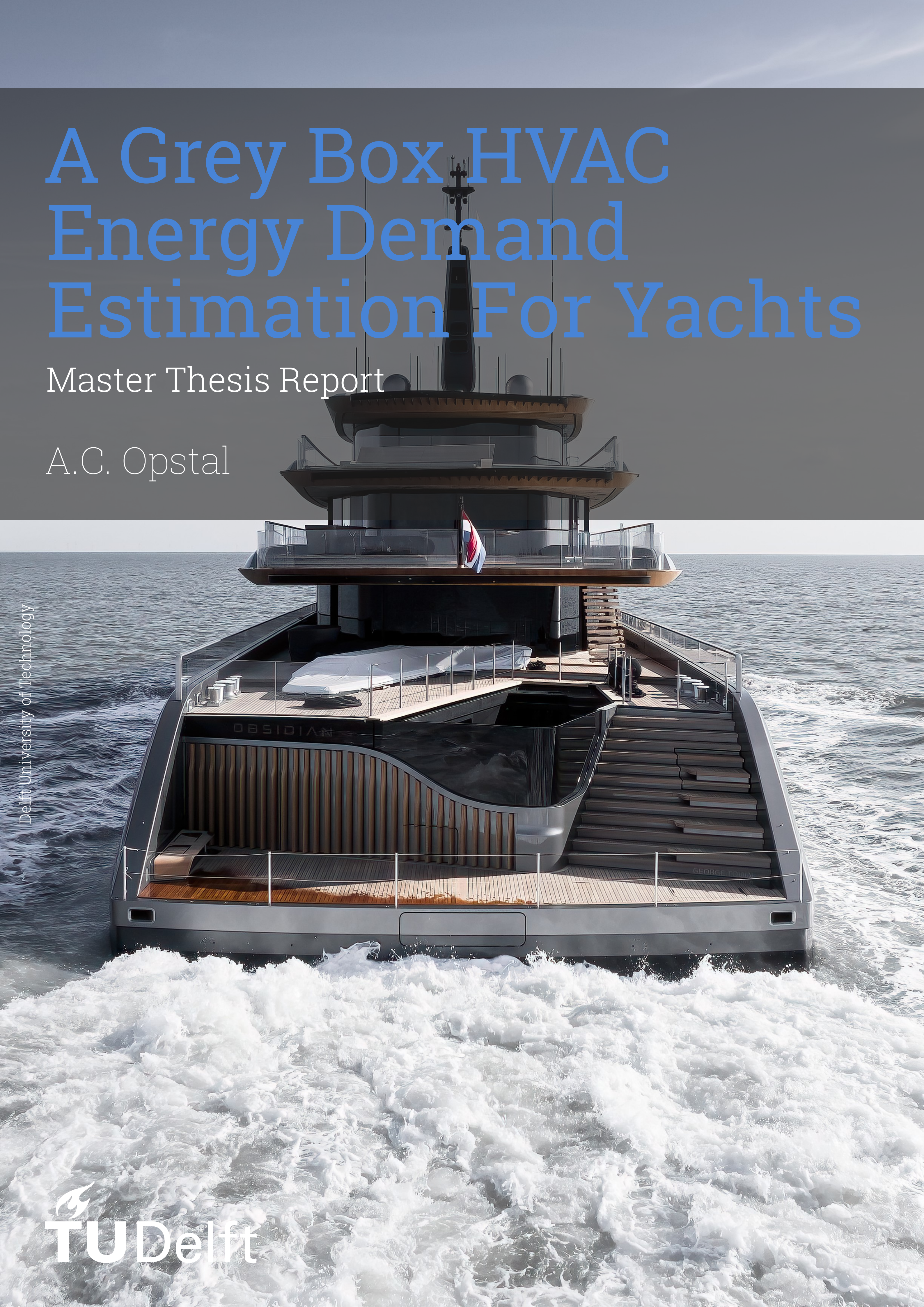
A Grey Box HVAC Energy Demand Estimation For Yachts

Master Thesis Report

A.C. Opstal

Delft University of Technology



A Grey Box HVAC Energy Demand Estimation For Yachts

Master Thesis Report

by

A.C. Opstal

To obtain the degree of Master of Science in Marine Technology
in the specialisation of Ship Design
at the Delft University of Technology,
to be defended publicly on Wednesday May 29th, 2024.

Student number:	4569024	
Report number:	MT.23/24.031.M	
Date:	Wednesday 29th May, 2024	
Supervisors:	Dr. A.A. Kana	TU Delft Supervisor
	Ing. A.A. El Mouhandiz	DVNA Daily Supervisor
	Dr.ir. L. van Biert	Committee

TU Delft

DE VOOGT
NAVAL ARCHITECTS

Preface

Having lived in Delft nearly my whole life, my enthusiasm for technology began with the proximity and influence of TU Delft. Growing up near this hub of innovation, I was naturally drawn to the field of technology. My passion for the maritime world, however, was sparked by my father, who introduced me to the offshore industry and supported my diving ambitions. This blend of technological curiosity and maritime fascination has guided my academic journey, culminating in this thesis.

First and foremost, I would like to extend my gratitude to De Voogt Naval Architects for providing me with the opportunity to work on this research. Special thanks to Abdel-Ali El Mouhandiz, whose assistance in acquiring the necessary data and supervision throughout my thesis has been essential. Your critical insights have helped me push the level of this research. To Bram Jongepier, Aaron Alkemade and Kjeld Broekhaus, thank you for making time to answer my questions and sharing your knowledge.

I am also incredibly thankful for the opportunities presented to me by the Delft University of Technology. Austin Kana, your guidance and kindness have been extremely helpful throughout my master and thesis. Your feedback and supervision have guided me in taking the next steps and elevating the quality of this research.

I would like to thank my friends and family, especially my parents, Ellen & Adri, your love and encouragement have guided me throughout my thesis. And lastly, to my girlfriend, Anna, thank you for being the source of distraction I needed, and my greatest support.

A.C. Opstal
Rotterdam, May 2024

Summary

The yachting industry is embracing sustainable practices in response to global environmental concerns, particularly those outlined in the Paris Agreement, which aim to reduce greenhouse gas emissions and limit global temperature increases. Although diesel engines are most commonly used in yachts and are a major contributor to greenhouse gas emissions, efforts are being made to minimise energy consumption through the use of alternative fuels. Auxiliary systems, such as heating, ventilation and air conditioning (HVAC), are also being studied for efficiency improvements. As yachts spend a considerable amount of time at anchor or in port, these systems account for a significant percentage of the total consumption.

This thesis focuses on estimating the energy demand of HVAC systems onboard yachts. In collaboration with De Voogt Naval Architects, part of Feadship, this research uses a grey box model approach, a combination of a white box and a black box, to estimate and evaluate HVAC systems. White box models are based on first principles or physics and are transparent models. This contrasts with black box models, which model a system based solely on observed data, without any prior knowledge of the system. These models are primarily machine learning algorithms such as artificial neural networks. Black box models perform better than white box models within the range of the trained data, but white box models excel at prediction outside this range. The integration of both models can combine the advantages of both into a grey box model.

The white box model described in this report is a type of predictive model that combines theoretical knowledge with empirical data to estimate the heat load for HVAC systems on yachts. The model uses theoretical principles derived from various methods, such as an ISO heat load balance, and incorporates empirical data collected from sensors installed on yachts, including temperature and humidity measurements. This data is used to simulate the performance of the HVAC system under different weather conditions.

The heat load estimation of the white box model is used as input for the greybox model combined with various sensor data. It makes use of a perceptron artificial neural network (ANN), that can learn from the data and adjust its predictions. The hyperparameters of the ANN are chosen and validated using a kfold cross validation. The final calculations are performed with the optimal configuration.

The results of the optimal model are validated against the power consumption recorded in the voyage data. The grey box model achieves a mean absolute percentage error (MAPE) accuracy of 91.3%, which is an improvement compared to the performance of the solo black box. Specifically, the MAPE accuracy of the grey box model is 0.9% higher along with similar improvements in other metrics.

Finally, the study provides an insight on the possible application of the grey box model. The solo white and black box models can also be used apart from the grey box model. A simple extrapolation analysis is conducted to provide insights on the capabilities of the model. With the implementation of these models, Feadship can predict the heat load and energy demand for varying weather conditions and ranges of data.

Contents

Preface i

Summary ii

Nomenclature viii

1 Introduction 1
- 1.1 Heating, Ventilation & Air Conditioning Systems . . . 3
- 1.2 Feadship & De Voogt Naval Architects . . . 4
- 1.3 Problem statement . . . 5
- 1.4 Research Objective . . . 5
- 1.5 Research Questions . . . 5
- 1.6 Report Outline . . . 6
- 1.7 Scope . . . 6

2 Problem Analysis 7
- 2.1 DVNA general & HVAC design process . . . 7
- 2.2 Significance of Accurate Prediction Methods . . . 9
- 2.3 Current Methods and Accuracy . . . 10
- 2.4 Modelling in the Maritime Industry . . . 12
- 2.5 Available data . . . 13
- 2.6 Problem Conclusion . . . 13
- 2.7 Method Requirements . . . 14

3 Solution Approach 15
- 3.1 White box modelling . . . 15
  - 3.1.1 Model Selection . . . 15
  - 3.1.2 White box limitations . . . 16
- 3.2 Black box modelling . . . 17
  - 3.2.1 Model Selection . . . 17
- 3.3 Data Preparation . . . 19
- 3.4 Grey box modelling . . . 20
- 3.5 Method Requirement Coverage . . . 21

4 Methodology 22
- 4.1 Data Collection . . . 22
- 4.2 Data insight . . . 23
- 4.3 Data Preparation . . . 26
- 4.4 White Box Methodology . . . 29
- 4.5 Grey Box & Black Box . . . 36
  - 4.5.1 ANN Basic Principles . . . 36
  - 4.5.2 Hyperparameters . . . 37
  - 4.5.3 Performance Evaluation . . . 40

5 Results and Validation 42
- 5.1 Data Preparation Results . . . 42
- 5.2 Power calculation from voyage data . . . 43
- 5.3 White Box Results . . . 44
- 5.4 White box Validation . . . 45
- 5.5 Grey Box Results . . . 50
- 5.6 Black Box . . . 52
- 5.7 Comparison of Black and Grey Box . . . 53

6 Model Application **54**
6.1 Application White box . . . 54
6.2 Application Black box . . . 55
6.3 Application Grey box . . . 56
6.3.1 General . . . 56
6.3.2 Extrapolation Capabilities . . . 57

7 Conclusion **59**
7.1 Conclusion Research Questions . . . 59

8 Discussion & Recommendations **61**
8.1 Contributions . . . 61
8.2 Limitations . . . 62
8.3 Recommendations . . . 62

A Absorption and convective heat coefficient **67**

B White Box Model Development **68**
B.0.1 General Arrangement Spreadsheet . . . 68
B.0.2 Python Calculation . . . 69

C Complete White Box Results **70**

D Reheater addition **71**

List of Figures

1.1 System breakdown of electrical components and systems (Boertz, 2020). 2
1.2 90m-plus fleet forecasting scenarios (The Superyacht Group, 2023). 2
1.3 Air Conditioning Unit (ACU) schematic of a HVAC system Odendaal (2021). 3
1.4 Obsidian, a Feadship Yacht delivered by their yard in Aalsmeer in June 2023 (Feadship, 2023). 4
1.5 Report Structure and Outline 6

2.1 Design Phases of DVNA (2022). 7
2.2 Traditional design spiral approach by Harvery (2009) 8
2.3 The relationship of design freedom, knowledge, and cost committed (Mavris and DeLaurentis, 2000). 9
2.4 Total power consumption and HVAC Loadlist, Power (kW) vs. Temperature (°C) (Feadship, 2022) 11
2.5 ISO-7547 calculation example (Feadship, 2022). 11
2.6 Probability of occurrence % of Relative humidity (%) vs Temperature (°C) (Feadship, 2022). 12

3.1 Schematic of a artificial neural network 18
3.2 Possible grey box configurations (Odendaal, 2021), based on data from Leifsson et al. (2008). 20
3.3 Proposed serial grey box configuration 21

4.1 Distribution of Fresh Air Temperature (7SEAS sensor data) 24
4.2 Distribution of Fresh Air Humidity (ECMWF, 2023) 24
4.3 Distribution of Beam Normal Irradiation (ECMWF, 2023) 24
4.4 Distribution of Diffuse Horizontal Irradiation (ECMWF, 2023) 24
4.5 Sea route of selected yacht. 25
4.6 Distribution of Vessel Heading (Logarithmic scale) 25
4.7 Distribution of Coordinates 25
4.8 Supply and Return Temperature and Humidity. 26
4.9 Spearman's Correlation Heatmap 28
4.10 Part of AC system layout (Feadship, 2022) 30
4.11 General Heat gain schematic (Odendaal, 2021) 30
4.12 Solar angle on a bow facing surface. 31
4.13 Solar angles with respect to a tilted surface. (Sarbu and Sebarchievici, 2017) 32
4.14 Solar irradiance on different orientated surfaces. 32
4.15 Solar radiation on different oriented surfaces of the yacht. 33
4.16 Mollier diagram of humid air (Engineering Toolbox, 2003) 35
4.17 Multilayer perceptron network (Nasruddin et al., 2019). 36
4.18 Model error comparison (left) and corresponding fitting relations (right), from Odendaal (2021), based on Silva et al. (2017) 38
4.19 Dropout technique according to Chollet (2017) 39
4.20 Sigmoid and ReLu Activation Functions and their derivative 39
4.21 Model selection using holdout validation, based on Géron (2023) 40
4.22 A k-fold cross-validation with k=5 (scikit-learn developers, 2024) 41

5.1 HVAC system schematic of the yacht from which the sensor data is obtained. 44
5.2 Example room from the General Arrangement spreadsheet 45
5.3 Chiller power vs. COP 46

5.4 Heat Load comparison, 13-09 to 01-10 47
5.5 Heat Load comparison, 13-06 to 24-06 47
5.6 Unshifted data 47
5.7 Data shifted for better fit. 48
5.8 Heat load prediction comparison (Original data) 49
5.9 Error distribution of MAPE (Original data) 49
5.10 Heat load prediction comparison (Shifted data) 49
5.11 Error distribution of MAPE (Shifted data) 49
5.12 Heat load prediction comparison (Shifted-WMA data) 49
5.13 Error distribution of MAPE (Shifted-WMA data) 49
5.14 Predictions made by the white box model displayed on time series basis. 50
5.15 Actual vs. Predicted Power Values 51
5.16 Error distribution of MAPE 51
5.17 Predictions made by the grey box model displayed on time series basis. 51
5.18 Solo black box model. 52
5.19 Actual vs. Predicted Power Values (Black Box) 53
5.20 Error distribution of MAPE (Black Box) 53

6.1 Final Grey Box Model 54
6.2 Temperature ($^{\circ}C$) vs. Power (kW) (Shifted data) 55
6.3 Grey box predictions selected time frame 56
6.4 Heat Load extrapolation white box model 57
6.5 Power extrapolation grey and black box model 58

A.1 Reflective coefficients for different colours (van Wijngaarden and El Mouhandiz, 2023) . 67

B.1 Example room from the General Arrangement spreadsheet 68
B.2 Boundary types with corresponding k and G values in the look up table. 69
B.3 Temperatures of different areas in the yacht at a certain data point, collected in a look up table 69

C.1 Complete White box model results with the fresh air temperature and humidity 70

D.1 Schematic of the HVAC system of the yacht. 71
D.2 Stacked Area Graph Power Demand. 72
D.3 Mean Percentage of Total Power Consumption per consumer 72
D.4 Fresh Air Temperature vs. Total Power Reheaters. 73
D.5 Average Hourly Power Consumption Over the Week. 73
D.6 Boxplot of Power Consumption by Each Reheater. 74

List of Tables

2.1 Simplified format of load balance sheet 10
2.2 Available dataset types, locations and descriptions. 13
2.3 Problem Definition. 14
2.4 Method requirement breakdown. 14

3.1 Current methods and relevant research. 16
3.2 Research on BBM in the maritime industry and on HVAC energy modelling for buildings. 18
3.3 Method Requirement Coverage 21

4.1 Required voyage data for calculation 23
4.2 Required voyage data for validation 23
4.3 Neural network architectures in residential HVAC and marine engineering applications . 38
4.4 Selected Hyperparameters 40

5.1 Data preparation results 43
5.2 Efficiency and COP of HVAC system components. 46
5.3 Comparison of White Box Model Performance Metrics 48
5.4 Hyperparameters for grid search 50
5.5 Final Configuration of the selected Grey Box model 50
5.6 Final Configuration of the Black Box model 52
5.7 Comparison of final configurations 53

6.1 Range and mean of the Black box trained data 55

Nomenclature

Abbreviations

Abbreviation	Definition
AIS	Automatic Identification Systems
ANN	Artificial Neural Network
AC	Air Conditioning
ACU	Air Conditioning Unit
ASHRAE	American Society of Heating, Refrigeration, and Air Conditioning Engineers
BNI	Beam Normal Irradiation
DHI	Direct Horizontal Irradiation
DVNA	De Voogt Naval Architects
ECMWF	European Centre for Medium-Range Weather Forecasts
FA	Fresh Air
GA	General Arrangement
GP	Gaussian Process
GHG	Green House Gasses
HVAC	Heating, Ventilation, and Air Conditioning
IMO	International Maritime Organisation
ISO	International Organization for Standardization
MLP	Multilayer Perceptron network
YETI	Yacht Environmental Transparency Index

Symbols

Symbol	Definition	Unit
ACH	Air Changes per Hour	[1/h]
FA	Proportion of fresh air in the supplied air	[-]
FAH	Fresh Air Humidity	[%]
FAT	Fresh Air Temperature	[$^\circ C$]
I	Current	[A]
I	Solar Radiation	[W/m²]
MAPE	Mean Absolute Percentage Error	[-]
N_{ACH}	Number of air changes per hour required	[1/h]
P	Power	[W]
$Q_{auxiliary}$	Heat gain from auxiliary systems (e.g., lighting)	[W]
Q_{person}	Heat gain from occupants	[W]
$Q_{radiation}$	Heat gain from solar radiation	[W]
$Q_{sensible}$	Total sensible heat	[W]
$Q_{transmission}$	Heat gain due to transmission through surfaces	[W]
R	Reflective coefficient	[-]
RAH	Return Air Humidity	[%]
RH	Relative Humidity	[%]
SAH	Supply Air Humidity	[%]
T	Temperature	[$^\circ C$]
U	Voltage	[V]
V	Volume	[m³]
v	Wind speed	[m/s]
x	Absolute humidity	[kg/kg]
α	Absorption coefficient	[-]
γ	Surface-solar azimuth angle	[°]
h	Enthalpy	[J/kg]
h_c	Convective heat transfer coefficient	[W/(m²·K)]
$\dot{V}$	Flow rate	[m³/h]
ρ_{air}	Air density	[kg/m³]
Σ	Solar elevation angle	[°]
θ	Angle of incidence of solar radiation on the surface	[°]
ΔT	Temperature difference	[K]
Δh	Enthalpy difference	[J/kg]
γ	Surface-solar azimuth angle	[°]
ρ	Density	[kg/m³]
Σ	Solar elevation angle	[°]
θ	Angle of incidence of solar radiation on the surface	[°]
φ	Relative Humidity	[%]

1 Introduction

In today's era of environmental consciousness and the critical need for sustainable practices, the yachting industry accepts their responsibility in embracing greener technologies and reducing energy consumption. As luxury vessels navigate the world's oceans, it becomes increasingly crucial for the industry to lead the way in preserving marine ecosystems and creating a path towards a more sustainable future. By investing in alternative fuels, minimizing energy usage and reducing emissions, the yachting industry can lay the foundation for a sustainable future.

The Paris Agreement (UNFCCC, 2016), established in 2015, is a global effort to address the challenges of climate change. The agreement sets long-term goals to guide all nations to substantially reduce global greenhouse gas (GHG) emissions to limit the global temperature increase in this century to 2 degrees Celsius while pursuing efforts to limit the increase even further to 1.5 degrees. Achieving these goals requires decisive action from various sectors, including the maritime industry, which plays a significant role in global emissions. According to IMO (2020), the GHG emissions including carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O), expressed in CO_{2e} – of total shipping (international, domestic and fishing) have experienced an 9.6 % increase from 977 million tonnes in 2012 to 1,076 million tonnes in 2018. The proportion of shipping emissions in the total global anthropogenic emissions also increased, climbing from 2.76% in 2012 to 2.89% in 2018 (IMO, 2020).

In response to the Paris Agreement's objectives, the Marine Environment Protection Committee of the International Maritime Organisation (MEPC) defined a strategy to combat climate change and reduce greenhouse gasses in their fourth GHG study (IMO, 2020). The IMO aims to achieve the long-term objectives of the Paris Agreement by striving to decrease the average CO_2 emissions per transport work in international shipping. The following contains the levels of ambitions for international shipping (MEPC, 2023), which are valuable guidelines for the yachting industry in the future.

1. The reduction of carbon intensity of vessels through implementation of further phases of the energy efficiency design index (EEDI) for new ships to review with the aim to strengthen the energy efficiency design requirements.
2. The reduction of carbon intensity of international shipping CO_2 emissions per transport work, as an average across international shipping, by at least 40% by 2030, pursuing efforts towards 70% by 2050, compared to 2008.
3. The uptake of zero or near-zero GHG emission technologies, fuels and/or energy sources to represent at least 5%, striving for 10%, of the energy used by international shipping by 2030.
4. GHG emissions from international shipping to reach net-zero by or around, i.e. close to, 2050, taking into account different national circumstances, whilst pursuing efforts towards phasing them out consistent with the long-term temperature goal set out in Article 2 of the Paris Agreement.

Despite the innovation of alternative fuels in the maritime industry, a diesel engine is still the most common engine used in yachts. The emission of greenhouse gasses by these engines are directly related to the energy consumption of these vessels. In order to minimise the energy consumption, it is important to calculate and categorise the different systems and components onboard of a ship. Boertz (2020) managed to distinguish these systems and their components for a cruise ship, shown in Figure 1.1. A yacht and a cruise ship are both designed for luxury leisure and transit, but they have distinct operational profiles and purposes. While a yacht is typically a private vessel intended for personal or small group use, focusing on luxury and comfort, a cruise ship is designed to accommodate a large number of passengers with an organized travel schedule. Both share the same electrical power consumers but a cruise ship is equipped to handle a larger scale of operations.

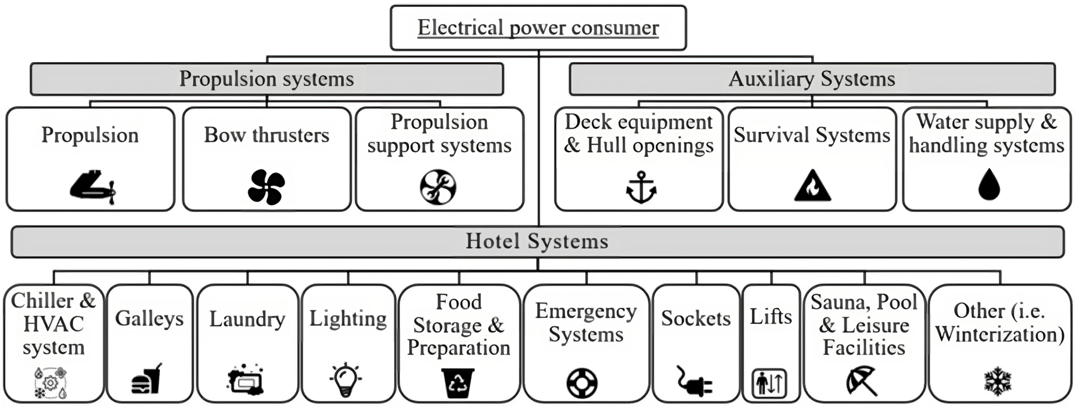


Figure 1.1: System breakdown of electrical components and systems (Boertz, 2020).

According to (Feadship, 2022), the HVAC system plays a significant role in the energy consumption of a yacht. Several factors contribute to the heightened energy demand of the HVAC system. According to its operational profile, analysed by Roy et al. (2011) a yacht is 75% of the time anchored or in port. During this time, the yacht is solely using its auxiliary and hotel systems, without using energy for the propulsion systems. The HVAC system contributes up to 60% of the total hotel load (Feadship, 2022) and is mainly due to its operational profile the largest energy consumer after the propulsion system with a significant impact on the vessel's overall energy consumption. The energy demand of the HVAC systems is heavily effected by the external environmental conditions. In both low and high temperatures, the HVAC system must adapt accordingly. A comfortable environment for the passengers necessitate the systems' continuous operation, leading to increased energy usage.

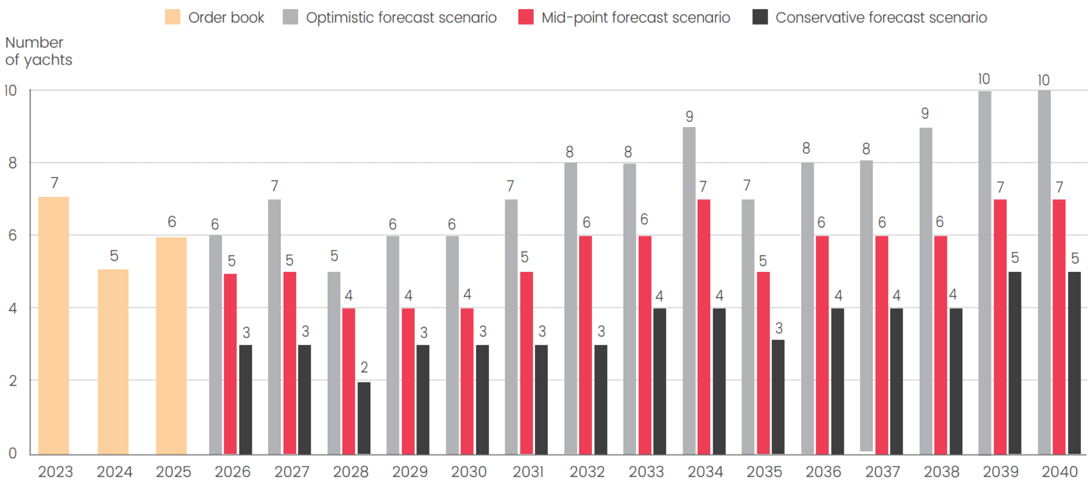


Figure 1.2: 90m-plus fleet forecasting scenarios (The Superyacht Group, 2023).

The increased energy usage is compounded by the fact that the size of worldwide yacht fleet is growing every year. Projections made by The Superyacht Group (2023) and shown in figure 1.2 indicate the forecasting scenarios of yachts with a minimum length of 90 meters. With no immediate interventions, the environmental impact of the yachting industry is expected to grow with the expanding fleet size. Recently, the Yacht Environmental Transparency Index (YETI) was developed (SuperyachtNews, 2023). YETI is a tool specifically designed for indicating the environmental impact of yachts. It is an industry collaboration between the most renowned shipyards, naval architects, and research institutes joining forces to accelerate sustainable change in a transparent and reliable manner. Moving towards more sustainable yachting, vessel optimisation is essential. Researching the energy demand of the HVAC system can give important new insights into this problem.

1.1. Heating, Ventilation & Air Conditioning Systems

Klein Woud and Stapersma (2016) provides guidance in the design of Heating, Ventilation & Air Conditioning (HVAC) systems in ships. HVAC systems are designed to create and maintain optimal environmental conditions for the well-being of humans, equipment, machines, and cargo. These conditions include air temperature, air humidity, the level of carbon dioxide (CO_2), and air purity in terms of dust, bacteria, and unwanted smells. Ventilation and air conditioning play key roles in achieving these conditions. Ventilation systems supply air to specific areas, making sure the air is circulated or exhausted if needed (ASHRAE, 2013). Ventilation can be achieved through mechanical ventilation with fans or natural ventilation. Heat can be added to the airflow through heat exchangers in the ventilation channels or radiators within the room. However, ventilation systems do not typically include air cooling or humidity control features. Air conditioning systems however, use various processes such as filtering, heating, cooling, humidifying, and dehumidifying to meet ASHRAE (2013) Standards for thermal comfort. Air conditioning systems are found throughout the whole ship in accommodations, operational spaces, rooms with sensitive equipment like electronics, and cargo areas with perishable goods.

According to Klein Woud and Stapersma (2016), there are different types of layouts for an air conditioning system. A system with a central Air Conditioning Unit (ACU) is a standard installation. This is when there is only one ACU, which is connected to all areas by ducting. This method lacks the possibility to control the temperature per area. A more effective system is a hybrid air conditioning system, where a central and zonal ACU's are used. It is possible to equip an area with local heaters or coolers if needed.

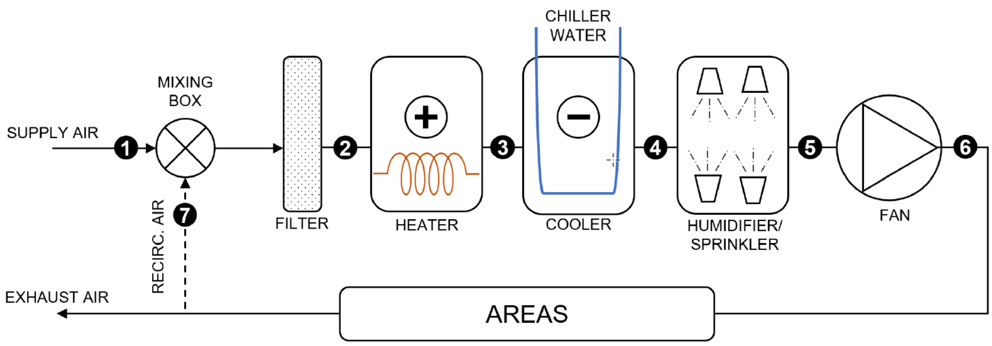


Figure 1.3: Air Conditioning Unit (ACU) schematic of a HVAC system Odendaal (2021).

An ACU consists of several components, as illustrated in figure 1.3. The system receives a combination of fresh ambient air and recirculated air. These air streams are mixed and filtered in the first section of the ACU. The use of recirculated air is often preferred as it reduces energy consumption. However, a drawback is that harmful substances like smoke and bacteria can also be recirculated, posing potential health risks. In some cases, only fresh air without re-circulation is implemented, for example in the kitchen and the dining room in order to get rid of the smell of food. The fresh air supply is necessary to control the CO_2-levels and maintain air purity within the spaces. The exhaust air effectively removes CO_2 and impurities.

Cooling in a HVAC system is achieved through a refrigeration machine, which supplies chilled water, chilled brine, or liquid refrigerant to the system. These refrigeration units use a vapor-compression or absorption cycle to remove heat from a liquid, typically water, which is then circulated to cool air in the ACU or equipment. Chillers can be air-cooled or water-cooled, with the latter being more common in maritime applications due to their efficiency. In ships, chillers help maintain comfortable conditions for occupants, protect sensitive equipment, and preserve cargo integrity by ensuring optimal temperature and humidity levels, despite the challenging marine environment. Heating within the system can be provided by hot water, steam, or electricity, using a heat exchanger. A hot water boiler, low-pressure steam boiler, or electrical generator acts as the heat producer, depending on the heat source chosen. The cooling of air with condensation of water vapour is a method used for dehumidifying the air. Humidification can be done by either a sprinkler system with water or steam humidification.

Marine HVAC systems are complex systems and play a critical role in maintaining a suitable environment within the unique and challenging conditions of a ship. To determine the most effective HVAC system for a particular ship, it is essential to carry out a detailed assessment of the ship's energy requirements. This involves a careful analysis of the existing HVAC systems. By identifying and focusing on the main components contributing to the energy demand, the efficiency and performance of the HVAC system can be significantly improved, ensuring that it meets the specific requirements of the ship's environment.

1.2. Feadship & De Voogt Naval Architects

De Voogt Naval Architects is a yacht design company based in the Netherlands. The company was founded by Henri Willem de Voogt, who built a shipyard on the banks of Haarlem's Spaarne River in 1912. Today, the company is part of Feadship. Feadship is a cooperative venture between Koninklijke De Vries and Royal van Lent. De Vries has their main yards located in Aalsmeer and Makkum, van Lent has their main yards located in Amsterdam and de Kaag.

De Voogt Naval Architects is the main naval architect and marine engineer on all projects for the Feadship group. The Knowledge & Innovation department for the Feadship group is also operated from De Voogt. Feadship collects operational data from their yachts to improve processes. This data can help them design yachts that are perfectly suited to their intended use. By using this valuable data, Feadship continues to innovate in naval engineering, creating more sustainable and efficient vessels.

The latest addition to Feadship's fleet is the 84.2-meter vessel named Obsidian (Feadship, 2023). This yacht marks the first of Feadships' line of large yachts, focusing on minimising carbon footprint. This is achieved through hulls optimised for cruising efficiency instead of maximum speed, weight control, electric propulsion, and the capacity to utilise engines powered by non-fossil diesel fuel known as HVO. Notably, during the yacht's trial runs, the onboard generators operated using this second-generation biofuel, resulting in a 90% reduction in carbon emissions compared to conventional fossil fuel-operated yachts.

Figure 1.4: Obsidian, a Feadship Yacht delivered by their yard in Aalsmeer in June 2023 (Feadship, 2023).

1.3. Problem statement

This thesis report identifies an issue in the design process of yachts: the currently employed methods for estimating the energy demand of HVAC systems are not sufficiently accurate. The primary issue of this is that the employed methods are mainly used to estimate the maximum load of the HVAC system under certain temperature conditions, which significantly differ from the actual operating conditions. This lack of precision in energy demand estimation can lead to inefficiencies and design flaws in yacht HVAC systems, impacting both their performance and sustainability. Addressing this gap is crucial for ensuring that new yachts are equipped with HVAC systems that not only adhere to energy efficiency standards but also fulfil operational and environmental performance criteria effectively.

1.4. Research Objective

The goal of this thesis is to obtain an understanding of the energy demand of the HVAC system onboard of yachts that are already in use, in order to contribute to improvements in the design of new yachts. This is done by creating a data-driven model that can estimate the power consumption of different yachts with the help of obtained data and machine learning. This model can be used to optimise the design process of these systems for future builds. This allows De Voogt to successfully apply this to a new build yacht in order to realise a calculated fitted HVAC system, aiming to reduce the energy demand.

1.5. Research Questions

In order to obtain the main research objective, the following research question is proposed:

To what extent can data from previously built DVNA yachts accurately predict the energy consumption of Heating, Ventilation and Air Conditioning (HVAC) systems in order to improve the design of new future yachts?

To answer this main research question, multiple key sub-questions are proposed. Combining these sub-questions will answer the main research question. The sub-questions answered in this thesis report are:

1. What is the state of the art in predicting the energy consumption of HVAC systems?
   (Chapter 2, section 2.1-2.3)
2. What are the method requirements to model the energy consumption of HVAC systems?
   (Chapter 2, section 2.7)
3. What methods are suitable and in what ways can machine learning and grey box modelling contribute to solving the problem?
   (Chapter 3)
4. How can the integration of the grey box approach optimise the prediction and understanding of HVAC system energy consumption?
   (Chapter 4)
5. How can the accuracy of the model be validated?
   (Chapter 5)
6. How can the proposed model be implemented in the design of new future yachts, and what criteria must be met to consider it to be successful?
   (Chapter 6)

1.6. Report Outline

This report outlines a structured approach to enhance the estimation of HVAC energy demand in ship design, unfolding across four key segments. The opening part introduces the subject and discusses the problem, setting the context for the challenges in ship design. In the second section, the literature investigation and methodology are presented, presenting the techniques used in the estimation calculation. The third section is dedicated to the results of the model and its validation, assessing the model's results to ensure model accuracy. The final section ends with the conclusions and engages in a discussion about the implications of the model. It acknowledges the limitations and details recommendations for future development. Figure 1.5 provides an overview of the report's structure.

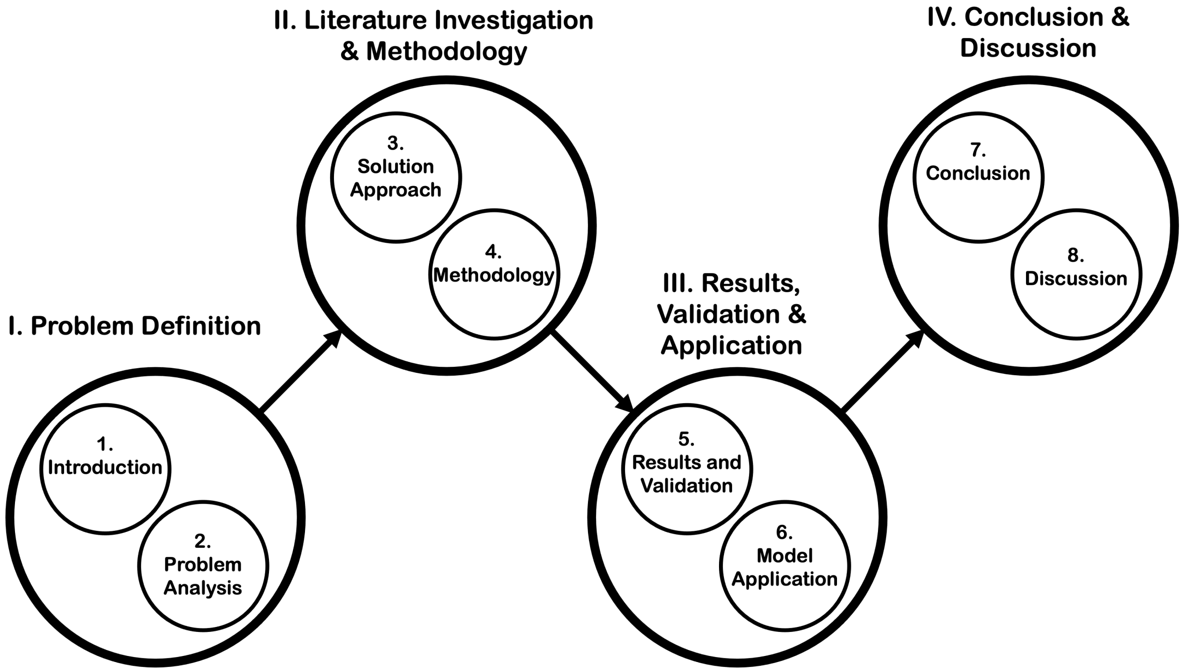


Figure 1.5: Report Structure and Outline

1.7. Scope

Within the scope of the project lies the evaluation of the HVAC systems and the estimation of their energy demand onboard yachts. It involves assessing the design, capacity, and energy consumption patterns of the existing HVAC systems and an identification of the problem. The project includes the creation and accuracy validation of a model that can predict the energy demand, in what way Machine Learning can advance the model and to what extend the model can influence the current design process. On the contrary, the project does not include detailed HVAC system design, procurement of equipment, execution of implementation activities or yacht structural modifications. These aspects fall outside the defined scope of the project.

2 Problem Analysis

This chapter will give an answer to the first research sub-question: *What is the state of the art in predicting the energy consumption of HVAC systems?* and the second research sub-question: *What are the method requirements to model the energy consumption of HVAC systems?*. It will provide a deeper analysis of the challenges with estimating the energy demand of HVAC systems onboard yachts. First in section 2.1, HVAC and its place within ship & yacht design will be discussed. Next, current methods used to predict the energy consumption of HVAC system used by DVNA will be presented in section 2.3. Section 2.3 will give an insight of the accuracy of these methods and their shortcoming. Section 2.2 will expand on the significance of accurate prediction methods. Section 2.5 elaborates on the available data provided by DVNA. After this, the problem definition will be concluded in section 2.6. With data available and problem definition known, in section 2.7, the method requirements will be set up for this research.

2.1. DVNA general & HVAC design process

De Voogt Naval Architects (DVNA) conducts its daily operations following the DVNA project phases, as depicted in Figure 2.1. This representation showcases a simplified version and indicates 'ideal' phase alignment and transition. All acquired information about the DVNA design processes are obtained via internal company documentation of DVNA (2022, 2023).

Figure 2.1: Design Phases of DVNA (2022).

DVNA's work starts at the Contract Design phase with a small project team. This phase receives its input from Concept Design, which is managed by the team at Studio De Voogt that starts after the mission requirements are determined in consultation with the new yacht owner. The transition into DVNA's involvement varies depending on the progress of the Sales phase of the Yard. The contract signing for the Build Number between the client and the Feadship Yard can take place before, during, or after the Contract Design phase. The decision regarding the timing of this contract signing is entirely at the discretion of the Yard and its Sales team. A more comprehensive understanding of these phases, their subsequent processes and what energy demand calculations are done in each phase is explained in the following sections.

Concept Design In this phase Studio De Voogt works in consultation with the Sales department, on the portfolio of Design Prospects for Feadship. There can be dozens of design prospects that can lead to a single build number. Studio De Voogt provides the Technical consultancy towards Sales on these Prospects, as well as offering Design services towards the Clients. Once the Sales department of the yard wants to move to the next phase of the design, they will approach DVNA to assemble a team dedicated to the Technical / Contract Design. Studio De Voogt will then prepare a Design Review and package of deliverables. This typically includes: General Arrangement, Design Data & Risk List; Design Review, Lines Plan, Preliminary Weight Budget & Height Stacking. The auxiliary energy demand is empirically estimated in this phase. An expected initial load balance is constructed by comparison with other yachts. These estimations have a low fidelity and are certain to change. It is only used as a basis for the design.

Technical / Contract Design The main goal of the Technical / Contract Design phase is to establish a technically feasible design for the yacht. This ensures that the yacht can be marketed with a clear understanding of the design's potential risks, and significant risks are appropriately addressed and mitigated. DVNA's role is to recommend on the design's feasibility and identify potential risks. Ultimately, the decision to sell the yacht rests solely with the Yard, regardless of the risk profile. It is possible for the yacht to be sold either before or during this phase. The technical general arrangement is the key deliverable, together with the naval architecture & structural reports. The HVAC power estimations are performed by an external HVAC contractor at this stage. The initial load balances are expanded and equipment lists are developed. Activity rates for the equipment are set to create a broader estimation. These load factors cause uncertainty in the estimations, since these are based on prior knowledge and reference yachts and will still differ from the actual load.

Basic Design The purpose of the Basic Design phase is to engineer the design until it reaches a Class approved level. It is also aimed at ensuring that there are no conflicts between various disciplines involved in the project, thereby achieving an "Integrated Design". All auxiliary power estimations and load lists are conducted by contractors and associated yards during this phase. The HVAC system power estimations are still performed by a HVAC contractor. The power estimations are done using first principles heat load balances on the supplied general arrangements. They are analyzed across different operational temperature levels to establish distinct load cases. Once HVAC assessments are finished, they are incorporated into the electrical load balance sheets. These projections are consistently refined with the emergence of more precise data regarding supplier equipment, area arrangements, and specifications of machinery.

Design Monitoring During this phase, a reduced team will be responsible for closely overseeing the detailed engineering and refinement of the design, from the point of initial development until the yacht's delivery. The primary objective is to ensure that the design integrity established in the Basic Design phase is maintained throughout the process. The naval architecture work will continue, encompassing tasks such as generating weight control reports and conducting a final stability assessment. The HVAC load balances are finished and only minor changes happen in this stage. The estimations are considered to have a high fidelity at this stage since all the equipment is incorporated.

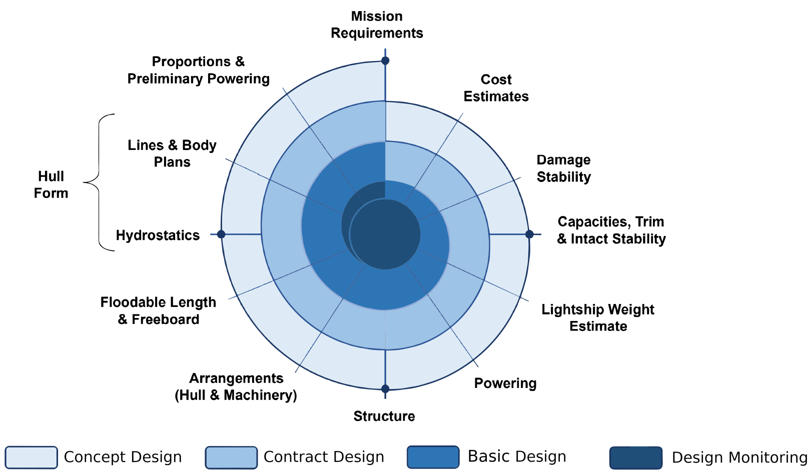


Figure 2.2: Traditional design spiral approach by Harvery (2009)

DVNA uses a design spiral approach during their design phases based on the design spiral by Harvery (2009), shown in figure 2.2. The purpose of the design spiral is to assist in organising the thought process, enabling ship design problems to be solved most efficiently. Harvery (2009) describes how it focuses on how to estimate and balance ship design parameters in a time efficient way using an iterative process. Educated guesses are made to select a HVAC system. When new information becomes available later and the HVAC system changes, the iterative process is used to modify the initial arrangements, balance estimations and power estimations.

Another concept that DVNA uses during the design of HVAC systems and their vessels in general is Concurrent design. Concurrent design is a collaborative design approach that involves the simultaneous cooperation of multiple disciplines, teams, or experts in the design process. The goal of concurrent design is to enhance communication, foster creativity, and streamline the design process by breaking down traditional sequential barriers. Different teams, such as naval architects, structural engineers, systems engineers, interior designers, and other specialists, work together concurrently on their respective aspects of the yacht's design. Due to this method changes in the design of the HVAC system are picked up by multiple disciplines. This approach allows them to exchange ideas, address potential conflicts, and make real-time adjustments in the design and power estimations, leading to a more integrated and cohesive overall design.

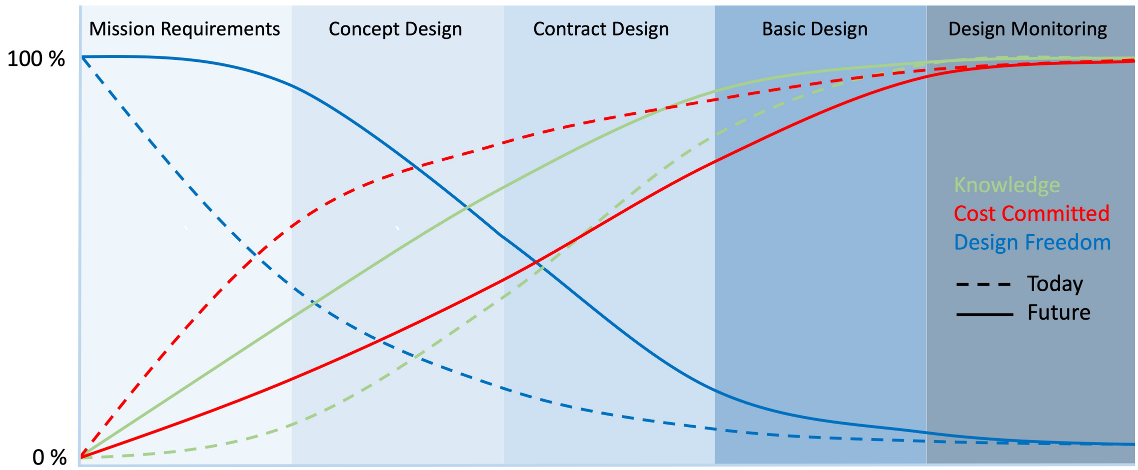


Figure 2.3: The relationship of design freedom, knowledge, and cost committed (Mavris and DeLaurentis, 2000).

Figure 2.3 shows the relationship between design freedom, knowledge and committed cost during the different design phases. During the initial phases of the design process, a considerable part of the design remains uncertain, although a significant part of the costs committed becomes fixed early on in the process. Ship performance evaluation typically starts during the initial phases of ship design. Achieving an accurate estimation of the performance of the vessel is a very challenging process since the ship's design must adhere to safety requirements, as well as cost-effectiveness in construction and operation (Logan, 2011). This causes models and data from previous ship designs to be exceptionally valuable. Without this data, physical laws or a deterministic modelling approach is the only way to evaluate the ship's performance. Constructing models based on vessel data and profiles substantially enhances future yacht designs.

2.2. Significance of Accurate Prediction Methods

The need for precise prediction methods in designing HVAC systems for yachts is crucial. As Feadship (2022) notes, these systems are second only to propulsion systems in terms of energy consumption on yachts, so the consequences of overestimating their requirements are significant. A HVAC system should be able to perform in extreme conditions, but an overestimation of the requirements of the system can lead to inefficiencies. A difference between the design condition and the actual condition a HVAC system operates in can have several consequences:

- **Inefficient operation:** This is the most crucial consequence of an oversized HVAC system. A HVAC system can for example have two chillers. The second chiller might cycle on and off more frequently depending on the demand, leading to inefficient energy use and reduced overall energy efficiency. The excess energy demand not only leads to higher operational costs but also contributes to an increased environmental impact.
- **Wasted space and weight:** Oversized systems will occupy more space than necessary. This can result in loss of valuable space onboard of the yacht and unnecessary additional weight. Heinen & Hopman (2022) produces and installs HVAC systems onboard ships. According to their systems, a 30% decrease in the maximum air quantity a standard AC unit can handle reduces the volume of the AC unit system up to 31%, which is a significant decrease.
- **Higher production & maintenance costs:** Bigger systems cost more to produce and to maintain. If the system is not working on its optimal point, it will put additional strain on its components and may require more frequent maintenance. Which may result in a reduced lifespan of the system.

It is due to these differences and consequences, valuable to gain a better understanding of what influences the energy demand of HVAC systems. With accurate predictions, engineers can design HVAC systems to the specific needs of a yacht, avoiding oversizing and ensuring efficient, cost-effective, and environmentally responsible operation.

2.3. Current Methods and Accuracy

As explained in 2.1, the energy demand estimation of the HVAC system during the design phases are done by an HVAC contractor and are determined empirically. All components onboard are listed comprehensively in load balance sheets. The average load of each component is manually determined by assigning activity percentages to each component under various operational conditions. Determining the activity rates is highly subjective and lacks in accuracy. This approach is implemented because the estimations are generally conservative compared to the operational targets. These methods remain popular due to the lack of explicit knowledge about the practical requirements of the systems. An example of the format of a load balance sheet is shown in table 2.1.

		Temperature	0 °C		10 °C		20 °C	
Component	Consumer	Installed Power (kW)	%	P (kW)	%	P (kW)	%	P (kW)
AC-1	Preheater							
	...							
	...							
Total				Σ		Σ		Σ

Table 2.1: Simplified format of load balance sheet

As part of the operational data that is collected from the yachts, DVNA collects data from their hotel system with their *Project HOTEL Initiative* (Feadship, 2022). Auxiliary loads are recorded including AC power (voltage, amperage and fan speed), air and sea temperature, and exterior relative humidity. This data also shows the energy demand of the HVAC systems at different temperatures and humidity, as depicted in figure 2.4. The blue points are data points retrieved at operational conditions from a single Feadship currently sailing. The red line is the loadlist of the HVAC system onboard estimated by the HVAC contractor at various operational conditions. This graph shows that the installed HVAC systems have a lower energy consumption than the estimated demand calculated by the load lists (Feadship, 2022). The loadlist line is at nearly every data point higher than the operational data. It can be concluded that the loadlist overestimates the power consumption of the HVAC system. The load factors used as activity percentages for the components onboard are standard values and estimated empirically. These values are highly sensitive for over estimations and it is expected that this is part of the problem in overestimating the energy demand.

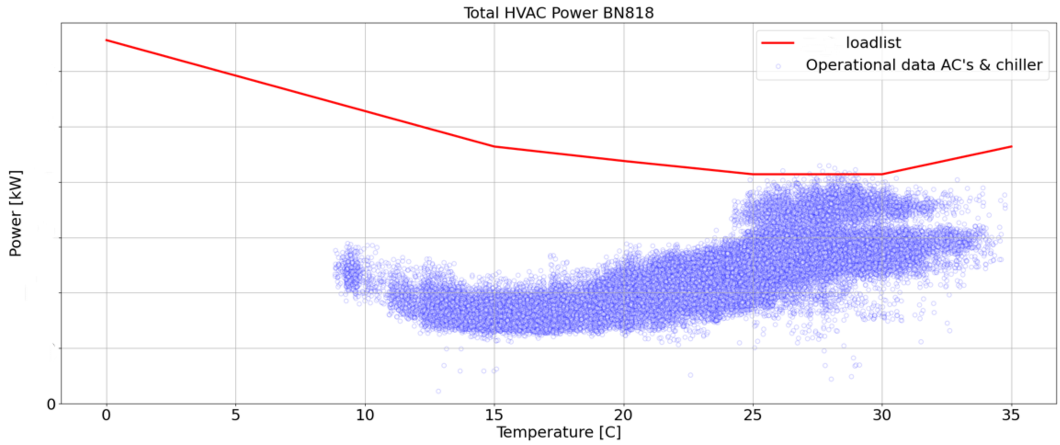


Figure 2.4: Total power consumption and HVAC Loadlist, Power (kW) vs. Temperature (°C) (Feadship, 2022)

Another method that the HVAC contractor uses to calculate the energy demand, is given by the International Organisation for Standardization (ISO). The ISO is a non-governmental worldwide federation of national standards. The main task of the technical committees is to prepare International Standards in all technical and nontechnical fields other than electrical and electronic engineering. ISO-7547 (2002) specifies design conditions and methods of calculation for air-conditioning and ventilation of accommodation spaces for all conditions except those encountered in extremely cold or hot climates. Opposed to the electrical components from the loadlist method, this method uses the heat production per area, number of air changes and the fresh air mixing ratio to calculate the total needed energy to maintain a certain temperature and humidity. An example of a part of the calculation is shown in figure 2.5. Shown here are the calculations where transmitted heat through walls and ceilings, solar radiation and emitted heat by persons and equipment is taken into account.

Guest Cabin 3					VOLUME	70.3 m3	PAGE:	1		
Maindeck	Area	29.3 m2		Ceiling height	:	2.4 m	Room temp	21		
				WIND. SURF.	A SURF.	K WATT/	dT	HEAT P	dT	COOL P
	l	x	b(h)	m2	m2	m2.K	K	WATT	K	WATT
Ceiling	1.0	x	20.1		20.10	0.50	27			
Windows (nxA)	1.0	x	4.9		4.88	4.80	27			
Outer bulkh. Long	4.7	x	3.4	4.9	11.11	0.50	27			
Outer bulkh. Trans.	0.0	x	0.0		0.00	0.50	0	0		
Inner bulkh. Long	7.7	x	3.4		0.00	0.80	3	0		
Inner bulkh. Trans	0.0	x	0.0		0.00	0.80	0	0		
Floor AC	1.0	x	12.8		12.80	0.80	6	0		
Floor crew bathroom	1.0	x	2.5		2.48	0.80	3	0		
TRANSMISSION LOSS----------------------TOTAL P-HEAT										
					K					
Heat of person sensible heat						75	x	2.0 =		
Solar radiation vertical light surface					0.5	12	x(A)	11.1 =		
Solar radiation vertical dark surface					0.5	29	x(A)	0.0 =		
Solar radiation horizontal light surface					0.5	16	x(A)	10.1 =		
Solar radiation horizontal dark surface					0.5	32	x(A)	0.0 =		
Solar radiation windows						150.5	x(A)	4.9 =		
Light						8	x(A)	29.3 =		
Apparatus etc.										
TRANSMISSION LOSS ECT.------------------------------TOTAL P-COOL										

Figure 2.5: ISO-7547 calculation example (Feadship, 2022).

ISO-7547 (2002) provides an accurate estimation of the energy demand per area at a certain condition. Each room is calculated at one moment with their maximum occupancy, maximum sun radiation, all lights on and all devices turned on at every moment. The design condition of this method is for summer an outdoor air temperature of 35 °C and humidity of 70% and for winter an outdoor air temperature of -20 °C and no specification for humidification. These conditions rarely occur however. Figure 2.6

shows the probability of occurrence of the relative humidity versus the temperature collected from 31 Feadships that are currently in use. The data shows that a temperature of 24.5 °C and a relative humidity of 75% has the highest occurrence. The occurrence of the ISO design condition of the HVAC system for summer is lower then 0.5%.

		Temperature											
		-5.5	-0.5	4.5	9.5	14.5	19.5	24.5	29.5	34.5	39.5	44.5	Sum
Relative humidity	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
	25	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.4
	35	0.0	0.0	0.0	0.2	0.3	0.3	0.4	0.3	0.1	0.0	0.0	1.6
	45	0.0	0.0	0.1	0.4	0.8	1.0	1.1	0.6	0.0	0.0	0.0	4.1
	55	0.0	0.0	0.1	0.9	1.8	2.8	3.1	0.9	0.0	0.0	0.0	9.7
	65	0.0	0.1	0.3	1.7	3.0	4.6	7.4	1.3	0.0	0.0	0.0	18.4
	75	0.0	0.1	0.4	2.4	4.3	6.1	14.3	2.3	0.0	0.0	0.0	29.9
	85	0.0	0.1	0.5	2.3	4.8	5.8	12.5	1.3	0.0	0.0	0.0	27.3
	95	0.0	0.0	0.3	1.1	2.1	2.4	2.4	0.2	0.0	0.0	0.0	8.5
	105	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
	115	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	125	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Sum	0	0.0	0.3	1.7	9.0	17.2	23.2	41.4	6.9	0.2	0.0	0.0	100

Figure 2.6: Probability of occurrence % of Relative humidity (%) vs Temperature (°C) (Feadship, 2022).

2.4. Modelling in the Maritime Industry

A physical system in the Maritime Industry is usually modelled using one of two types of mathematical approaches: white box models or black box models. A white box model approach is a model in which the structure is perfectly known and it has been possible to construct it entirely from prior knowledge, physical laws and deterministic first principle relations (Ljung, 2001). A black box model approach involves modelling a system solely based on observed data, without prior knowledge of the system. It primarily refers to learning algorithms, such as regression techniques and neural networks (Haranen et al., 2016). These models can be regarded as systems that describe the relationships between input variables and their corresponding outputs (Ljung, 2001).

A third option is to integrate both white and black box model approaches, which is known as grey box modelling. This grey box approach uses a model for which some physical insight is available but some parameters need to be estimated from the observed data (Ljung, 2001). Grey box methods can be considered as a combination of white and black box approaches, where the theoretical principles and physical laws of white box models are complemented by knowledge extracted through black box techniques from experimental data.

Studies by de Haas (2022), Odendaal (2021), Zwart (2020) and Bakker (2021) have demonstrated the potential of combining white box and black box models into grey box models. They collectively explore modelling techniques in ship design, comparing these three types of modelling. Key takeaways of their studies include:

- Data quality and size: All studies underline the importance of data quality and training data size. With smaller datasets, white box modelling appears more effective, while larger datasets show similar performance between grey box and black box modelling. Careful consideration must be given to data quality and focus, as poor quality input data leads to unreliable output.
- Accuracy consistency: Bakker (2021) and Zwart (2020) conclude that grey box modelling outperforms other model variants in accuracy, but the accuracy reduces when conditions change to more complex scenarios. They point out that traditional methods can still be relevant in these cases to create consistent accuracy.
- Interpolation and extrapolation: Black box modelling excels in interpolation and the white box component ensures precise predictions beyond the scope of the used training data. Black box lacks the ability to extrapolate accurately due to a lack of concrete knowledge about the problem, which white box modelling does have.

Overall, these studies suggest a need for more comprehensive and nuanced modelling approaches in ship design and recognising the importance of data quality and size. They collectively point towards a future where ship design is increasingly informed by advanced modelling techniques and data-driven insights.

2.5. Available data

For this study, DVNA provides a substantial amount of data collected from their yachts that are currently sailing, containing yacht parameters, engine and motion data, voyage reports and auxiliary power data. An overview of the accessible data is provided in Table 2.2. The provided data contains ship design specifications and the layout of all onboard systems. Additionally, new yachts are fitted with a variety of sensors and monitoring equipment. This allows for continuous data collection on parameters such as engine and generator power demand and experienced ship motions, respectively. The weather and climate is something that is continuously monitored and stored as well. While these are relatively new and still in their early stages, the amount of real-time data of operations of the Feadship fleet is currently growing a lot. As more data becomes available, research is needed to look at the applicability of this information. This data presents the opportunity for comprehensive modelling and potential fresh insights into the problem. Defining the parameters of this dataset and assessing the available versus unavailable data is crucial for establishing the requirements for the methodology.

Table 2.2: Available dataset types, locations and descriptions.

Data Type	Data Source	Description
Yacht parameters	DVNA design department	Ship design parameters: hull shape information (L, B, C_b, C_p), general arrangements of systems.
Engine and Motion data	Sensor monitoring	Main engine and generator power (shaft power), tank levels (consumption), ship motions (pitch and roll).
Voyage report data	7SEAS Portal Initiative	Yacht location, speed and heading. Wave and wind conditions. Weather related parameters.
Auxiliary power data	Project HOTEL Initiative	Recorded total auxiliary loads, AC power (voltage, amperage, fan speed), air and sea temperature, relative humidity.

2.6. Problem Conclusion

At its core, the fundamental challenge lies in the fact that current energy demand estimation method in the early stages of design fixate on conservative design points and extreme operating conditions. This choice is made by the HVAC contractor and solely relies on the experience of the company. Consequently, DVNA encounters a substantial uncertainty in the estimation compared to the energy consumption during real-world vessel operation. This difference jeopardizes the advancement of sustainable vessels and the implementation of optimal design methodologies. Additionaly, a continiously expanding voyage dataset of currently sailing Feadships is available. The available information on the HVAC systems is not integrated yet in the energy demand calculation. A grey box modelling method has the possibility to improve the level of accuracy and certainty with the help of the integration of operational voyage data. The main problems to solve are listed in the table below.

Table 2.3: Problem Definition.

Problem	Definition
P. 1	The use of activity rates of devices and other standard values in the loadlist calculation cause an overestimation of the energy demand of the HVAC system.
P. 2	The ISO-7547 method is proportioned on extreme weather conditions instead of the condition with the highest occurrence.
P. 3	A substantial database of operational data is available and is not incorporated yet into the design of HVAC systems.

2.7. Method Requirements

With the introduction in chapter 1 and the problem analysis in chapter 2, a clear image is drawn of the issue with the estimation of the energy demand of HVAC systems. A model is proposed to generate accurate predictions for the early stages of design. The goal is to improve the existing estimations and combine them with data collected by DVNA from their yachts. A new method taking the current problems, mentioned in section 2.6, into account can provide a more extensive estimation of the energy demand. Before the potential approaches and methodology are discussed in the next chapters, the method requirement are given below:

Table 2.4: Method requirement breakdown.

Method Requirement	Description
MR. 1	A model that predicts the energy demand of the HVAC system across diverse operational scenarios.
MR. 2	A white box model that includes a revision of the standard values in order to create certainty in this part of the estimation.
MR. 3	A white box model that proportions the calculations to weather conditions which actually occur for a representative estimation.
MR. 4	A black box model that includes a trained model that incorporates the available voyage data to predict the energy demand of the HVAC system.
MR. 5	A grey box model that combines white box model as input for the black box model to enhance the accuracy of the energy demand estimation.
MR. 6	A method that is based on real voyage data provided by Feadship & DVNA.
MR. 7	The method should possess the capability to manage and identify inconsistencies or inaccuracies within the voyage dataset.

3
Solution Approach

This chapter will provide an answer to the third research sub-question: *What methods are suitable and in what ways can machine learning and grey box modelling contribute to solving the problem?* A method is presented that will fit the method requirements discussed in section 2.7. In section 3.1, the white box model will be explored and in section 3.2 the black box model will be explored. The data preparation is given in section 3.3. In section 3.4 the grey box modelling configuration is explained. Finally, the coverage of the method requirements is shown in 3.5

3.1. White box modelling

After completing the problem analysis, the initial selection of methods for the white box modelling process is next. The white box model consists of the physical processes occurring in the HVAC system, granting a clearer understanding of the dynamics of the system in and around the yacht. In subsection 3.1.1, the different methods are evaluated to determine their suitability for the current research. For the white box modelling approach, the objective is to include a revision of the standard values in order to create certainty in this part of the model, as mentioned in MR.1 of the method requirements, section 2.7. Method requirement MR.2 also needs consideration in the white box modelling part. In subsection 3.1.2, an evaluation is made in the extent to which the white box model covers the problem. This assessment will help identify the components of the problem that remain unaddressed.

3.1.1. Model Selection

The emphasis of a white box model is on understanding the system's physical behavior. However, the current white box models used for estimating the energy demand of the HVAC systems onboard use a high amount of estimated coefficients and values (ISO-7547, 2002) (Feadship, 2022). These coefficients are calculated in a statistical manner and include errors as mentioned in section 2.3 (Feadship, 2022). In table 3.1, the different research done on the configuration of different white box models and its applicability for this research are given.

Table 3.1: Current methods and relevant research.

Source	Method	Applicability
Estimated Electrical Load Balance (Feadship, 2022)	Load balance sheet containing all components onboard with manually assigned running activity percentages.	Low applicability, focuses on electrical demand. Listing the influencing components is useful.
ISO-7547 (2002)	Determines the required power to maintain a state of equilibrium in an area.	Basis of this estimation is highly applicable, adjustments are needed.
Enhancing early-stage energy consumption predictions using dynamic operational voyage data (Odendaal, 2021)	Based on ISO-7547 (2002) and Boertz (2020), expands the ISO method with Feadship standards.	Shows potential with ISO for yachts. But uses simplified values for transmission factors and a static calculation which is not time dependent.
Internal documentation by (Kamstra, 2020) and (de Vroet, 2018)	Uses the same principles as ISO-7547 (2002), improves the calculation on the parts of sun radiation, external temperature and time dependency.	Very applicable, but both methods make limited use of real voyage data. Also the estimation is only for one area of the yacht.

The basis for the white box model are the methods of calculation for air-conditioning and ventilation of accommodation spaces given by ISO-7547 (2002). It determines the required power to maintain a state of equilibrium when considering dynamic factors such as heat transmission through walls, ceilings and floors, solar radiation, equipment and persons in the room. Odendaal (2021) made use of this method to estimate the power of a HVAC system. In cooperation with De Voogt, K. Odendaal makes use of a grey box model approach to estimate the total power usage of a yacht including the HVAC system. For the HVAC system particularly a white box model approach is used. The estimations show improvement compared to the empirical load list, but still employs a lot of assumptions and standardised values. For example, heat transmission values are taken from standard values (ISO-7547, 2002).

Research by de Vroet (2018) makes use of the same principles as ISO-7547 (2002) but improved the estimation at different parts in the estimation. For example, taking the solar radiation into account at certain times of the day at different sailing routes and climates. In the next section, the white box model method for this research will be discussed with the differences and improvements compared to the existing methods.

3.1.2. White box limitations

A white box model calculation, which involves developing a detailed physics-based or mechanistic model based on known principles and equations, is not sufficient to accurately calculate the energy demand of a HVAC system. The current white box models based on ISO-7547 (2002) show the following limitations:

- **Complexity of a HVAC system** Yachts have highly complex HVAC systems that involve various interconnected components, such as compressors, heat exchangers, fans, valves, and control systems. Modelling all these components accurately in a physics-based manner can be exceptionally complex and time-consuming. A white box model will nearly always be a simplification of the real situation.
- **Variable Operating Conditions** The operating conditions of the HVAC system can vary significantly based on factors like outdoor temperature, humidity, occupancy, insulation levels, and system load. Incorporating these variations accurately into a physics-based model can be challenging.

- **Real-World Uncertainties** Real-world uncertainties, such as opening doors, non-uniformity in material properties, or variations in the quality of components, can significantly affect the performance of the HVAC system. It's difficult to account for all these uncertainties accurately in a white box model.
- **Interaction with Other Systems** The HVAC system is interconnected with various other systems, such as electrical, plumbing, and energy storage systems. Capturing the dynamic interactions and feedback loops between these systems is a significant challenge for a white box model.
- **Human Behavior and Occupancy Patterns** Human behavior, occupancy patterns, and user preferences also play a critical role in determining the HVAC system's energy demand. Predicting these aspects accurately and incorporating them into a white box model is difficult.

MR. 2 and 3 are met with this white box model approach. To address the remaining gaps, another approach that combines white box modelling with data-driven techniques like machine learning is proposed.

3.2. Black box modelling

A black box gives a functional relationship between system input and output, without using any physical insight about the situation. According to Huotari et al. (2020) this modelling technique is especially valuable when the behaviour of a system is not clear or when a white box model is not accurate or predictable enough. A black box starts to recognise patterns and relations in the data by training the black box to give the right output. This is called machine learning and is part of a broader, more known term artificial intelligence.

Machine learning principles have led to enhancements in prediction and modelling techniques across different research fields. Instead of attempting to model all the underlying phenomena and physics related to a problem, the black box approach uses data for prediction without a deep understanding of the involved elements such as ships, water, resistance, or powering. Numerous studies showcase the application of black box modelling in predicting essential ship performance parameters, including propulsion power, fuel consumption, and speed. Diverse methods, including artificial neural networks and Gaussian processes (GP) show impressive results. Examples of such successful applications include studies by Pedersen and Larsen (2009), Petersen et al. (2011) and Yuan and Wei (2019).

A black box model can be more accurate than a white box model, but it also suffers from some clear disadvantages. Black box models require large amounts of data for training, and due to this extrapolation is limited to the datasets they are derived from (Leifsson et al., 2008). All patterns and relations found by the model are hard to assess since these are not visible and only happen inside the black box.

3.2.1. Model Selection

The objective now is to choose an appropriate machine learning algorithm based on the problem and available data to enable learning from the available data. As mentioned in 3.1.2, the white box model is unable to encompass every aspect, so a black box model is essential. The large amount of data available makes it possible to incorporate this type of method. Choosing the ideal model method presents a significant challenge. The research is shown in Table 3.2.

Table 3.2: Research on BBM in the maritime industry and on HVAC energy modelling for buildings.

Source	Description	Applicability
Prediction of Full-Scale Propulsion Power using Artificial Neural Networks (Pedersen and Larsen, 2009)	The paper gives an insight into the application of ANN to ship data collection. Propulsion power is considered instead of HVAC energy demand.	+/-
A machine-learning approach to predict main energy consumption under realistic operational conditions (Petersen et al., 2011)	Investigates two approaches; ANN and GP. Focuses mainly on fuel efficiency in ship propulsion.	+/-
Comparison of using artificial neural network and Gaussian process in ship energy consumption evaluation (Yuan and Wei, 2019).	Compares both ANN & GP to evaluate the ship energy consumption. But also mainly focuses on the propulsion system.	+/-
Physical energy and data-driven models in building energy prediction: A review (Chen et al., 2022)	Reviews multiple applications of black box modelling for the prediction of the energy demand of buildings.	+
Energy analysis of a building using artificial neural network: A review (Kumar et al., 2013)	Reviews multiple applications of ANN in analyzing the energy demand of a building including HVAC.	+
Optimization of HVAC system energy consumption in a building using artificial neural network and multi-objective genetic algorithm (Nasruddin et al., 2019)	Presents the optimization of HVAC system operations to minimise energy consumption for a residential building. Including the use of ANN. Shows the most promising method for the black box model.	++

One of the most applied methods within the maritime industry and for energy demand estimations is the artificial neural networks (ANN) method (Table 3.2). An ANN is a machine learning model inspired by the brain (Silva et al., 2017). It is made up of interconnected nodes with adjustable weights. Input data is fed into an input layer, processed through hidden layers using activation functions, and produces an output in the output layer (Silva et al., 2017). During training, the network adjusts the weights to minimise prediction errors using a learning algorithm. After training, it can make predictions or classifications based on new data. A schematic of an ANN is shown in figure 3.1.

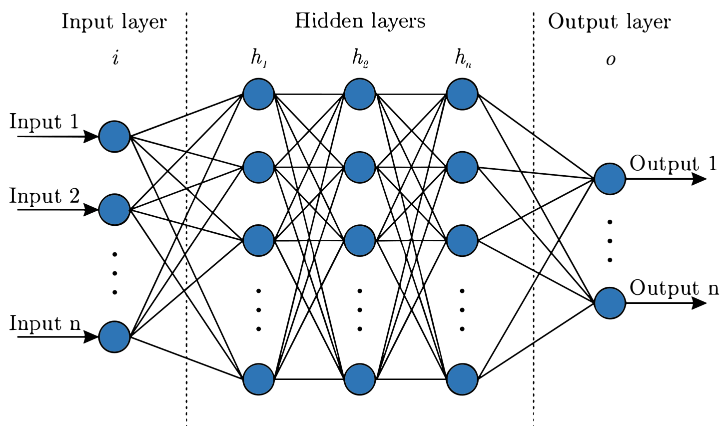


Figure 3.1: Schematic of a artificial neural network

One of the first to use this method to predict the propulsion power of container vessels is Pedersen and Larsen (2009). Additionally, Petersen et al. (2011) conducted a comparative analysis of ANN and Gaussian Process (GP) for modelling ship propulsion efficiency. Yuan and Wei (2019) did the comparative analysis between ANN and GP for ship energy consumption. Looking at this reviews there are multiple reasons ANN is more applicable then GP:

1. GP is more computationally demanding than a ANN (Yuan and Wei, 2019).
2. ANNs can capture non-linear relationships between input and output variables (Runge and Zmeureanu, 2019), making it more suitable for complex energy demand estimation tasks like the HVAC sytem.
3. ANNs scale better with large datasets compared to GPs (Petersen et al., 2011).

Looking at the use of ANN in the prediction of energy demand of HVAC systems, there is research done by Kumar et al. (2013), Chen et al. (2022) and Nasruddin et al. (2019) for residential buildings. These researches show promising results for the prediction of energy demand using ANN and are highly applicable since the layout of a yacht is comparable to a building.

3.3. Data Preparation

Both the white box model and the black box model use data as input for the model. This section outlines a methodology by García et al. (2016) and Zwart (2020) for preparing the raw data in modelling, emphasising its importance due to the data-driven nature of the process. Zwart (2020) adopted a novel cleaning approach developed by García et al. (2016) which is commonly applied within computer science applications. It details eight critical steps:

1. **Data integration:** Involves aligning multiple data sources, focusing on common features which in this case are timestamps. Once alignment is confirmed, datasets with higher frequencies can be used as a foundation for interpolating the remaining sources. Interpolation can cause errors in the dataset due to the creation of new data points that do not exist in the real environment. It is recommended to use varying datasets to reduce this error.
2. **Data transformation:** This converts non-numeric data (like strings) into numerical forms, using methods like ordinal encoding or one-hot encoding. Ordinal encoding is a standard method that assigns a distinct numerical value to each unique character string. These numerical values represent a ranked order, ranging from 0 up to the total number of entries. With one-hot encoding, each unique entry becomes its own data feature with a binary value of either 0 or 1. This is not necessary since all parameters are either continuous or binary already.
3. **Missing value imputation:** This addresses missing data. Strategies include removing incomplete data points or using interpolation and imputation techniques. In this case the data is selected on having at least one data point in a certain time frame. After this selection interpolation methods are applied to keep valuable data points without discarding missing data points.
4. **Data cleansing:** Data cleansing focuses on removing unrealistic data points using engineering insights, aiming to reduce noise and outliers. An example is the orientation of a vessel, where unrealistic values beyond the 360 degrees are eliminated.
5. **Noise identification:** Involves detecting outliers using methods like Standard Deviation. This assumes that data is normally distributed and requires an understanding of the data on which this is applied. Outliers might also be relevant for understanding the operations of the modelled system.
6. **Data selection** When the size of datasets increases, data models encounter a significant challenge due to the increased computational costs. It is therefore important to select an appropriate dataset size.

At this point the processed data is used for the white box model. The following steps are applied to the data for the black box model and ultimately the grey box model.

6. **Feature selection:** Involves selecting the most relevant features for the ANN. The goal is to obtain a subset of features from the original problem that still appropriately describe it (García et al., 2016). A good understanding of the white box model is therefore helpful.
7. **Data normalisation:** Finally, Silva et al. (2017) recommends scaling the input and output variables to prevent the saturation of neurons. Saturation happens when neurons operate at their maximum or minimum capacity, leading to a loss of sensitivity to changes in input. This scaling is based on the proportional segment principle, as illustrated by Equation 3.1. In this equation, 'z' represents the scaled value, while 'x' refers to the original value from the data.

$$z = 2 \cdot \left(\frac{x - x_{min}}{x_{max} - x_{min}} \right) - 1 \tag{3.1}$$

Each step is designed to refine the data quality for data modelling, underscoring the importance of accurate and clean data in achieving reliable model outputs.

3.4. Grey box modelling

Now the white box and the black box method are discussed, lastly the grey box method will be elaborated. As outlined by Leifsson et al. (2008), there are two primary categories of grey box models that differ based on their application: serial modelling and parallel modelling. Figure 3.2 shows the distinct compositions of these configurations.

A serial configuration entails a white box model and black box model arranged in series. The inputs are routed to both the white box model and black box model, however, the initial prediction (P') of the white box model is directly integrated into the black box model. In this scenario, an internal development of a mapping between the applied physics and operational data can be achieved. The parallel modelling approach includes a white box model estimation, where a black box model is simultaneously used to decrease the residual (R') between predictive and target data. The results are then combined to ascertain the final prediction. Series configuration uses data computed by one box to partly power the other. Parallel configuration adds different results of both boxes or finds a desirable average.

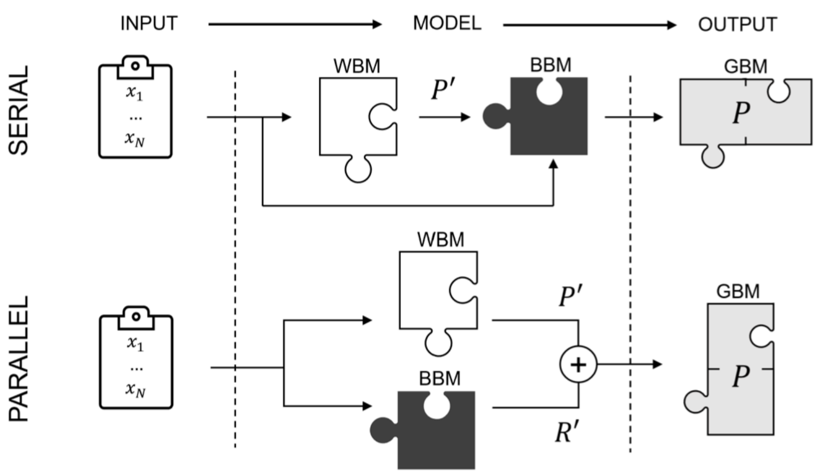


Figure 3.2: Possible grey box configurations (Odendaal, 2021), based on data from Leifsson et al. (2008).

Research has been conducted by several graduates of the TU Delft on grey box modelling and have shown promising results. A trim optimisation has been conducted by Zwart (2020) and an early stage energy demand predictions have been made by Odendaal (2021). Research on grey box modelling using ANN for HVAC systems is limited. The majority of existing studies, such as the one by Talib et al. (2023), focus on temperature prediction or control in HVAC systems through grey box modelling. Consequently, there is a need to propose a novel method in this area.

As discussed in subsection 3.1.2, the white box model does not cover every aspect of the energy demand calculation problem. The black box covers these remaining aspects. Firstly, the general arrangement and the voyage data of the yacht are both collected and processed for use in the model. The white box model uses this data to calculate an energy demand estimation. The heat load estimation of the white box model is used as additional input with the other data for the black box model to give a better prediction of the energy demand of the HVAC system of the yacht. This serial approach can be effective because each model captures different aspects of the systems behaviour. The result is the grey box as shown in figure 3.3.

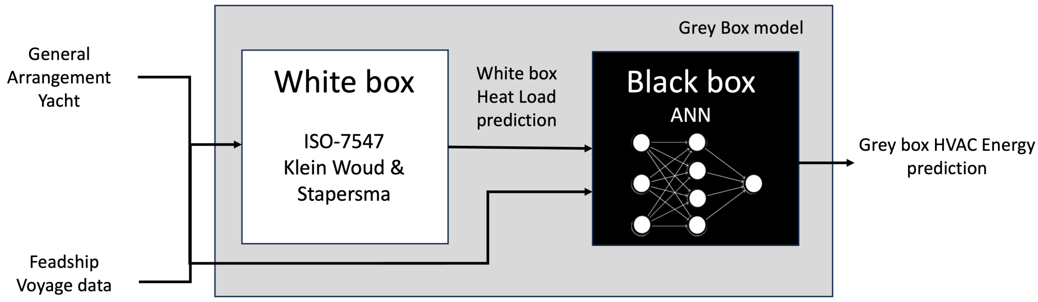


Figure 3.3: Proposed serial grey box configuration

3.5. Method Requirement Coverage

The method requirements discussed in 2.7 are evaluated of their coverage in the proposed method.

Table 3.3: Method Requirement Coverage

Method Requirement	Description	White box	Black box	Grey box
MR. 1	A model that predicts the energy demand of the HVAC system across diverse operational scenarios.	–	✓	✓
MR. 2	A white box model that includes a revision of the standard values in order to create certainty in this part of the estimation.	✓	–	–
MR. 3	A white box model that proportions the calculations to weather conditions which actually occur for a representative estimation.	✓	–	–
MR. 4	A black box model that includes a trained model that incorporates the available voyage data to predict the energy demand of the HVAC system.	–	✓	✓
MR. 5	A grey box model that combines white box model as input for the black box model to enhance the accuracy of the energy demand estimation.	–	–	✓
MR. 6	A method that is based on real voyage data provided by Feadship & DVNA.	-✓	✓	✓
MR. 7	The method should possess the capability to manage and identify inconsistencies or inaccuracies within the voyage dataset.	✓	✓	✓

4 Methodology

This chapter explores the methodology used for optimising the prediction of HVAC energy consumption, directly addressing the research sub-question: *How can the integration of the grey box approach optimise the prediction and understanding of HVAC system energy consumption?.* The research begins with section 4.1 which outlines the essential voyage data and operational parameters that form the basis of the models. This is followed by Section 4.2, which analyses the data to gain a better understanding of the operations op the yacht. Section 4.3 details the processes involved in integrating, cleaning, selecting and scaling the data to prepare it for effective modelling. Section 4.4 discusses the methodology of the white box model.

Section 4.5 focuses on how the grey box model incorporates an ANN. Within this section, subsection 4.5.1 introduces the basic principles of an ANN. Subsection 4.5.2 discusses the selection and optimisation of ANN hyperparameters, which are important in maximising the model's performance. Lastly, Subsection 4.5.3 describes the techniques used to assess the accuracy of the ANN configurations and the metrics used to confirm the models' validity.

4.1. Data Collection

The data collection primarily revolves around the use of Automatic Identification Systems (AIS) and various onboard sensors to gather crucial maritime information. AIS is a vital tool for tracking vessel movements, providing data on location, course, speed, and navigational status by broadcasting signals to nearby ships and shore-based stations. Together with the sensor data a comprehensive dataset is created in the 7SEAS portal of Feadship, which supports better decision-making in research and development. The weather data used in this methodology comes from the ECMWF database (ECMWF, 2023).

In Table 4.1, a list of the voyage data from Feadship that is required for the grey box calculation is presented. It is important to use proper data preparation for the data entered into the white box and grey box model. This will be detailed in the following sections.

Table 4.1: Required voyage data for calculation

Symbol	Unit	Description	Source - Data name
FAT	$^{\circ}C$	Fresh air temperature at time t	POLESTAR Science - AC.SENS.FAT
RAT	$^{\circ}C$	Return air temperature at time t	POLESTAR Science - AC.SENS.RAT
SAT	$^{\circ}C$	Supply air temperature at time t	POLESTAR Science - AC.SENS.SAT
FAH	%	Fresh air relative humidity at time t	7SEAS (ECMWF) - RH
RAH	%	Return air relative humidity at time t	POLESTAR Science - AC.SENS.RAH
SAH	%	Supply air relative humidity at time t	POLESTAR Science - AC.SENS.SAH
ϕ & λ	$^{\circ}$	Latitude and longitude of location of the vessel at time t	7SEAS - LAT LON
O	$^{\circ}$	Orientation of the vessel at time t	7SEAS - TRACK TRUE
u & v	m/s	Windspeed components at time t	7SEAS (ECMWF) - u & v
I_{BNI}	J/m^2	Beam Normal Irradiationat time t	7SEAS (ECMWF) - BNI
I_{DHI}	J/m^2	Direct Horizontal Irradiation at time t	7SEAS (ECMWF) - DHI

Table 4.2: Required voyage data for validation

Symbol	Unit	Description	Source
P_{AC}	kW	AC Power at time t	POLESTAR Science
$P_{Chiller}$	kW	Chiller Power at time t	POLESTAR Science
P_{Fans}	%	Fan Power at time t	POLESTAR Science
$P_{Reheaters}$	%	Reheater Power at time t	POLESTAR Science

4.2. Data insight

It is essential to get an insight of the available data before starting the data preparation. As García et al. (2016) outlines, this insight starts with a thorough exploration of the data to ascertain its structure, content, and the relationships within. Important initial steps include examining the types of data and identifying any missing or unusual data points. Visualisation techniques such as charts, graphs, and plots play a crucial role. These tools help to uncover patterns, trends, and outliers, providing a visual interpretation of complex relationships that can help with further analysis.

A few figures are provided in the following section in order to give an overview of the operational profile of the yacht in question and its use. First the climate which the yacht operates in is shown. The temperature of the fresh air is shown with the relative humidity. These variables contribute the most to the energy demand of the HVAC system, since these are the two parameters that need to be regulated for a comfortable climate. It can be noted that the mean of both of the fresh air parameters is above the comfort zone of 21 $^{\circ}$ and 50% relative humidity, stated by Feadship (2015).

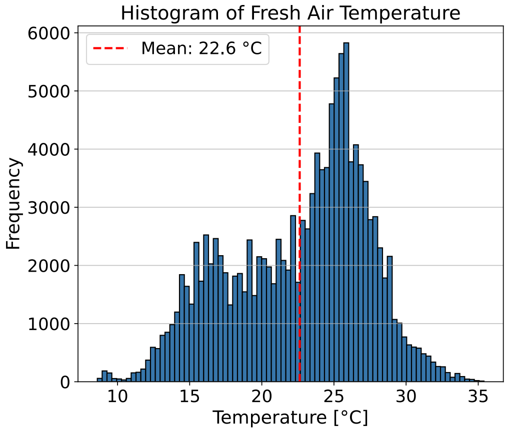


Figure 4.1: Distribution of Fresh Air Temperature (7SEAS sensor data)

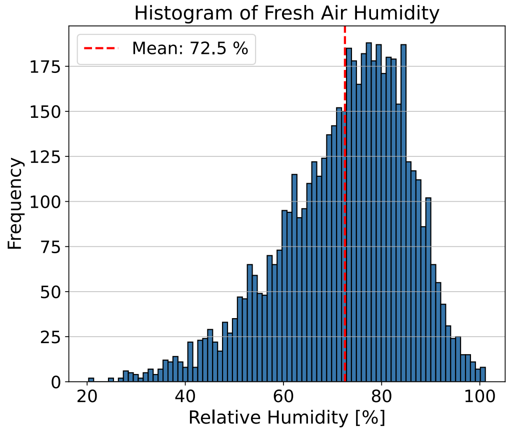


Figure 4.2: Distribution of Fresh Air Humidity (ECMWF, 2023)

Another variable that contributes to the temperature inside the yacht is the solar irradiation. This data is taken from weather models by ECMWF (2023) based on observations. In order to calculate the solar irradiation on a tilted surface, the Beam Normal Irradiance (BNI) and Diffuse Horizontal Irradiance (DHI) are needed. BNI quantifies the direct solar energy on a surface perpendicular to the sun, while DHI measures scattered sunlight reaching the Earth. More on this is explained in Section 4.4. The solar irradiation is put on a logarithmic scale since at night there is no sun in the sky, which would cause a high frequency at zero in the histogram of the solar irradiation.

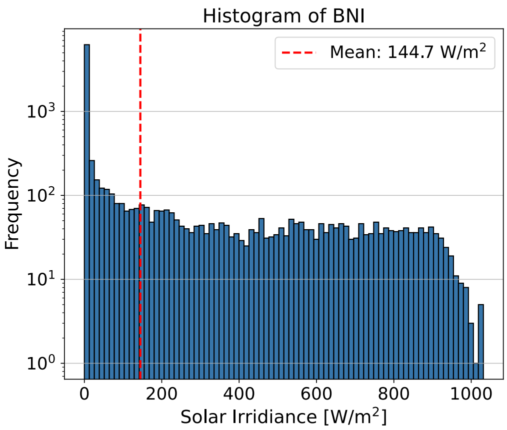


Figure 4.3: Distribution of Beam Normal Irradiation (ECMWF, 2023)

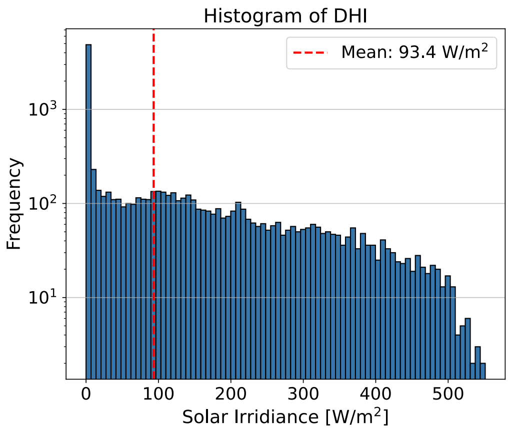


Figure 4.4: Distribution of Diffuse Horizontal Irradiation (ECMWF, 2023)

To be able to understand the use of the yacht, the heading of the vessel and its location are portrayed in Figure 4.6 and 4.5. It is visible in which areas the yacht operates in. This is also useful in understanding the climate that matches with these areas. It can be seen that the data is not always available and some part of the voyage are missing. This should be taken into account in the preparation.

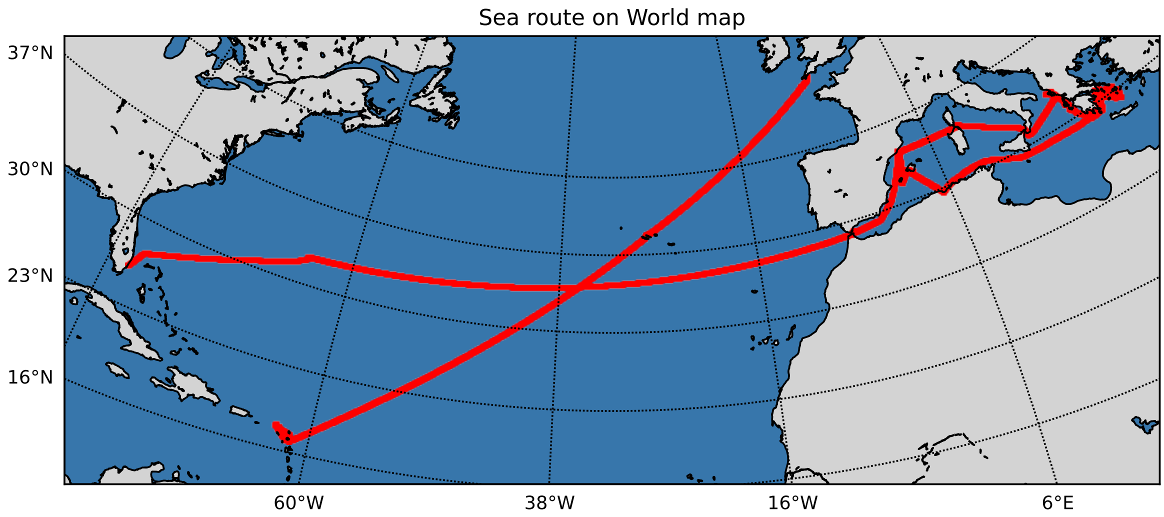


Figure 4.5: Sea route of selected yacht.

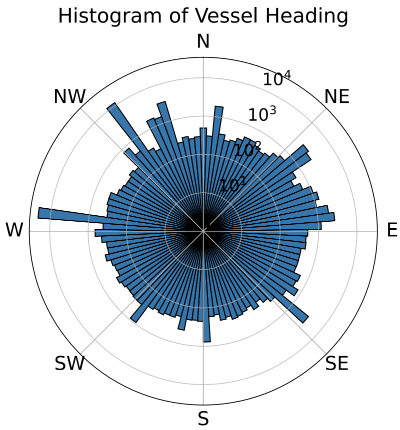


Figure 4.6: Distribution of Vessel Heading (Logarithmic scale)

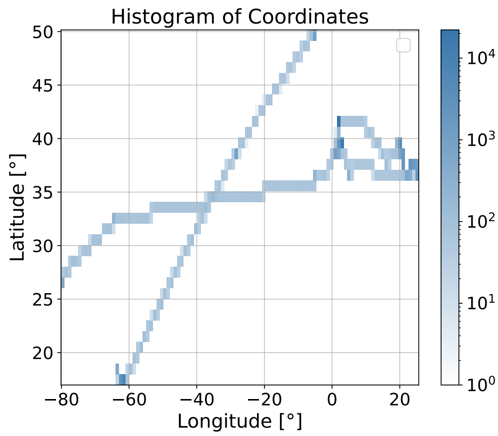


Figure 4.7: Distribution of Coordinates

Figure 4.8 presents the return and supply air temperature and humidity. This data is viewed to gain an understanding of the operating settings of the system. The HVAC system is consistently warming and humidifying the air throughout the days. The observed behaviour in the HVAC system's data, showing gradual increases and decreases in both return and supply air temperatures and humidity, could be attributed to several factors. Changes in weather conditions, such as rising outdoor temperatures and humidity levels, could be influencing the indoor climate, necessitating the HVAC system to work harder to maintain comfort levels. Internally, the presence of additional heat and moisture sources, such as people, electronic devices or other activities, could contribute to the trends in temperature and humidity.

The efficiency of the HVAC system is influenced by its operational strategies, such as temperature and humidity setpoints, which determine its performance. Some of the setpoints are clearly visible, this is where the temperature or humidity stabilises on a certain value.

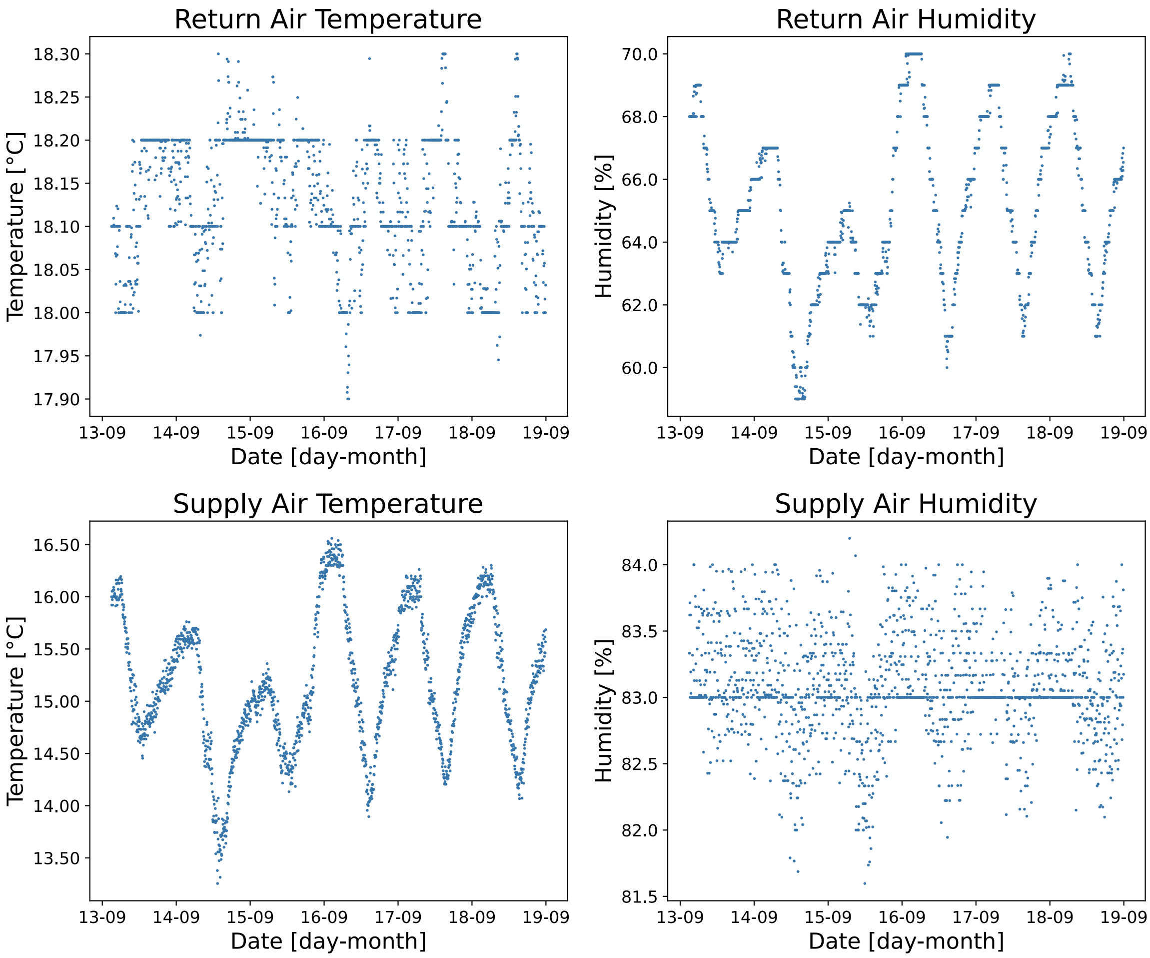


Figure 4.8: Supply and Return Temperature and Humidity.

4.3. Data Preparation

In section 3.3 the different steps of data preparation by García et al. (2016) are outlined. This section will go into more detail about what type of preparation is used specifically on the data sets.

Data Integration

To prepare the data for integration, it's important to merge various sources of information. The sensor and AIS (Automatic Identification System) data that is collected from the yacht directly and the weather data of ECMWF are all time-dependent. Sensor data includes variables like temperatures, humidity's wind speed, and heading. Combining the AIS data with the ECMWF database offers predictions on weather conditions like DHI, BNI, relative humidity and wind speed based on the ship's location, with a forecast being generated every hour. This approach allows for the integration of sea state information with known sensor data points, providing insights into the conditions the ship was experiencing at those times. The primary focus for this integration is on the timing of power measurements, as the goal is to estimate the energy demand and compare these to the actual power consumption. The power and sensor data time step is every three minutes, with the value representing a 3-minute average value. This already minimises the chance on outliers.

Data Transformation

Data transformation is a crucial step in data preprocessing, where raw data is converted into a format more suitable for analysis. One-hot coding is a specific type of data transformation commonly used to handle categorical variables by converting them into a binary format. However, when dealing with datasets that are already numerical, there is no need for one-hot coding. All of the data mentioned in section 4.1 is already numerical data and can be used directly in the model without the need for this transformation.

Missing Value Imputation

The imputation of missing values can be done by interpolation. The data is selected to have at least one data point in a certain time frame. If the surrounding data points are present, interpolation is performed. This is done to keep a substantial amount of valuable data points. Otherwise a lot of data points will be lost because of the merging of multiple data variables from different sources.

Data Cleansing

Data cleansing is performed by removing unrealistic data using engineering insights. During the interpolation process the yacht's heading is taken into account, because normal interpolation might cause an error with values no interpolating the right way in the range of 360-degrees. This consideration is vital for accurately merging sensor data with weather forecasts, as the direction in which the yacht faces can greatly affect the environmental conditions it encounters. Another example are negative solar radiation values that need filtering, since a negative solar radiation is not possible and would cause an error in the calculation.

Noise Identification

In the solution approach, specifically in subsection 3.3, the data preparation mentions noise identification (García et al., 2016). There are multiple ways to filter noise out of data using Standard Deviation, IQR or Chauvenet's criterion (Lin and Sherman, 2023). Chauvenet's criterion is particularly useful with larger datasets, because it filters less datapoints out compared to IQR. The Chauvenet's criterion is designed to identify and filter out outliers that deviate significantly from a Gaussian distribution, using the standard deviation of the data columns. In equation 4.1, the term x_i is a chosen i-th data point. The term μ stands for the mean and σ for the standard deviation of the whole dataset. The number of data points in the set is given by n and $erfc$ is the complementary error function. The criterion states that the data points for which the expression is true are outliers and should be removed.

$$n \cdot erfc\left(\frac{|x_i - \mu|}{\sigma}\right) < 0.5 \tag{4.1}$$

Data Selection

Finally, the whole data set is structured and selected. Different subsets of the data are made to make initial calculations less computational expensive. A subset of ten days is selected to perform calculations with and test the model. After the model is finalised, the full data set is used for the calculation which is detailed in section 4.4.

Feature Selection

For the grey box and black box model further data preparation is needed. A feature selection most be performed, before the configuration of the ANN is determined. Selecting relevant features is crucial for the ANN to perform well. According to Parkes et al. (2018), features with a low correlation to the target variable can cause unnecessary complexity, which will effect the performance and generalization capabilities of the ANN model. A first selection is made of potential model input features. The prediction of energy consumption of the HVAC system should be able to be performed for future situations, this is why the first selection is made according the following criteria:

- The goal is to predict the energy of the HVAC system at certain environmental conditions. So these should be included in the feature selection. These are the primary features since they directly influence the HVAC system's performance.
- Certain features like power output of components, coolwater temperatures, compressor speeds, and valve outputs are closely tied to the HVAC system's energy use but cannot be predicted in advance. Excluding these helps focus the model on environmental predictors and makes the model more generally applicable.

- Navigation features should be considered. The heading determines which side of the yacht catches the most solar radiation and warms the inside temperature. The location is excluded. By excluding the precise location of the yacht, It ensures that the model's predictions are based on immediate environmental conditions and the orientation of the yacht, making it versatile and applicable in any geographic location.
- The hour of the day or month should be excluded since the model needs to perform consistently regardless of the time. This approach focuses the model on current environmental conditions rather than time-based patterns, which can vary significantly across different days or seasons.

For the final input selection a Spearman's correlation is used. Parkes et al. (2018) and Zwart (2020) show that a Spearman's correlation can provide insight on the correlation between features. The heatmap in figure 4.9 displays the correlations between the different variables.

The heading and the FAH from the ECMWF database show a very low correlation with the Power Consumption. This weak correlation is likely to degrade the performance of the ANN, resulting in the exclusion of these features for the model. Although the Mixed Air Temperature and the Fresh Air Temperature show a high correlation to eachother, they might both be relevant since the Mixed Air Temperature contains specific information about the mixing amount with the Return air. The final selected features and the Spearman's correlation are given below:

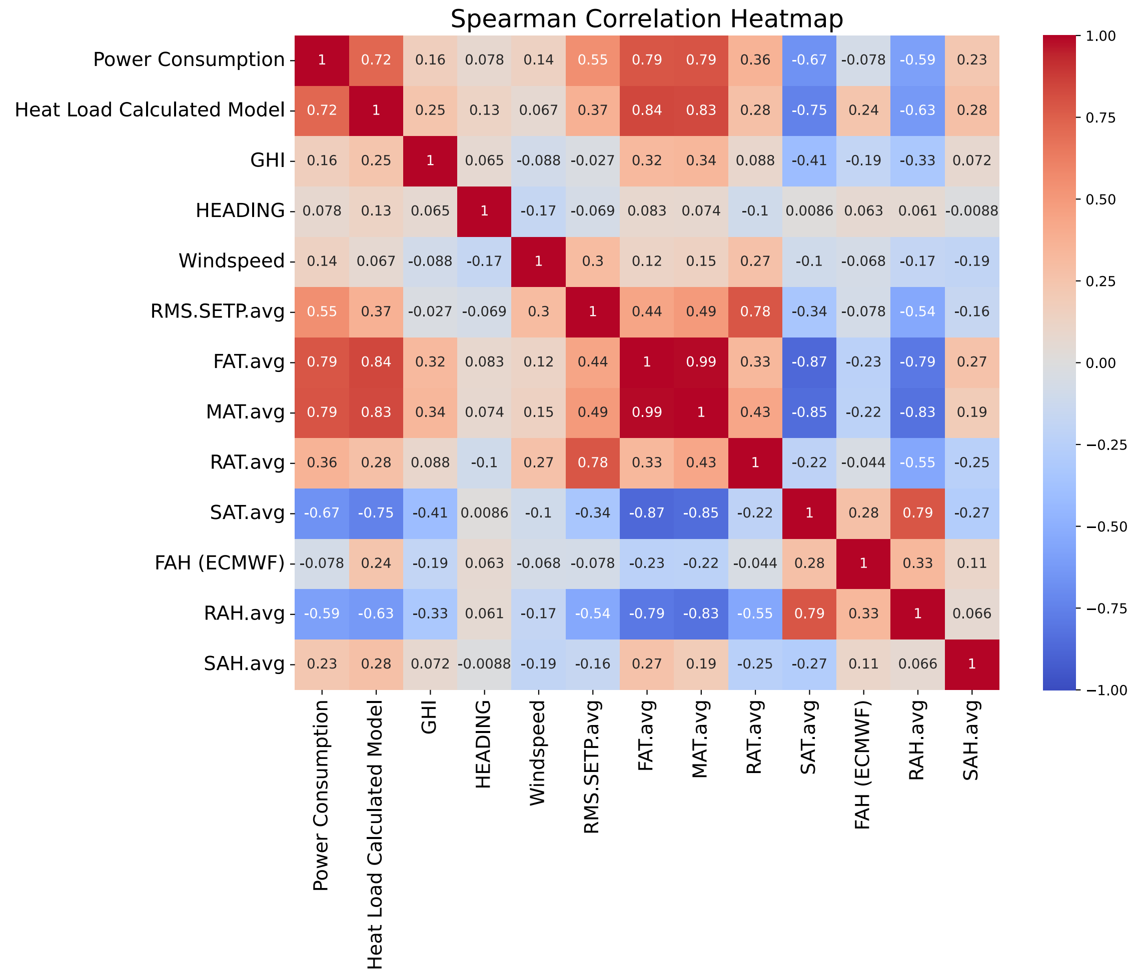


Figure 4.9: Spearman's Correlation Heatmap

- Heat Load Calculated by white box Model
- Global Horizontal Irradiance
- Windspeed
- Average Room Setpoint Temperature (RMS.SETP.avg)
- Average Return Air Temperature
- Average Fresh Air Temperature
- Average Mixed Air Temperature
- Average Supply Air Temperature
- Average Return Air Humidity
- Average Supply Air Humidity

Data Normalisation

With machine learning, data scaling uses techniques like normalisation and standardisation, which are essential during the preparation stage. Normalisation adjusts the data values to a common scale without distorting differences in the ranges of values, ideal for when the algorithm must recognise relationships between data points for prediction purposes. It typically involves rescaling the values into a range of [0,1]. Standardisation transforms data to have a mean of zero and a standard deviation of one, thus conforming to a standard normal distribution. This is particularly useful when the data has outliers or a non-uniform distribution, as it enables the model to better understand the importance of each feature by giving a balanced weight, irrespective of the original distribution's scale.

Uniform scaling through normalisation ensures all values are equally proportioned, thereby allowing the model to emphasise more significant variables without needing to account for scale variances—streamlining the training process. Conversely, standardisation can be more robust to outliers, making the features more comparable and often improving algorithm performance. The method employed to achieve this normalised scaling is shown in equation 4.3 .

$$x_{\text{norm}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{4.2}$$

The target variable is also normalised. The ANN will have a scaled output. These can be transformed back to real values with equation 4.3, taking the maximum and minimum value of the original used grey box dataset for the target vector:

$$x = x_{\text{norm}} \cdot (x_{\text{max}} - x_{\text{min}}) + x_{\text{min}} \tag{4.3}$$

4.4. White Box Methodology

For the white box model the methodology is detailed in the following section. The calculation consists of the following steps and is based on the methods used by ISO-7547 (2002) and Klein Woud and Stapersma (2016):

1. Prediction of heat gain per cabin or area
2. Obtaining required number of air changes
3. Obtaining fresh air (FA) ratio of supply air volume
4. Calculation of intermediate conditions in the fan coil unit or air handling unit

These different steps are explained in the following part. The calculation is done for every ACU that controls multiple rooms. Figure 4.10 shows multiple rooms in the blue area controlled by one ACU and the green area controlled by another ACU. The example yacht used in this research has a total of 5 ACU's.

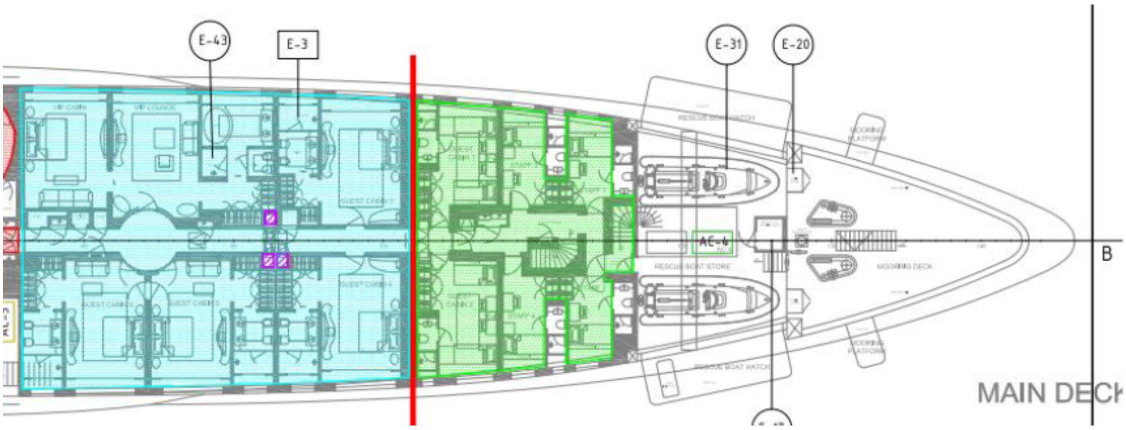


Figure 4.10: Part of AC system layout (Feadship, 2022)

1. Heat Gain per cabin

The sensible heat gain for an area i can be assessed by the total sensible heat (Q_{sensible}). This is a summation of all the components that influence the heat flow. These components can be categorized as the transmission of heat through surfaces, the heat gain from crew and guests, the heat gain from sun radiation and the heat gained by auxiliary systems like lighting and power devices (ISO-7547, 2002). A schematic image of the heat gain influences of equation 4.4 is shown in figure 4.11.

$$Q_{\text{sensible},i} = Q_{\text{transmission},i} + Q_{\text{person},i} + Q_{\text{radiation},i} + Q_{\text{auxiliary},i} \tag{4.4}$$

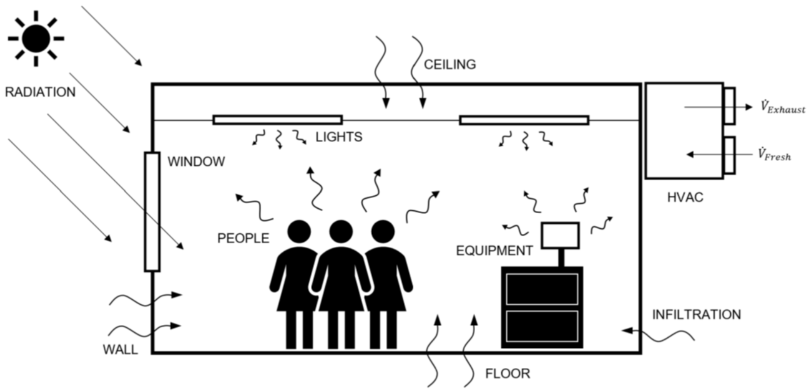


Figure 4.11: General Heat gain schematic (Odendaal, 2021)

The heat gain due to the transmission of heat through surfaces can be calculated for every individual area by equation 4.5. Where A_s is the surface of area i with a heat transmission factor k_s. ΔT is given by the difference in temperature between both sides of the surface. The total of all surfaces of an area is then given by:

$$\dot{Q}_{\text{transmission},i} = \sum_{s} (\Delta T \cdot A_s \cdot k_s) \tag{4.5}$$

The temperatures of the different areas are given by data that is recorded in the voyage data of the yacht or acquired from the ECMWF Weather Database (ECMWF, 2023). Instead of a constant temperature, the temperature will depend on the current outside temperature given by the Fresh Air Temperature T_{FAT} given by the voyage data at a certain time.

The heat gain by a person at an indoor temperature of 27 $^{\circ}C$ is 70 W sensible heat at rest (ISO-7547, 2002). The amount of crew or guests in each area differs. An assumption is made for each area based on the function and size of that area.

$$Q_{\text{person},i} = (n_{person})_i \cdot 70[W] \tag{4.6}$$

Solar radiation is also of influence on the heat gain of an area. Solar radiation reaches the Earth's surface in three forms: direct (beam) solar radiation, diffuse solar radiation, and reflected radiation. Reflected radiation can often be neglected for simplicity. The total radiation received by a horizontal surface at ground level on a day is primarily the sum of direct and diffuse radiations as shown in equation 4.7 . Direct solar radiation's intensity significantly depends on the orientation of the receiving surface, whereas diffuse solar radiation is considered nearly uniform across different orientations, despite minor variations in reality.

$$I_{Total} = I_{Direct} + I_{Diffuse} = BNI + DHI \tag{4.7}$$

The beam normal irradiance (BNI) and diffuse horizontal irradiance (DHI) are both collected from the ECMWF weather database. The calculation of the solar radiation on a surface contains several steps:

Relative solar position For each time step, the solar azimuth and elevation angles are calculated using the coordinates of the yacht which are stored in the navigation data of the yacht. This data also shows the true heading of the yacht, which is used to determine the relative position of the sun, as shown in Figure 4.12.

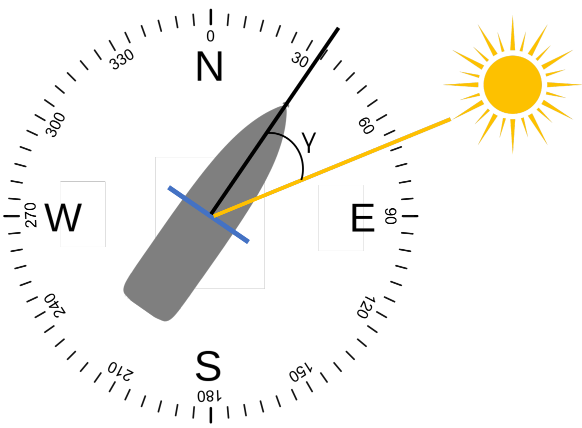


Figure 4.12: Solar angle on a bow facing surface.

Angle of incidence The angle of incidence (θ) of solar radiation on the surface is determined, considering both the solar elevation angle and the surface tilt. This angle is crucial for calculating the effective solar irradiance on the surface. Five different surfaces are considered. The four vertical surfaces facing towards the bow, starboard, stern and port side. The fifth surface is a horizontal surface, where the azimuth angle does not influence the radiation on the surface. The surface-solar azimuth angle is given by γ and calculated by 4.8. In order to calculate the angle of incidence, equation 4.9 is used. The corresponding angles can be seen in Figure 4.13.

$$\gamma = \phi - \psi \tag{4.8}$$

$$cos(\theta) = cos(\beta) \cdot cos(\gamma) \cdot sin(\Sigma) + sin(\beta) \cdot cos(\Sigma) \tag{4.9}$$

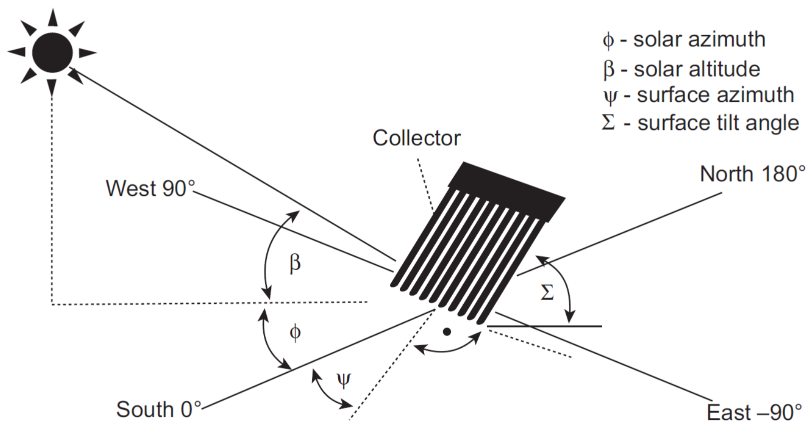


Figure 4.13: Solar angles with respect to a tilted surface. (Sarbu and Sebarchievici, 2017)

The total radiation on a surface is then calculated by equation 4.10 (Sarbu and Sebarchievici, 2017). This results in a varying solar radiation on each surface of the yacht portraited in Figure 4.14.

$$I_{surface} = cos(\theta) \cdot BNI + \frac{(1 + cos(\Sigma))}{2} \cdot DHI \tag{4.10}$$

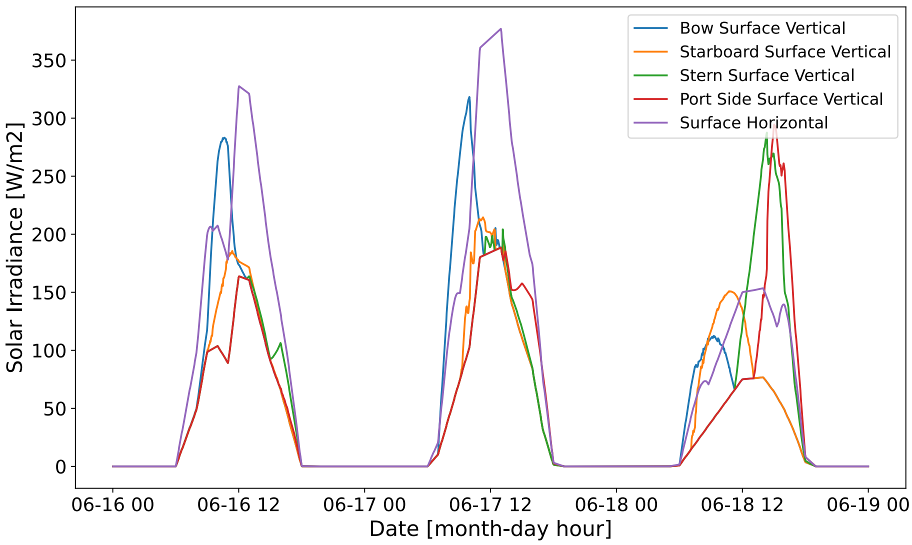


Figure 4.14: Solar irradiance on different orientated surfaces.

The solar radiation calculations are extended by incorporating the physical geometry of shadowing objects to adjust direct solar irradiance on the target surface, enabling a more accurate estimation of solar exposure for surfaces under various environmental conditions. The shadow length due to overhangs is calculated using the overhang depth and the sun's elevation angle. This length determines the proportion of the window shaded by the overhang. The height of the shadow cast on the window by the railing is computed based on the railing's height, its distance from the window, and the sun's elevation angle. This calculation provides the height of the shadow on the window, allowing for an estimation of the shaded portion of the window. The shadow effects from both the overhang and the railing are combined to determine the total shadowed portion of the window. This step is critical for estimating the amount of direct solar radiation reaching the window.

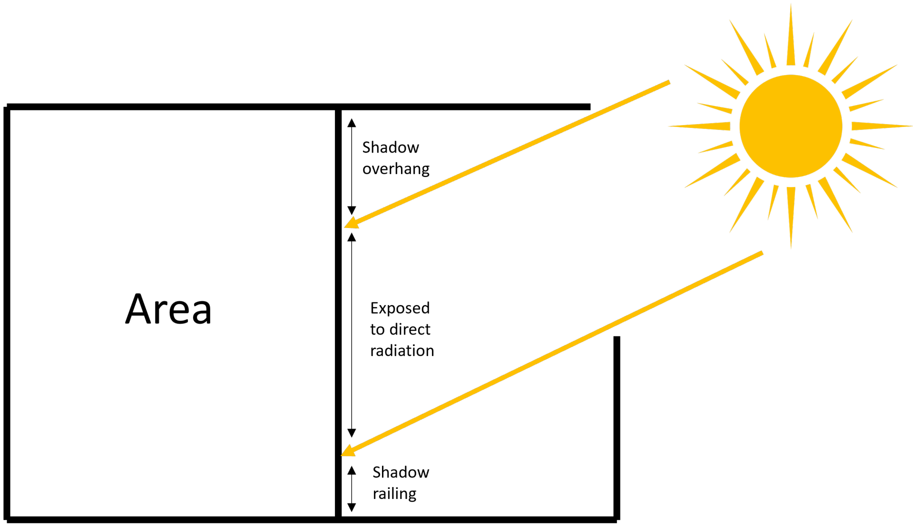


Figure 4.15: Solar radiation on different oriented surfaces of the yacht.

$$I_{surface} = cos(\theta) \cdot BNI \cdot (1 - \%_{shadow}) + \frac{(1 + cos(\Sigma))}{2} \cdot DHI \tag{4.11}$$

According to ASHRAE (2013), solar radiation affects the outer surfaces of walls and roofs, which results in raised temperature of the surfaces compared to that of the ambient air temperature. This temperature increase of the exterior surface is recognised as Sol-air temperature (T_s). The Sol-air temperature is influenced by various factors such as the structure's properties, the material and colour of the exterior surface and the intensity of solar radiation perpendicular to the outside surface. The Sol-air temperature can be calculated with equation 4.12 using T_o as the outside temperature, α as the absorption coefficient of the outside surface, $I_{surface}$ as the total net solar radiation on the surface and h_c convective heat transfer coefficient. The absorption and convective heat transfer coefficient are further explained in Appendix B.

$$T_s - T_o = \alpha * \frac{I_{surface}}{h_c} = \Delta T_{solar} \tag{4.12}$$

Equation 4.13 considers all surfaces of an area that are in direct contact with the sun. Parameter c_{shadow} is the percentage of the total surface that is covered by a shadow due to a balcony or overhang, ΔT_{solar} is the difference between the surface temperature T_s and the inside temperature T_in and $A_{s,\text{normal}}$ is the area of the surface. The equation also includes the solar heat gain through glass surfaces. This is calculated with the Solar Heat Gain Coefficient (SHGC) of glass which are given by Feadship (2015).

$$Q_{\text{solar},i} = \left(\sum_s (c_{\text{shadow}} \cdot (k_s \cdot \Delta T_{solar} \cdot A_{s,\text{normal}}) + SHGC_{glass} \cdot I_{solar} \cdot A_{s,\text{glass}}) \right)_i \tag{4.13}$$

The auxiliary heat flux arises from equipment and lights present in the area, generating a certain amount of heat in the surrounding area. Consequently, the overall expression can be divided into two distinct parameters.

$$Q_{\text{auxiliary},i} = \sum Q_{\text{light},i} + \sum Q_{\text{equipment},i} \tag{4.14}$$

According to ISO-7547 (2002), when areas have no light, the heat gain caused by lighting is determined based on the rated wattage of the lights. Generally, LED lights are commonly used, although the

ISO-7547 (2002) only offers guidance for incandescent or fluorescent lights. Nonetheless, LED and fluorescent lighting produce a similar level of illumination. While precise heat gains can be obtained by analysing individual spaces and lighting types, the general guidelines are considered sufficient for an estimate.

In addition to lighting, the heat gains per square meter are also influenced by powered equipment. This equipment is known and listed per room to provide a comprehensive overview of all parameters necessary for the heat load balance.

2. Number of air changes

Once the heat gain per cabin has been determined at a certain time step, it is possible to predict the minimum number of air changes necessary to maintain the desired environmental conditions within a specific area. The number of air changes per hour (ACH) required is dependent upon various factors: the prescribed air exchanges needed to deliver adequate warm or cool air, the minimum air change requirements set by ISO regulations (ISO-7547, 2002), and any specifications required by the owner.

$$N_{ACH,i} = max\{N_{ACH-Heat}, N_{ACH-ISO}, N_{ACH-Feadship}\}_i \tag{4.15}$$

Initially, the heat gain that has been previously calculated needs to be counterbalanced by the supplied air. In the case of cooling, the inflow temperature of the air supplied should not be more than 10 $°C$ lower than the average temperature, whereas during heating, the temperature difference should be limited to a maximum of 23 $°C$ ISO-7547 (2002). By incorporating air density (ρ_{air}), the specific heat constant (c_p), and a maximum area temperature difference parameter (ΔT), the total sensible heat gain of equation 4.4 can be converted into the number of ACH due to heat $N_{ACH-Heat}$ and a supply air flow rate $\dot{V}_{Supplyair-heat}$ in m^3/h.

$$N_{ACH-Heat} = \left(\frac{\dot{V}_{Supplyair-heat}}{V_{Room}}\right)_i = \left(\frac{Q_{sensible}}{\rho_{air} c_{p,a} \Delta T V_{room}}\right)_i \tag{4.16}$$

The volume of the supply can be calculated with the number of air changes and the volume of the specific area.

$$N_{ACH,i} = \left(\frac{V_{Supplyair}}{V_{Room}}\right)_i \tag{4.17}$$

3. Fresh air mixing ratio

With the number of air changes it is now possible to predict the proportion of fresh air (FA) in the supplied air. The maximum value must be selected to meet the maximum CO_2 level, comply with the ISO standard, and satisfy the owner's specifications.

$$FA_i = max\{FA_{CO_2}, FA_{ISO}, FA_{Feadship}\} \tag{4.18}$$

The CO_2-level can only be controlled by adjusting the proportion of fresh air inflow. The CO_2-level increases due to people in the area. The maximum allowable level of CO_2 in the room is 0.1%. Fresh air contains 0.03-0.04% of CO_2 (Klein Woud and Stapersma, 2016). The minimum required proportion of FA can be calculated with the new supplied air volume $\dot{V}_{\text{SupplyAir}}$.

$$FA_{CO_2,i} = \frac{\dot{V}_{FA_{CO_2},i}}{\dot{V}_{\text{SupplyAir},i}} = \frac{\dot{V}_{CO_2 2,i}}{(y_{CO_2,area} - y_{CO_2,FA})\,\dot{V}_{SupplyAir,i}} \tag{4.19}$$

The ISO regulations specify a minimum FA proportion of 0.008 m^3/s per person for which the space is designed that must be met in all conditions ISO-7547 (2002).

$$FA_{ISO,i} = \frac{0.008\,(n_{\text{People},i})}{\dot{V}_{\text{SupplyAir},i}} \tag{4.20}$$

The owner's specification usually sets the highest value. For example, Feadship (2015) sets a minimum of 70 % of fresh air in guest and even higher in dining rooms and the kitchen.

4. Intermediate HVAC conditions

Once the heat gain, air exchanges, and fresh air proportions are established, the power requirements for the intermediate HVAC conditions can be derived. A methodology outlined by Klein Woud and Stapersma (2016) is used for the calculations of the intermediate conditions. The thermodynamic properties depend on the heating, cooling, or humidification processes in the air flow cycle. The calculation of the different processes are given in order of calculation.

To assess these thermodynamic properties effectively, a Mollier diagram, also known as the enthalpy-entropy diagram, is employed. The Mollier diagram provides a graphical representation of the thermodynamic state of air and allows for a straightforward analysis of the intermediate air conditions, starting from the inlet air and tracing through to the exhaust air conditions.

The temperature (T), relative humidity (RH), enthalpy (h), and absolute humidity (x) are of importance for the use of the Mollier diagram. By knowing any two of these properties, the diagram enables the precise determination of the remaining two properties.

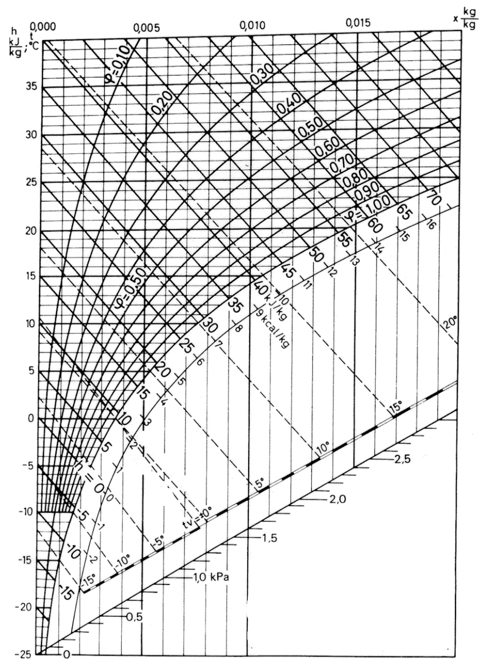


Figure 4.16: Mollier diagram of humid air (Engineering Toolbox, 2003)

Enthalpy determination

The temperature (T) and relative humidity (RH) are monitored in this process by Feadship and stored in the voyage data. These properties are used to calculate the enthalpy at a time step and is outlined by Klein Woud and Stapersma (2016). The fresh air enthalpy and the supply air enthalpy can both be determined with the Mollier diagram.

$$\Delta h = h_{FreshAir} - h_{SupplyAir} \tag{4.21}$$

Total Thermal Heat

The total thermal demand can be calculated with the mass flow and the enthalpy difference between the conditions before and after the Air Handling Unit.

$$Q_{Total} = \sum \left(Q_{sensible,i} + \frac{\dot{V}_{SupplyAir,i} \cdot \rho_{air} \cdot \Delta h}{3600} \right) \tag{4.22}$$

4.5. Grey Box & Black Box

The output of the white box model is used as input for the grey box model. To compare the performance of the grey box model, a solo black box model is also constructed. Both these models make use of an Artificial Neural Network. In the following subsections, the basic princples of an ANN will be discussed along with the Hyperparameter determination and performance evaluation of ANNs.

4.5.1. ANN Basic Principles

Géron (2023) describes Artificial Neural Networks as a foundational component of modern artificial intelligence and machine learning. ANNs are inspired by biological neurons, which are mostly found in animal brains. The main use of these models is to approximate functions that can depend on a large number of inputs that are unknown and complex. The basic principle of an ANN revolves around its structure, which consists of layers of interconnected nodes called neurons. The inputs and output of these neurons are numbers, and each input connection is associated with a weight. This type of ANN is called a Perceptron network, which was invented already in 1957 by Frank Rosenblatt (Géron, 2023). Each neuron is a simple processor that performs a weighted sum of its inputs and passes the result through an activation function. The layers of neurons include an input layer, one or more hidden layers, and an output layer as depicted in figure 4.17.

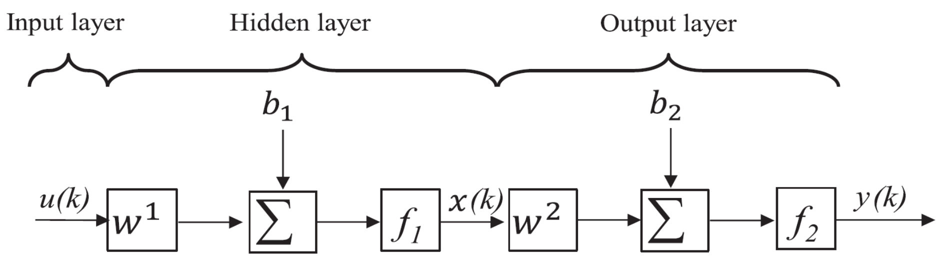


Figure 4.17: Multilayer perceptron network (Nasruddin et al., 2019).

The ANN used by Nasruddin et al. (2019) in his research on energy demand for residential buildings is a multilayer perceptron network (MLP). It consists of:

Hidden Layer Output $x(k)$:
The output from the hidden layer at time step k is given by:

$$x(k) = f_1(w_1 \cdot u(k) + b_1) \tag{4.23}$$

Here, f_1 represents the transfer function of the hidden layer, w_1 is the weight matrix connecting the input layer to the hidden layer, $u(k)$ is the input vector at time step k, and b_1 is the bias in the hidden layer.

Output Layer $y(k)$:
The output of the network at time step k is computed as:

$$y(k) = f_2(w_2 \cdot x(k) + b_2) \tag{4.24}$$

In this equation, f_2 is the transfer function of the output layer, w_2 is the weight matrix connecting the hidden layer to the output layer, $x(k)$ is the output vector from the hidden layer, and b_2 is the bias in the output layer.

The learning process in an ANN involves adjusting the weights w_i to minimise the difference between the predicted output and the actual output. This process is typically carried out using a backwards propagation algorithm combined with an optimisation method such as the Adaptive Moment Estimation (Adam). According to Géron (2023), the Adam Optimiser is considered the preferred optimiser for many researchers due to its ability to handle sparse gradients and adapt its learning rate for different parameters.

The weight update formula in the Adam optimizer is expressed as follows:

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon}\hat{m}_t \tag{4.25}$$

- w_t and w_{t+1} are the current and updated weights.
- η is the learning rate, which influences the size of the steps taken in the weight update.
- $\hat{m}_t$ is the first moment estimate, essentially a decayed average of past gradients. It helps accelerate the updates in the correct direction.
- $\hat{v}_t$ is the second moment estimate, a decayed average of past squared gradients. It adjusts the learning rate dynamically for each parameter.
- ϵ is a small constant added to prevent division by zero, ensuring numerical stability.

Equation 4.25 adjusts the weights by moving in the direction of the optimised gradient. This scaling adjusts the step size, making it smaller where the gradient varies more, resulting in more stable updates.

4.5.2. Hyperparameters

Now that the principles of the ANN have been explained and defined, the next step is to determine the different hyperparameters to be included in the hyperparameter optimisation. This is a critical part of the model selection process, which includes choosing between different learning algorithms, determining the hyperparameters of a chosen algorithm, and defining the structure of the algorithm itself. The aim is to identify a set of hyperparameters that will produce the most effective neural network. The hyperparameters involved are the number of hidden layers, the number of neurons, the activation functions and the regularisation coefficient. This is a challenging and time-consuming task, as there is no universal configuration that can optimally address all data related problems.

Number of Hidden Layers

In the context of hyperparameter optimisation for ANNs, determining the optimal number of hidden layers is crucial, as it significantly influences the model's ability to learn and generalise from the data. Typically, ANNs start with a single hidden layer, which can effectively capture linear relationships in the data. However, more complex data structures often require additional layers. The choice between a shallow network (fewer hidden layers) and a deep network (more hidden layers) depends on the specific problem, the complexity of the data, and the computational resources available. While deeper networks can model more complex patterns, they are also more prone to overfitting and may require more data and training time to achieve optimal performance. Overfitting occurs when a model learns the noise in the training data rather than generalising from it (Géron, 2023). This problem leads to high performance on the training data but poor performance on unseen data, making the model less effective for practical applications. Figure 4.18 shows on the left what the loss curve looks like while overfitting, with the corresponding fitting relations on the right. Underfitting is the opposite, when the model is too simple and more complexity needs to be added.

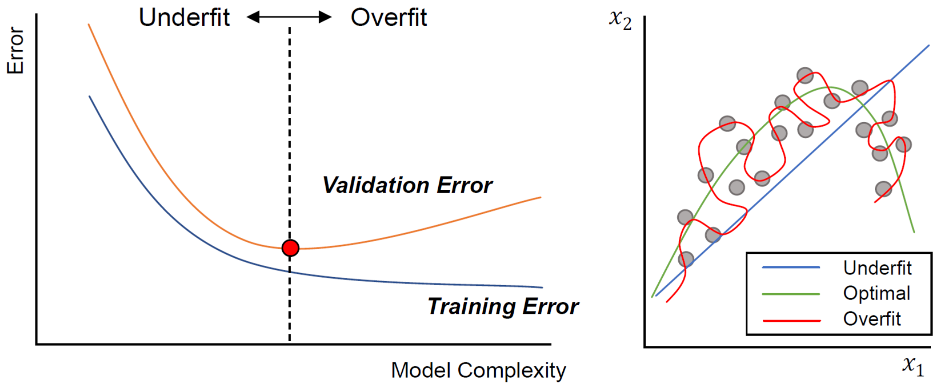


Figure 4.18: Model error comparison (left) and corresponding fitting relations (right), from Odendaal (2021), based on Silva et al. (2017)

Finding the right number of hidden layers is often a trade-off between the depth of the network and its efficiency and generalisation capabilities. This choice can be approached systematically using techniques such as cross-validation, where different configurations are empirically tested to find the best structure, which will be explained further on in this section. Heaton (2015) and Bishop (2006) provide comprehensive overviews of the architectural decisions in neural network design. With the insights provided in the previous discussion and the comparable neural network models for residential buildings explored, the number of hidden layers considered in the hyper-parameter optimisation process includes configurations of 1, 2 and 3 layers.

Number of Neurons per Layer

In the process of designing ANNs, selecting the appropriate number of neurons in each hidden layer is a critical decision that impacts the performance. The number of neurons determines the capacity of the network to learn from complex datasets. If too few neurons are used, the network may not capture all the underlying patterns in the data, resulting in underfitting as shown above in figure 4.18. On the other side, too many neurons can lead again to overfitting. Determining the ideal number of neurons typically involves a trial-and-error approach using the grid search technique. To establish a baseline for ANN configurations, a review of neural networks from previous marine engineering studies is shown in Table 4.3. This review provides an initial understanding of how many neurons per hidden layer are used in these applications.

Table 4.3: Neural network architectures in residential HVAC and marine engineering applications

Reference	Input Features	Hidden Layers	Neurons	Output Targets
Nasruddin et al. (2019)	10	1	3	2
Kalogirou et al. (2001)	12	3	18 → 18 → 18	1
Odendaal (2021)	5	2	30 → 40	1
Zwart (2020)	13	1	15	1
Pedersen and Larsen (2009)	9	1	12	1
Parkes et al. (2018)	6	3	50 → 50 → 50	1

Regularisation Techniques

To mitigate overfitting, regularisation techniques are effective measures that still allow a model to perform well (Silva et al., 2017). One effective strategy is to use early stopping during training. Early stopping involves monitoring the model's performance on a validation set and stopping the training process when performance begins to deteriorate or fails to improve significantly. For example, in Figure 4.18, the training of the model would be stopped at the dotted line. This technique not only helps prevent overfitting, but also optimises training time, making the process more efficient.

Another technique, called dropout, is a widely used regularisation method that randomly deactivates a subset of neurons during training (Chollet, 2017). According to Géron (2023), the dropout rate is typically set between 10% and 50%. If the rate is set to 20%, the model will drop an average of 20% of neurons during training. This temporarily simplifies the network and prevents the model from becoming too dependent on one neuron. This random disappearance of input features during training iterations encourages the network to develop more robust features that are useful in combination with many different random subsets of the other neurons. In Figure 4.19, it is depicted how this works. Together, dropout and early stopping are powerful tools for improving the generalisation ability of ANNs.

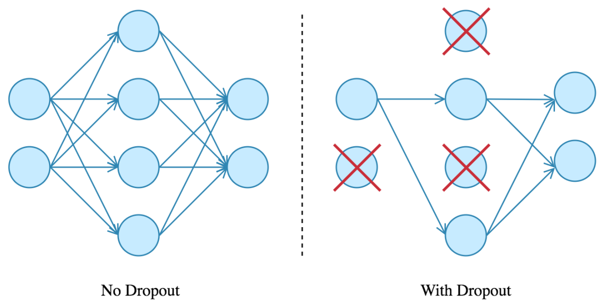


Figure 4.19: Dropout technique according to Chollet (2017)

Activation Function

In ANNs, activation functions are crucial for introducing non-linearity into the model, allowing it to learn complex patterns in the data. Two common activation functions are the rectified linear unit (ReLU) and the Sigmoid function (Bishop, 2006). The ReLu is favoured in many neural network architectures due to its computational efficiency and its ability to reduce the likelihood of the vanishing gradient problem when the gradient of a neuron becomes extremely small (Chollet, 2017).

Unlike the ReLu function, the Sigmoid function maps the input values to the range $(0, 1)$. Both functions have different characteristics that make them suitable for different types of neural network layers (Géron, 2023).

- Rectified Linear Unit (ReLU):
$$f(z) = max(0, z) \tag{4.26}$$
- Sigmoid Function:
$$f(z) = \frac{1}{1 + e^{-z}} \tag{4.27}$$

The sigmoid function is used in several successful researches with ANNs on the energy demand of HVAC (Moayedi et al., 2019) (Nasruddin et al., 2019). The function helps to introduce complexity and non-linearity, allowing the MLP to understand more complicated patterns. Figure 4.20 shows what the two functions look like and their derivatives. The derivative of the activation function is a key part of ANN training. It shows how much a change in input weight will affect the change in output. In other words, it shows how the weights are updated during training.

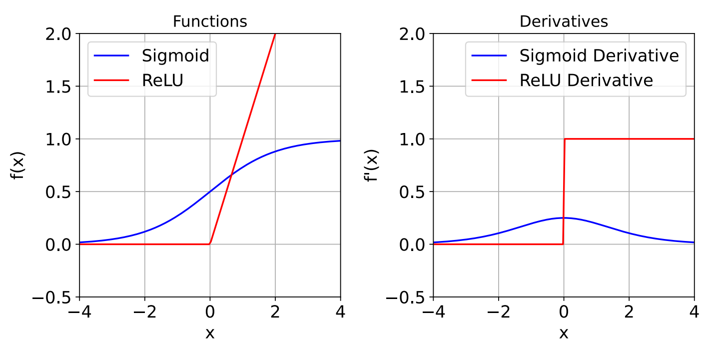


Figure 4.20: Sigmoid and ReLu Activation Functions and their derivative

Selected Hyperparameters

A summary of the selected hyperparameters for the optimisation is given in Table 4.4. These parameters are chosen based on the information above and earlier performed researches (Table 4.3). A grid search method (Géron, 2023), will be performed to find the best configuration.

Table 4.4: Selected Hyperparameters

Hyperparameter	Selected inputs
Input Features	10
Number of hidden layers	[1 2 3]
Number of neurons	[5 - 60]
Activation Functions	[Sigmoid ReLu]
Dropout	[0 0.1 0.2 0.3]
Optimizer	Adam
Output variables	1

4.5.3. Performance Evaluation

The performance evaluation of an ANN can be done by the holdout validation (Géron, 2023). In this technique, the data set is divided into two distinct parts: a training set and a test set. Typically, a larger portion of the data is used for training and the remaining is used as the test set. The ANN is trained on the training set only and then evaluated on the test set. This approach provides a straightforward assessment of how well the model performs on unseen data. A schematic of this method is shown in Figure 4.21.

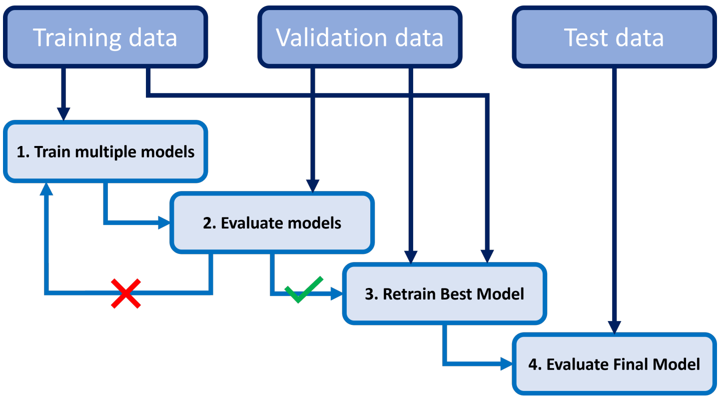


Figure 4.21: Model selection using holdout validation, based on Géron (2023)

Another method, called k-fold cross-validation, extends the holdout method to provide a more comprehensive evaluation. A k-fold cross-validation iteratively tests the model on different subsets, as shown in Figure 4.22. This reduces the potential for overfitting and increases the validation of the model's predictive power (Kohavi, 2001). The k-fold cross-validation is chosen since it randomises the data opposed to a cross validation on a time series basis (Shrivastava, 2020). This will make sure the model is trained on different climates of different locations of the yacht.

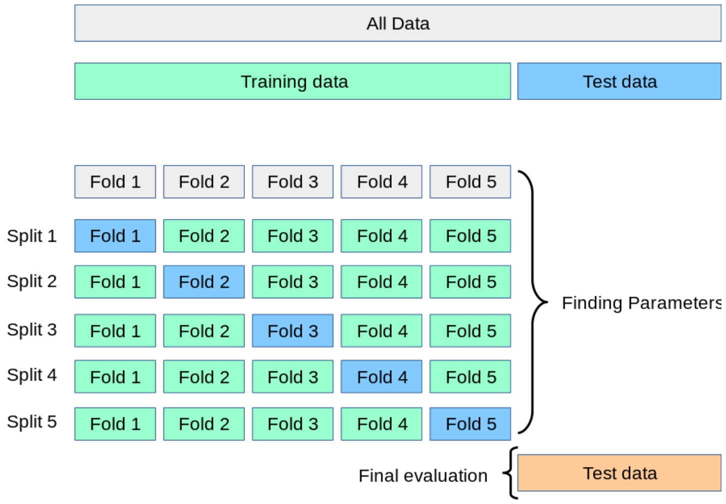


Figure 4.22: A k-fold cross-validation with k=5 (scikit-learn developers, 2024)

According to Botchkarev (2019), several key metrics are crucial in assessing model performance, including mean square error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The MSE measures the average of the squares of the errors, which is the average squared difference between the estimated values and the actual value. This metric gives a sense of how far off the predictions are from the actual values. The MAE measures the average size of the errors in a set of predictions, regardless if it is a positivie or negative error. It is calculated as the average over the test sample of the absolute differences between prediction and actual observation, with all individual differences given equal weight. The Mean Error (ME), calculates the average of all the prediction errors, giving an indication of the overall bias in the prediction whether they tend to be overestimates or underestimates. Lastly, the MAPE expresses accuracy as a percentage and is particularly useful in contexts where it is desirable to compare the performance of prediction models across different data scales. This simplicity and direct interpretation often make MAPE, along with the MAE, a preferred metric.

- Mean Squared Error:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{4.28}$$

- Mean Absolute Error:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{4.29}$$

- Mean Error:

$$\text{ME} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i) \tag{4.30}$$

- Mean Absolute Percentage Error:

$$\text{MAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{4.31}$$

5 Results and Validation

In the previous chapters the methodology is explained and the model is developed. In this chapter the research sub-question *How can the accuracy of the model be validated?* will be answered along with the results. The results of the data preparation are discussed in section 5.1. Secondly, in section 5.2 the power calculations needed for the validation are given. After this, the white box model results are displayed and validated in 5.3. In the end the results of the grey box and black box model are discussed in 5.5.

5.1. Data Preparation Results

In table 5.1 the results of the data preparation are shown per variable. The AC data is only shown from one AC unit, since there are five AC units on the yacht. The weather data from the ECWMF database has a different time step, one per hour instead of one per three minutes, so fewer datapoints and is interpolated to the power as mentioned in 4.3. Noteworthy are the duplicates in the solar radiation data, which are due to the double time stamp in the data retrieval.

The data is eventually selected on the same time frame and interpolated towards the chiller data. This data is the most representive of the power signal and is used for the validation. Every data point gathered is valuable, in order to preserve as much information as possible. However, it is not always possible to have complete datasets when numerous data points are missing. The data is only interpolated to fill missing values if the two surrounding data points are present.

Table 5.1: Data preparation results

Variables	Remove duplicates		Noise Identification		Data Integration	
	Removed	Remaining	Removed	Remaining	Removed	Remaining
AC Power	0	131172	35	131137	96604	34533
Chiller Power	0	69846	7466	68691	32824	35867
Fresh Air Temperature	0	131172	0	131172	96638	34534
Fresh Air Humidity	0	5378	94	5284	4724	560
Return Air Temperature	0	131172	194	130978	96444	34534
Supply Air Temperature	0	131172	378	130965	96431	34534
Return Air Humidity	0	131172	1323	128717	94183	34534
Supply Air Humidity	0	131172	3745	129691	95157	34534
DHI	7466	10006	520	9486	7862	1624
BNI	7466	10006	296	9710	7568	1602
Longitude	0	89296	0	89296	54808	34488
Latitude	0	89296	0	89296	54808	34488
Heading	0	89336	0	89336	54808	34488

5.2. Power calculation from voyage data

To assess the overall thermal energy in relation to the electrical demand of the HVAC system, an evaluation of the power data must be conducted.

The dataset includes:

1. **Electrical signal for each AC system:** The power is calculated by equation 5.1. The factor 0.9 is the power factor $cos(\phi)$, I_1 to I_3 are the different line currents and U is the line voltage. This includes the cumulative power usage of the heater, humidifier, fans, and reheaters. This equation is given by the external contractor that suppllies the data of the HVAC system and it complies with information provided by Sen (2013).

$$P_{AC} = 0.9 * (I_1 + I_2 + I_3) * \frac{U}{\sqrt{3}} \tag{5.1}$$

2. **Power signal for the chiller:** This represents the total power consumed by the chiller. This signal gives the total current in Amperes. The factor 0.9 is the power factor $cos(\phi)$, I_1 to I_3 are the different line currents and U is the line voltage.

$$P_{Chiller} = 0.9 * I_{total} * U \tag{5.2}$$

3. **Power signal for reheaters:** This signal gives the power of the reheaters as a percentage. Equation 5.3 gives the power usage of each individual reheater. $Pmax$ is the maximum power of the reheater. This is obtained from the manufacturer.

$$P_{reheater} = \%_{reheater\ power\ level} * P_{max} \tag{5.3}$$

4. **Power signal for fans:** This accounts for the power consumed by each fan, including both exhaust and supply fans. $Pmax$ is the maximum power of the fan. This is also obtained from the manufacturer of the fans.

$$P_{fan} = \%_{fan\ power\ level} * P_{max} \tag{5.4}$$

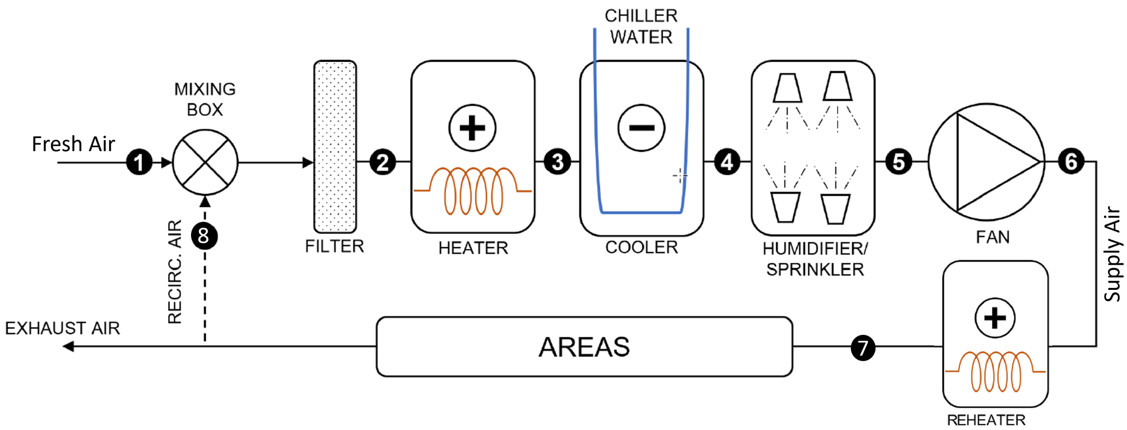


Figure 5.1: HVAC system schematic of the yacht from which the sensor data is obtained.

The heater, humidifier, and chiller are the components responsible for mitigating the total thermal heat generated in the areas that are supplied by the HVAC system. The total system can be seen 5.1 Since the power data is recorded of all the systems together, the total power consumed by these components can be calculated by equation 5.5

$$P_{HVAC} = P_{chiller} + (P_{heater} + P_{humidifier}) = P_{chiller} + (P_{AC1-5} - P_{fan} - P_{reheater}) \tag{5.5}$$

5.3. White Box Results

Figure 5.2 displays the thermal load calculated by the model versus the actual power consumption of the system. A certain period is chosen to make the data more visible. Appendix A show the results of the whole data set. The top subplot illustrates two datasets: the calculated heat load (from the model) and the measured power consumption of the HVAC system. The calculated heat load is depicted with a blue line, providing insights into expected thermal power requirements based on the white box calculation. The measured HVAC power is represented by the orange line, showing the actual power consumption data recorded from the system derived from the calculation in section 5.2.

In the lower subplot, two environmental parameters are plotted: the average fresh air temperature (FAT) and the fresh air relative humidity (FAH), each relevant to the air conditioning system's performance. The FAT data is illustrated with a green line, reflecting the ambient temperature conditions, which are critical in the calculation of the HVAC system's thermal load. The relative humidity data, is plotted on a secondary y-axis to the right. This variable offers additional context for the HVAC system's operation, as humidity levels significantly impact the system performance. Both the temperature and humidity data are crucial for correlating environmental conditions with the power and heat load requirements, providing a comprehensive view of the system's operation under varying atmospheric conditions.

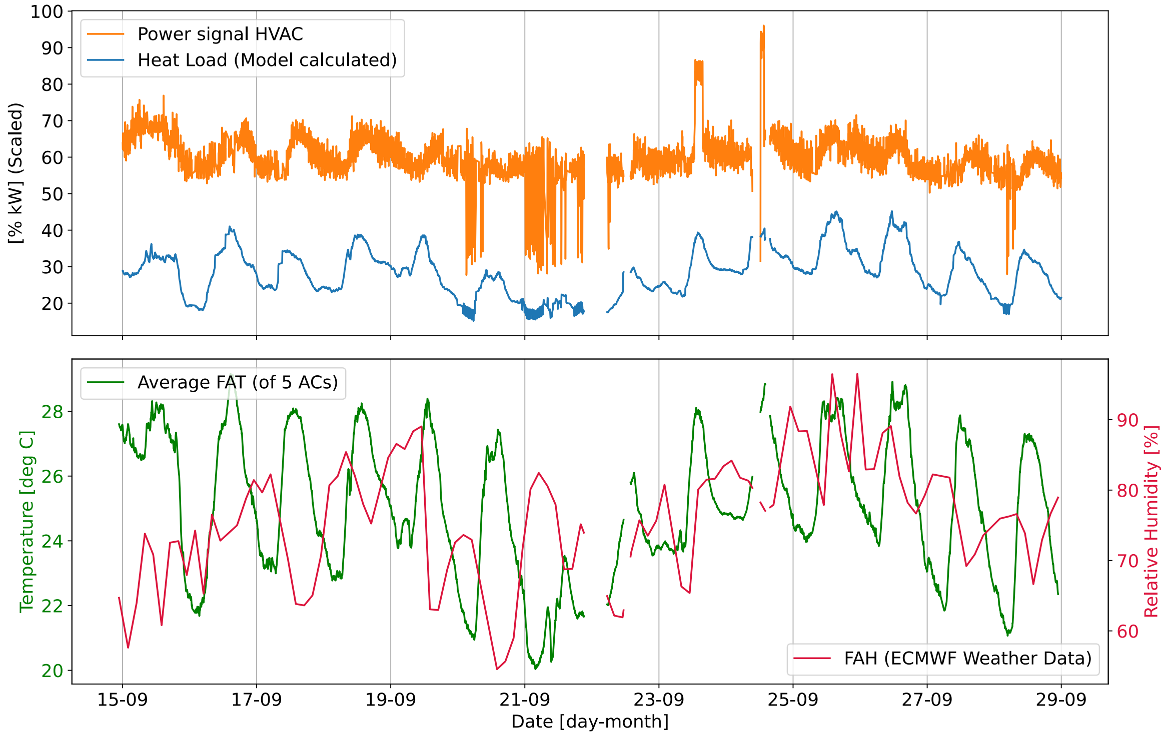


Figure 5.2: Example room from the General Arrangement spreadsheet

The plot depicts data over a 14-day period, several observations can be made. The average fresh air temperature represented by the green line shows the day cycles of the eleven days and are clearly distinguishable by the eleven peaks in the temperature. The fresh air humidity in red show the same variability. The dips in humidity don't always correspond to the peaks in temperature, suggesting that other factors may be influencing humidity levels, or that the relationship between temperature and humidity is not straightforward. In general, the relationship between temperature and relative humidity in the atmosphere is inversely related, but relative humidity is not solely determined by temperature. Relative humidity is the amount of moisture in the air relative to what the air can hold at that temperature. As the temperature increases, the air can hold more moisture, so if the amount of moisture stays the same while the temperature rises, the relative humidity will decrease. However, specific local weather patterns like rain, time of day, and geographic factors can affect this relationship.

At certain points, changes in the temperature correspond with changes in the HVAC power usage. For example, a peak in temperature is often matched by an increase in power usage, since the HVAC system works harder to cool the environment when external temperatures are higher. The total thermal load seems less reactive to the higher peaks in temperature compared to the actual power consumption. An example being the difference in the trough and peak of the heat load compared to the power signal before 17-09. Which would mean that at higher temperatures the system has a higher efficiency and better COP.

5.4. White box Validation

To be able to validate the total heat load of the yacht calculated by the white box model, the efficiencies of the system need to be explored. It is hard to estimate the individual efficiencies within the the total heat load, since it is overcome by the chiller, heater and humidifier together and it is not clear what percentage is delivered by the individual systems. From the power side however is more data available per individual system. The actual total heat load can be calculated using equation 5.6.

$$Q_{total} = P_{chiller} \cdot COP_{chiller} \cdot \eta_{hex} + P_{heater} \cdot \eta_{heat} + P_{humidifier} \cdot \eta_{hum} \tag{5.6}$$

The coefficient of performance (COP) is a value representing the extracted heat divided by the systems net work delivered. The COP is derived using the chiller's data. This includes monitoring the return temperature of the chilled water and its supply temperature. Additionally, the variable frequency drive VFD for the two installed chilled water pumps is recorded as a percentage and the mass flow is calculated with equation 5.7. $\dot{m}_{max}$ is part of the information given by the supplier of the Chiller System. It also states that the coolwater contains 10 % Glycol, which has a c_p of 4079 J/kgK (Toolbox, 2023). Using this information, the COP can be determined through the following formulas according ASHRAE (2013).

$$\dot{m} = VFD \cdot \dot{m}_{max} \tag{5.7}$$

$$Q_{cw} = \dot{m} * c_p * (T_{return} - T_{supply}) \tag{5.8}$$

$$COP = \frac{Q_{cw}}{P_{chiller}} \tag{5.9}$$

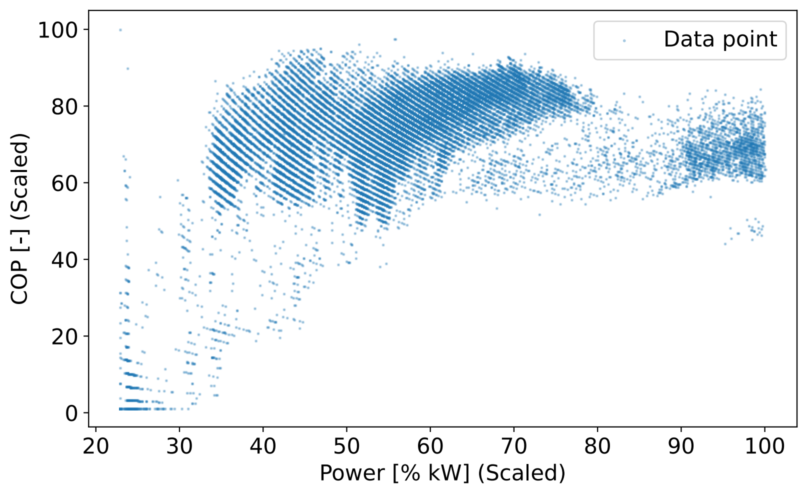


Figure 5.3: Chiller power vs. COP

The intermediary efficiencies and coefficient of performance (COP) of the chiller for the associated HVAC components are detailed in Table 5.2. It is important to acknowledge that these values are largely estimative and may differ from actual operational outcomes. Therefore, the presented efficiencies are only suitable for preliminary design phases, where a higher level of variability and uncertainty is expected.

Efficiency or COP	Symbol	Value	Source
Chiller COP	$COP_{chiller}$	-	Equation 5.9
Heater efficiency	η_{heat}	80%	Klein Woud and Stapersma (2016)
Heat exchanger efficiency	η_{hex}	80%	Klein Woud and Stapersma (2016)
Humidifier pump efficiency	η_{hum}	80%	Klein Woud and Stapersma (2016)

Table 5.2: Efficiency and COP of HVAC system components.

The result is shown in Figure 5.4 and 5.5. The heat load calculated by the model follows the trend of the actual heat load but at some points overestimates the peaks and troughs. This is mainly due to the relative humidity as shown in 5.2. The calculated heat load reacts strongly to peaks in the relative humidity. This relative humidity comes from a weather database instead of the voyage data. This can explain the error in the calculation.

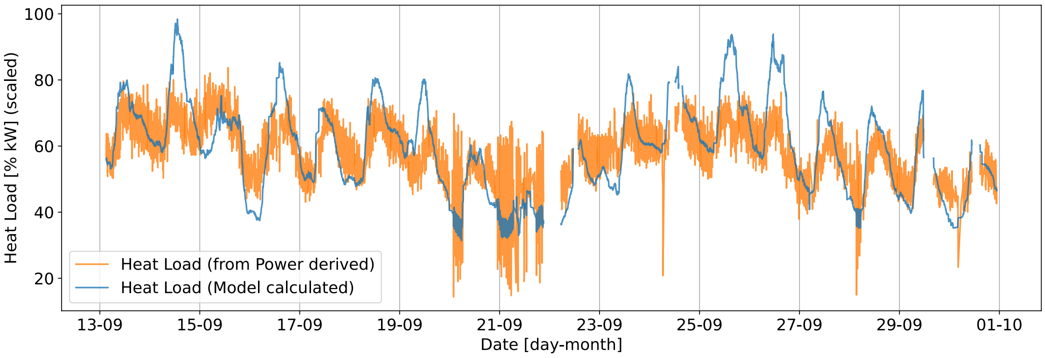


Figure 5.4: Heat Load comparison, 13-09 to 01-10

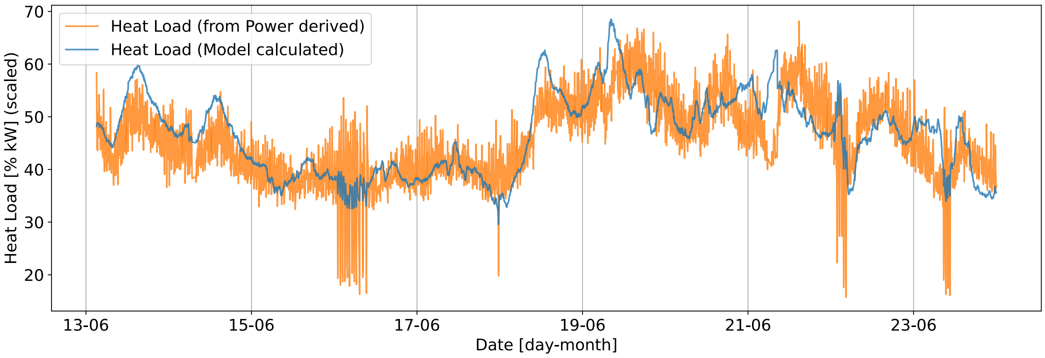


Figure 5.5: Heat Load comparison, 13-06 to 24-06

Figure 5.6 shows that the calculated heat load is lagging and it shows a trough or peak a few data points later then the heat load that is derived from the power. This is mainly due to the fact that the calculated heat load is also based on the supply air temperature. This causes a lag in the calculated data. This is clearly visible if the power of the HVAC system cycles between a high and a low point. This is why this is also visible in the heat load.

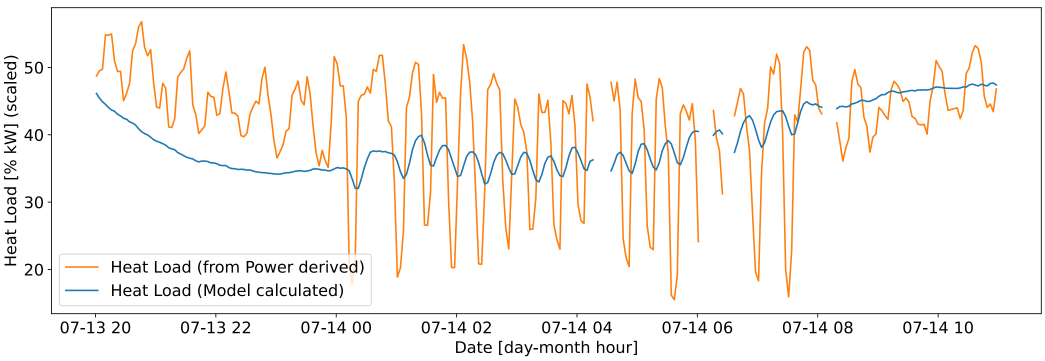


Figure 5.6: Unshifted data

Figure 5.7 shows the data when the calculated heat load is shifted 6 minutes (2 data points) earlier to match the troughs and peaks.

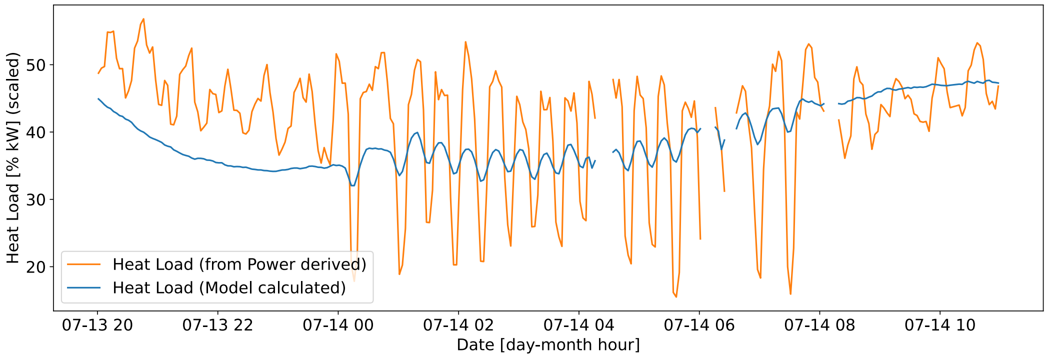


Figure 5.7: Data shifted for better fit.

The heat load derived from the power shows a high variability. A Weighted Moving Average (WMA) function can smooth out the noise created by the power of the HVAC system.

Table 5.3 presents the performance metrics across the different datasets. The comparison reveals that the datasets with shifted data, and those further processed with a WMA, show improved error values relative to the original dataset. Specifically, the dataset with just the shifted data marginally outperforms the original dataset in terms of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). However, the dataset that combines shifted data with the application of a WMA stands out with the best result. It demonstrates the lowest MAE, RMSE, and MAPE values, thereby indicating the highest level of predictive accuracy and optimal model fit compared to the other datasets.

Table 5.3: Comparison of White Box Model Performance Metrics

	MAE [kW]	RMSE [-]	MAPE [%]
Original data	10.09	14.24	15.31
Shifted data	9.86	13.81	14.84
Shifted and WMA applied data	8.31	11.59	11.38

Figures 5.8, 5.10 and 5.12 show the predicted values versus the actual values of the heat load for the three types of data. Figure 5.9, 5.11 and 5.13 show the distribution of the MAPE for the three types of data. The shifted data shows a small change in the position of the data points. This is mainly noticeable in the area around (30, 60) $\%kW$ in Figure 5.8 and 5.10. These data points are also visible in Figure 5.5, where the heat load from power derived shows a high variance around 16-06, with very low values. This is where the heat load calculated by the model overestimates the values as seen in the comparison figures. The WMA function cancels out this high variance caused by the power of the HVAC system and shows that this results in a better prediction.

The shifted-WMA data still shows a bigger error above 100 $\%kW$ of the predicted values. In figure 5.14 this overestimation of the white box model more obvious. It mainly happens when the relative humidity and the temperature of the fresh air are both at a high point. The temperature seems to have the most influence, but the model overestimates the calculated heat load when both parameters are at a peak. This error is most likely caused by the different data source of the fresh air humidity. The values out of the sensor data of this variable were unusable and were taken from the ECMWF (2023) weather database because of this. This data has one data point every hour and is significantly less precise.

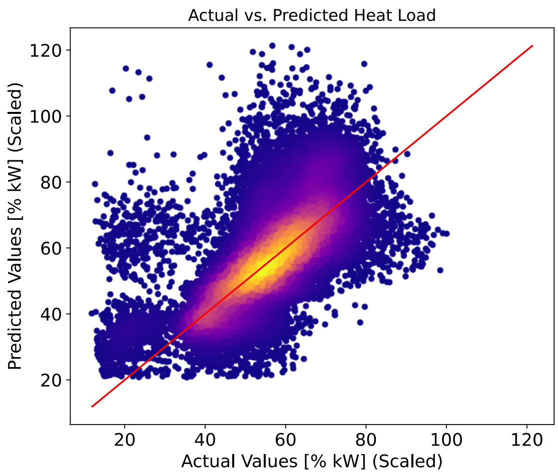


Figure 5.8: Heat load prediction comparison (Original data)

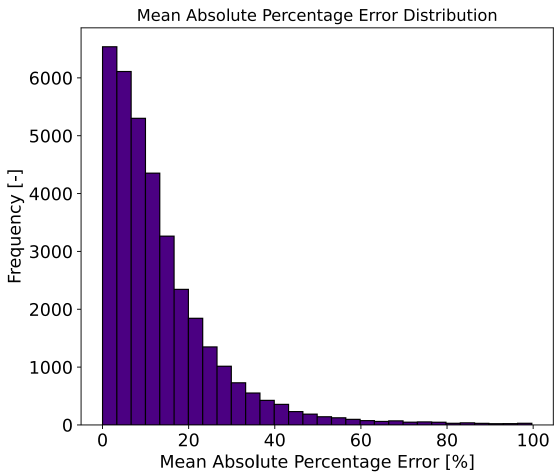


Figure 5.9: Error distribution of MAPE (Original data)

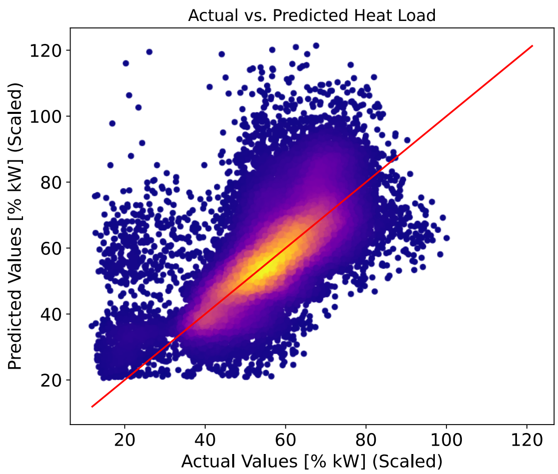


Figure 5.10: Heat load prediction comparison (Shifted data)

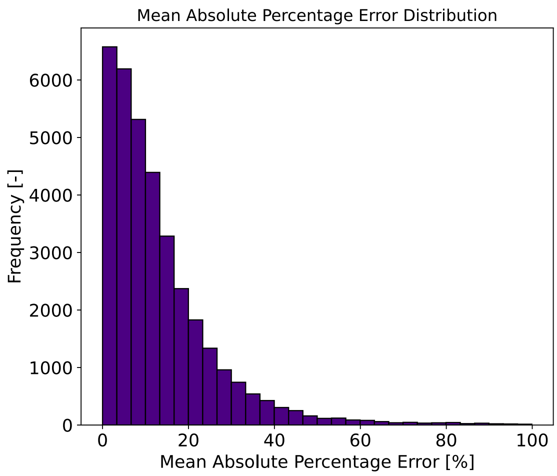


Figure 5.11: Error distribution of MAPE (Shifted data)

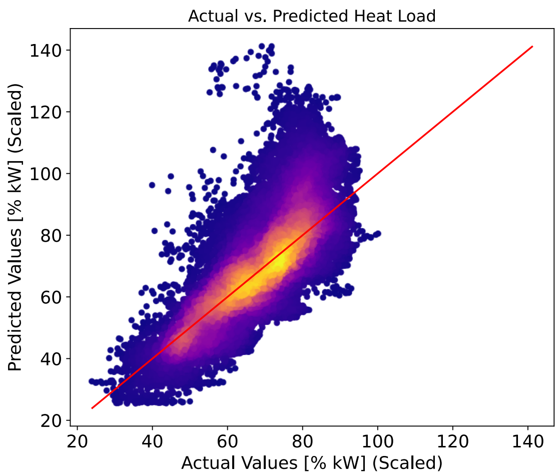


Figure 5.12: Heat load prediction comparison (Shifted-WMA data)

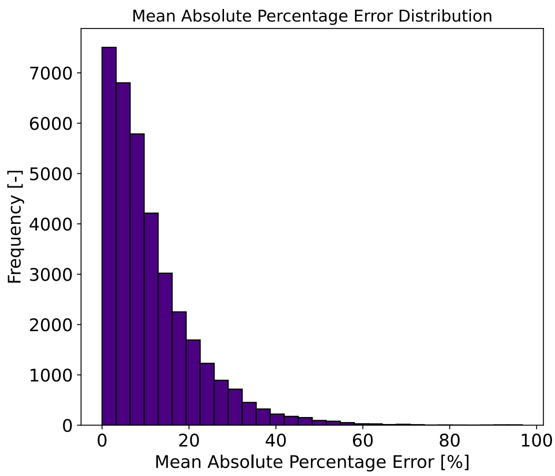


Figure 5.13: Error distribution of MAPE (Shifted-WMA data)

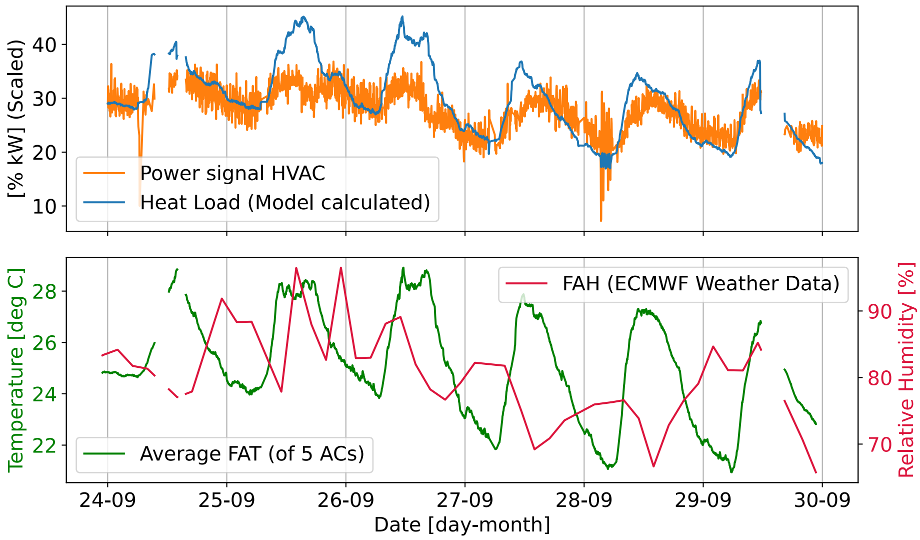


Figure 5.14: Predictions made by the white box model displayed on time series basis.

5.5. Grey Box Results

As mentioned in section 4.5.2, a grid search is applied to find the optimum hyperparamaters. As Table 5.4 shows, a total of $3 \cdot 12 \cdot 2 \cdot 4 = 288$ configurations will be tested five times according the kfold validation. As discussed in 4.5.2, the results are evaluated after this based on different error metrics.

Table 5.4: Hyperparameters for grid search

Hyperparameter	Grid search inputs
Number of hidden layers	[1 2 3]
Number of neurons	[5 - 60]
Activation Functions	[Sigmoid ReLu]
Dropout	[0 0.1 0.2 0.3]

After evaluating the different results, the configuration with 2 layers of 20 neurons each and a dropout rate of 0.1, had the best perfomance taking overfitting into account. The final hyperparameters are shown in Figure 5.5, with the error metrics of the test data set that is not used during the kfold cross-validation. It performs with a MAPE of 8.71%. This translates to an accuracy of 91.29%, which suggest that the model's predictions are about 91.29% close to the actual values.

Table 5.5: Final Configuration of the selected Grey Box model

Hyperparameter	Selected
Input Variables	10
Number of hidden layers	2
Number of neurons	20
Activation Functions	ReLu
Dropout rate	0.1
MSE [kW^2]	364.64
MAE [kW]	11.76
ME [kW]	-2.87
MAPE [%]	8.71

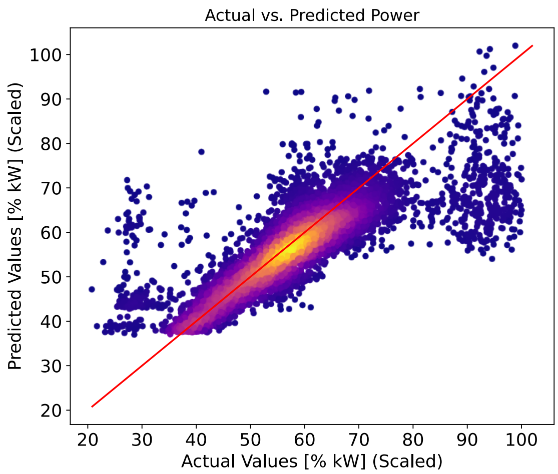


Figure 5.15: Actual vs. Predicted Power Values

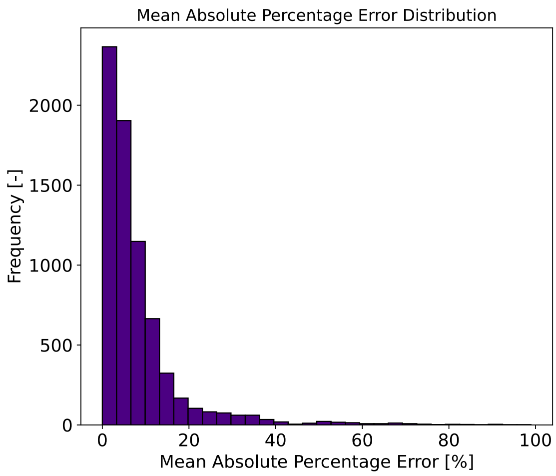


Figure 5.16: Error distribution of MAPE

Figure 5.15 and 5.16 show the comparison of the predicted values to the actual values, and the distribution of the MAPE. There is a visible cloud above 80 $\%kW$ of the actual values. This is where the grey box model underestimates the actual values. In Figure 5.17, these data points are clearly visible in between 05-08 and 01-09. The blue dots represent the actual values and the red cross the predicted values. It shows that the grey box model has trouble predicting the high variance in the data points. This high variance occurs when the second system of the chiller cycles on and off. Within this range, from 01-08 to 01-09, the grey box model struggles to predict values accurately, resulting in a MAPE of 17.03%. Outside of this high variance range, the MAPE is 6.82%. These findings indicate that further research into the chiller system's cycling behaviour is valuable.

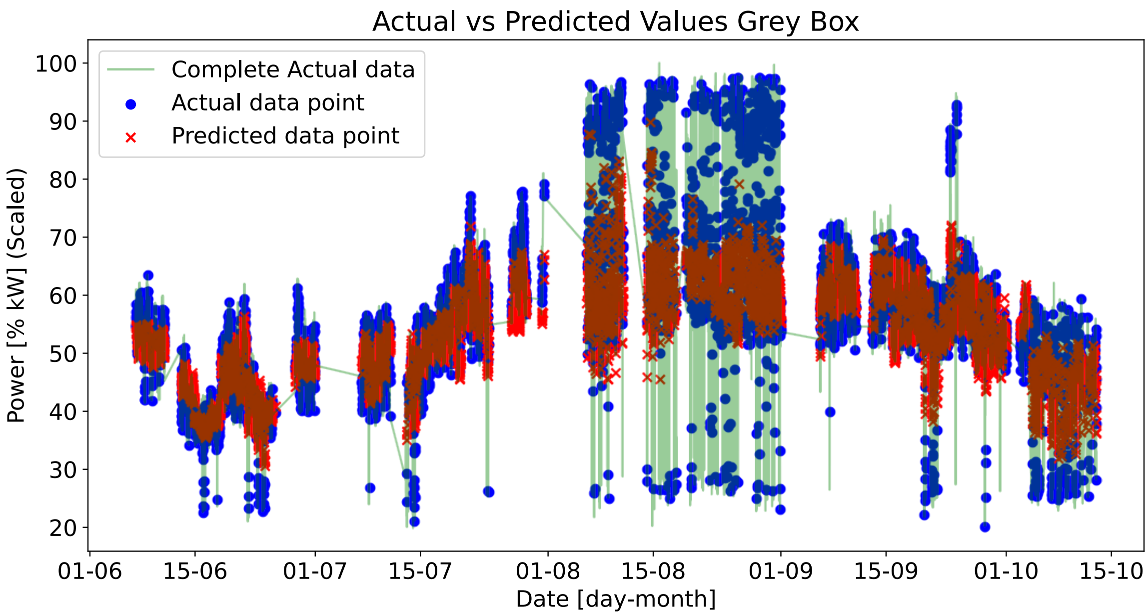


Figure 5.17: Predictions made by the grey box model displayed on time series basis.

5.6. Black Box

A solo black box model is developed in the same way as the grey box model, but without the input heat load data of the white box model. This configuration is done to compare it with the performance of the grey box model and is shown in Figure 5.18. This configuration only uses the following input variables:

- Global Horizontal Irradiance
- Windspeed
- Average Room Setpoint Temperature
- Average Return Air Temperature
- Average Fresh Air Temperature
- Average Mixed Air Temperature
- Average Supply Air Temperature
- Average Return Air Humidity
- Average Supply Air Humidity

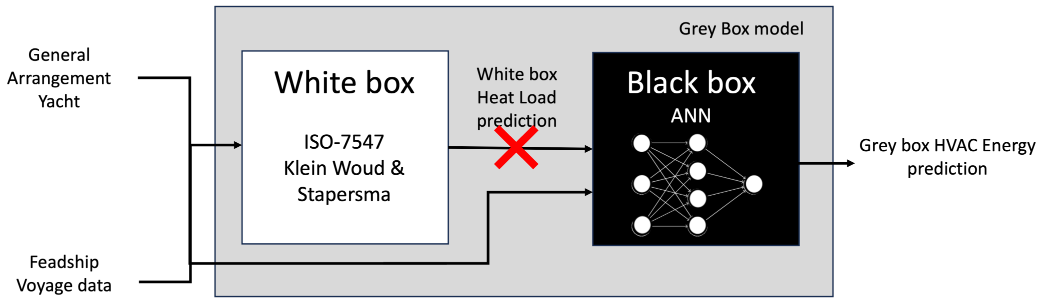


Figure 5.18: Solo black box model.

The model selection process is consistent with the approach of the grey box, to evaluate the performance of black box models. The same grid search is performed on the hyperparameters of Table 5.4. Also the k-fold cross-validation is again applied. This method ensures that both types of models are evaluated under the same conditions, with their best set of hyperparameters for comparison.

After evaluating various configurations, the ANN with two layers of 15 neurons each and a dropout rate of 0.1 performed best across the five validation folds. The chosen hyperparameters are shown in Figure 5.6, along with error metrics on the remaining test set that was not used in the kfold cross-validation. The average MAPE on this test set is 9.58%, which corresponds to a MAPE accuracy of 90.42%.

Table 5.6: Final Configuration of the Black Box model

Hyperparameter	Selected
Input Variables	9
Number of hidden layers	2
Number of neurons	20
Activation Functions	ReLu
Dropout rate	0.1
MSE [kW^2]	424.41
MAE [kW]	13.21
ME [kW]	-4.35
MAPE [%]	9.58

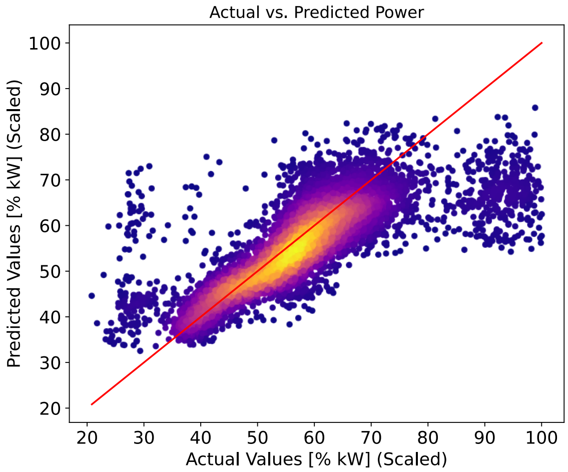


Figure 5.19: Actual vs. Predicted Power Values (Black Box)

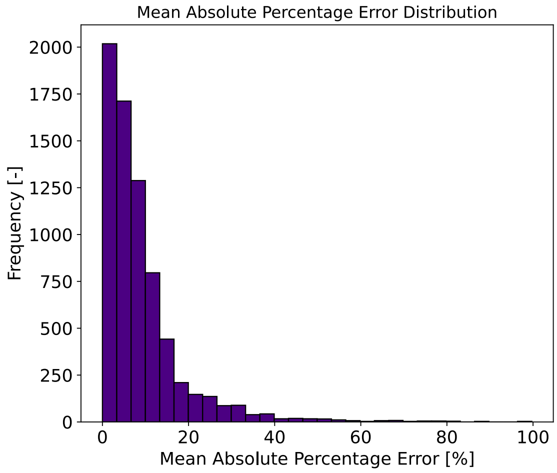


Figure 5.20: Error distribution of MAPE (Black Box)

5.7. Comparison of Black and Grey Box

A comparison between the grey box model and the black box model is done in order to establish if the grey box model is an improvement. Notable differences in performance and configuration can be observed in Table 5.7. Both models have a two-layer configuration and use a consistent dropout rate of 0.1 to mitigate overfitting; however, the grey box model uses 20 neurons per layer compared to the 15 neurons per layer in the black box model. In addition, the black box model has one input variable less than the grey box model, because it lacks the input data of the white box model. The difference in inclusion of the white box data results in a MAPE of 8.71% for the Grey Box model, corresponding to a prediction accuracy of 91.29%, compared to a MAPE of 9.58% and an accuracy of 90.42% for the Black Box model. The grey box model also demonstrates an overall better performance on other key metrics such as MSE, MAE and ME. To enhance the analysis, the performance of the grey box model is shown with plots in the previous sections comparing actual versus predicted values and the distribution of errors, providing a clearer insight into the model's effectiveness and error characteristics. These results show the importance of the input data in improving model performance.

Table 5.7: Comparison of final configurations

Hyperparameter	Grey Box	Black Box
Input Variables	10	9
Number of hidden layers	2	2
Number of neurons	20	15
Activation Functions	ReLu	ReLu
Dropout rate	0.1	0.1
MSE [kW^2]	364.64	424.41
MAE [kW]	11.76	13.21
ME [kW]	-2.87	-4.35
MAPE [%]	8.71	9.58

6

Model Application

In this chapter the following research sub-question will be answered: *How can the proposed model be implemented in the design of new future yachts, and what criteria must be met to consider it to be successful?*. The final grey box model can be summarised as shown in Figure 6.1. In this chapter the application of the grey box model is detailed, but also the applicability of the solo white box and solo black box model are explored.

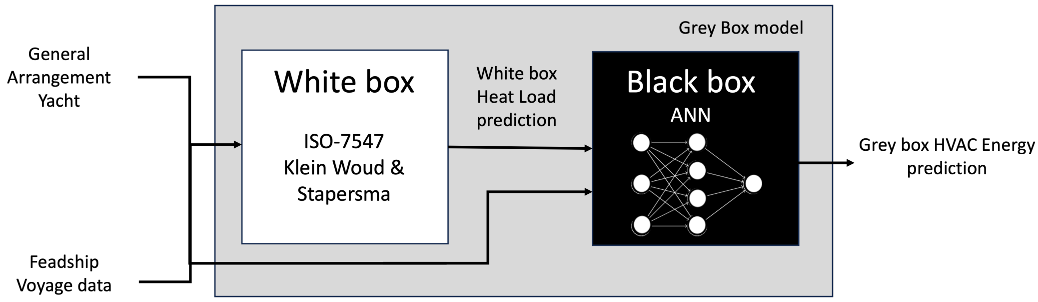


Figure 6.1: Final Grey Box Model

6.1. Application White box

As shown in 5.7, the grey box model outperforms the solo black box model by a small margin across different error metrics. The question can be asked whether this small margin is worth the large computational cost of the white box model. However, the white box model is valuable on its own. If a yacht does not yet have voyage data, or the desired output from the model is outside the range of trained data, a full white box approach is still possible to estimate the heat load of the yacht.

Figure 6.2 shows the relationship between temperature on the x-axis and heat load on the y-axis. Two data sets are shown: The heat load calculated by the white box model and the heat load derived from the power consumption as described in section 5.2. The distribution of data points shows the correlation between temperature and heat load, with an increase in heat load as temperature rises. The loadlist calculation by the external contractor is also shown in the figure, derived from the power using the same principles to get the heat load. The mean COP at certain temperatures is used for this. It shows that the actual environmental conditions have a significantly lower demand than the maximum loadlist calculation by the external contractor.

This data can help engineers decide what size of HVAC system is needed and what heat load it must deliver at different temperatures. Understanding this relationship is critical to designing and operating HVAC systems more efficiently. It can inform energy management strategies to optimise energy use against predicted thermal loads, resulting in cost savings and reduced energy demand.

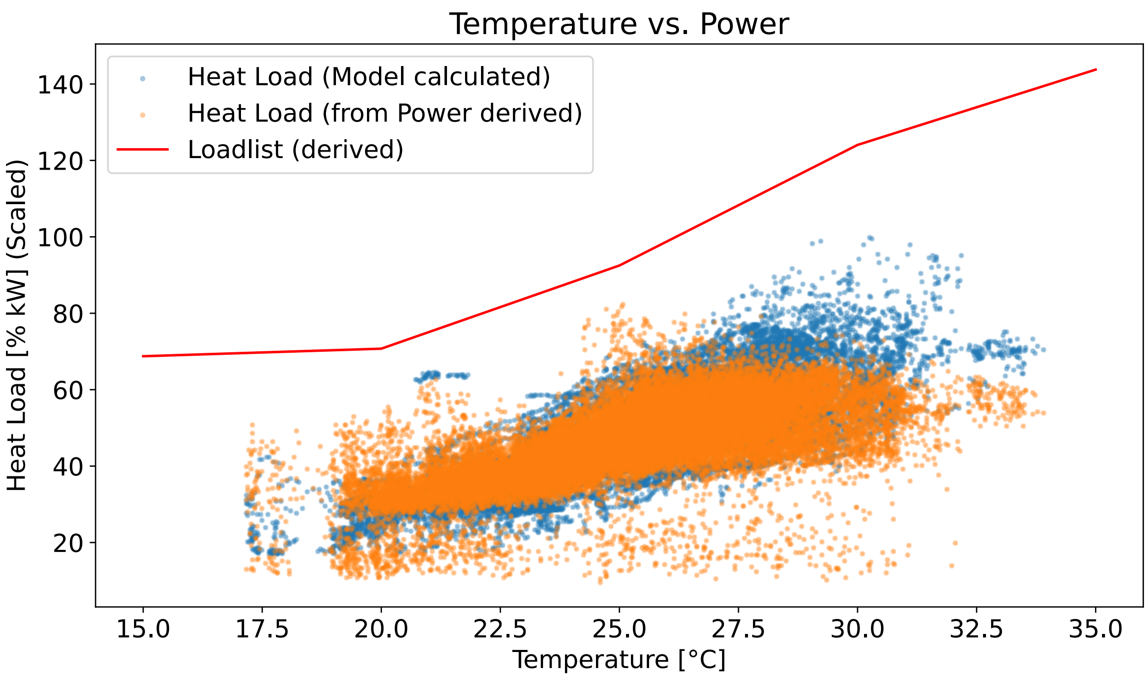


Figure 6.2: Temperature ($°C$) vs. Power (kW) (Shifted data)

6.2. Application Black box

The solo black box model shows a great performance. With the nine input variables discussed in 5.6, the black box performs with a mean absolute percentage accuracy of 90.42%. This model can be used for this particular yacht for a possible refit in the future. It can help optimise the operation of the HVAC system by predicting power consumption under various conditions. This can guide the external contractor to adjust settings like temperature and airflow to minimise power usage without compromising comfort. Effective energy management is crucial for the sustainable operation of yachts. The model can provide insights into the power consumption patterns of the HVAC system, facilitating better integration with onboard energy sources like generators and batteries.

This is all on the condition that the values of these conditions are within the range of the trained data. Table 6.1 shows the data ranges where the black box model is currently trained on.

Table 6.1: Range and mean of the Black box trained data

Input Variable	Minimum	Mean	Maximum	Unit
Global Horizontal Irradiance	0.0	173.1	764.3	$[W/m^2]$
Windspeed	0.0	4.6	12.6	$[m/s]$
Average Room Setpoint Temperature	19.8	20.3	21.0	$[°C]$
Average Return Air Temperature	17.9	18.4	19.8	$[°C]$
Average Fresh Air Temperature	17.2	25.2	33.9	$[°C]$
Average Mixed Air Temperature	18.7	23.0	27.8	$[°C]$
Average Supply Air Temperature	11.3	13.8	16.0	$[°C]$
Average Return Air Humidity	48.2	59.7	68.2	[%]
Average Supply Air Humidity	58.3	81.9	88.4	[%]

6.3. Application Grey box

The Grey box model shows an improved performance to the black box model with an accuracy of 91.29%. Once additional voyage data from currently sailing Feadship yachts becomes available, the grey box model can be trained using the input from the white box model. Due to this, the relationship between vessel size and HVAC energy demand can become more evident. As a result, yachts in the design phase for which no voyage data is available will also be able to use the grey box model.

6.3.1. General

In assessing the performance of the prediction model, as shown in Figure 6.3, several key observations are worth discussing. The plot shows a comparison between actual power values and those predicted by the model over a selected time frame in June.

Firstly, the predictive model shows a commendable level of accuracy, closely matching the actual data points over the majority of the observed period and it is capable of capturing the general trends in the power consumption. However, the model's accuracy is less consistent at the extremes of the data range, with notable discrepancies around the 16th and between the 22nd and 24th of June. During these intervals, the predicted power levels significantly overshoot the actual measurements, indicating that the input variables do not have the information for this behaviour of the HVAC system's power. The model lacks of the ability to account for these anomalies or outlier data points, while proper data preparation is performed.

Consistency is another aspect to consider, with the model showing greater stability in its predictions within the mid-range of the power values. However, reliability appears to decrease when predicting lower and higher power requirements. This variability could limit the practicality of the model in later design phases like the 'Basic Design' phase detailed in 2.1, where accurate power estimates are more critical. To mitigate this, a safety factor could be incorporated into the design based on the model's performance on previous data. It is essential to consider the difference in the data range where the second chiller system is operational compared to when only one system is active, as the error rates differ significantly, as mentioned in 5.5.

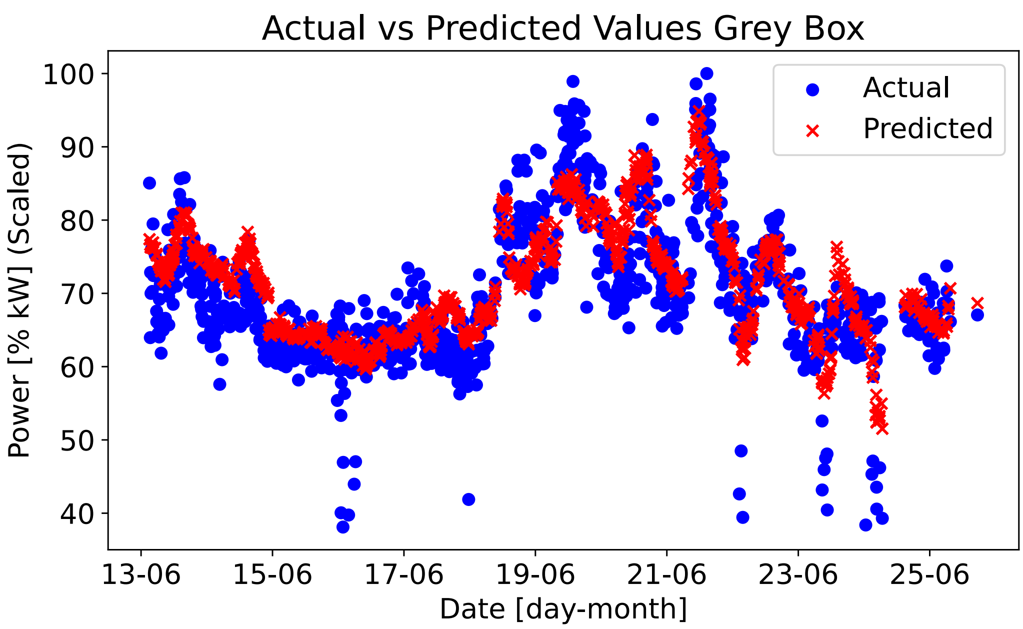


Figure 6.3: Grey box predictions selected time frame

6.3.2. Extrapolation Capabilities

Another application of the grey box model is improved extrapolation. To evaluate the extrapolation capabilities of a grey box model, a simple extrapolation is compared to a standalone black box model extrapolation. A subset of the original data is selected and only the fresh air temperature is raised by 5 degrees Celcius while keeping other variables constant. This is done to create new data that both the grey box and black box models are not familiar with. Only the fresh air temperature is changed because the data of the relative humidity is less reliable as mentioned in 5.4.

Another reason this approach is chosen, is because the outdoor air temperature is a critical variable that has a significant impact on the heat load and power demand of the HVAC system. By changing only this variable, a controlled scenario is created that isolates the effect of temperature changes on the models' predictions. This method allows an assessment of how well the models handle conditions beyond the original data, providing insight into their reliability under potential future climate scenarios or unexpected temperature increases.

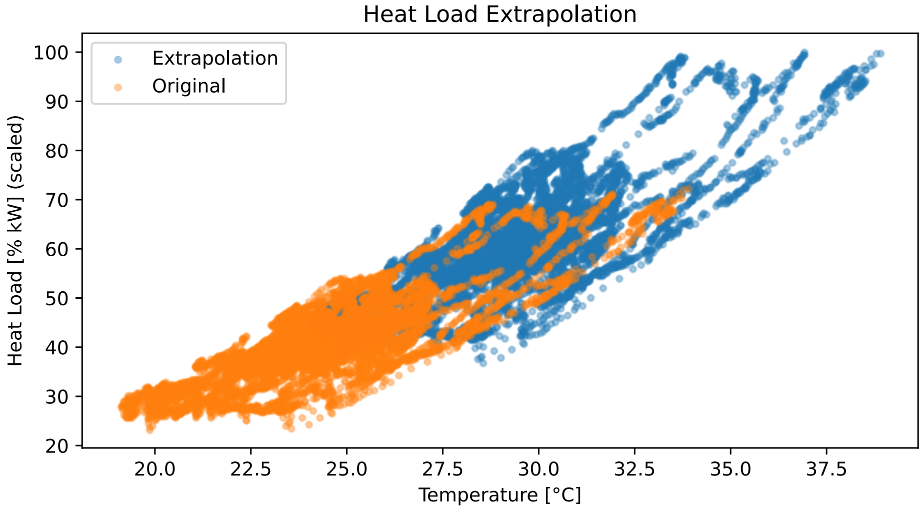


Figure 6.4: Heat Load extrapolation white box model

Figure 6.4 analyses the extrapolation of the white box model. The graph shows the relationship between temperature and heat load, compares the original data set and the extrapolated data set. The heat load shows a positive correlation with temperature in both datasets. In particular, the extrapolated data extends the heat load range to higher temperatures, while maintaining a similar trend to that observed in the original dataset.

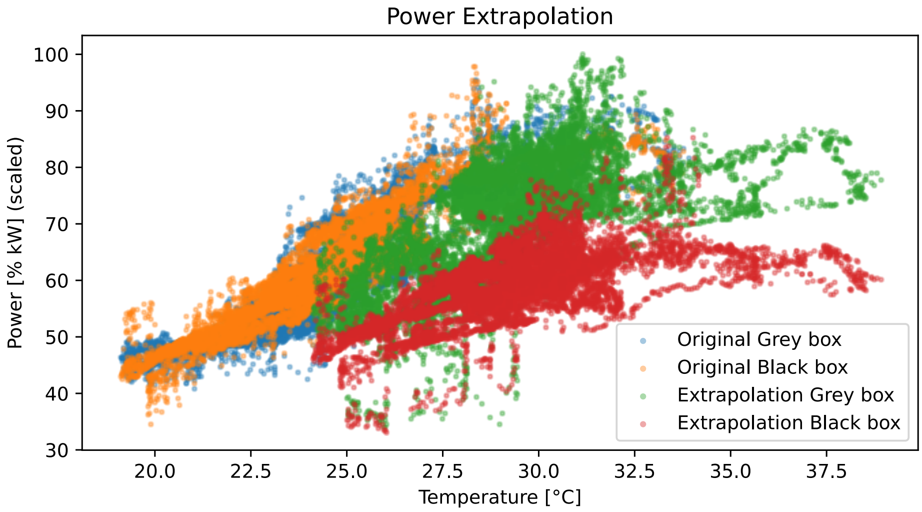


Figure 6.5: Power extrapolation grey and black box model

Figure 6.5 compares temperature and power demand, the original grey box model data and black box model data is shown. Their extrapolated counterparts are shown in green and red. The extrapolated grey box model shows an increased power demand at higher temperatures, following the trend observed in the original data. The extrapolated data from the black box model also follows the trend, but at a lower power level.

Analysing the extrapolation capabilities of the models reveals several key points. The white box model maintains the trend observed in the original dataset during extrapolation, with the increased fresh air temperature resulting in higher heat loads as expected.

For the grey box model, the extrapolated power demand aligns well with the increased temperature, maintaining the trend observed in the original data. The result can not be validated since data in this range is missing. Looking at the orginal data, it slightly underestimates the expected trend, but this can also be due to the fact that the relative humidity is not changed in the extrapolation.

The black box model, while capable of following the general trend, shows an underestimation of the power over the whole data set. This suggests that the black box model alone is less accurate then the grey box model in extrapolation.

Overall, the grey box model demonstrates good extrapolation capabilities, effectively combining the strengths of both the white box and black box approaches. The black box model alone, is less reliable, suggesting that a hybrid approach provides more reliable results for predicting HVAC energy demand under varying conditions.

7 Conclusion

In the previous chapter the possible applications of the model are outlined. In this final chapter, conclusions will be drawn based on the results of Chapter 5. First the research questions that are detailed in section 1.5 will be answered. Second, the model requirements are investigated to determine if the objectives have been met. It continues with the discussion where the contributions

7.1. Conclusion Research Questions

To develop a feasible model, several sub-questions are proposed in section 1.5. These questions are answered throughout this report and can be summarised as below.

1. **What is the state of the art in predicting the energy consumption of HVAC systems?**
   Different methods for HVAC energy demand estimations have been developed. The currently employed method is an empirical method with load lists. This method is used to only estimate the maximum power at different temperatures. However, this method shows large inaccuracies and the weather conditions used for this method rarely occur. The method is highly subjective and depends on the experience of the external contractor. Current calculation methods based on ISO-7547 (2002) show promising results for future use and improvement. Together with the data that is collected by Feadship of their currently sailing yachts, a data-driven modelling approach could be the solution.
2. **What are the method requirements to model the energy consumption of HVAC systems?**
   The proposed method is a white box model that proportions the calculations to the actual weather conditions to make an estimation of the heat load and integrating a black box model that incorporates the available voyage data to predict the power consumption of the HVAC system. The method requirements are detailed in section 2.7.
3. **What methods are suitable and in what ways can machine learning and grey box modelling contribute to solving the problem?**
   The proposed grey box model provides a solution for predicting the HVAC energy demand using both data-driven and transparent approach. The solo white box model only calculates the heat load, but does not cover every aspect of the estimation problem. The black box is implemented to cover these remaining aspects with the help of Machine learning, an artificial neural network in particular. The proposed model is detailed in section 3.4 and its application is detailed in section 6.3.
4. **How can the integration of the grey box approach optimise the prediction and understanding of HVAC system energy consumption?**
   The integration of the grey box approach into the prediction of HVAC system energy consumption optimises performance by combining the strengths of white box and black box methods. Specifically, the grey box model combines the precision of the white box model, which is a heat load calculation based on ISO-7547 (2002) with solar radiation and actual sensor data included, and

incorporates this with the adaptability of black box models. An artificial neural network is proposed and a grid search to determine the hyperparameters of the ANN is explored. By using both empirical data and theoretical principles, the grey box approach provides a balanced, efficient solution.

5. **How can the accuracy of the model be validated?**
The white box model can be validated by deriving the heat load from the power data using efficiencies of the system and a calculated COP as detailed in 5.2. Evaluating the two results reveals that the heat load calculated by the model is lagging. Adjusting this shows improvements across the whole data. The final grey box model calculates the total power consumption using an ANN. This ANN makes use of a kfold cross-validation to select the optimal hyperparameters of the configuration of the ANN. The results of the optimal model are validated against the power consumption recorded in the voyage data. A comparison of these results in section 5.7 shows that the grey box model outperforms the black box model by a slight improvement, with a MAPE accuracy that is 0.87 % higher.

6. **How can the proposed model be implemented in the design of new future yachts, and what criteria must be met to consider it to be successful?**
When evaluating the predictive models, the solo Black Box, Grey Box and White Box models each offer distinct advantages based on their computational complexity and data handling capabilities. The solo white box model, although computationally intensive, is essential when no prior voyage data is available or when operational parameters are beyond the range of the trained data. The ability of this model to make predictions based on environmental conditions is critical for initial estimates of HVAC loads. The solo black box model has a high level of accuracy at 90.42%, making it suitable for optimising the systems of the existing yacht within known operating conditions. The grey box model, slightly more accurate at 91.29%, benefits from additional input from the white box model and promises improved predictive power for yachts still in the design phase. The extrapolation capabilities show improved results compared to the solo white box model. When more data becomes available, the grey box model can improve its extrapolation capabilities and it can be trained using the input from the white box model from different ships.

The main research question of this research was:

’To what extent can data from previously built DVNA yachts accurately predict the energy consumption of Heating, Ventilation and Air Conditioning (HVAC) systems in order to improve the design of new future yachts?’

The development of a grey box model to analyse sensor data and predict energy demand is a notable improvement within the limitations of the available data. This model allows accurate predictions of the HVAC system heat load and power consumption for the early stages of the design of yachts, provided that it falls in the ranges of the dataset that it is trained on. When more data of the energy demand of the HVAC system of different yachts becomes available, the grey box model can be expanded. Once the parameters of a yacht have been estimated and a operational profile has been set, predictions can be made at a more accurate level than is currently possible. In addition, a white box method has been developed to give insight in the heat load of the HVAC systems. This method improves the understanding of energy consumption trends and identifies areas where system loads are high and energy distribution is inefficient. These models achieve the primary objectives of this thesis, but also make further research possible.

8

Discussion & Recommendations

The research centred on the development of a predictive grey box model that improves the accuracy of energy consumption predictions by incorporating both theoretical principles and empirical data. The contributions of this research are detailed in section 8.1. Although significant progress has been made, the applicability of the model is limited by some constraints, which are discussed in section 8.2. Addressing these limitations with recommendations is essential to refine the model's effectiveness and ensures its practical use in yacht design and operation. These potential improvements are proposed in section 8.3.

8.1. Contributions

This report presents a comprehensive method for modelling the energy consumption of Heating, Ventilation and Air Conditioning systems for yachts. Through the development of a grey box predictive model, contributions have been made to the field of marine HVAC energy demand estimations. The main contributions of this study are outlined below:

- Development of a grey box model: One of the major contributions of this thesis is the development and refinement of a grey box model that combines the transparency of white box models with the predictive power of black box approaches. This model uses both empirical data from yacht voyages and theoretical principles derived from standard protocols by ISO-7547 (2002). By integrating machine learning techniques, specifically artificial neural networks, the model adapts to varying weather conditions and operational parameters, providing a tool for predicting HVAC energy requirements with improved accuracy for the early stages of the design of new yachts.
- Improving prediction accuracy: The grey box model developed in this study demonstrated an improvement in prediction accuracy over a solo black box model. In a comparison with the solo black box model, the grey box approach has shown a small but notable increase in accuracy, with a Mean Absolute Percentage Error (MAPE) that is 0.87% better than the black box model. This improvement is crucial for optimising energy management on yachts, leading to better resource allocation and operational efficiency.
- Application to yacht design: The white box model developed, is directly relevant to the design of new yachts and the optimisation of existing vessels. The heat load calculation at various weather conditions can be used to during the design stages of a yacht to ensure that HVAC systems are tailored to the unique conditions each yacht will face.
- Methodological contributions: This report has detailed the integration of multiple methods in the white box model. The methodological advances made here, including the use of methods such as the integration of the solar radiation model and the surface temperature caused by the solar radiation, pave the way for future research and development. These methodologies provide a scalable solution that can be adapted to different types of vessels beyond yachts.

8.2. Limitations

While the development of predictive models for HVAC system energy consumption has led to these contributions, there are limitations that must be acknowledged to fully understand the scope and application of the results. These limitations affect the overall effectiveness and applicability of the models in specific operational scenarios.

- Dependence on external data for fresh air humidity: One of the notable limitations is the reliance on the ECMWF (2023) weather database for fresh air humidity data, which was not present in the original sensor dataset. This dependency potentially introduces variability and uncertainty into the models, as the external data may not perfectly match the specific conditions experienced by the HVAC systems on the yachts. The accuracy of the humidity data is critical for accurate energy consumption predictions and any discrepancies in this data can affect the model outputs.
- Exclusion of reheat and fan coil contributions: The predictive models do not currently include the contributions of reheaters and fancoils, as the supply air measurements are taken after these components in the five AC units. These can be included if the supply temperatures per room are measured in the sensor data.
- Challenges in modelling: Initially, a method was proposed to split the energy demand across the five AC zones. This would already advance the method to five different zones and give a zonal design instead of the whole yacht. However, this approach proved impractical due to a lack of data on how the power from the chiller is allocated to each ACU. Without this critical information, it is not possible to accurately model and predict the specific energy consumption for each zone.
- Omission of actual occupancy data: The white box model, designed to predict HVAC energy demand based on theoretical and environmental parameters, does not incorporate actual occupancy data. Instead an estimation is made per room. Occupancy levels have an impact on the heat load, as the presence and activity levels of people affect the need for heating, cooling and ventilation.

These limitations highlight the need for improved data collection methods and model adjustments to improve the accuracy and reliability of the model. Addressing these issues will be critical for future iterations of the models to enable more accurate solutions for yacht HVAC systems.

8.3. Recommendations

A number of recommendations are proposed to address the limitations identified in the previous section. These recommendations aim to improve data accuracy, model comprehensiveness and overall prediction reliability. Addressing these areas is critical to refining the model and ensuring its practical application in yacht design and operation.

- FAH sensor data: As humidity sensors are already installed on the yachts, but are reporting inaccurate data (reading zero), it is important to address this issue. A diagnostic should be carried out to determine whether the problem is with the sensor hardware, software or data transmission processes. The availability of this data can further enhance the model.
- Inclusion of reheaters and fan coils in the model: The current exclusion of reheaters and fan coils from model calculations is a significant limitation. It is recommended to modify the sensor setup to capture data before and after these components within the HVAC system. This data will allow the contribution of reheaters and fancoils to be accurately assessed and included in the energy consumption predictions, providing a more comprehensive understanding of the total energy demand. A analysis of the power consumption of the reheaters is provided in D.
- Energy distribution for AC zones: To address the challenge of distributing energy demand across the five air conditioning zones, an approach using additional data is recommended. Specifically, the temperatures of the chilled water entering and leaving each ACU can provide critical insight into the thermal energy transfer within each zone. The mass flow of the air is also needed for this method, which is not measured at this point. Implementing this measurement setup will not only allow validation of energy consumption per zone, but will also support a zonal design approach for the yacht, predicting energy consumption based on the size of these zones.

- Collecting occupancy data through crew reports: The crew possibly maintains a daily log or itinerary already that records the number of occupants. This method would use the crew's routine activities to collect real-time data without additional hardware costs. The advantage of this is that the crew may already have historical data, making other historical data usable.
- Analysis of power sensor data variance: The high variance observed in power sensor data, which contributes to errors in energy consumption estimates, needs to be thoroughly investigated. A detailed analysis is recommended to identify the sources of this variance. Factors such as sensor calibration, placement and maintenance should be investigated to ensure data integrity. The environmental conditions included in the black box can not clarify this variance at this point, which makes it highly likely that something else is causing the variance.

Implementing these recommendations will significantly improve the ability of predictive models to accurately estimate the heat load and HVAC energy consumption, thereby improving energy management and operational efficiency on the yachts. These improvements are essential for the development of more sustainable and efficient marine HVAC systems.

Bibliography

ASHRAE (2013). *ASHRAE Handbook: Fundamentals*. American Society of Heating, Refrigeration, and Air Conditioning Engineers.

Bakker, M. (2021). A reference-based design approach: in preliminary ship design. Master's thesis, TU Delft.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.

Boertz, C. (2020). Energy demand of a fuel cell-driven cruise ship. Master's thesis, TU Delft.

Botchkarev, A. (2019). A new typology design of performance metrics to measure errors in machine learning regression algorithms. *Interdisciplinary Journal of Information, Knowledge, and Management*, 14:045–076.

Chen, Y., Guo, M., Chen, Z., Chen, Z., and Ji, Y. (2022). Physical energy and data-driven models in building energy prediction: A review. *Energy Reports*, 8:2656–2671.

Chollet, F. (2017). *Deep Learning with Python*. Manning.

de Haas, M. (2022). Early stage fouling effects prediction for yacht design: A grey-box model approoach using operational voyage data. Master's thesis, TU Delft.

de Vroet, J. (2018). Dimensioneren van hvac-systemen op basis van omgevingsdata. Internal Documentation.

DVNA (2022). De voogt organisation. Technical report, De Voogt Naval Architects.

DVNA (2023). Concurrent design feadship. Technical report, De Voogt Naval Architects.

ECMWF (2023). Datasets.

Engineering Toolbox (2003). Moist air - the mollier diagram. Accessed: 2023-11-28.

Feadship (2015). Ac feadship standard. Technical report, Feadship.

Feadship (2022). Hvac power consumption. Technical report, Feadship.

Feadship (2023). Obsidian: Feadship's first delivery of 2023 raises the bar on carbon reduction.

García, S., Ramírez-Gallego, S., Luengo, J., et al. (2016). Big data preprocessing: methods and prospects. *Big Data Analysis*, 1(9).

Géron, A. (2023). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc., Sebastopol, CA.

Haranen, M., Pakkanen, P., Kariranta, R., and Salo, J. (2016). White, grey and black-box modelling in ship performance evaluation.

Harvery, J. (2009). Basic design concepts. *Journal of the American Society for Naval Engineers*, 71:671–678.

Heaton, J. (2015). *Artificial Intelligence for Humans: Deep learning and neural networks*. Artificial Intelligence for Humans Series. Heaton Research, Incorporated.

Heinen & Hopman (2022). Air handling units. Accessed: 2023-11-20.

Huotari, J., Ritari, A., Vepsäläinen, J., and Tammi, K. (2020). Hybrid ship unit commitment with demand prediction and model predictive control. *Energies*.

IMO (2020). Fourth greenhouse gas study 2020. Technical report, International Maritime Organization, 4 Albert Embankment, London SE1 7SR.

ISO-7547 (2002). Ship and marine technology - airconditioning and ventilation of accomodation spaces - design conditions and basis of calculation. Technical report, International Organization of Standardization.

Kalogirou, S. A., Florides, G. A., Neocleous, C., and Schizas, C. N. (2001). Estimation of the daily heating and cooling loads using artificial neural networks.

Kamstra, T. (2020). Eload analyse op basis van realdata en het polestar systeem. Internal Documentation.

Klein Woud, H. and Stapersma, D. (2016). *Selected Chapters of 'Design of Auxiliary Systems, Shafting and Flexible Mounting'*. IMarEST.

Kohavi, R. (2001). A study of cross-validation and bootstrap for accuracy estimation and model selection. 14.

Kumar, R., Aggarwal, R., and Sharma, J. (2013). Energy analysis of a building using artificial neural network: A review. *Energy and Buildings*, 65:352–358.

Leifsson, L. P., Sævarsdóttir, H., Sigurðsson, S. □., and Vésteinsson, A. (2008). Grey-box modeling of an ocean vessel for operational optimization. *Simulation Modelling Practice and Theory*.

Lin, L. and Sherman, P. D. (2023). Cleaning data the chauvenet way. In *Name of Conference Goes Here*, San Mateo, CA; San Jose, CA. MedFocus; Independent Consultant.

Ljung, L. (2001). Black-box models from inputs-output measurements. *18th IEEE Instrumentation and Measurements Technology Conf., Budapest*.

Logan, K. (2011). Using a ship's propeller for hull condition monitoring. Technical report, ASNE Intelligent Ships Symp. IX.

Mavris, D. and DeLaurentis, D. (2000). Methodology for examining the simultaneous impact of requirements, vehicle characteristics, and technologies on military aircraft design. *In Proc. 22nd Congress of the Int. Coun. on Aeronautical Sci. (ICAS)*, page 145.1–145.10.

MEPC (2023). 2023 IMO Strategy on Reduction of GHG Emissions from Ships. Accessed: 2023-11-28.

Moayedi, H., Bui, D. T., Dounis, A., Lyu, Z., and Foong, L. K. (2019). Predicting heating load in energy-efficient buildings through machine learning techniques. *Applied Sciences*, 9(20).

Nasruddin, Sholahudin, Satrio, P., Mahlia, T. M. I., Giannetti, N., and Saito, K. (2019). Optimization of hvac system energy consumption in a building using artificial neural network and multi-objective genetic algorithm. *Sustainable Energy Technologies and Assessments*, 35:48–57.

Odendaal, K. (2021). Enhancing early-stage energy consumption predictions using dynamic operational voyage data. Master's thesis, TU Delft.

Parkes, A. I., Sobey, A. J., and Hudson, D. A. (2018). Physics-based shaft power prediction for large merchant ships using neural networks. *Ocean Engineering*, 166:92–104.

Pedersen, B. P. and Larsen, J. (2009). Prediction of full-scale propulsion power using artificial neural networks. *COMPIT '09: 8th International Conference on Computer and IT Applications in the Maritime Industries*.

Petersen, J. P., Jacobsen, D., and Winther, O. (2011). A machine-learning approach to predict main energy consumption under realistic operational conditions. *10th Int. Conf. Computer and IT Applications in the Maritime Industries (COMPIT),*.

Roy, J., Shallcross, P., Hardy, A. M., and Burnay, S. (2011). Reducing the environmental impact of large yachts. Technical report, Genoa, Italy : The Royal Institution of Naval Architechts. Design Conference: Design, Construction & Operation of Super & Mega Yachts.

Runge, J. and Zmeureanu, R. (2019). Forecasting energy use in buildings using artificial neural networks: A review. *Energies*, 12(17).

Sarbu, I. and Sebarchievici, C. (2017). *Chapter 2 - Solar Radiation*. Academic Press.

scikit-learn developers (2024). Cross-validation: evaluating estimator performance. Accessed: 2024-04-16.

Sen, P. C. (2013). *Principles of electric machines and Power Electronics*. John Wiley and Sons, Inc.

Shrivastava, S. (2020). Cross validation in time series. Medium. Retrieved May 18, 2022.

Silva, I., Spatti, D., Flauzino, R., Bartocci Liboni, L., and Reis Alves, S. (2017). *Artificial Neural Networks*. Springer.

SuperyachtNews (2023). Yeti 1.0 goes live. https://www.superyachtnews.com/operations/yeti-10-is-now-live-, revised on 2023-08-07.

Talib, A., Park, S., Im, P., and Joe, J. (2023). Grey-box and ann-based building models for multistep-ahead prediction of indoor temperature to implement model predictive control. *Engineering Applications of Artificial Intelligence*, 126:107115.

The Superyacht Group (2023). The superyacht new build report. Issuu.

Toolbox, E. (2023). Ethylene glycol heat-transfer fluid properties. Accessed: 2023-02-02.

UNFCCC (2016). Paris agreement. Technical report, United Nations, United Nations Framework Convention on Climate Change.

van Wijngaarden, M. and El Mouhandiz, A. A. (2023). Direct calculations - technical report. Internal Documentation.

Yuan, J. and Wei, S. (2019). Comparison of using artificial neural network and gaussian process in ship energy consumption evaluation. *DEStech Transactions on Environment Energy and Earth Science*.

Zwart, R. (2020). Trim optimization for ships in service: A grey-box model approach using operational voyage data. Master's thesis, TU Delft.

A

Absorption and convective heat coefficient

Kleur		R	ΔT in K Horizontaal	ΔT in K Verticaal
White		0.80	7	5
Aluminum Zinc (GL)		0.67	14	11
Polar White (PW)		0.66	15	11
Snow White (SW)		0.65	15	12
Almond (AL)		0.63	16	13
Oyster White (WH)		0.52	22	17
Light Stone (LS)		0.50	23	18
Copper Metallic (CM)		0.46	25	20
Brownstone (BS)		0.44	26	20
Scarlet Red (SR)		0.42	27	21
Ash Grey (AS)		0.37	29	23
Sahara Tan (ST)		0.36	30	24
Colony Green (GR)		0.34	31	24
Hawaiian Blue (BL)		0.32	32	25
Harbor Blue (HB)		0.28	34	27
Burnished Bronze (BR)		0.28	34	27
Hunter Green (HG)		0.28	34	27
Fern Green (FG)		0.28	34	27
Black		0.05	45	36

Figure A.1: Reflective coefficients for different colours (van Wijngaarden and El Mouhandiz, 2023)

The calculation of the absorption coefficient according to van Wijngaarden and El Mouhandiz (2023) is as follows:

$$\alpha = 1 - R \tag{A.1}$$

With R as the reflective coefficient and α as the absorption coefficient. For the convective heat transfer coefficient the windspeed v_a is needed. According to (Feadship, 2015) convective heat transfer coefficient is calculated by:

$$h_c = \begin{cases} (6.2 + 4.2 \cdot v_a) & \text{for } v_a \leq 5\,\text{m/s} \\ (7.53 \cdot v_a^{0.78}) & \text{for } v_a > 5\,\text{m/s} \end{cases} \tag{A.2}$$

B

White Box Model Development

At the beginning of the development of the model, a specific yacht was selected for which data is readily available. The chosen yacht has an existing layout and is currently operational, ensuring the availability of sufficient data for an energy demand estimation. An Excel spreadsheet was created to compile all relevant information about the yacht, including its dimensions, materials, and volumes, among other details. This spreadsheet is then used within a Python script to perform the final calculations, allowing the methodology to be applied to other yachts as well.

B.0.1. General Arrangement Spreadsheet

The general arrangement (GA) of all the rooms that are connected to the yachts HVAC system are collected in a spreadsheet built by DVNA. All connected surfaces to the area are listed with their respected area, boundary type, adjoining space and exterior finish.

Design number		YN 818	
date		15-12-2023	
room name		MD pantry	
room type		Pantry	crew
occupancy (daytime)		1	persons
Air handler		AC1	
zone type		interior zone1	
design temperature		21	degrees
# of fan coils		0	
nominal capacity		6	x1000 BTU/h (ref fc type)
overall dimensions			
	floor area	9	m2
	headroom	2.1	m2
	volume	18.9	m3

name	area [m2]	Boundary type	adjoining space	exterior finish
Ceiling AC-room	4.9	deck alu A0/A60 +10-6-10 floor + ceiling (SS dec	tech space	
Windows (nxA)	1.4	glass, standard for this yacht	Air	vertical glass, tinted
Outer bulkh. long.	5.8	ext shell alu steel equivalent + finish	Air	surface temperature vertical white surface
Outer bulkh. trans.	0			
Inner bulkh. Casing	6	int bhd A60 + finish	engine room	
Inner bulkh. Foodlift/el.stc	16.8	int bhd A60 + finish	unconditioned (zone1)	
Fl/all.w/Sanit.space	0			
Floor ER	8.5	int deck steel A60 + floor 10-6-10 (ER)	engine room	

Figure B.1: Example room from the General Arrangement spreadsheet

B.0.2. Python Calculation

The script is designed to calculate the thermal loads and air conditioning requirements for various areas or rooms within the yacht. The application processes the GA spreadsheet that contains various sheets with relevant data for these calculations.

To be able to use the script for other yachts with different dimensions a generic calculation method is used:

- At the start, the script requires the user to specify the number of rooms on the yacht and the number of air conditioning units are in operation on the yacht. Additionally, the user must provide the file paths for the voyage data and the GA spreadsheet.
- The script processes the GA spreadsheet. This data forms the backbone of the thermal load analysis, providing the essential attributes and coefficients required for accurate calculations.
- For each data point within the voyage data, the script updates a dedicated temperature table, adjusting the values to reflect the current environmental conditions. It sets the temperatures and humidity's for fresh air, return air, and supply air based on the specific data point under consideration. Subsequently, it calculates the enthalpy of these conditions.
- The analysis extends to evaluating the surface temperatures for both vertical and horizontal surfaces, incorporating variations for light and dark finishes.
- The calculation detailed in subsection 4.4 is then performed for every room in the yacht. For each AC system and data point, the script calculates various thermal parameters such as the temperature differences on both sides of the surfaces, solar heat, thermal heat by windows, and total heat for each area. It iterates through the surfaces of every room and uses a function to fetch values based on matching conditions from the GA spreadsheet. It works like a look up function. Some tables used for reference, like the boundary type table shown in Figure B.2, stay the same for the whole analysis. It gives each surface a k value and each window a k and G value. In contrast, tables like the zone temperature table in Figure B.3 get updated with new data each time.

	type	k [W/m2K]	G [-]
0	int deck steel no insul + ply floor (TT)	0.49000	0.00000
1	int deck steel no insul + 10-6-10 floor + ceiling (LD/MD)	0.39000	0.00000
2	int deck steel A60 + floor 10-6-10 (ER)	0.73000	0.00000
3	int deck alu steel equivalent +10-6-10 floor + ceiling (SS decks)	0.54000	0.00000
4	int deck alu A0/A60 +10-6-10 floor + ceiling (SS decks)	0.39000	0.00000
5	int bhd C = finish	0.60000	0.00000
6	int bhd B0 + finish	0.40000	0.00000
7	int bhd A60 + finish	0.30000	0.00000
8	int bhd A0 + finish	0.35000	0.00000
9	ext shell steel + finish	0.55000	0.00000
10	ext shell alu steel equivalent + finish	0.50000	0.00000
11	ext deck steel + mascoat + teak + ceiling	0.55000	0.00000
12	ext deck alu steel equivalent + teak + ceiling	0.39000	0.00000
13	ext deck alu steel equivalent + ceiling (dodgers)	0.50000	0.00000
14	wooden cns	0.31000	0.00000
15	glass, standard for this yacht	5.30000	0.40250
16	glass single, clear	5.50000	0.70000
17	glass single, tinted	5.50000	0.40000
18	glass single, coated	5.50000	0.53000

Figure B.2: Boundary types with corresponding k and G values in the look up table.

	type	temp [deg C]	humidity	enthalpy
0	738390.2465283459	12.80410	85.33300	32.73587
1	Air	28.33767	52.49943	60.99968
2	sea water	24.33767	nan	nan
3	interior zone1	18.64581	51.00000	36.07945
4	interior zone2	50.00000	50.00000	155.36460
5	interior zone3	21.00000	50.00000	40.83878
6	tech space	28.33767	80.00000	nan
7	engine room	43.33767	80.00000	nan
8	unconditioned (below wl)	24.33767	nan	nan
9	unconditioned (zone1)	24.00000	nan	nan
10	unconditioned (zone2)	53.00000	nan	nan

Figure B.3: Temperatures of different areas in the yacht at a certain data point, collected in a look up table

- After computing all these parameters across different data points and areas, the script aggregates this data to provide an overview of the heat load and cooling requirements per AC unit.

In summary, the script is a tool for analysing thermal loads and determining the air conditioning requirements for different areas within a yacht, factoring in various dynamic and static inputs to provide detailed and customised cooling profiles for each area.

C

Complete White Box Results

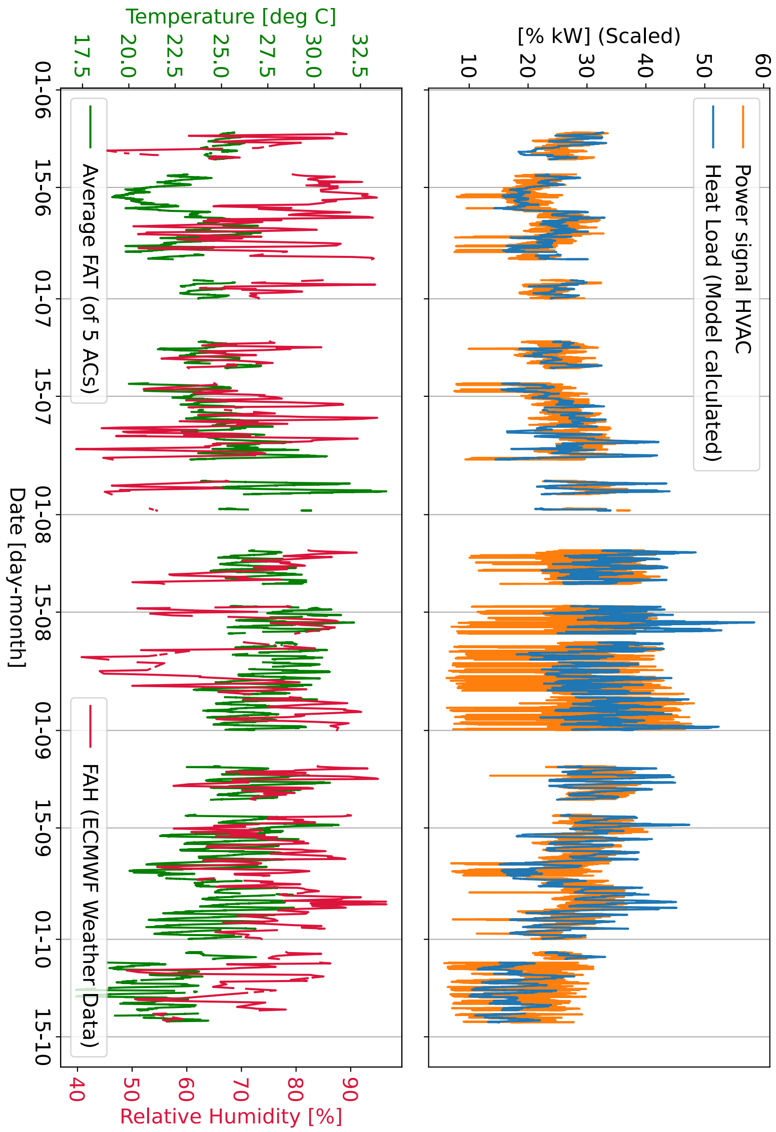


Figure C.1: Complete White box model results with the fresh air temperature and humidity

D
Reheater addition

In addition to the recommendation an analysis is done of the contribution of the reheaters tot the total HVAC energy demand. Figure D.1 shows the the reheaters in question. The reheaters are excluded from the model due to the fact that there is no information available on the eventual supply air temperature per area. The power of the AC units is the combined power of the heater, humidifier, switchboard and fans.

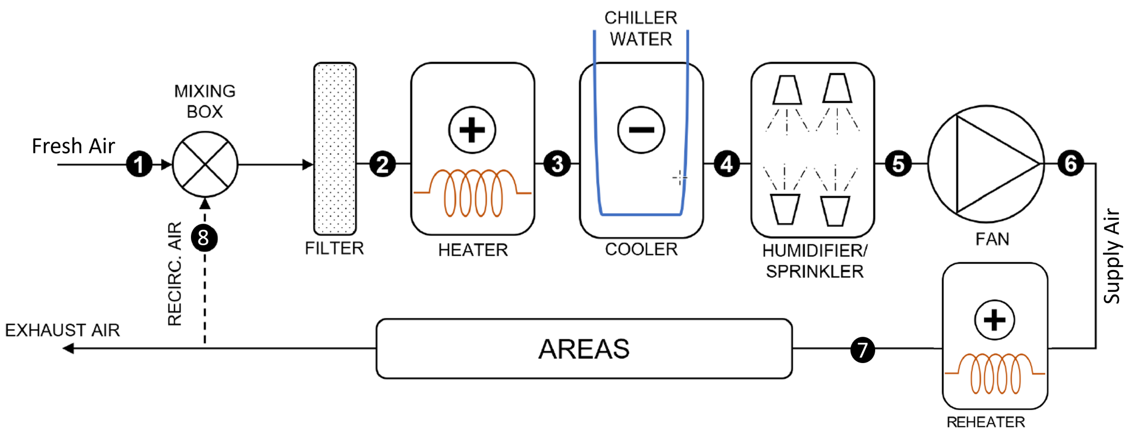


Figure D.1: Schematic of the HVAC system of the yacht.

A way to reduce the energy consumption of the ship is to replace these reheaters by heaters that work on waste heat. The following data is analysed to understand how much energy can be saved. Figure D.3 and D.2 show the total power demand of the reheaters, and the contribution of the reheaters on the total of the HVAC system. The reheaters contribute to 24.1% of the total HVAC consumption on average.

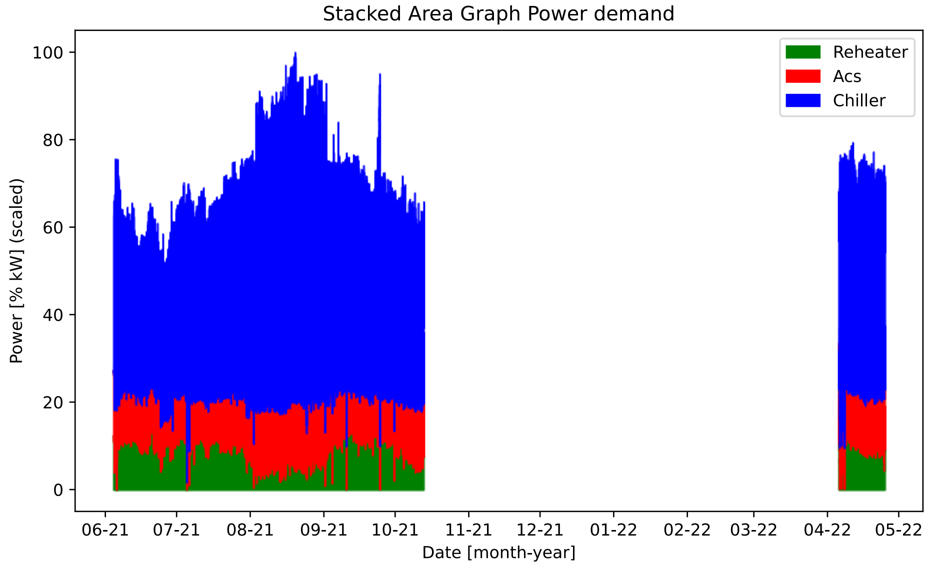


Figure D.2: Stacked Area Graph Power Demand.

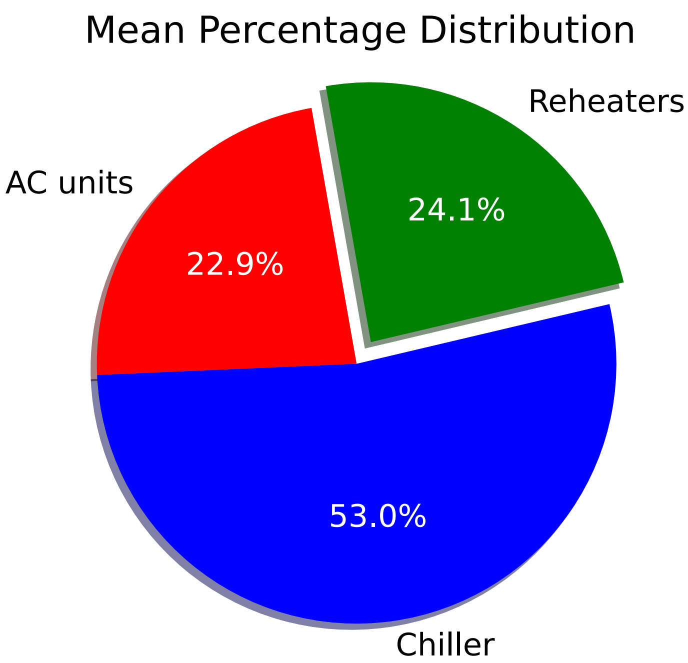


Figure D.3: Mean Percentage of Total Power Consumption per consumer

There is not much correlation visible between the fresh air temperature and the total power of the reheaters, but the average seems to decrease at higher temperatures.

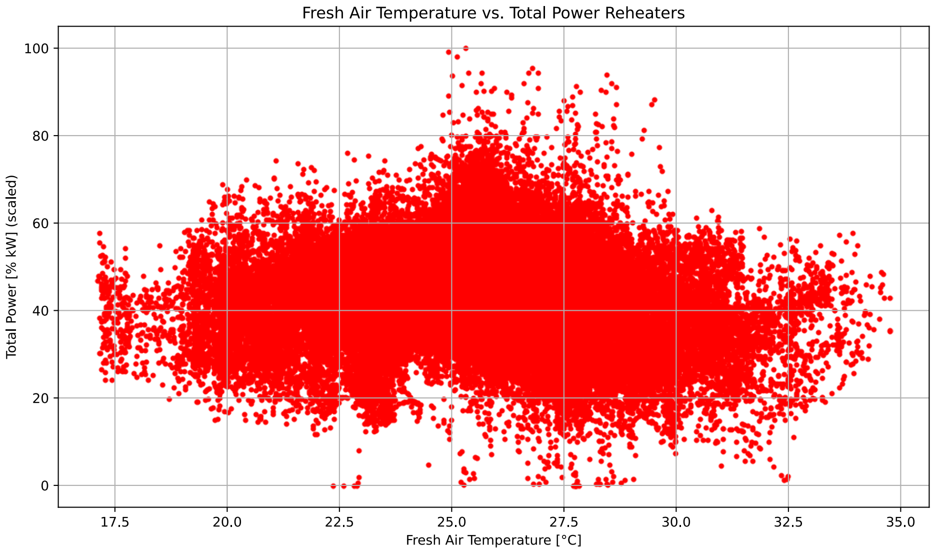


Figure D.4: Fresh Air Temperature vs. Total Power Reheaters.

[redacted]

Figure D.5: [redacted]

[redacted]

Figure D.6 shows a box plot for every reheater. It shows that there are several rooms with a high demand. These are the larger rooms of the yacht like the lounge, dining, owners stateroom or bridge. These rooms have multiple reheaters. The box plot reveals multiple outliers for the power consumption of individual reheaters. These outliers are attributed to the varying usage patterns of the rooms. For instance, the owner's stateroom only consumes power when the owner is on the yacht. Consequently, during the owner's infrequent visits, the power usage spikes, creating outliers in the data, as the system is typically turned off when the room is unoccupied.

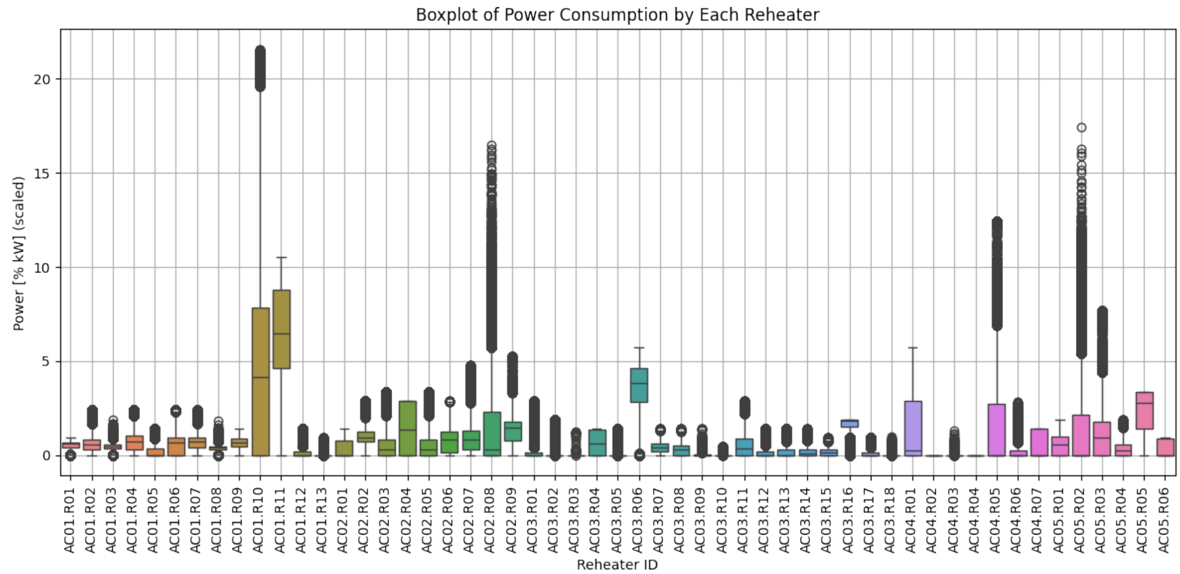


Figure D.6: Boxplot of Power Consumption by Each Reheater.

As stated before, reheaters in the HVAC system of the yacht account for 24.1% of total power consumption, presenting a significant opportunity for energy savings. [redacted]

Larger rooms such as the lounge, dining area, owner's stateroom, and bridge have the highest reheater power demands.

Replacing reheaters with waste heat heaters can significantly reduce HVAC energy consumption. It is recommended to prioritise this replacement in the larger, high-demand areas to maximise energy savings and improve overall efficiency.