresponding comparison results and outcomes are seen in table 6.9. It should be noted that although the auxiliary power BBM models were the best performing models, the GBM's had a near-identical outcome; as such, for consistency, all models use the GBM outcomes for consistent Validation and Verification.

Table 6.9: Modelling results validation and verification summary

Developed Models	RMSE	$\overline{P}_{target}$	nRMSE	MAPE	V&V
Sailing Propulsion $GBM + P_{t,cw}$	121.145	3461.0	0.0350	2.2%	✓
Anchor Auxiliary (ΨØ) $GBM + P_{hvac}$	35.094	381.8	0.0919	7.4%	✓
Combined Auxiliary (ΨØ) $GBM + P_{hvac}$	47.959	373.2	0.1285	10.3%	✓
Combined Total $GBM + P_{t,cw+hvac}$	120.343	1533.9	0.0785	6.3%	✓
Propulsion					
Zwart [108]			0.0842	6.63%	
Parkes et al. [83]			0.126	7.80%	
Bal Beşikçi et al. [11]			0.102	6.00%	
Pedersen and Larsen [84]			-	2.70%	
Average			0.104	5.78%	✓
Auxiliary					
Karatasou et al. [64]			0.024	1.50%	
Neto and Fiorelli [79]			-	16.50%	
Kalogirou and Bojic [63]			-	9.00%	
Average			0.024	9.00%	?✓

While the absolute relative errors are transparent in each literature investigation, using only one performance metric, as previously indicated in section 5.4.2, does not adequately capture the total modelling outcomes. As such, the additional RMSE is implemented. However, since this metric is a dimensional unit, a normalization of the parameter must be conducted to compare the various modelling solutions unilaterally. As such, the following relation can be applied,

$$\text{nRMSE} = \frac{\text{RMSE}}{\overline{P}_{target}} \tag{6.8}$$

Where the RMSE performance metric is normalized using the mean of the target parameters found within each literature investigation. While most literature sources are transparent in their results, not all reports have this metric listed. Thus, the average of the corresponding results is used as a comparison baseline for each operational condition.

First, looking at the verification component, the literature results can be compared directly with the developed models. It can be seen that the corresponding outputs from each operational case align within the same order of magnitudes of each literature result. The propulsion cases behave moderately better than most of the listed sources; however, this can also be attributed to the amount of data provided for model training. The auxiliary case also falls within the expected modelling outcome orders of magnitudes. However, in this case, it should be noted that the nRMSE is much higher than the average results. This deviation is attributed to the fact that only a singular solution entry was listed in the

associated literature. In general, Karatasou et al. [64] results exhibited excellent performance metrics. Unfortunately, the only comparative literature is found within building engineering case studies, as the maritime industry has yet to apply either BBM or GBM solutions to such investigations. Nonetheless, the remaining entries, while only a singular metric is extracted, exhibits similar magnitudes. Ultimately, the models behave as expected for similar cases; thus, they can also be considered verified in the context of model development.

The second criteria, Validation, is slightly more complex to evaluate. As detailed in section 5.4.3, each developed model is inherently tested and assessed using an independent holdout test dataset. These unbiased results are then evaluated and compared using the key performance metrics, as seen in section 6.3. However, to validate whether the correct model is built for real-world purposes and client expectations, the associated results must be measured against the established modelling requirements. In this case, the accuracy requirement ($< 15\% \pm 95\%CI$) is used. In all cases, it can be seen that the models exhibit excellent estimation ability. However, the complexity of validating the GBM solution lies within the test datasets. While these subsets indicate generalization and unbiased performance, they are only a portion of an infinite input domain. In other words, while the test set explores pieces of the data feature input range, it does not fully explore every input possibility. Thus, it is impossible to validate each model thoroughly. Nevertheless, based on the established validation criteria and the modelling requirements, the developed models can be considered well validated within their design range.

7

Model Application

The following chapter expands on the details and evaluations seen within chapter 6. Ultimately, the developed models were both validated and verified within the training stage. However, these conclusions were only drawn within the modelling working data ranges. As such, the full capabilities and capacity of each developed model are ultimately neglected. Thus, this chapter investigates the GBM deployment stage via an extrapolation and exploitation overview by addressing the following research question,

'What estimation capabilities does the proposed solution exhibit outside the design domain? And how can the total performance be leveraged to isolate and extract hidden relationships such as,

(a) *Fouling effects overtime on total propulsion power?*

(b) *Daylight cycle effects on total auxiliary power?'*

Section 7.1 outlines an extrapolation investigation using the best-developed propulsion and auxiliary models. This section provides insight into the relative performance of each operation via a direct modelling comparison between GBM, BBM, and WBMs using data entries outside the range of the developed input ranges. Section 7.2 further explores the GBM model's solution capabilities via an exploitation study. This investigation focuses on extracting and isolating key parameters such as fouling and daylight cycle contributions on the power demands. Ultimately, this section provides additional insight into the general versatility and modelling strengths of the proposed solution.

7.1. Extrapolation Investigation

As found within the literature investigation (see section 3.3 and 3.4), GBM and BBM models are highly effective within the developed model's training data ranges. This performance has been demonstrated in section 6.3, where a high degree of prediction accuracy has been verified and validated. However, it can ultimately be noticed that both the GBM and BBM behave similarly with one another. Thus, a question remains as to why a more complex GBM modelling solution would be favoured over the more straightforward BBM approach.

The goal of the GBM is to introduce an aspect of the foundational physics attained from the WBM to aid in the internal dependencies and mappings. It is hoped that this introduction of physics allows for the prevention of unreasonable results when nearing the extrapolation regions. However, the previous performance and comparative studies alone are inadequate to evaluate such conditions. Thus, to assess these unknown regions, data entries outside the training bounds must be applied to the developed models. At this point, the performance metrics, as outlined in section 5.4.2, can be used to evaluate each modelling category.

In the following sections7.1.1 and 7.1.2, both the best-developed propulsion and auxiliary model extrapolation capacities are evaluated. It should be noted that to evaluate the extrapolation capacity, information beyond the input data ranges is required. Unfortunately, this imposes a challenge for each

operational dataset since all available information was implemented to ensure the best possible models were developed.

One approach to overcome this challenge is the re-development of each modelling solution with a reduced data region and manually creating an extrapolation dataset with the remaining entries (Bakker [10]). However, this has the consequence of completely changing the model's dynamics and thus would require an entirely new data preparation and evaluation process. Luckily, this can be avoided by taking advantage of the IQR outlier detection's conservative nature. As alluded to in section 5.2.4, the IQR approach is an indiscriminate cleaning approach. While it eliminates extreme outliers, actual good operational data is often eliminated as well. Therefore, by investigating the residual IQR data, valid extrapolation points can be collected and applied to the best-developed models without inherently changing any internal dependencies.

7.1.1. Propulsion Power Performance

The residual IQR dataset must first be investigated and manually cleaned to evaluate the propulsion extrapolation capacity successfully. This additional preparation stage is required to separate the potentially valid entries from clear-cut outlier points found within the eliminated dataset.
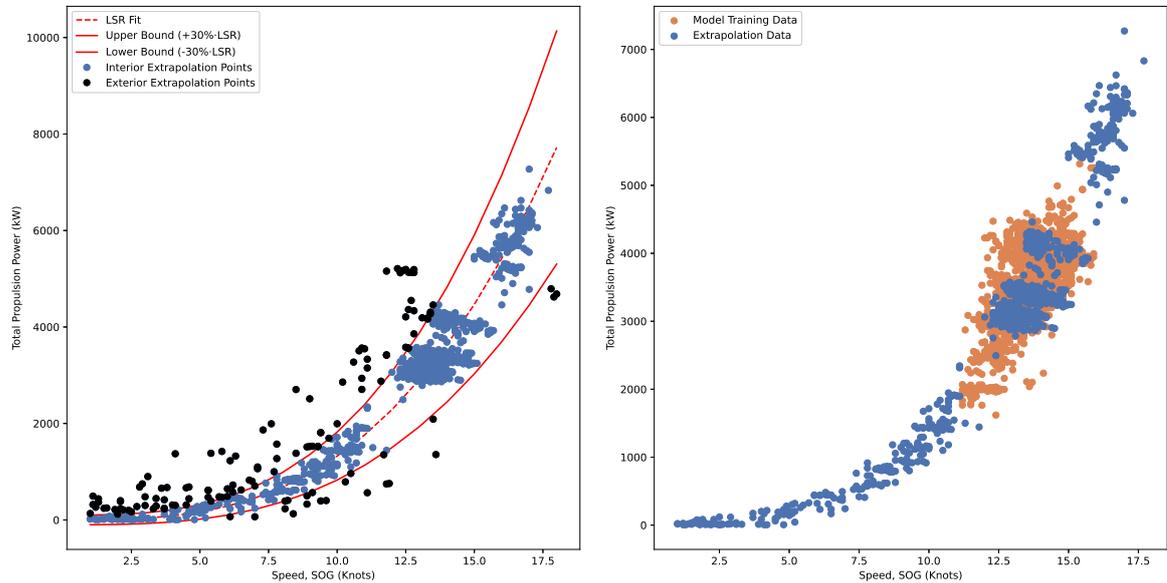
Figure 7.1: Residual IQR Sailing data reduction bounds (left) and training versus extrapolation dataset comparison (right)

The procedure to eliminate the existing outliers is done using a least-squares regression (LSR) methodology based on a conventional resistance-power cubic relationship as seen below,

$$P_{fit} = \frac{P_E}{\eta_{chain}} = \frac{R_T \, V_s}{\eta_{chain}} = \frac{C_E \, \rho^{1/3} \, \Delta^{2/3} \, V_s^3}{\eta_{chain}} = C_{fit} \cdot V_s^3 \qquad (7.1)$$

In this case, the unknown lumped constant, C_{fit}, is optimized to determine the optimal fit within the existing dataset. Once the LSR fit is obtained, an elimination bound strategy is applied. This approach utilizes a $\pm 30\%$ upper and lower bounding error domain to remove any extreme outliers. It should be noted that, while this strategy eliminates any severe irregularities, the set limitations are arbitrarily established. As such, outliers within each feature data set may still be found within the collection. Therefore, each data feature is investigated independently to ensure no extreme hidden entries are found in addition to the bound methodology. After preparation and elimination of the excess outliers, the propulsion training dataset and residual IQR set can be directly compared, as seen within figure 7.1. The results show that an extrapolation region above, below, and within the training set can be analyzed.

Once the residual data is prepared, the best-developed IQR Sailing models can be evaluated. The

propulsion extrapolation potential can be seen summarized in table 7.1. Additionally, the corresponding target and prediction relationships for each modelling solution can be seen in figure 7.2.

Table 7.1: IQR Sailing Propulsion extrapolation performance summary comparisons between GBM, BBM and WBM

	Sailing Propulsion - IQR				
Test Values	1039				
Perf. Metric	$GBM + P_{s,t}$	$GBM + P_{s,cw}$	BBM	$P_{s,t}$	$P_{s,cw}$
R^2	0.722	0.788	0.694	0.83	0.704
ME (kW)	262.016 (U)	214.277 (U)	328.704 (U)	-271.502 (O)	721.204 (U)
MAE (kW)	416.864	373.148	445.594	491.395	728.574
RMSE (kW)	835.68	729.215	877.31	654.338	862.575
	Percent Error				
MPE	89.042%	81.330%	356.547%	-25.151%	17.496%
MAPE	162.739%	126.162%	365.733%	33.940%	30.746%
RMSPE	26.543%	23.162%	27.865%	20.783%	27.397%
CI95% Lower	1203.1	1146.23	1093.91	-	-
CI95% Upper	2202.43	1991.77	2233.92	-	-
Cover%	93.74%	94.51%	93.36%	-	-
	Propulsion Model Percent Relative Change				
ΔR^2	-24%	-17%	-27%	+168%	+128%
ΔME	$O \rightarrow U$	$O \rightarrow U$	$O \rightarrow U$	$O \rightarrow O$	$U \rightarrow U$
ΔMAE	+470%	+413%	+506%	-18%	-20%
ΔRMSE	+590%	+508%	+617%	+175%	-13%

Figure 7.2: Propulsion extrapolation prediction performance direct comparison

Based on the results, the overall extrapolation capacity of each modelling solution can be directly compared and investigated. Ultimately, it can be seen that the WBM solution exhibits the best overall performance. This is to be expected, as one of the most significant benefits of using a WBM is its rooted fundamentals, allowing for improved extrapolation. However, the GBM is seen as the next best performing model, whereas the BBM is the global worst performer by a substantial margin. Therefore, it can be clearly observed that the GBM has successfully retained some physics induced by the integrated WBM. Ultimately, this can be related to the dependencies seen in section 6.2.4. The WBM data features all exhibited a high degree of correlation. Thus, solid internal mappings between the physics model and the BBM portion could be made.

Nevertheless, apparent drawbacks of these modelling solutions are exposed within the investigation. While the GBM does exhibit good relative performance, the global performance is substantially lower than experienced within the modelling training data ranges. For the GBM and BBM models, all performance indicators have shown a reduction. Not only has the relative performance decreased, but

the average uncertainty has drastically increased by an approximate average factor of six. This rise indicates that, although the mean performance is relatively okay, substantial variation amongst the ensemble models exists. In comparison, the WBM managed to improve some of its performance metrics, such as the general fitting and absolute errors.

However, arguably the most critical drawback is a direct consequence of the learning capacity and is highlighted within figure 7.1. In a physical sense, an estimation below zero is nonsensical in terms of power estimation and consumption. The WBM inherently manages this constraint and provides feasible and realistic estimates. However, this is not seen within the GBM and BBM components. As a result, although the model performance shows good results, unrealistic data estimations exist. The extreme numerical increase of the MAPE is a direct consequence of these non-feasible solutions, which indicate poor low power estimation capabilities. Therefore, while it is acknowledged that the GBM has managed to induce some physic within its internal structure, the GBM solution still demonstrates that substantial risk and uncertainty in the modelling behaviour exists outside of the training region.

7.1.2. Auxiliary Power Performance

As seen in section 7.1.1, the residual IQR dataset must first be investigated and manually cleaned to evaluate the modelling extrapolation capacity. However, in this case, the preparation procedure deviates quite drastically, as previously demonstrated. Within the propulsion operation, exploitation of the known empirical cubic relation could help guide the refinement process. However, no such clear relationship exists within the auxiliary power target parameter.
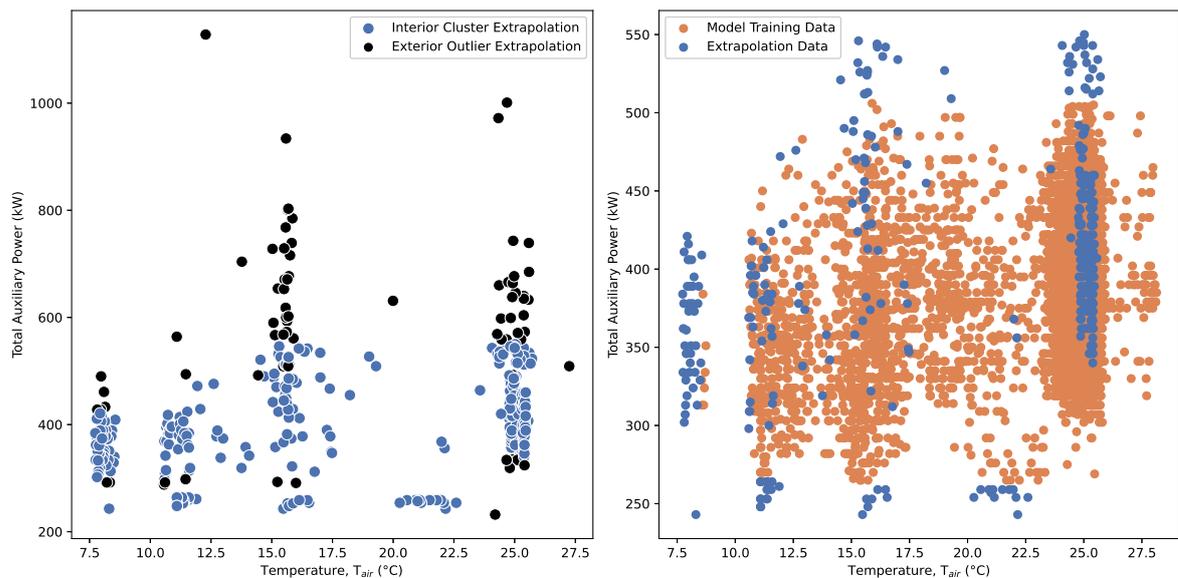
Figure 7.3: Residual IQR Anchor data reduction clusters (left) and training versus extrapolation dataset comparison (right)

An alternative method known as density-based (DB) clustering is applied to identify likely groupings. These groupings are based on proximity and dense regions in the data space, containing a minimum number of established points. Ultimately, the approach attempts to isolate likely sets with non-grouped outliers using each specific data point circumferential euclidean distances related to all neighbouring points. Fortunately, the method is commonly applied within computer science; thus, open-source packages within the Python environment allow for easy and seamless implementation, (Pedregosa et al. [87]).

The results of the DB clustering method can be seen in figure 7.3, where a default minimum of 15 nearest points is considered a unit. It should be noted that the modification of the minimum number of points can significantly influence the outcomes. Thus, this parameter was selected to obtain a moderate degree of extrapolation within the auxiliary power dataset. Much like the above scenario, this approach, while advanced, does not guarantee the complete elimination of all extreme outliers amongst the data

features. Therefore, a manual investigation of each feature was undertaken to eliminate outlier entries in addition to the clustering approach. After preparation and elimination of the excess outliers, the auxiliary training dataset and residual IQR set can be directly compared, as seen within figure 7.3.

Once the residual data is prepared, the best-developed IQR Anchor ($\Psi\emptyset$) models can be evaluated. The auxiliary extrapolation potential can be seen summarized in table 7.2. Additionally, the corresponding target and prediction relationships for each modelling solution can be seen in figure 7.4.

Table 7.2: IQR Anchor Auxiliary ($\Psi\emptyset$) extrapolation performance summary comparisons between GBM, BBM and WBM

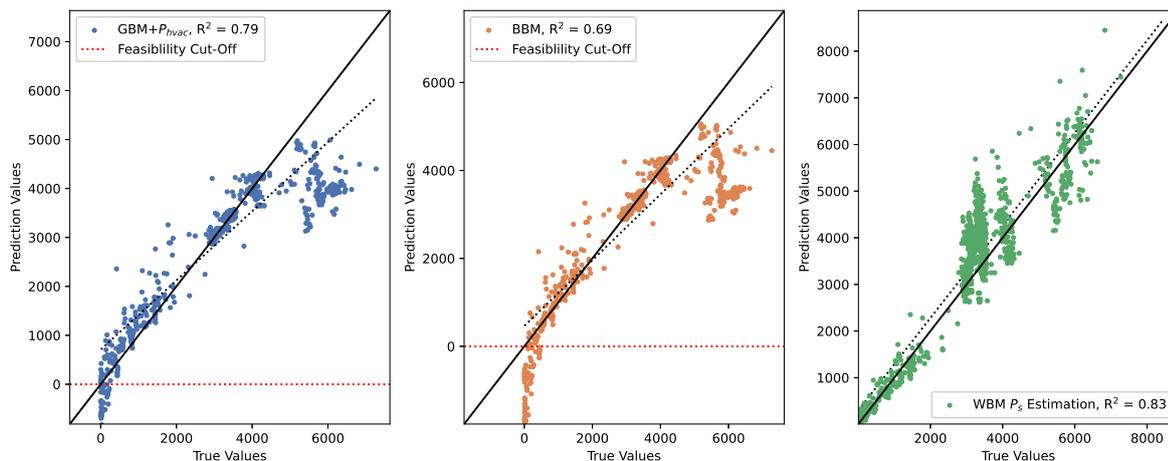| Anchor Auxiliary - IQR ($\Psi \neq 0$)				
Test Values	374			
Perf. Metric	$GBM + P_{hvac}$	BBM	$Load\ List$	P_{hvac}
R^2	0.285	0.284	-0.500	-13.58
ME (kW)	14.912 (U)	18.048 (U)	-51.337 (O)	269.31 (U)
MAE (kW)	47.627	48.55	72.369	269.31
RMSE (kW)	61.612	61.615	89.135	278.233
Percent Error				
MPE	1.338%	2.201%	-17.162%	67.310%
MAPE	11.938%	12.128%	21.213%	67.031%
RMSPE	15.455%	15.460%	22.358%	69.792%
CI95% Lower	102.84	103.98	-	-
CI95% Upper	127.44	121.41	-	-
Cover%	94.12%	94.39%	-	-
Auxiliary Model Percent Relative Change				
ΔR^2	-4%	-5%	+82%	+65%
ΔME	$U \rightarrow U$	$U \rightarrow U$	$O \rightarrow O$	$U \rightarrow U$
ΔMAE	+69%	+73%	+3%	+5%
ΔRMSE	+76%	+76%	+12%	+7%

Figure 7.4: Auxiliary extrapolation prediction performance direct comparison

Based on the results, the extrapolation capacity of each modelling solution can be similarly compared and investigated. It can be seen that, unlike the propulsion case, the auxiliary models behave radically differently. The GBM and BBM solutions, while a slight reduction in performance is seen, exhibit the best overall extrapolation performance. The WBM models, on the other hand, were inferior models to begin with, and as such, not much of a performance reduction be directly seen. On the other hand, the load list indicates a near-zero fitting, with an overall performance of approximately 21% error, whereas the GBM shows a 15% error. While the GBM and BBM models are superior, the outcomes are not much improved over a perfectly linear mean result. Thus, it can be inferred that the associated error of both models is optimistically low since the general relationship, while dynamically complex, is

not globally challenging to interpret (linear relationship). When comparing the GBM to the BBM, not many direct differences are observed, as illustrated in figure 7.4. It can be seen that ultimately, the associated prediction performance is closely mirrored between both modelling categories. As such, this alludes to the unfortunate conclusion that the GBM was not successfully able to retain much of the physics induced by the integrated WBM. While there is a slight improvement over the BBM model, the overall performance is relatively weak to begin. These mediocre results can be related to each feature's internal correlations to the target power parameter, as seen in section 6.2.4 and elaborated within section 6.3.2. While the WBM did consistently have the most significant correlation, the magnitude of these relations is much lower and spread out than the propulsion cases. Thus, it can be inferred that limited relationships could be adequately developed to aid the model's ability to extrapolate beyond its design ranges.

While it is unfortunate that the model seemingly did not retain much of the WBM physics, important insight into the GBM behaviour is achieved. Between the two extrapolation evaluations, it was demonstrated that when correlations between the WBM data feature and the target parameter are significant, inherent retention of the fundamental physics can be achieved. This learning ability allows for a clear improvement of general performance and nears the WBM's ability to extrapolate. However, when the correlations are low, no such learning is achieved, and the modelling behaviour performs similarly to a pure BBM. While this intuitively makes sense, it enforces the need for appropriate WBM's where the dynamic contributions instead of direct accuracy play the most critical role. Thus, the need for not just large data quantities, but also suitable data features that capture each target parameter's operational nature is required.

7.2. Exploitation Investigation

In section 6.3, the development methodology and performance evaluations demonstrated each GBM's exceptional capability to interpolate and predict within its input data ranges. Whereas section 7.2 has been confirmed that, under specific conditions, the GBM can learn inherent foundational physics for an improved degree of extrapolation.

However, one of the main shortcomings of using such a modelling solution is the limited interpretability. Luckily, this can be improved by taking advantage of specific input-target variable internal dependencies and effects. Ultimately, most WBM models are limited to a single-use application; however, BBM or GBM are not restricted in these regards. Thus, an exploitation analysis can be conducted by utilizing both the interpolation and extrapolation capabilities of a GBM solution approach to extract and isolate critical relationships.

In the following sections 7.2.1 and 7.2.2, both the best-developed propulsion and auxiliary models are used to exploit and isolate fouling and daylight cycle contributions to the power demand. It should be noted that various other data features can be investigated and varied to the target parameters for improved interpretability. However, such investigations fall outside the proposed investigation's scope (see section 3.8). Nonetheless, the subsequent study provides initial insight into the additional modelling capabilities and applications intrinsically available when applying a GBM solution.

7.2.1. Fouling Contribution

The fouling contribution to the total shaft power can only be considered when isolating the Hull cleaning interval (HCI) input feature. This separation can be achieved by investigating the effect varying HCI has on the model estimations while maintaining fixed constants for the other input features. The developed propulsion model (IQR Sailing) must be set up for success to ensure that reliable results are obtained. Since the developed relationships can exhibit complex non-linear behaviours, the associated input parameters can significantly influence the overall prediction outcomes. Thus, the remaining parameters should be fixed to lay in a region in which the model is known to exhibit its best prediction performance. As such, the median of each corresponding data feature is used to ensure each additional feature falls well within the data limitations and modelling training ranges. The corresponding limits for the following model can be seen in Table E.16.

It should be noted that the associated WBM estimation is evaluated using the selected data feature inputs, as each associated combination does not necessarily result in a median output. Nevertheless, the corresponding exploitation results when considering a hull cleaning interval over two years can be seen in figure 7.5.
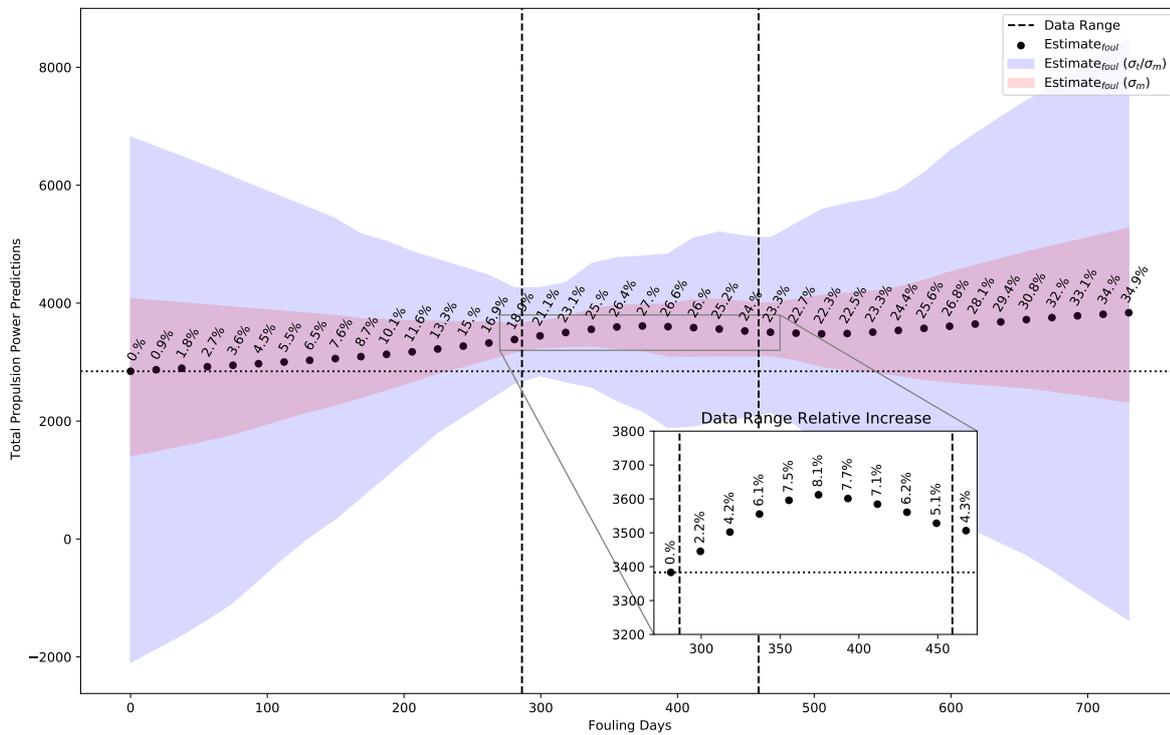
Figure 7.5: IQR Sailing Propulsion power increase due to fouling contribution

Based on the results, the isolated fouling contribution can be directly estimated in terms of total propulsion shaft power as well as a percent relative increase to the zeroth day contribution. Based on these estimations, it can be seen that over the course of two years, estimated growth of 35% is expected. However, the modelling training range is limited to between 280 to 450 days since a known hull cleaning has occurred. This range limitation indicates that the majority of the estimations fall within an extrapolation region. Therefore, while the model evaluations indicate excellent performance within the ranges, extrapolation studies, as seen in section 7.1.1, has shown potentially large deviations. This uncertainty is reflected within the estimated modelling and inherent error confidence bands. Furthermore, as the estimations become further from the training range, the overall confidence bands quickly increase. As such, concrete conclusions cannot be drawn.

Nevertheless, the interior region, although limited, does allow for a general performance within 5% of the actual operational results (see section 6.3.1). Therefore, the relevant region can be magnified and investigated. Here, it is observed that an increase of 8.1% can be seen over 95 days. However, beyond that point, a dip in the fouling contribution occurs, whereas, in a global sense, the trend seemingly exhibits something of a 3rd order fitting. Intuitively, it is assumed that without any hindrance, fouling continues to increase the power consumption over time, as opposed to decreasing. To understand and identify this irregularity, the recorded propulsion power can be directly analyzed via time-series groupings. The associated study results can be seen highlighted in figure 7.6.

By decomposing all recorded propulsion results within monthly groupings, individual least-squared
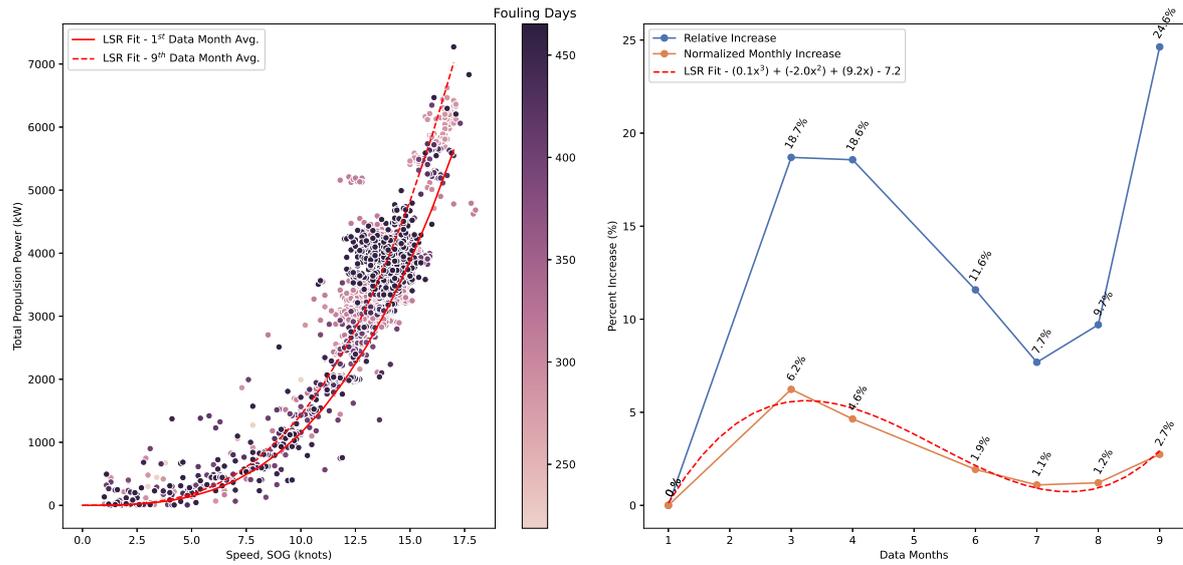
Figure 7.6: Recorded power-fit trend variation over time (left) and recorded monthly and normalized percent increases (right)

regression fittings using the Admiralty relation can be applied as,

$$
\left.\begin{array}{l}
P_0 = \dfrac{\Delta^{(2/3)} \cdot V_s^3}{C_0} \\[3mm]
P_x = \dfrac{\Delta^{(2/3)} \cdot V_s^3}{C_x}
\end{array}\right\}
\quad
P_x = \left(\dfrac{C_0}{C_x}\right) \cdot P_0
\tag{7.2}
$$

Since each monthly time-slice formulation considers a constant displacement (Δ) and velocity (V_s), the associated constants can then be compared directly. Thus, a relationship between the relative increase in the fitting constants for each consecutive month, *x* related to the initial month *0*, can be directly plotted. These percent increases are then normalized with the corresponding number of months since a hull cleaning has occurred to evaluate the monthly contributions per time grouping. Ultimately, the shape of the trend is quite like what is learned by the developed-model estimations.

There are multiple reasons for the existence of such an irregular relation. As noted by Uzun et al. [103] and Zwart [108], the fouling growth rate on hulls and propellers is highly dependent on the geographical regions. Based on the latitude and longitudes indicated within figure 7.6, it is known that the associated dataset considers multiple diverse environmental regions. As such, the associated external conditions play a prominent role in the effectiveness of the fouling contributions. As the longitude approaches the warmer equator (0°), the monthly fouling contribution is relatively significant; however, as the vessel climbs to the more northern and cooler regions, the normalized percent increase decreases, respectively. As such, an apparent connection between geographical location and fouling growth is observed.

However, other factors can also play a prominent role. While no hull cleaning has been recorded, undocumented propeller cleanings may be present. Such cleanings are likely to demonstrate a reduction in the fouling contributions. Although there is no clear indication, the suggested dip in the 4th data month may indicate an unrecorded external cleaning of sorts. Unfortunately, such a lack of data availability may limit the modelling usage as the developed model considers only one component of many, contributing to the fouling contributions.

While these observations are merely conjectured to establish apparent links, it does not take away from the fact that the inherent developed HCI-power relationships are traceable within the collected dataset. As such, it stands to reason that the following solution approaches can develop inherent relationships amongst smartly chosen input features. Unfortunately, while this is a highly desirable trait, it allows introduces significant associated risk. For instance, if the quality of the gathered results is incor-

rect, critical features are unknown, or multiple operational conditions are merged, the interpretability of the associated results becomes impossible to distinguish and decompose. Thus, great care must be considered when implementing such exploitation techniques. Ultimately, the GBM-ANN is a powerful learning tool; however, what it sees, is what you get.

7.2.2. Day Cycle Contributions

Similar to the fouling exploitation, day cycle contributions can only be obtained when isolated the sunlight factor (α_{sun}) input feature. However, unlike the previous input, which was a continuous variable, the sunlight factor is a binary classification feature that varies between night and day. As such, the developed relationship will directly investigate the auxiliary power contribution associated with each classification at various temperatures. As indicated in section 7.1.1, the developed auxiliary model (*IQR Anchor* $\Psi\emptyset$) must be set up to ensure that reliable results are obtained. Therefore, all remaining parameters are to lay in a region in which the model is known to exhibit its best prediction performance. As such, the median of each corresponding data feature is used to ensure each additional feature falls well within the data limitations and modelling training ranges. The corresponding limits for the following model can be seen in table E.17. The corresponding exploitation results when considering a binary day cycle and variable temperature ranges can be seen in figure 7.7.
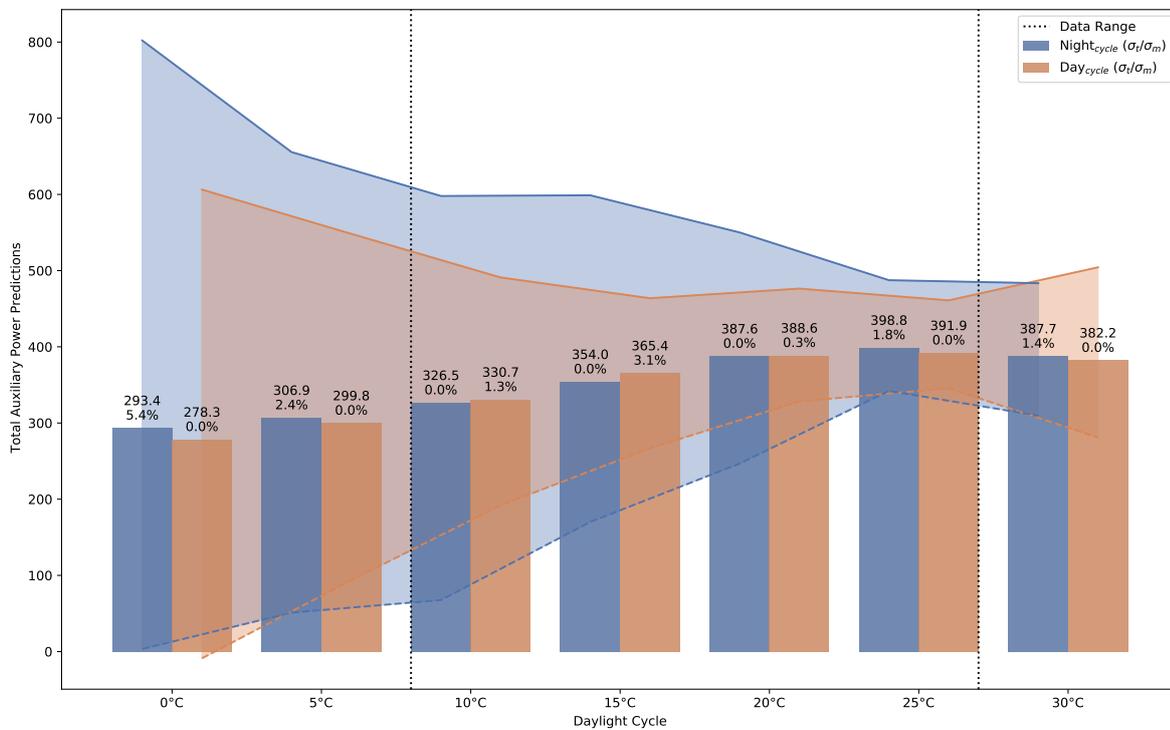
Figure 7.7: IQR Anchor ($\Psi\emptyset$) Auxiliary power comparisons between daylight cycle contributions for varying temperatures

Based on the results, the isolated day cycle contribution can be directly estimated in terms of total auxiliary power and a localized percent relative difference between the two classifications for each corresponding temperature. The considered temperature ranges between 0°C to 30°C, respectively. In a global sense, each factor has the same general parabolic shape. Therefore, as the temperature increases, the associated power required to maintain the indoor set temperatures also increases. Intuitively this makes sense when only considering the HVAC powering component; however, the challenge with this investigation is that the corresponding power is the total auxiliary and not the individual HVAC proportion. As such, multiple other systems are also included with the estimation. While this is difficult to interpret directly, relative comparisons between the multiple cases can provide a great deal of insight into experienced heat transmission and radiation effects.

Auxiliary power contributions due to heat transmission can be determined when solely investigating a single classification for each associated temperature. This comparison indicates the expected powering increases between various external conditions while maintaining fixed parameters for all other input features. In contrast, the radiation component can be directly investigated between the degree difference of both daylight factor classifications. It should be noted that while this can give a powerful indication of the relative contributions of each component, considerations such as equipment loads and onboard persons are unknown and thus can inherently alter the overall contribution proportions. Nevertheless, operational insight into such behaviours can be beneficial within initial design studies.

Unfortunately, a high degree of uncertainty is observed in the determined confidence intervals. Within figure 7.7, it can be seen that even within the data training regions, the uncertainty bounds substantially increase as the temperature decreases. Thus, an investigation into the data collection can be conducted to ascertain whether the model has learned any sensible relationship between the corresponding input features. The dataset total auxiliary power contributions for varying temperatures can be seen in figure 7.8. The associated results can be seen summarized in table 7.3.
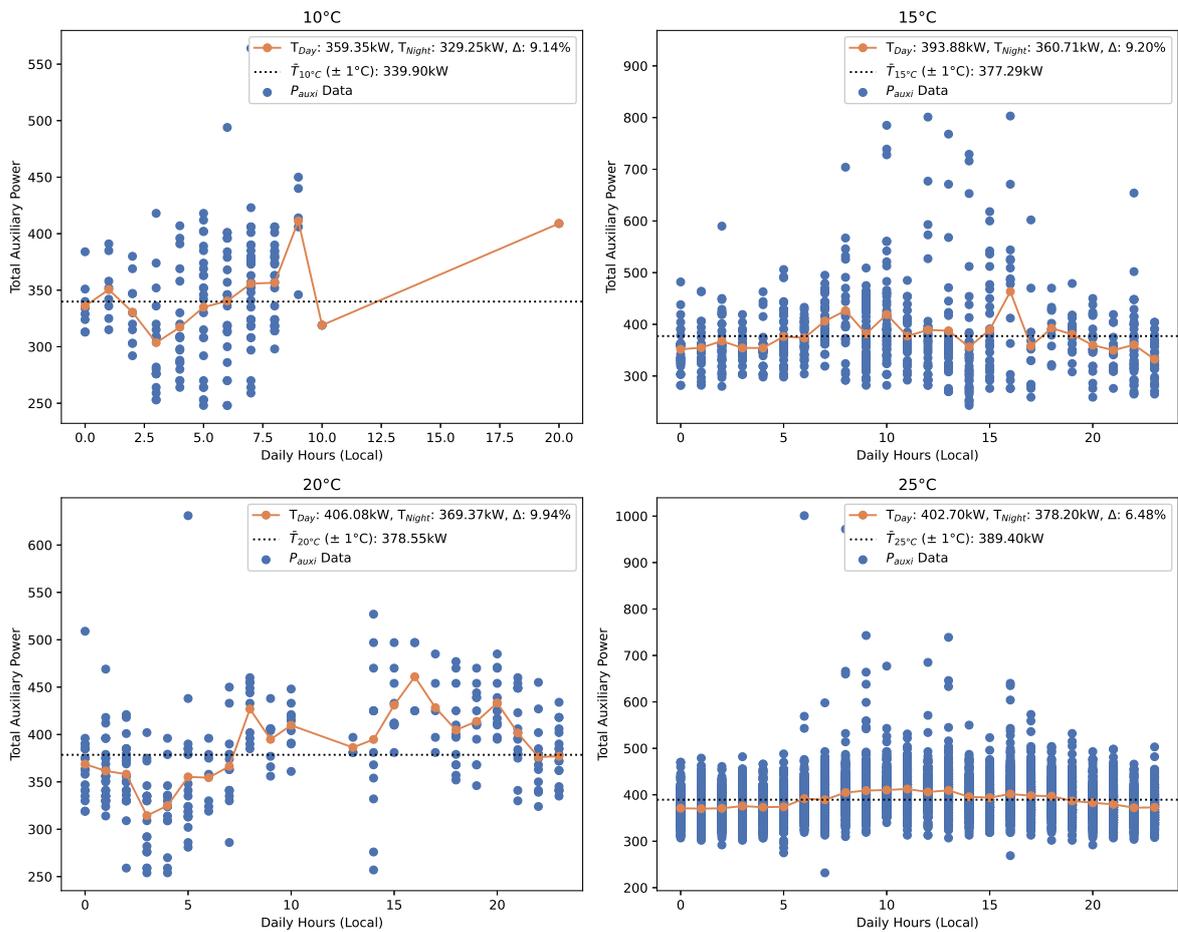
Figure 7.8: Recorded local hourly auxiliary power contributions for varying temperatures

It should be noted that the full *IQR Anchor* input data temperature range was considered. However, to ensure enough data per temperature was investigated, a $\pm 1°C$ bandwidth is considered. Additionally, to make sure the corresponding time is accurately mirrored in the various regions, a time-zone conversion between UTC and local time is undertaken, as detailed in section 6.2.1.

As seen from the operation dataset, as the temperature decreases, the overall sparsity of the dataset begins to decrease. In the lowest temperature region, 10°C, the collected data does not sufficiently cover the entire day cycle spectrum. As such, a direct comparison between this temperature zone is insufficient as the results are highly skewed. Ultimately, the irregularity in the highly increasing

uncertainty can be correlated to the associated data scarcity within the dataset. Nonetheless, when comparing the other regions to the estimated predictions, it can be seen that model behaves well. In each instance, the category with the highest contribution is mirrored in both the estimations and the collected results. Additionally, it can be seen that the mean absolute degree difference between each valid contribution is a mere 3.9%. While this estimation error is relatively minimal, most results remain highly uncertain, and therefore great care and consideration must be used when attempting to exploit such a model.

Table 7.3: Daylight cycle exploitation results comparison and summary

Day Cycle Contributions									
$GBM + P_{hvac}$			Dataset			Model Comparison ($E = 9.2\%$)			
T_{air}	P_{night}	P_{day}	$\%_{inc}$	P_{night}	P_{day}	$\%_{inc}$	Category	ΔP_{night}	ΔP_{day}	$\Delta_{\%inc}$
0°C	293.4	278.3	5.4%	-	-	-	-	-	-	-
5°C	306.9	299.8	2.4%	-	-	-	-	-	-	-
10°C	326.5	330.7	-1.3%	359.4	329.3	9.1%	Day/Night	-9.1%	0.4%	10.4%
15°C	354	365.4	-3.1%	360.7	393.9	-8.4%	Day/Day	-1.9%	-7.2%	-5.3%
20°C	387.6	388.6	-0.3%	369.4	406.1	-9.0%	Day/Day	4.9%	-4.3%	-8.8%
25°C	398.8	391.9	1.8%	402.7	378.2	6.5%	Night/Night	-1.0%	3.6%	4.7%
30°C	387.7	382.2	1.4%	-	-	-	-	-	-	-
MAPE [1]								2.6%	5.1%	6.3%

Unfortunately, as seen in section 6.3.2, the developed model has a MAPE and RMSPE of 7.2% and 9.2%, respectively. While this amount is sufficient regarding the modelling requirements, the degree difference between classifications falls near or below this error range. Therefore, while not explicitly displayed in the following study, the model has a higher potential to indicate incorrect classification features. Nevertheless, the corresponding research indicates and enforces the GBM's ability to learn inherent relations among the various input data features. As such, the general application of such an approach immensely increases, which ultimately allows for a high degree of flexibility and creativity in usage as long as modelling error, uncertainty bounds, and data range limitations are fully considered and understood.

7.3. General Remarks on GBM Performance

Chapters 6 and 7 investigated the interpolation and extrapolation potential of the GBM approach. Ultimately, a detailed comparison of the outcomes for total energy demand and subsequent sub-powering categories (propulsion and auxiliary power) has been presented for each modelling category. The corresponding outcomes are seen as,

- Interpolation: $GBM \geq BBM > WBM$
- Extrapolation: $WBM > GBM \geq BBM$

These conclusions have ultimately been reflected within similar GBM studies to a certain degree. Bakker [10] thoroughly investigated modelling data gap performances and the extrapolation potential of GBM using a sub-BBM random forest regression (RFR) approach. While the study focused on lightship weight estimations, the general regional performance conclusions indicated similar outcomes. The GBM (as well as the BBM) performed well within the interior data ranges. However, when extrapolation capacity was investigated, the proposed RFR approach proved inadequate compared to the WBM. Nevertheless, the GBM performed better than its singular BBM counterpart. Unfortunately, the inherent differences between ANNs and RFR make further direct comparisons a challenge.

---

[1]Only considers the viable dataset regions of 15°C , 20°C , and 25°C

The Zwart [108] study most closely aligns with a portion of the current work as propulsion power estimations using GBMs with artificial neural networks (ANN) were investigated. However, instead of using the results to decompose the total energy demand, an input-output relationship between trim-power was investigated for operational performance. Ultimately, the interpolation results similarly indicate good performance (see table 6.9); however, an extrapolation investigated was never performed. As such, the advantages of a GBM over a conventional BBM were not proven within the study. As such, conclusions regarding the GBM-ANNs extrapolation capabilities can not be compared, and the advantages of preventing unreasonable results and using less historical data than a pure BBM are only inferred from alternative studies (see Bakker [10], Leifsson et al. [68]).

Nevertheless, based on the current work, the investigation of the interpolation, extrapolation, and exploitation capacity indicates that the GBM approach, while not a perfect solution, demonstrates a great deal of potential. However, it is acknowledged that further investigation into the extrapolation capacity is required. While the initial conclusions suggest that WBM can provide improvements, dedicated manual data range augmentation should be applied to identify the potential of both internal sparse data gaps and exterior data-bound extrapolation ability of artificial neural networks.

8
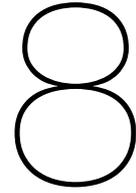
Design Process Implementation

All chapters and sections up until this point have focused on the general case objective of energy estimation, which highlighted the many modelling solutions and limitations. However, a big question remains about how and when these solution approaches can be implemented within complex design processes. Therefore, this chapter aims to provide a practical overview of Simon Sinek's Golden Circle [97] concept, focusing on the *Why, How, and What* of data-driven design modelling and process implementation by answering the following research question,

'How can the proposed solution approach be integrated within conventional ship design processes, and what criteria must be considered for successful implementation?'

Section 8.1 highlights the relevancy of the proposed solution, providing a clear inspiration for Naval Architects to consider such methods within conventional design processes. In addition to motivation, Section 8.2 illustrates how and when these solutions should be potentially applied. Here, the key considerations the Naval Architect needs to consider are outlined and addressed. Thirdly, section 8.3, details on future application opportunities. This section elaborates and focuses on the traditional design spiral design approach commonly applied within the Maritime and Yachting industry. Ultimately, it seeks to provide insight to Naval Architects about usage and application within a conventional design procedure. Finally, section 8.4 provides awareness of the sociological and ethical concerns inherent within machine learning applications. While technical limitations are presented throughout, data security and privacy concerns must also be considered and handled responsibly.

8.1. Why?

The *Why* stands for the purpose, such as what is the motivation behind any action. In the corresponding thesis investigation, this can be related to,

Why is the Grey-Box or Black-Box Modelling approach relevant, and why does it matter to Naval Architects?

In the last decade, the area of computer learning and data-driven modelling to improve and optimize vessel efficiency has increased significantly. This rise is mainly attributed to the considerable sustainability shifts across the globe and the ever-increasing technological advances allowing for easier availability of big data via advanced measuring systems.

As suggested by Gougoulidis [46], traditional calculation methods used in naval architecture and marine engineering mainly depend on statistical and regression analysis. Ultimately, the purpose of the Naval Architect is to produce the most efficient ship design possible. Unfortunately, these methods are often limited in capturing extremely complex non-linear problems where little information about real-life relations between inputs and outputs exists. Furthermore, using all the knowledge available in mathematical modelling of hydrodynamic phenomena for ship design is a highly complex task. Therefore, data-driven approaches such as Grey-box or Black-box modelling are becoming increasingly

favourable. Amarel and Steinberg [6] further suggests that the direct design benefits can be found in the following areas,

- Ability to adapt quickly to operational changes and technological improvements.
- Ability to rapidly transition from design concept to operational digital-twin prototypes,
- Improved quality of designs and lower life-cycle costs, and;
- Increased insight into design options and their hydrodynamic environment.

Therefore, the effect on committed cost, design freedom and the gathered problem knowledge through the design stages can be seen highlighted in figure (8.1). Ultimately, the GBM/BBM solution approach allows for increased problem knowledge as early as possible, increasing design freedom and reducing the overall committed cost. Thus, the risk in the design process can be reduced, and the process as a whole is made more efficient and cheaper.
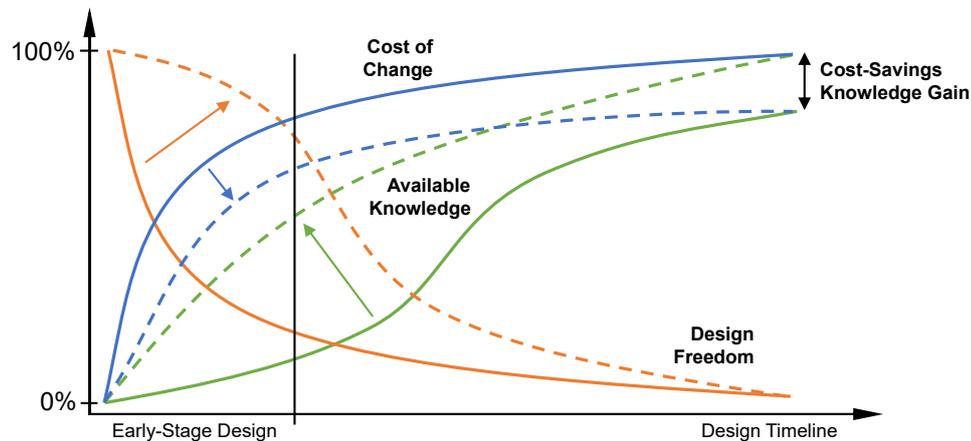
Figure 8.1: Solution approach influence on design cost-knowledge-freedom dependence (based on Mavris et al. [74])

Furthermore, both the Grey-box and Black-box models leverage advanced machine learning algorithms, such as neural networks, to tackle complex multi-variable problems. However, as the studies indicated in chapters 6 and 7, the GBM solution approach has the potential to exhibit several advantages over the pure BBM, such as,

- Adjustable and can improve performance when new data become available
- Excellent interpolation prediction tools within the trained data ranged
- Potential to inherently learn existing physics-based on integrated WBM
    - Allows for improved extrapolation capacity over pure BBM
    - Allows for a lesser amount of training data to perform at the same level of accuracy [10]
    - Allows for a greater degree of interpretability based on residual differencing

Ultimately, based on these advantages, such solution approaches make them highly attractive within ship design processes to aid naval architects and, as such, should provide sufficient motivation to genuinely explored to their full capabilities.

8.2. How?

The *How* indicates the actual process. It concerns the actions that have been taken to (be able to) realize the *Why* component. However, just because we can implement such approaches does not necessarily mean we should. Therefore, to address the *How*, we must also consider the *When*, which can be related as,

How and when should the Grey-box modelling or Black-box modelling approach be implemented within a design process?

To successfully answer the above question, a general decision-making process is implemented to give a clear road map of the necessary considerations for use, as shown in figure 8.2. Before the idea of modelling implementation and application, the problem must first be decomposed entirely and defined. This problem definition is critical in understanding the details before any design study. As such, this portion is typically involved,

- What design phase or level of fidelity is required?
- What information and tools are available?
- When is this information available?
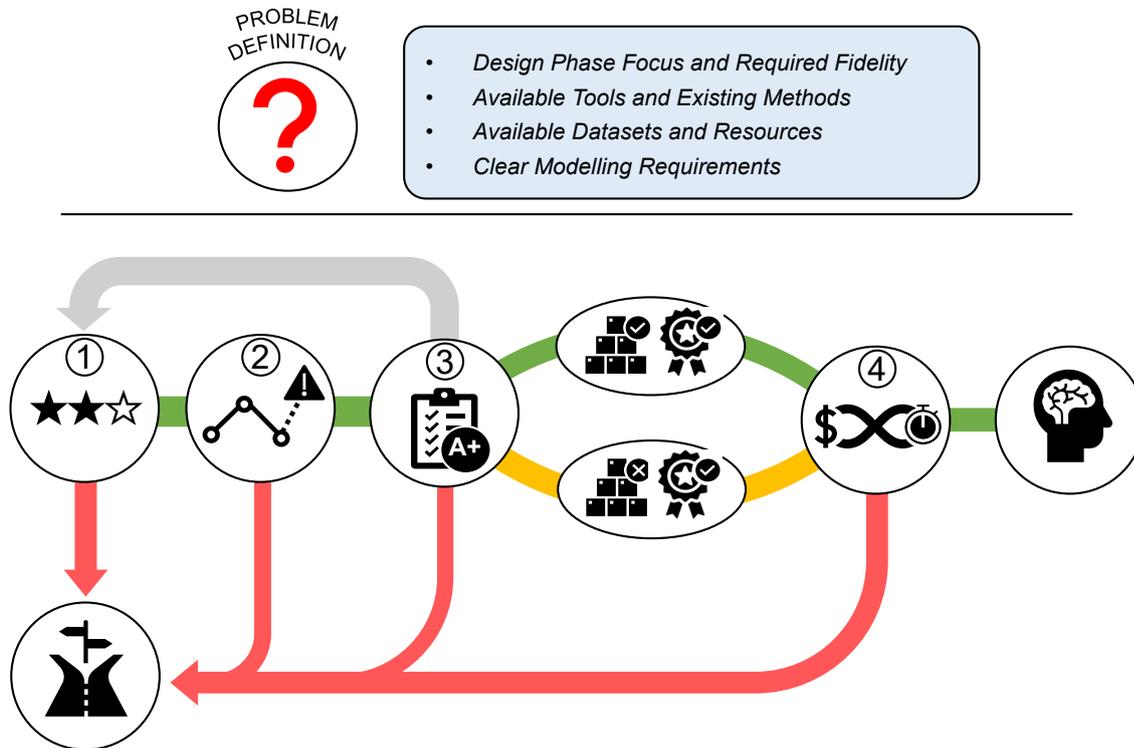
Figure 8.2: General solution approach considerations for successful design implementation

By addressing these three initial considerations, clear and transparent modelling requirements can be developed. These requirements encompass the foundation of any modelling investigation, and therefore must be carefully considered. Chapter 2 provides a detailed look into how the corresponding problem has been initially investigated and converted into detailed modelling requirements. Once established, each critical step can be detailed as follows,

1. *Is the problem highly complex?*

   If the defined problem can be solved simply with alternative methods, solution approaches such as GBM or BBM should not be implemented.

   Ultimately, these solution approaches are beneficial techniques when determining functional patterns between independent and dependent variables, even when no such pattern is discernible using conventional methods. However, the overarching complexity of the methods can significantly outweigh the complexity of the solution. However, if the proposed problem is highly complex, non-linear, or requires a high degree of estimation accuracy, the GBM or BBM maybe be a reasonable approach. Therefore, it is recommended first to investigate the application based on existing literature of such solution approaches. Considerations such as similar data dimensions, complexity, characteristics, and even expected performance should be studied. This study should provide a preliminary indication of whether or not the established requirements can be achieved.

If no literature or reference is available, advice from experts should be sought.

2. ***What region of the design space is to be investigated?***

There are three modelling categories: WBM, BBM, and GBM. Each has its strengths and weaknesses in regards to interpolation and extrapolation performance. Based on the investigation, the following conclusions can be drawn for each criterion.

- Interpolation (Section 6.3): $GBM \geq BBM > WBM$
- Extrapolation (Section 7.1): $WBM > GBM \geq BBM$

Therefore, if the data range fully covers the feasible design space, GBM or BBM can be useful as the modelling falls within an interpolation region, where excellent prediction performance is exhibited. However, if the feasible design space falls outside of the modelling data ranges, these approaches should not be implemented. While the GBM does suggest some improvements over the conventional BBM, entirely feasible solutions are not guaranteed. In contrast, the WBM, while not necessarily the most accurate, is much more reliable. It should be noted that the study indicates that the GBM performs equal or better than the BBM approach in all cases. As such, it is recommended that the former is applied over the latter. However, benefits are only obtained if a corresponding WBM is available and sufficiently captures most of the dynamical contributions of the targeted dependent variable. Nevertheless, if such a model does not exist, a pure BBM approach can still be applied as the technique is a mere subset of the GBM.

3. ***Is there sufficiently 'good' data available?***

GBM or BBM solutions require large amounts of data before they even begin to output valuable results. However, not all data is useful and can even harm the modelling capabilities and accuracy. Data augmentation and preparation methodologies do indeed help to eliminate noise and outlier entries. However, the corresponding solution approach might begin to memorize these irregularities if there are too many erratic entries. Thus, general modelling capabilities are not only highly dependent but limited to the original quality of the source. As such, having more data is always the preferred solution; having quality data is a paramount priority.

Unfortunately, for each application quantity and quality of each associated dataset can vary drastically. Therefore, it is nearly impossible to introduce a lower bound for each criterion. Therefore, the following considerations can be used to determine whether available data is suitable,

- *Is GBM or BBM common for similar tasks?*
- *Has GBM or BBM been used for very similar datasets, or is your dataset comparable?*
- *Is it good enough if you achieve comparable results?*
- *If not, is your dataset larger/better/broader or are you an expert?*

It should be noted that a great deal of overlap exists with the first step. If any of the corresponding considerations are not met, it does not necessarily mean that such solution approaches are inadequate. However, a re-evaluation of the existing literature should be undertaken to ensure the above limitations are still in line with the established modelling requirements and expectations.

4. ***Is there sufficient time and relevant resources available?***

The machine learning component within both the GBM and BBM is heavily time- and resource-intensive. The process is an iterative learning procedure that requires continuous collection, monitoring, and adjustments to ensure the developed solutions are performing at their maximum capacities. Such a methodology is proposed in chapter 5. While the study demonstrated the general success of the approach, the continuous iteration and tuning of the models were neglected due to time constraints. Nonetheless, this iterative learning cycle is a critical component and ultimately exhibits the underlying strengths of such approaches.

The relevancy of the available datasets is also a critical consideration. It does not matter how 'good' the available information is; if the operational situations are vastly different from prior datasets, it is irrelevant. Therefore, in addition to continuously updating models with recent data, elimination and removal of irrelevant entries should also be conducted to ensure data significance is maintained.

If all criteria are met, the application of such methods is seen as promising. However, even when a straightforward procedure is applied, these methods do not always behave as expected. As such, the Naval Architect must additionally consider and weigh the benefits with the negatives concerning expected outcomes and urgency of the results before implementation.

8.3. What?

Finally, *What* stands for the results and the outcomes of *Why* and *How*. This criterion can be related to,

What applications in a typical design process can the GBM or BBM be incorporated by a Naval Architect?

Traditionally, the ship design process can be seen as a highly iterative cycle, often represented as a spiral-like refinement procedure. However, as noted by Gougoulidis [46], while this point-based design approach has evolved and adapted over time, the design principles and steps typically remain the same. Therefore, as seen within figure 2.2, the many design elements can be identified and orientated towards a GBM or BBM application within the traditional design sense.

- *Principal dimensions:* The first stage of the design analysis is the determination of the principal dimensions of a ship. Clausen et al. [25] investigated and completed applications of this phase by using simple independent inputs such as capacity, breadth, or top speed. The corresponding dependent outputs would estimate the main particulars: length, breadth, draft, and displacement. As remarked by Gougoulidis [46], it was found that when comparing the approaches with conventional regression techniques, slightly better estimations were acquired because of the model's inherent flexibility and adaptability.

- *Hydrostatics (seakeeping):* Initial seakeeping characteristics have also been investigated. In this case, Cepowski and Szelangiewicz [22] used basic ship dimensions such as breadth, draft, metacentric height, and B/L ratios as inputs. The corresponding outputs were roll, pitch, and heave frequency transfer coefficients. Gougoulidis [46] noted that while each component could be determined, various models with varying degrees of complexity were required. Nevertheless, an extremely high degree of accuracy was recorded as compared to complex and exact numerics.

- *Powering:* Powering and resistance is arguably one of the most studied applications and is the focus of the corresponding investigation. The performance results seen in chapter 6 have demonstrated the applicability and feasibility of this design component application. Additional insight into alternative powering applications for various uses can be found within chapter 3.

- *Weight estimate:* Initial weight estimations have also been performed and investigated. Bakker [10] used reference-vessel principal characteristics and corresponding regression-based weight estimates to predict vessel lightship weight. It was ultimately determined that, while performance was heavily influenced by the relevancy of data and WBM models, the GBMs performed very well compared to traditional approaches.

While not explicitly outlined, features such as general arrangements, structures, costs, stability, and holistic design analysis have also been investigated and studied with success. Therefore, the potential of such approaches is limited only to the creativity of the Naval Architect. However, as noted in section 8.2, just because an approach can be applied does not necessarily mean it should. Nevertheless, after carefully considering the *Why, How, and What* of design process implementation, solution approaches such as GBM or BBM can greatly aid future Naval Architects in modernizing future ship design to tackle

the complex sustainability challenges. Therefore, to solve the problems of tomorrow, we need the tools of today.

8.4. Sociological and Ethical Challenges of Machine Learning

The current study has extensively elaborated and demonstrated the advantages and technical challenges the GBM or BBM face. In addition, three criteria were considered and elaborated to aid Naval Architects in understanding the *Why, How, and What* of potential design process implementation. However, each solution's inherent machine learning components also introduce natural sociological and ethical concerns that must be carefully considered and understood to ensure the benefits outweigh their disadvantages. One such problem that has a direct impact can be found in the form of data privacy.

Machine learning allows us to extract information from data, discover new patterns, and turn seemingly harmless data into sensitive, personal records. Depending on the available data features, models can be exploited, as proved in section 7.2, in potentially unintended ways, which can significantly risk client security and trust. Unfortunately, the lack of transparency in how these models use information raises many data privacy concerns (European Parliamentary Research Service [36]). For instance, patterns related to fuel use, vessel location and even grey- and black-water tank levels can help estimate overall vessel performance; however, it can also indicate whether guests are onboard or not. Unfortunately, applications based on machine learning require large amounts of data, but data subjects/clients have limited rights and understanding over how their data is used. As such, considering the sensitive and exclusive nature of the Yachting industry, significant repercussions for both privacy and anonymity infringements can exist.

Ultimately, the ethical and societal implications of data and machine learning are broad and complex. These tools have the ability to influence society directly and, as such, require a high degree of moral responsibility and understanding, not just of technical capabilities but also of the other implications such as privacy. Nevertheless, such tools can significantly benefit future ship design processes within the Maritime and Yachting industries. While it can be acknowledged that ethical and social concerns exist, through awareness, transparency, and regulation, the risks can be mitigated to leverage the tremendous benefits.

Conclusions, Contributions, and Recommendations

The investigation's primary purpose was to determine and apply a solution approach to improve early-stage energy consumption estimations using real-operational data. To accomplished this goal, seven research questions were addressed to help guide the course of the investigation. Section 9.1 elaborates on the findings and conclusions of each question, respectively. After the questions are considered, the overall research goal is addressed and concluded in section 9.2. Here, the established modelling requirements are investigated to determine if the overall objectives are suitably met. The general discussion then continues with a contribution toward science and observed limitations of the corresponding research. Finally, Section 9.4 presents the proposed future recommendations.

9.1. Research Questions

1. How do the DVNA design process, calculation methods, and data availability influence the overall modelling requirements when estimating the EC?

These three criteria directly influenced the modelling requirements for estimating total energy consumption. Based on each analysis, associated problem definitions were directly converted into method requirements.

The design process is necessary to understand which phase in the design process needs to be considered. The earlier the stage, the less fidelity and accuracy of the modelling solution is required and vice-versa. The calculation methods are influential in understanding what resources are available and what might be needed. Having a clear picture of what could be used or what is currently being used, the associated method could either integrate or avoid redundancy. Finally, since the overall goal is to incorporate real-operation information to improve existing processes, understanding what datasets were available was paramount. It was determined that various sources of expanding datasets were available, and as such, the method requirements should be able to account for and incorporate these growing collections.

Thus, the solution approach needed to adhere to a $< 15\% \pm 95\%CI$ accuracy threshold while allowing a modular framework to incorporate multiple available dynamic models. Additionally, the method needed to be based on readily available and growing *DVNA* data collections.

2. What methods currently exist to predict EC for both propulsive and auxiliary loading accurately, and which approaches are most suited to achieve the modelling requirements?

Potential solutions to energy consumption prediction were explored and identified. It was found that three main modelling categories can be considered: White-box, Black-box, and Grey-box modelling. Each of these approaches had its own associated strengths and weaknesses.

WBM's are based on deterministic equations and require a high degree of system knowledge. These are commonly applied in all design applications within the Maritime Industry and are typically semi-empirical-based expressions. These methods are usually used for first approximations as they allow for a high degree of extrapolation with moderately low accuracy. On the other hand, BBMs, such as machine learning algorithms, are purely data-driven with an input-output representation. These methods, while exhibiting high prediction accuracy, have very poor extrapolation and interpretability capacity. Artificial neural network (ANN) algorithms are most commonly applied to estimate energy consumption within the Maritime Industry. Finally, the GBM attempts to blend both a WBM with a BBM to retain some physics within a data-orientated solution. The advantages of such a model are high prediction accuracy, improved extrapolation capacity, and potentially reduced input data amounts.

Based on a careful qualitative comparison between all modelling categories, it was found that the GBM-ANN had the highest likelihood of meeting all established method requirements and filling the identified literature gaps. However, associated risks of success, such as data quantity and quality, as well as WBM accuracy, completeness, and relevance, were also acknowledged.

3. How do the proposed technical solutions operate, and which modelling conditions, limitations, and assumptions are necessary for optimal performance?

As a GBM solution was judged most suitable, both WBM and BBM model limitations and assumptions must be considered for optimal performance. Ultimately, four WBMs and one BBM solution were investigated and outlined, respectively.

- **Calm-Water Resistance Calculation (WBM Propulsion):** The calm-water calculation has two main limitations: vessel type and efficiencies. The model is limited to monohull (semi-) displacement vessels with moderately slender bodies and low Froude numbers. Additionally, early-stage efficiencies are limited to low-order empirical formulations heavily dependent on propeller design and vessel shape. Usually, such information is not available in the early stages; thus, literature and model-scale results are considered to validate the results.

- **Wind Resistance Calculation (WBM Propulsion):** The wind resistance component is based on *Feadship* CFD wind tunnel profiles. These profiles have been verified and validated with corresponding model wind tunnel tests. While the studies are limited to vessel lengths of 66.5m to 100m only, the overall influence of the windage profiles is narrowly banded and shows minimal drag impacts.

- **Added Thrust in Waves Calculation (WBM Propulsion):** The wave thrust component is similarly based only on *Feadship* mode-scale tests. This model is limited to regressions based on waterline lengths of 51m to 108m. While the model has been verified and validated, only a single-load (design draft) condition is considered.

- **HVAC Power Calculation (WBM Auxiliary):** The HVAC power model is based on area classification requirements assigned by ASHRAE/ISO codes and standards. However, the model does not take into account any fitting losses and personnel movements. Instead, room-by-room evaluations with maximum persons distributed per room classifications are used to evaluate required powering demand.

- **Artificial Neural Network (BBM):** ANN's are susceptible to both over- and under-fitting depending on the added complexity of the model structure and inputs. Therefore, hyperparameter tuning could be used to control the learning processes for optimal solutions. However, the grid searching approach is limited in terms of computational demand as it can take a very long time. Therefore, literature and similar cases were used as an initial starting position for the search.

4. What current data preparation methodologies exist to incorporate raw operational information within the modelling approach?

Based on previous works of Zwart [108], a sequential data-preparation process was outlined and extensively detailed. This process consisted of eight complete steps to ensure the available raw data could be successfully converted into BBM-ready information. Noise Identification and Feature Selection methods were seen as the dominant preparation steps.

Noise Identification is the process of identifying and removing any outlying data entries. These irregular points have the potential to skew the output results significantly. Two approaches are commonly

applied: simple parametric methods such as IQR elimination and complex non-parametric techniques such as clustering or density grouping. Since the available data is being continuously monitored from real-world operations, the law of large numbers and central limit theorem suggests that the feature distribution is expected to be normally distributed if the samples are large enough. Therefore, while a highly conservative approach, the parametric IQR approach eliminates outlier noise throughout the investigation.

Feature Selection determines which independent features are closely related to the dependent target parameter. It was found that additional unnecessary features have the disadvantage of adding complexity and reducing modelling accuracy. Therefore, the Spearman Correlation method was introduced to identify the most relevant inputs. The approach quantifies non-linear monotonically increasing or decreasing correlations to help guide the ANN training procedure. Unfortunately, non-monotonic dependencies such as parabolic or oscillatory relationships can not be accurately measured, as such engineering sense still needs to be applied. Ultimately, the larger the correlations between the input and output features, the higher likelihood of capturing the induced WBM physics.

5. How is the general performance influenced by varying modelling categories, data-preparation procedures, and vessel-specific operational usage?

The performance of the best-developed models was directly compared and investigated for each vessel-specific condition, namely: Sailing, Anchor, and Combined. It was found that performance greatly varies between modelling categories, outlier detection applications, and associated vessel operations.

Amongst the three modelling categories, the GBM solution had the best performance metrics whereas, WBM's were typically least effective within the dataset training ranges. Based on the modelling requirements, all GBM solutions successfully achieve the established accuracy conditions. The observed ranking of performance was as follows,

$$GBM \geq BBM > WBM$$

A direct comparison between the GBM and WBM indicated that the former had an average of 15% improvement over the latter. Whereas the performance between the GBM and BBM was either slightly improved (1%) or equal. When comparing the IQR outlier detection methodology, an average performance improvement of 3% was observed. However, in the Total Energy Consumption Combined analysis, the IQR approach resulted in decreased performance. This case was related to the adjoining of the two vastly different operational datasets between Propulsion and Auxiliary power, as well as Sailing and Anchoring operations. Thus, the required parametric distribution assumptions were no longer directly applicable. Finally, for each corresponding best-developed model, a comparison between each associated operational dataset was conducted. The observed performance metric rankings were seen as follows,

Propulsion Sailing > Total Combined > Auxiliary Anchor > Auxiliary Combined

Where the MAE for each condition was 2%, 6%, 7%, and 10%, respectively. Ultimately, it was recognized that feature dependency correlations played a significant role in the early identification of modelling accuracy. However, it also demonstrated that the performance of a combined model reflects each independent case. Thus, while slight improvements were seen due to the propulsion power serving as a dominant proportion of the total consumption, the individual operational dataset limitations are mainly retained and reflected in the combined model. In other words, the zero-speed auxiliary uncertainty behaved similarly to the auxiliary anchor-only uncertainty. In comparison, the combined operations demonstrated uncertainty propagation principles.

6. What estimation capabilities does the proposed solution exhibit outside the design domain, and how can the total performance be leveraged to isolate and extract hidden relationships such as,

(a) **Fouling effects overtime on total propulsion power?**

(b) **Daylight cycle effects on total auxiliary power?**

While the initial study focused on interpolation within the interior-domain data regions, the modelling extrapolation capacity was also of interest. By utilizing the over-conservative nature of the IQR

approach, both bound limit and clustering methods were applied to extract valid outside-domain data entries from the residual datasets.

The best-developed propulsion model indicated a significant learning degree, demonstrating an improved extrapolation capacity compared to the BBM by approximately a 20% difference in MAE. While performance metrics made improvements over the BBM, the WBM indicated even more superior extrapolation performance. The best-developed auxiliary model demonstrated profoundly different results, where the GBM showed only a tiny amount of physical sense was retained over the BBM.

From the study, a significant limitation of the GBM was its apparent inability to distinguish between infeasible energy demand regions as opposed to the WBMs. Thus, two conclusions were drawn:

- If models are not appropriately trained to identify such areas, the ability to combat these infeasibilities is purely based on the learning capacity.
- The degree of learning is related to the internal dependencies, where the need for not just large data quantities but also suitable data features that fully capture each target parameter's dynamic operational nature is paramount.

Therefore, the observed performance metric rankings in the extrapolation regions were seen as follows,

$$WBM > GBM \geq BBM$$

A secondary study was also conducted to isolate and exploit the inherent pattern developments between the input and output features. Two studies were investigated, the contribution of continuous fouling on the propulsion power and the contribution of binary daylight on the auxiliary power. Both studies provided exciting insight into the ability to develop non-linear relationships from seemingly simple input features.

The first study indicated that an expected 26% increase in propulsion power is expected without any cleaning in a year. However, the fouling contribution exhibited an unexpected declining behaviour within the inside-domain data region beyond a year's time. However, after carefully decomposing and dissecting the corresponding dataset, similar features were noticed, verifying the learned non-linear output relationship results. The second study demonstrated the effects of a binary input relation. Here it was demonstrated that both radiation and transmission contributions could be isolated and analyzed. However, apparent uncertainty within the training data regions was seen as significant. Similarly, a dataset investigation indicated that the sparse dataset regions were echoed with the growing uncertainty bounds. Additionally, while it was seen that the model typically categorized the dominant feature correctly, a 6.3% mean absolute difference was recorded. Unfortunately, the developed model's expected performance error is 9%. Therefore, attempting to quantify a relationship where the actual deviations fall below the error threshold can lead to highly uncertain results.

Ultimately it was concluded that both investigations provided evidence of the solution approach's ability to learn, isolate and extract characteristics and patterns that are not inherently present. However, great care must be considered when implementing such exploitation techniques. While the GBM-ANN is a powerful and innovative tool, data quality and focus can directly influence the corresponding target outputs. In other words, garbage in is equal to garbage out and vice-versa.

7. How can the proposed solution approach be integrated within conventional ship design processes, and what criteria must be considered for successful implementation?

Three criteria, *Why, How, and What* were considered to provide a practical overview of the proposed GBM solution within the conventional ship design process.

Why was related to the relevancy and motivation to incorporate such data-driven solutions. Currently, sustainability trends and challenges enforce the need to improve efficiency. However, current design methods are challenged by capturing the complete dynamic responses vessels undergo within operational usage. Thus, direct design benefits can include,

- Rapid transition between concept development and detailed operational digital twins, thus,
- Reducing design life-cycle costs while improving quality of designs.

How was related to the necessary limitations and considerations when such methods could be useful. Ultimately, it was concluded that, while the GBM solution approach is versatile, there are situations

when the method should not be implemented, such as simple objectives, outside domain range, low data quantity, poor data quality, and/or limited resources.

Finally, the *What* provided a practical overview of what applications within a typical design iteration process could be investigated. Such design elements included principal dimensions, powering, weight estimates, hydrostatics, costing, etc. Ultimately, the limiting factor is the creativity of the Naval Architect. A general remark about sociological and ethical challenges such as Data Privacy was acknowledged as a final design implementation consideration.

9.2. Research Goal

From the above research questions, the findings within each could be used to formulate a general road map to address the final research goal,

Develop an approach to accurately predict total dynamic Energy Consumption (EC) using real operation voyage data for the improved early-stage design of new future yachts

It can be concluded that an essential first step is made in extracting, identifying, and implementing operational datasets to estimate dynamic operational energy consumption. However, to ensure that the goal has been thoroughly met, a set of method requirements were established to assess whether the approach can fully quantify whether the condition *'accurately'* has been achieved. As such, a total of six method requirements were established, where the corresponding conclusions of each can be seen summarized in table 9.1.

Table 9.1: Method requirements and overall solution conclusions

Estimate power for propulsion and auxiliary systems under dynamic conditions within $\pm15\%$ with 95%C.I.	✔
Ability to proportion both auxiliary and propulsion power consumption independently	✔
Be based on available data within *De Voogt Naval Architects* databases	✔
Be based on a modular methodology to easily incorporate various estimation tools and results	✔
Be able to deal with discrepancies and errors in voyage report data	✔
Be able to incorporate a range of ship sizes within the *De Voogt* fleet	❓

Ultimately, it can be concluded that the first five requirements have been confidently achieved. Section 6.3 demonstrated the incredibly accurate modelling evaluation capabilities for each corresponding operational consideration, all of which fell below the required 15% threshold. As indicated in section 2.4, all models are developed based on readily available *De Voogt* datasets (summarized in table 5.1). Section 6.3.4 demonstrated the flexible nature of a GBM solution, allowing for simple model aggregation and proportioning based on a singular output using a multiple model scenario. In addition, and falling in line with the framework's flexibility, section 7.1 proved WBMs can be simply included as direct inputs to aid in the input-output learning capacity if suitable models are selected. Finally, while not inherently present within the models, the established modelling methodology seen in section 5.2 considers and outlines proven techniques to handle irregular data entries. The study ultimately confirmed methodology repeatability potential for various investigation applications.

Unfortunately, a multiple ship analysis has not been investigated during the study to focus on the modelling details and performance on a singular vessel. As such, the final requirement cannot be confidently addressed.

Ultimately, the developed models and the proposed modelling methodology has been inherently developed to handle and deal with growing datasets. As such, the approach should be able to quickly scale to other vessels within the *Feadship* fleet. While this broadened dataset may allow for improved application, it can also create sparse data regions. As evidenced within the corresponding study seen in section 7.2, while the model favours interpolation over extrapolation, uncertainty significantly increases

when limited training data is available to the GBM solutions. Additionally, new data features such as Froude number, ship length, displacement, or hull shape coefficients might be required to link the GBM inputs with the associated target outputs when including more vessels. These added features have the potential to increase complexity and thus, introduce the added risk of overfitting. Nevertheless, supposing the design domain is sufficiently broad and fully populated with the correct input data features, no evidence suggests that the approach would not be capable of meeting the required method constraints again. Therefore, it is highly suggested to expand the current investigation to consider additional vessels within the *Feadship* fleet to maximize the application potential.

9.3. Contributions within Industry and Academia

In the corresponding research, advanced analytical data science and maritime operational information and design have been combined to address the need to improve early-stage energy consumption estimations. Typically, the marine industry has always been considered more conservative than the rapidly advancing data science industry, which is currently at the forefront of the digital revolution. Ultimately, the lack of studies associated with applying advanced BBM or GBM methods related to auxiliary power estimation and the lack of general yachting research were recognized and addressed.

Therefore, a comprehensive two-fold contribution was realized to allow for immense value within both industry and academia (see section 3.5). While the associated research has both considered and filled the associated gaps, additional contributions have developed as a natural consequence in search of realizing the global research goal of improved early-stage energy consumption estimations. As such, the following relevant industry and academic contributions can be recognized,

1. **Identified strengths and weaknesses via modelling comparison studies for optimal usage:** A comparative investigation of all modelling categories was used to address the full capabilities of the GBM solution approach by examining both inside- and outside-design data range performance. While, it was acknowledged that GBM does exhibit a degree of learning and, thus, improved extrapolation capacity over the pure BBM. The degree of learning is highly dependent on the relationship between the WBM and the operational target parameter. Nevertheless, it was observed that a GBM performed either the same or substantially better than a BBM. Also, the GBM was a much more capable solution than the WBM counterparts within the trained data ranges. Unfortunately, outside these ranges, accuracy and total confidence quickly decreased, thus limiting overall utilization capabilities. Additional meaningful comparisons include,

   - Multiple WBM model comparisons:
     - Influence due to the difference in accuracy versus dynamic influence
   - Multiple operational profile comparisons:
     - IQR cleaning influence and differences
     - Spearman correlation influence and differences
     - Hyperparameter optimization influence and differences

2. **Applied uncertainty quantification for improved solution interpretability:** The applied methodology does not deviate much from conventional data science (mining) suggestions: data cleaning, feature selection, input scaling, etc. However, one of the most considerable contributions was related to the quantification of GBM and BBM modelling and inherent error uncertainty within the methodology framework. A statistical Bootstrapped Aggregation (Bagging) methodology was adapted to compute the 95% confidence intervals of the solution predictions. This application moves away from the limited point-based estimations to provide a solution with a much larger degree of interpretability. Such improvements allow for direct visualization and identification of regions where the model is inadequately trained, or sparse data regions are located. Therefore, the Bagging approach not only gives Naval Architects an indication of overall modelling performance but also an understanding of regional performance within the interior and exterior design space. However, it is acknowledged that challenges in computational demand and associated time necessity are fundamentally introduced.

3. **Demonstrated the importance of feature selection and outlier detection:** Additional insight into the contribution of feature selection and outlier detection techniques on the overall model performance has been demonstrated. The Spearman Correlation feature selection method provides an early indication of both modelling performance and the GBM's ability to learn based on the WBM contributions. Unfortunately, while the technique can be applied to quantify non-linear relationships, it can only capture monotonic interactions. Nevertheless, a direct link between correlation strength and modelling performance was observed and identified. Outlier detection strategies, allowing for identifying irregular or noisy data, were also proven to influence global modelling performance positively. The parametric IQR approach would improve performance if distribution exhibited a Gaussian-like structure. However, if this criterion was not met, limited performance gains were noticed. Thus, while these methods increase performance, applying the correct approaches is the driver of effectiveness for such routines. Ultimately, the contributions are the tried-and-true performance improvements and observable early-stage insight into modelling performance when applying outlier detection and feature selection methods, respectively.

4. **Identified WBM accuracy versus WBM dynamic influences:** Grey-box models are composed of Black-box and White-box sub-models to obtain the benefits each modelling category offers. Thus, the WBM(s) quality is known to influence the overall performance, where two criteria were considered within the associate study: WBM accuracy and WBM dynamic correlation. Ultimately, it was found that the importance of capturing WBM dynamical effects as opposed to WBM accuracy was more influential to the overall GBM performance. A WBM may be more accurate in terms of performance metrics; however, the BBM component is more so influenced by the internal relationships and dependencies between the input-output variables. While modelling accuracy is usually tied with improved dynamical relations, this connection is not guaranteed. Therefore, it has been demonstrated that even though WBM accuracy and relevance play a role, capturing the underlying dynamical links is a driving force in GBM performance success.

5. **Provided a transparent methodology and design implementation support structure:** Finally, arguably one of the most significant contributions of the following research is in providing a clear, transparent, and repeatable workflow to demystify machine learning aspects and applications. While familiar within data science, many of the presented concepts have not yet been encountered within the Maritime industry and, as such, can provide a degree of uncertainty that can inherently lead to resistance or reluctance in exploration. Thus, in addition to a straightforward sequential process, awareness of the limitations and conditions for appropriate application are thoroughly addressed to aid Naval Architects in the early adoption decision processes based on the thesis learnings. Unfortunately, this is a rapidly evolving field, and innovations are found daily; thus, the methods and concepts may quickly become outdated and irrelevant. Nevertheless, it is believed that this practical framework can provide universal contributions to multiple maritime sectors and design applications by raising user understanding and awareness.

9.4. Future Recommendations

Ultimately, the current research and findings have only scratched the surface of the possibilities available for a GBM approach. Therefore, both analysis and modelling recommendations can be identified when considering the current focus on powering and data-driven GBM modelling. Both the observed limitations and associated recommendations can be seen below,

1. **Analysis - Multi-vessel incorporation study:** The current study investigated a single vessel – ultimately developing a digital twin model. Therefore, a significant limitation can be related to the immediate implementation of the developed GBMs since only a single vessel can be accurately estimated. As such, it is recommended that multiple vessels and their associated performance be investigated to determine the true feasibility of such solution approaches for early-stage energy consumption of an entire fleet. This study would provide Naval Architects with valuable insight into the GBM's ability to consider and estimate new vessel powering demand, which can be utilized for quick exploratory design optimization capabilities and early-stage feasibility studies.

2. **Analysis - Manual extrapolation study:**   The current extrapolation approach takes advantage of the over-conservative nature of the IQR technique. This method allowed the best-developed models to be directly compared under extrapolation conditions. Unfortunately, this approach does not allow for a robust investigation into all sparse data regions and exterior domain regions. Therefore, it is recommended to consider an alternative analysis approach, where a hold-out extrapolation dataset is extracted manually beforehand. Unfortunately, this would require the development of a model based on reduced information. However, the manually extracted points could then be smartly selected to evaluate sparse data regions, extrapolation regions, and interpolations regions. Ultimately, this would allow for more user control over the investigations, thus providing more flexibility and certainty of GBM within the various design regions.

3. **Analysis - Feature selection sensitivity study:**   It was acknowledged that feature selection and associated dependencies play a critical part in the learning capabilities of the GBM solutions. However, the actual influence of the number of features was not considered within the study. Therefore, understanding the direct impact the number of features has on the modelling outcomes is ultimately unknown. It is recommended that a direct sensitivity study where the number of input features is varied and the impact on performance in terms of fitting and accuracy is measured and recorded. This study would provide valuable insight into the importance of targeted feature selections. In addition to varying features, alternative feature selection methods should also be comparatively investigated due to the recognized importance of the process.

4. **Modelling - Operational profile decomposition:**   The investigation focused on orientating the dataset to consider three operations: Sailing, Anchor, and Combined. Unfortunately, this breakdown is rather crude; thus, challenges were observed in clearly distinguishing between actual outlier data points and factual situational data since yachting functions are never clearly defined (pleasure versus purpose). Therefore it is recommended to further decompose each operation into more detailed operational profile data orientations such as; Sailing-low speed, Sailing-medium speed, Sailing-high speed), Anchor, Harbour, Guest onboard/off-board, and Manoeuvring conditions. Such detailed decompositions can provide further insight and connection between input-output relations and improved IQR cleaning benefits.

5. **Modelling - Missing input features:**   The study indicated that the auxiliary power estimation models lacked the appropriate features to quantify dynamic input-output relationships fully. Unfortunately, this detrimentally influenced the overall modelling performance. As such, further decomposition within the auxiliary power should also be considered. Since auxiliary power is a complex summation of multiple dynamic systems, univariate input-output analysis is likely inadequate. Therefore, main contributors such as HVAC and Stabilization/Rudder powering target features should be collected and applied to develop dedicated GBMs. These will help more closely define the relationships between existing parameters and further enforce whether other unknown inputs exist. In addition, Acceleration/Deceleration and Speed-Through-Water (STW) are not considered within the study. These parameters could significantly influence the total powering demand and, as such, should be quantified in the following investigations.

Personal Reflection

I have always appreciated a challenge; as such, I was excited for the opportunity to live and learn in a country halfway across the world. In addition to finally pursuing one of my lifelong passions, I have always loved to travel and experience new cultures. Therefore, it was a no-brainer to complete my Master's degree abroad.

Unfortunately, starting within the year 2020, the COVID-19 pandemic quickly presented enormous challenges and roadblocks. While the initial 100% digital transition saw a dip in quality and clarity, this was quickly overcome, and the educational value remained high throughout. As the old saying goes, *'the first through the wall always gets a little bruised.'* However, the most significant challenge and what could not be accounted for is the missed social interaction and the inherent learning that comes along with it. While many support services were provided and a great deal was invested in addressing the global challenges, obvious limitations remained. Exploration and experience of Dutch and neighbouring cultures quickly proved impossible. Furthermore, sleeping, eating, working all in the same place for 24-hours a day presented motivation and demoralization problems I have never experienced before! As such, I am incredibly proud of myself for; not only completing the entirety of my masters online but completing my study slightly ahead of schedule without (I believe) a significant drop in quality.

However, I honestly can't complain too much as I had a great support system! While physical interactions were reduced, online video and phone calls quickly provided a satisfactory substitution. Although not a perfect replacement, I found comfort in the fact that I was not alone in the struggle. Nevertheless, through adversity comes learning, thus I'd like to provide a few experiences to help the next generation of students,

- **Look good, feel good, write good:** Having converted to everything online, sometimes it was challenging to find a nice routine to get into the thesis writing mood. However, I noticed that a degree of self-care, even something as small as putting on moderate work attire, can boost motivation. As such, if you are ever feeling demoralized or unmotivated, a slight change in routine can go a long way!

- **Plan, plan, and… you guessed it, plan:** Having a thorough and detailed plan early within the graduation project truly helped organize the days. Without any concrete objectives or plans, I often felt a high degree of uncertainty in direction, which resulted in the proverbial *'spinning of the wheels.'* Nevertheless, I found that if you take 15 to 20 minutes a day to establish a task per day, no matter how big or small, it can push you in the right direction. Any progress always gives a great sense of accomplishment! *PS. Write things down!!! I can't mention enough how often I had a good idea and then forgot about it, which is super frustrating but easily avoidable!*

- **The famous 'Enjoy the small things' cliché:** A Master's degree is not easy, and a lot of work and mental power is required. Therefore, make sure you take time to celebrate and enjoy the small things that come along with it. Since it is a long journey, you should never be afraid to look back at what you have already done and feel proud. Sometimes, I would feel guilty about not accomplishing much within a day to instead do something entirely unrelated to the graduation project. However, I found that a bit of *RnR* often resulted in new perspectives and new inspiration! So my advice is to enjoy (obviously within moderation... thesis duh) the small things, small wins, and small pleasures along the way! They might even improve your performance (what a cliché!).

- **The good, the bad, and the ugly:** Finally, to put it bluntly, bad things happen. Sometimes they are avoidable; sometimes, they are not. Some days are better than others; some are not. I found the journey had a lot of twists and turns. So remember to keep an open mind and try not to let one hangup influence the remainder of the project and experience. Easier said than done, I know, but as they say, *'smooth seas do not make skillful sailors.'*

I hope that students can use my experiences (and bad humor) to improve their own, Success!

A
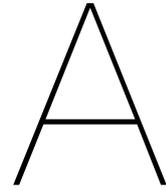
Appendix

A.1. De Voogt Naval Architects Interviews - CONFIDENTIAL

The following section contains *Feadship* classified information; therefore, it is removed from the report to maintain confidentiality.

B
Appendix

B.1. White Box Model Calculation Procedures

B.1.1. Holtrop and Mennen - Calm Water Resistance Estimation

Table B.1: Required and optional input parameters for Holtrop and Mennen Method

Parameter	Symbol	Remarks
Ship speed (SOG)	V_s	
Length in waterline	L_{WL}	
Molded beam	B	
Molded mean draft	T	Typically $T = \frac{1}{2}(T_A + T_F)$
Molded draft at aft perpendicular	T_A	
Molded draft at fore perpendicular	T_F	
Volumetric displacement	∇	Alternatively use the block coefficient as $C_B = \nabla/(B\,T\,L_{WL})$
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}	May have to be estimated in early design stages
Longitudinal center of buoyancy	l_{cb}	Positive forward; with respect to $L_{WL}/2$ in percent of L_{WL}
Area of ship and cargo above waterline	A_V	Projected in direction of V_s
Immersed transom area	A_T	Measured at rest
Transverse are of bulbous bow	A_{BT}	Measured at forward perpendicular
Height of center of A_{BT} above basis	h_b	Has to be smaller than $0.6\,T_F$
Propeller diameter	D	
Propeller expanded area ratio	A_E/A_O	
Stern shape coefficient	C_{Stern}	
Optional Parameters		
Wetted surface (hull)	S	
Wetted surface of appendages	$S_{APP,i}$	Bilge keels, stabilizer fins, rudders, etc.
Half angle of waterline entrance	i_E	
Diameter of bow thruster tunnel	d_{TH}	

The general Holtrop and Mennen [51, 52] method computation computes the non-dimensional total resistance, which is broken into multiple individual resistance components.

$$R_{T,CW} = (1+k)R_F + R_{APP} + R_W + R_A + R_B + R_{TR} + R_{AA} \tag{B.1}$$

These components include frictional resistance, R_F with form factor k for the hull variations, the resistance of appendages R_{APP}, a wave-making and wave-breaking resistance R_W, a model-ship correlation resistance R_A, pressure resistance due to bulbous bow R_B, an additional pressure resistance of the immersed transom $R_T R$ and an air resistance component R_{AA}. Each resistance element can be calculated using the following independent relations,

$$R_F = \tfrac{1}{2}\, \rho_{SW}\, V_s^2\, S_{BH}\, C_F \tag{B.2}$$

$$R_{TH} = \rho_{SW}\, V_s^2\, \pi\, d_{TH}^2\, C_{D,TH} \tag{B.3}$$

$$R_{APP} = \tfrac{1}{2}\, \rho_{SW}\, (1 + k_2)\, C_F \sum_i S_{APP,i} + \sum_i R_{TH} \tag{B.4}$$

$$R_w = c_1\, c_2\, c_5\, \rho_{SW}\, g\, V\, e^{(m_1\, Fr^d + m_4\, cos(\lambda\, Fr^{-2}))} \tag{B.5}$$

$$R_B = 0.11\, \rho_{SW}\, g\, \left(\sqrt{A_{BT}}\right)^3 \left(\frac{Fr_i^3}{1 + Fr_i^2}\right) e^{(-3.0\, P_B^{-2})} \tag{B.6}$$

$$R_{TR} = \tfrac{1}{2}\, \rho_{SW}\, V_s^2\, A_T\, c_6 \tag{B.7}$$

$$R_A = \tfrac{1}{2}\, \rho_{SW}\, V_s^2\, (C_A + \Delta C_A) \left[S_{BH} + \sum_i S_{APP,i}\right] \tag{B.8}$$

$$R_{AA,ITTC} = \tfrac{1}{2}\, \rho_{air}\, V_s^2\, C_x\, A_v \quad \text{(see section B.1.4)} \tag{B.9}$$

To evaluate these expressions, multiple required input parameters are necessary as outlined in table B.1. In general, vessel characteristics, such as length, width, and draft, must be known. However, more intricate properties such as the length of run and percent center of buoyancy must be known as well.

$$L_R = L_{WL} \left(\frac{1 - C_P + 0.06\, C_P\, l_{CB}}{4\, C_P - 1}\right) \tag{B.10}$$

$$l_{cb} = -(0.44\, Fr_{design} - 0.094) \tag{B.11}$$

The hull and appendage form factors can be determined using the following relations,

$$(1 + k_1) = 0.93 + 0.4871\, c_{14} \left[\left(\frac{B}{L_{WL}}\right)^{1.0681} \left(\frac{T}{L_{WL}}\right)^{0.4611} \left(\frac{L_{LW}}{L_R}\right)^{0.1216} \left(\frac{L_{LW}^3}{\nabla}\right)^{0.3649} (1 - C_P)^{-0.6042}\right] \tag{B.12}$$

$$(1 + k_2) = \frac{\sum_i (1 + k_{2,i})\, S_{APP,i}}{\sum_i S_{APP,i}} \tag{B.13}$$

Where k_1 and k_1 are the from factors for the hull and appendages, respectively. It should be noted that the appendages form factor is based on a single, equivalent relation accounting for many individual components. The appendages which are considered are rudders and stabilizers. The overall bare-hull wetted surface can be estimated using the following Holtrop and Mennen relation,

$$S_{BH} = L_{WL}\, (2T + B)\, \sqrt{C_M} \left[0.615989\, c_{23} + 0.111439\, C_M^3 + 0.0000571111\, C_{stern} + 0.245357\, \frac{c_2 3}{C_M}\right]$$
$$+ 3.45538\, A_T + \frac{A_{BT}}{C_B} \left(1.4660538 + \frac{0.5839497}{C_M}\right) \tag{B.14}$$

Where, each area component and critical height parameters can be determined as,

$$A_{APP} = n_{stabi.} A_{stabi.} + n_{rudder} A_{rudder} \qquad (B.15)$$

$$A_{BT} = A_M C_{ABT} \qquad (B.16)$$

$$A_M = B T C_M \qquad (B.17)$$

$$h_b = 2.4 \text{ assumed based on } h_b < T_F \cdot 0.6 \qquad (B.18)$$

$$h_F = C_p C_M \frac{B T}{L_{WL}} (136 - 316.3 \, Fr) \, Fr^3 \qquad (B.19)$$

$$h_W = \frac{i_E V_s^2}{400 \, g} \qquad (B.20)$$

$$P_B = 0.56 \frac{\sqrt{A_{BT}}}{T_F - 1.5 \, h_b + h_F} \qquad (B.21)$$

A_M represents the immersed transverse section area of the transom at the aft perpendicular. h_f and h_W, represent the forward sinkage and local wave height at the bow, respectively. The parameter P_B quantifies the emergence of the bulb from the still water line. Additionally, the Waterline entrance angle can be determined as,

$$i_E = 1 + 89 \, e^a \qquad (B.22)$$

where the exponent a is found through,

$$a = -\left[\left(\frac{L_{WL}}{B} \right)^{0.80856} (1 - C_{WP})^{0.30484} \left[1 - C_p - 0.0225 \, l_{CB} \right]^{0.6367} \left(\frac{L_R}{B} \right)^{0.34574} \left(\frac{100\nabla}{L_{WL}^3} \right)^{0.16302} \right] \qquad (B.23)$$

The additional critical form and resistance coefficients can be determined as,

$$C_{ABT} = 0.075 \qquad (B.24)$$

$$C_F = \frac{0.075}{(\log_{10} Re - 2)^2} \qquad (B.25)$$

$$C_B = \frac{\nabla}{(B T L_{WL})} \qquad (B.26)$$

$$C_P = \frac{\nabla}{L_{WL} A_M} \qquad (B.27)$$

$$C_M = 0.8 + 0.21 \, C_B \qquad (B.28)$$

$$C_{WP} = \frac{2}{3} C_B + \frac{1}{3} \qquad (B.29)$$

$$C_{stern} = +10 \qquad (B.30)$$

$$C_A = 0.00546 \, (L_{WL} + 100)^{-0.16} - 0.002 + 0.003 \sqrt{\frac{L_{WL}}{7.5}} \, C_B^4 \, c_2 \, (0.04 - c_4) \qquad (B.31)$$

$$\Delta C_A = \frac{0.105 \, k_s^{1/3} - 0.005579}{L_{WL}^{1/3}} \quad \text{if } k_s > 150\mu m \quad \text{else } \Delta C_A = 0 \qquad (B.32)$$

The Holtrop and Mennen regression formulation coefficients for the many resistance components can

be additionally summarized as,

$$c_1 = 2223105 \, c_7^{3.78613} \left(\frac{T}{B}\right)^{1.07961} (90 - i_E)^{-1.37565} \tag{B.33}$$

$$c_2 = e^{-1.89\sqrt{C_3}} \tag{B.34}$$

$$c_3 = 0.56 \, \frac{A_{BT}^{1.5}}{\left[B \, T \, (0.31\sqrt{A_{BT}} + T_F - h_b)\right]} \tag{B.35}$$

$$c_4 = \frac{T_F}{L_{WL}} \tag{B.36}$$

$$c_5 = 1 - 0.8 \, \frac{A_T}{B \, T \, C_M} \tag{B.37}$$

$$c_6 = 0.2 \, (1 - 0.2 \, Fr_T) \tag{B.38}$$

$$c_7 = \frac{B}{L_{WL}} \qquad \text{if } 0.11 < B/L_{WL} \le 0.25 \tag{B.39}$$

$$c_{14} = 1.0 + 0.011 \, C_{stern} \tag{B.40}$$

$$c_{15} = -1.69385 + \frac{\frac{L_{WL}}{\nabla^{1/3}} - 8}{2.36} \tag{B.41}$$

$$c_{16} = 8.07981 \, C_P - 13.8673 \, C_P^2 + 6.984388 \, C_P^3 \tag{B.42}$$

$$c_{23} = \left[0.453 + 0.4425 \, C_B - 0.2862 \, C_M - 0.003467 \, \frac{B}{T} + 0.3696 \, C_{WP}\right] \tag{B.43}$$

$$d = -0.9 \tag{B.44}$$

$$\lambda = 1.446 \, C_P - 0.03 \, \frac{L_{WL}}{B} \tag{B.45}$$

$$m_1 = 0.014407 \frac{L_{WL}}{T} - 1.75254 \frac{\nabla^{1/3}}{L_{WL}} - 4.79323 \frac{B}{L_{WL}} - c_{16} \tag{B.46}$$

$$m_4 = 0.4 \, c_{15} \, e^{\left(-0.034 \, Fr^{-3.29}\right)} \tag{B.47}$$

Here, most of the coefficients are related to the wave resistance computation where Froude numbers (Fr) are less than 0.40. The respective Froude number variations and Reynold's relations can be seen below,

$$Fr = \frac{V_s}{\sqrt{g \, L_{WL}}} \tag{B.48}$$

$$Fr_i = \frac{V_s}{\sqrt{g \, (T_F - h_B - 0.25\sqrt{A_{BT}} + h_F + h_W}} \tag{B.49}$$

$$Fr_T = \frac{V_s}{\sqrt{\frac{2 \, g \, A_T}{B + B \, C_{WP}}}} \tag{B.50}$$

$$Re_{L_{WL}} = \frac{L_{WL} \, V_s}{\nu} \tag{B.51}$$

B.1.2. Holtrop and Mennen - Hull-Propeller Interaction Parameters

In addition to determining the total resistance, Holtrop and Mennen also provide estimates for the associated hull-propeller interaction parameters. The full-scale wake fraction component can be determined as,

$$w = 0.3095 \, C_B + 10 \, C_V \, C_B - 0.23 \, \frac{D}{\sqrt{B \, T}} \tag{B.52}$$

Where C_v is considered the viscous resistance coefficient. Ultimately, this parameter combines all friction related components of the resistance and the correlation resistance from the evaluated Holtrop and Mennen components.

$$C_V = \frac{(1+k)\,R_F + R_{APP} + R_A}{{}^{1}\!/_2\,\rho_{SW}\,V_s^2(S_{BH} + \sum_i S_{APP,i})} \tag{B.53}$$

The thrust deduction fraction for a twin screw vessel can be estimated as,

$$t = 0.325\,C_B - 0.1885\,\frac{D}{\sqrt{B\,T}} \tag{B.54}$$

Additionally, the relative rotative efficiency component can be determined as,

$$\eta_R = 0.9737 + 0.111\,(C_P - 0.0225\,l_{CB}) - 0.06325\frac{P}{D} \tag{B.55}$$

The above relations are all related to twin-screwed vessels. It should be noted that the single screw propeller relations are much more complex and should be appropriately evaluated using suitable relations if such vessels are to be analyzed.

B.1.3. Holtrop and Mennen - Propulsion Chain Efficiencies

The associated propulsion chain efficiencies can be seen listed in table B.2. It should be noted that many of these parameters are mere estimates and can very from real world operation. As such, the listed efficiencies are appropriate for early stage design where a degree of variation and uncertainty exists.

Table B.2: Propulsion chain efficiencies

Parameter	Symbol	Efficiency	Remarks
Hull efficiency	η_H	varies (~95%)	Wake fraction varies with speed
Open-water efficiency	η_O	varies (~63%)	Open-water eff. varies with speed
Rotative efficiency	η_R	1.01%	Equation B.55
Propulsive efficiency	η_D	~60%	
Shaft efficiency	η_H	99%	Stapersma and Klein Houd [98]
Gear-box efficiency	η_O	95%	Stapersma and Klein Houd [98]
Transmission efficiency	η_{TRM}	94%	

The open water efficiencies are extracted from MARIN model test results. These parameters are highly dependent on the vessel speed and propeller type. Nonetheless, the parameters typically vary between 55% to 65%.

B.1.4. ITTC - Wind Resistance Estimation

This ITTC added wind resistance component is defined as the difference between the total wind resistance in seas and the air resistance force in calm water as a result of the ships relative speed.

$$\Delta R_{AA} = {}^{1}\!/_2\,\rho_{air}\,A_v\,C_x(\beta_{wr,ref})\,V_{wr,ref}^2 - {}^{1}\!/_2\,\rho_{air}\,A_v\,C_x(\beta_{wr,ref} = 0°)\,V_s^2 \tag{B.56}$$

Where, the wind force coefficient, C_x, is the parameter related projected front relative motion. This parameter is a function of the relative wind angle, $\beta_{wr,ref}$. This parameter is measured at the corresponding measuring reference height. A_v is the corresponding maximum area of the exposed frontal transverse section. ρ_{air} is the air density at the ambient sea level conditions. $V_{wr,ref}$ is the apparent wind speed which can be determined using the following relation,

$$V_{wr,ref} = \sqrt{V_{wt,ref}^2 + V_s^2 + 2\,V_s\,V_{wt,ref}\,cos(\beta_{wt})} \tag{B.57}$$

Table B.3: Required and optional input parameters for ITTC Method

Parameter	Symbol	Remarks
True wind velocity	V_{wt}	
True wind direction	B_{wt}	Global reference (β_{wt} local reference)
Ship speed (SOG)	V_s	
Ship heading (COG)	$\psi_{heading}$	
Area of ship and cargo above waterline	A_v	Projected in direction of V_s
Wind force coefficient	C_x	

In this case the true wind speed ($V_{wt,ref}$) and the ship speed over ground (V_s) is taken as a vector sum projection with respect to the true wind angle (β_{wt}). The relative wind angle can be further denoted as,

$$\beta_{wr,ref} = cos^{-1} \left(\frac{V_{wt,ref} \, cos(\beta_{wt}) + V_s}{V_{wr,ref}} \right)$$ (B.58)

The corresponding true wind velocity, $V_{wt,ref}$ is usually measured at a reference height (Z_{ref}) of 10 meters. However, when the measurements are taken at various heights, a simple conversion can be applied to account for the varying height velocity changes.

$$V_{wt,ref} = U_{wind} \left(\frac{Z_{ref}}{10} \right)^{1/9}$$ (B.59)

This reference measurement is important as a typical wind boundary layer can generally exists. This corresponding boundary layer shape, much like plate-friction boundary layers shape changes as the distance from the flat-plate (earth) increases. This effect can have a dramatic influence on the overall wind velocity. As such, care must to taken to ensure the corresponding reference heights are considered.

B.1.5. VoogtWAVE - Wave Added Thrust Estimation - CONFIDENTIAL

The following section contains *Feadship* classified information; therefore, it is removed from the report to maintain confidentiality.

B.1.6. Hotel Heat Gain - HVAC Power Estimation

Table B.4: Required and optional input parameters for Heat Balance Method

Parameter	Symbol	Remarks
Inside air temperature	T_{in}	
Inside relative humidity	RH_{in}	
Outside air temperature	T_{out}	
Outside relative humidity	RH_{out}	
Number of people	n_{people}	
Room wall, window, floor surface area	A_s	
Fresh Air Proportion	F	
Maximum allowable temp. difference	ΔT	

The general prediction process, as proposed by Stapersma and Klein Houd [99] and implemented by Boertz [15], can be described as follows,

1. Estimate the internal heat gain ($\dot{Q}$) in cabins and areas individually

2. Obtain the maximum number of air exchanges (N_{ACH})

3. Obtain the maximum fresh air (F) ratio of the supply air volume

4. Calculate the transitional conditions within the fan coil units (FCU) to obtain:

    (a) Heating demand

    (b) Cooling demand

    (c) Humidification demand (pump as an electrical component)

    (d) Fan power required to supply volumetric air flow

The sensible heat gain for an area i can be categorized as the total sensible heat flow, $\dot{Q}_{sensible,i}$ within that specific area. The total heat load can be further decomposed into four individual components. These components include heat transmission through surfaces, heat gain from people, heat gain due to solar radiation, and the heat gained by auxiliary equipment such as lights and electrical devices. A general schematic of the sensible and latent heat loads within a space can be seen in figure B.1.

$$\dot{Q}_{sensible,i} = \dot{Q}_{transmission,i} + \dot{Q}_{people,i} + \dot{Q}_{radiation,i} + \dot{Q}_{auxiliary,i} \tag{B.60}$$
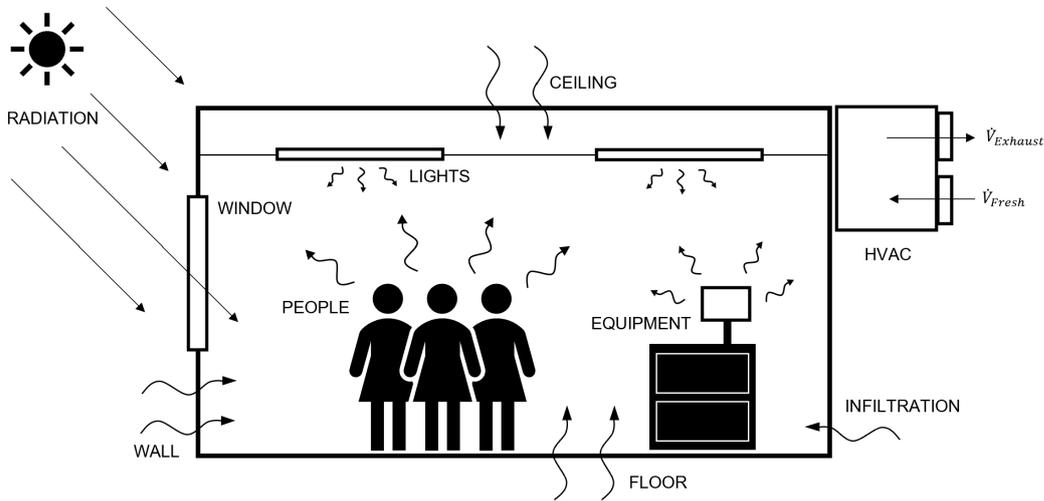
Figure B.1: General heat gain schematic

Heat gain or losses due to transmission can be determined using the following relation for each individual calculated area,

$$\dot{Q}_{transmission,i} = \sum_s (A_s K_s \Delta T)_i \tag{B.61}$$

Where A_s is the associated walls, floor, ceiling, and wind areas. k_s is the pre-defined heat transmission factor. ΔT is the associated temperature difference between the outside and inside adjoining spaces. Using the ISO 7547 [59], transmission factors and temperature differences are defined. These associated criteria can be seen in table B.5 for each corresponding surface. It should be noted that the transmission gain can either be positive or negative, which represents heat gain or heat loss to the surroundings, respectively.

Much like heat transmission, solar radiation can also influence the associated room heat gain. This parameter considers all surfaces in direct contacts with the sun, such as exterior walls and windows. As such, the general radiation heat flux consists of two main components representing each independent surface.

$$\dot{Q}_{radiation,i} = c_{sun} \, \alpha_{sun} \left(\sum_s (k_{rad} \, \Delta T_r \, A_{s,outer})_i + \sum_s (G_s \, A_{s,window})_i \right) \tag{B.62}$$

Table B.5: Heat transmission factors and minimal temperature differences for solar and nominal conditions, [7, 15, 59]

Area	Transmission Factor, k $(W/(m^2 \cdot K))$	ΔT (K)
Ceiling	0.6	5
Floor	0.6	5
Inner Wall	2.5	2
Outer Wall	0.6	$T_{out} - T_{in}$
Window	3.5	$T_{out} - T_{in}$

Solar Area	Solar Transmission Factor $(W/(m^2 \cdot K))$	ΔT_r (K)
Window	$G_s = 240$ (clear glass surface w/ interior shade)	(-)
Outer Wall	$k_{rad} = 0.6$	12

Much like the transmission factors, table B.5 provides the general parameter design guidance outlined by the ISO 7547 [59]. However, additional parameters are included for the sake of accurate modelling. α_{sun} represents a binary factor, which indicates whether it is day-time (the sun is present) or night (the sun is not present). The shadowing factor, c_{sun}, accounts for any partial balcony overhang and average sun penetration angle. In most cases, this factor ranges depending on the overhang distance, balcony length, and also the time of day. As such, an average factor of 0.40 will be applied to account for the uncertainties and further simplify the prediction model.

The heat gain per passenger, when at rest, can be assumed to increase the sensible heat by 70 W/Person and the latent heat by 50 W/Person. The number of people varies for each area and is highly dependent on the personnel movement and passenger behaviours on board. However, for simplicity, the number of guests for each room is dependent on the ratio of room category area and total area. The proportion is then applied to the maximum number of possible guests and distributed to the corresponding number of rooms per category type (see table B.6).

$$\frac{\left(\frac{A_{cat.}}{n_{room,cat.}}\right)}{A_{total}} = \frac{\left(\frac{n_{cat.}}{n_{room,cat.}}\right)}{n_{people,total}} \tag{B.63}$$

$$\frac{n_{people,cat.}}{n_{room,cat.}} = \frac{n_{people,total} \cdot \left(\frac{A_{cat.}}{n_{room,cat.}}\right)}{A_{total}} \tag{B.64}$$

Ultimately, this approach allows that the maximum onboard occupancy is never exceeded. However, cabins and state rooms will be limited to two persons as per regulations [59].

$$\dot{Q}_{sens_{people,i}} = (n_{people})_i \cdot 70 \text{ W/Person} \tag{B.65}$$

$$\dot{Q}_{lat_{people,i}} = (n_{people})_i \cdot 50 \text{ W/Person} \tag{B.66}$$

The auxiliary heat flux results from the relevant equipment and lights that produce a degree of heat to the surrounding area. As such, the general expressions can be decomposed into two independent parameters.

$$\dot{Q}_{auxiliary,i} = \dot{Q}_{light,i} + \dot{Q}_{equip.,i}$$
$$\dot{Q}_{auxiliary,i} = \dot{q}_{light,i} A_{s,floor} + \dot{Q}_{equip.,i} \tag{B.67}$$

As outlined by the ISO 7547 [59], in spaces without light, the heat gain from lighting shall be determined from the lighting's rated wattage. In general, LED lights are the norm; however, ISO 7547 [59] provides only incandescent or fluorescent lights as a guide. Nonetheless, LED and fluorescent lighting provide a similar degree of illumination. While exact heat gains can be obtained by analyzing individual required spaces and lighting types, the general guidance is deemed sufficient for a first estimate. Much like lighting, equipment loads also influence the associated gains per meter area coverage. Unfortunately, ISO 7547 [59] does not indicate such equipment loads. However, Boertz [15] has compiled a general

area summary for a passenger vessel based on consultation with literature and practicing engineers. It was determined that a fixed equipment load of 200 W for occupied state rooms is sufficient. Table B.6 highlights all parameters required for heat load balance and HVAC sizing. It should be noted that the indoor relative humidity will be maintained at 50% for all areas following the ISO 7547 [59].

Table B.6: General area HVAC design parameters based on typical owner specifications and international standards, [7, 15, 59]

Area	Heat Lights (W/m^2)	Heat Equip. (W/m^2)	Air Changes $(^-/h)$	FA Prop. (%)	Inside T $(^\circ C)$
Corridors	10	0.5	6	50%	22
Crew Cabins	8	3	6	40%	22
Crew Public Area, Offices, Mess Hall	10	2	10	40%	22
Disco, Gym, Wellness	20	2	12	100%	22
Engine Room (Primary)	10	80	15	100%	S: 27, W: 22
Galley, Pantries	20	4	25	100%	22
Garbage Rooms, Stores	10	3	60	40%	-5
Hospital	20	2	10	100%	22
Laundry	20	3	20	40%	22
Public Spaces	10	1	12	40%	22
Restaurants	10	3	10	50%	22
State Rooms	8	200 (Fixed)	$max(N_{ACH})$	$max(F)$	22
Tech. Rooms, Workshops, AC Rooms	10	4	20	100%	25
Wheelhouse	10	4	20	40%	22

Once the heat load has been determined, the sizing of the HVAC system can begin with determining the required number of air changes (N_{ACH}) necessary to main the set environmental conditions. Ultimately, many conflicting regulations exist in the determination of the parameter. These can vary from heat balance, set specifications, and owner preference. As such, the maximum number of air changes to maintain a respectable level of energy use can be determined by comparing each. The required air exchanges can be further converted into the necessary supply airflow rate considering the room volume.

$$N_{ACH,i} = \max\{N_{Heat}, N_{Spec}, N_{Owner}\} \tag{B.68}$$

$$N_{ACH,i} = \left(\frac{V_{supplyair}}{V_{room}}\right)_i \tag{B.69}$$

The total sensible heat gain can be converted to supply air flow rate through the incorporation of air density (ρ_{air}), the specific heat constant ($c_{p,a}$), and a maximum area temperature difference parameter (ΔT).

$$N_{Heat,i} = \left(\frac{\dot{V}_{supplyair-heat}}{V_{room}}\right)_i = \left(\frac{\dot{Q}_{sensible}}{\rho_{air}\,c_{p,a}\,\Delta T\,V_{room}}\right)_i \tag{B.70}$$

As outlined by ISO 7547 [59], the maximum temperature difference should not exceed $10^\circ C$ when cooling and should not exceed $8^\circ C$ when heating spaces. While a minimum number of air changes is preferred to reduce energy use, the number of air changes cannot be lower than six. As highlighted by Boertz [15], these regulations have been placed by the Maritime Labour Convention (MLC) to ensure a safe working environment. Additionally, The American Bureau of Shipping (ABS [1]) enforces that a minimum air exchange rate of six should be set for all enclosed spaces. Whereas, *Feadship* specifications require a minimum of eight air changes for all guests state rooms.

Once the number of air exchanges is known, the minimum fresh air ratio (F) of supplied air can be estimated. This parameter evaluates the maximum amount of outside air required depending on CO_2 content, ISO 7547 [59] specifications, and owner requirements.

$$F_{FA,i} = \max\{F_{CO2,i}, F_{ISO,i}, F_{Owner,i}\} \tag{B.71}$$

Ultimately, rising CO_2-levels emitted by people must be replaced with incoming fresh air. As outlined by Stapersma and Klein Houd [99], the maximum CO_2-level in a room should be limited to 0.1%, whereas the CO_2-level of fresh air is assumed as 0.035%. Additionally, it is estimated that an average person produces 0.02 m³/h of CO_2. The associated fresh air ratio can be determined by the required fresh air over the previously determined total supply air.

$$F_{CO2,i} = \frac{\dot{V}_{FA-CO2,i}}{\dot{V}_{supplyair,i}} = \frac{\dot{V}_{CO2,i}}{(y_{CO2,room} - y_{CO2,FA}) \, \dot{V}_{supplyair,i}} \tag{B.72}$$

The ISO 7547 [59], states that the minimum proportion of fresh air should not be lower than 0.008 m³/(s·Person). However, they continue to state that the minimum quantity of outdoor air should not be less than 40% of the total air supplied.

$$F_{ISO,i} = \frac{0.008 \, (n_{people})}{\dot{V}_{supplyair,i}} \tag{B.73}$$

Furthermore, *Feadship* specifications suggest that the fresh air ratio should be set at a minimum of 70% for all guest state rooms.

Once the heat gain, air exchanges, and fresh air proportions are known, the intermediate HVAC conditions' powering requirements can be determined. The prediction model evaluates these intermediary stages using a novel procedure outlined by Stapersma and Klein Houd [99]. Thermodynamic properties depend on whether heating, cooling, or humidification processes are undergone in each section. A Mollier diagram is used to trace the intermediate air conditions from the inlet air to the exhaust air conditions to evaluate these properties. The thermodynamic properties of importance are temperature (T), relative humidity (ϕ), enthalpy (h), and absolute humidity (x). If two properties are known, the diagram allows for the determination of the remaining two.

FCU Mixing Box (1-2-7)   The fan coil units (FCU), used to heat the cabins, state-rooms, and areas, use mixing boxes to provide fresh air proportions with the recirculated ambient air. This merging of air flows allows for the determination of the mixed properties using the previously determined fresh air ratios (equation B.71).

$$T_{mix} = F_{FA} \cdot T_{FA} + (1 - F_{FA}) \cdot T_{recirc} \tag{B.74}$$

$$h_{mix} = F_{FA} \cdot h_{FA} + (1 - F_{FA}) \cdot h_{recirc} \tag{B.75}$$

$$x_{mix} = F_{FA} \cdot x_{FA} + (1 - F_{FA}) \cdot x_{recirc} \tag{B.76}$$

Humidification/Sprinkler System (4-5)   During winter or cold outside conditions, the air can be quite dry thus requiring humidification. Whereas, during summer or warm periods, the air does not any added humidity. Instead water may be sprinkled to reduce the temperature without affecting the enthalpy. Depending on the situation, either adds steam or water vapour to the supply air. For humidification, the process does not added any additional sensible heat, however the change in enthalpy can be determined using water evaporation thermodynamic properties such as the evaporation constant (r_W), specific heat of water vapour (c_p, v), and the evaporation temperature (T_{vap}). For simplicity this temperature is generally assumed as 100°C.

$$\Delta h_{hum} = \Delta x_{hum} \cdot (r_W + c_{p,v} \, \Delta T_{vap}) \tag{B.77}$$

For sprinkling, the wet bulb temperature remains the same, while the dry bulb temperature is reduced. Ultimately, the wet bulb temperature and enthalpy follow a similar line on the Mollier diagram. As such, no change in enthalpy will be considered.

Fan Power (5-6)   The required fan power can be estimated through consideration of the volumetric flow rate of supply air ($\dot{V}$), associated fan pressure drop (ΔP_{fan}), and the overall fan and electric motor efficiency (η_{fan} and η_{em}).

$$P_{fan} = \sum \left(\frac{\dot{V}_{supplyair,i} \, \Delta P_{fan}}{\eta_{fan} \, \eta_{em}} \right)_i \tag{B.78}$$

The latter parameters are generally determined based on supplier specifications. However, for simplicity, fan and electric motor efficiency will be held constant at 0.6 and 0.9, respectively [99]. Additionally, the fan pressure head will be assumed as 0.02 bar [99].

Heating Power (2-3)   The heating power is a function of the heat flux across the heating unit and the transfer efficiency. The FCU is used to heat the cabins, state rooms, and areas individually. As such, they require a high degree of control. Thus, an electric heater is used with an associated heater efficiency of 80%. For simplicity, a constant heater efficiency will be used as indicated in table B.7. The overall heating power demand can be determined when the change in thermodynamic properties is known as,

$$P_{heat} = \sum_{FCU} \left(\frac{\dot{Q}_{heat-FCU,i}}{\eta_{heat}} \right)_i \tag{B.79}$$

Cooling Power (3-4)   Cooling within the the FCU relies on chiller fluid to reduce the supply air temperature. These systems generally act as low and high-temperature reservoirs where the coefficient of performance (COP) is represented by the extracted heat over the system's net work within the heating cycle [99]. The general reservoir process simplification can be seen highlighted in figure B.2. These

BOILER SYSTEM   CHILLER SYSTEM
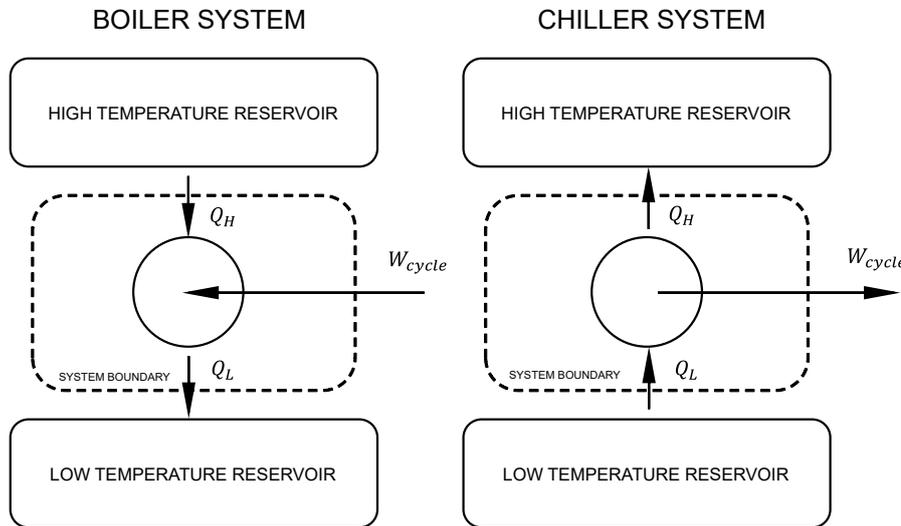
Figure B.2: High and low reservoir schematic simplification, (Stapersma and Klein Houd [99])

Therefore, the chiller process is reversed from that of the boiler system, thus heat is instead extracted. The chiller system parameters are a function of many dynamic factors such as multi-stage condensing efficiencies, and refrigerant usage. Thus, for the sake of simplicity a constant chiller COP and heat exchanger efficiency will be considered as seen in table B.7.

$$P_{cool} = \sum_{FCU} \left(\frac{\dot{Q}_{cool-FCU,i}}{\eta_{HE}\,COP_{chiller}} \right)_i \tag{B.80}$$

Humidifier Pump Power (4-5)   Moisture must be added to the environment to ensure the ambient conditions are not too dry. As such, a fluid must be pumped and converted to steam to achieve this. While these mechanical systems are generally not a main contributor , the electrical demand must still be considered. Much like fan power, the power estimation considers the required volumetric flow rate of steam ($\dot{V}_{steam}$), the pressure pumping head loss (ΔP_{hum}), and a humidification efficiency factor (η_{hum}).

$$P_{humid} = \sum \left(\frac{\dot{V}_{steam,i}\,\Delta P_{hum}}{\eta_{hum}} \right)_i \tag{B.81}$$

Where the pump pressure losses, as investigated by Boertz [15], are assumed to be 80bar based on supplier data. As well general humidification efficiencies, based on Stapersma and Klein Houd [99], have general efficiencies around 80%. Since the water flow rate is considered the determined heat gain can be converted into a volumetric flow through the used of the general heat flux formulation,

$$\dot{Q}_{steam} = \dot{m} \cdot \Delta h_{hum} = \rho_{steam} \dot{V}_{steam} \cdot \Delta h_{hum} \tag{B.82}$$

Where the measured density of saturated steam at 100°C is 0.6 kg/m^3.

B.1.7. Hotel Heat Gain - Equipment Efficiencies

The associated HVAC equipment intermediary efficiencies and coefficients of performances (COP) can be seen listed in table B.7. It should be noted that many of these parameters are mere estimates and can vary from the real-world operation. As such, the listed efficiencies are appropriate for an early-stage design where a more considerable degree of variation and uncertainty exists.

Table B.7: HVAC equipment efficiencies and coefficient of performances

Parameter	Symbol	Efficiency & System COP	Remarks
Fan efficiency	η_{fan}	60%	Stapersma and Klein Houd [99]
Fan motor efficiency	η_{em}	90%	Stapersma and Klein Houd [99]
Heater efficiency	η_{heat}	80%	Stapersma and Klein Houd [99]
Heat exchanger efficiency	η_{HE}	80%	Stapersma and Klein Houd [99]
Humidifier pump efficiency	η_{hum}	80%	Stapersma and Klein Houd [99]
Chiller COP	$COP_{Chiller}$	5	Cruise Ship Reference, Boertz [15]

B.2. Black Box Model Calculation Procedures

B.2.1. Artificial Neural Networks - Activation Functions

Simply put, an activation function is a function that is added into an artificial neural network in order to help the network learn intricate patterns in the data. Ultimately, these are transform functions that convert and introduce non-linearities within the neural network. A broad range of activation functions exists for various applications such as pattern classification or regression (curve fitting). Figure B.3 highlights some commonly applied transformations, which are further detailed below.

(a) **Linear Activation Function**
The linear activation function produces an output result equal to the activation potential of x.

$$f(x) = x \quad \text{for all } x \tag{B.83}$$

da Silva et al. [29] outlines that this type of function is typically applied when performing universal curve fitting (function approximation), as it directly maps the input/output variables of a particular process. As such, the final output layer of a regression ANN is typically comprised of a linear activation function.

(b) **Binary Step (Sign) Activation Function**
The step activation function produces a unitary value depending on the neuron activation potential threshold. If the input exceeds its threshold, then the output is positive; otherwise, it is negative.

$$f(x) = \begin{cases} 1 & \text{for } x \geq 0 \\ -1 & \text{for } x < 0 \end{cases} \tag{B.84}$$

These functions are partially differentiable and are commonly applied within the pattern classification ANNs.
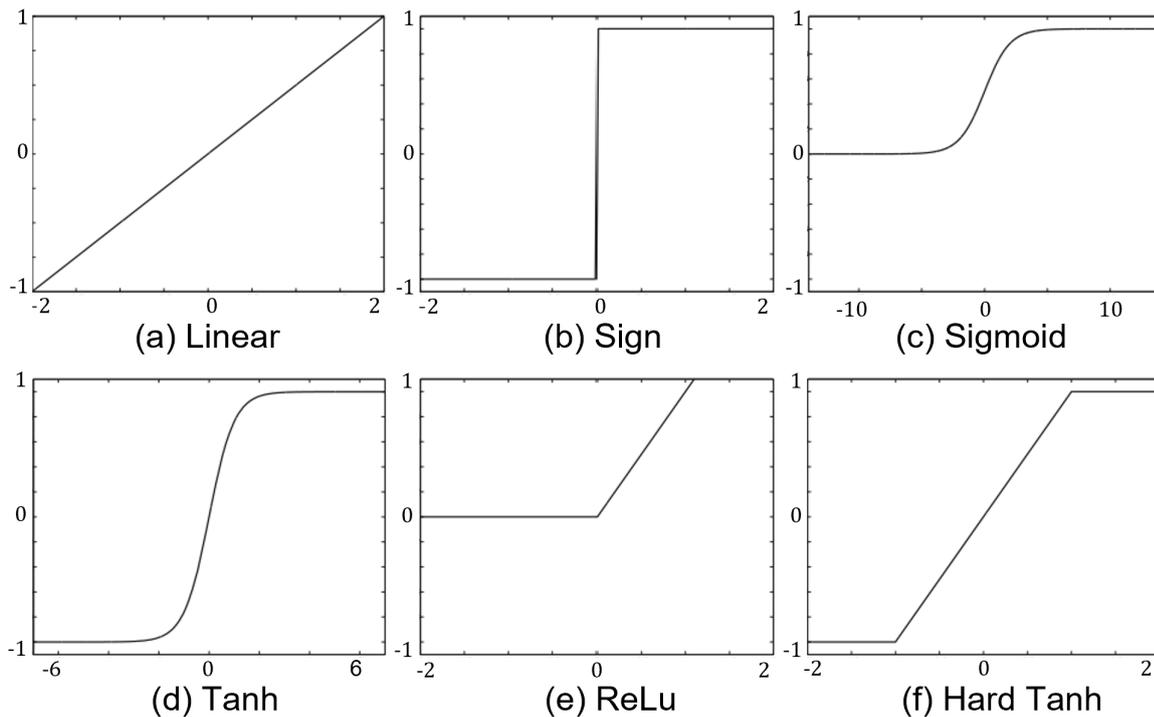
Figure B.3: Commonly applied activation functions, (Aggarwal [3])

(c) **Sigmoid (Logistic) Activation Function**
The Sigmoid or Logistic activation function is fully differentiable and always assumes real values between 0 and 1, (da Silva et al. [29]). The function is similar in form to the binary step function when the slope variable, β tends towards infinity.

$$f(x) = \frac{1}{1 + e^{(-\beta \cdot x)}} \tag{B.85}$$

This function is one of the most common activation functions used within a neural network's hidden layers.

(d) **Hyperbolic Tangent (Tanh) Activation Function**
The Tanh activation function behaves similarly to the sigmoid; however, the bounding values range between [-1,1] rather than [0,1].

$$f(x) = \frac{e^x - e^{(-\beta \cdot x)}}{e^x + e^{(-\beta \cdot x)}} \tag{B.86}$$

As detailed by da Silva et al. [29], both logistic and hyperbolic tangent functions belong to a sigmoidal family of functions. As such, due to their fully differentiable forms, they have been commonly applied within regression-based ANNS.

(e) **Rectified Linear Unit (ReLU) Activation Function**
The rectified linear unit activation function is a modern solution to deep neural architectures due to faster and practical training on large and complex datasets.

$$f(x) = \begin{cases} 0 & \text{for } x \geq 0 \\ x & \text{for } x < 0 \end{cases} \tag{B.87}$$

The function has found much success in computer vision and speech applications.

(f) **Hard Tanh Activation Function**

As detailed by Aggarwal [3], the hard tanh function is very similar to the regular tanh function. However, it is sometimes favoured as it is computationally cheaper than the former.

$$
f(x) = \begin{cases} -1 & \text{for } x < 0 \\ x & \text{for } -1 \leq x \leq 1 \\ 1 & \text{for } x > 1 \end{cases}
\tag{B.88}
$$

Unlike the regular tanh function, the hard tanh saturates for thresholds [-1,1]; thus, offering a more limited range of application.

B.2.2. Artificial Neural Networks - Backward Propagation Calculation Procedure

As further explored by da Silva et al. [29], the backward propagation stage can be decomposed into two main parts:

Part 1: Adjusting the synaptic weights of the output layer

$$
\nabla E_{MSE}^{(L)} = \frac{\partial E}{\partial W_{j,i}^{(L)}} = \frac{\partial E}{\partial \hat{Y}_j^{(L)}} \cdot \frac{\partial \hat{Y}_j^{(L)}}{\partial I_j^{(L)}} \cdot \frac{\partial I_j^{(L)}}{\partial W_{j,i}^{(L)}}
\tag{B.89}
$$

Where each individual component can be decomposed as,

$$
I_j^{(L)} = \sum_{i=0} \hat{Y}_j^{(L)} \cdot W_{j,i}^{(L+1)}; \quad \frac{\partial I_j^{(L)}}{\partial W_{j,i}^{(L)}} = \hat{Y}_j^{(L)}; \quad \frac{\partial \hat{Y}_j^{(L)}}{\partial I_j^{(L)}} = g'^{(L)} \cdot (I_j^{(L)}); \quad \frac{\partial E}{\partial \hat{Y}_j^{(L)}} = \sum_{j=1} \frac{\partial E}{\partial I_j^{(L)}} \cdot \frac{\partial I_j^{(L)}}{\partial \hat{Y}_j^{(L)}}
\tag{B.90}
$$

For the outer layer, when considering the three-layer schematic seen in figure 4.8, the following relations can be derived,

$$
\frac{\partial E}{\partial \hat{Y}_j^{(3)}} = -(Y_j - \hat{Y}_j^{(3)})
\tag{B.91}
$$

Thus applying and replacing all individual components of the above gradient relation we have,

$$
\frac{\partial E}{\partial W_{j,i}^{(3)}} = -(Y_j - \hat{Y}_j^{(3)}) \cdot g'^{(3)} \cdot (I_j^{(3)}) \cdot Y_j^{(2)}
\tag{B.92}
$$

From which the adjustment of the weight matrix, $W_{j,i}^{(3)}$, must be made in the opposing gradient direction to minimize the respective error as,

$$
\Delta W_{j,i}^{(3)} = -\eta \cdot \frac{\partial E}{\partial W_{j,i}^{(3)}} = \eta \cdot \delta_j^{(3)} \cdot Y_j^{(2)}
\tag{B.93}
$$

Where, δ consist of both the error and activation components and η is defined as the variable learning rate of the back-propagation algorithm. This defines the local gradient related to the jth neuron in the output layer. Following an iterative updating procedure the above expression can be rewritten as,

$$
W_{j,i}^{(3)}(t+1) = W_{j,i}^{(3)}(t) + \eta \cdot \delta_j^{(3)} \cdot Y_j^{(2)}
\tag{B.94}
$$

Therefore, by taking into account the differences between the observed responses of the network and the desired targets, adjustments of the neuron weights can be completed iteratively for the outer layer.

Part 2: Adjusting the synaptic weights of the intermediate layers The output layer has direct access to the desired values for their output difference measurements. However, the intermediate layers do not. As such, adjustment of their synaptic weights is performed through estimation of the output

error related to the neurons in the advance positions which have already been adjusted. Ultimately, the backward propagation procedure seen in part 1, is similarly followed with slight differences,

$$\frac{\partial E}{\partial \hat{Y}_j^{(L)}} = \sum_{j=1} \frac{\partial E}{\partial I_j^{(L+1)}} \cdot \frac{\partial I_j^{(L+1)}}{\partial \hat{Y}_j^{(L)}} = \sum_{j=1} \frac{\partial E}{\partial I_j^{(L+1)}} \cdot \frac{\partial (\sum_{i=0} \hat{Y}_j^{(L)} \cdot W_{j,i}^{(L+1)})}{\partial \hat{Y}_j^{(L)}} \tag{B.95}$$

Where the value of the partial derivative of the second fraction term with respect to $Y_j^{(L)}$, can be simplified as,

$$\frac{\partial E}{\partial \hat{Y}_j^{(L)}} = \sum_{j=1} \frac{\partial E}{\partial I_j^{(L+1)}} \cdot W_{j,i}^{(L+1)} = -\sum_{j=1} \delta_k^{(L+1)} \cdot W_{j,i}^{(L+1)} \tag{B.96}$$

Where, $delta_k$ is a condensed parcel composed of the partial error gradient. Therefore, the full intermediate error-weight gradients can be determined as,

$$\frac{\partial E}{\partial W_{j,i}^{(L)}} = -\left(\sum_{j=1} \delta_k^{(L+1)} \cdot W_{j,i}^{(L+1)} \right) \cdot g'^{(L)} \cdot (I_j^{(L)}) \cdot Y_j^{(L-1)} \tag{B.97}$$

Repeating the similar gradient adjustment procedure seen in part 2, the minimization of the forward error can be made using,

$$\Delta W_{j,i}^{(L)} = -\eta \cdot \frac{\partial E}{\partial W_{j,i}^{(L)}} = \eta \cdot \delta_j^{(L)} \cdot Y_j^{(L-1)} \tag{B.98}$$

Where, δ_j is the local neuron gradient of the corresponding intermediate layer,

$$\delta_j^{(L)} = -\left(\sum_{j=1} \delta_k^{(L+1)} \cdot W_{j,i}^{(L+1)} \right) \cdot g'^{(L)} \cdot (I_j^{(L)}) \tag{B.99}$$

At this point an iterative updating procedure the above expression can be rewritten as,

$$W_{j,i}^{(L)}(t+1) = W_{j,i}^{(L)}(t) + \eta \cdot \delta_j^{(L)} \cdot Y_j^{(L_1)} \tag{B.100}$$

It should be noted that the initial input layer, x_i, is similarly used in place of the intermediate $Y_j^{(L_1)}$ error function when analyzed first layer propagation. At this final stage, the error has been propagated through the whole network, and all synaptic weights have been adjusted towards the more optimal solution. As further expressed by da Silva et al. [29], the procedures for adjusting the weight matrices can be applied for any topology of the MLP neural network, independent of the number of intermediate layers.

$$C$$
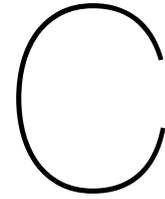
Appendix

C.1. BBM Small-Scale Model

The following appendix details the small-scale proof-of-concept model seen within section 4.4. Therefore, the following sections present the details and additional visualizations further to enhance the BBM – Artificial Neural Network (ANN) approach.

C.1.1. Small Scale Model Admiralty Coefficient - CONFIDENTIAL

The following section contains *Feadship* classified information; therefore, it is removed from the report to maintain confidentiality.

C.1.2. Small Scale Model Data

In addition to the common Admiralty relation, a random variable function was added to introduce a degree of noise on the system. The actual estimation and added noise component for speed-power and displacement-power can be seen in figure C.1, respectively.
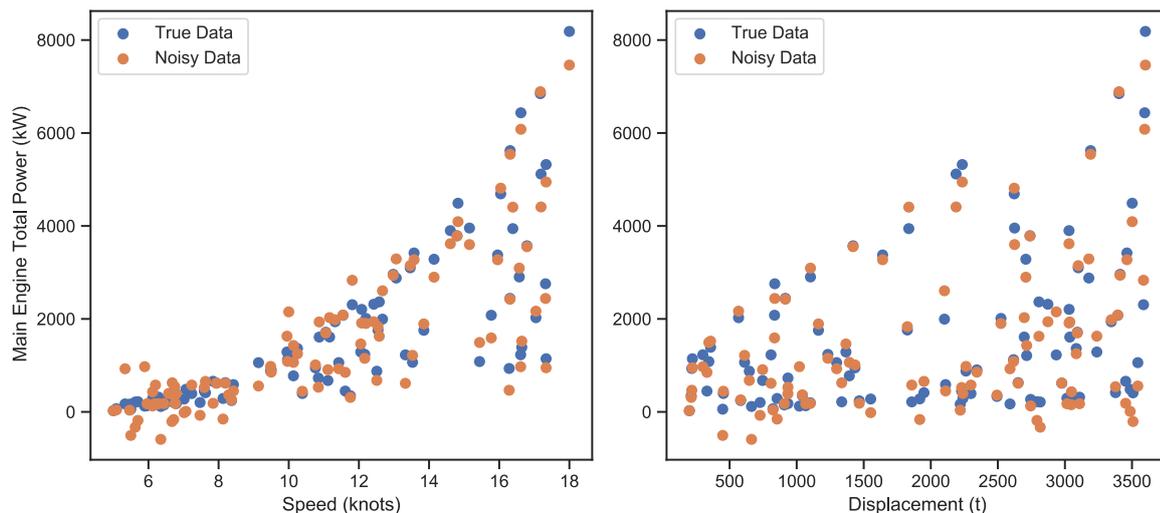
Figure C.1: True and noisy data visualization for both speed (left) and displacement (right)

From the above results, it can be seen that the random noise component actually provides some

estimation that falls into the negative regions. Intuitively, this does not make sense from a physical perspective. Nevertheless, these data entries are used to train the model to help understand the capabilities. The detailed breakdown of each input and output data feature can be seen within figure C.2, where a total of 100 data points are split using a 70%-15%-15% training scheme.
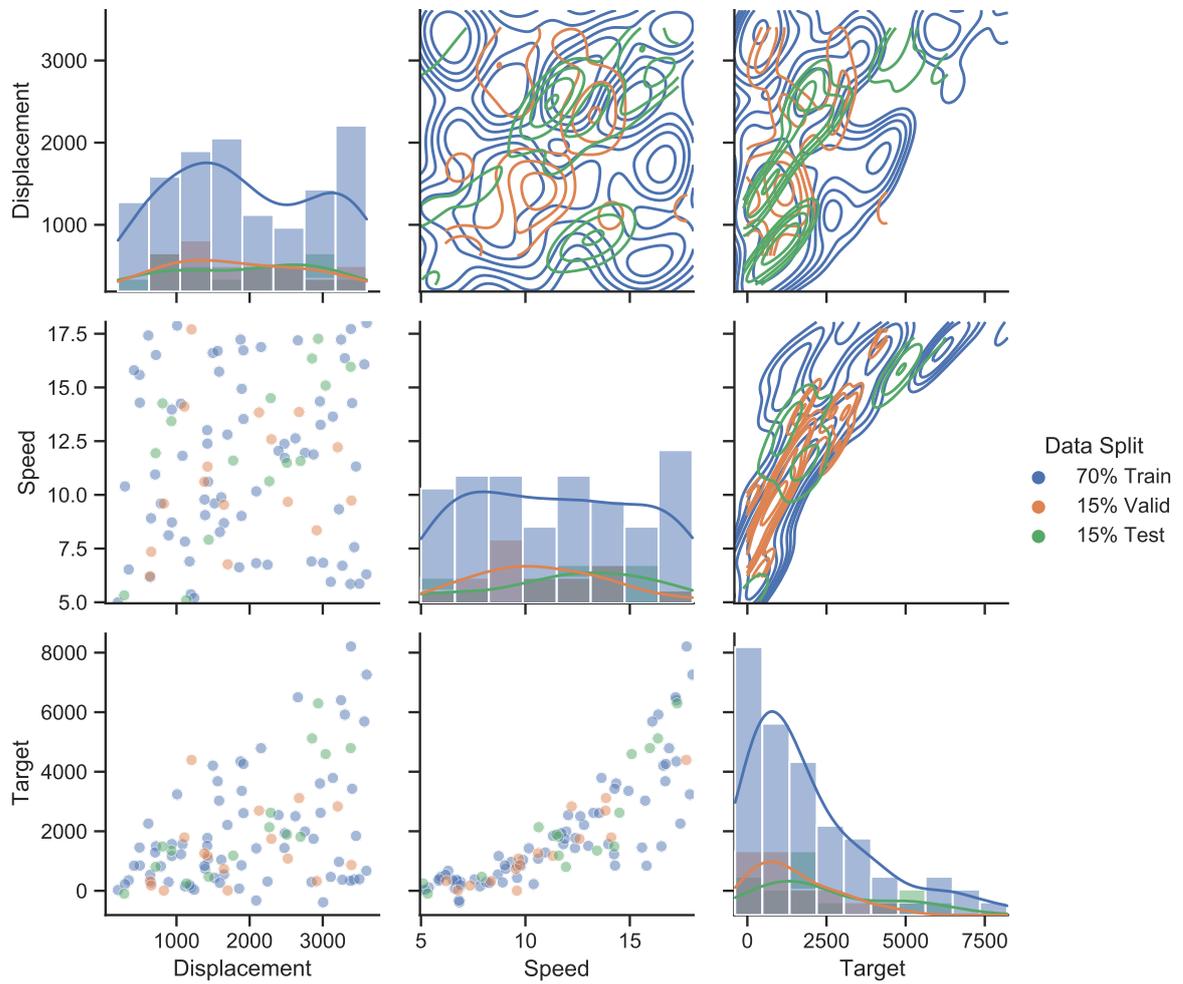
Figure C.2: Data visualization with data split percentage (n=100)

C.1.3. Small Scale Model Bootstrap Uncertainty

The bootstrapped developed probability distribution functions for both the modelling and inherent error uncertainty can be seen in figures C.3 and C.4, respectively.
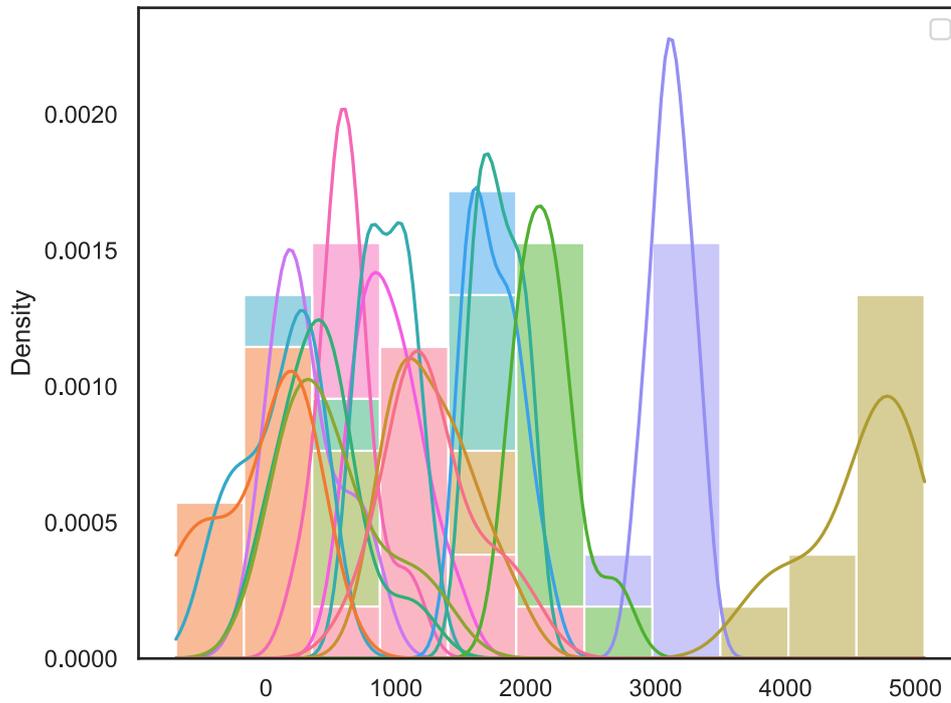
Figure C.3: Modelling prediction error probability distribution functions (B=10)
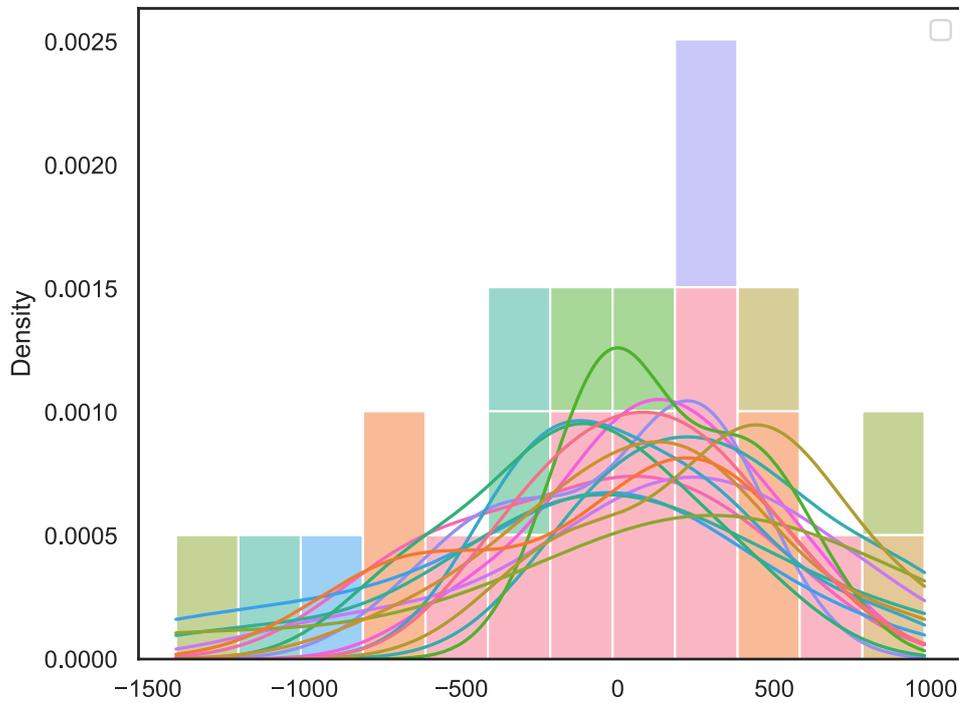
Figure C.4: Inherent residual error probability distribution functions (B=10)

C.1.4. Small Scale Model Results

The BBM-ANN prediction outcomes, including the developed uncertainty bands for both the displacement-power and speed-power relations, can be seen in figures C.5 and C.6, respectively.
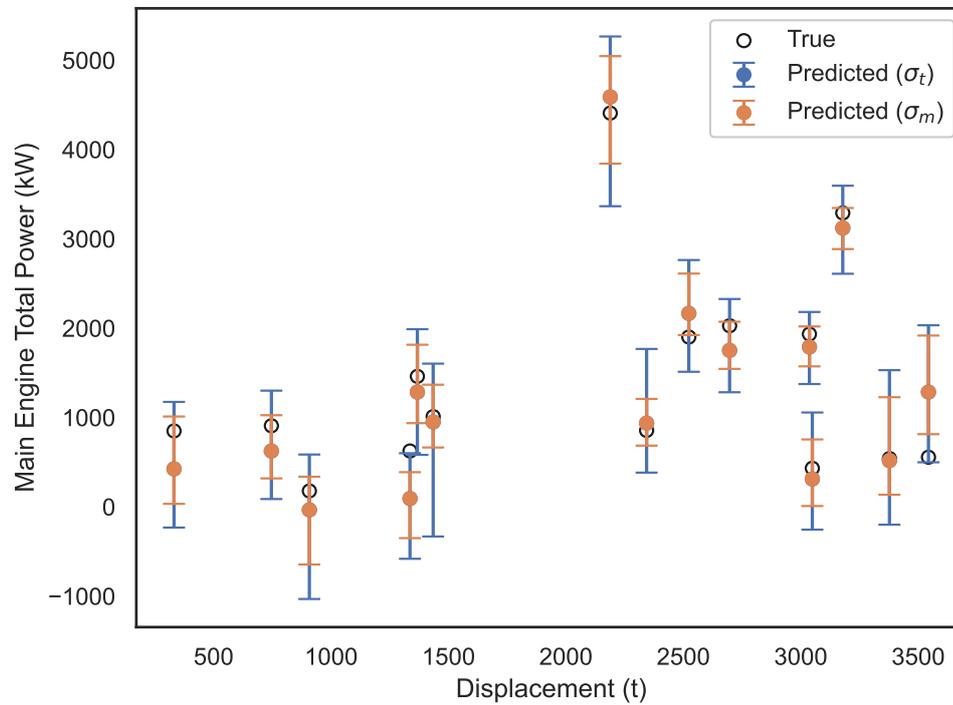
Figure C.5: Prediction performance of the trained small-scale neural network for main engine power and displacement
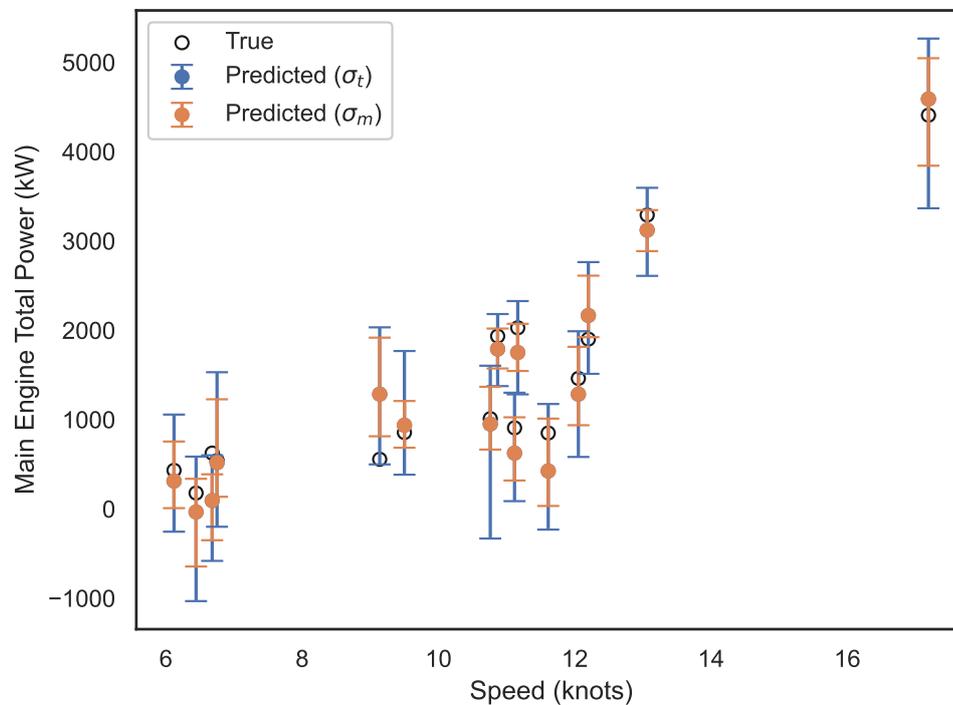
Figure C.6: Prediction performance of the trained small-scale neural network for main engine power and ship speed

D
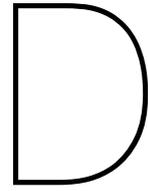
Appendix

D.1. Additional Visualization

The following appendix details the associated evaluation process and each associated visualization seen within Chapter 6. Therefore, the following sections present the figures of operational datasets, feature correlation matrices, WBM contributions, and developed prediction outcomes (without confidence bands).

D.1.1. Dataset Input Features
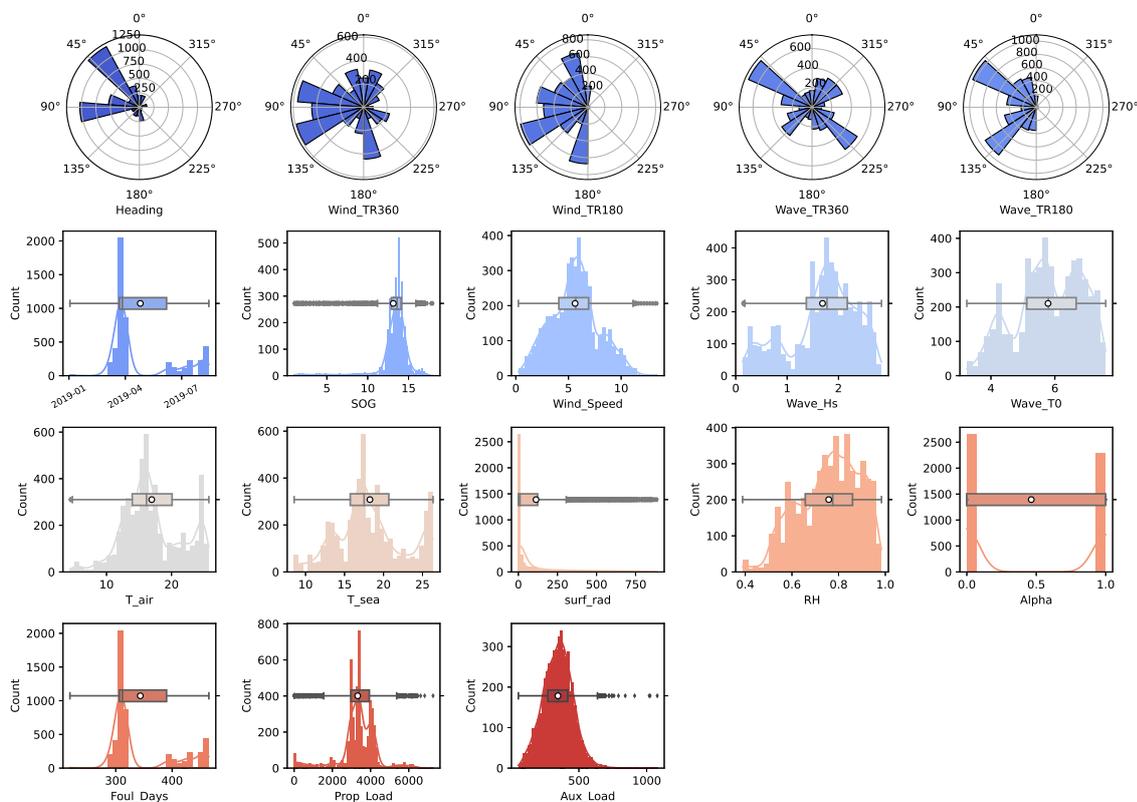
Figure D.1: Sailing operational polar, histogram, and box-plot dataset visualization
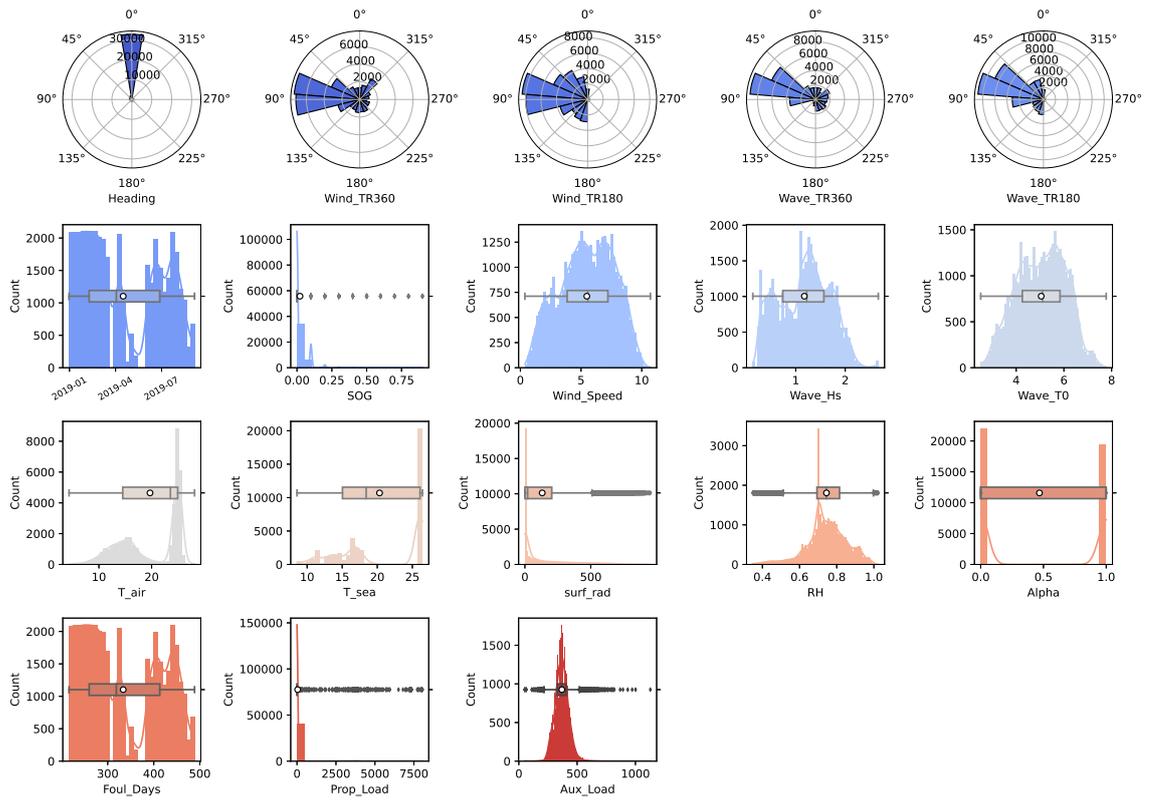
Figure D.2: Anchor operational polar, histogram, and box-plot dataset visualization
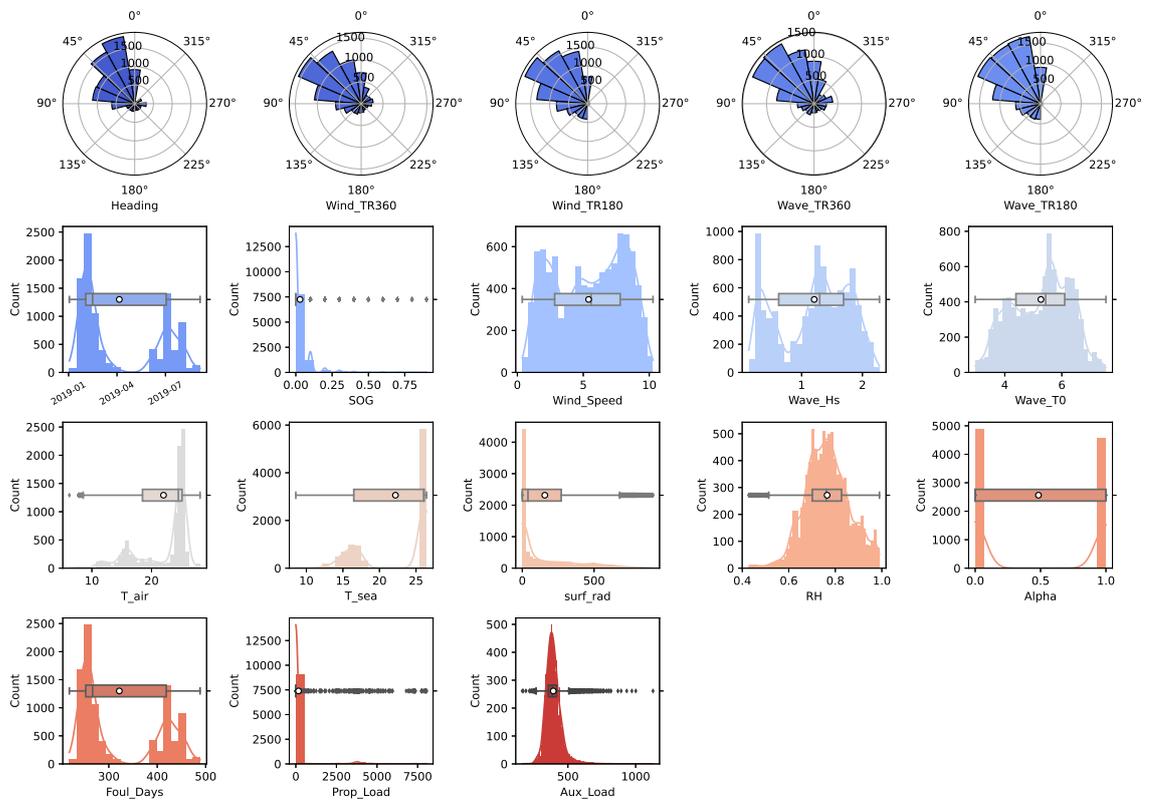
Figure D.3: Anchor (Ψ∅) operational polar, histogram, and box-plot dataset visualization
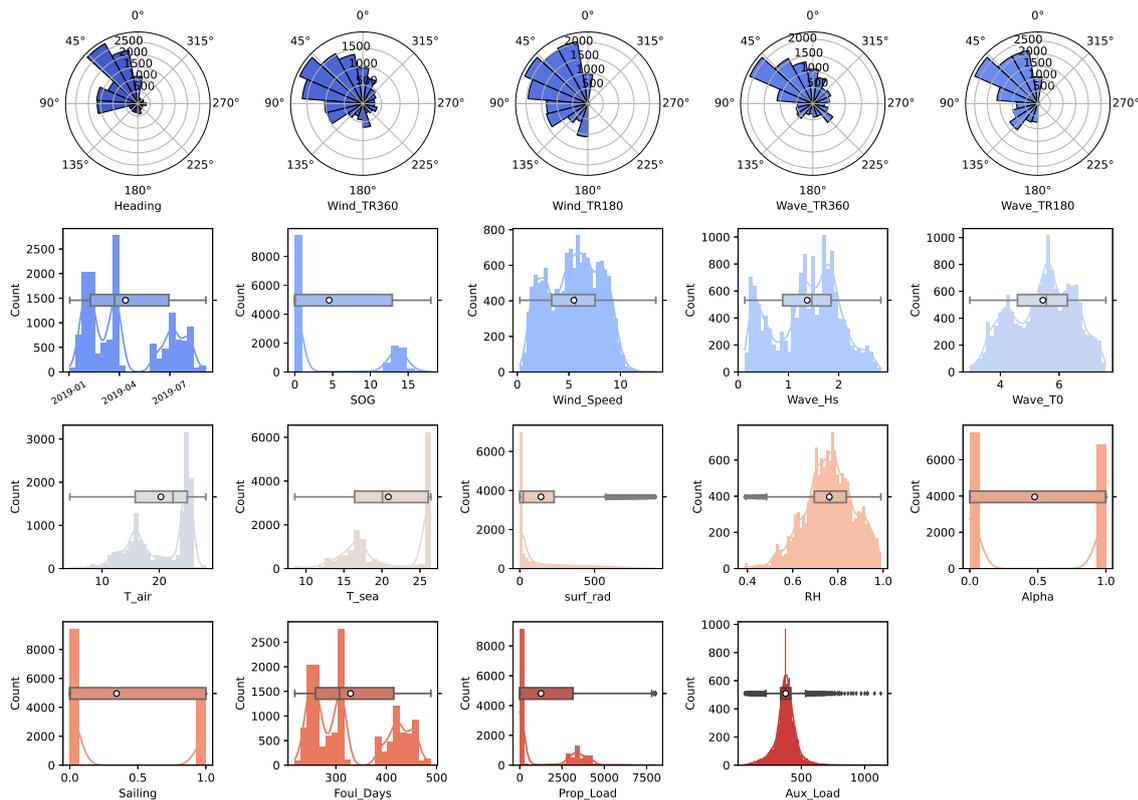
Figure D.4: Combined operational polar, histogram, and box-plot dataset visualization

Figures D.1, D.2, D.3, and D.4 identifies the individual feature visualizations for each operational dataset. While each operation requires different input variables, many of the same features are ultimately retained within each dataset. Nevertheless, clear observations among all the datasets can be observed.

The first row presents the polar directional features of the true vessel heading and the associated wind and wave relative incoming directions. Here, within the Anchor operation, apparent irregularities are noticed within the heading parameter. It can be observed that an overwhelming number of null placeholders ultimately dominate the heading data feature. Unfortunately, it isn't easy to distinguish between pure null and actual 0-degree heading. Nevertheless, upon elimination of all null headings, a more appropriate heading distribution can be identified as shown in figure D.5.
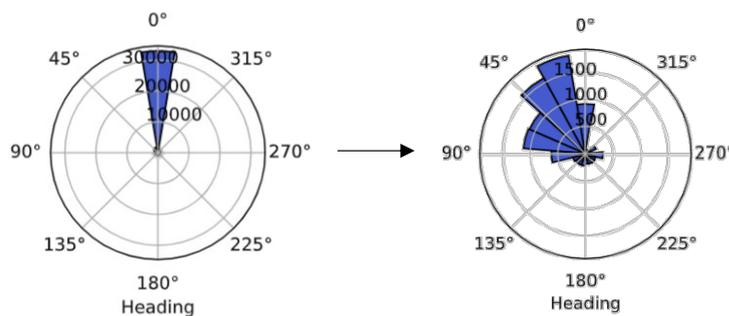
Figure D.5: Irregular heading input feature (left) cleaned heading (right)

Additionally, two binary classification features are also seen within the datasets. These relate to the sailing and daylight factor classification. Ultimately, it can be seen that the amount of time at anchor dramatically outweighs the time spent in motion for the sailing operation. In comparison, the daylight

factor is generally evenly divided. It should be noted that only for the combined condition is it required to include a binary input sailing classification. The remaining features all exhibit continuous input feature data. Ultimately it can be seen that as the datasets are orientated for different operations, the corresponding distributions are also influenced. Interesting features to note are surface radiation and temperatures. The former presents a significant data skew across all operations; thus, the number of outliers within the data feature is consistently considerable. Additionally, the temperatures indicate two prominent peaks. These can potentially be traced back to the geographical regions of the vessel and the associated environmental conditions.

D.1.2. Feature Correlations Rank Matrices

The Spearman Coefficient Rank Matrices for each associated operation (Sailing, Anchor, Anchor (H0), Combined) and outlier detection application (no IQR, IQR) can be seen in the figures D.6, D.7, D.8, D.9, D.10, D.11, D.12, and D.13, respectively.

Ultimately it can be seen that for each operation, a varying degree of possible input features is available for use to evaluate each input-output relationship univariately. Nevertheless, the highest positive correlators are represented as red squares, whereas the largest negative correlators are represented as blued squares. Based on these relations, it can be seen that while input-output features can be quantified, input-input relations can also be identified. If two input features are highly related to one another, the elimination of the lower correlator should be applied. Since these features exhibit a degree of redundancy, the overall new contributions to the artificial neural network are negligible. For instance, a relatively large degree of correlation is often noticed between the air and sea temperatures. While both features are often seen to correlate to the targeted output, only one feature should be retained as both exhibit similar dynamic contributions.
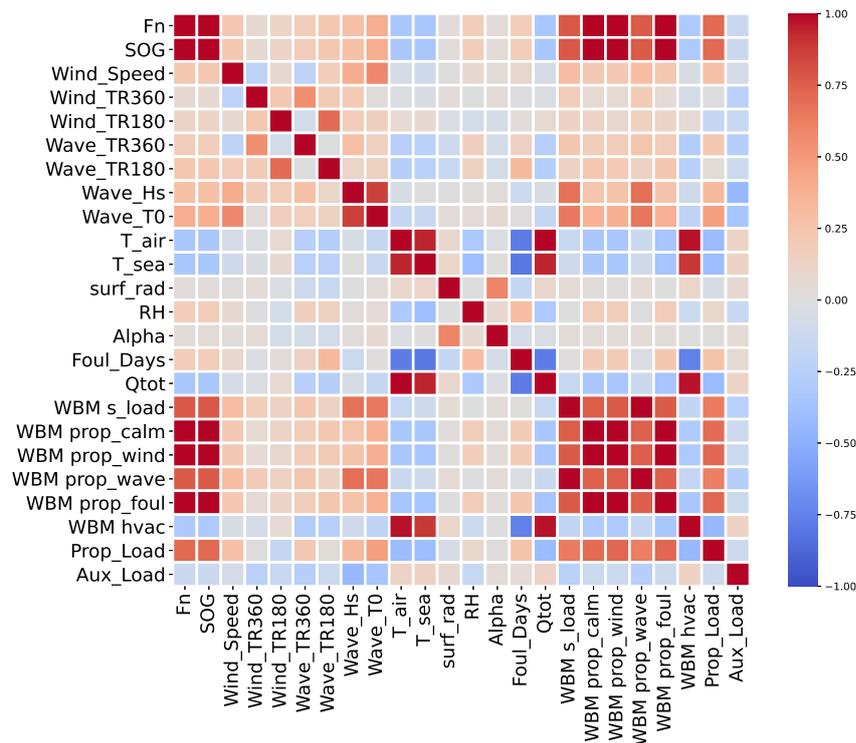
Figure D.6: Sailing (non-IQR) operation Spearman correlation rank matrix
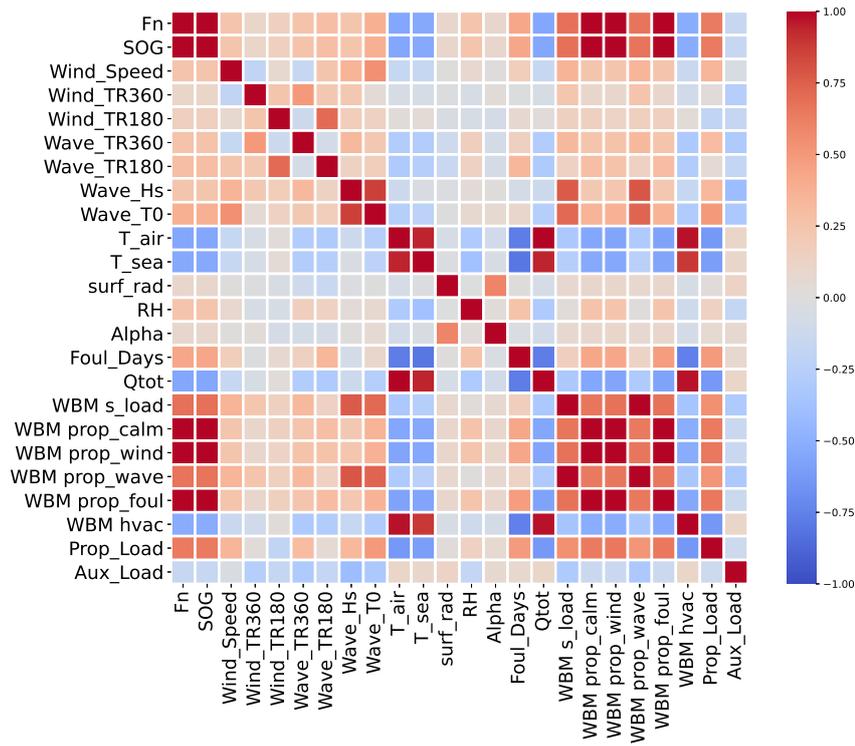
Figure D.7: Sailing (IQR) operation Spearman correlation rank matrix
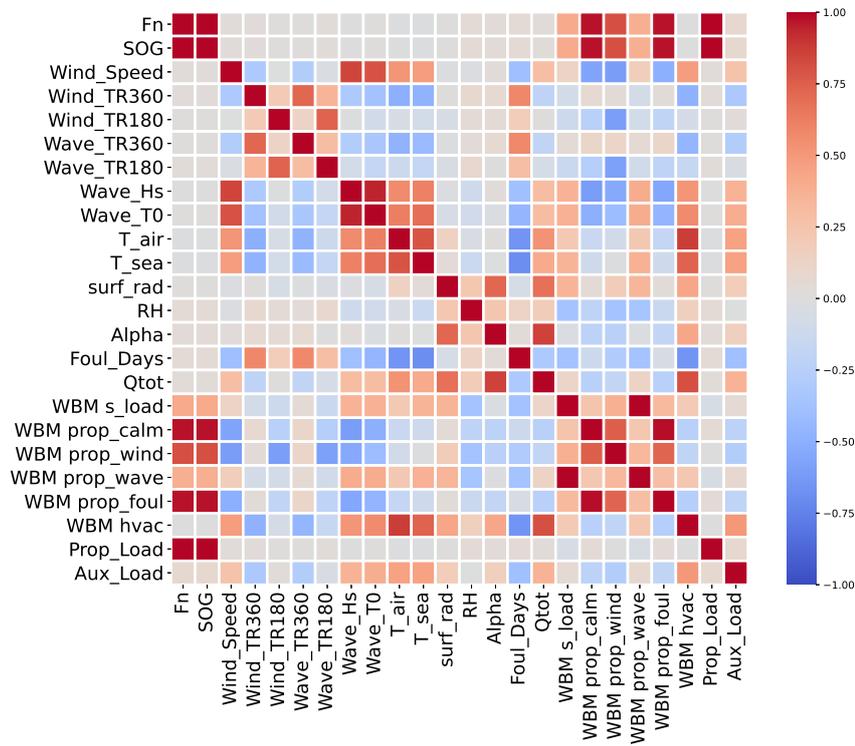
Figure D.8: Anchor (non-IQR) operation Spearman correlation rank matrix
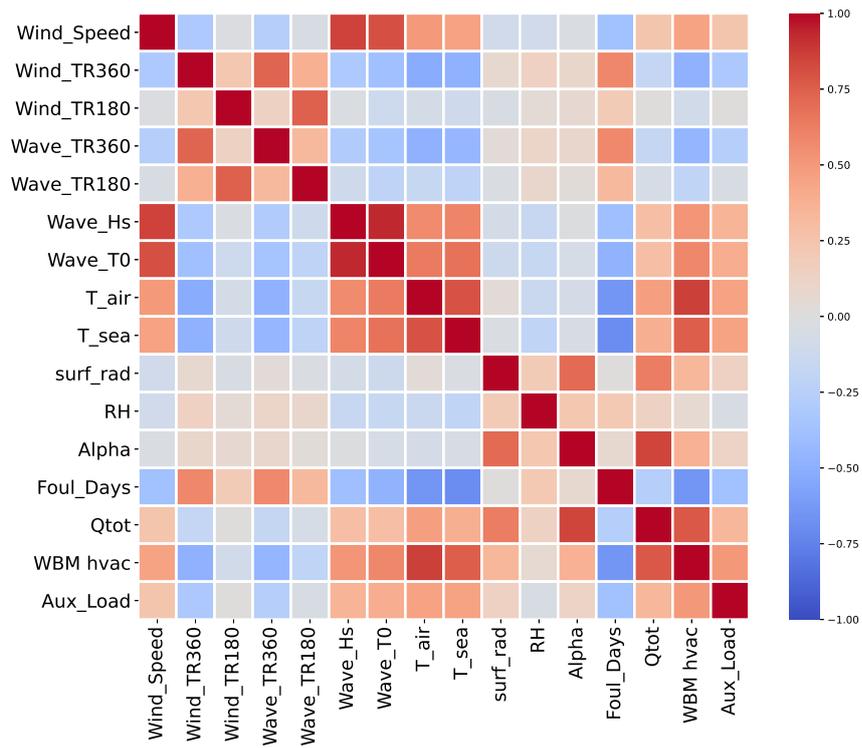
Figure D.9: Anchor (IQR) operation Spearman correlation rank matrix

Figure D.10: Anchor (Ψ∅ non-IQR) operation Spearman correlation rank matrix

Figure D.11: Anchor (Ψ∅ IQR) operation Spearman correlation rank matrix
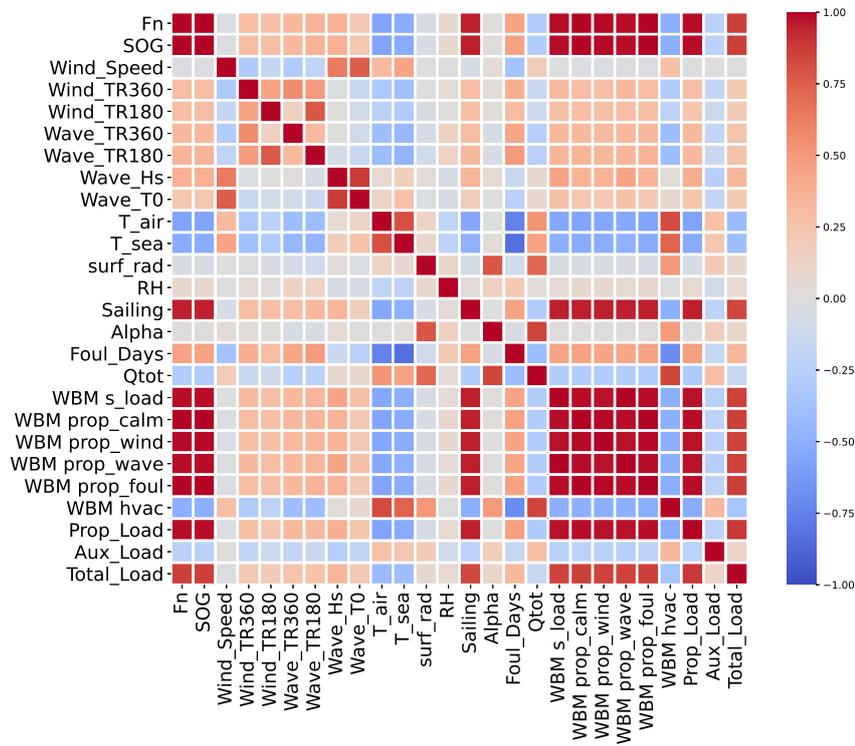
Figure D.12: Combined (Ψ∅ non-IQR) operation Spearman correlation rank matrix
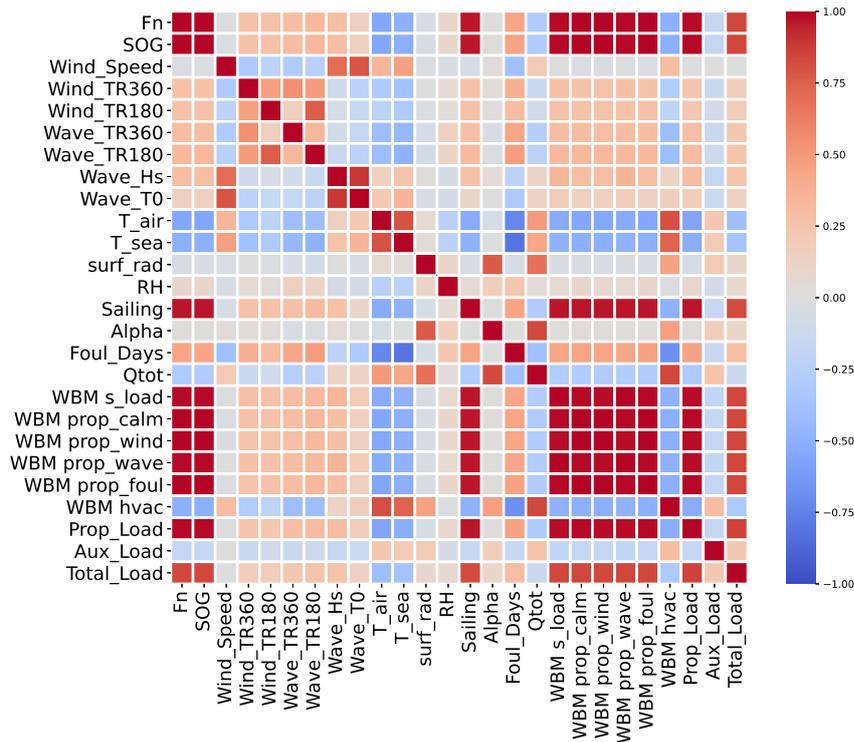
Figure D.13: Combined (Ψ∅ IQR) operation Spearman correlation rank matrix

D.1.3. WBM Contribution Comparisons

The WBM evaluations for both propulsion and auxiliary energy demand can be seen in figures D.14, D.15, D.16, D.17, D.18 and D.19, respectively. Each of these figures includes the observed effects and differences between the application of the outlier detection method.

The IQR approach can significantly reduce the data ranges. This consequence is evidently clear within the Sailing propulsion case. Ultimately, the full observed speed range reduces to a mere 5 knot prediction range. However, when applying the approach to the Anchor auxiliary cases, the effects become more favourable. Here, it can be observed that extreme outliers are ultimately eliminated. While qualitatively, the overall distributions seem more realistic without the extreme data entries, significant variance in the datasets is still observed. Therefore, genuine underlying irregularities may still exist within the data samples not captured by the IQR approach. Ultimately, the operational profiles are decomposed into three simple conditions: Sailing, Anchor, and Combined datasets. However, additional operational conditions to further decompose both Sailing and Anchoring conditions may help to further distinguish between irregular data entries.

An additional observation can be related to the degree difference in the amount of data for each operational dataset. While the anchor case has a substantial amount of entries, upon removal of the null heading irregularities, the associated data quickly reduces. Interestingly, the majority of the irregularities existed between the 4th and 7th data months. This loss may indicate a substantial system error or be related to geographical locations and lack of available GPS or system recognition. Nevertheless, a substantial portion is unfortunately eliminated due to the uncertainty behind the measurement devices.
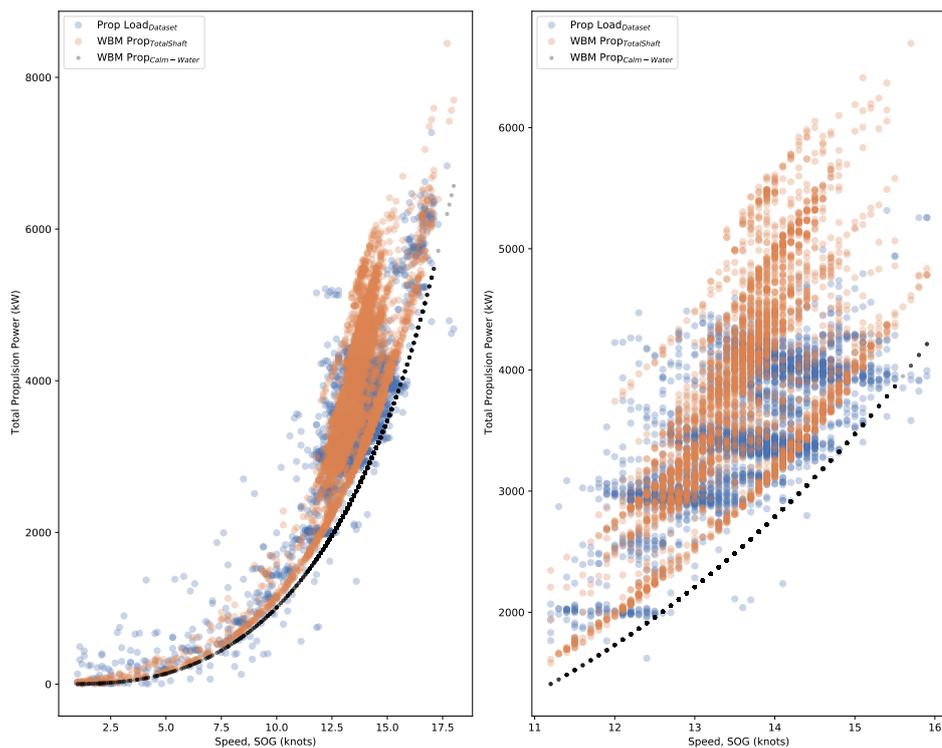
Figure D.14: WBM Sailing Propulsion outlier detection comparison between non-IQR (left) and IQR (right)
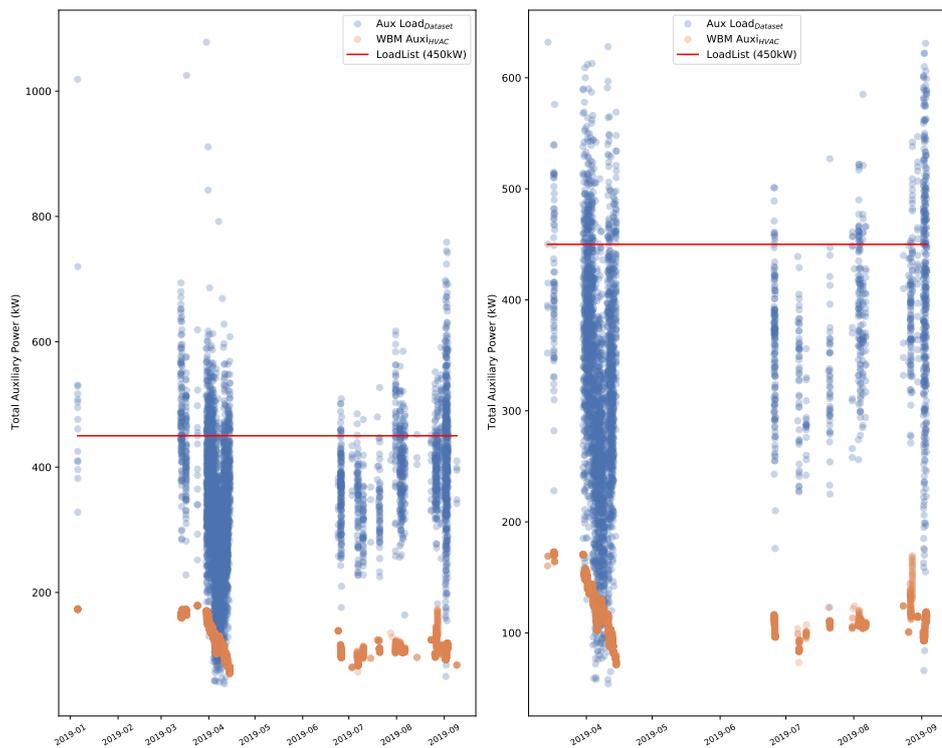
Figure D.15: WBM Sailing Auxiliary outlier detection comparison between non-IQR (left) and IQR (right)
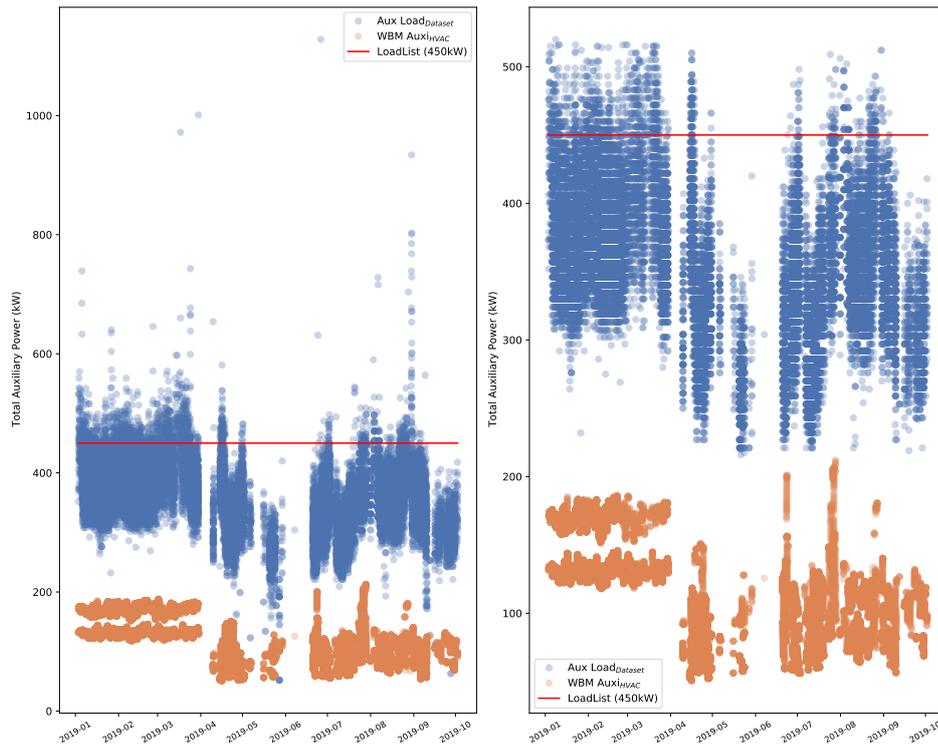
Figure D.16: WBM Anchor Auxiliary outlier detection comparison between non-IQR (left) and IQR (right)
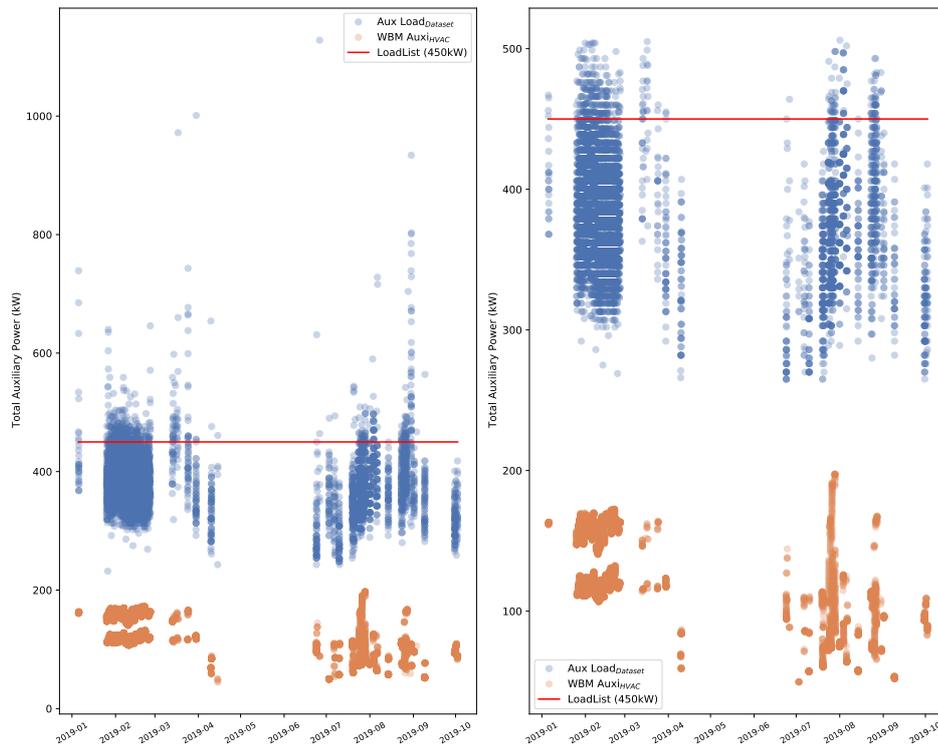
Figure D.17: WBM Anchor (Ψ∅) Auxiliary outlier detection comparison between non-IQR (left) and IQR (right)
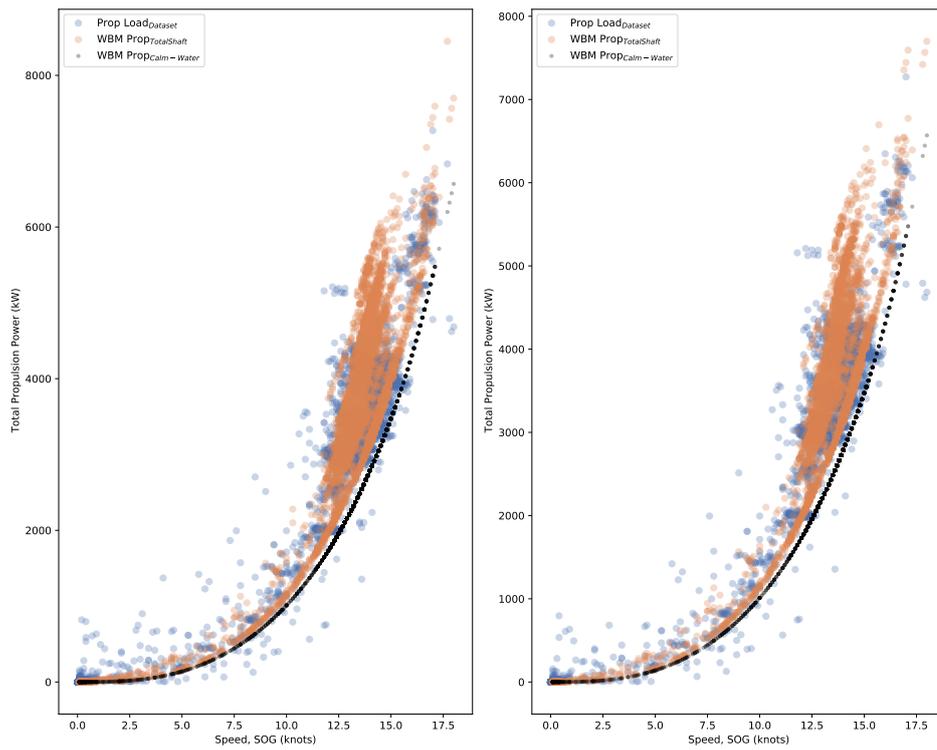
Figure D.18: WBM Combined Propulsion outlier detection comparison between non-IQR (left) and IQR (right)
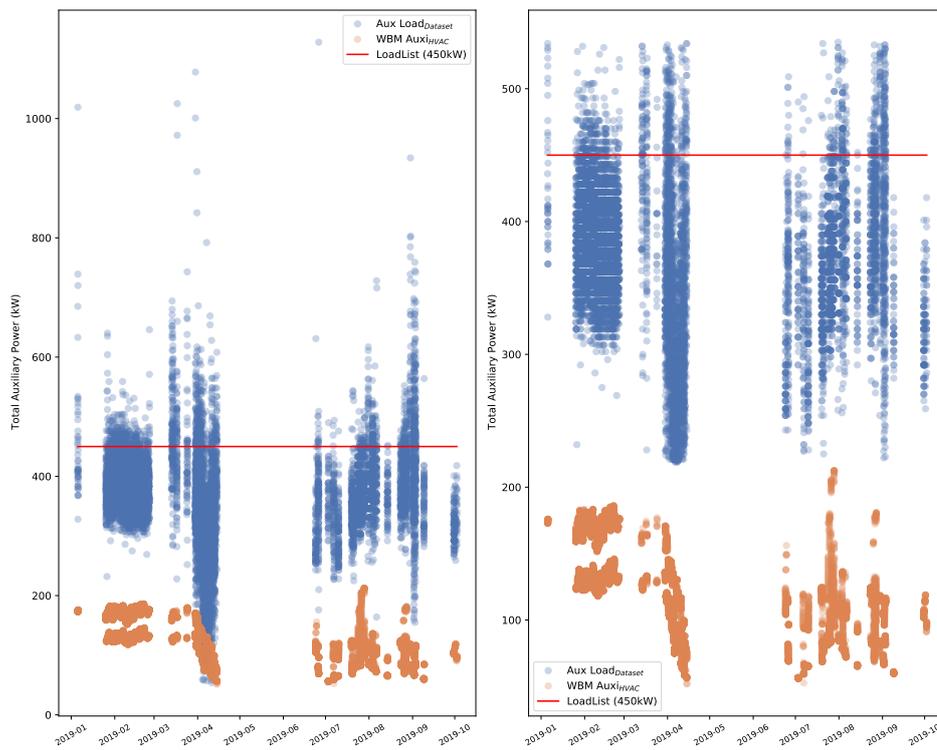
Figure D.19: WBM Combined Auxiliary outlier detection comparison between non-IQR (left) and IQR (right)

D.1.4. Modeling Outcomes with No Confidence Intervals

Figures D.20, D.21, D.22, and D.23 present visualizations of the best-developed models as evaluated within section 6.3. In this case, the bootstrapped 95% confidence intervals (modelling and inherent uncertainty) are removed for a more transparent look into the mean prediction results.
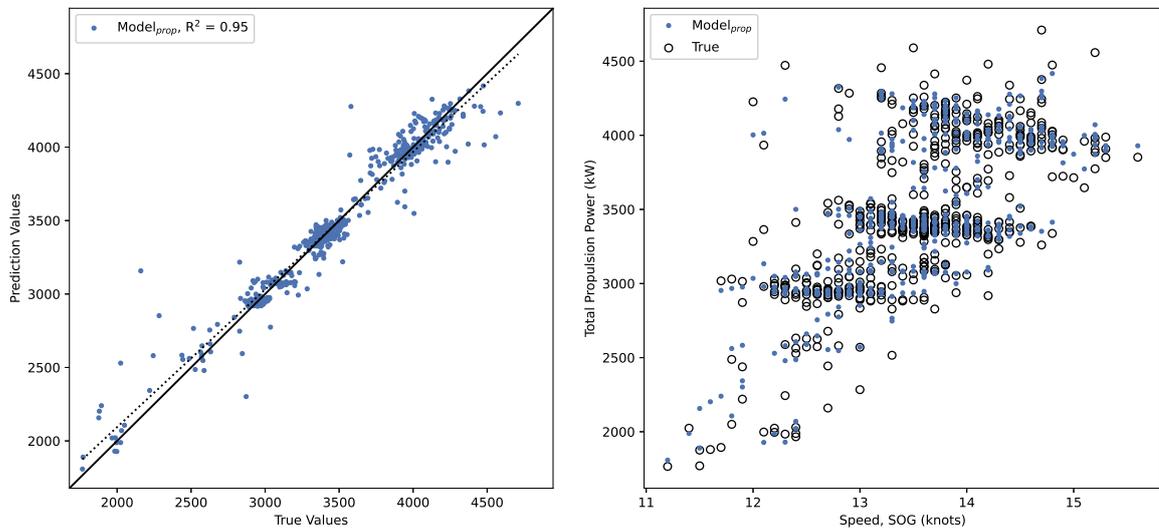
Figure D.20: IQR Sailing Propulsion $GBM + P_{s,t}$ prediction performance (left) and speed-power relationship (right) without 95%CI
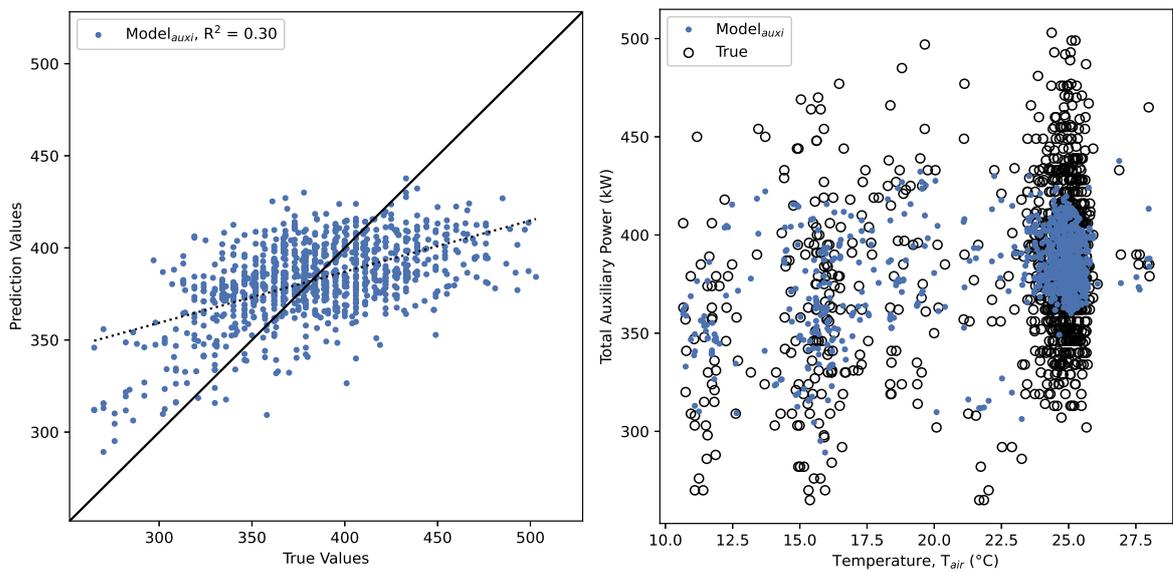
Figure D.21: IQR Anchor ($\Psi\emptyset$) Auxiliary BBM prediction performance (left) and temperature-power relationship (right) without 95%CI
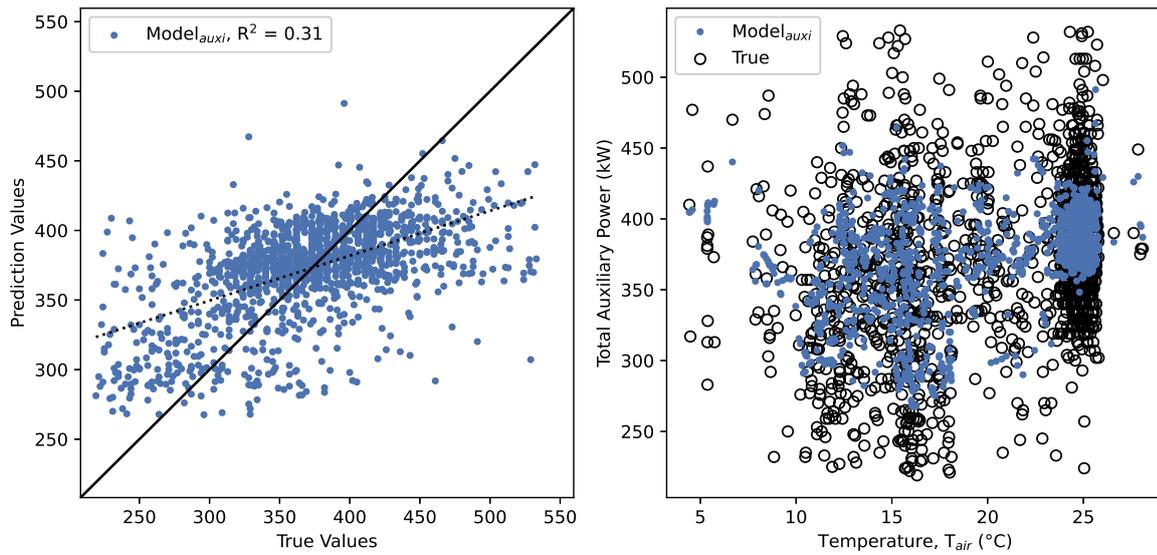
Figure D.22: IQR Combined ($\Psi\emptyset$) Auxiliary BBM prediction performance (left) and temperature-power relationship (right) without 95%CI
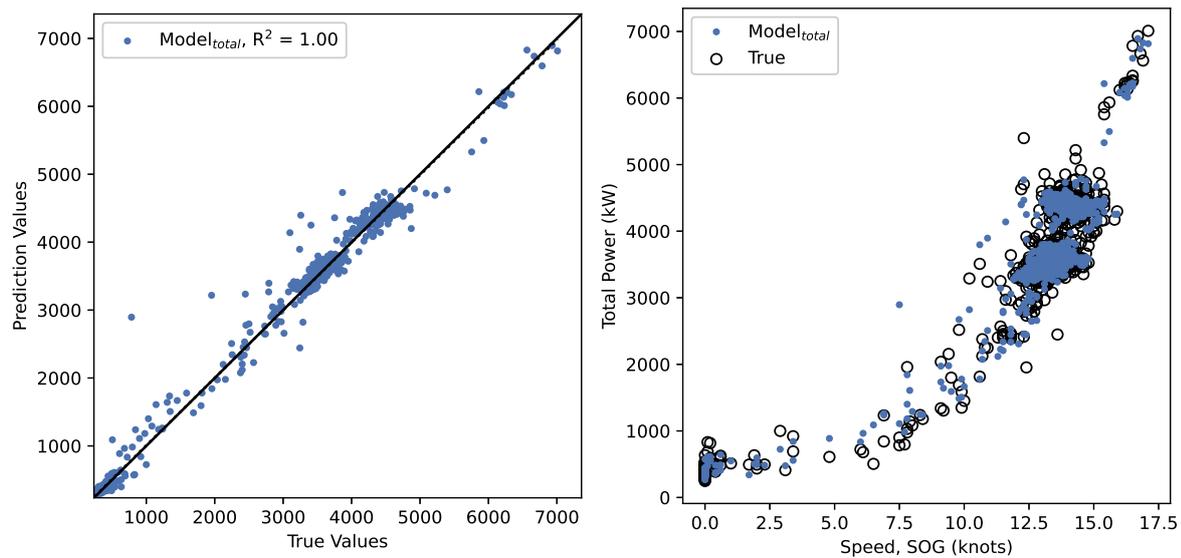
Figure D.23: None Combined Total $GBM + P_{s,cw+hvac}$ prediction performance (left) and speed-power relationship (right) without 95%CI
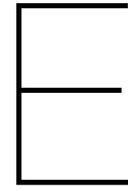
Appendix

E.1. Additional Result Summaries

The following appendix provides additional details and a summary of results related to the evaluation process seen within Chapter 6. Therefore, the following sections present a series of table summaries addressing data preparation results, additional results summaries, applicable model data ranges, and feature selection correlations.

E.1.1. Detailed Data Preparation Results

Section 6.2.2, presents an overview of the corresponding data preparation results as identified in table 6.2. However, each operation can be further detailed in tables E.1, E.2, E.3, and E.4, respectively. Here, additional details on retained interpolation and data specifications are presented. It should be noted that approximately eight self-established specifications are applied to each corresponding dataset to ensure feasible operating conditions are obtained within each input data feature.

Table E.1: Sailing operation data preparation detailed summary

Pre - Processing Step	Outlier Detection: None		Outlier Detection: IQR	
	Amount Dropped	Data Remaining	Amount Dropped	Data Remaining
Initial Datapoints	0	61944	0	61944
Duplicated Rows	0	61944	0	61944
Category Transform	0	61944	0	61944
Sailing Only	55999	5945	55999	5945
Missing Data	993	4952	993	4952
Interpolated points	236		236	
Data Specifications	185	4767	111	4841
$Data < 0$ (exl. T_{air})	0		0	
$T_{air} > 40°C$	0		0	
$Radial > 360°$	0		0	
$RH > 95\%$	153		111	
$SOG > SOG_{max}$	0		0	
$Power > Power_{max}$	0		0	
$SOG \leq 0 \& PropLoad > 0$	0		0	
$SOG > 0 \& PropLoad \leq 0$	32		0	
Outlier Drop	0	4767	1245	3596
Final Datapoints	**92.3%**	**4767**	**94.2%**	**3596**

Table E.2: Anchor operation data preparation detailed summary

	Outlier Detection: None		Outlier Detection: IQR	
Pre - Processing Step	Amount Dropped	Data Remaining	Amount Dropped	Data Remaining
Initial Datapoints	0	61944	0	61944
Duplicated Rows	0	61944	0	61944
Category Transform	0	61944	0	61944
Anchor Only	5945	55999	5945	55999
Missing Data	14506	41493	14506	41493
Interpolated points	252		252	
Data Specifications	8076	33417	552	40941
$Data < 0$ (exl. T_{air})	30		30	
$T_{air} > 40°C$	0		0	
$Radial > 360°$	0		0	
$RH > 95\%$	893		522	
$SOG > SOG_{max}$	0		0	
$Power > Power_{max}$	0		0	
$SOG \leq 0 \, \& \, PropLoad > 0$	352		0	
$SOG > 0 \, \& \, PropLoad \leq 0$	6801		0	
Outlier Drop	0	33417	11491	29450
Final Datapoints	**46.1%**	**33417**	**52.5%**	**29450**

Table E.3: Anchor operation ($\Psi \neq 0$) data preparation detailed summary

	Outlier Detection: None		Outlier Detection: IQR	
Pre - Processing Step	Amount Dropped	Data Remaining	Amount Dropped	Data Remaining
Initial Datapoints	0	61944	0	61944
Duplicated Rows	0	61944	0	61944
Category Transform	0	61944	0	61944
Anchor Only	5945	55999	5945	55999
Missing Data	14506	41493	14506	41493
Interpolated points	252		252	
Data Specifications	34158	7335	32177	9316
$\Psi \neq 0$	32031		32031	
$Data < 0$ (exl. T_{air})	0		0	
$T_{air} > 40°C$	0		0	
$Radial > 360°$	0		0	
$RH > 95\%$	354		146	
$SOG > SOG_{max}$	0		0	
$Power > Power_{max}$	0		0	
$SOG \leq 0 \, \& \, PropLoad > 0$	352		0	
$SOG > 0 \, \& \, PropLoad \leq 0$	1448		0	
Outlier Drop	0	7335	2431	6885
Final Datapoints	**88.2%**	**7335**	**88.9%**	**6885**

Table E.4: Total combined operation ($\Psi \neq 0$) data preparation detailed summary

Pre - Processing Step	Outlier Detection: None		Outlier Detection: IQR	
	Amount Dropped	Data Remaining	Amount Dropped	Data Remaining
Initial Datapoints	0	61944	0	61944
Duplicated Rows	0	61944	0	61944
Category Transform	0	61944	0	61944
Anchor + Sailing	0	61944	0	61944
Missing Data	15512	46432	15512	46432
Interpolated points	222		222	
Data Specifications	34341	12091	33976	12456
$\Psi \neq 0$	32040		32040	
$Data < 0$ (exl. T_{air})	30		30	
$T_{air} > 40°C$	0		0	
$Radial > 360°$	0		0	
$RH > 95\%$	504		448	
$SOG > SOG_{max}$	0		0	
$Power > Power_{max}$	0		0	
$SOG \leq 0$ & $PropLoad > 0$	325		245	
$SOG > 0$ & $PropLoad \leq 0$	1472		1243	
Outlier Drop	0	12091	1964	10492
Final Datapoints	**80.5%**	**12091**	**83.1%**	**10492**

E.1.2. Feature Selection Detailed Dependencies

Section 6.2.4 presents an overview of the corresponding feature selection input-output correlation results as seen within table 6.3. However, this summary only presents the top 5 correlators for each summary case. In reality, each feature relationship can be univariately quantified for each operation, as presented in tables E.5, E.6, and E.7. Here, a colouring scheme is applied to identify the associated relationships quickly. Green indicates a positive correlation, whereas red indicates a negative one. Additionally, the WBMs are also highlighted orange for improved clarity. In this case, a mean absolute Spearman Correlation of the top 10 features is presented.

Table E.5: Spearman correlation results summary for sailing operation

	Outlier Detection: None				Outlier Detection: IQR			
Sailing:	Propulsion		Auxiliary		Propulsion		Auxiliary	
Rank	Input	SCo	Input	SCo	Input	SCo	Input	SCo
1	V_s	+0.71	H_s	-0.43	V_s	+0.64	H_s	-0.41
2	$P_{s,cw}$	+0.71	T_0	-0.34	$P_{s,cw}$	+0.63	T_0	-0.33
3	$P_{s,t}$	+0.64	α_{360}	-0.26	T_{air}	-0.63	α_{360}	-0.30
4	T_0	+0.48	β_{360}	-0.23	$P_{s,t}$	+0.53	β_{360}	-0.26
5	T_{air}	-0.42	RH	-0.16	T_0	+0.49	RH	-0.18
6	H_s	+0.33	P_{hvac}	+0.14	HCI	+0.49	V_s	-0.15
7	V_{wi}	+0.28	V_s	-0.13	V_{wi}	+0.35	E_e	+0.12
8	HCI	+0.26	T_{air}	+0.13	H_s	+0.33	T_{air}	+0.11
9	α_{360}	+0.22	V_{wi}	-0.07	α_{360}	+0.31	P_{hvac}	+0.11
10	β_{360}	+0.00	E_e	+0.06	β_{360}	+0.02	α_{sun}	+0.07
11	-	-	α_{sun}	+0.04	-	-	V_{wi}	-0.04
Target	P_s	1.00	P_a	1.00	P_s	1.00	P_a	1.00
MASCo T10	0.41		0.20		0.44		0.20	

Table E.6: Spearman correlation results summary for anchor operation

	Outlier Detection: None				Outlier Detection: IQR			
Anchor:	Auxiliary		Auxiliary ($\Psi \neq 0$)		Auxiliary		Auxiliary ($\Psi \neq 0$)	
Rank	Input	SCo	Input	SCo	Input	SCo	Input	SCo
1	P_{hvac}	+0.51	P_{hvac}	+0.32	P_{hvac}	+0.50	P_{hvac}	+0.33
2	T_{air}	+0.46	E_e	+0.31	T_{air}	+0.46	E_e	+0.31
3	T_0	+0.40	α_{sun}	+0.27	T_0	+0.39	α_{sun}	+0.24
4	H_s	+0.36	T_{air}	+0.15	H_s	+0.35	T_{air}	+0.15
5	β_{360}	-0.33	β_{360}	-0.06	β_{360}	-0.33	β_{360}	-0.07
6	α_{360}	-0.27	α_{360}	-0.05	α_{360}	-0.27	α_{360}	-0.06
7	V_{wi}	+0.26	RH	+0.03	V_{wi}	+0.25	RH	+0.03
8	E_e	+0.18	V_{wi}	+0.03	E_e	+0.15	T_0	-0.02
9	α_{sun}	+0.16	H_s	+0.02	α_{sun}	+0.12	H_s	+0.00
10	RH	-0.01	T_0	+0.00	RH	-0.04	V_{wi}	+0.00
Target	P_a	1.00	P_a	1.00	P_a	1.00	P_a	1.00
MASCo T10	0.29		0.13		0.29		0.12	

Table E.7: Spearman correlation results summary for combined operation

	Outlier Detection: None				Outlier Detection: IQR			
Combined:	Auxiliary ($\Psi \neq 0$)		Total ($\Psi \neq 0$)		Auxiliary ($\Psi \neq 0$)		Total ($\Psi \neq 0$)	
Rank	Input	SCo	Input	SCo	Input	SCo	Input	SCo
1	P_{hvac}	+0.34	V_s	+0.86	P_{hvac}	+0.30	V_s	+0.84
2	T_{air}	+0.27	$P_{s,cw}$	+0.86	T_{air}	+0.23	$P_{s,cw}$	+0.84
3	H_s	-0.25	$P_{s,t}$	+0.86	E_e	+0.21	$P_{s,t}$	+0.84
4	V_s	-0.23	S_{oper}	+0.84	α_{sun}	+0.17	S_{oper}	+0.82
5	S_{oper}	-0.22	T_{air}	-0.43	H_s	-0.17	T_{air}	-0.39
6	E_e	+0.20	P_{hvac}	-0.35	V_s	-0.17	P_{hvac}	-0.31
7	T_0	-0.19	H_s	+0.34	S_{oper}	-0.16	HCI	+0.30
8	β_{360}	-0.18	HCI	+0.33	α_{360}	-0.14	H_s	+0.27
9	α_{360}	-0.18	α_{360}	+0.25	β_{360}	-0.14	α_{360}	+0.22
10	α_{sun}	+0.17	β_{360}	+0.21	T_0	-0.13	β_{360}	+0.18
11	RH	-0.09	T_0	+0.21	RH	-0.07	T_0	+0.15
12	V_{wi}	+0.01	α_{sun}	+0.10	V_{wi}	+0.01	α_{sun}	+0.11
13	-	-	E_e	+0.07	-	-	E_e	+0.09
14	-	-	RH	+0.06	-	-	RH	+0.07
15	-	-	V_{wi}	+0.00	-	-	V_{wi}	+0.00
Target	P_a	1.00	P_{Total}	1.00	P_a	1.00	P_{Total}	1.00
MASCo T10	0.22		0.53		0.18		0.50	

E.1.3. Additional Modelling Case Results

Section 6.3 presents the detailed best-model performance comparison between the three modelling categories: White-box, Black-box, and Grey-box modelling using the critical performance metrics. The remaining model evaluations can be seen presented in tables E.8, E.9, E.10, E.11, E.12, E.13, E.14 and E.15, respectively.

It should be noted that although many of the models meet the established method requirements, they are not the best. As such, they are not included in the main report. Additionally, while most models fall just outside of the error range, the Sailing Auxiliary cases present a substantial performance drop compared to the other modelling cases.

- Table E.8 does not meet the method requirement (<15%); however, the MAPE is subject to numerical inflation as the percent formulation approaches zero. The percent increase of the MAE is 42%; thus, a more accurate estimation would fall around 3.13%.

- Tables E.9 and E.10 do not meet the method requirements. Ultimately, these cases present a substantial performance drop compared to the other modelling cases. Likely causes are,

    - White-Box HVAC model only considers non-dynamic stationary HVAC conditions; thus, input-output correlations are much lower than average.
    - Much fewer data points are available compared to the other cases; thus, there are not enough entries to adequately train the model for the corresponding conditions.
    - Input features cannot adequately capture the appropriate dynamic responses. Probably, rudder and stabilization control systems play a more crucial role when the vessel is in motion. However, no feature inputs are available to quantify these portions.

- Table E.14 does not meet the requirements. However, it should be noted that this model is just on the verge of being considered sufficient for the application. Nevertheless, the combined model includes both sailing and anchoring conditions. As such, many of the limitations and deficiencies of the above case are apparent here as well.

Table E.8: non-IQR Sailing Propulsion performance summary comparisons between GBM, BBM and WBM

Sailing Propulsion - None				
Test Values	715			
Perf. Metric	$GBM + P_{s,t}$	$GBM + P_{s,cw}$	BBM	$P_{s,t}$	$P_{s,cw}$
R^2	0.946	0.949	0.935	0.349	-0.03609
ME (kW)	-7.403 (O)	-9.828 (O)	-3.92 (O)	-306.215 (O)	864.187 (U)
MAE (kW)	106.683	104.764	112.312	582.053	873.543
RMSE (kW)	222.861	216.812	244.256	777.22	980.794
Percent Error					
MPE	-10.647%	-10.892%	2.224%	-12.796%	24.386%
MAPE	14.682%	15.327%	21.257%	22.477%	28.373%
RMSPE	6.655%	6.474%	7.294%	23.234%	29.320%
CI95% Lower	453.17	441.16	469.73	-	-
CI95% Upper	393.4	393.53	423.67	-	-
Cover%	95.80%	96.22%	96.92%	-	-

Table E.9: IQR Sailing Auxiliary performance summary comparisons between GBM, BBM and WBM

Sailing Auxiliary - IQR			
Test Values	539		
Perf. Metric	$GBM + P_{hvac}$	BBM	$Load\ List$	P_{hvac}
R^2	0.334	0.328	-1.291	-4.354
ME (kW)	2.779 (U)	2.275 (U)	-116.839 (O)	214.872 (U)
MAE (kW)	64.208	64.163	129.621	215.716
RMSE (kW)	82.869	83.21	155.651	237.96
Percent Error				
MPE	-7.719%	-8.003%	-53.796%	59.901%
MAPE	23.44%	23.52%	56.204%	60.94%
RMSPE	24.86%	24.96%	46.720%	71.43%
CI95% Lower	153.6	152.68	-	-
CI95% Upper	158.96	157.63	-	-
Cover%	93.32%	92.02%	-	-

Table E.10: non-IQR Sailing Auxiliary performance summary comparisons between GBM, BBM and WBM

Sailing Auxiliary - None				
Test Values	715			
Perf. Metric	$GBM + P_{hvac}$	BBM	$Load\ List$	P_{hvac}
R^2	0.353	0.348	-0.925	-4.05
ME (kW)	-6.751 (O)	-9.226 (O)	-106.25 (O)	223.143 (U)
MAE (kW)	66.893	67.165	125.876	223.984
RMSE (kW)	86.596	86.924	153.291	248.249
Percent Error				
MPE	-12.090%	-13.017%	-50.03%	60.26%
MAPE	26.153%	26.447%	53.483%	61.29%
RMSPE	25.619%	25.716%	44.594%	72.22%
CI95% Lower	167	167.03	-	-
CI95% Upper	160.52	160.69	-	-
Cover%	92.59%	92.31%	-	-

Table E.11: non-IQR Anchor Auxiliary (Ψ∅) performance summary comparisons between GBM, BBM and WBM

Anchor Auxiliary - None ($\Psi \neq 0$)				
Test Values	1100			
Perf. Metric	$GBM + P_{hvac}$	BBM	$Load\ List$	P_{hvac}
R^2	0.338	0.332	-1.66	-26.146
ME (kW)	-0.098 (O)	-1.383 (O)	-65.34 (O)	259.258 (U)
MAE (kW)	30.492	30.612	71.348	259.258
RMSE (kW)	40.52	40.686	82.709	264.213
Percent Errors				
MPE	-1.070%	-1.423%	-18.821%	67.237%
MAPE	7.932%	7.995%	19.846%	67.237%
RMSPE	10.543%	10.586%	21.502%	68.687%
CI95% Lower	73.01	71.7	-	-
CI95% Upper	83.37	82.15	-	-
Cover%	94.91%	94.73%	-	-

Table E.12: IQR Anchor Auxiliary performance summary comparisons between GBM, BBM and WBM

Anchor Auxiliary - IQR				
Test Values	4418			
Perf. Metric	$GBM + P_{hvac}$	BBM	$Load\ List$	P_{hvac}
R^2	0.525	0.523	-3.184	-22.068
ME (kW)	0.256 (U)	-0.238 (O)	-91.595 (O)	242.223 (U)
MAE (kW)	28.434	28.541	93.083	242.223
RMSE (kW)	35.743	35.829	104.996	246.525
Percent Error				
MPE	-0.958%	-1.107%	-28.295%	67.620%
MAPE	8.052%	8.098%	28.604%	67.620%
RMSPE	9.991%	10.015%	29.295%	68.784%
CI95% Lower	62.12	62.57	-	-
CI95% Upper	72.05	72.33	-	-
Cover%	92.17%	93.35%	-	-

Table E.13: non-IQR Anchor Auxiliary performance summary comparisons between GBM, BBM and WBM

	Anchor Auxiliary - None			
Test Values	5013			
Perf. Metric	$GBM + P_{hvac}$	BBM	$Load\ List$	P_{hvac}
R^2	0.515	0.512	-2.709	-16.703
ME (kW)	0.142 (U)	0.744 (U)	-89.94 (O)	224.54 (U)
MAE (kW)	28.274	28.936	92.694	224.57
RMSE (kW)	37.757	37.861	105.241	229.931
	Percent Error			
MPE	-1.048%	-0.871%	-28.123%	62.191%
MAPE	8.081%	8.082%	28.638%	62.244%
RMSPE	10.465%	10.494%	29.229%	63.859%
CI95% Lower	65.840	66.270	-	-
CI95% Upper	78.190	78.670	-	-
Cover%	93.64%	93.86%	-	-

Table E.14: non-IQR Combined Auxiliary (Ψ∅) performance summary comparisons between GBM, BBM and WBM

	Combined Auxiliary - None (Ψ ≠ 0)			
Test Values	1814			
Perf. Metric	$GBM + P_{hvac}$	BBM	$Load\ List$	P_{hvac}
R^2	0.429	0.431	-0.983	-8.781
ME (kW)	-3.737 (O)	-2.57 (O)	-81.516 (O)	244.658 (U)
MAE (kW)	43.314	43.249	92.827	244.819
RMSE (kW)	61.479	61.353	115.766	257.082
	Percent Error			
MPE	-5.373%	-5.006%	-31.131%	64.926%
MAPE	14.287%	14.200%	33.105%	65.129%
RMSPE	16.766%	16.731%	31.417%	69.768%
CI95% Lower	117.590	117.820	-	-
CI95% Upper	122.550	124.760	-	-
Cover%	93.00%	93.11%	-	-

Table E.15: IQR Combined Total (Ψ∅) performance summary comparisons between GBM, BBM and WBM

	Combined Total - IQR (Ψ ≠ 0)				
Test Values	1574				
Perf. Metric	$GBM + P_{s,t+hvac}$	$GBM P_{s,cw+hvac}$	BBM	$P_{s,t} + LL$	$P_{s,cw} + LL$
R^2	0.992	0.992	0.991	0.929	0.898
ME (kW)	-1.856 (O)	-11.11 (O)	-1.621 (O)	-156.548 (O)	224.043 (U)
MAE (kW)	65.106	65.9	66.248	241.402	319.484
RMSE (kW)	155.704	153.545	161.562	458.422	532.033
	Percent Error				
MPE	-1.436%	-3.702%	-2.015%	-15.809%	-5.610%
MAPE	6.993%	7.590%	7.076%	18.554%	20.398%
RMSPE	9.882%	9.681%	10.186%	29.886%	34.685%
CI95% Lower	295.31	279.88	293.41	-	-
CI95% Upper	285.52	283.42	297	-	-
Cover%	97.52%	97.01%	97.78%	-	-

E.1.4. Applicable Data Ranges of Best Grey-Box Models

Tables E.16, E.17, E.18, and E.19 present the best-developed models corresponding inputs and the associated data summaries.

Table E.16: IQR Sailing Propulsion developed data range limits

Sailing Propulsion - IQR					
Details	Units	Minimum	Median	Mean	Maximum
V_s	knots	11.2	13.6	13.6	15.9
T_{air}	$°C$	4.7	16.1	16.7	25.2
V_{wi}	knots	0.2	5.8	5.8	11.1
β_{360}	°	0.0	117.7	153.0	359.8
H_s	m	0.2	1.9	1.8	2.8
T_0	s	3.3	6.0	5.9	7.6
α_{360}	°	0.0	184.5	181.9	359.9
HCI	Days	286.3	311.8	340.6	459.3
$P_{s,cw}$	kW	1408.0	2543.5	2553.5	4213.5
$P_{s,t}$	kW	1581.5	3781.8	3804.2	6695.3
P_s	kW	1620.0	3406.0	3461.0	5316.0

Table E.17: IQR Anchor Auxiliary (Ψø) developed data range limits

Anchor Auxiliary - IQR ($\Psi \neq 0$)					
Details	Units	Minimum	Median	Mean	Maximum
T_{air}	$°C$	8.6	24.7	22.8	28.1
RH	%	0.5	0.8	0.8	0.9
V_{wi}	knots	0.4	6.4	5.9	10.3
β_{360}	°	0.0	62.5	102.6	360.0
T_0	s	3.0	5.6	5.5	7.5
H_s	m	0.2	1.4	1.3	2.3
α_{360}	°	0.0	59.2	114.9	360.0
E_e	W/m^2	0.0	27.0	130.1	678.2
α_{sun}	Night/Day	0.0	0.0	0.4	1.0
P_{hvac}	kW	49.7	121.4	125.7	197.4
P_a	kW	265.0	379.0	381.8	506.0

Table E.18: IQR Combined Auxiliary (ΨØ) developed data range limits

Details	Units	Minimum	Median	Mean	Maximum
Combined Auxiliary - IQR ($\Psi \neq 0$)					
V_s	knots	0.0	0.0	4.6	18.0
T_{air}	$°C$	4.4	23.9	20.5	28.1
RH	%	0.5	0.8	0.8	0.9
V_{wi}	knots	0.2	5.9	5.7	13.4
β_{360}	°	0.0	78.7	119.1	360.0
T_0	s	3.0	5.6	5.6	7.6
H_s	m	0.1	1.5	1.4	2.8
α_{360}	°	0.0	77.5	135.5	360.0
E_e	W/m^2	0.0	11.7	97.5	573.2
α_{sun}	Night/Day	0.0	0.0	0.4	1.0
S	Anchor/Sail	0.0	0.0	0.4	1.0
P_{hvac}	kW	51.8	125.6	123.1	212.4
P_a	kW	219.0	375.0	373.2	535.0

Table E.19: non-IQR Combined Total (ΨØ) developed data range limits

Details	Units	Minimum	Median	Mean	Maximum
Combined Total Power - None ($\Psi \neq 0$)					
V_s	knots	0.0	0.0	4.6	18.0
T_{air}	$°C$	4.4	23.9	20.5	28.1
RH	%	0.5	0.8	0.8	0.9
V_{wi}	knots	0.2	5.9	5.7	13.4
β_{360}	°	0.0	78.7	119.1	360.0
H_s	m	0.1	1.5	1.4	2.8
T_0	s	3.0	5.6	5.6	7.6
α_{360}	°	0.0	77.5	135.5	360.0
HCI	Days	218.6	303.9	319.5	488.5
E_e	W/m^2	0.0	11.7	97.5	573.2
α_{sun}	Night/Day	0.0	0.0	0.4	1.0
S	Anchor/Sail	0.0	0.0	0.4	1.0
P_{hvac}	kW	51.8	125.6	123.1	212.4
$P_{s,cw}$	kW	0.0	0.0	859.8	6568.7
$P_{s,t}$	kW	0.0	0.0	1240.4	7699.3
P_{total}	kW	232.0	411.0	1533.9	7718.0

Bibliography

[1] ABS. Guide For: Passenger Comfort on Ships. *American Bureau of Shipping*, 2014(September 2014), 2015.

[2] ABS. Guide For: Crew Habitability on Ships. *American Bureau of Shipping*, 2016.

[3] Charu C. Aggarwal. *Neural Networks and Deep Learning: A Textbook*. Springer International Publishing AG, 1 edition, 2018. ISBN 9783319944623.

[4] Rob Akershoek. Design of a Fully Electric Super Yacht: Research of design to find driveres, enablers, and barriers of electric yachting. *MSc. Thesis*, 2018.

[5] Jasper Alberts and Martin Jacoby. The Effect of Biofouling on the Performance of Feadships. *Msc. Internship Thesis*, 2019.

[6] Saul Amarel and Louis Steinberg. Artificial Intelligence and Marine Design. *Proceedings of the Jerusalem Conference on Information Technology*, 11(1):315–333, 1990. doi: 10.1109/jcit. 1990.128300.

[7] ASHRAE. *ASHRAE Handbook: Fundamentals*. American Society of Heating, Refrigeration, and Air Conditioning Engineers, 2013. ISBN 9781936504466.

[8] Sammie Bae. *JavaScript Data Structures and Algorithms*. Apress Media LLC, 2019. ISBN 9781484239872. doi: 10.1007/978-1-4842-3988-9.

[9] Andro Bakica, Inno Gatin, Vuko Vukčević, Hrvoje Jasak, and Nikola Vladimir. Accurate assessment of ship-propulsion characteristics using CFD. *Ocean Engineering*, 175(February):149–162, 2019. ISSN 00298018. doi: 10.1016/j.oceaneng.2018.12.043.

[10] Michael R. Bakker. A Reference-based Design Approach in Preliminary Ship Design. *Msc. Internship Thesis*, 2021. ISSN 1098-6596.

[11] E. Bal Beşikçi, O. Arslan, O. Turan, and A. I. Ölçer. An artificial neural network based decision support system for energy efficient ship operations. *Computers and Operations Research*, 66: 393–401, 2016. ISSN 03050548. doi: 10.1016/j.cor.2015.04.004.

[12] Francesco Baldi, Fredrik Ahlgren, Tuong Van Nguyen, Marcus Thern, and Karin Andersson. Energy and exergy analysis of a cruise ship. *Energies*, 11(10), 2018. ISSN 19961073. doi: 10.3390/en11102508.

[13] Werner Blendermann. Wind Loads on Moored and Manoeuvring Vessels. *Conference on Offshore Mechanics & Arctic Engineering*, 1(1993):183–190, 1993.

[14] Werner Blendermann. Parameter identification of wind loads on ships. *Journal of Wind Engineering and Industrial Aerodynamics*, 51(3):339–351, 1994. ISSN 01676105. doi: 10.1016/ 0167-6105(94)90067-1.

[15] Clemens Boertz. Energy demand of a fuel cell-driven cruise ship. *MSc. Thesis*, 2020.

[16] Alexei Botchkarev. Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology. *Interdisciplinary Journal of Information, Knowledge, and Management*, 14, 9 2018. doi: 10.28945/4184. URL https://arxiv.org/ abs/1809.03006.

[17] James E. Braun and Nitin Chaturvedi. An inverse gray-box model for transient building load prediction. *HVAC and R Research*, 8(1):73–99, 2002. ISSN 10789669. doi: 10.1080/10789669. 2002.10391290.

[18] Christine Bressy and Marlène Lejars. Marine fouling : An overview marine fouling. *Journal of Ocean Technology*, 9(4):19–28, 2014.

[19] R. Burger. Improving the Predictions of Ship Speed and Fuel Consumption for Heavy Lift Vessels. *MSc. Thesis*, page 89, 2017.

[20] John S. Carlton. *Marine Propellers and Propulsion, 2nd edition*, volume 163. Elsevier Ltd, 2010. ISBN 978-0-75-068150-6.

[21] CE Delft. Study on methods and considerations for the determination of greenhouse gas emission reduction targets for international shipping. *European Commission*, 1(April):87, 2019.

[22] T Cepowski and T Szelangiewicz. Application of Artificial Neural Networks to investigations of ship seakeeping ability. Part I. *Polish Maritime Research*, Vol. 8, nr:11–15, 2001.

[23] Tomasz Cepowski. The prediction of ship added resistance at the preliminary design stage by the use of an artificial neural network. *Ocean Engineering*, 195(October 2019):106657, 2020. ISSN 00298018. doi: 10.1016/j.oceaneng.2019.106657. URL https://doi.org/10.1016/j.oceaneng.2019.106657.

[24] François Chollet. Keras. \url{https://github.com/fchollet/keras}, 2015.

[25] H. B. Clausen, M. Lützen, A. Friis-Hansen, and N. Bjørneboe. Bayesian and neural networks for preliminary ship design. *Marine Technology and SNAME News*, 38(4):268–277, 2001. ISSN 19453582. doi: 10.5957/mt1.2001.38.4.268. URL https://linkinghub.elsevier.com/retrieve/pii/B9780080439501500442.

[26] Andrea Coraddu, Luca Oneto, Francesco Baldi, and Davide Anguita. Vessels fuel consumption: A data analytics perspective to sustainability. *Studies in Fuzziness and Soft Computing*, 358: 11–48, 2018. ISSN 14349922. doi: 10.1007/978-3-319-62359-7{_}2.

[27] Andrea Coraddu, Luca Oneto, Francesco Baldi, Francesca Cipollini, Mehmet Atlar, and Stefano Savio. Data-driven ship digital twin for estimating the speed loss caused by the marine fouling. *Ocean Engineering*, 186(January):106063, 2019. ISSN 00298018. doi: 10.1016/j.oceaneng.2019.05.045. URL https://doi.org/10.1016/j.oceaneng.2019.05.045.

[28] Elizabeth Johanna Cozijnsen. The footprint of yacht production. *MSc. Thesis*, 2019.

[29] Ivan Nunes da Silva, Danilo Hernane Spatti, Rogerio Andrade Flauzino, Luisa Helena Bartocci Liboni Liboni, and Silvas Franco dos Reis Alves. *Artificial neural networks*. Springer Nature, 2011. ISBN 9781617615535. doi: 10.4324/9781315154282-3.

[30] Henk de Vries. Feadship | There are yachts and there are Feadships., 2019. URL https://www.feadship.nl/chapters/building-outside-the-box/henk-de-vries.

[31] Yigit Kemal Demirel, Mahdi Khorasanchi, Osman Turan, Atilla Incecik, and Michael P. Schultz. A CFD model for the frictional resistance prediction of antifouling coatings. *Ocean Engineering*, 89(2014):21–31, 2014. ISSN 00298018. doi: 10.1016/j.oceaneng.2014.07.017.

[32] Yigit Kemal Demirel, Osman Turan, and Atilla Incecik. Predicting the effect of biofouling on ship resistance using CFD. *Applied Ocean Research*, 62(2017):100–118, 2017. ISSN 01411187. doi: 10.1016/j.apor.2016.12.003. URL http://dx.doi.org/10.1016/j.apor.2016.12.003.

[33] G. B. Deng, P. Queutey, and M. Visonneau. RANS prediction of the KVLCC2 tanker in head waves. *Journal of Hydrodynamics*, 22(5 SUPPL. 1):476–481, 2010. ISSN 10016058. doi: 10.1016/S1001-6058(09)60239-0. URL http://dx.doi.org/10.1016/S1001-6058(09)60239-0.

[34] Ke-Lin Du and M. N. S. Swamy. *Neural Networks and Statistical Learning.* Springer-Verlag London Ltd., 2 edition, 2019. ISBN 9781447174516.

[35] Bradley Efron and Robert J. Tibshirani. *An Introduction to the Boostrap*. Springer-Science + Business Media B.V., 1993. ISBN 9780412042317.

[36] European Parliamentary Research Service. The Ethics of Artificial Intelligence : Issues and Initiatives. *Scientific Foresight Unit (STOA)*, 1(March), 2020.

[37] J Harvey Evans. Basic Design Concepts. *Journal of American Society fo Naval Engineers*, 71 (4):671–678, 1957.

[38] Feadship. Concurrent Design at Feadship, 2016. ISSN 1098-6596.

[39] Feadship. De Voogt - Scope integer ontwerp, 2019.

[40] Feadship. The Working Process De Voogt, 2020.

[41] E. Ferrario, N. Pedroni, E. Zio, and F. Lopez-Caballero. Bootstrapped Artificial Neural Networks for the seismic analysis of structural systems. *Structural Safety*, 67(November):70–84, 2017. ISSN 01674730. doi: 10.1016/j.strusafe.2017.03.003.

[42] Hitoshi Fujii and Takeshi Takahashi. Experimental Study on the Resistance Increase of a Large Full Ship in Regular Oblique Waves. *Journal of the Society of Naval Architects of Japan*, 137 (1975):132–137, 1975.

[43] Toshifumi Fujiwara, Michio Ueno, and Tadashi Nimura. Estimation of Wind Forces and Moments Acting on Ships. *Journal of the Society of Naval Architects of Japan*, 1998.

[44] Toshifumi Fujiwara, Michio Ueno, and Yoshiho Ikeda. A New Estimation Method of Wind Forces and Moments acting on Ships on the Basis of Pysical Component Models. *Journal of the Japan Society of Naval Architects*, 2(2005):243–255, 2005.

[45] Salvador García, Sergio Ramírez-Gallego, Julián Luengo, José Manuel Benítez, and Francisco Herrera. Big data preprocessing: methods and prospects. *Big Data Analytics*, 1(1):1–22, 2016. ISSN 2058-6345. doi: 10.1186/s41044-016-0014-0.

[46] George Gougoulidis. The Utilization of Artificial Neural Networks in Marine Applications: An Overview. *Naval Engineers Journal*, 120(3):19–26, 2008. ISSN 00281425. doi: 10.1111/j. 1559-3584.2008.00150.x.

[47] Rob Grin. On the prediction of wave-added resistance with empirical methods. *Journal of Ship Production and Design*, 30(4):1–11, 2014. ISSN 21582874. doi: 10.5957/JSPD.30.4.130060.

[48] Rob Grin. Empirical prediction method of wave added thrust for yachts. *MARIN*, 1(September), 2020.

[49] M. R. Haddara and C. Guedes Soares. Wind loads on marine structures. *Marine Structures*, 12 (3):199–209, 1999. ISSN 09518339. doi: 10.1016/S0951-8339(99)00023-4.

[50] Michael Haranen, Pekka Pakkanen, Risto Kariranta, and Jouni Salo. White, Grey and Black-Box Modelling in Ship Performance Evaluation. *1st Hull Performance & Insight Conference*, 1(April): 115–127, 2016.

[51] J. Holtrop. A Statistical Re-analysis of Resistance and Propulsion Data. *International Shipbuilding Progress*, 363(28):272–276, 1984.

[52] J. Holtrop and G. G. J. Mennen. An Approximate Power Prediction Method. *International Shipbuilding Progress*, 29, 1982.

[53] Zhihui Hu, Yongxin Jin, Qinyou Hu, Sukanta Sen, Tianrui Zhou, and Mohd Tarmizi Osman. Prediction of Fuel Consumption for Enroute Ship Based on Machine Learning. *IEEE Access*, 7: 119497–119505, 2019. ISSN 2169-3536. doi: 10.1109/access.2019.2933630.

[54] Janne Huotari, Antti Ritari, Jari Vepsäläinen, and Kari Tammi. Hybrid ship unit commitment with demand prediction and model predictive control. *Energies*, 13(18), 2020. ISSN 19961073. doi: 10.3390/en13184748.

[55] IMO. Initial IMO Strategy on Reduction of GHG Emissions from Ships, 2018. ISSN 10004882.

[56] IPCC 2007. Climate Change 2007: Impacts, Adaptation, and Vulnerability. *IPCC*, 2007. doi: 10.1016/B978-008044910-4.00250-9.

[57] IPCC 2013. Principles Governing IPCC Work. *IPCC*, 2013.

[58] R. M. Isherwood. Wind Resistance of Merchant Ships., 1973.

[59] ISO 7547. Ships and Marine Technology - Air-conditioning and Ventiliation of Accomodation Spaces - Design Conditions and Basis of Calculations. *ISO International Standard*, 2, 2002.

[60] Jachtbouw Actueel. Profile: De Voogt Naval Architects, 2019. URL `https://www.jachtbouwactueel.nl/profile-de-voogt-naval-architects/`.

[61] W. D. Janssen, B. Blocken, and H. J. van Wijhe. CFD simulations of wind loads on a container ship: Validation and impact of geometrical simplifications. *Journal of Wind Engineering and Industrial Aerodynamics*, 166(March):106–116, 2017. ISSN 01676105. doi: 10.1016/j.jweia.2017.03.015.

[62] V. Jinkine and V. Ferdinande. A Method for Predicting the Added Resistance of Fast Cargo Ships in Head Waves. *International Shipbuilding Progress*, 21(238):149–167, 1974.

[63] Soteris A. Kalogirou and Milorad Bojic. Artificial neural networks for the prediction of the energy consumption of a passive solar building. *Energy*, 25(5):479–491, 2000. ISSN 03605442. doi: 10.1016/S0360-5442(99)00086-9.

[64] S. Karatasou, M. Santamouris, and V. Geros. Modeling and predicting building's energy use with artificial neural networks: Methods and results. *Energy and Buildings*, 38(8):949–958, 2006. ISSN 03787788. doi: 10.1016/j.enbuild.2005.11.005.

[65] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

[66] Arjen Koop, Bruno Rossin, and Guilherme Vaz. Predicting Wind Loads on Typical Offshore Vessels using CFD. *International Conference of Ocean, Offshore, and Arctic Engineering*, 31 (2012):1–12, 2012.

[67] Lars Larrson and Hoyte C. Raven. *The Principles of Naval Architecture Series: Ship Resistance and Flow*. The Society of Naval Architects and Marine Engineers, 2010. ISBN 9780939773756.

[68] Leifur Th Leifsson, Hildur Sævarsdóttir, Sven Th Sigurdsson, and Ari Vésteinsson. Grey-box modeling of an ocean vessel for operational optimization. *Simulation Modelling Practice and Theory*, 16(8):923–932, 2008. ISSN 1569190X. doi: 10.1016/j.simpat.2008.03.006.

[69] M E Letschert. A Process Design Towards the Yacht Environmental Transparency Index. *MSc. Thesis*, 2020.

[70] Haakon Lindstad, Ruud Verbeek, Merle Blok, Stephan van Zyl, Andreas Hübscher, Holger Kramer, Joko Purwanto, Olga Ivanova, and Hettie Boonman. GHG emission reduction potential of EU-related maritime transport and on its impacts. *TNO*, page 130, 2015.

[71] Ruihua Lu, Osman Turan, Evangelos Boulougouris, Charlotte Banks, and Atilla Incecik. A semi-empirical ship operational performance prediction model for voyage optimization towards energy efficient shipping. *Ocean Engineering*, 110(July 2014):18–28, 2015. ISSN 00298018. doi: 10.1016/j.oceaneng.2015.07.042.

[72] MAN Diesel & Turbo. Basic Principles of ship propulsion. *Ship Propulsion*, 2006.

[73] MARIN. Power Prediction by Statistical Methods. *Hydrodynamics in Ship Design*, 2019.

[74] Dimitri N. Mavris, Daniel A. DeLaurentis, Oliver Bandte, and Mark A. Hale. A Stochastic Approach to Multi-disciplinary Aircraft Analysis and Design. *Aerospace Sciences Meeting and Exhibit*, 1998.

[75] Ehsan Mazloumi, Geoff Rose, Graham Currie, and Sara Moridpour. Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in bus travel time prediction. *Engineering Applications of Artificial Intelligence*, 24(3):534–542, 2011. ISSN 09521976. doi: 10.1016/j.engappai.2010.11.004. URL http://dx.doi.org/10.1016/j.engappai.2010.11.004.

[76] Anthony F. Molland. *The Maritime Engineering Reference Book: A Guide to Ship Design, Construction, and Operation*. Elsevier Ltd, 2008.

[77] Naval Sea Systems Command. Electric Power Load Analysis (EPLA) for Surface Ships, 2012.

[78] Oliver Nelles. *Nonlinear System Identification: From Classical Approaches to Neural Networks, Fuzzy Models, and Gaussian Processes*. Springer-Verlag Berlin Heidelberg, 2 edition, 2020. ISBN 9783030474386.

[79] Alberto Hernandez Neto and Flávio Augusto Sanzovo Fiorelli. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. *Energy and Buildings*, 40(12):2169–2176, 2008. ISSN 03787788. doi: 10.1016/j.enbuild.2008.06.013.

[80] K. Niklas and H. Pruszko. Full-scale CFD simulations for the determination of ship resistance as a rational, alternative method to towing tank experiments. *Ocean Engineering*, 190(July):106435, 2019. ISSN 00298018. doi: 10.1016/j.oceaneng.2019.106435. URL https://doi.org/10.1016/j.oceaneng.2019.106435.

[81] Naya Olmer, Bryan Comer, Biswajoy Roy, Xiaoli Mao, and Dan Rutherford. Greenhouse Gas Emissions From Global Shipping, 2013-2015. *The International Council on Clean Transportation*, 2017. ISSN 10638652. URL https://www.theicct.org/sites/default/files/publications/Global-shipping-GHG-emissions-2013-2015_ICCT-Report_17102017_vF.pdf.

[82] Guofei Pang, Liu Yang, and George Em Karniadakis. Neural-net-induced Gaussian process regression for function approximation and PDE solution. *Journal of Computational Physics*, 384 (May):270–288, 2018. doi: 10.13140/RG.2.2.35191.11681.

[83] A. I. Parkes, A. J. Sobey, and D. A. Hudson. Physics-based shaft power prediction for large merchant ships using neural networks. *Ocean Engineering*, 166(February):92–104, 2018. ISSN 00298018. doi: 10.1016/j.oceaneng.2018.07.060.

[84] Benjamin Pjedsted Pedersen and Jan Larsen. Modeling of Ship Propulsion Performance. *World Maritime Technology Conference*, 1(2008):1–10, 2009. URL http://imare.in/media/30638/paper-no-4a-3-mr-bp-pedersen.pdf.

[85] Benjamin Pjedsted Pedersen and Jan Larsen. Prediction of Full-Scale Propulsion Power using Artificial Neural Networks, 2009.

[86] Benjamin Pjedsted Pedersen and Jan Larsen. Gaussian Process Regression for Vessel Performance Monitoring. *COMPIT 13*, 12, 2013.

[87] F Pedregosa, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, M Blondel, P Prettenhofer, R Weiss, V Dubourg, J Vanderplas, A Passos, D Cournapeau, M Brucher, M Perrot, and E Duchesnay. Scikit-learn: Machine Learning in {P}ython. *Journal of Machine Learning Research*, 12: 2825–2830, 2011.

[88] Jóan Petur Petersen, Daniel J. Jacobsen, and Ole Winther. Statistical modelling for ship propulsion efficiency. *Journal of Marine Science and Technology*, 17(1):30–39, 2012. ISSN 09484280. doi: 10.1007/s00773-011-0151-0.

[89] Propulsion Committee of the 28th ITTC. 1978 ITTC Performance Prediction Method. *ITTC - Recommended Procedures and Guidelines*, 2017.

[90] Claudia C.Gutiérrez Rodríguez and Sylvie Servigne. Managing sensor data uncertainty: A data quality approach. *International Journal of Agricultural and Environmental Information Systems*, 4(1):35–54, 2013. ISSN 19473192. doi: 10.4018/jaeis.2013010103.

[91] Jason Runge and Radu Zmeureanu. Forecasting energy use in buildings using artificial neural networks: A review. *Energies*, 12(17), 2019. ISSN 19961073. doi: 10.3390/en12173254.

[92] Hamid Sadat-Hosseini, Ping Chen Wu, Pablo M. Carrica, Ho Kim, Yasuyuki Toda, and Frederick Stern. CFD verification and validation of added resistance and motions of KVLCC2 with fixed and free surge in short and long head waves. *Ocean Engineering*, 59:240–273, 2013. ISSN 00298018. doi: 10.1016/j.oceaneng.2012.12.016.

[93] Hyatt Saleh. *The Deep Learning with PyTorch Workshop*. Pacht Publishing Ltd., 1 edition, 2020. ISBN 9781838989217.

[94] Seakeeping Commitee of the 29th ITTC. Calculation of the weather factor fw for decrease of ship speed in wind and waves. *International Towing Tank Conference*, 2018.

[95] Michael J. Siemann. Performance and Applications of Residential Building Energy Grey-Box Models. *MSc. Thesis*, 2013.

[96] Dana Simian and Laura Florentina Stoica. Communications in Computer and Information Science. *Modelling and Development of Intelligent Systems*, 2019. URL http://link.springer.com/10.1007/978-3-030-39237-6.

[97] Simon Sinek. *Start with why : how great leaders inspire everyone to take action*. London: Portfolio/Penguin, 2009. ISBN 9781591846444 1591846447.

[98] Douwe Stapersma and Hans Klein Houd. *Design of Propulsion and Electric Power Generation Systems*. IMarEST, 2012.

[99] Douwe Stapersma and Hans Klein Houd. *Selected Chapters of "Design of Auxiliary Systems, Shafting, and Flexible Mounting"*. IMarEST, 2016.

[100] R L Townsin. The Ship Hull Fouling Penalty. *Biofouling*, 19(sup1):9–15, 2003. ISSN 0892-7014. doi: 10.1080/0892701031000088535.

[101] Michio Ueno, Fumitoshi Kitamura, Naoto Sogihnara, and Toshifumi Fujiwara. A Simple Method to Estimate Wind Loads on Ships. *Advances in Civil, Environmental, and Materials Research*, 1 (2005):2314–2322, 2012.

[102] United Nations. Decisions Adopted by the Conference of the Parties: Adoption of the Paris Agreement. *Framework Convention on Climate Change*, 2016. ISSN 0378777X.

[103] Dogancan Uzun, Yigit Kemal Demirel, Andrea Coraddu, and Osman Turan. Time-dependent biofouling growth model for predicting the effects of biofouling on ship resistance and powering. *Ocean Engineering*, 191(September):106432, 2019. ISSN 00298018. doi: 10.1016/j.oceaneng.2019.106432. URL https://doi.org/10.1016/j.oceaneng.2019.106432.

[104] Martijn van Wijngaarden, Sven Oosterbaan, and Pieter van Loon. Superyacht Construction Design Exploration. *HISWA Symposium*, 2018.

[105] Haiying Wang and Huiru Zheng. Model Validation, Machine Learning. In Werner Dubitzky, Olaf Wolkenhauer, Kwang-Hyun Cho, and Hiroki Yokota, editors, *Encyclopedia of Systems Biology*, pages 1406–1407. Springer New York, New York, NY, 2013. ISBN 978-1-4419-9863-7. doi: 10.1007/978-1-4419-9863-7{_}233. URL https://doi.org/10.1007/978-1-4419-9863-7_233.

[106] Liqian Yang, Gang Chen, Niels Gorm Malý Rytter, Jinlou Zhao, and Dong Yang. A genetic algorithm-based grey-box model for ship fuel consumption prediction towards sustainable shipping. *Annals of Operations Research*, 2019. ISSN 15729338. doi: 10.1007/s10479-019-03183-5. URL `https://doi.org/10.1007/s10479-019-03183-5`.

[107] Claudio Zeni, Kevin Rossi, Aldo Glielmo, and Francesca Baletto. On Machine Learning Force Fields for Metallic Nanoparticles. *Advances in Physics: X*, 4(1):28, 2019.

[108] R H Zwart. Trim Optimization for Ships in Service. *MSc. Thesis*, 2020.