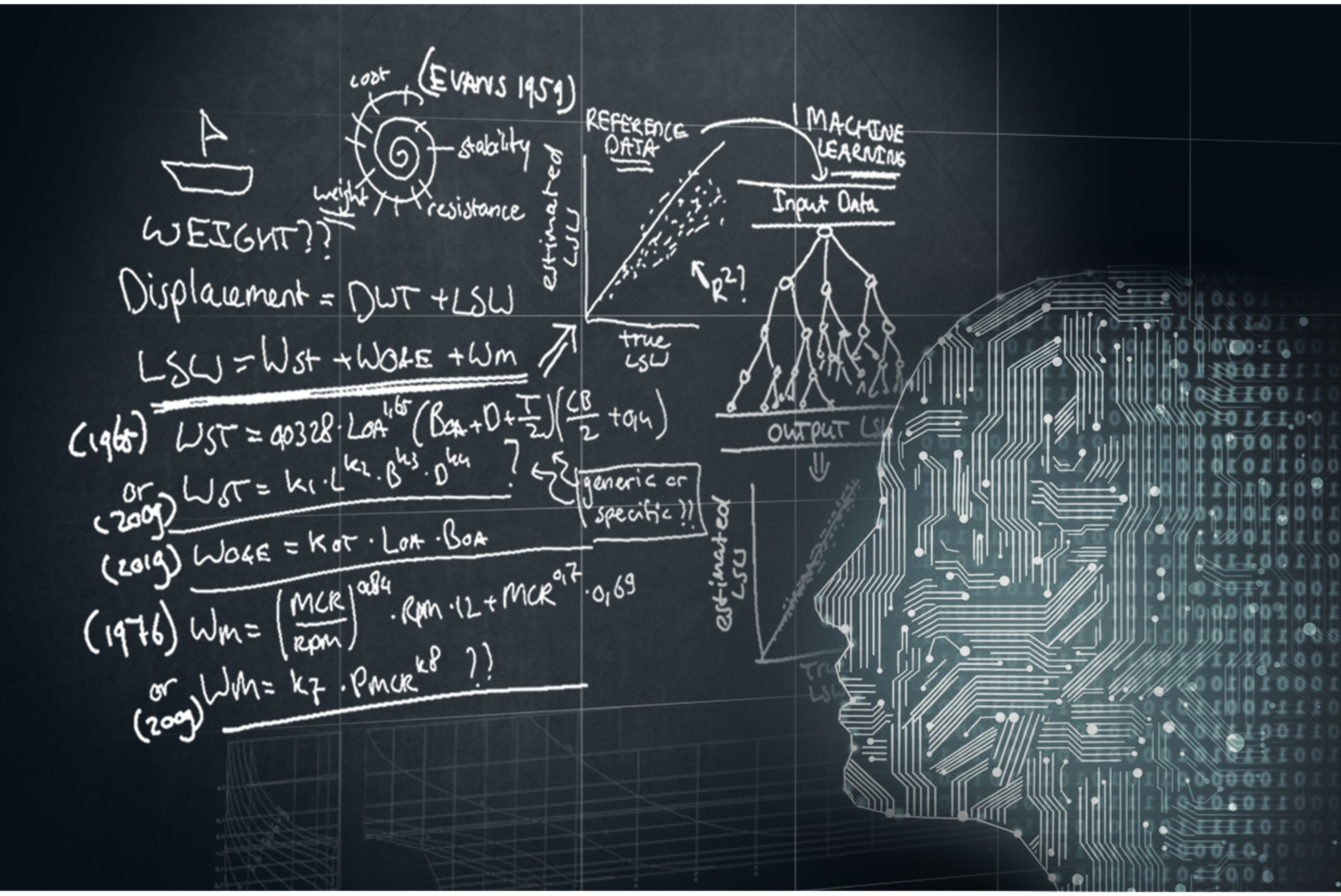
A Reference-based Design Approach

in Preliminary Ship Design

M.R. Bakker

SDPO.21.007.m



A Reference-based Design Approach

in Preliminary Ship Design

by

M.R. Bakker

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Wednesday March 31, 2021 at 01:00 PM.

Student number:	4168976	
Project duration:	May 29, 2020 – March 31, 2021	
Thesis committee:	Prof. dr. ir. J.J. Hopman,	TU Delft, Supervisor
	Dr.ir. A.A. Kana,	TU Delft
	Dr.ir. X. Jiang,	TU Delft
	MSc. R. de Winter,	C-Job Naval Architects

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

TU Delft

Abstract

Design decisions which are made in the preliminary ship design phase have a significant influence on the performance and total cost of a ship. These design decisions are mostly made with very little knowledge of the ship design problem. In fact, it is the personal experience of the naval architect which plays a significant role in this phase of the design process. In the world of rapidly increasing possibilities with artificial intelligence it is hard to imagine that these decisive design choices are based on a naval architect's personal experience. Especially when one takes into account the large capital and operational costs of ships.

The development of C-Job's Maritime Intelligence Tool (MIT) in 2019 has shown that reference data can better be exploited in this phase of the ship design process. As a result, theoretical reductions of the resistance and weight of the vessel go up to 19% and 10% respectively, using this tool. When the availability of reference data is limited, the trustworthiness of this tool cannot be guaranteed. This is especially unfavourable at the boundaries of a design space, as it is expected that novel and innovative ship design can be found here. Thus, in order to support naval architects in all regions of a design space, a solution must be found.

First, research is done into design approaches in the preliminary ship design phase. In this research, naval architects of C-Job with different backgrounds were interviewed. During these interviews it became clear that time, budget and customer ambitions are important motives in this phase. As a result, a lot of ship design decisions in the preliminary ship design phase are based on the naval architect's personal experience. The more the design is developed, the more insight is gained into the complexity of that ship design. As a result, more design decisions could be based on this gained insight, instead of the personal experience.

During these interviews, challenges were identified and discussed that the naval architects face, before the Maritime Intelligence tool can be used in practice. Based on these challenges a list of tool requirements was determined and potential solutions were sought. Three solutions were found to be promising for this thesis. These were serial hybrid modelling, parallel hybrid modelling and constrained black box identification. The parallel hybrid model is chosen, primarily because of the independent operations of the data-driven sub-model (Black Box model) and the knowledge-driven sub-model (White Box model) in a parallel hybrid model.

There are two requirements for parallel hybrid modelling. The first requirement is a method to estimate a ship design parameter. The second requirement is the availability of the true values of the same design parameter of reference vessels. These were both only available for the design parameter lightship weight. In the proposed parallel hybrid model, the white box model is used to estimate the lightship weight. Thereafter the black box model is trained to predict the difference between this estimation and the actual lightship weight, based on reference data.

The proposed parallel hybrid model is subjected to multiple experiments to assess the performance. The R^2-score and 10-fold cross validation are used to determine the performance. First the performance of the white box, black box and parallel hybrid model is discussed. Thereafter, the relation between the availability of reference data and the predicting capability is researched. The final experiment was to research the performance of the three different models in interpolation and extrapolation gaps.

Based on these experiments it was concluded that for a training data set of 50 reference vessels or smaller, the parallel hybrid model was the best model. For larger training data sets, the black box and parallel hybrid model performed similarly. For interpolation and extrapolation problems the white box model should be chosen. Additionally, a method is presented to update the used white box model. As a result, it is expected that high performance scores can also be obtained without the use of artificial intelligence tools.

Preface

As a little kid, I always loved to watch Discovery Channel's *How it's made?*, *Myth Busters* and *Mighty Ships*. On the first day that I went to the TU Delft with my father, a professor told us "If you have watched the following TV shows when you were young, than you probably belong here" and he mentioned these exact three TV shows. I already decided that I wanted to study at TU Delft, but the professor's words were so accurate, that I never questioned this decision again.

Now, quite some years later, I see how important this decision was for me. A 1000 words can't describe how I have experienced my time as a student and I will also not try to do so. The most important thing is that I have met a lot of incredible people along the way, with whom I share great memories. Because of them, I've had an amazing time as a student and I've developed from an 18 year old kid to a TU Delft Engineer.

The report that lies before you is my master thesis, where I have been working on the past months. During this thesis I learned a lot about data science. This brought me to quite some new territories and it provided my with some new insights on preliminary ship design. The result of this research is a new method for estimating the lightship weight in the preliminary design phase based on reference data, which outperforms any of the known estimation methods.

I am very proud of this result and I could not have done it without good supervision. Roy, thank you for supervising this research on behalf of C-Job Naval Architects. I really appreciate your directness, both when the project was not really moving forward and also when I made good progress. A data scientist versus a shipbuilder, that was sometimes quite challenging, but I enjoyed it! I think that, because of our discussions, we have shown that some ship design methods are very much outdated, but also that a ship design problem is not an easy formula to solve.

Thijs, your weekly enthusiastic calls in February 2020 definitely contributed in my choice for C-Job. As a working student in 2017, I had a great time at C-Job. Also then, we had meetings regularly, discussed some results and determined the focus for the coming weeks. These meetings motivated me very much and also during the graduation project these meetings really helped me moving forward.

Roy, Thijs and everybody at C-Job, thank you for the great time, your support and the opportunity that I was able to start my internship right in the middle of COVID. I know very few students that managed to get internship in the COVID period.

And last but not least, Austin, thank you for your supervision on behalf of the TU Delft. I have enjoyed the progress meetings we've had and I can say that I always was more positive after the meeting than before. Your feedback has always been very clear and constructive. Especially with the new working from home situation this had great value to me.

The last words of appreciation will go to my family. Mum, Dad, Dennis, Jordy, grandparents Hermanns and Bakker, thank you for your continuous and unconditional support. Because of you, I was able to enjoy my time as a student as much as I did. Especially this last year your support helped me a lot to complete this master thesis. Thank you for everything.

M.R. Bakker
Den Burg, March 2021

Contents

1
Introduction

In this master thesis a solution is sought to improve the preliminary ship design phase by expanding the capabilities of designing ships based on data of reference vessels. This research is conducted at C-Job Naval Architects. In this introduction first an overview is given of C-Job's projects. Secondly, the importance of the preliminary ship design phase will be explained as well as the potential improvements of this design phase using a new developed tool. Thereafter a problem statement will follow which has led to this research. Finally, the research questions will be presented and the lay-out of this thesis will be given.

1.1. About C-Job Naval Architects

C-Job Naval Architects (C-Job) is the largest independent ship design and engineering company in the Netherlands. The company has been involved in a broad range of sectors like dredging, heavy lift, offshore (wind), ferries and superyachts. In both new-build and conversions or modifications C-Job has played a leading or supportive role in the design process. This could be for the entire design process but also for a part of the process. Some recent projects are shown in Figures 1.1a and 1.1b. C-Job was responsible for the full design scope of the Atlantic Dawn 4400 DWT heavy lift vessel for CIG Shipbuilding. For the Texelstroom project, C-Job was responsible for the initial, concept and basic design. Initial and concept design can be considered as the preliminary design and basic design is also referred to as contract design. Multiple definitions are used in literature to describe the different phases of the design process. In Section 1.2, three design phases are defined that will be used in this thesis.

(a) Atlantic Dawn

(b) Texelstroom

Figure 1.1: Some recent projects of C-Job Naval Architects (from C-Job (2020))

1.2. Ship design process

The ship design process consists of three phases; the Preliminary Design Phase, Contract Design Phase and Detailed Design Phase. Each phase has a specific goal.

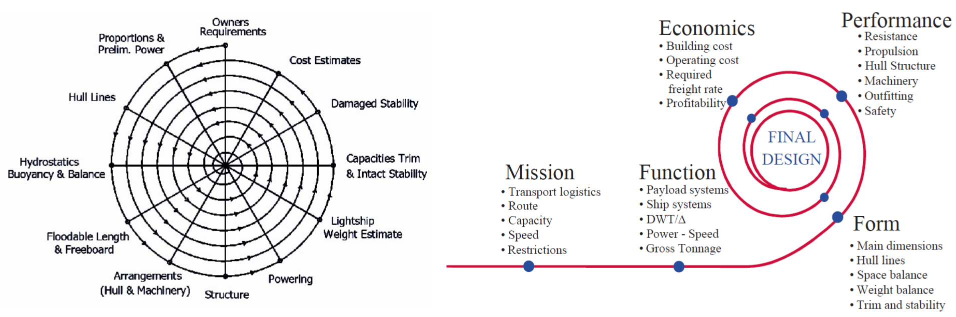
The goal of the **Preliminary Design Phase** is to find a balance between customer ambition (needs), available budget and possible design solutions [14]. Based on the requirements from the client, a concept exploration is performed. This is to explore possible design solutions and to get an idea of how design solutions relate to design requirements from the client. This is done to be able to select the most desirable design solution for the client. This selection is mainly based on technical and economical aspects of the design, but also company policy can play a large role [13].

This design solution is then worked out in more detail the **Contract Design Phase**. This means sufficient detail to describe a contract and determine a contract price [14]. Also the contracts of materials and equipment are discussed in this phase. This is done after a preliminary lay-out of systems has been designed. As the shipyard has its own suppliers of materials and systems, this more detailed design phase is mostly undertaken by, or in close co-operation with, the shipyard who assesses producibility and cost [14].

In the final phase, the **Detailed Design Phase**, the contract design is translated into a design definition that is suitable for production [14]. That means that production drawings have to be made for every detail. Therefore this phase claims a large part of the entire design process.

The preliminary ship design process can be described by the design spiral of *Evans (1959)* [17]. This is shown in Figure 1.2a. The design spiral starts with the general arrangement (G.A.) of the vessels as input for the design spiral, i.e. the main particulars of the vessel have already been chosen. This spiral addresses multiple aspects of the ship design during each iteration. After some iterations a final design solution can be determined by the naval architect.

A slightly different design spiral is presented by *Erikstad and Levander (2012)* [15] in Figure 1.2b. This design spiral is a more comprehensive presentation of the preliminary ship design process. In this spiral, determining the Mission and Function of the vessel is part of the design process, after which the general arrangement can be determined. This indicates that multiple general arrangements or design solutions can satisfy the requirements from the customer. As initial sizing is the most crucial phase in determining the overall configuration [3], a thorough design space exploration should be done in order to find a configuration which bests fulfils the (Mission & Function) requirements. The determination of suitable main dimensions, block coefficient and the arrangement concept is considered as the most important step in this phase [40]. *van Bruinessen et al. (2013)* [37] mention that with *Evans'* design spiral, the design choices which have the most impact are made before the design spiral commences. Therefore, a naval architect should use the design spiral provided by *Erikstad and Levander* where it is more likely that multiple different design configurations will be explored.



(a) Evans (1959)

(b) Erikstad & Levander (2012)

Figure 1.2: General ship design spirals

Generally, based on the requirements of the customer a few design solutions are explored by the naval architect. The customer can then choose a concept which will go to the next phase, the contract phase. It is also possible that only one design solution is explored, when the budget for a project is limited. This will be continued in Section 2.2. This means that the project success is very dependent on the experience of the naval architect. *Andrews (1998)* [3] describes that the manner in which an individual selects, creates or produces his initial ideas of the overall design can have a significant bearing on the end product. A designer's personal experience can thus have a significant influence on the end product and its success. The project success is defined by the performance and the total cost of the vessel. This is visualised by *Mavris and Delaurentis (2000)* in Figure 1.3. As can be seen, most of the performance and committed cost of the design are locked-in in the preliminary design phase. This is all done with little design problem knowledge. Therefore, it is essential to have good design methods in the preliminary ship design phase that support the naval architect in making the right decisions.

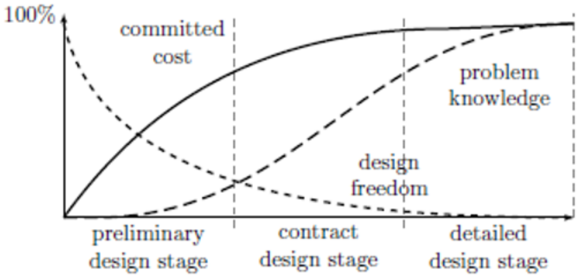


Figure 1.3: A generic design timeline (from [11])

1.3. Maritime Intelligence Tool

To support its naval architects, C-Job has an in-house developed marine reference tool called the Maritime Intelligence Tool (MIT). This tool consists of a database of around 170.000 ships and approximately 130 data fields. Its purpose is to analyse and compare the ship data from the database, with the ship data of a novel ship design. Using this tool, an analysis of existing ship designs can be done quickly. This provides insight in the initial sizing of the new vessel. Secondly, the results of calculations and simulations of C-Job's naval architects can be benchmarked with data from the database. A recent addition to this tool is a machine learning algorithm, which is able to predict ship parameters for a novel ship design based on the data of existing vessels.

The first step of this tool is reference vessel selection. This is done by the naval architect. Here, a type of ship or multiple types of ship can be selected. Corresponding data of these vessels is then loaded into a design space. This data consists of main particulars, engine power, lightship weight, deadweight, but also more specific parameters such as hopper volume. Secondly, based on data from these ships, polynomial regression models can be determined. These regression models will be used to estimate various parameters of a design solution.

Thereafter, a genetic evolutionary algorithm is used to find advantageous parameters to optimise design solutions. The proposed solutions of this multi-objective optimisation problem are then plotted on a Pareto-front. An example of this can be seen in Figure 1.4. In this figure the green dots represent 508 existing passenger vessels. First, a machine learning model is trained based on this data. After that, two objectives were determined for a new passenger vessel. The first is to minimise the Maximum Continuous Rating (MCR) and the second objective is to maximise the deadweight. The minimal MCR was chosen because there is no data available about the resistance of a vessel. But as lower resistance means a lower required MCR, it is sufficient to take MCR instead of resistance in this (low-detail) stage of the design process. A lower MCR means lower cost, because of lower fuel usage and lower engine purchasing cost. The other objective, to maximise the deadweight, or the payload, is chosen because a higher cargo capacity means more earnings per transport. Therefore, both objectives are important in the overall performance of the vessel. It can be seen that, considering these two objectives, the proposed solutions out-perform the existing vessels.

Figure 1.4: Existing vessels versus design solution proposed by the MIT

There are two important advantages of such a tool. It is less time-consuming compared to a naval architect doing the calculations. And secondly, this tool is able to explore large amounts of design solutions. Therefore it is more likely that a global optimal design solution will be found, instead of a local optimal design solution [30, 31].

The MIT provides an optimised design solution for a certain design problem by learning from previous designs. This can be done in a few minutes. The optimised design solution can be used as a starting point for the design process. In Figure 1.5 this starting point is described as the initial design. The difference with the traditional design method is significant. Currently, a starting point or initial design is based on the experience of the naval architect, or based on (a) reference vessel(s). This will be discussed in Section 2.2. Instead, the initial design can be based on data, calculations and optimisation techniques by using the MIT.

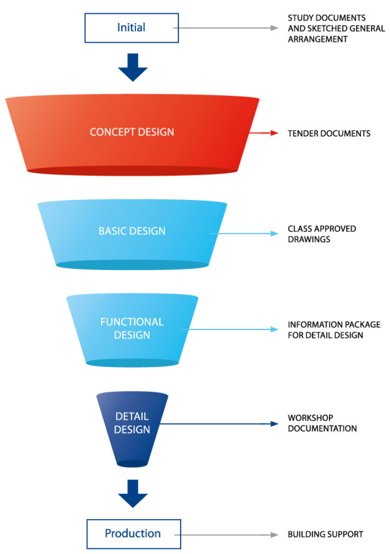


Figure 1.5: Design Phases at C-Job

1.4. Problem Statement

The Maritime Intelligence Tool from C-Job is able to predict advantageous ship parameters of a novel ship design based on the data of existing ships. As already mentioned, the tool uses polynomial regression models to determine the relations between ship parameters. These models are quite accurate when there is enough data in the design space. Unfortunately, for regions of the design space where data is limited, the regression models give incorrect solutions. In order to find novel design solutions, it is important to explore many regions of the design space, even though there are less existing ships in these regions. Therefore possible errors or gaps in the data should be sufficiently evaluated and dealt with.

A gap in the data can exist because of two reasons. The first reason is because of practical issues. For example, inland waterway ships have to deal with limitations caused by bridges and locks. Therefore one will find a lot of vessels concentrating in regions with same width, length or depth. In Figure 1.6 this can be seen. In this figure the Draught and the Breadth (Moulded) can be seen of crude oil tankers. The ships in the upper right corner of this graph are limited by the Panamax dimensions. In this graph the Panamax maximum width of 32.3 can be seen. This is a popular region of the design space. In between these popular regions one can find a gap in the design space, with the result that the polynomial regression models are inaccurate in this region.

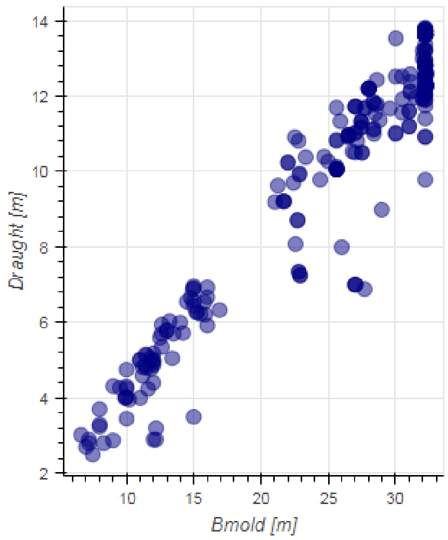


Figure 1.6: Gaps in data

The second reason is that a region is located at the boundaries of a design space. These areas can be considered as more challenging design solutions. Novel ships with a specific purpose are often found on the boundaries of the physical possibilities and thereby on the boundaries of the available data (i.e. the empty regions in the design space). In order to explore novel design solutions it is important that the tool also produces accurate results at the boundaries of a design space.

1.5. Research goal and focus

The goal for this research is to develop a method that deals with gaps in the design space and helps make predictions more accurate in those regions. With this method the MIT should be much more trustworthy and usable for the naval architect. Also, the integrity of the proposed design solution towards the boundaries of the design space can thus be ensured and exploited.

It is expected that a method that supports the current Maritime Intelligence Tool with additional data or known formulas will improve the results of the MIT. It is essential that the feasibility of these results or design solutions is ensured. Thereafter, the naval architect should be provided with sufficient knowledge of the presented design solutions to make well-advised design decisions.

To support this research goal, the following main research question is determined.

"How can a ship designer better explore gaps in a design space, generated using data of reference vessels, to make predictions of the main parameters of a novel ship design solution more accurate?

The following sub-questions will support the main research question and will also form the basis for the plan of approach.

1. How does a naval architect make well-advised design decisions in an early stage of C-Job's current design process?
2. How can this be done in C-Job's future reference based design approach?
3. What are the important gaps in the design space?
4. What are leading principles in designing a model that deals with these gaps?
5. How can these principles be converted into a model, that improves the quality of the design solutions produced by C-Job's current Maritime Intelligence Tool and using its database as a design space?
6. How can one determine if a design solution is feasible and optimal?

1.6. Thesis lay-out

This section describes how this thesis is structured. In Chapter 2 the design approaches are discussed. These are approaches that have been described in literature or approaches that are currently used in practice. Also C-Job's vision for its future design approach will be discussed together with the challenges that lie ahead. In Chapter 3 a list of requirements will be presented for the new tool. Thereafter several potential solutions will be explained and evaluated based on the requirements. One solution will be chosen and further developed into a model. How this model works exactly will be described in Chapter 4. This chapter will also describe some performance metrics that are used to evaluate the model. Chapter 5 describes the experiments that have been done and the results of these experiments. These results will be discussed quantitatively and qualitatively. Lastly, in Chapter 6 the main research question will be answered and a conclusion of this research will be given. Also, the contribution of this research, the limitations of the research and recommendations for future research are given in this chapter.

2
Design Approach

This chapter aims to answer the first research question. First, preliminary ship design methodologies from literature will be described. From this, information will be derived which is considered important for a naval architect to make well-advised decisions in an early stage of the ship design process. Section 2.2 then discusses how preliminary ship design is done at C-Job. This is done for the current design approach, but also for the future approach. For the future approach there are still some challenges to overcome. In Section 2.4, additional literature will be discussed that could be useful in this research. After this, a conclusion will be drawn about the main challenges for this research.

2.1. Preliminary ship design approach in literature

The goal of the preliminary design phase is to find a balance between customer ambition (needs), available budget and possible design solutions [14]. After the requirements of the client are set, naval architects start developing technically feasible design solutions based on the operating environment and the requirements. In *Molland (2008)* [28] a very brief description of the design process of a ship is given (See Figure 2.1).

To determine whether or not a design can be considered as technically feasible, *Molland (2008)* describes a preliminary design path This can be seen in Figure 2.2. As can be seen, in a ship design process, several aspects of a ship design are covered sequentially. At several design aspects a feedback arrow can be seen. Here it is checked if the design meets the requirements or that the design needs to be adapted. Once a design meets the requirements for all these design aspects, a design can be considered as technically feasible. Other methods like *Andrews (1998)* [3], *Molland (2008)* [28] and *Papanikolaou (2019)* [29], are about the same.

According to *Molland (2008)*, a design is considered technically feasible if the design meets the following principle requirements [28]:

1. **Adequate in size and arrangement for intended service -** Implies ability to carry a specific ed volume of cargo and have adequate space for machinery, fuel and crew etc.
2. **Floats at correct draught -** Implies sum of weights of lightship and deadweight equals force due to buoyancy (function of ship form)
3. **Floats upright -** Implies adequate stability
4. **Achieves correct speed -** Implies satisfactory estimates of resistance and propulsive power (plus margins) and installation of suitable engine(s).
5. **Is structurally safe/ sound -** Implies structural design with the ability to withstand forces in the marine environment; typically built to the requirements of a classification society
6. **Meets requirements for manoeuvring, coursekeeping and seakeeping -** Implies choice of suitable hull form

7. **Meets international standards of safety and reliability -** Meets requirements of IMO

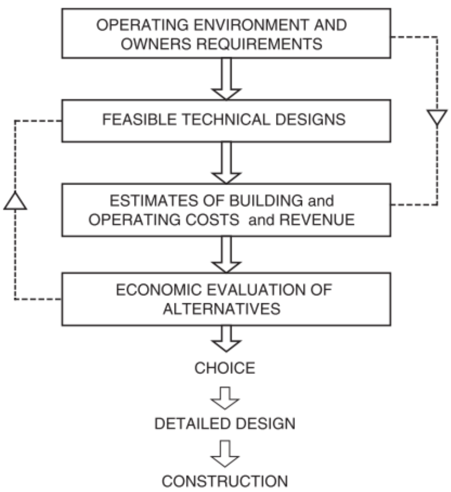


Figure 2.1: Overall flow path ship design process (from [28])

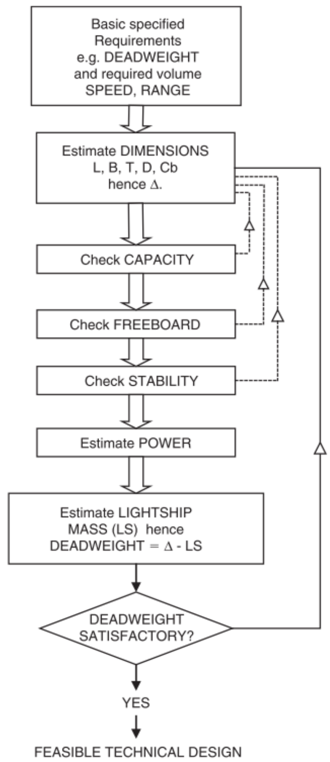


Figure 2.2: Preliminary Design Path (from [28]

Empirical versus parametric methods

With an approach as described in Figure 2.2 the main dimensions and the basic form characteristics can be determined. This can be done with two methods, the relational (or empirical) method and the parametric method [29].

In the **empirical** method estimations of the main dimensions is based on data of similar built vessels. *Papanikolaou (2019)* [29] describes that a variation of this method is the use of empirical design formulas deduced through regression fitting. Interpolation between comparative data is in general seamless, whilst extrapolation may prove problematic; unless for small exceedance of boundary limits it is possible.

Watson (1998) [40], *Schneekluth and Bertram (1998)* [18], *Andrews (1998)* [3], *Molland (2008)* [28] and *Papanikolaou (2019)* [29] all describe such empirical design equations. With these equations the naval architect can rapidly estimate certain values in the preliminary ship design phase. The variables in the equations usually have a physical justification, and the relationships will be ‘ calibrated ’ for different ship types. An interesting remark is made by *Watson (1998)*. He mentions that some relations have an economical justification (e.g. the ratio L/B), instead of a physical justification (e.g. the ratio’s B/D and T/D).

As empirical relations are based on knowledge which is derived from previous work, it is important to handle these relations carefully. It is the duty of a naval architect to update such empirical relationships whenever possible [28]. Therefore, the date of publishing should be taken into account when using these equations.

When comparative data from similar ships is lacking, a study needs to be conducted from scratch to find the best main dimensions and form characteristics. In this case the naval architect builds a **parametric** model in which different aspects of the design have been taken into account. In such a model the ship's main design parameters are rationally related to the ship's performance (physical and economic characteristics). This multi-objective model is then optimised to determine the desirable main dimensions and form characteristics. [29]

Deadweight versus capacity determined designs

A distinction is made in literature between deadweight determined designs and capacity / space determined designs. A deadweight design approach is determined by the weight of the cargo. This is typical for oil tankers, bulk ore carriers and most cargo vessels. For a capacity design approach, the dimensions of the vessels are primarily determined by the volume of the cargo. This is the case for container vessels, ferries and most naval vessels. [3, 28]

In the deadweight design approach in preliminary ship design the deadweight, range and speed are considered as the main requirements. Capacity, stability, freeboard and others are treated as 'checking' or constraint requirements [28]. In the capacity design approach the capacity and the deadweight are swapped. *Molland (2008)* [28] gives design strategies for all of these requirements in his book. This includes empirical formulas and ship type specific values, such as steel mass divided by deadweight.

2.2. Preliminary ship design approach at C-Job

This section addresses the preliminary ship design phase in practice at C-Job. First the current design methods that are used by naval architects (NA) will be discussed in Section 2.2. These methods are derived from interviews. These 30 minutes to 1 hour interviews were held via a video-call. The content of this section has been verified by the naval architects. This section aims to answer research question : "*How does a naval architect make well-advised design decisions in an early stage of C-Job's current design process?*". Secondly, the future methods will be discussed in Section 2.3. Due to recent research it is clear that current preliminary ship design methodologies can be partly automated and improved.

The interviews were carried out with the following naval architects from C-Job.

Name	Function	Years of experience	Date of communication	Reference in text
de Vries, N.	Lead NA	6	2020, June 4	[1]
van den Ing, A.	Lead NA	15	2020, June 17	[2]
Houwaart, K.	Lead NA	6	2020, June 19	[3]
Frontera, R.	NA	5	2020, July 23	[4]

Table 2.1: Background information interviewees

Determine starting point

To get insight in the main dimensions of the new vessels multiple methods are used at C-Job, namely:

- Perform reference study and determine trend lines
- Start with requirements and perform own calculations
- Study literature and select relevant equations
- Adapt a convenient reference vessel

In the first method, a reference study is conducted to determine trend lines and variances [1,3]. This reference study is done in RefWeb, the database of C-Job. Using these trend lines the first parameters can be estimated. Depending on the experience of the naval architect and the expected uncertainty of some parameters, an estimation of the parameters could be done conservatively, optimistically or somewhere in between.

An other method is to start with one or more of the requirements from the customer [2,3]. For example, the payload that the vessel should be able to transport and the stability that needs to be guaranteed. With basic calculations, that are mainly based on rules of thump and the experience of the naval architect, the main dimensions can be determined as well.

A third method is to study literature [4]. An advantage of studying literature instead of reference vessels to design a new vessel, is that the new design is not limited by what has already been done before. In literature a lot of empirical formulas are given that can be used in preliminary ship design. These equations can be based on a certain ship type, but they can also be based on a data-set. This last feature means that one can customise a certain parameter based on a set of selected reference vessels. These empirical formulas can already give the naval architect an accurate idea of the magnitude of these parameters in the early stage of the design process.

A final method, which can be considered as a low-budget method, was to take one reference vessel as a starting point and to adapt this design until all requirements for the novel design are met [4].

A conclusion can be drawn about these different methods for determining the main dimensions of the design. Performing a thorough concept exploration (i.e. exploring the entire design space) is too expensive in terms of time, but is sometimes required. This results in the fact that most of the time one, the most promising, design solution is worked out in more detail. The method that is used to do this highly depends on the background and skills of the naval architect. This results in four main approaches which are used in practice; reference study, own calculations, empirical formulas and adaptation of a reference vessel.

Design strategy

Thereafter the naval architect adapts the initial design in an iterative process until all the requirements from the customer are met. As explained in Section 1.2, this iterative process can be described by the design spirals of *Evans (1959)* [17] and *Erikstad and Levander (2012)* [15]. In practice, this iterative process depends on the requirements and the type of vessel [2,3]. For example the hopper volume is important for a dredging company. Therefore, in the design process of a dredger the first requirement to meet is the hopper volume [3]. For a passenger vessel an important requirement is the seakeeping of the vessel and therefore the hydrodynamics should be evaluated in an early stage of the design process [4]. Also the different load-cases that are applied to a vessel should be taken into account [2]. The load-cases can differ significantly for a heavy lift vessel with a crane for example. This influences the stability of the vessel. A last example of this is given by *Watson (1998)* [40]; the cargo handling of a Ro-Ro ship must be considered at an early stage in the design process, whereas most aspects of the cargo handling of a tanker can be dealt with quite late in the design process. Thus, it is up to the naval architect to understand which design requirements are critical for a ship design and to determine which design strategy best fits this design problem.

This strategy depends on the experience and skills of the naval architect. *Andrews (1998)* [3] describes this as the nature of the designer's personal 'stamp' which is relevant to the overall design process. This differs amongst the naval architects at C-Job. Some naval architects choose to build a model right away. This is done in NAPA [3] or DELFTShip [4] for example. NAPA can be used for hull form design and stability analysis, but also structural design. DELFTShip focuses more on hull modelling and stability analysis. Other naval architects did their calculations in EXCEL from which a model can be derived.

Thus, the design method that is used in this iterative design process depends on the type of ship, the requirements from the customer and the experience and skills of the naval architect. That means that design methods can differ significantly per project.

Determine final design

At some point in the iterative design process the naval architect determines that a final design has been reached. During the interviews it was clear that once all the requirements from the customer are met, a final design is reached [2,3,4]. A design could be further optimised, if the clients budget would allow it [3,4]. Also the evaluation of a design solution is done in this way. If the requirements are met, and the calculations are correct, then the design is considered as a feasible design [3]. Based on insight that is acquired during the project, a naval architect can determine that a design is both feasible and desirable [1,2,3]. This is acquired by evaluating the requirements from the customer[2,3], performing own calculations [1,3] and from iteratively developing the design [3]. From these activities a naval architect gets a feeling of how requirements and possible design solutions are interrelated.

Also the assessment of a design solution was based on the experience of the naval architect. All naval architects wanted insight into the design problem by looking at the calculations. Different aspects of the design were evaluated, because of personal background and interest. These different aspects were power generation and distribution [1], stability[2] and seakeeping [4]. For example, the installed power can be estimated from a reference study and trend lines. To determine if such an estimation is reasonable, one should look at the power demand from the main components. If the distribution between propulsion, systems and margins is reasonable, then a naval architect can determine that a design is feasible [1]. The same goes for stability, which has already been explained. The naval architect can only determine whether the stability is sufficient,when insight has been gained about the load-cases that are taken into account [2].

A conclusion can be drawn about assessing design solutions and selecting a final design solution. For both assessing the feasibility and selecting a design solution applies, this is done on the basis of insight which is gained during the design process. When a design is feasible and all requirements from the customer have been met, then a design is considered as the final design solution.

Conclusion

After conducting the interviews with naval architects from C-Job, it was clear that the design process is highly dependent on the experience of the naval architect. This determines how one chooses a starting point for the design process and also what the design strategy will be. A final design is found when all the requirements from the customer have been met. The feasibility of this design is determined by the naval architect, which is done on the basis of own calculations and modelling. To reduce time and cost, the naval architect focuses on the most promising concepts, rather than to fully develop multiple concepts and to compare the outcome [2]. Although understandable, this can be seen as a disadvantage, as it is likely that innovative or better design solutions are missed using the current design methods.

After Sections 2.1 and 2.2, where preliminary design methods in literature and in practice at C-Job have been discussed respectively, the first research question can be answered.

Research question 1

How does a naval architect make well-advised design decisions in an early stage of C-Job's current design process?

In the preliminary design phase the naval architect needs an understanding of the requirements and their effect on a design solution. As requirements are often conflicting, it is this understanding which aids the designer in making appropriate trade-offs. Based on the requirements and the personal experience of the naval architect a design strategy is chosen that best fits a certain design project. Four methods have been described that are used to determine a starting point of the design process:

- Perform reference study and determine trend lines
- Start with requirements and perform own calculations
- Study literature and select relevant equations
- Adapt a convenient reference vessel

These different methods all represent a different level of how wide and thorough the initial search for a design solution is. As is concluded after some interviews, most of the naval architects at C-Job start their design process with an already focused initial search, which is directed towards the most promising areas of a design space. By performing own calculations the naval architects gain insight into the design process and the design itself. Literature describes a lot of empirical and parametric calculation methods that can be used in this stage of the design process. Next to gaining insight, based on these calculations and by comparing the results with reference vessels, the naval architects can determine if a design solution is feasible or not. In literature some requirements have been described to determine the feasibility of a design solution. To conclude, every naval architect determines his own design strategy. This will lead to different design processes and different designs, depending on who is designing. Also the limited time and budget contribute to this.

2.3. Future design approach at C-Job

As mentioned in Section 2.2, exploring the entire design space is too time-expensive. Therefore only a few design solutions are explored in the current situation. The current method is based on traditional design spirals presented by *Evans (1959)* [17] and *Erikstad and Levander (2012)* [15]. A naval architect performs each calculation by hand and after a few iterations a possible design solution has been reached. To find an optimal design solution it is essential to explore a significant part of the design space. But as there are many variables in a ship design the current method quickly becomes impossible for a naval architect to do by hand. C-Job has developed the Accelerated Concept Design method [31] to solve this process more efficiently.

Accelerated Concept Design

The Accelerated Concept Design (ACD) methodology is represented by the design circle in Figure 2.3. In this method the design spiral is replaced by a circle with multiple layers. These layers represent a level of accuracy. The more one enters the centre of the circle, the more accurate the calculations will be. On the other hand, the computational time increases significantly with the accuracy of those calculations. An example of the meaning of these levels is given for the resistance; Level 1 - Holtrop Mennen estimates, Level 2 - Potential flow calculations, Level 3 - Viscous flow calculations, Level 4 - Model testing in towing tank.

The ACD circle represents a holistic design methodology that deals with decision variables, constraints and objectives simultaneously. This methodology is different compared to the traditional design spiral, where different aspects of the design are addressed sequentially. With the use of automated software and a multi-objective optimization algorithm it is possible to make decisions about these design aspects simultaneously. *De Winter (2019)* [31] mentions that this algorithm considers the entire design space which makes it more likely that a globally optimal solution will be found. This is clearly an advantage compared to the current design methods that are described in Section 2.2. Also *Parsons (2009)* [30] mentions that such an algorithm has the major advantage that they can have a very high probability of locating the global optimum and not just one of the local optima if they are present in a particular problem. The algorithm which is used by *De Winter* is the Constrained Efficient Global Optimization (CEGO) algorithm [41]. This algorithm is designed for optimizing multiple objectives at the same time.

De Winter (2019) [31] mentions that using such a holistic multi-objective optimisation method, de-

Figure 2.3: Accelerated Concept Design Circle [31]

signs can be improved significantly. In his paper an experiment is described where a Trailing Hopper Suction Dredger (THSD) design is optimised by the ACD framework. In this experiment the generated design solutions had a Level 2 accuracy. The proposed design solutions are then compared with the original design. The results are plotted in Figure 2.4. In this plot the green dots represent the optimised design solutions. One can see that the optimised solutions out-perform the original design in terms of resistance coefficient and steel weight. The objectives for this optimisation problem were to minimise both the resistance and the steel weight. The marked designs, which are the original design and the most interesting optimised design solution, are compared. The proposed solution had a 19% smaller resistance coefficient and a steel weight that was 14% less compared to the original design. This shows the potential of the holistic multi-objective optimisation method.

Figure 2.4: Results of optimization experiment by the ACD framework [31]

Maritime Intelligence Tool

The MIT is an elaboration of the ACD methodology. Looking at Figure 2.3 the MIT can be seen as a tool in the outer layer of the design circle, Level 1. As mentioned, the multi-objective optimisation algorithm considers the entire design space when starting a new design project. As calculation time rapidly increases with each level it is not possible to evaluate every design solution with a high level of accuracy. Therefore it is essential to have accurate estimation methods, belonging to an outer (high speed) level of the design circle. With accurate methods and high calculation speed it is possible to explore and evaluate a large amount of design solutions. This makes it more likely that a globally optimal design solution is found. In this way, the MIT provides a design solution to the naval architect which has already been optimised for certain design objectives.

There also some disadvantages of such a tool. The MIT can be considered as a black box model. A black box model extracts knowledge directly from data, with a few assumptions about the true underlying process behaviour [5]. Therefore this is an empirical method. *Duarte et al. (2004)* describes two disadvantages of a black box tool. The first is that such a purely empirical method only permits limited extrapolation beyond the domain of the data from which they were derived. The second is that any mechanistic knowledge that may be available about the process and its underlying physics are ignored, thereby potentially resulting in unreasonable results. *Duarte et al.* propose a method with increased extrapolation capabilities which takes into account any available (mechanistic) knowledge. This solution will be discussed in Section 3.2.

Challenges faced by C-Job naval architects

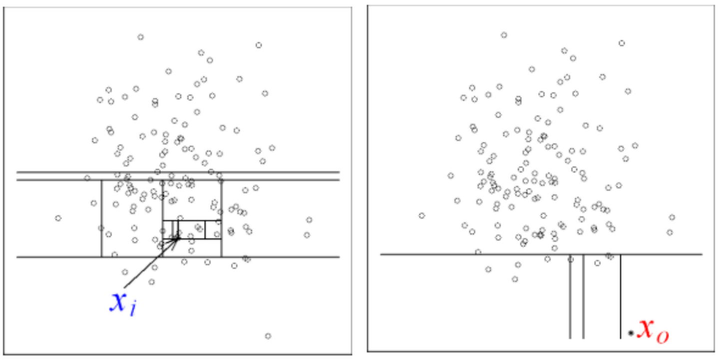
Despite the potential of the MIT, there are some challenges to overcome before the tool can be used in practice. These challenges are pointed out by naval architects from C-Job during interviews. The statements made in this paragraph have been verified by the naval architects.

First, the proposed solutions of the tool should comply with the laws of physics, basic principles of naval architecture and other governing ship design rules and regulations. An example of this is the law of Archimedes in formula 2.1. Currently, for some design solutions this equation is not satisfied. The reason is that such formulas or knowledge are not included in the MIT method. In the current MIT method the displacement of a ship design is not calculated with this formula, but it is predicted. Therefore, a method should be found that ensures that such equations are satisfied, so that the proposed design solutions comply with the rules of physics.

$$\nabla = L \cdot B \cdot T \cdot C_b \tag{2.1}$$

Secondly, the lack of used equations and calculations in the MIT-approach also limits the naval architect's ability to asses a design solution and its technical feasibility [1,2,3,4]. As mentioned in Section 2.2, it is essential to have insight in the design process in order to assess a design solution. This means assessing whether a design solution is technically feasible or not, but also to determine how well requirements from the customer are met. Currently, the MIT performs a feasibility assessment based on an isolation algorithm. The idea behind this algorithm is that deviating data points, or anomalies, are more susceptible to isolation and hence have short path lengths [21]. A 'path length' can be interpreted as the number of criteria one has to determine in order to isolate a data point. An example can be seen in Figures 2.5a and 2.5b. Each of the horizontal and vertical lines in these figures can be seen as a criterion to separate the data. It is clear the data point x_0 is easier to isolate. Based on its short 'path length' this data point is classified as deviating and therefore as infeasible. Deviating design solutions are isolated using this algorithm. However, a deviating design solution can be feasible. An example of such a ship can be seen in Figure 2.6. One can imagine that because of the large breadth (approx. 70m.) of the *Ramform Titan* compared to its length (approx. 104 m.), this design deviates from other designs. Thus, such a design is easily isolated, and thereby incorrectly classified as infeasible, using the isolation algorithm method. Hence, a disadvantage of this method is that innovative design solutions, which are per definition deviating from other designs, are easily filtered out.

Also some remarks were made about the presentation of the results. The results are currently plotted on a Pareto-front. The interpretation of such a result should be done carefully [2]. For example, the two objectives 'minimise resistance' and 'minimise steel weight' have already been mentioned in this report. An optimal design solution is not only about minimal resistance and minimal steel weight. A design solution with minimal steel weight and resistance could be interpreted as the most basic version of all design solutions [2]. That means that the design has no additional features like a crane or extra living space



(a) Isolating data point x_i (b) Isolating data point x_0

Figure 2.5: Different path lengths using the isolation algorithm (from [21]

Figure 2.6: Ramform Titan (from gcaptain.com)

for the crew. For some features, this problem could be solved by redefining the optimisation problem. At the reference selection step, the naval architect can select vessels that have a crane on-board, as the crane-option is one of the data fields in the database. In this way, the MIT takes into account the extra space and steel weight that come with the crane-option. For extra living space for the crew on the other hand, this is more difficult. This is because in order to say something about such features, a 3D model is often required. One can imagine that building a 3D model of every design solution in the design space is impossible as computational time would increase significantly.

2.4. Other relevant literature

In this section, other literature that is deemed relevant for this thesis will be discussed. This is literature regarding mathematical methods that improve the preliminary ship design phase. Mathematical methods are chosen as these make it possible to explore a significant amount of design solutions within a reasonable time. Together with the conclusions of Sections 2.1, 2.2 and 2.3, a gap in literature and current available methods will be pointed out.

Artificial intelligence definitions

In this section a brief introduction will be given on artificial intelligence and its different methods. In Figure 2.7 these methods can be seen as well as their relation to each other.

All techniques that try to mimic human intelligence are considered as artificial intelligence (AI). This can be done using logic, if-then rules, decision trees and machine learning (ML) for example. Other than logic, if-then rules and decision trees, machine learning does not make use of predetermined rules.

A machine learning model structures data, learns from the data and then applies what it has learnt to make informed decisions. Using statistical techniques this model is able to improve tasks with experience. Although a ML model is able to improve itself, it only looks at data in the way it is programmed to do so. This is a difference with deep learning. In deep learning (DL) an algorithm adapts itself, when exposed to different patterns in data. As deep learning is based on neural networks, it takes a lot of processing power to get trained. Secondly a lot of training data is needed to feed a neural network. [20]

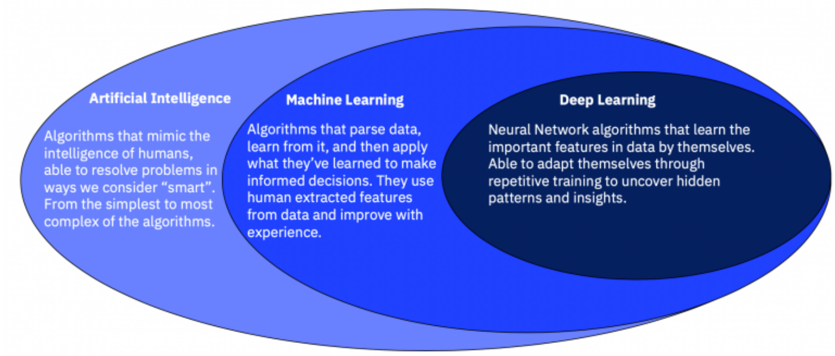


Figure 2.7: Artificial Intelligence (from IBM (September 2020))

Determination of main particulars based on reference data

In *Claussen et al. (2001)* [10] a method is described how the determination of the main particulars can be eased in the initial ship design stage given a type of ship and a few parameters. Empirical relations are derived from a database of 87.000 ships, using three different methods. In the first method a simple regression analysis is carried out to fit functions to statistical data. In the second and third method, the data is used to learn a Bayesian network and a neural network, respectively, to encode the relations between the characteristics.

A case study is done for the determination of the main parameters of a container vessel. Based on data of existing container vessels, relations are derived that relate TEU capacity to a main characteristic. For the simple regression and the neural network this relation is given in terms of a continuous function based on the least squares method. For the Bayesian network the relation between parameters is given in terms of a network topology and corresponding probability tables.

Claussen et al. [10] show that the neural network has the smallest average percentage of error of all three methods. However, due to the scattered and sparse data in a range, the prediction of parameters differs for the three methods. In such a range, for example at the boundaries of a data set, the neural and Bayesian network have the best predicting capabilities.

A disadvantage of a neural network is that it must be trained for each combination of input data separately [10]. This means that neural networks need a lot of training data. As the MIT has to deal with regions of the design space, where (training) data is limited, this is undesirable.

A Bayesian network is a probabilistic description of the problem. Such a description can be seen in Figure 2.8. In this network, the characteristics of a container vessel are predicted based on the cargo capacity (TEU). In a Bayesian network each variable is discretised and each interval holds an equal amount of ships. As a result, dense parts of the distribution is finely discretised, whereas sparse regions are represented by fewer and longer intervals [10]. This means that the accuracy of the Bayesian network is lower in these regions. To obtain a high accuracy, also for the Bayesian Network, a lot of (training) data is needed.

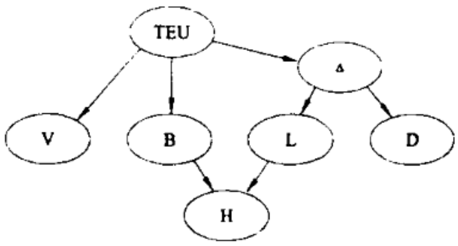


Figure 2.8: Bayesian network for a container vessel (from [10])

Optimisation methods in preliminary ship design

In *Zalek (2007)* [43] a hull form is optimised for both smooth water powering and sea-keeping performance using an advanced evolutionary algorithm. In this research an initial design is provided by the U.S. Navy's Advanced Surface Ship Evolution Tool synthesis model (ASSET, 2005). This design is then optimised for smooth water powering and sea-keeping performance, two conflicting criteria. To maintain the validity of the parent ship analysis performed by ASSET and to make sure that the final hull would still meet the mission effectiveness provided by ASSET, a maximum allowable deviation was set for hull form parameters and variables [30].

First the objectives are transformed into mathematical minimisation criteria. For the power minimisation criterion, formula 2.2 is used for example. The actual meaning of these parameters is irrelevant, but what is important is the following: this formula is predetermined by the user and the actual values are to be derived from solving the optimisation problem. This is a clear difference compared to the current method of the MIT, where such formulas are not predetermined. In the current MIT, for every objective and constraint a machine learning model is trained based on reference data. This model is significantly more complex than formula 2.2, as more variables are taken into account.

$$F_{PWR}(x) = w_1 \frac{P_{BEreq}(x)}{P_{BEreq}(x_o)} + w_2 \frac{P_{BSreq}(x)}{P_{BSreq}(x_o)} + w_3 \frac{V_{max}(x_o)}{V_{max}(x)} \tag{2.2}$$

The optimisation problem that was used can be seen in formula 2.3. Five criteria are used, from which two are based on the two objectives; minimal required power (F_{PWR}) and minimal inoperability (F_{SK}). One of the additional criteria was a penalty term to force weight to equal the displacement. In this way, Archimedes law is always satisfied.

$$minF(x) = min\{F_{PWR}(x), F_{SK}(x), D(x), H(x), G(x)\} \tag{2.3}$$

Solutions are obtained using a multi-criterion evolutionary algorithm. That means that an initial population was randomly generated. The highest ranking design solutions were then added to an archive. Via mutation and a survival of the fittest method new non-dominated solutions were derived. With the final non-dominated solutions a Pareto front can be approximated. A global optimal design solution is determined by finding a design solution nearest to the Utopian design solution. This can be seen in Figure 2.9.

Using this method a global optimal design solution is found within the created design space. This design space is already a specific region of a wider design space as it is based on the (detailed) design solution provided by ASSET and taking into account a 15% allowable deviation for certain variables and parameters. This limitation was imposed to provide assurance that the final hull would still meet the mission effectiveness provided by ASSET [30].

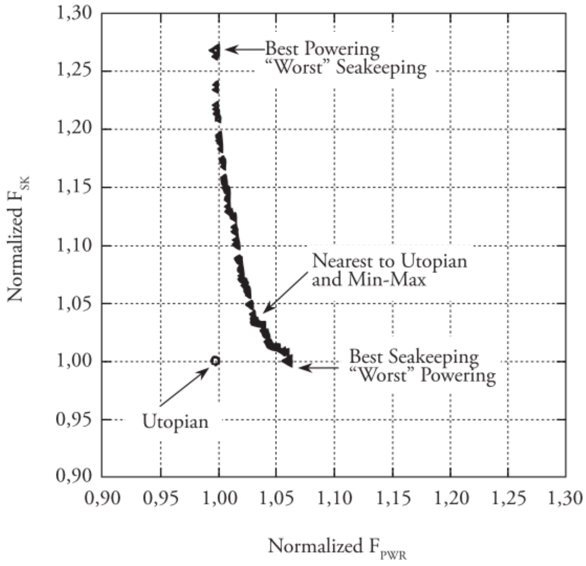
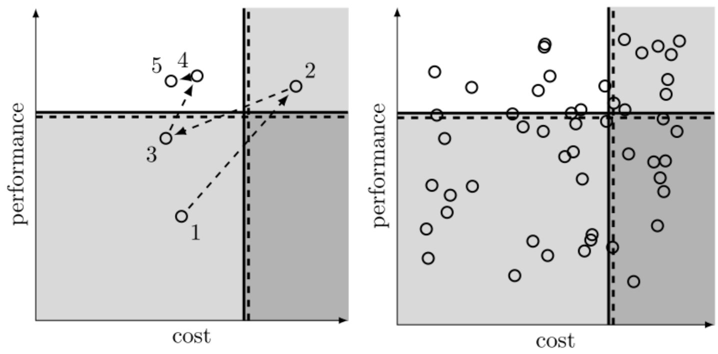


Figure 2.9: Pareto front in normalised criterion space (from [43])

Duchateau (2016) [14] describes a method for interactive evolutionary concept exploration in preliminary ship design phase. The goal of this research is to find better design solutions in the preliminary ship design phase, by performing an interactive concept exploration, where human and algorithms work together. Because of this interactivity, the human expert is able to gradually build up knowledge about the design problem and adjusts or expands the requirements accordingly. A search algorithm or optimisation algorithm can then refine its search for interesting design solutions.

In this thesis, first a distinction is made between sequential (point-based) exploration and concurrent (set-based) exploration. These two methods can be seen in Figure 2.10. *Duchateau* mentions that advances in computational power and ship synthesis models have shifted concept exploration methods from point-based approaches, where only few design solutions can be generated and explored, to more automated concurrent approaches, where many solutions can be generated and assessed simultaneously. The proposed progressive method is both sequential and concurrent. Concurrent in the way that design solutions are being generated simultaneously and sequential because, after one round of exploration, requirements are adjusted and an algorithm runs a second sound of exploration. This will continue until the most desirable design solution is found. An example of this is shown in Figure 2.11. As can be seen a wide area of a design space is covered in the initial global search. Then, in two steps, this search is narrowed down to a local search. Through human evaluation and feedback, requirements can be adjusted or added after the initial global search.

The implemented genetic search/optimisation algorithm is the NSGA-II algorithm by *Deb et al. (2002)* [1]. This algorithm is used together with constraints and objectives to search for and generate design solutions [14]. *Duchateau* mentions that the randomness introduced by the search algorithm's genetic operations, ensures sufficient diversity in overall size and arrangement of the vessels. This contributes in generating a large and diverse set of design solutions from which a designer can explore and pick interesting options. This corresponds to the goal of this thesis.

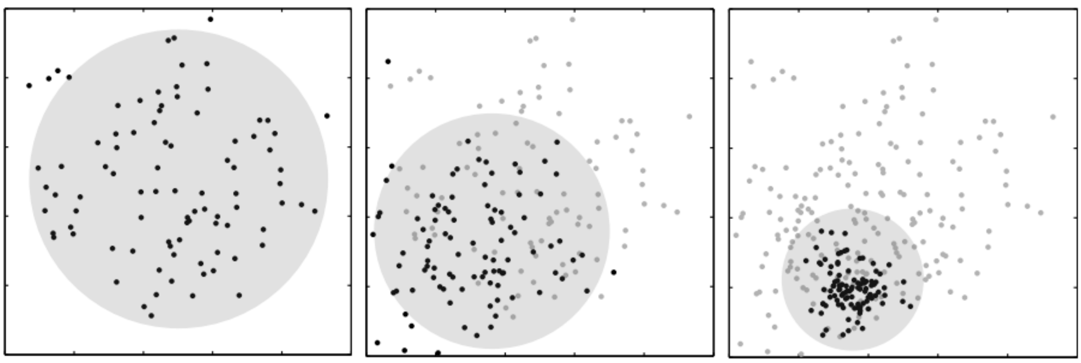


(a) Sequential (point-based) exploration (b) Concurrent (set-based) exploration

Figure 2.10: Two main approaches to concept exploration (from [14])

An advantage of the progressive search method can be related to one the challenges that is mentioned in Section 2.3. To asses a design solution it is essential to have insight in the design process. A progressive method makes use of insight which the user or decision maker has gained during the search algorithm's progress [14]. Therefore, with every step insight is gained, the search area is narrowed down and new design solutions are assessed.

A difference with this thesis is that the design space exploration is based on generated data or design solutions, instead of data of reference vessels. The method presented by *Duchateau* focuses on gaining insight into the design problem and deriving "What it is we are looking for", by performing an interactive design space exploration. In this thesis, the exploration phase focuses on learning from previously designed vessels and generating new designs based on that knowledge. As this should be done as fast as possible, thereby allowing a large number of design solutions to be explored, the human interaction, as proposed by *Duchateau*, is undesirable.



(a) Initial global search (b) Refined global search (c) Local search

Figure 2.11: Graphical representation of a type of progressive search process (from [14])

Incremental versus radical innovative design solutions

In *Garcia and Calantone (2002)* [32] a distinction is made between incremental and radical innovation. Based on these terms, two examples are given in *van Bruinessen et al. (2013)*. These examples can be seen in Figures 2.12a and 2.12b respectively.

The difference between incremental innovation and radical innovation is that incremental innovation is developed from existing knowledge and creates very little design space with known and familiar boundaries. Radical innovation on the other hand starts with a clean sheet of paper. Therefore the duration and the number of iterations of radical design projects is high [37].

For this thesis, a conclusion can be drawn about the level of innovation of the proposed design solutions. As the basis of this thesis is to design novel ships based on data of existing ships, the resulting design solutions are always a derivation of previous vessels. For radical innovation, a different approach is required, which is based on the functional requirements of the vessel only [37]. Therefore, in this thesis only incremental innovative design solutions are sought.

(a) Incremental innovation (b) Radical innovation

Figure 2.12: Different levels of innovation (from [37])

2.5. Conclusion

The current design methods at C-Job are strongly dependent on the experience of the naval architect, the type of ship, the customer requirements, time and customer budget. To decrease the human effort in the design process, the Accelerated Concept Design methodology has been developed. In this holistic design methodology all objectives, constraints and variables are dealt with simultaneously. The first step in this methodology, should be covered by the MIT, which performs a thorough reference study and makes predictions about advantageous main characteristics of a novel vessel.

The current method which is used in the MIT, does not meet the required predicting accuracy, especially in regions of the design space where data is limited. Also the neural and Bayesian method as described in [10] do not meet this requirement.

A second disadvantage is the difficulty to accept a design solution if it is produced by a black box model, which is per definition hard for a human to understand. To improve this, a naval architect should be able to see that a design solution meets certain ship design rules and regulations. If this is the case, a design can be considered as feasible. Literature describes some feasibility requirements. The method proposed by *Duchateau (2016)* [14] solves this problem, but the need for human interaction is a disadvantage.

The naval architect should search for incremental innovative design solutions only, using the MIT, because the proposed design solutions are always a derivation of previous work. For radical innovative designs a different approach is required. The second remark is that the results of the MIT should be handled with care. A naval architect should understand the requirements and the (different) purpose(s) of the vessel in order to choose which design is desirable.

To conclude, the method which is currently used in the MIT is a fast reference-study method with significant potential to improve the design process. To use this method in practice the gap between the data-based MIT method and the traditional knowledge-based method needs to be addressed.

3
Method Exploration

This chapter aims to find a method that will potentially solve the problem that is stated in Section 1.4 and deals with the challenges as discussed in Section 2.3. To do this, first a list of requirements is given. In Section 3.2, several potential methods are discussed. Finally, in Section 3.3, these methods will be compared to the requirements and one method is selected to be used for the rest of this research.

3.1. Requirements

In order to deal with the challenges that are mentioned in Section 2.3 and to improve the usability of the MIT, a list of requirements is set up. With the newly added method the MIT should meet these requirements. The requirements and a short explanation can be found in this section.

Ability to deal with data-sparse and data-abundant regions of a design space - As the tool should be able to cover the entire design space, it must be able to deal with regions with sufficient data, but also regions with limited data.

The results of the MIT should comply with the laws of physics and other governing (basic) ship design rules - The law of Archimedes is an example of a governing equation that should be satisfied. As mentioned in Section 2.3, this is currently not the case. Also other more empirical equations or basic ship design rules should be complied with. These equations can be found in the literature, as mentioned in Section 2.1. This must ensure that the boundary between technically feasible and infeasible design solutions is clear. In literature this requirement is used more often, but the term ’technically feasible solution’ is also referred to as ’reasonable solution’ [9] or ’believable solution’ [4].

The new method should provide insight - This requirement is twofold. First, insight in the design method should be obtained. This is possible when it is clear to the naval architect which calculations are done to come to a solution. Secondly, as mentioned in Section 2.1, the requirements can differ per type of ship. Based on these requirements, a naval architect should assess a design solution. If insight is given into how well the primary requirements and constraint requirements are met, it is easier to accept a design solution.

The new method should be a fast method - As the new method will be located in the outer layer of the design circle, as presented in Figure 2.3, it should have a high speed. Only with low computational time it is possible to consider a significant amount of design solutions. The method should provide the naval architect with the results within a reasonable time. The naval architect should be able to adjust and re-run the optimisation problem a couple of times a day. Approximately 15 minutes is considered to be a reasonable duration.

Ability to deal with feedback - Once the results have been plotted, it could be that the naval architect wants to adjust several designs and to see what the consequence are. For example, the length can be increased, to increase the cargo hold volume. As the machine learning model has already been trained, it is possible to use this model to generate a design with predetermined dimensions. In this case the naval architect would bypass the optimisation routine, to derive a specific design solution. The generated results can than be plotted and compared to the initially generated results.

3.2. Potential solutions

In this section possible solutions are discussed. As mentioned in Section 2.3, there are some disadvantages of a black box tool. *Duarte et al. (2004)* [5] describe that approaches that combine the mechanistic ("White Box") models with those of the empirical ("Black Box") techniques, integrating the best of both paradigms, could be useful. Such approaches are called "Grey Box" modelling. They aim to achieve good extrapolation properties, some degree of process behaviour rationalisation, ease of model development and focused phenomenological parameter fitting [5]. Because grey box modelling seems to cover the gap between the traditional design (White Box) method and the current (Black Box) MIT method, this will be the area to search for possible solutions in this thesis.

This is supported by *Estrada-Flores et al. (2006)* [16]. They state that grey box modelling or a semi-physical neural model may be regarded as a trade-off between a knowledge-based model and a black box model. They mention a neural model specifically, but this can also be a machine learning model. In this approach a priori knowledge concerning the process is used and it relies on parameterised functions, whose parameters are estimated from experimental data [42]. This estimation is done to deal with the unknown parts of the process [6]. In *Estrada-Flores et al.* a refrigeration system is modelled. This can be done in a very complex manner; considering all the variables and factors affecting refrigeration plants. This might result in a highly accurate model but also high development cost, long computation times and/or considerable amounts of data required from the user. A simplified model of the refrigeration system may provide less insight in the process and may be less accurate than the previous complex model. In the proposed solution a balance is found for the complexity/accuracy trade-off. Hence, this solution benefits from both models' advantages. In the proposed solution Fourier's Law (see formula 3.1) is used as a first-principle heat transfer equation. This is part of the white box model. The authors state that the true process is not covered by this formula only. To model lesser-understood relationships in the process, a black box model is used. In this black box equations are developed from statistical techniques and experimental data. In this thesis, the experimental data is the data of existing vessels.

$$q = -kA\frac{d\theta}{dx} \tag{3.1}$$

An advantage of this type of modelling is mentioned by *Oussar and Dreyfus (2001)* [42]. Since a larger amount of prior knowledge is used in the design of a semi-physical model than in the design of a black-box model, a smaller amount of experimental data is required to estimate its parameters reliably. This is interesting for two reasons. The first is that it could possibly make the MIT more accurate and faster, as some knowledge has already been included before the black box model starts learning the relation between input data and output. Secondly, in regions of the design space, where less data of existing ships is available, it could be possible to still produce accurate results, as less data is required with grey box modelling.

Sohlberg and Jacobsen (2008) [6] divide grey box modelling into five main branches. These are constrained black box modelling, semi-physical modelling, mechanistic modelling, hybrid modelling and distributed parameters systems. In literature these terms alternate and there is also not a clear difference between them. Therefore, in the next section only two clearly different branches are described. The first is hybrid modelling, which consists of serial and parallel hybrid modelling [16, 6] and the second is constraint black box identification [6].

Serial hybrid modelling

The serial hybrid modelling approach involves an empirical model which is fed with operating data used to estimate parameters and a mechanistic model [5]. Within serial hybrid modelling two approaches can be found.

In the first approach the use of a black box sub-model precedes the use of a white box sub-model [16] (see Figure 3.1). This approach is illustrated in Figure 3.1. The first sub-model is fed with data. This data is used to estimate parameters which are then provided to the white box sub-model. Here the estimated parameters are the input for a knowledge-based formula that describes the process [5, 16].

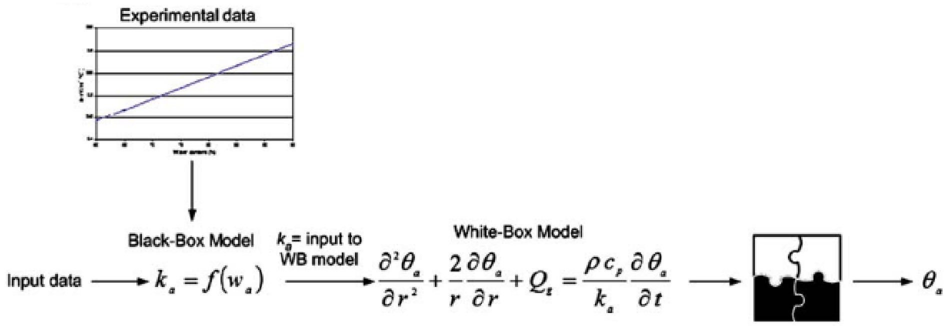


Figure 3.1: Serial hybrid modelling approach 1 (from [16])

In the other approach the use of the white box model precedes the use of the black box model (see Figure 3.2). In this approach the aim is to obtain a correlation between a white box parameter and a second parameter not included in the white box. First, a knowledge-based formula is used to estimate a characteristic Y_i, for multiple input values X_i. The resulting values are then regressed against an unknown variable (Z_i) and a black box model relating Y_i and Z_i would be obtained [16].

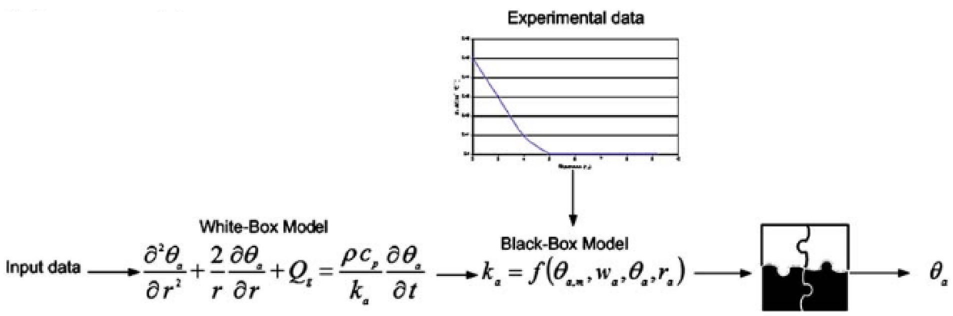


Figure 3.2: Serial hybrid modelling approach 2 (from [16])

Estrada-Flores et al. [16] mention one disadvantage of serial hybrid modelling. Serial hybrid modelling consists of two stages. First a white box model and then a black box model, or vice versa. The uncertainty in the individually estimated input variable is not carried through to the second stage. This causes a loss of information, which results in less accurate estimates of the regression parameters.

Several examples of grey box modelling can be found in literature. Often this type of modelling is used to model biochemical or mechanical processes. *Schubert et al. (1994)* [35] model yeast cultivation with a serial hybrid modelling approach. In *Acuña et al. (1999)* [2] the same approach is used to model bioprocesses kinetic rate expressions. In *Zwart et al. (2020)* [44] a neural network is fed with an initial estimate based on the physics-based model. Using this method the trim of a vessel is optimised based on operational voyage data.

Parallel hybrid modelling

The parallel hybrid modelling approach consists of a knowledge-based white box model, which is used to estimate process behaviour, while an empirical black box model aims to forecast the corrections that have to be added to the white box model predictions to obtain the true process behaviour [5, 16, 6]. As the white box and black box sub-model are both fed with the same data simultaneously, there is no loss of information between these two sub-models [5, 16]. As mentioned, this latter is a disadvantage in serial hybrid modelling. The parallel hybrid structure can be seen in Figure 3.3. For the output of this approach it is clearly visible that the white box estimate θ_a and the black box correction $\hat{\psi}$ are only depending on the input data t.

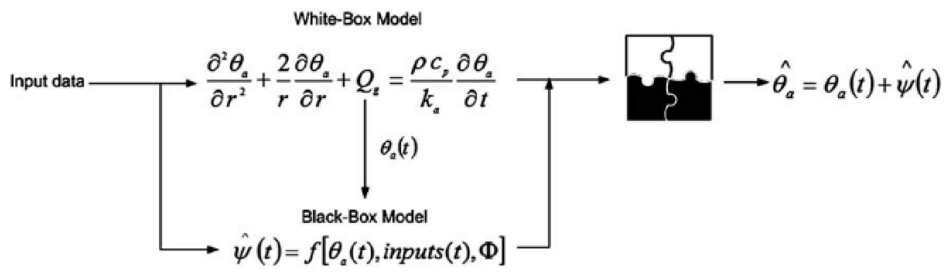


Figure 3.3: Parallel hybrid approach (from [16])

Thompson and Kramer (1994) [25] present a more extensive structure, as can been in Figure 3.4. This structure can be seen as both parallel and serial. Compared to Figure 3.3, the black box model (Artificial Neural Network) is based on the input data and on the output of the white box model (Default Mathematical Model).

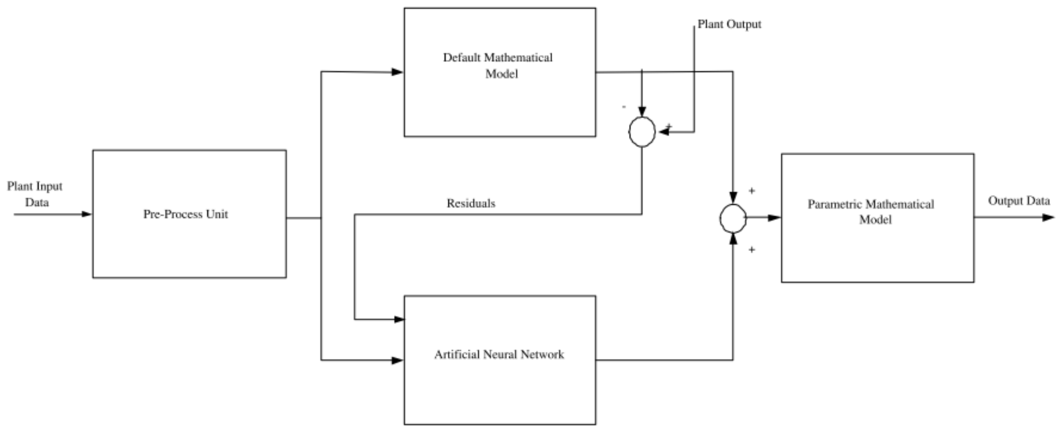


Figure 3.4: Parallel hybrid approach (from [25])

In literature some other examples of parallel hybrid modelling are given. *Su et al. (1992)* [19] use this method to model a chemical reactor system. *Shum and Myers (1996)* [33] propose parallel structure for octane control in platforming units. In *Van Can et al. (1996)* [38] it is used to model and control a laboratory pressure vessel. In *Duarte (2004)* [5] a difference is mentioned about how the correction can be done in parallel hybrid modelling. In *Su et al.* the correction is done at the end of a time horizon of interest, whereas in *Duarte* this correction is done continuously. In this case the mechanistic model captures the essence of the dynamics of the system (e.g. speed of response etc.) while the main function of the empirical model is to correct the actual values predicted over the, often, wide range of operating conditions encountered during transient operation [5]. In *Leifson et al. (2008)* [23] both serial and parallel hybrid modelling are used to predict the vessel fuel consumption. Both show more accurate results compared to the white box model.

Constrained black box identification

Constrained black box identification (CBBI) uses a black box model where specific parameters, which are estimated using measured data, are constrained based on physical relations [6]. This could be interesting for this research project. An example of such a physical relation is the law of Archimedes. *McDonald (2010)* [27] describes that unacceptable combined options are removed to ensure that options are 'balanced'. As an example he mentions that a constraint on weight vs. displacement could be used to remove Archimedially unbalanced designs. This could be useful in this thesis. Also other physical constraints, as mentioned by *Watson (1998)* [40], can be included. These are for example the ratios B/D and T/D, as mentioned in Section 2.1.

3.3. Evaluation of solutions

The goal of combining white and black box models is to improve the accuracy and to make sure that generated design solutions will correctly be classified as feasible or infeasible. The new grey box model should be a fast method.

In literature a lot of equations are described. Some of these equations are meant to describe a process as well as possible. Other equations define a bandwidth or range in which a process will occur. The first type of equations have a lot in common with the serial or parallel hybrid modelling. The goal of these approaches is to describe a process as well as possible using First Principle rules, and to estimate the unknowns with an empirical black box method. As the second type of equations are about constraining the process, the black box identification method seems useful.

In Table 3.1 the potential solutions are qualitatively assessed based on the requirements. For some requirements there are some clear differences in the performance of the potential solutions e.g. the ability to deal with data limited areas. The CBBI model is comparable to the current model used in the MIT; both are only based on data. The difference is that the CBBI model is constraint, which means that some knowledge is taken into account in this model. In regions of a design space with limited data, the CBBI model will, although constrained, still perform poorly. The serial hybrid model is assumed to perform better as it takes into account both the data and knowledge. On the other hand, as these two models are connected in series, they are also very dependent. In the parallel hybrid modelling approach the data and knowledge model are connected in parallel. This means that they work independently, which is an advantage. As the amount of available data differs per situation, also the applicability of the models differ. For regions with limited data the model should rely more on knowledge instead of data and for regions with sufficient data vice versa.

Both the hybrid modelling methods contain a black box model and a separate white box model. The white box model contains all the relevant knowledge, described in a mathematical form. As this model is understandable to a naval architect, this increases the insight in the design process.

Method requirements	Serial hybrid modelling	Parallel hybrid modelling	Black box identification
Deal with data limited areas	medium	good	bad
Comply with laws and rules	medium	medium	medium
Provide insight in design method	good	good	medium
Speed	medium	good	good
Ability to deal with feedback	good	good	good

Table 3.1: Evaluation of solutions based on method requirements

Mainly because of the independent operations in the parallel hybrid model, this approach is preferred, compared to the other two models. This independence makes sure that a parallel hybrid model can perform well in both data-sparse and data-abundant regions of a design space. As the constrained black box identification method is a very simple and clear method to include constraints, this can be used to take into account constraints like the law of Archimedes.

Based on the tool requirements, several potential solutions have been explored. Parallel hybrid modelling is chosen as the most promising method to solve the problems that C-Job's naval architects face, in order to make well-advised design solutions in the future reference-based design approach. With this, research question 2 can be answered.

Research question 2

How can a naval architect make well-advised design decisions in C-Job's future reference-based design approach?

Based on Chapter 2, a list of requirements is determined for the new tool. In these requirements, the current design approaches have been taken into account as well as the challenges that lie ahead for the future design approach. These requirements are the following

1. Ability to deal with data-sparse and data-abundant regions of a design space
2. Results should comply with the laws of physics and other governing (basic) ship design rules
3. The new method should provide insight
4. The new method should be a fast method
5. Ability to deal with feedback

Based on these requirements some potential solutions have been explored and evaluated. It shows that parallel hybrid modelling is the most promising solution for the problems defined in this thesis. The advantage of this type of modelling is that the available knowledge can be included in a white box model. Using this white box model an estimation can be done about the magnitude of certain parameters. Based on data of reference vessels a machine learning model can be trained to learn about the differences between this estimation and the 'true' values (i.e. the data of reference vessels). This knowledge can then be used to correct the first-principle white box model.

An advantage of this method is that the white box model is constructed by a naval architect. In this model the naval architect can define all the relevant equations and rules. This provides the insight in the design process that is required. Next to that, by including these equations and rules, the naval architect can ensure the feasibility of the proposed design solution.

By requiring that the new method must be a fast method, the naval architect is able to run the optimisation problem, explore the results, redefine and re-run the optimisation problem. Using this method the first step in the current design method, which is finding a starting point, is done by the MIT. Therefore it is possible to explore a significant amount of feasible design solutions in a reasonable amount of time, which increases the possibility of finding better design solutions compared to the current method.

4

Methodology

This chapter describes the method that is used in this thesis. First a general description of the model will be given in Section 4.1. Secondly, the resource of the reference data is discussed in Section 4.2. After that, every component of this model will be explained in more detail. This is first the White Box model in Section 4.3. In this section different empirical methods are compared qualitatively and quantitatively. The best methods are then chosen to be implemented in the white box model. Thereafter a description of the black box model will follow in Section 4.4. This section elaborates on which technique is used to train the black box model and how this works. In Section 4.5, a brief explanation will follow about how the white box and black box model work together. Finally, Section 4.6 describes different methods to assess the performance of a model.

4.1. General model description

As can be seen in Figure 3.3, the result of a parallel hybrid model is a summation of the results of two models; the white box result and the black box result. Equation 4.1 [5] represents this summation. The white box result is an estimated value of a parameter, based on formulas. The black box model corrects this estimation, based on statistics. In literature different symbols are used for each component of the parallel hybrid model. In this thesis, the same symbols are used as those in Figure 3.3.

$$\hat{\theta}_a = \theta_a(t) + \hat{\psi}(t) \tag{4.1}$$

The first term $\theta_a(t)$ is the estimated value of a certain parameter, determined by the white box model. The second term $\hat{\psi}(t)$ is the black box model correction. The hat (ˆ) above this term means that it is based on statistics, instead of formulas. In order to predict $\hat{\psi}(t)$, first a black box model needs to be trained with data before it can be used. This means that the proposed parallel hybrid model works in two phases; a training phase and a prediction phase.

Training Phase - This phase can be seen in Figure 4.1. First, training data is selected. This training data contains data of the independent variables and the dependent variable, or the design parameter, that one wants to predict. First, based on the independent variables, a white box model is used to calculate $\theta_a(t)$. This term contains the estimated values of the dependent variable. A black box model is then trained to learn the difference between the estimated value of a design parameter and the true value of this parameter. For the training data, the estimated values of a design parameter $\theta_a(t)$ are subtracted from the true values Φ, according to Formula 4.2 [5]. Thus, $\hat{\psi}(t)$ contains the absolute differences between the true and the estimated value of a design parameter.

A black box model can then be trained. First, the independent variables and the dependent variables need to be defined. The dependent variable is the $\hat{\psi}(t)$, this is the term that should be predicted.

The independent variables are all the variables that have influence on the $\hat{\psi}(t)$. These variables are defined by the naval architect. Now that the independent and dependent variables are defined, a black box model can be trained. How this works exactly is described in Section 4.4. The result is a model that makes a prediction $\hat{\psi}(t)$, only based on the independent variables.

$$\psi(t) = \Phi - \theta_a(t) \tag{4.2}$$

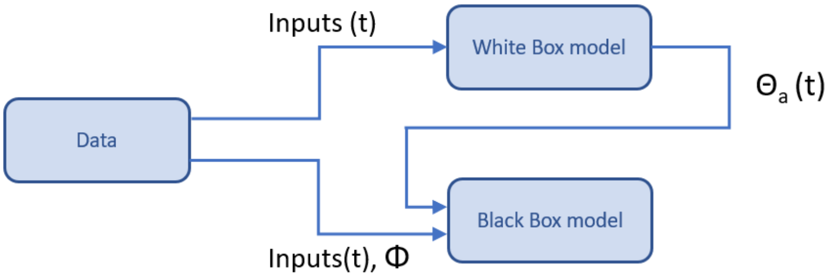


Figure 4.1: General description of a Parallel Hybrid model - Training phase

Prediction Phase - The trained model can now be used in the prediction phase. This phase can be seen in Figure 4.2. In the prediction phase, new data is used. This is the test-data. Which portion of the data is used for the training phase and the prediction phase will be explained in the following sections. The values of the independent variables of the test data-set are used as input for both the white box and the black box model. The white box model again estimates $\theta_a(t)$, but now the black box model immediately predicts the correction $\hat{\psi}(t)$. The estimation and the predicted correction are then summed according to equation 4.1. This gives the term $\hat{\theta}_a$, which is the Parallel Hybrid model prediction.

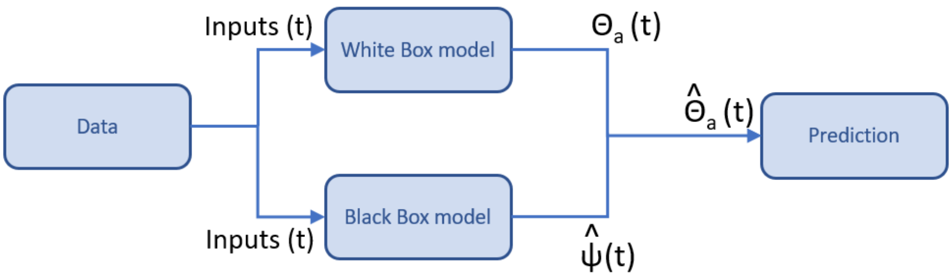


Figure 4.2: General description of a Parallel Hybrid model - Prediction phase

Example

To make the parallel hybrid model more clear, the lightship weight (LSW) of the vessel will be used as an example. This aspect is chosen because the true values as well as knowledge of the formulas are required for parallel hybrid modelling. For the lightship weight aspects, both the formulas and the true LSW values are available.

In the training phase, first an estimation of the lightship weight is done based on formulas. This is done for every reference vessel that is selected for the training data-set. The input data which is required for these white box formulas differ per formula, but for the lightship weight it is mostly dependent on the length, breadth, depth and block-coefficient. The white box formulas are further described in Section 4.3. Secondly, the difference between the true and the estimation lightship weight is found by using equation 4.3.

$$\psi_{LSW} = \Phi_{LSW} - \theta_{LSW} \tag{4.3}$$

The ψ_{LSW} will be fed to the black box model as the dependent variable. The independent variables are defined by the naval architect, but it is obvious to use variables that define the dimensions of the vessels, such as length, breadth, depth and block-coefficient. Also, variables that have significant influence on the lightship weight are selected as independent variables, such as the Maximum Continuous Rating (MCR). The MCR has influence on the engine size and therefore also contributes to the total lightship weight. Thus, for the dependent variable [*lightship weight*] the independent variables are [*length, breadth, depth, block-coefficient, MCR*].

This trained model can be used in the prediction phase. In this phase the independent variables of new and unseen vessels are fed to the white box model and black box model simultaneously. The white box model makes an estimation for the lightship weight based on formulas and the black box model predicts the difference between this estimation and the true value. The total prediction is calculated by summing the results of both models as is shown in equation 4.4.

$$\hat{\theta}_{LSW} = \theta_{LSW}(t) + \hat{\psi}_{LSW}(t) \tag{4.4}$$

4.2. Reference data

The reference data that will be used in this thesis is collected from Sea-web, a database from IHS Markit [26]. During this thesis several problems were found in this database, including:

1. Database contained duplicates and sister vessels
2. Certain parameters were calculated instead of registered
3. Reference data was found unreliable after performing calculations

Each of these problems will be described in the following sub-sections and a solution will be given about how to deal with these problems.

4.2.1. Duplicates and sister vessels

During this project, it became clear that the database contained a lot of similar vessels. This could for example mean that the database contained multiple sister vessels, but also that the exact same vessel was in the database multiple times. Thus, there are two types of vessels that have been removed from the database:

1. Exact duplicates
2. Sister vessels

The reason to delete exact duplicates from the database is clear. A certain vessel should be in the database just once. There are a couple of arguments for deleting the sister vessels. The first argument is that there is no extra knowledge to learn from sister vessels. A second argument has to do with potential overfitting of the black box model. If a black box model learns from data of ten sister vessels and one other vessel, one can imagine that this automatically puts more weight on the data of the sister vessels, leading to an overfitted model. A third argument has to do with how the model's performance will be assessed. If a model has been trained to predict the lightship weight, with data of reference vessels, it basically means that the model already knows what the lightship weight is of all those reference vessels. Therefore, it is easy to predict the lightship weight of a sister vessel, because it has already seen the lightship weight of a sister vessel during the training phase. Thus, to make sure that training- and test-data do not contain similar vessels, the first of the sister vessels will be retained and the rest will be deleted. This argument is further described in Section 4.6.2.

The exact duplicates are deleted using a Python function. In this function, the user defines the variables where duplicate values should be found. If a vessel is an exact duplicate of another vessel, for all these variables, the first vessel in the database is retained and the rest is deleted. The variables that are used to find duplicates are

1. Gross tonnage
2. Length between perpendiculars
3. Breadth overall
4. Depth
5. Draught
6. MCR

The margins which are used to find similar vessels are shown in Table 4.1. If the difference between two vessels was smaller than the margins for all the mentioned variables, then only the first vessel was retained. The variables are chosen because vessel owners are obliged to report Gross tonnage, length between perpendiculars, breadth moulded and depth. Therefore it was assumed that these values in the database are correct.

It was noted that a certain vessel can be available in different versions. This could mean that the dimensions are identical with another vessel, but that the MCR was different. The same vessel is in that case available with different engines. The difference in engine type was also observed in the lightship weight value, as the weight of the engine is part of the lightship weight. Therefore, duplicate vessels that have a different MCR, and therewith a different engine weight, are not removed from the database.

As can be seen in Table 4.1, the margins are constant. They are independent of the ship's size. These values are determined by looking at the dataset and searching for sister vessels. After a few iterations, it was noted that, by using these values as margins, similar sister vessels were filtered out of the database. An improvement of this method is to make the margins dependent of the size of the vessels, but for the database that is used in this thesis, the constant values are sufficient.

Variable	**Margin**
Gross tonnage	15 tonnes
Length	1 meter
Breadth	0.5 meter
Depth	0.5 meter
MCR	10 kW

Table 4.1: Margins for finding similar vessels

The result of this cleaning method can be seen in Table 4.2. First, the initial amount of reference data for a certain ship type is given. Thereafter follows how much the data set reduced in size after the two cleaning steps; delete the duplicates and delete the sister vessels. Thus, the amount of data that is left, after deleting the sister vessels, is the data set that is usable for the MIT.

This means that more than half of the database consisted of bulk carriers. For other ship types this is of the same order.

	Bulk carrier	**General Cargo**	**Oil Tanker**	**Container ship**
Initial data set	2903	3030	187	964
Duplicates deleted	883	474	22	175
Sister vessels deleted	804	989	33	336
Remaining vessels	1216	1567	132	453

Table 4.2: Reducing size of the data set due to cleaning

4.2.2. Registered or calculated values

Not all the values in the database were officially registered values. During this project, it became clear that some of the parameters were calculated values. Two important parameters for this thesis, the block-coefficient and the lightship weight, were calculated using Formulas 4.5 and 4.6 respectively. These calculations were done by C-Job's naval architects. The ∇ in these formulas is the displacement in cubic meters, the L_{pp} is the length between perpendiculars, the $B_{moulded}$ is the moulded breadth and the draught is represented by T. The $\triangle$ - term in Equation 4.6 is also the displacement, but now expressed in tonnes.

$$C_{\mathrm{B}} = \frac{\nabla}{L_{\mathrm{pp}} \cdot B_{\mathrm{moulded}} \cdot T} \tag{4.5}$$

$$\mathrm{LSW} = \triangle - \mathrm{DWT} \tag{4.6}$$

To deal with this problem, first, the source of the different parameters was checked. Tables 4.3 and 4.4 describe which parameters were given in the database from IHS Markit. The Deadweight Tonnage (DWT) and the displacement $\triangle$ were both registered in the IHS Markit database. Using Formula 4.6, the lightship weight can be calculated. As this formula is the definition for lightship weight, the obtained value for lightship weight is assumed to be the true value as well.

Term	Definition
Length Overall	The extreme length of the ship including any bow protrusion.
Length (Reg)	The registered length as given on the ship's certificates
Length (BP)	The length between perpendiculars is the distance on the summer load waterline from the fore side of the stem to the after side of the rudder post, or to the centre of the rudder stock if there is no rudder post
Breadth Extreme	This is the maximum breadth to the outside of the ship's structure.
Breadth Moulded	This is the greatest breadth at amidships from heel of frame to heel of frame
Draught	In most cases this is the maximum draught amidships, but in some ships of special construction the maximum draught is at the after end, and this measurement is recorded.
Depth	This is the vertical distance at amidships from the top of the keel to the top of the upper deck beam at side.
Height	Height measured from the keel to the highest fixed point.
Displacement	Vessel's laden displacement in tonnes.
T/CM	This is the number of tonnes needed to immerse the vessel by one centimetre. The value shown is that corresponding to the maximum summer draught.

Table 4.3: Registered dimension data from IHS Markit from (from [26])

Term	Definition
Gross/Net Tonnage	Gross and Net Tonnages (GT and NT) are defined in The International Convention on Tonnage Measurement of Ships, 1969 adopted by the International Maritime Organisation in 1969 and came into force in July 1982. These measurements replaced Gross and Net Register Tonnage (GRT and NRT). Gross Tonnage (GT) is a unitless function calculated from the moulded volume of all enclosed spaces of the ship.
Net Tonnage (NT)	is produced by a formula which is a function of the moulded volume of all cargo spaces of the ship. The net tonnage shall not be taken as less than 30 percent of the gross tonnage. Gross Tonnage forms the basis for manning regulations, safety rules and registration fees. Both gross and net tonnages are used to calculate port dues.
Deadweight	The weight in tonnes (1000 kg) of cargo, stores, fuel, passengers and crew carried by the ship when loaded to her maximum summer loadline
Formula Deadweight	This is calculated using the following formula Length between Perpendiculars x Moulded Breadth x Moulded Depth divided by 2,265.

Table 4.4: Registered tonnage data from IHS Markit (from [26])

This poses a challenge for this thesis as these values contradict two requirements for parallel grey box modelling, namely:

Knowledge-based formulas - These are required to estimate a design parameter, using the white box model.

True values - These are required for training a black box model to learn the difference between the white box model estimation of a design parameter and the true value of that design parameter.

For other interesting design parameters, such as resistance and stability, these knowledge-based formulas were available. The problem is that the true values of these design parameters are not available. The true resistance can be obtained from sea trials, and the stability can be obtained from an inclining test. The true values might be available for ship owners themselves, but for competitive reasons these are obviously not shared. Therefor, the decision was made to focus on predicting the lightship weight in this thesis. In the rest of this thesis, only the lightship weight design parameter will be discussed.

4.3. White Box model

Multiple white box models have been assessed in this thesis. Three groups of methods have been distinguished in literature, namely:

1. Generic method
2. Ship type specific methods
3. Generic ship type specific method

These different methods will be described below. First, the generic method will be discussed in Section 4.3.1. Section 4.3.2 will describe many ship type specific methods. In this section a conclusion will be drawn about what ship types can and should be tested with the parallel hybrid modelling approach. Thirdly, a method is presented which is both generic and ship type specific in Section 4.3.3. In Section 4.3.4 a method is described on how to validate the results of all the weight estimation methods.

4.3.1. Generic method

First, a generic formula is used. This formula, Formula 4.7 *(Indian Maritime University Visakhapatnam, 2017)* [39] is only dependent on the deadweight (DWT) of the vessel. An advantage of this formula is that it is applicable to every ship type. Because of that, it is expected that this method is not as accurate as ship type specific methods.

$$W_{\text{LS, generic}} = 1128 \cdot \left(\frac{\text{DWT}}{1000}\right)^{0.64} \tag{4.7}$$

4.3.2. Specific methods

Literature describes multiple empirical formulas which are more specific to a certain ship type, or a certain component of the lightship weight. These components can be seen in Formula 4.8. The steel weight consists of structural steel of the hull, the superstructure and the weight of the outfit steel. Outfit steel consists of the machinery foundations, supports, masts, ladders and handrails for example. The outfitting and equipment weight includes deck machinery. The machinery weight consists of the main engine, auxiliary machinery, propeller and shaft.

Table 4.5 provides information of these formulas specific for a weight component. The gaps in the table mean that methods to estimate a specific weight component corresponding to a specific ship are not found in literature.

$$\begin{aligned} W_{\text{Lightship}} &= W_{\text{Steel}} + W_{\text{Outfitting \& Equipment}} + W_{\text{Machinery}} \quad \text{, or} \\ W_{\text{LS}} &= W_{\text{S}} + W_{\text{O\&E}} + W_{\text{M}} \end{aligned} \tag{4.8}$$

Criteria for selecting weight estimation methods

The methods from Table 4.5 were assessed based on three criteria, namely:

Applicability in wide range of ship dimensions - The method should be applicable for a large range of ship dimensions. For example the method from *Det Norske Veritas (DNV) (1972)* [29] for calculating the steel weight of tankers is only applicable for ship with a length in between 150 and 480 meters, a length/breadth ratio between 5 and 7 and a length/depth ratio between 10 and 14. This is the case for most of the tankers in the database, but these limitations should be taken into account.

The steel weight calculation method from *Sato* [29] seems advantageous in this case, because it doesn't have these dimension-limitations.

Applicability in an early stage of the ship design process - Some methods are based on variables that are not yet known in this stage of the design process. The *Watson & Gilfillan (1976)* [29] method to calculate the steel weight for example uses the number, height and length of the deck-houses and superstructures. Another example is the method from *Buxton (1976)* [29] that can be used to calculate the machinery weight. For this method, information is required about the propeller (single-propeller / twin-propeller). This is not yet available in this stage of the design process, nor is it available in the database. Therefore, these methods are less suitable to be implemented in the MIT.

Availability of reference data - The amount of available reference data for a ship type is important, because in order for a black box model to perform well, sufficient data is required. 100 reference vessels were chosen as the minimum amount of reference data per ship type. In the chosen black box model the reference data set will be distributed across 100 different samples. These samples need to be different in order to get a well performing black box model. For a reference vessel data set of 100 and higher, it is expected that these samples are indeed different. This will be explained in detail in Section 4.4.

Ship type	Structural steel	Outfiting & Equipment	Machinery
Generic	Watson & Gilfillan (1976), Watson (1998)	Watson & Gilfillan (1977)	Munro-Smith (from [39]), Watson & Gilfillan (1976), Buxton (1976)
General cargo	Schneekluth (1985)	Schneekluth (1985), Papanikolaou (2019)	Munro-Smith (from [39]), Watson & Gilfillan (1976), Buxton (1976)
Dry Cargo	Schneekluth (1985), Wehkamp-Kerlen (1985), Watson & Gilfillan (1976), Harvald & Jensen (1992), Danckwardt(1961)		Munro-Smith (from [39]), Watson & Gilfillan (1976), Buxton (1976)
Bulk carrier	Hurrey J.M. (from [39]), DNV (1972), Murray (1965)	Papanikolaou (2019)	Munro-Smith (from [39]), Watson & Gilfillan (1976), Buxton (1976)
Container	Schneekluth (1985), Chapman K.R. (1969), Miller D. (1968)	Jensen (1992), Papanikolaou (2019)	Schneekluth (1985), Munro-Smith (from [39]), Watson & Gilfillan (1976), Buxton (1976)
Oil tanker	DNV (1972), Sato (1967)	Schneekluth (1985) Papanikolaou (2019)	Munro-Smith (from [39]), Watson & Gilfillan (1976), Buxton (1976)
Ropax	Papanikolaou (2019)	Papanikolaou (2019)	Munro-Smith (from [39]), Watson & Gilfillan (1976), Buxton (1976)
Tanker	DNV (1972), Harvald & Jensen (1992)	Papanikolaou (2019)	Munro-Smith (from [39]), Watson & Gilfillan (1976), Buxton (1976)
Reefer		Carreyette (1976)	Munro-Smith (from [39]), Watson & Gilfillan (1976), Buxton (1976)

Table 4.5: Empirical formulas for lightship weight components in literature

Selected methods

Based on the criteria, methods for four ship types were selected to be further assessed. The reason for this is that, for these ship types, there were useful weight estimation methods available and there was sufficient reference data available. The ship types that were chosen are

1. Bulk carrier
2. General cargo ship
3. Oil tanker
4. Container ship

For each weight group, the paragraphs below will give the formulas that are used per ship type. The symbols that are used in these formulas are explained in Table 4.6. The goal is to find the best formulas to estimate the lightship weight for each ship type. Thus, after assessing these formulas quantitatively, one estimation method per ship type will be implemented in the white box model as a part of the parallel hybrid model. How these formulas are assessed quantitatively, will be explained in Section 5.2.
To calculate the **steel weight** (W_S) multiple formulas have been used. These formulas per ship type can be found in Table 4.7.

Symbol	Meaning	Symbol	Meaning
L_{OA}	Length overall	C_B	Block-coefficient
L_{BP}	Length between perpendiculars	$\triangle$	Displacement (tonnes)
B_{OA}	Breadth overall	BHP	Brake Horse Power (kW)
D	Depth of hull	RPM_{ENG}	Engine rpm
T	Draught	MCR	Maximum Continuous Rating (kW)

Table 4.6: Used symbols and their meaning

Shiptype	Method	Formula
Bulk carrier	Murray (1965)	4.9
General cargo ship	Wehkamp-Kerlen (1985)	4.10
	Watson & Gilfilan 1976	4.11
Oil tanker	Det Norske Veritas (1972)	4.12
	Sato (1967)	4.13
Container ship	Chapman (1969)	4.14
	Miller (1968)	4.15

Table 4.7: Formulas for steel weight of different ship types

$$W_{\text{S, Bulk}} = 0.0328 \cdot L_{OA}{}^{1.65} \cdot \left(B_{OA} + D + \frac{T}{2}\right) \cdot \left(\frac{C_B}{2} + 0.4\right) \tag{4.9}$$

$$\begin{aligned} W_{\text{S, General cargo}} &= 0.0832 \cdot A \cdot \mathrm{e}^{-5.73 \cdot A \cdot 10^{-7}} \quad \text{, where} \\ A &= \frac{L_{BP}{}^2 \cdot B_{OA} \cdot C_B{}^{\frac{1}{3}}}{12} \end{aligned} \tag{4.10}$$

$$W_{\text{S, General cargo}} = C_B{}^{\frac{2}{3}} \cdot \frac{L_{BP} \cdot B_{OA}}{6} \cdot D^{0.72} \cdot 0.002 \cdot \left(\frac{L_{BP}}{D}\right)^2 + 1 \tag{4.11}$$

$$W_{\text{S, Tanker}} = \triangle \cdot \left(\alpha_L + \alpha_T \cdot \left(1.009 - 0.004 \cdot \frac{L_{OA}}{B_{OA}}\right) \cdot 0.06 \cdot \left(28.7 - \frac{L_{OA}}{D}\right)\right) \quad \text{, where}$$

$$\begin{aligned} \alpha_T &= 0.029 + 0.00235 \cdot \triangle \cdot 10^{-5} \quad \text{, for } \triangle < 6 \cdot 10^5 \\ \alpha_T &= 0.0252 \cdot \left(\triangle \cdot 10^{-5}\right)^{0.3} \quad \text{, for } \triangle => 6 \cdot 10^5 \\ \alpha_L &= \frac{\left(0.054 + 0.004 \cdot \frac{L_{OA}}{B_{OA}}\right) \cdot 0.97}{0.189 \cdot \left(100 \cdot \frac{L_{OA}}{D}\right)^{0.78}} \end{aligned} \tag{4.12}$$

$$W_{\text{S, Tanker}} = \left(\frac{C_B}{0.8}\right)^{\frac{1}{3}} \cdot \left(5.11 \cdot {L_{BP}}^{3.3} \cdot \frac{B_{OA}}{D} + 2.56 \cdot {L_{OA}}^2 \cdot (B_{OA} + D)^2\right) \cdot 10^{-5} \tag{4.13}$$

$$W_{\text{S, Container}} = 0.0209 \cdot {L_{BP}}^{1.759} \cdot {B_{OA}}^{0.712} \cdot T^{0.374} \tag{4.14}$$

$$W_{\text{S, Container}} = 0.000435 \cdot (L_{OA} \cdot B_{OA} \cdot T)^{0.9} \cdot (0.675 + 0.5 \cdot C_B) \cdot \left(0.00585 \cdot \left(\frac{L_{OA}}{T} - 8.3\right)^{1.8} + 0.939\right) \tag{4.15}$$

For the **outfitting & equipment weight ($W_{O\&E}$)** component a general formula is used. This is formula 4.16, which has been described in *Papanikolaou (2019)* [29]. The K_{OT}-term is given for multiple ship types. These can be found in Table 4.8. As can be seen, the K_{OT}-term for the bulk carrier is described as a formula. *Papanikolaou* gives the K_{OT} for bulk carrier with a length around 140 meters and a length around 250 meters. The corresponding values for K_{OT} are up to 0.25 and 0.18 respectively. In this thesis the K_{OT} is assumed to be linearly dependent of the length. Based on this assumption, a linear relation is determined, which can be seen in Table 4.8.

A second remark is that all the values of K_{OT} in this table are conservative values. In *Papanikolaou (2019)* [29], a range is given for K_{OT}, instead of one value. One reason to choose one value instead of a range is that the model would not become too complex. The second reason is that a step in a later stage of the Accelerated Concept Design process is the optimization step. As is shown in *De Winter (2019)* [31], reductions of steel weight can go up to 14% by using the ACD method. This is a result of the optimization step. Thus, this step has much more influence on the steel weight than for example the outfit coefficient. Therefore, in this stage of the design process, a conservative value was chosen for K_{OT}, i.e. the maximum value of the range given by *Papanikolaou.*

$$W_{O\&E} = K_{OT} \cdot L_{OA} \cdot B_{OA} \tag{4.16}$$

Shiptype	**K_{OT}**
Bulk carrier	$0.3391 - 0.000636364 \cdot L_{OA}$
General cargo ship	0.45
Oil tanker	0.28
Container ship	0.38

Table 4.8: Outfit & Equipment coefficients

Two formulas have been used for the **machinery weight (W_M)** component. These are Formulas 4.17 and 4.18. These formulas are given by *Watson & Gilfillan (1976)* [29] and by *Murrirosmith* [39] respectively. Formula 4.17 is a summation of the main machinery weight, the first term, and the auxiliary weight component, the second term.

$$W_M = \left(\frac{\text{MCR}}{\text{RPM}_{\text{ENG}}}\right)^{0.84} \cdot \text{RPM}_{\text{ENG}} \cdot 12 + \text{MCR}^{0.7} \cdot 0.69 \tag{4.17}$$

$$W_M = \frac{\text{BHP}}{10} + 200 \tag{4.18}$$

Assessment of different methods

Different combinations of the above mentioned ship type specific weight estimation methods will be tested. As there is no data available of the individual weight components, one cannot test the performance of each formula individually. Only data of the total lightship weight is available. Therefore the approach to find the best formulas is to try different configurations of formulas, and measure how well the total lightship weight can be estimated.

4.3.3. Generic ship type specific method

Thirdly, a method was found that is both generic and ship type specific. This is a method from *D'Almeida (2009)* [12]. *D'Almeida* gives a generic formula for each component, independent of the ship type. Thus, an advantage of this method is that is is applicable to all ship types. This method is also ship type specific, because the formula also consists of ship type specific components. The coefficients are statistically determined. For the machinery weight component, the coefficients are based on the type of propulsion plant and not on the type of ship. A second advantage is that the method from *D'Almeida* is from 2009. Thus, it is the most up-to-date formula and therefore it is expected to be the most accurate.

For the steel weight Formula 4.19 is used. The coefficients $k1$, $k2$, $k3$ & $k4$ are ship type specific and can be found in Table 4.9. For L, the length overall is used and for B the moulded breadth is used.

$$W_{\text{Steel}} = k1 \cdot L^{k2} \cdot B^{k3} \cdot D^{k4} \tag{4.19}$$

The weight for the equipment can be found using formula 4.20. The coefficients $k5$ & $k6$ are also ship type specific and can be found in Table 4.9.

$$W_{\text{Equipment}} = k5 \cdot (L \cdot B \cdot D)^{k6} \tag{4.20}$$

Coefficient	Bulk	Oil tanker	Container	General Cargo
k1	0.0328	0.0361	0.0293	0.0313
k2	1.6000	1.6000	1.760	1.675
k3	1.0000	1.000	0.712	0.850
k4	0.2200	0.2200	0.374	0.280
k5	6.1790	10.820	0.1156	0.5166
k6	0.48	0.41	0.85	0.75

Table 4.9: Coefficients for weight calculations (D'Almeida (2009) [12])

Finally, the machinery weight can be calculated with formula 4.21. This formula is dependent on the propulsion plant. *D'Almeida* gives four options, which can be seen in Table 4.10. P_{MCR} is the propulsive power [bhp].

$$W_{\text{Machinery}} = k7 \cdot P_{MCR}{}^{k8} \tag{4.21}$$

4.3.4. Validation of estimated weight components

Multiple formulas have been described in this chapter. These formulas will be tested with the available data to find the best formulas. To ensure the reliability, the estimation of each weight component will be compared to a weight ratio. Figure 4.11, from Strohbusch (1971), Schneekluth (1985) and updated by Papanikolaou (2011) [29], gives typical percentages for weight groups.

Type	**k7**	**k8**
Diesel (2 stroke)	2.41	0.62
Diesel (4 stroke)	1.88	0.60
2 x Diesel (2 stroke)	2.35	0.60
Steam Turbine	5.00	0.54

Table 4.10: Coefficients for machinery weight formula based on type of propulsive plant (D'Almeida (2009) [12]

First, in column 1 and 2, the lower and upper limits are given in terms of deadweight, unless stated otherwise. For vessels within this range of deadweight, typical percentages are given for the percentages deadweight : displacement (column 3), steel weight : lightship weight (LSW) (column 4), outfitting weight : LSW (column 5) and machinery weight : LSW (column 6).

Based on column 3, the typical range for lightship weight : displacement can be calculated, according to Formula 4.22. To be clear, to calculate the upper limit for the percentage lightship weight : displacement, one should use the lower limit for the percentage deadweight : displacement in Formula 4.22. This calculation provides a typical range for the percentage lightship weight : displacement.

$$\frac{W_{\mathrm{LS}}}{\triangle} = 100 - \frac{\mathrm{DWT}}{\triangle} \tag{4.22}$$

Thereafter, the typical ranges for each of the weight components can be calculated according to Formula's 4.23, 4.24 and 4.25 . To be clear, to calculate the lower limit of the percentage steel weight : lightship weight for example, one should use the lower limit for the percentage lightship weight : displacement in Formula 4.23.

$$\frac{W_{\mathrm{ST}}}{\triangle} = \frac{W_{\mathrm{LS}}}{\triangle} \cdot \frac{W_{\mathrm{ST}}}{W_{\mathrm{LS}}} \tag{4.23}$$

$$\frac{W_{\mathrm{O\&E}}}{\triangle} = \frac{W_{\mathrm{LS}}}{\triangle} \cdot \frac{W_{\mathrm{O\&E}}}{W_{\mathrm{LS}}} \tag{4.24}$$

$$\frac{W_{\mathrm{M}}}{\triangle} = \frac{W_{\mathrm{LS}}}{\triangle} \cdot \frac{W_{\mathrm{M}}}{W_{\mathrm{LS}}} \tag{4.25}$$

Thus, the white box model, containing the weight estimation formulas, will be validated using Table 4.11. This can be done for the lightship weight and for each of the weight components.

Table 2.1 Typical sizes and percentages of weight groups for main merchant ship types (compilation of data from Strohbusch (1971), Schneekluth (1985), updated by Papanikolaou using IHS Fairplay World Shipping Encyclopedia, v. 12.01, 2011)

	1	**2**	3	4	5	6
Ship type	Limits		DWT/Δ (%)	W_{ST}/W_L (%)	W_{OT}/W_L (%)	W_M/W_L (%)
	Lower	Upper				
General cargo ships (t DWT)	5,000	15,000	65–80	55–64	19–33	11–22
Coasters, cargo ships (GRT)	499	999	70–75	57–62	30–33	9–12
Bulk carriers[a] (t DWT)	20,000	50,000	74–85	68–79	10–17	12–16
	50,000	200,000	80–87	78–85	6–13	8–14
Tankers[b] (t DWT)	25,000	120,000	78–86	73–83	5–12	11–16
	200,000	500,000	83–88	75–88	9–13	9–16
Containerships (t DWT)	10,000	15,000	65–74	58–71	15–20	9–22
	15,000	165,000[c]	65–76	62–72	14–20	15–18
Ro-Ro (cargo) (t DWT)	$L \cong 80$ m	16,000 t DWT	50–60	68–78	12–19	10–20
Reefers[d] (ft^3) of net ref. vol.	300,000	500,000	45–55	51–62	21–28	15–26
Passenger Ro-Ro/ferries/ RoPax	$L \cong 85$ m	$L \cong 120$ m	16–33	56–66	23–28	11–18
Large passenger ships (cruise ships)	$L \cong 200$ m	$L \cong 360$[e] m	23–34	52–56	30–34	15–20
Small passenger ships	$L \cong 50$ m	$L \cong 120$ m	15–25	50–52	28–31	20–29
Stern Trawlers	$L \approx 44$ m	$L \cong 82$ m	30–58	42–46	36–40	15–20
Tugboats	$P_B \cong 500$ KW	3,000 KW	20–40	42–56	17–21	38–43
River ships (towed)	$L \cong 32$ m	$L \cong 35$ m	22–27	58–63	19–23	16–21
River ships (self-propelled)	$L \cong 80$ m	$L \cong 110$ m	78–79	69–75	11–13	13–19

W_L light ship weight, W_{ST} weight of steel structure, W_{OT} weight of outfitting, W_M weight of machinery installation

[a] Bulk carriers without own cargo handling equipment

[b] Crude oil tankers

[c] Triple E class of containerships of Maersk, DWT = 165,000 t, first launched 2013

[d] Banana reefers

[e] Oasis class cruise ship of Royal Caribbean Int., $L = 360$ m, 225,282 GT, launched 2009

Table 4.11: Typical sizes and percentages of weight groups of main merchant ship types (from [29])

4.4. Black Box model

In order to choose a black box model, the tool requirements of Section 3.1 need to be taken into account. One of the requirements is that the tool should provide insight into the design process to naval architects. Therefore, it is beneficial to not use a too complex machine learning techniques in the black box model.

Another requirement is that the method should be a fast method.

Thereafter, the model should be robust for outliers. An example of such an outlier is the vessel *Ramform Titan*, which is shown in Figure 2.6. The machine learning model should not put too much weight on these somewhat odd vessels. A second advantage of this robustness, is that a machine learning model is not sensitive to errors in the database.

Multiple machine learning models have been explored. These models can be found in Table 4.12. These models were assessed based on the requirements. As can be seen, the random forest regression model is expected to be the appropriate model based on these requirements.

	Random forest model	Decision Tree	k-Nearest Neighbour
Complexity	Ok	Ok	Complex
Speed	Good	Good	Not good
Robustness	Good	Not good	Not good

Table 4.12: Assessment of different black box models

Random Forest Regression model

The black box model that is used is the Random Forest Regression model [7]. It is the Author's first known application of the Random Forest Regression model in ship design.

In a random forest model, many decision trees are used to predict a design parameter, such as lightship weight. Each decision tree gives a prediction of this design parameter. All the predictions are thereafter averaged to get one ensemble-averaged prediction. By averaging the results of each individual decision tree the variance of the prediction will reduce [36]. Decision trees are ideal for this method, because they can capture complex interaction structures in data [36]. As mentioned in *Duchateau (2016)* [14], early stage ship design is complex because of the large amount of continuous and discrete interrelated variables, constraints and objectives. It is expected that, using a decision tree method, these interrelations can be captured in a relatively simple manner.

Breiman (2001) [7] mentions that random forests are an effective tool in prediction. He mentions two advantages. The first is that because of the Law of Large Numbers, Random Forests do not overfit. This has to do with the large number of uncorrelated decision trees that are used in this model. The decision trees are uncorrelated because for each decision tree, a different sample of the data is used. Another advantage is that the model is accurate and stable, when the right kind of randomness is applied when constructing these decision trees. This means a sufficiently random distribution of the data across the different samples and secondly, randomness when the splits in a decision tree are determined. This latter will be further explained in sub-section *Decision Tree*. The accuracy of a random forest regression model is also a result of the large amount of decision trees that are used. In general, the more trees in the forest, the more robust the prediction is and thus, the higher the accuracy. By default, 100 decision trees are constructed in a random forest regression model [22].

Decision Tree

A decision tree can be used for regression and classification problems. The aim of a regression tree is to partition the data into smaller, more homogeneous groups [24]. In the context of this thesis, homogeneous groups mean a group with similar reference vessels. An example of a decision tree can be found in Figure 4.3, given in *James et al. (2009)* [36]. This tree is used for a regression problem. The goal of this tree is to predict the salary of a baseball player, based on the number of years that he has played in the major leagues and the number of hits that he made in the previous year. The salary can be seen as the dependent variable and the number of years and hits can be seen as the independent variables. The decision tree consists of a series of splitting rules, starting at the top of the tree. The splitting rules are based on the independent variables as can be seen in Figure 4.3. The prediction of the salary follows at the bottom of the decision tree.

In the same manner a data set of reference vessels can be split. An example of this can be seen in Figure 4.4. The variables *years* and *hits* are now replaced by the variables *length*, *breadth* and *draught* for example. The dependent variable *salary* is replaced by the design parameter *lightship weight*. Each decision tree will hereby be a reflection of a sample of the reference data. Based on this sample data, an algorithm determines how data can be split in two, at each node. For example, there are 100 vessels in the sample. 50 vessels are larger than 75 meter length and the other 50 are smaller. Then the algorithm can define the split-condition: L>75. This example can be seen in Figure 4.4. Now the length is chosen as a first split, but this can be any of the independent variables. A decision tree algorithm determines the 'best

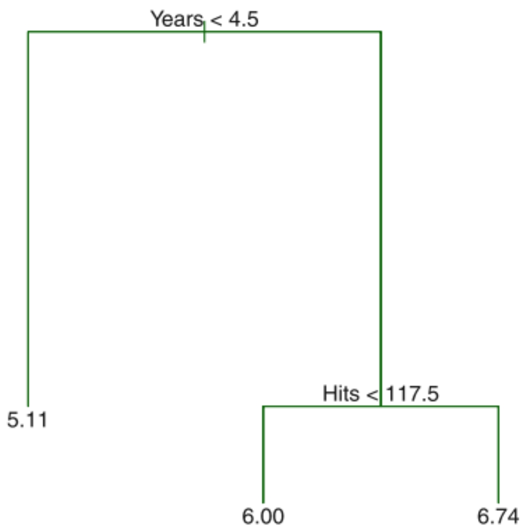


Figure 4.3: Example Decision Tree (from [36])

splits' using a Gini index [8]. By default, the Scikit Learn module uses the Gini index [22]. Basically, the Gini index method makes sure that the data is homogeneously distributed amongst the leaves of the decision tree. The Gini index varies between zero and one, where a Gini index score of 0 means that all elements of the data belong to a certain class. A Gini index of 1 means that the data is randomly distributed amongst the leaves. The 'best split' is the split with the lowest Gini index. The following example illustrates how this works. Let's say we have a splitting rule which puts 10 balls in a leaf and these 10 balls all have the same colour, red. Then the Gini index is zero, because all the balls belong to a certain class, in this case colour. But if we have a splitting rule which puts 5 red balls and 5 blue balls in a leaf, then the Gini index is 1. This means that data is distributed randomly. This is undesired as the aim for regression groups is to partition data into smaller and more homogeneous groups [24].

In this manner an algorithm determines how data should be split at each node. This goes on until a leaf of the decision tree only contains one data point, or one reference vessel in this case. Thus, according to the split-rules that the algorithm has defined, it can now make a prediction for the lightship weight, given that sample of the reference data.

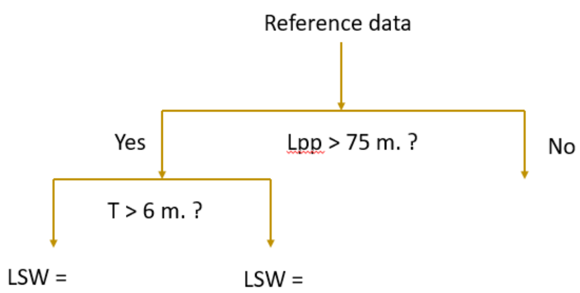


Figure 4.4: Decision tree example 2

Ensemble averaged prediction

The prediction for an individual decision tree is thus very dependent on the sample of training data. Therefore, in random forest regression models a large number of decision trees is used, all based on a random sample of training-data. The predictions of all the individual trees are ensemble-averaged, which makes the prediction significantly less dependent of one particular training data sample. The formula for the ensemble-averaged prediction can be seen in Formula 4.26 [36]. In this manner, such a model can predict the lightship weight of a novel vessel for example, based on reference data.

$$\hat{f}^{B}_{RandomForest}(x) = \sum_{b=1}^{B} T_b(x) \tag{4.26}$$

The variance of the ensemble averaged prediction of all trees is lower than the variance of the prediction of each individual tree [7, 24]. This means that the difference between the predicted value and the actual value is smaller. In other words, the 'Average All Predictions' as is shown in Figure 4.5 outperforms each individual prediction, in this case prediction 1 until prediction 600. As already mentioned, in the proposed hybrid model, 100 decision trees are constructed in each random forest regression model. This is a default value.

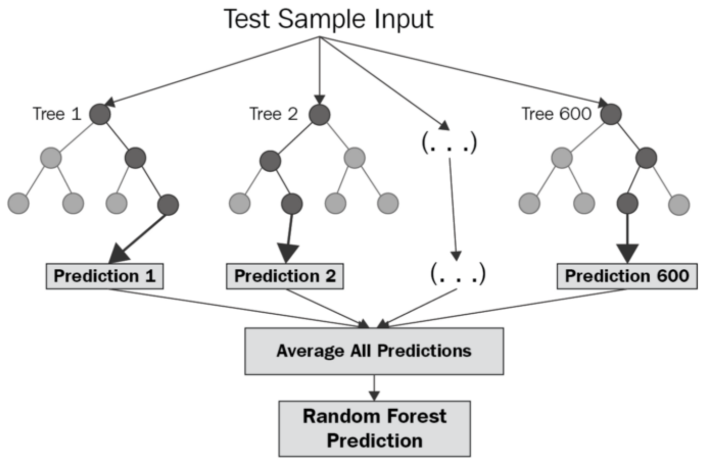


Figure 4.5: Random Forest mode (from: medium.com (2019))

4.5. Parallel hybrid model

In the proposed parallel hybrid model the above mentioned white box- and black box model are used. In this model the first white box model is used to provide an estimation of a parameter, such as lightship weight or resistance. Thereafter the model calculates the difference between the estimation and the true values of this parameter using Formula 4.2.

The calculated differences are fed to the black box model as the dependent variable. The independent variables are defined by the naval architect and depend on the parameter that will be predicted. For lightship weight the independent variables should at least contain the dimensions of the vessel. For the resistance of the vessel one can imagine that the service speed and the frontal - and wetted surface are important. Therefore the independent variables should at least contain the waterline length, the breadth, the draught and the service speed. With both the dependent and independent variables fed to the black box model, the model can be trained. This is the training phase.

In the prediction phase, the calculated estimation of the white box model and the predicted correction of the black box model are summed according to Formula 4.1. The result is the parallel hybrid model prediction of a certain design parameter, such as the lightship weight and the resistance.

4.6. Performance of model

To measure the performance of the white box, black box and parallel hybrid model two performance measurements techniques have been used. These will be explained in the following sub-sections.

4.6.1. Performance metrics for regression models

There are three main metrics to measure the performance of a regression model. These are

1. R^2 [24]
2. Root Mean Square Error (RMSE) / Mean Square Error (MSE) [24]
3. Mean Absolute Error (MAE) [34]

The **R^2-method** shows how much of the variation in the dependent variable can be explained by the model. Thus, an R^2 value of 0.75 implies that the model can explain three-quarters of the variation in the outcome [24], i.e. the dependent variable. The R^2 is calculated with Formula 4.27. In this formula y_i represents the actual value of one dependent variable, $\bar{y}$ represents the average value of all the dependent variables and $\hat{y}_i$ is the estimated value of ship i. The closer the R^2 score is to one, the smaller the difference between the actual and the predicted value and thus the better the model. Therefore, the R^2-method is a good method to evaluate how good a model is, i.e. how good a model fits the data. The R^2-score is not a measurement for accuracy of the model.

However, one of the limitations of the R^2-method is that it is possible that a worse model leads to a higher score. This has to do with the potential overfitting of a model. This is explained in Figure 4.6. The blue line represents the true relation between two variables, and the dots represent true values. The values can be better approximated with the red curve, and thus this model has a higher R^2-score. Still, it is not a better model than the blue line as it doesn't capture the true (linear) relation. As a result, a model can fit very well to the (training) data that is used to build the model, but may fit very poorly to new (testing) data. Basically, overfitting means that a model becomes more complex than the process that one wants to model. The possibility of overfitting needs to be taken into account when the R^2-scores are assessed. An indication of overfitting is when the R^2-score for training data is significantly higher than the R^2-score for unseen testing data.

An advantage of this method on the other hand is that one is able to compare the performance of different models, because the performance is presented in a relative manner. This means that the performance of a lightship weight model and a resistance model can be compared immediately.

$$\hat{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{4.27}$$

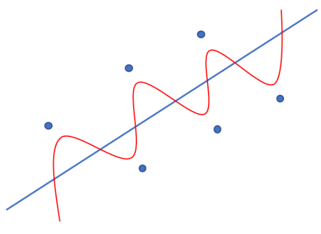


Figure 4.6: Overfitting model

The **Root Mean Square Error-method** (RMSE) is a method that represents the standard deviation of the difference between the predicted values and the true values. Compared to the R^2-method this method provides an absolute value on how much the predicted values deviate from the true values, instead of a relative value. Therefor the RMSE-method is a good method to determine the accuracy of a model. To calculate the RMSE the equation 4.28 is used.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \cdot \sum_{j=1}^{n} \left(y_j - \hat{y}_j\right)^2} \tag{4.28}$$

The **Mean Absolute Error-method** (MAE) represents the average of the absolute difference between the predicted values and the true values. The MAE is calculated via formula 4.29. Compared to the MSE or RMSE method, the MAE is more robust to outliers or large errors. This means that a model that produces some large errors can have a good MAE score. This is undesirable, because a model is desired that has little error for all the estimations.

$$\mathrm{MAE} = \frac{1}{n} \cdot \sum_{j=1}^{n} |y_j - \hat{y}_j| \tag{4.29}$$

Conclusion
R^2- method is a common method to compare the performance of different models. This is possible because the performance is given as a relative value. This is an advantage. Thereafter, the R^2-method is relatively easy to interpret. A goal of this thesis is that the proposed parallel hybrid model will be accepted by the naval architects. Thus, good interpretability is an advantage.

Therefore the R^2- method will be used to assess the performance of the empirical formulas. The formulas with the best performance will be selected for the further development of the parallel hybrid model.

4.6.2. Performance metrics for machine learning models

In machine learning models, data is often split in training-data and test-data. The reason for this is that if a model has been trained to predict the dependent variable based on the independent variables, it already knows the answers to that particular data set, the training data set. Therefore, to tell something about how well a model can predict the dependent variable, it should be tested with data that it doesn't already know. Three validation methods that deal with this problem will be discussed.

The first is **k-fold cross-validation** [24]. In k-fold cross validation samples are randomly partitioned into k sets of roughly equal size. The model is trained based on the data of the first fold or subset of the data. The rest of the subsets are used to test the data. For k=10, thus in 10-fold cross validation, 90% of the data is used for training and 10% for testing. As k gets larger, the larger the portion of the data that is used for training. This leads to a more accurate prediction. Literature describes that 5 and 10 are usual values for k [24]. The predicted values for the test-data are then compared to the true values. Using the R^2-method, a score for this prediction can be obtained. In 10-fold cross validation, this is done 10 times for 10 different sets of training and test data. An example of this can be seen in Figure 4.7. This is an example of 4-fold validation. As can be seen, each data point is used both for training and testing.

The scores for each prediction are thereafter averaged. This is done to make sure that the predictive capability converges to a constant value and is not dependent on the set of training and test data that is used. A luckily chosen combination of training and test-data can, for instance, lead to a low error and high predictive capability and vice versa.

The second validation method is the **Leave One Out-method** [24]. In this method one data point is left out of the training-set. This means that the model uses nearly all the available data to train. More training data generally means a more accurate model. On the other hand, the performance of the model using this method, is very dependent on how well one data point fits in this model. So this method calculates the performance of a model, but with low certainty that it is the true performance. 10-Fold cross validation deals with this problem. First, by taking a large data-set as testing data and secondly, by using different combinations of training and test data.

Finally, data can be **manually split into Train & Test data**. The goal of this method is to force a particular distribution of training data and test data. This method is interesting because one of the

Figure 4.7: k-fold cross validation

research questions is about how well the model performs in data sparse areas of design space. In other words, in using the Train & Test split method one can select an area in the design space and put all the available data in the test data set. A model can thereafter be trained with all other available data. The resulting performance gives a good indication of how well the model performs in data-sparse areas. A naval architect can then determine if this performance it good enough, and if this model can be used in the ship design process.

Conclusion

The 10-fold cross validation will be used to determine how good the black and grey box models are. This is because the performance is measured multiple times for different sets of training and test data. Therefore this is a good method to measure the performance of the model, independent of the data that is used for training or testing.

Based on the Train & Test split method, different sets of training and test data will be chosen to measure the performance of the model in different situations. Situations that reflect some ship design problems that are mentioned already in Chapter 2. For example, designing a new vessel in a data-abundant region of the design space.

4.7. Conclusion

In this chapter, an overview is given about how the proposed parallel hybrid model is constructed. All the different sub-models have been discussed.

First of all the white box model was discussed. Different methods were explored that are able to predict the lightship weight of a ship, or a certain weight component of the lightship weight. Methods were found for different ship types. Based on the applicability of these methods and the availability of reference data, it was chosen to test four ship types. These are the bulk carrier, the general cargo ship, the oil tanker and the container ship. The different weight estimation methods for these ship types will be assessed with the R^2-method. This method expresses the performance of the weight estimation methods as a relative value. Based on this performance, one weight estimation method is selected for each ship type. These methods are implemented in the white box model. The performance of this model describes the performance of methods that were already available in literature. This is important when the performance of the black box and parallel hybrid model are assessed as well. Based on this one can tell if the ship design process has thus been improved by this research.

Secondly the black box model was described in detail. As a black box model the Random Forest Regression Model is used. In this method 100 uncorrelated decision trees are randomly constructed. The individual predictions of each tree are ensemble averaged in order to derive the prediction of the entire random forest model. This prediction out-performs any of the predictions of the individual decision trees.

The white box can now be used to estimate the lightship weight for four ship types and the black box model can be trained to predict the correction that should be applied, based on the data. This approach is the parallel hybrid modelling approach.

The performance of both the black box model and the parallel hybrid model are assessed with the 10-fold cross validation. This gives a good and robust indication of the models' performances. The Test & Train split method will be used to simulate certain situations in the ship design process. For example, how well can a model predict the lightship weight in a region of a design space where data is lacking.

5

Experiments & Results

This chapter will describe the experiments that have been conducted in order to test the proposed model. First, an explanation is given about each experiment and their goal. Secondly, the performance of the current design approach will be determined, i.e. the white box model. The same will be done for the black box model approach and the proposed parallel hybrid model approach. Based on this performance assessment a conclusion can be drawn about whether or not the proposed solution is an improvement of the current design approach and an improvement of the initial Maritime Intelligence Tool.

5.1. Design of experiments

To evaluate the model and its components, the following questions have been determined that should be answered based on experiments.

What is the performance of the white box model? - The answers to this question will give insight into how accurate the methods are that are currently used to estimate the lightship weight. Based on these results, methods will be selected that will be used in the white box model. This experiment is described in Section 5.2.

What is the performance of the black box model and the parallel hybrid model? - The performance of the black box model and parallel hybrid model should be determined to compare it with the performance of the white box model. In this way one can compare three different design approaches, namely

1. A design approach based on knowledge (White box model)
2. A design approach based on statistics (Black box model)
3. A design approach based on statistics and knowledge (Parallel hybrid model)

A description of this experiment and the results are given in Section 5.3.

How do these models perform with a smaller training set? - The current MIT performs well when sufficient data is available, but when the available data is lacking, the performance drops significantly. This should be examined and compared to the performance of the proposed parallel hybrid model in a similar situation. It is expected that a model that is based on both data and on knowledge, i.e. the parallel hybrid model, performs better in this case. That would mean that a parallel hybrid model performs better at the boundaries of a design space. That is beneficial, because it is expected that novel and innovative ship designs are located in these boundary regions. This experiment is described in Section 5.4.

What is the performance of the model in interpolation and extrapolation gap areas? - This experiment is similar to the previous experiment, namely to measure the performance of the model in a region where data is lacking. The difference is that in this experiment specific data is selected for the training and the test data set, based on one of the variables, e.g. length or deadweight. This means that the focus is not on reducing the size of the training set, but on predicting the lightship weight of vessels without any similar vessels in the training data set. In this way, gaps in a design space are manually created. These are interpolation gaps and extrapolation gaps. This experiment will be further explained in Section 5.5.

When should a naval architect rely on which model? - This questions will be answered qualitatively, based on all the experiments that are conducted. Therefore, first the results of the first two questions will be discussed in a different perspective in Section 5.6. It is important to compare the results of these different questions, because that will give a good indication of the overall performance of the different models. After that, a naval architect is able to choose a model that best deals with a certain problem in ship design. This question will be answered in Section 5.7.

For each experiment, first a description of the experiment will be given. Thereafter follow the results of that experiment and a discussion of these results. Finally, a conclusion will be given for each experiment containing the most important findings.

5.2. Performance white box model

The goal of this experiment is to determine the performance of multiple weight estimation methods. As mentioned in Section 4.6, to measure the performance of the estimation methods, the R^2-method is used.

Experiment

In this experiment the choice was made to test multiple combinations of weight estimation methods and measure the total performance. To do this, a method was chosen for each of the three weight components (steel, outfitting & equipment and machinery), a prediction was done for each component and the results were summed. The result is a prediction for the lightship weight, for which a R^2-score can be determined. The different combinations of methods and their R^2-score can be found in Table 5.1 for each ship type. Based on these scores, methods will be selected for implementation in the white box model.

Ideally, one would measure the performance of each weight estimation method individually. As a result, a naval architect is able to select the best weight estimation method for each weight component. However, this was not possible, because there was no reference data available in the database for each weight component.

Results

The results of this experiment can be found in Table 5.1. The best R^2-scores are printed in bold. There are three remarks to this table.

First of all, it was expected that the values for the R^2-scores varied between 0 and 1. As can be seen in Table 5.1, this is not the case for experiment 12, 13, 14, 15, 20 and 21. The combinations of estimation methods in these experiments lead to a negative R^2-score. A negative R^2-score means that the model does not follow the trend in the data and the model fits worse than a horizontal line.

An explanation for the negative scores is that the methods are outdated. These methods are the methods from *DNV (1972)*, *Sato (1967)* and *Miller (1968)* for steel weight. These methods are empirically derived and based on ship data that was available at the time. Since that time, ship sizes have increased significantly. This development can be seen in Figure 5.1 for container ships. With this increase in ship size, also the longitudinal bending moments increase significantly, and thus, significantly more structural steel is required. In other words, the lightship weight of today's large ships are very hard to estimate based on reference data from approximately 1970. As the steel weight is the largest component of the light ship weight, the performance of a model is also mostly determined by the performance of the steel weight estimation method. For a crude oil tanker the steel weight is approximately 80% of the lightship weight and for a container ship this is around 65% [29].

Secondly, it is noticeable that the method from *D'Almeida (2009)* also performs relatively well for all ship types. This method is also the youngest as it is from 2009. Thus, unlike the methods from *DNV (1972), Sato (1967)* and *Miller (1968)*, most of the new and larger ships are taken into account in *D'Almeida (2009)*. Not surprisingly, this leads to a better fitting model and thus to a higher R^2-score.

Thirdly, the generic method has a higher performance than was expected. As mentioned in Section 4.3.1, it is expected that a generic method, which is thus applicable to every ship type, would perform worse than a ship type specific method. As can be seen in experiment 5 and 11, this is not the case.

Experiment	W_{ST}	$W_{O\&E}$	W_M	R^2-score
Bulk carrier				
1	Generic method	-	-	0.5774
2	Murray	Papanikolaou	Watson	0.714
3	Murray	Papanikolaou	Munro-Smith	**0.7649**
4	D'Almeida	D'Almeida	D'Almeida	0.6017
General cargo ship				
5	Generic method	-	-	**0.8322**
6	Wehkam-Kerlen	Papanikolaou	Watson	0.6999
7	Wehkam-Kerlen	Papanikolaou	Munro-Smith	0.7528
8	Carreyette	Papanikolaou	Watson	0.0679
9	Carreyette	Papanikolaou	Munro-Smith	0.1755
10	D'Almeida	D'Almeida	D'Almeida	0.8077
Oil tanker				
11	Generic method	-	-	**0.8843**
12	DNV	Papanikolaou	Watson	-2.1399
13	DNV	Papanikolaou	Munro-Smith	-1.9702
14	Sato	Papanikolaou	Watson	-0.3219
15	Sato	Papanikolaou	Munro-Smith	-0.2133
16	D'Almeida	D'Almeida	D'Almeida	0.6997
Container Ship				
17	Generic method	-	-	0.6084
18	Chapman	Papanikolaou	Watson	0.4934
19	Chapman	Papanikolaou	Munro-Smith	0.5949
20	Miller	Papanikolaou	Watson	-1.6282
21	Miller	Papanikolaou	Munro-Smith	-1.371
22	D'Almeida	D'Almeida	D'Almeida	**0.8942**

Table 5.1: R^2-scores for different combinations of ship type specific methods and the generic method

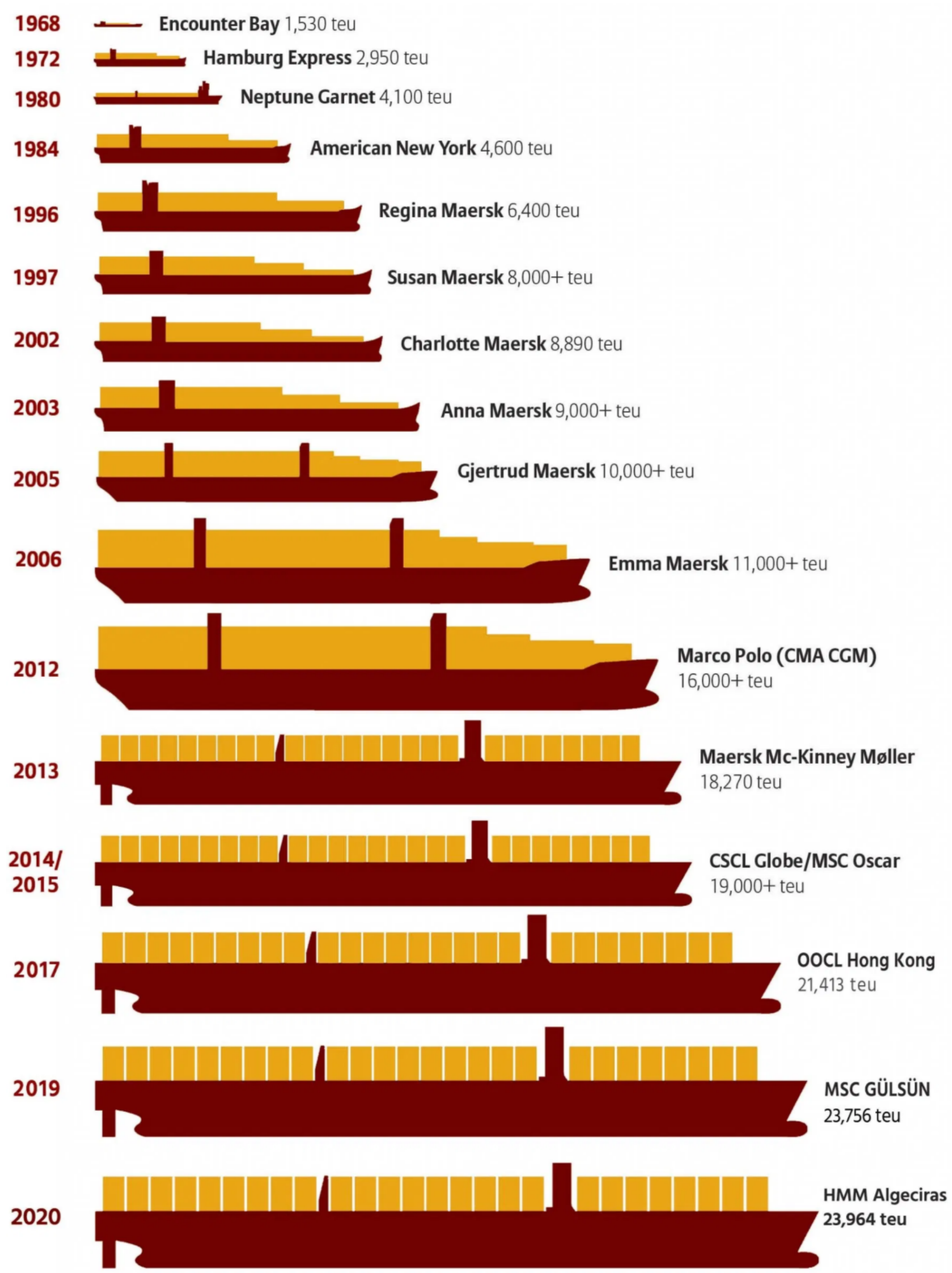


Figure 5.1: Development of container ship size (from Shipping and Freight Resource, 2021)

Finally, to give an idea of the accuracy of the white box model, in Figures 5.2, 5.3, 5.4, 5.5 the estimated values for lightship weight are plotted against the real values. Ideally, the estimated values and the real values are equal, which would put a dot on the diagonal line in the scatter plot. As can be seen in the figures, for all the ship types most of the dots are below the diagonal line. This means that the white box model mostly underestimates the lightship weight.

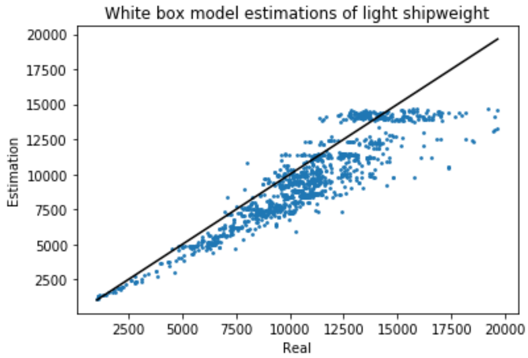


Figure 5.2: White box model results - Bulk carrier

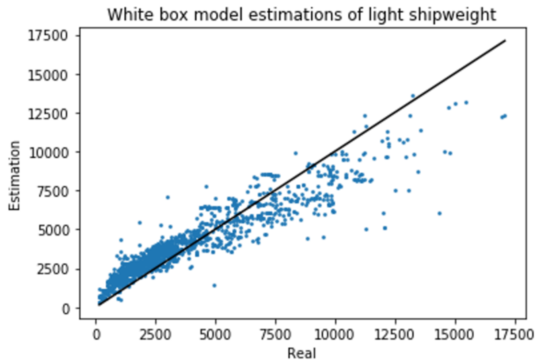


Figure 5.3: White box model results - General cargo ship

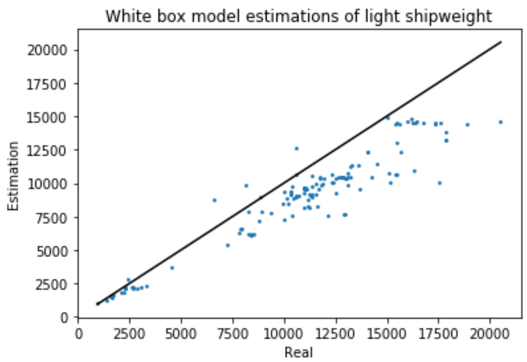


Figure 5.4: White box model results - Oil tanker

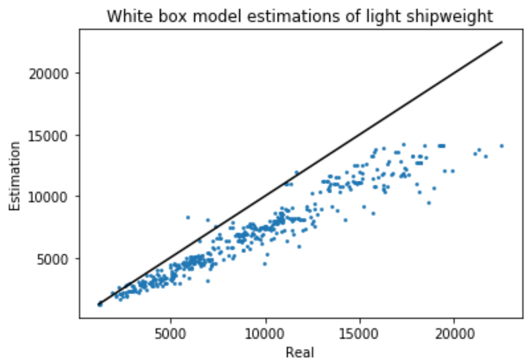


Figure 5.5: White box model results - Container ship

Conclusion

Based on the results of the performance tests of the weight estimation methods, the best methods are selected for further implementation in the white box model. The selected methods per ship type can be found in Table 5.2. These methods are referred to as the white box model for the rest of this report.

Ship type	W_{ST}	$W_{O\&E}$	W_M
Bulk carrier	Murray	Papanikolaou	Munro-Smith
General Cargo Ship	Generic method	-	-
Crude Oil Tanker	Generic method	-	-
Container Ship	D'Almeida	D'Almeida	D'Almeida

Table 5.2: Selected methods per ship type

5.3. Performance of black box model and parallel hybrid model

In this section the performance of the black box model and the parallel hybrid model will be discussed. The goal is to determine how the performances of the white box model, black box model and parallel hybrid model relate to each other.

Experiment

As is mentioned in Section 4.6.2, a model should not be tested with the same data that it has been trained with. In k-fold cross validation, multiple random combinations of training-data and test-data are used to measure the predicting capability or performance of the model. As mentioned in *Kuhn & Johnson (2016)* [24], by choosing $k = 10$, it can be expected that the measurement for the performance is sufficiently accurate.

The black box model and the parallel hybrid model are tested with 10-fold cross validation. For each ship type all the available reference data is used in the 10-fold cross validation method. In this method, 90% of the input data will be used to train the model and 10% of the data will be used for testing. This is done 10 times, for different sets of training and test data and the resulting performance scores will be averaged.

Results

In Table 5.3 the results of this experiment can be found. First the size of the training data set and test data set are given. Secondly, the 10-fold cross validation scores are given for the black box model and the parallel hybrid model. The white box R^2-score is added for comparison.

Ship type	Train data	Test data	Black Box	Parallel Hybrid	White Box
Bulk carrier	1095	121	0.936	0.942	0.765
General Cargo Ship	1411	156	0.967	0.952	0.832
Crude Oil Tanker	119	13	0.923	0.922	0.884
Container Ship	408	45	0.932	0.943	0.894

Table 5.3: 10-fold cross validation scores for different models and ship types

First of all, looking at the scores in Table 5.3 for the black box model and the parallel hybrid model, it can be seen that those are all quite high. As mentioned in Section 5.2, an R^2-score of 1 means that a 100% of the variance of the outcome can be explained by the model, which means that it is a good model. All the black box and parallel hybrid model R^2-scores are higher than the score for the white box model, although it is not far off.

Secondly, for both the general cargo ship and the container ship, the score of the parallel hybrid model is higher. It was expected that the parallel hybrid model had a higher score than the black box model, for each ship type. This seems a bit strange, but it should be noted that the differences between the scores are very small. In fact, it is so small that the differences can be neglected. The reason for this is that the distribution of training and test data still has got influence on the 10-fold cross validation score. The influence is limited, by choosing k = 10 in k-fold cross validation, but still it can be seen. In Section 5.4 an example of this is given.

Thirdly, comparing the size of the training data set with the 10-fold cross validation scores for the black box model, it can be seen that the larger the training data set, the higher the score. The 10-fold cross validation scores of the black box model are plotted against the size of the training data set in Figure 5.6. This result was expected. The relation between the size of the training set and the performance of the model will be described in more detail in Section 5.4

And finally, these scores for performance are still a bit abstract. The scores are determined by a 10-fold cross validation method, which uses the R^2-method to calculate the error. The R^2-score is a measurement for the fitness of the model to the reference data, but it is not a measurement for absolute accuracy. To give a better idea of the accuracy, the predictions for lightship weight by the black box model and

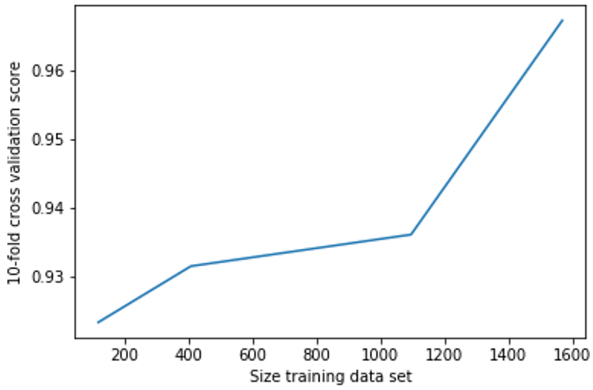
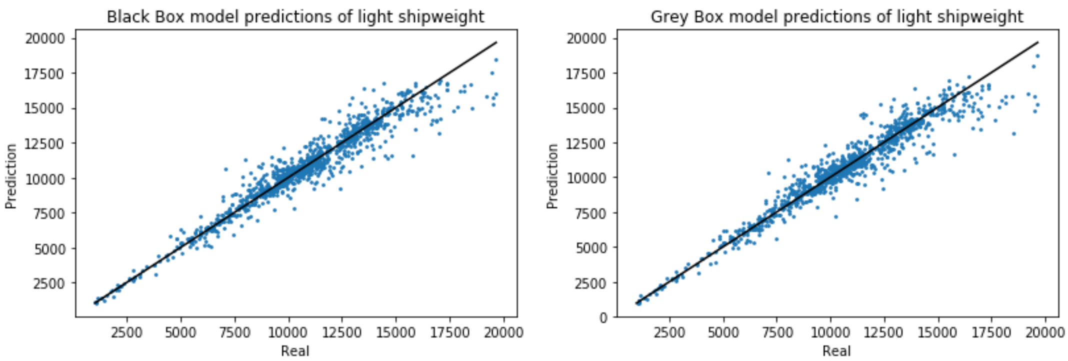


Figure 5.6: Size of training data set versus the 10-fold cross validation score for a black box model

the parallel hybrid model are plotted against the true values for lightship weight. This can be seen in Figures 5.7, 5.8, 5.9, and 5.10, for the bulk carrier, general cargo ship, oil tanker and the container ship respectively. It can be seen that the plots for the black box model and the parallel hybrid model are quite similar. This is also expected because the 10-fold cross validation scores are similar. It is especially interesting to compare these plots with the white box model plot.

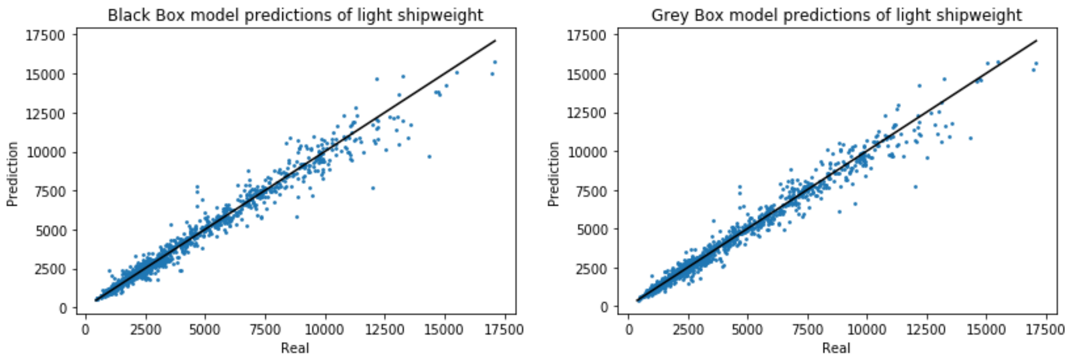
Thus, taking the bulk carrier as an example, Figures 5.7a and 5.7b are compared to Figure 5.2. It can be seen that the dots are much closer to the diagonal line, which represents a perfect model. This is also the case for the general cargo ship, oil tanker and the container ship. For these figures the same reference data is used to make estimations with the white box model, as is used to make predictions with the black box model and parallel hybrid model. Thus, by comparing these figures it is clear that both the black box model and the parallel hybrid model outperform the white box model, as the data in these figures is much more concentrated near the diagonal. Based on Figures 5.7a and 5.7b, significant differences between the black box model and parallel hybrid model cannot be seen.



(a) Black box model result

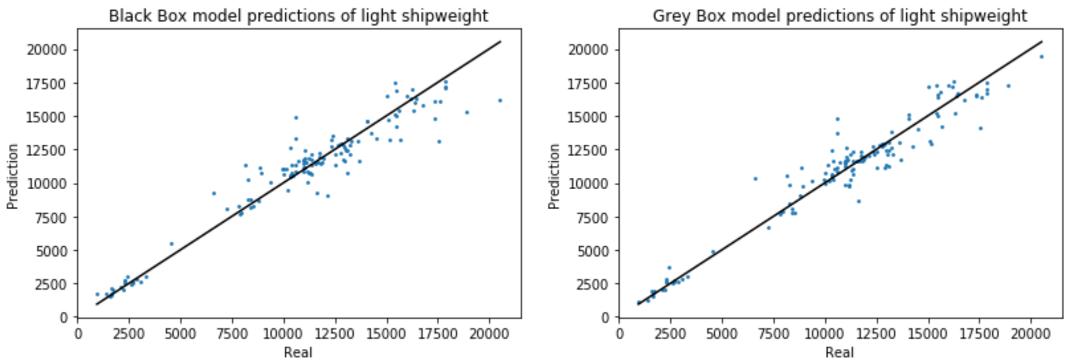
(b) Parallel Hybrid model result

Figure 5.7: Lightship weight predictions - Bulk carrier



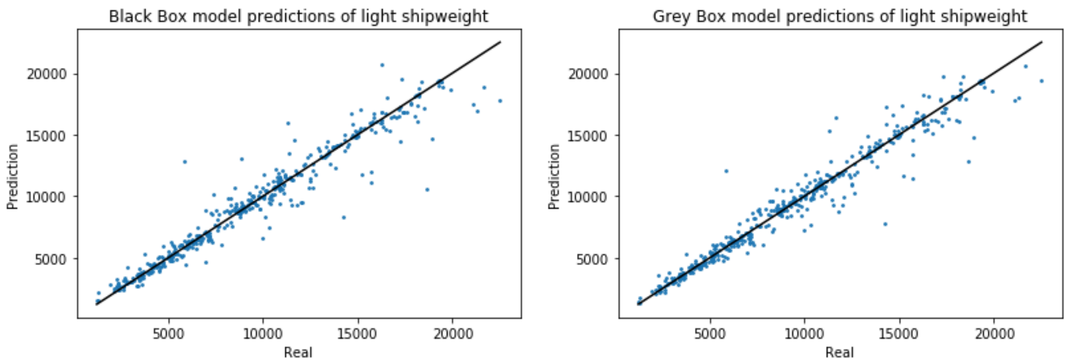
(a) Black box model result (b) Parallel Hybrid model result

Figure 5.8: Lightship weight predictions - General cargo



(a) Black box model result (b) Parallel Hybrid model result

Figure 5.9: Lightship weight predictions - Oil tanker



(a) Black box model result (b) Parallel Hybrid model result

Figure 5.10: Lightship weight predictions - Container ship

Conclusion

The 10-fold cross validation scores for the black box model and the parallel hybrid model are quite high. By comparing these scores to the white box model score and by comparing the plotted results for the three models, it can be concluded that the black box model and the parallel hybrid model outperform the white box model for each ship type.

Thereafter, by comparing the scores and plotted results from black box model and the parallel hybrid model scores, little difference can be seen. It is expected that these differences are negligible as it is expected that it is primarily due to the distribution of training and test data. This distribution is about which reference vessels are in the training data set and test data set. It is not about the ratio between training data set size and test data set size. In Section 5.4 this further explained.

5.4. Influence of the training data set size

As is mentioned in Section 1.4, the initial MIT, which is a data-based model, performs badly when data is lacking. In this experiment the goal is to determine how the performance of this model is influenced by the amount of available data. This will be done for the black box model and the parallel hybrid model.

Experiment

In this experiment the size of the training data set is varied from 1% until 99% of the data. In the first iteration, 1% of randomly selected data is used to train a model. The trained model will be tested with all other available data and the R^2-score is calculated. This is done ten times with different samples of training data and the R^2-scores are averaged. This is similar to 10-fold cross validation. As mentioned in Section 4.6.2, this is done to make the performance independent of which data is used as training data and test data. This is repeated until the last iteration where 99% of the data is used for training and 1% of the data is used for testing the trained model. The distribution of data is shown in Figure 5.11. The blue curve represents the size of the training data set. In the first few iterations the step was 1%. The step size was increased for larger size of the training set. This was done in order to reduce calculation time. It was possible to do this, because, after some experiments, it was clear that the performance converged to a certain value. Therefore, a small step size was not required.

The size of the training data set will be referred to as N_{Train} and the size of the test data set will be referred to as N_{Test}

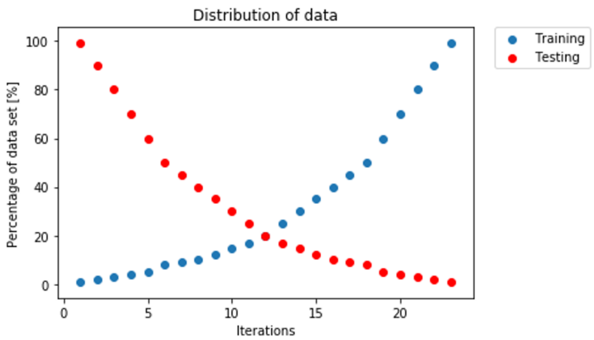


Figure 5.11: Distribution of data over the training- and test data set

Results

The results of this experiment are plotted in Figures 5.12, 5.13, 5.15 and 5.16. These are the results for the bulk carrier, container ship, oil tanker and general cargo ship respectively. In these graphs the averaged R^2-scores are plotted against N_{Train}. The blue curve represents the black box model and the red curve the parallel hybrid model.

Figure 5.12 shows what one would expect. With a small N_{Train}, the parallel hybrid model performs significantly better that the black box model. As N_{Train} increases, the performances of both models converge to a constant value. Although the figure shows that the parallel hybrid model performs better over the entire range, it should be noted that the differences are very small for a larger N_{Train}.

Figure 5.13 shows the result for container ships. It is similar to Figure 5.12, only in this graph the curves go up a little bit at an N_{Train} of around 400. This increase in performance can be neglected, because it is a result of the random distribution of data across the training and test set. When either N_{Train} or N_{Test} becomes small, the influence of the random distribution increases.

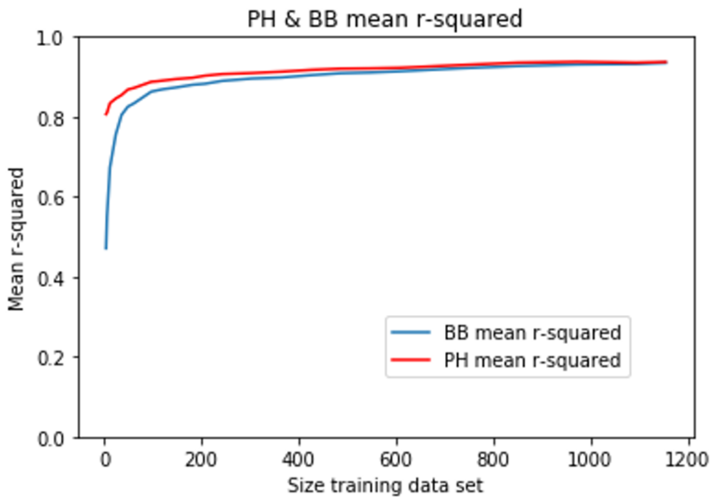


Figure 5.12: Increasing the N_{Train} of the bulk carrier

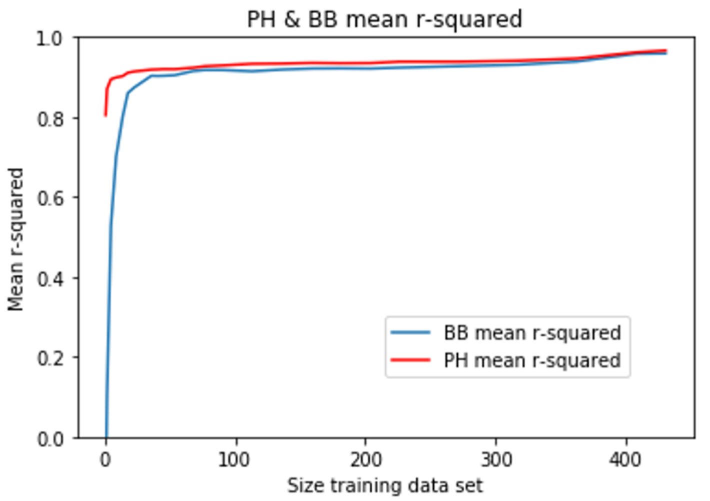
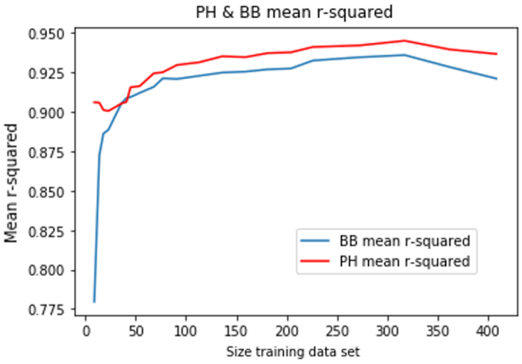
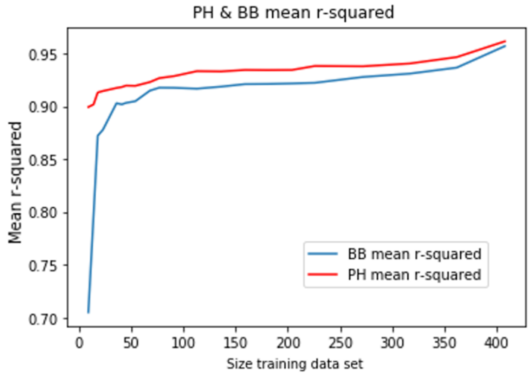


Figure 5.13: Increasing the N_{Train} of the container ship

This influence of the random distribution is shown in Figures 5.14a and 5.14b. These figures are the results of the same experiment as Figure 5.13, but the y-axis scale differs. As can be seen in Figure 5.14a, at an N_{Train} of approximately 25 the performance decreases first before it increases. In general, the higher the N_{Train}, the higher the performance so this decrease in performance should not be there. Secondly, this happens again for N_{Train} of approximately 325 and higher. Here the performance decreases again. Figure 5.14b shows the results of the same experiment as in Figure 5.14a, but now a different random distribution is chosen. In this graph the decrease in performance is gone. Thus, when assessing the performance for small training or N_{Test} one should take into account the influence of the random distribution. Based on Figures 5.14a and 5.14b, it is clear that for either a small training or test data set size (< 75 reference vessels), the random distribution has influence on the performance scores. Therefore, in this case, performance score differences lower than 0.03 should be neglected.



(a) Random state 1



(b) Random state 2

Figure 5.14: Increasing the N_{Train} of the container ship

For the oil tanker results in Figure 5.15, this influence can also be seen. The decrease in performance at an N_{Train} of 120 is a result of the random distribution. For an N_{Train} higher than 25 approximately, the two models perform equally. For small N_{Train} it is clear that the parallel hybrid model outperforms the black box model for small N_{Train}.

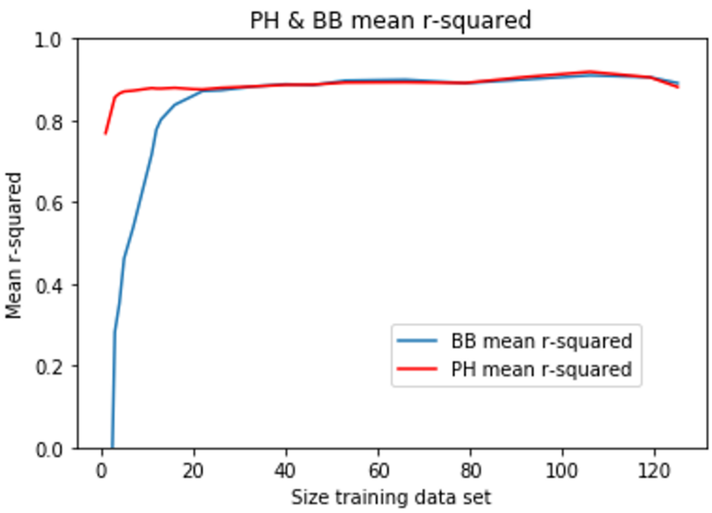


Figure 5.15: Increasing the N_{Train} of the oil tanker

Finally, the result for the general cargo ship in Figure 5.16. In this experiment the black box model outperforms the parallel hybrid model on almost the entire range of N_{Train}. Even though the difference in performance is little, this was not expected. Different random states have been tried, but each time the black box model outperformed the parallel hybrid model. Thus, the random state was not the cause. It is expected that the chosen white box model is a reason for this behaviour. As can be seen in Figure 5.3, the white box model for the general cargo ship first overestimates the lightship weight. For larger vessels, this model underestimates the lightship weight. Thus, the correction that should be applied, should first be negative and then positive. This inconsistency might result in that the parallel hybrid model performs worse than the black box model. It should be noted that the difference between the models is very small and might be negligible.

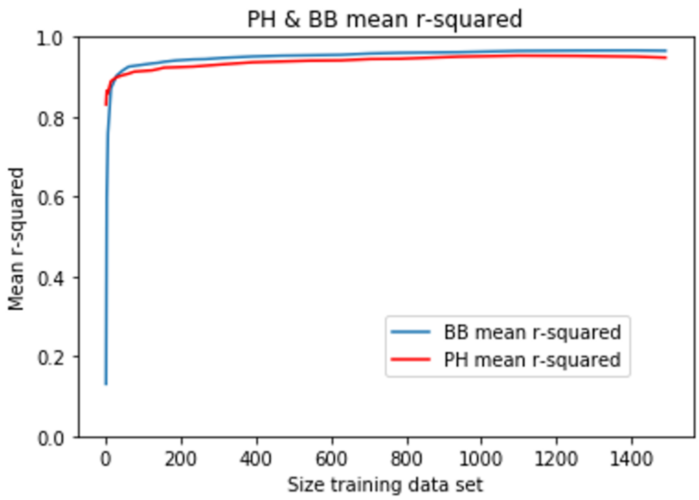


Figure 5.16: Increasing the N_{Train} of the general cargo ship

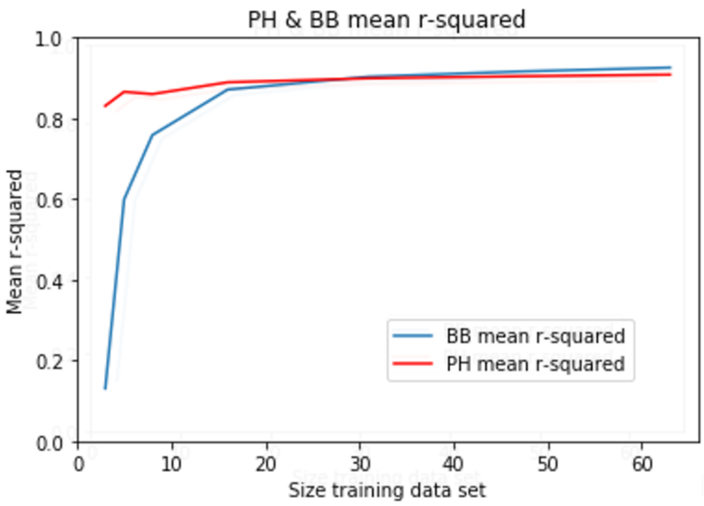


Figure 5.17: Increasing the N_{Train} of the general cargo ship - only small N_{Train}

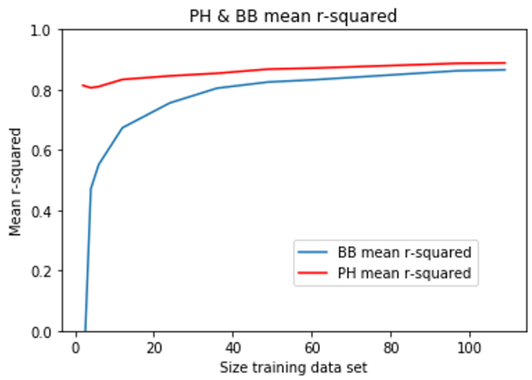
Conclusion

Based on the results a few conclusions can be drawn from the performance of the black box model and the parallel hybrid model for different sizes of the training data set.

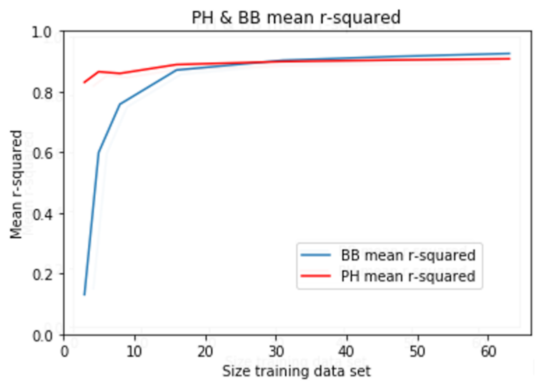
First of all, the parallel hybrid model performs better for small sizes of the training data set. This can be seen for all the experiments, thus for all ship types. Even for the smallest size of the training set that is used in the experiments, the performance of the parallel hybrid model is at least 0.8 approximately. This in contrast to the black box model which shows a significant decrease in performance for smaller sizes of the training data set. Thus, the parallel hybrid model is more consistent over the entire range of training data set sizes than the black box model.

Secondly, when the size of the training data set increases the performance of the black box model will converge to the performance of the parallel hybrid model. For the general cargo ship the black box model outperforms the parallel hybrid when the size of the training data set increases. The differences between the performance of the black box model and the parallel hybrid model on the other hand are relatively small. Therefor is it difficult to draw a hard conclusion on this matter or to point out what the actual cause of these differences is. Thus, in general for all ship types, it is clear that the performances of both models converge to a maximum value.

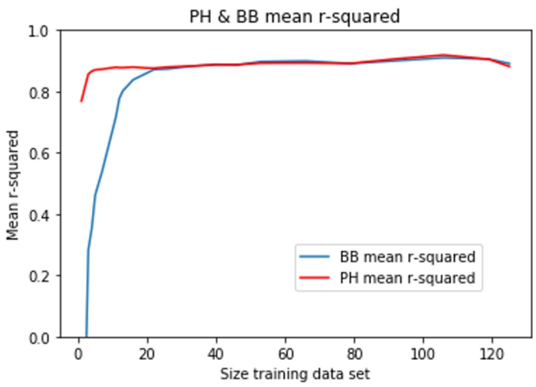
Finally, Figure 5.18 shows the results for the four different ship types. These are the same results as is shown already in the section, but the performances of larger training data set size are left out. Based on the graphs in this figure one can determine when to use the black box model and when the parallel hybrid model. In general, for training data set size smaller than approximately 50 reference vessels, a naval architect should use the parallel hybrid model. For higher training set sizes both models perform practically the same. The black box model performance for all ship types is stable for training set sizes higher than approximately 35. Thus, this is the smallest size for a training data set when a naval architect wants to use the black box model.



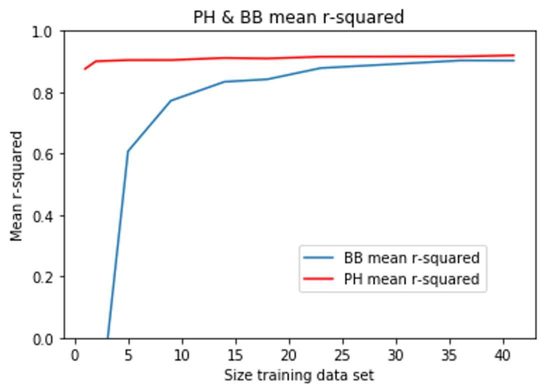
(a) Bulk carrier



(b) General cargo ship



(c) Oil tanker



(d) Container ship

Figure 5.18: Results for four ship types, focused on the small training data set sizes

5.5. Performance of model in gaps of the design space

The goal of this experiment is to assess the performance of the black box model and parallel hybrid model in areas of the design space where there is no data. In this experiment the focus is not on the size of the training data set, but on predicting the lightship weight of innovative and novel ships. These vessels are considered to be outside the boundaries of a design space. Thus, in this experiment the lightship weight of ships outside these boundaries will be predicted and the performance of the model will be determined.

Experiment

In this experiment specific reference vessels are selected for the train data set and test data set. This is the 'manually split into Train & Test Data' - method that is mentioned in Section 4.6.2. Two types of gaps will be used.

Interpolation As mentioned in Section 1.5, a goal of this thesis is to ensure the feasibility of design solutions in the entire design space. The interpolation gap experiment focuses on gaps within the boundaries of the design space. Because of ship size limitations, due to bridges and locks for example, it is often found in the data that ships are concentrated near a certain limit. Just above this limit, one will not find a lot of vessels. In this experiment a gap will be created in between groups of data. A few examples are shown in Figure 5.19. A model will be trained with data located in the blue areas. Data in the red area will be used for testing. The area in between the blue and red areas is added to force that test data is outside the boundaries of the train data set.

In Figure 5.19, multiple interpolation gap areas can be seen. From left to right, it can be seen that gap area increases. It is expected that the larger the interpolation gap area, the smaller the performance of the model. This will be tested.

Extrapolation As mentioned in Section 1.4 it is expected that novel and innovative design solutions are located at the boundaries of a design space. The extrapolation experiment focuses on these boundary regions and especially on regions outside the boundaries of the design space. In this experiment a limit will be determined which represents a boundary. A model will be trained with all available data up to this limit. The model will be tested with data above this limit. This can be seen in Figure 5.20. The red areas represent the gap areas, where the available reference data will be used for testing.

In Figure 5.20, multiple extrapolation gap areas can be seen. From left to right, it can be seen that the distance from training area and the testing area is increased. It is expected that when the distance between the training area and the testing area increases, the performance of the model will decrease and it will decrease at a higher rate as well.

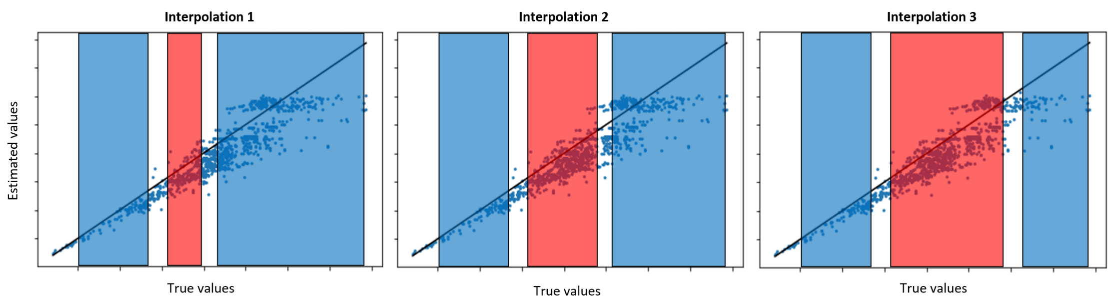


Figure 5.19: Interpolation: Increasing the gap

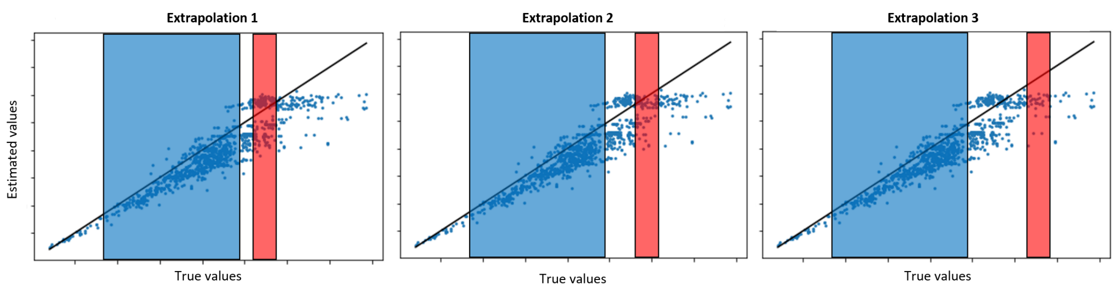


Figure 5.20: Extrapolation: Increasing the distance between train and test data

Instead of performing this experiment for multiple ship types, one ship type is chosen for this experiment. The first reason for this is that the black box and parallel hybrid models for the different ship type perform quite similar. Therefor it is expected that the results for this experiment for different ship types is similar as well. The bulk carrier will be used in this experiment. This is done because there is a lot of data available for the bulk carrier. Secondly, looking at the figures in Section 5.4, it can be seen that the models for the bulk carrier have the most smooth curves (see Figure 5.12) and hence they are considered the most stable models as well.

For this experiment it is more interesting to vary the limits that will be used to divide data into the training or test data set. In this way the inter- and extrapolation power of the models can be measured. The choice was made to divide the data based on two variables.

Length between perpendiculars This is a ship size variable. As already mentioned, the ship size is influenced by size limitations. As the lightship weight is significantly influenced by the length between perpendiculars, the choice was made to vary this ship size variable.

Deadweight The biggest component of the deadweight is the payload. The payload, i.e. the cargo, can be containers, crude oil, dredged sand or heavy lift cargo. Shipowners want to maximise the payload, but they want to minimise the dimensions of the vessel. The latter is to reduce the purchasing cost, due to less steel, and to reduce the fuel cost, due to less resistance. The deadweight will be used to divide data in the extrapolation gap test. After this test it will be clear how well the model performs when it predicts the lightship weight of ships that have an unseen and higher deadweight.

To eliminate the influence of the random distribution, as is discussed in Section 5.4, in this experiment only training and test data set sizes are used from 50 reference vessels or larger.

Results

Interpolation

First, the results will be given for the interpolation experiments. For each experiment, first, a table will be given containing the settings of that experiment and the resulting performance scores. The settings are the limitations that are used to divide the data. This leads to a distribution of the data in the sets *Train A*, *Test,* or *Train B*. *Train A* represents the part of the data set that is left from the test area, i.e. left from the red area in Figure 5.19. *Train B* represents the part of the data right from test area. Basically, reference vessels located left from the test area are smaller vessels and reference vessels located right from the test area are larger vessels. This distribution is given so that it is clear that both *Train A* and *Train B* contains a sufficient amount of reference vessels, namely at least 50.

In Tables 5.4, 5.5 and 5.6 the settings and results of the three interpolation tests are given. These correspond with Figures 5.21, 5.22 and 5.23. The performance scores in these tables are determined by calculating the R^2-score of the predicted values for the lightship weight of reference vessels located in the gap area. These are all the reference vessels in the red areas in Figure 5.19.

Settings			
Set	**Lower limit (m)**	**Upper limit (m)**	**Size data set**
Train A	0	140	114
Test	150	165	102
Train B	175	∞	687
Results			
Model	**Performance score**		
Black box model	-0.023		
Parallel hybrid model	0.168		

Table 5.4: Interpolation L_{pp} - Test 1

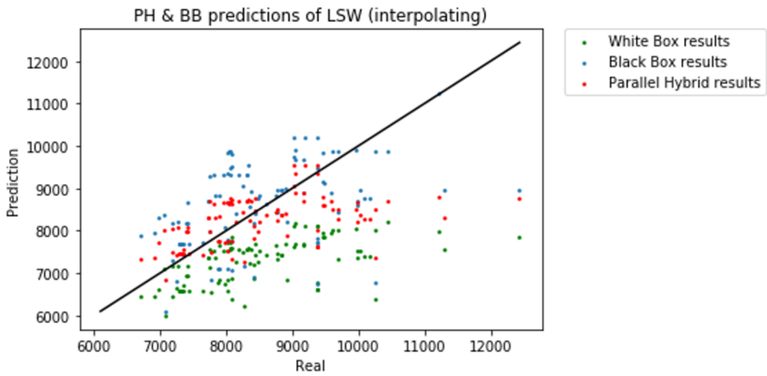


Figure 5.21: Interpolation L_{pp} - Test 1

Settings			
Set	**Lower limit (m)**	**Upper limit (m)**	**Size data set**
Train A	0	140	114
Test	150	175	350
Train B	185	∞	369
Results			
Model	**Performance score**		
Black box model	0.286		
Parallel hybrid model	0.478		

Table 5.5: Interpolation L_{pp} - Test 2

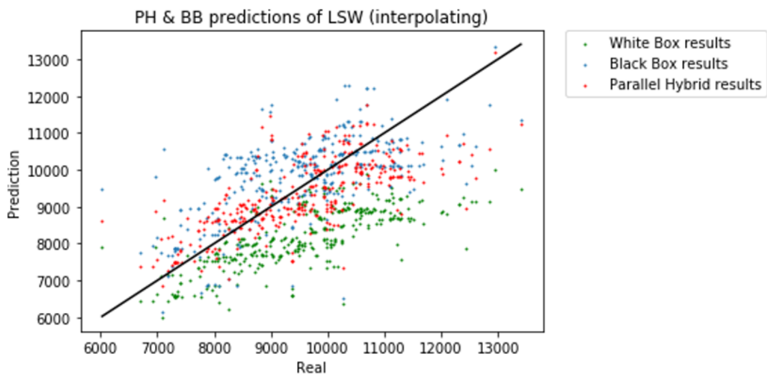


Figure 5.22: Interpolation L_{pp} - Test 2

Settings			
Set	**Lower limit (m)**	**Upper limit (m)**	**Size data set**
Train A	0	140	114
Test	150	200	808
Train B	210	∞	210
Results			
Model	**Performance score**		
Black box model	-0.169		
Parallel hybrid model	0.598		

Table 5.6: Interpolation L_{pp} - Test 3

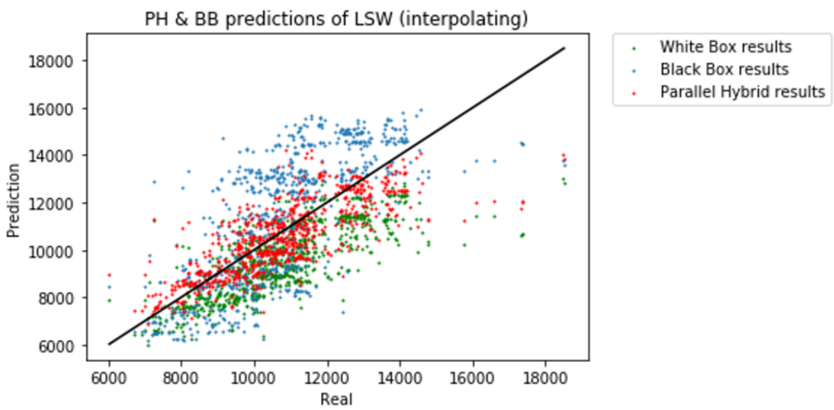


Figure 5.23: Interpolation L_{pp} - Test 3

The results of the three interpolation tests are given in Table 5.7. As can be seen, for all the tests the parallel hybrid model outperforms the black box model, as the performance scores are higher. A surprising result is that the performance score for the parallel hybrid model increases from interpolation test 1 to interpolation test 3. This is surprising because the gap is enlarged across these tests, as is shown in Figure 5.19. It was expected that the performance would decrease.

An explanation for this possibly lies in the size of data set *Train A* and *Train B*. Looking Tables 5.4, 5.5 and 5.5, it can be seen that the ratio of data set sizes Train A : Train B develops from approximately 1 : 6 in test 1, to 1 : 2 in test 3. As a result, the model is trained mainly with reference vessels in set Train B. Basically, too much weight is put on the reference vessels in Train B relative to Train A, leading to a bad score. In test 3 the distribution of data across Train A and Train B is more equal. This means that the model sees about the same amount of smaller reference vessels (Train A) during the training phase as it sees larger reference vessels (Train B), reducing the chance of overfitting. This should lead to a better prediction for both the black box model and the parallel hybrid model.

However, as can be seen in Table 5.7, this effect cannot be seen for the black box model. The performance of the black box does not improve across test 1, 2 and 3. So the previous explanation does not explain the results in this table.

A second reason for these results is that the random forest regression model is not the appropriate model for interpolation test. This will be further explained in the following extrapolation experiment results paragraph.

Model	Interpolation 1	Interpolation 2	Interpolation 3
Black box model	-0.023	0.286	-0.169
Parallel hybrid model	0.168	0.478	0.598

Table 5.7: Performance scores of 3 interpolation tests L_{pp}

Extrapolation

For the extrapolation gap tests, first, results are given for gaps that are created based on the length between perpendiculars variable. In Tables 5.8, 5.9 and 5.10, the settings and the results can be found. These correspond with Figures 5.24, 5.25 and 5.26, respectively. The performance scores in these tables are determined by calculating the R^2-score of the predicted values for the lightship weight of reference vessels located in the gap area. These are all the reference vessels in the red area's in Figure 5.20.

Settings			
Set	**Lower limit (m)**	**Upper limit (m)**	**Size data set**
Train A	0	140	114
Test	150	170	186
Results			
Model	**Performance score**		
Black box model	-3.092		
Parallel hybrid model	0.494		

Table 5.8: Extrapolation L_{pp} - Test 1

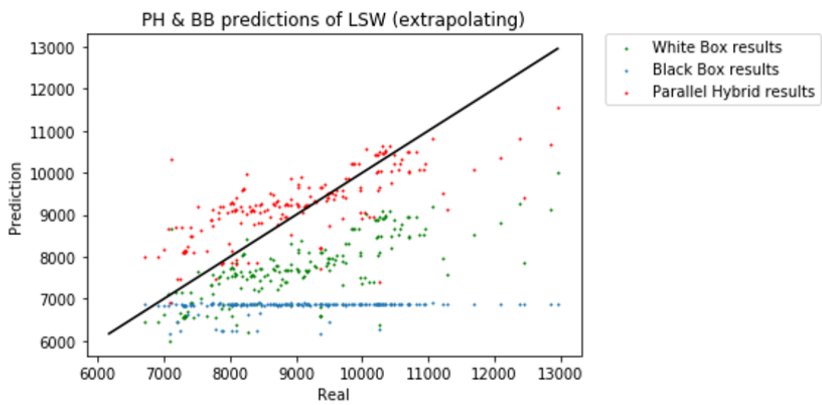


Figure 5.24: Extrapolation L_{pp} - Test 1

Settings			
Set	**Lower limit (m)**	**Upper limit (m)**	**Size data set**
Train A	0	140	114
Test	170	190	543
Results			
Model	**Performance score**		
Black box model	-7.453		
Parallel hybrid model	0.386		

Table 5.9: Extrapolation L_{pp} - Test 2

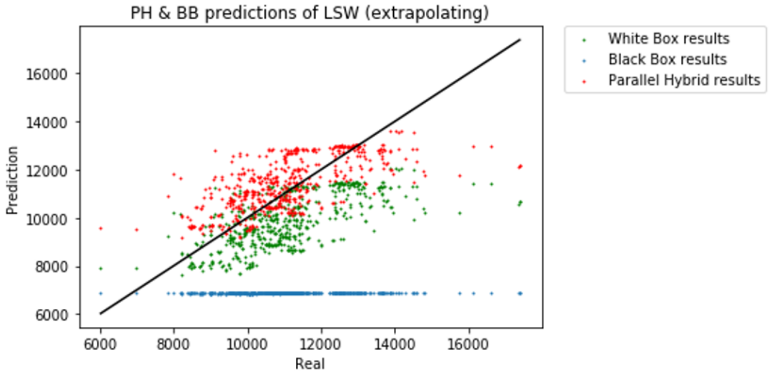


Figure 5.25: Extrapolation L_{pp} - Test 2

Settings			
Set	**Lower limit (m)**	**Upper limit (m)**	**Size data set**
Train A	0	140	114
Test	190	210	93
Results			
Model	**Performance score**		
Black box model	-8.434		
Parallel hybrid model	0.475		

Table 5.10: Extrapolation L_{pp} - Test 3

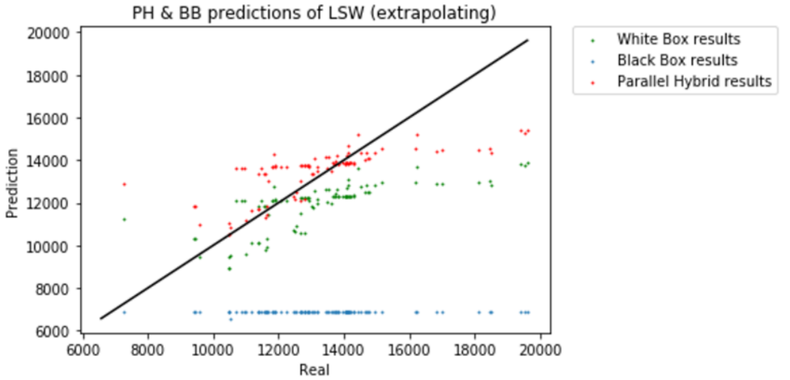


Figure 5.26: Extrapolation L_{pp} - Test 3

In Table 5.11 the results of these three experiments are summarised. It can be seen clearly that for the extrapolation test the performance of both models decreases, when the distance between the training data set and the test data set is increased. Secondly, it is noticeable that the black box model performance is significantly worse than the parallel hybrid model.

Model	**Extrapolation 1**	**Extrapolation 2**	**Extrapolation 3**
Black box model	-3.092	-7.453	-8.434
Parallel hybrid model	0.494	0.386	0.475

Table 5.11: Performance scores of 3 extrapolation tests

By comparing Figures 5.24, 5.25 and 5.26 an explanation for this is found quickly. Looking at the results for the black box model, it can be seen that the predicted values for lightship weight are constant, approximately 7000 tonnes. It can be concluded that the black box model has no extrapolating power at all. Thus, the chosen black box model, the random forest regression model, is not the appropriate model for this type of test.

This can also be seen when the results for the white box model and the parallel hybrid model are compared. The distribution of the green and red dots, which correspond to the white box model and the parallel hybrid model respectively, is similar. The only difference is a translation upwards. Thus, in the parallel hybrid model, the black box model correction which is applied to the white box model estimation, is constant. This is also a result of the absent extrapolating power of the random forest regression model.

An explanation can be found in Section 4.4. Assume that for an extrapolation test data is split into a training set and a test set, based on a length between perpendiculars of 100 meters. A random forest regression model distributes the training data in different samples and starts building the decision trees. At each node of these trees, training data is split based on one of the independent variables. The length between perpendiculars is one of the independent variables. Thus, a splitting rule at a node of a decision tree can be "*Length between perpendiculars > 99 m.*". In the testing phase for an extrapolation problem, all reference vessels meet this criteria and thus, most of these vessels will follow the same path in the decision tree, leading to the same prediction for lightship weight. This prediction is equal to the lightship weight of the largest vessel that the random forest regression model has seen in the training phase.

For interpolation this works the same. The only difference is that in interpolation, there are two splits based on independent variables. In the testing phase, the predictions that are made for the testing reference vessels are equal to the ensemble average lightship weight of training reference vessels located at either the upper limit of training set *Train A*, or the lower limit of training set *Train B*. Thus, also for interpolation problems the random forest regression model is not the appropriate model if one splits the data based on an independent variable.

Thus, it can be concluded, that a random forest regression model is not the appropriate model for interpolation and extrapolation tests if one splits the data based on an independent variable. Therefore, further discussion of the results in Table 5.11 is not necessary.

Secondly, the results are given for gaps that are created based on the deadweight variable. The settings and the resulting performance scores can be found in Tables 5.12 and 5.13. These tables correspond with Figures 5.27 and 5.28.

The results confirm the conclusion about the extrapolating power of a random forest regression model, that is drawn based on the previous extrapolation tests. Also in Figures 5.27 and 5.28, it can be seen that the distribution of green and red dots, corresponding with the white box model and the parallel hybrid model results, is similar. This means that only a constant correction is applied to the white box model estimation in the parallel hybrid model.

Although the Deadweight is not one of the independent variables, one can imagine that there is a strong relationship between the deadweight and the independent variables [*Length between perpen-*

diculars, Breadth, Depth, Block-coefficient]. Thus, the results for the extrapolation of deadweight and extrapolation of length between perpendiculars are comparable.

Comparing the performance scores in Tables 5.12 and 5.13, it can be seen that the parallel hybrid model does not outperform the black box model. This is in contrast to the previous extrapolation tests. The explanation is that the white box model does not perform well either.

Settings			
Set	**Lower limit (tonnes**	**Upper limit (tonnes)**	**Size data set**
Train A	0	30000	370
Test	35000	40000	145
Results			
Model	**Performance score**		
Black box model	-0.201		
Parallel hybrid model	-0.055		

Table 5.12: Extrapolation DWT - Test 1

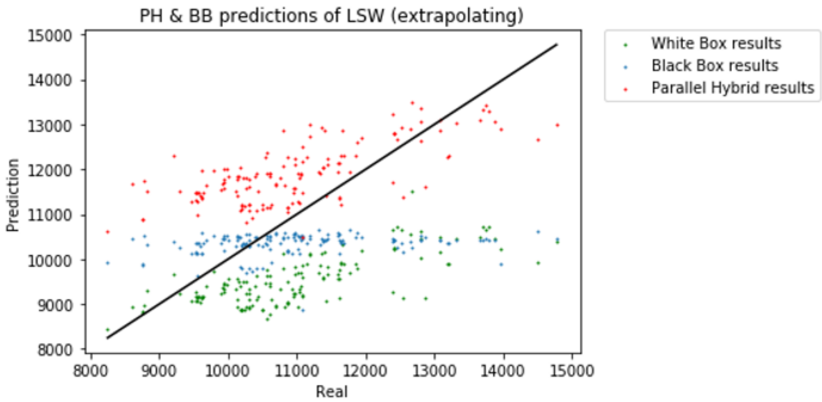


Figure 5.27: Extrapolation DWT - Test 1

Settings			
Set	**Lower limit (tonnes**	**Upper limit (tonnes)**	**Size data set**
Train A	0	30000	370
Test	40000	50000	145
Results			
Model	**Performance score**		
Black box model	-0.163		
Parallel hybrid model	-0.643		

Table 5.13: Extrapolation DWT - Test 2

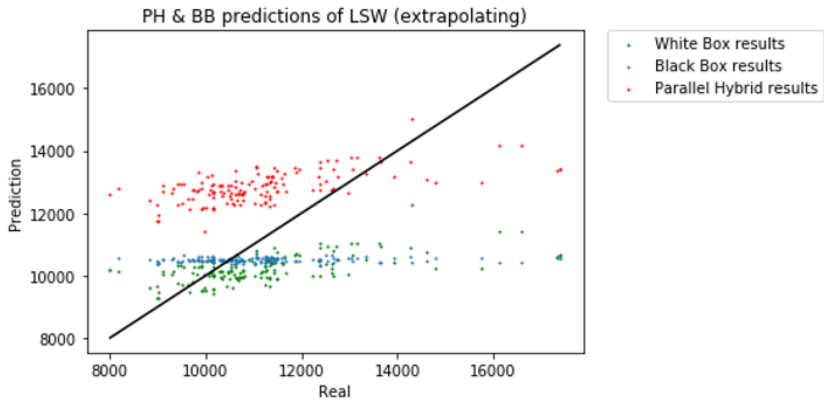


Figure 5.28: Extrapolation DWT - Test 2

Conclusion

Based on the interpolation gap test it can be concluded that the parallel hybrid model outperforms the black box model. A surprising result was that the performance of the parallel hybrid model increased when the gap was enlarged. Because of a well performing white box model, as part of the parallel hybrid model, a performance score in interpolation test 3 of approximately 0.6 seems reasonable. The performance score for interpolation test 1 is much lower than expected, namely approximately 0.17.

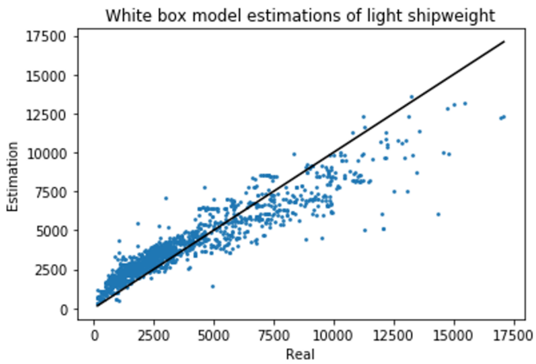
It is expected that this is a result of overfitting of the parallel hybrid model. This means that, in the parallel hybrid model, the estimation by the white box model, is over-corrected by the black box model. It is expected that this is a result of the unequal data set sizes of Train A and Train B. Further testing is required to confirm this.

Looking at the results for the extrapolation gap test there is only one conclusion to be drawn. The random forest model is not the appropriate model for extrapolation problems. This is because this model has no extrapolating power. Therefore, a well performing parallel hybrid model can only be the result of a well performing white box model.

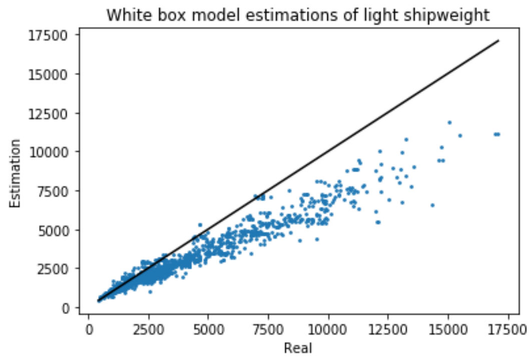
5.6. Reflection

In this chapter the performance of different models, the white box, black box and the parallel hybrid model is assessed under different circumstances. The current design approach as mentioned in Section 2.2 is mostly represented by the white box model, as this model consists of empirical formulas. The results of this chapter show that, by using an advanced machine learning tool, a higher performance can be obtained. Although this is a good result of this research, the results should also be placed into perspective.

Comparing the R^2-scores for experiment 5 and 10, in Section 5.2, it is noticeable that the scores are quite close. These are the scores for the generic method and the *D'Almeida* method. A deeper look is taken into these results. In Figures 5.29a and 5.29b, the estimated values for lightship weight are plotted against the real values, for the generic method and the *D'Almeida* method respectively. It can be seen that, although the R^2-score is higher for the generic method, the spreading of data is also higher. The relation between the estimated values and the true values is more constant for the *D'Almeida* method.



(a) Generic method result (R^2-score = 0.8322)



(b) D'Almeida method result (R^2-score = 0.8077)

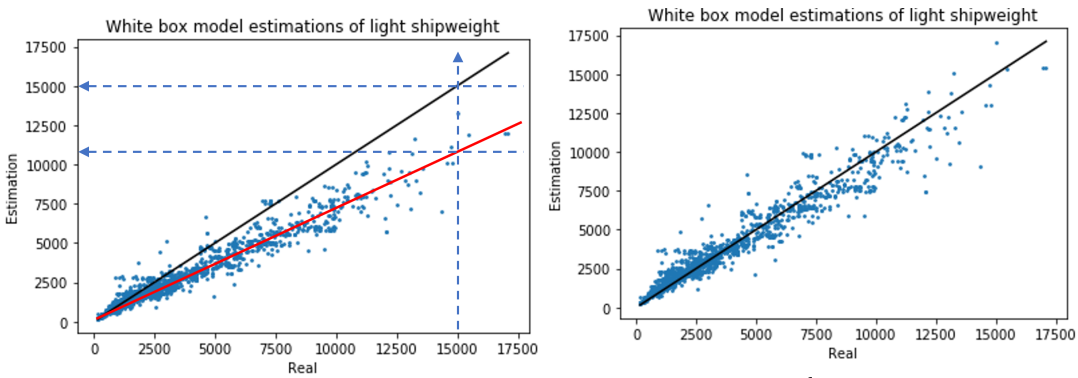
Figure 5.29: Deeper look into the white box model performance for the general cargo ship

In the 'old fashioned' design approach, a naval architect will therefore prefer the *D'Almeida* method. By manually tuning the *D'Almeida* method, a method can be obtained that has a higher accuracy. In other words, more dots near the diagonal line. As mentioned, the relation between the estimated values and the true values seems to be constant. Thus, this constant, or the correction factor, can be calculated. This is shown in Figure 5.30a. First, the red line is placed over the data. This line represents a corrected *D'Almeida* model. Assuming that the red line is the diagonal, it can be seen that for a true value of 15000 ton, the corresponding estimation is approximately 11650 ton (lower horizontal arrow). Based on the black diagonal the corresponding estimation is 15000 ton (upper horizontal arrow). The correction factor can be found with formula 5.1.

$$\text{Correction factor} = \frac{15000}{11650} = 1.28755 \tag{5.1}$$

Thus, by using the *D'Almeida* method to estimate the lightship weight and by multiplying the results with the correction factor, a better model is obtained. This model can be seen in Figure 5.30b. It can be seen that this model outperforms the generic method (Figure 5.29a) and the original *D'Almeida* method (Figure 5.29b). The corresponding R^2-scores are given in Table 5.14. It can be seen that the R^2-score for the correction *D'Almeida* method is the highest.

Comparing the score for the corrected *D'Almeida* method with the black box and parallel hybrid scores in Table 5.3, it can be seen that the performance is quite similar. Thus, by using the *D'Almeida* method and by applying a simple correction, it is also possible to obtain a high performance score for the general cargo ship.



(a) Derive a correction factor

(b)
Bakker corrected D'Almeida method result (R^2-score = 0.9328)

Figure 5.30: Correcting the D'Almeida method

Method	R^2-score
Generic method	0.8322
D'Almeida method	0.8077
Corrected D'Almeida method	0.9328

Table 5.14: R^2-scores for the generic method, D'Almeida method and the corrected D'Almeida method

5.7. Which model to choose ?

As mentioned in Section 5.1, it must be clear for a naval architect which model performs best in a certain ship design situation. Based on the previous experiments the following conclusions are drawn.

Based on experiments 1 and 2 it can be concluded that the black box and parallel hybrid model clearly outperform the white box model when there is sufficient reference data available. Thus, in that case, a naval architect should rely on the black box or parallel hybrid model.

After experiment 3 it was clear that when the training data set size reduced, the performance of the black box model reduced significantly, whereas the performance of the parallel hybrid model was relatively constant. It was concluded that for a training data set smaller than 50 reference vessels, a naval architect should rely on the parallel hybrid model. For a training data set larger that 50 reference vessels, the naval architect can choose either the black box model or the parallel hybrid model as the performances were similar.

In experiment 4, it became clear that the random forest model was not the appropriate model for interpolation and extrapolation problems. Therefore, a naval architect should rely on the white box model.

Thus, depending on the situation, the advanced machine learning methods (i.e. black box model and parallel hybrid model) which are presented in this thesis are an improvement of the current ship design approach that is used by C-Job's naval architects (i.e. the white box model), because one is able to obtain higher performance scores. Therefore, these methods should be used in practice to design future ships.

In Table 5.15 and Figure 5.31 an example of this is shown for future bulk carriers. In Table 5.16 and Figure 5.32 an example of this is shown for future container ships. In both tables it can be seen that a model has been trained with reference vessels built between 1980 and 2005. Thereafter, the model was tested with reference vessels built between 2010 and 2018. The sizes of the training and test set can be seen in the tables as well. The performance scores are given for the white box model, black box model and the parallel hybrid model. To be clear, the white box model performance score is calculated for the reference vessels in the test set only. Thus, these are slightly different than the white box model performance scores that were given in Table 5.3.

As can be seen in these table, there is a gap in build year between the training data set and the test data set. One can image that the vessels that are build in 2006 do not differ a lot of the vessels that are build in 2005. Therefore, it is assumed to be relatively easy to predict the lightship weight of vessel build in 2006, based on reference vessels build before 2006. By including a significant gap between the test and train data set, it is assumed that vessels in the test data set are also slightly different from the reference vessels in the training data set. This is because research and developments between 2005 and 2010 has been taken into account in the design of reference vessels that are built from 2010.

In the results section of these tables, it can be seen that the performance score for the black box and the parallel hybrid model are indeed higher than the performance of the white box model. Thus, naval architects should use the black box or the parallel hybrid model, instead of the white box model, to design future ships based on reference data of existing ships.

Settings			
Set	**Lower limit (build year)**	**Upper limit (build year)**	**Size data set**
Train A	1980	2005	552
Test	2010	2018	187
Results			
Model	**Performance score**		
White box model	0.709		
Black box model	0.733		
Parallel hybrid model	0.851		

Table 5.15: Extrapolation build year - Bulk carrier

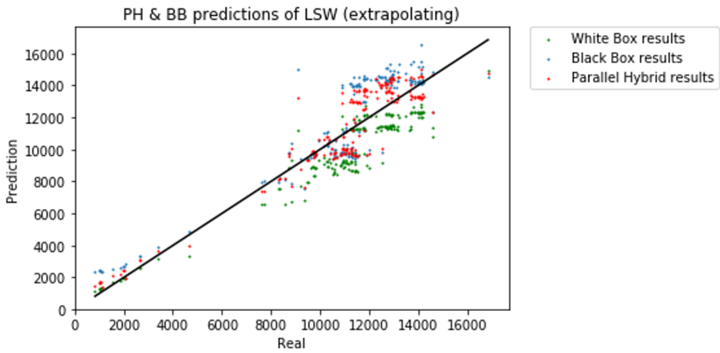


Figure 5.31: Predicting the lightship weight of future ships - Bulk carrier

Settings			
Set	**Lower limit (build year)**	**Upper limit (build year)**	**Size data set**
Train A	1980	2005	278
Test	2010	2018	31
Results			
Model	**Performance score**		
White box model	0.716		
Black box model	0.823		
Parallel hybrid model	0.838		

Table 5.16: Extrapolation build year - Container ship

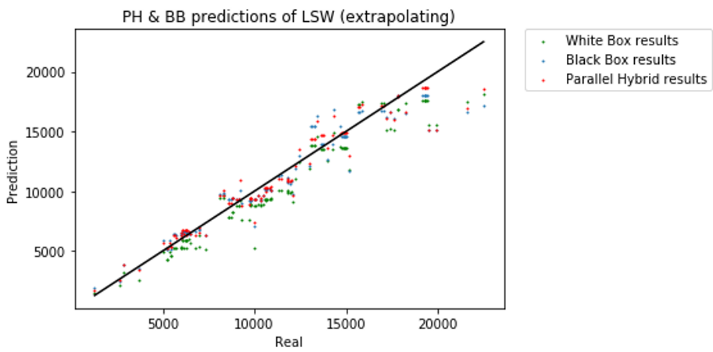


Figure 5.32: Predicting the lightship weight of future ships- Container

5.8. Validation

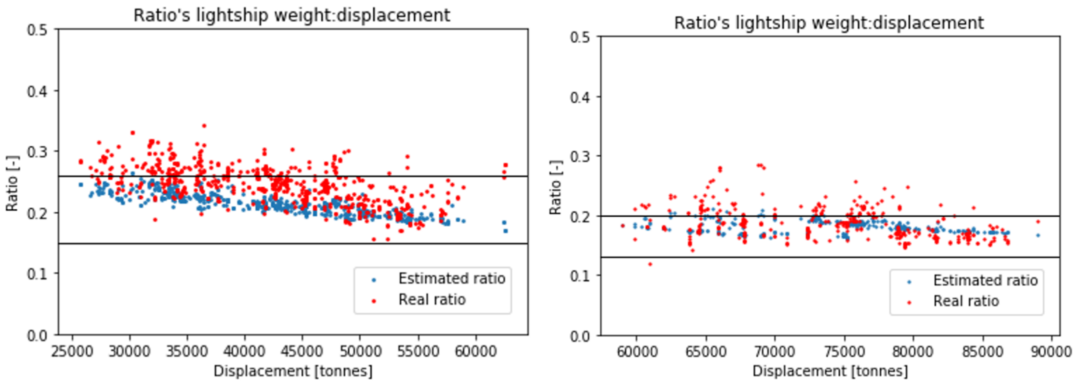
White box model

The results of the white box model can be validated using the typical range of weight groups that are given in Table 4.11. The range for the percentage lightship weight : displacement, of the four ship types are given in Table 5.17.

Ship type	Limits (tonnes)		Lightship weight/△ (%)
	Lower	Upper	
Bulk carrier A	20000	50000	15-26
Bulk carrier B	50000	200000	13-20
General cargo ship	5000	15000	20-35
Oil tanker	25000	120000	14-22
Container ship A	10000	15000	26-35
Container ship B	15000	165000	24-35

Table 5.17: Ratio lightship weight- displacement. Based on Table 4.11 from [29]

These are used to see if the estimated values of the lightship weight are within these ranges. This is also done for the true values for lightship weight. In Figures 5.33, 5.34, 5.35 and 5.36 the results can be seen for the bulk carrier, general cargo ship, oil tanker and container ship, respectively. In these figures the percentage of lightship weight : displacement is plotted against the displacement. The black horizontal lines indicate the typical range of the percentage lightship weight : displacement. The red dots represent the true values of the lightship weight of the reference vessels and the blue dots represent the estimated value for those vessels, according to the white box model.



(a) Deadweight tonnage [20000, 50000]

(b) Deadweight tonnage [50000, 200000]

Figure 5.33: Bulk carrier

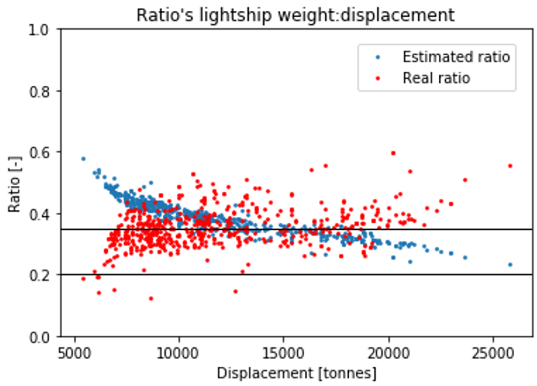


Figure 5.34: General cargo - Deadweight tonnage [5000, 15000]

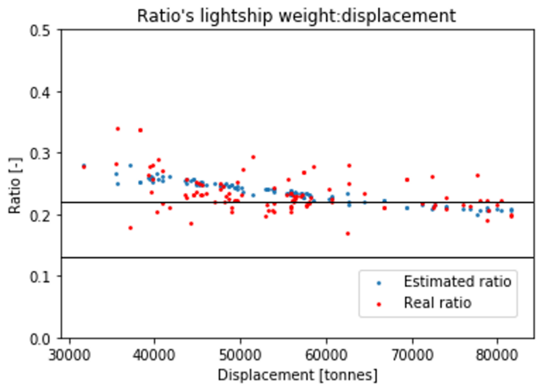
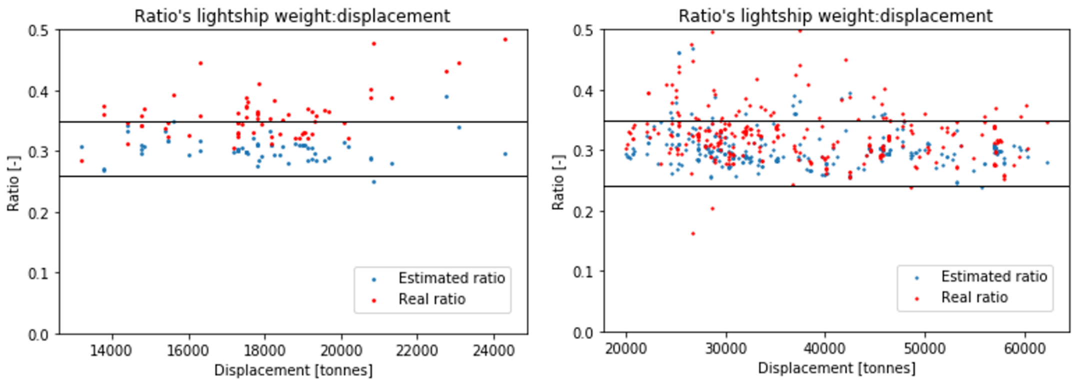


Figure 5.35: Oil tanker - Deadweight tonnage [25000, 120000]



(a) Deadweight tonnage [10000, 15000]

(b) Deadweight tonnage [15000, 165000]

Figure 5.36: Container ship

Comparing these figures, it is noticeable that for a lot of reference vessels, the actual lightship weight : displacement ratio is higher than the upper limit for this ratio, mentioned in Table 5.17. Thus, the upper limit of this ratio is not representative for the available reference data.

In Table 5.18, the results are given. In this table a portion is given for how much of the real data is in the typical range for the ratio lightship weight : displacement. This is also done for the estimated data. It can be seen that for the bulk carrier and the container ship, almost all estimated ratios lightship weight : displacement is in the typical range. Therefore, according to Table 4.11, these estimations are feasible and the estimation method is validated. For the oil tanker and the general cargo a smaller portion of the estimated values is in the typical range. As already mentioned, the upper limit does not represent the available reference data very well.

For the oil tanker and the general cargo ship the choice was made to validate the results based on the comparison between the true ratios and the estimated ratios. Thus, if the estimated ratio lightship weight : displacement are similar to the true ratios, than this estimation is considered as feasible. As can be seen, for both the oil tanker and the general cargo ship, most of the estimated ratio can be considered as feasible using this method. Only for the smaller general cargo ships the estimated ratios lightship weight : displacement are significantly higher than the true values for this ratio.

Thus, the white box model overestimates the lightship weight for smaller general cargo ships. This was already seen in Figure 5.3. Based on Figure 5.34 it is clear that the white box model for smaller general cargo ships can not be validated according to Table 4.11. Further research is required to validate these results.

Ship type	Total[1]	Validation set[2]	Real data	Estimated data
Bulk carrier A	1216	649	0.534	0.997
Bulk carrier B	1216	410	0.734	0.980
General cargo ship	1567	818	0.595	0.496
Oil tanker	132	104	0.356	0.192
Container ship A	453	65	0.477	0.969
Container ship B	453	281	0.815	0.922

[1] The total amount of reference vessels per ship type that are available in the database. These can be found in Table 4.2

[2] This is the amount of reference vessels that were within the range for deadweight tonnage given in Table 5.17

Table 5.18: Reference vessels and the portion of those vessels that are within the typical range for lightship weight, as mentioned in Table 5.17

Black box model and parallel hybrid model

The black box and the parallel hybrid model are validated using the 10-fold cross validation method. The results of this method can be found in Table 5.3. Based on these high performance scores, both the black box model and the parallel hybrid model are validated.

5.9. Conclusion

In Section 5.1 a brief explanation was given of which questions will be answered in this chapter. Four questions were answered quantitatively by performing experiments and one question was answered in a qualitative manner.

First the performance of the white box model was discussed. For the white box model, several weight estimation methods were found in literature. The performance of these method was expressed by a R^2-score. The methods with the highest score were chosen to be implemented in the white box model. The performance of these methods, and thus the performance of the white box model, gives an indication of the performance of weight estimation methods that are currently used in ship design. Primarily because of the age of several methods, the performance score was quite low. The generic weight estimation method and the *D'Almeida* method from 2009 performed best in general.

Secondly, the performance of the black box model and the parallel hybrid model was discussed. The performance of these models was determined with the 10-fold cross validation method. It was clear that both the black box and the parallel hybrid model outperformed the white box model, as the performance scores were higher. This is mainly because there was sufficient training data available for all four types of ship. The performance of the black box and parallel hybrid was similar for the amounts of training data that was used in this experiment.

In the third experiment, the size of the training data set was varied. In this experiment it was clear that the performance of the black box and parallel hybrid model differed when the size of the training data set became smaller. In general, the conclusion was drawn that for a training data set smaller than 50 reference vessels, the parallel hybrid model clearly outperformed the black box model. Where the performance of the black box model clearly decreases for smaller training data sets, it can be seen that the performance of the parallel hybrid model is much more consistent for different sizes of the training data set. For a training data set size larger than 50 reference vessels, the performance of these two models was similar.

Thereafter, the performance of the black box and the parallel hybrid model was assessed in both interpolation gaps and extrapolation gaps. In these experiments the manually split Train & Test data method was used, which is described in Section 5.3. It became clear that the Random Forest Regression model is not the appropriate model for extrapolation problems. To determine whether this is also the case for the interpolation problem, some extra research is required. In general the parallel hybrid model performed better than the black box model for interpolation problems. But, the performance score increased when the size of the gap increased. This is the opposite of what one would expect. It is expected that the distribution of data across *Train A* and *Train B* has influence on the performance scores which were presented for the interpolation test.

Section 5.6 placed the results of these four experiments into a different perspective. In this section an 'old fashioned' design approach was described. The result was that with a relatively simple correction of the *D'Almeida* method, it was also possible to obtain a higher performance score. This performance score was comparable with the performance of the black box and parallel hybrid model. One can argue that there is no need for an advanced machine learning tool if high performance scores can also be obtained by applying simple corrections to the white box models.

Taking this different perspective into account as well, the last question could be answered qualitatively. That question was "When should a naval architect rely on which model?". Based on the experiments the answer to this question can be summarised as follows:

Training data set size < 50 reference vessels:	Parallel hybrid
Training data set size > 50 reference vessels:	Black box or parallel hybrid
Interpolation or extrapolation problems:	White box

This means that the current design approach, which is represented by the white box model, is outdated and can be replaced by a more advanced method like black box modelling or parallel hybrid modelling.

On the other hand, some additional research is required in order to determine whether the parallel hybrid model and black box model also clearly outperform the white box models, which are corrected using a simple method. In Section 5.6, it is shown that this is not the case for the general cargo ship. It is expected that updating the ship type specific coefficients of the *D'Almeida* method, higher performance scores can easily be obtained.

Finally, to validate the results in this chapter two methods were used. To validate the white box model results, the predicted lightship weight : displacement ratio was compared with a typical range for this ratio, depending on the ship type. Most of these predicted ratios were in the typical range and thus, these predictions can be considered as validated. Secondly, if it was outside the typical range, the predicted ratio was compared with actual lightship weight : displacement ratios. If these values were comparable, than these predictions were also considered as validated. Almost all predicted lightship weight: displacement ratios were validated in this manner. Only the validation of the predictions for the smaller general cargo vessels should be questioned. By choosing the (corrected) *D'Almeida* method instead of the generic method, it is expected that this problem is solved.

Based on the results of this chapter, research question 5 and 6 can be answered.

Research question 5

How can these principles be converted into a model, that improves the quality of the design solutions produced by C-Job's current Maritime Intelligence Tool and using its database as a design space?

As is shown in Chapter 5, the quality of the predictions of the lightship weight has been improved, by using a parallel hybrid model. When less reference data is available this model has proven to perform better than the white box model and the black box model. As mentioned in Section 5.7, based on a training data set size of 50 reference or less, the parallel hybrid model outperforms the black box model and the white box model. Thus, in this situation the parallel hybrid model is an improvement of both the current Maritime Intelligence Tool (MIT) prediction and the current estimation methods.

With a training data set size larger than 50 reference vessels, the black box model and parallel hybrid model perform similarly. This was also expected because the black box model performs better when there is more reference data available.

For interpolation and extrapolation problems another black box model should be found in order to improve the current MIT. In this thesis, the random forest regression model is used as the black box model. Unfortunately, this model turns out to be an inappropriate model for interpolation and extrapolation problems. Thus, according to the results of Chapter 5 a white box model should be chosen for these problems.

Besides the proposed parallel hybrid model, another method is presented in Chapter 5. This method is the method of *D'Almeida*, but now the $k1$ coefficient is corrected. It is shown that with a simple correction it is possible to obtain high performance scores without using a black box model or a parallel hybrid model. Therefore it is expected that

Research question 6

How can one determine if a design solution is feasible and optimal?

Papanikolaou (2019) [29] describes several typical ranges for certain weight ratios. These are given per type of ship. These ratios are lightship weight : displacement, but also steel weight : lightship weight, machinery weight : lightship weight and outfitting&equipment weight : lightship weight.

First of all, the feasibility of the predicted ratio lightship weight : displacement is assessed based on the typical range of this ratio, provided by *Papanikolaou*. As can be seen in Section 4.3.4, most of the predicted ratios are within the typical range. However, also some predicted ratios are outside this typical range.

As can be seen, this is also the case for some actual ratios of lightship weight : displacement. Therefore, a second feasibility check was done by comparing the predicted ratios with actual ratios. It can be concluded that these were quite similar. Therefor most of the predicted ratios lightship weight : displacement are considered feasible.

Only for the smaller general cargo vessels the predicted ratios had different values compared to the actual ratios lightship weight : displacement. Thus, the feasibility of these results should be question. To solve this problem, a different white box model should be chosen, namely the corrected *D'Almeida* method. This is discussed in Section 5.6.

Unfortunately, in this thesis only the lightship weight design parameter has been discussed. The optimal design solution is therefore a design solution with a small value for the lightship weight, or a low lightship weight : displacement ratio. This means that the deadweight : displacement is high. As most of the deadweight consists of the payload, a high deadweight : displacement is beneficial for a ship owner.

6 Conclusion

This chapter will conclude this research. First, the research questions will be restated in Section 6.1. Thereafter, in Section 6.2 the major findings are discussed for each chapter. After that, the contribution of this research to science is discussed in Section 6.3. Section 6.4, gives the limitation of this research. Lastly, based on the research questions, the major findings, the contribution and limitations of this research, some recommendations will be done for future research in Section 6.5.

6.1. Research questions

In Chapter 1, it was discussed that advanced machine learning tools are very promising tools to improve the preliminary ship design phase. The development of C-Job's Maritime Intelligence Tool (MIT) has shown that the data of reference vessels can better be exploited in this phase of the ship design process. Therewith, naval architects can better be supported in making design choices for a novel ship design.

Although the MIT shows significant potential, its accuracy rapidly decreases when less reference data is available. This is especially unfavourable at the boundaries of a design space, as it is expected that novel and innovative ship designs are located at the boundaries. To better support the naval architect in all areas of the design space, the following main research question was determined for this thesis.

"How can a ship designer better explore gaps in a design space, generated using data of reference vessels, to make predictions of the main parameters of a novel ship design solution more accurate?

To answer this question, six sub-research questions were determined. These questions and their answers can be found in Appendix A.

6.2. Major findings

Chapter 2 - Design Approach

The current design methods at C-Job are strongly dependent on the experience of the naval architect, the type of ship, the customer requirements, time and customer budget. To decrease the human effort in the design process, the Accelerated Concept Design methodology has been developed. In this holistic design methodology all objectives, constraints and variables are dealt with simultaneously. The first step in this methodology, should be covered by the MIT, which performs a thorough reference study and makes predictions about advantageous main characteristics of a novel vessel.

The current method that is used in the MIT does not meet the required predicting accuracy, especially in regions of the design space where data is limited. Also the neural and Bayesian method as described in [10] do not meet this requirement.

A second disadvantage is that it is difficult to accept a design solution if it is produced by a black box model, which is per definition hard for a human being to understand. To improve this, a naval architect should be able to see that a design solution meets certain ship design rules and regulations. If this is the case, a design can be considered as feasible. Literature describes some feasibility requirements. The method proposed by *Duchateau (2016)* [14] solves this problem, but the need for human interaction is a disadvantage.

The naval architect should search for incremental innovative design solutions only, using the MIT. This is because the proposed design solutions are always a derivation of previous work. For radical innovative designs a different approach is required. The second remark is that the results of the MIT should be handled with care. A naval architect should understand the requirements and the (different) purpose(s) of the vessel in order to choose which design is desirable.

To conclude, the method that is currently used in the MIT is a fast reference-study method with significant potential to improve the design process. To use this method in practice the gap between the data-based MIT method and the traditional knowledge-based method needs to be addressed.

Chapter 3 - Method Exploration

Based on Chapter 2, a list of requirements is determined for the new tool. In these requirements, the current design approaches have been taken into account, as well as the challenges to lie ahead for the future design approach. These requirements are the following

1. Ability to deal with data-sparse and data-abundant regions of a design space
2. Results should comply with the laws of physics and other governing (basic) ship design rules
3. The new method should provide insight
4. The new method should be a fast method
5. Ability to deal with feedback

Based on these requirements some potential solutions have been explored and evaluated. It shows that parallel hybrid modelling is the most promising solution for the problems defined in this thesis. The advantage of this type of modelling is that the available knowledge can be included in a white box model. Using this white box model an estimation can be made about the magnitude of certain parameters. Based on data of reference vessels a machine learning model can be trained to learn about the differences between this estimation and the 'true' values (i.e. the data of reference vessels). This knowledge can then be used to correct the first-principle white box model.

An advantage of this method is that the white box model is constructed by a naval architect. In this model the naval architect can define all the relevant equations and rules. This provides the insight in the design process that is required. Next to that, by including these equations and rules, the naval architect can ensure the feasibility of the proposed design solution.

By requiring that the new method must be a fast method, the naval architect is able to run the optimisation problem, explore the results, redefine and re-run the optimisation problem. Using this method, the first step in the current design method, which is finding a starting point, is done by the MIT. Therefore it is possible to explore a significant amount of feasible design solutions in a reasonable amount of time, which increases the possibility of finding better design solutions compared to the current method.

Chapter 4 - Methodology

In Chapter 4, an overview is given of how the proposed parallel hybrid model is constructed. All the different sub-models have been discussed.

First of all, the white box model was discussed. Different methods were explored which are able to predict the lightship weight of a ship, or a certain weight component of the lightship weight. Methods were found for different ship types. Based on the applicability of these methods and the availability of reference data, it was chosen to test four types of ships. These are the bulk carrier, the general cargo ship, the oil tanker and the container ship. The different weight estimation methods for these ship types will be assessed with the R^2-method. This method expresses the performance of the weight estimation methods as a relative value. Based on this performance, one weight estimation method is selected for each ship type. These methods are implemented in the white box model. The performance of this model describes the performance of methods that were already available in literature. This is important when the performance of the black box and parallel hybrid model are assessed as well. Based on this one can tell if the ship design process has thus been improved by this research.

Secondly the black box model was described in detail. As a black box model the Random Forest Regression Model is used. In this method 100 uncorrelated decision trees are randomly constructed. The individual predictions of each tree are ensemble averaged in order to derive the prediction of the entire random forest model. This prediction out-performs any of the predictions of the individual decision trees.

The white box can now be used to estimate the lightship weight for four ship types and the black box model can be trained to predict the correction that should be applied, based on the data. This approach is the parallel hybrid modelling approach.

The performance of both the black box model and the parallel hybrid model are assessed with the 10-fold cross validation. This gives a good and robust indication of the models' performances. The Test & Train split method will be used to simulate certain situations in the ship design process. For example, how well a model can predict the lightship weight in a region of a design space where data is lacking.

Chapter 5 - Experiments & Results

In Chapter 5, a brief explanation was given of which questions will be answered in this chapter. Four questions were answered quantitatively by performing experiments and one question was answered in a qualitative manner.

First, the performance of the white box model was discussed. For the white box model, several weight estimation methods were found in literature. The performance of these method was expressed by a R^2-score. The methods with the highest score were chosen to be implemented in the white box model. The performance of these methods, and thus the performance of the white box model, gives an indication of the performance of weight estimation methods that are currently used in ship design. Primarily because of the age of several methods, the performance score was quite low. The generic weight estimation method and the *D'Almeida* method from 2009 performed best in general.

Secondly, the performance of the black box model and the parallel hybrid model was discussed. The performance of these models was determined with the 10-fold cross validation method. It was clear that both the black box and the parallel hybrid model outperformed the white box model, as the performance scores were higher. This is mainly because there was sufficient training data available for all four ship types. The performance of the black box and parallel hybrid was similar for the amounts of training data that was used in this experiment.

In the third experiment, the size of the training data set was varied. In this experiment it was clear that the performance of the black box and parallel hybrid model differed when the size of the training data set became smaller. In general, the conclusion was drawn that for a training data set smaller than 50 reference vessels, the parallel hybrid model clearly outperformed the black box model. Where the performance of the black box model clearly decreases for smaller training data sets, it can be seen that the performance of the parallel hybrid model is much more consistent for different sizes of the training data set. For a training data set size larger than 50 reference vessels, the performance of these two models was similar.

Thereafter, the performance of the black box and the parallel hybrid model was assessed in both interpolation gaps and extrapolation gaps. In these experiments the manually split Train & Test data method was used, which is described in Section 5.3. It became clear that the Random Forest Regression model is not the appropriate model for extrapolation problems. To determine whether this is also the case for the interpolation problem, some extra research is required. In general the parallel hybrid model performed better than the black box model for interpolation problems. But, the performance score increased when the size of the gap increased. This is the opposite of what one would expect. It is expected that the distribution of data across *Train A* and *Train B* has influence on the performance scores which were presented for the interpolation test.

Section 5.6 placed the results of these four experiments into a different perspective. In this section an 'old fashioned' design approach was described. The result was that with a relatively simple correction of the *D'Almeida* method, it was also possible to obtain a higher performance score. This performance score was comparable with the performance of the black box and parallel hybrid model. One can argue that there is no need for an advanced machine learning tool if high performance scores can also be obtained by applying simple corrections to the white box models.

Taking this different perspective into account as well, the last question could be answered qualitatively. That question was "When should a naval architect rely on which model?". Based on the experiments the answer to this question can be summarised as follows:

Training data set size < 50 reference vessels:	Parallel hybrid
Training data set size > 50 reference vessels:	Black box or parallel hybrid
Interpolation or extrapolation problems:	White box

This means that the current design approach, which is represented by the white box model, is outdated and can be replaced by more a more advanced method like black box modelling or parallel hybrid modelling.

On the other hand, some additional research is required in order to determine if the parallel hybrid model and black box model also clearly outperform the white box models, which are corrected using a simple method. In Section 5.6, it is shown that this is not the case for the general cargo ship. It is expected that by updating the ship type specific coefficients of the *D'Almeida* method, higher performance scores can easily be obtained.

Lastly, to validate the results in this chapter two methods were used. To validate the white box model results, the predicted lightship weight : displacement ratio was compared with a typical range for this ratio, depending on the ship type. Most of these predicted ratios were in the typical range and thus, these predictions can be considered as validated. Secondly, if it was outside the typical range, the predicted ratio was compared with actual lightship weight : displacement ratio's. If these values were comparable, than these predictions were also considered as validated. Almost all predicted lightship weight: displacement ratio's were validated in this manner. Only the validation of the predictions for the smaller general cargo vessels should be question. By choosing the (corrected) *D'Almeida* method instead of the generic method, it is expected that this problem is solved.

Conclusion on main research question

Based on the main findings and conclusions of each part of this research, the main research question can be answered.

A ship designer can better explore the design space by using a parallel hybrid model, instead of a white box model or a black box model.

But, in interpolation or extrapolation gap areas, a naval architect should use a white box model, mainly because the random forest regression model is not the appropriate model for these problems.

Thereafter, the white box models are outperformed, primarily because of the age of the formulas that are used.

6.3. Contribution of this research

In this research, advanced machine learning tools have been used to improve the ship design process in the preliminary ship design phase. This means that this *ship design* thesis has a lot of common ground with *data science*. Throughout this research, decisions were made based on these two perspectives; the ship design perspective and the data science perspective.

By taking both perspectives into account at each decision, it became clear that design methods that are common in ship design, are outdated from a data science point of perspective. One of these design methods is to find reference vessels and to determine a regression model (i.e. trend-line) between parameters. This is for example the trend-line between length and the lightship weight. Based on reference data a trend-line is easily found in Excel using the TREND-function. This function also presents a R^2-score for the trend-line that is determined, to indicate the performance of this regression model.

Based on a data science perspective, some improvements for this method are found in this thesis. In the above mentioned example, a regression model is trained with reference data. The performance score that is presented, indicates how well the model fits the training data. It tells a naval architect how well the lightship weight of the *already seen* reference vessels can be predicted, but it doesn't say anything about the predicting capability of the model for *unseen* vessels. Thus, by assessing a model's performance based on unseen vessels, a much more realistic value of the performance can be derived. Therefore, 10-fold cross validation is a very useful performance measurement tool.

With 10-fold cross validation and the R^2-score, three models are compared in this thesis, which represent three different design approaches, namely:

1. A design approach based on knowledge (White box model)
2. A design approach based on statistics (Black box model)
3. A design approach based on statistics and knowledge (Parallel hybrid model)

It is concluded that a design approach only based on knowledge is outdated in most situations. With more advanced machine learning techniques (i.e. the black box model and the parallel hybrid model), the reference-based design approach in preliminary ship design is taken to a higher level.

Although it is still hard to derive what exactly happens inside a machine learning model, with sufficient testing it is possible to validate this model. This is done by 10-fold cross validation and the train test split method. By averaging the performance scores of the model for different distributions training and test data, it is possible to approximate the actual performance of this model very well.

Lastly, this thesis describes the conditions that a naval architect should use in order to choose the appropriate model to solve a certain ship design problem.

6.4. Limitations

This research also has some limitations that should be taken into account. First, the limitations with regards to the performance of the models are discussed.

In Section 5.3, it was shown that the white box models are outperformed by the black box model and the parallel hybrid model. As is shown in Table 5.1, it is clear that the white box model is primarily outperformed, because its estimation methods date from 1970 approximately. Vessels that are built before 1970 are significantly smaller than today's vessels. A reason that these methods do not perform very well is that the estimation methods are empirically derived from reference vessels of that time. Thus, it should be questioned whether it is the difference in type of modelling, or the difference in the age of the method, that leads to a better performance. Although the white box model is based on knowledge, its estimation methods are in fact empirically derived from reference vessels older than approximately 1970. Therefore, it is expected that the age of these estimation methods limits the strength of the conclusion in Section 5.3, that the black box and parallel hybrid model outperform a white box model.

Thereafter, the conclusion on the interpolation and extrapolation gap experiment in Section 5.5, is limited by the fact that the random forest regression model is not the appropriate model for interpolation and extrapolation problems. This doesn't mean that any machine learning model is inappropriate for these problems. It is expected that other machine learning models have a better extrapolation power than the random forest regression model.

Currently, the proposed model is also limited based on the type of ship. For only four types of ships there was sufficient reference data available, as well as estimation methods. These were the bulk carrier, the general cargo ship, oil tanker and container ship. On the other hand, between the different ship types,

not a significant difference in performance was found. Therefore it is expected that also the lightship weight for other types of ship can be predicted. Section 6.5 describes how this can be done.
Secondly, there are some limitations of the proposed parallel hybrid model in general. In this thesis, only the lightship weight design parameter is discussed. The reason for this is the availability of estimation methods of a design parameter and the availability of the true values of the same design parameter. Estimation methods are available for more design parameters, such as the resistance, stability, motions and cost. Unfortunately the actual values of these design parameters are not publicly shared for competitive reasons. Therefore, the proposed parallel hybrid model is mainly interesting for companies, which have sufficient data available of the true values of these design parameters.

Secondly, there is a difference in how much data is required to obtain a good performing black box model or parallel hybrid model and how much data is required to validate these models. Although the parallel hybrid model performs relatively well for all sizes of the training data set, this doesn't mean that it can also be used when little data is available. Validation of these advanced machine learning models can only be done when there is sufficient data available. As it is difficult for humans to understand what actually happens inside these machine learning models, extensive testing the quality of predictions is required in order to validate such a model.

6.5. Recommendations for future research

Based on the Section 6.2, 6.3 and 6.4 some recommendations are determined for future research.

First, the white box model should be updated. In Section 5.6, a method is described about how to update the estimation methods that are currently used in the white box model. It is expected that with the updated estimation methods, it is possible to obtain high performance scores with the white box model. In Section 5.6 a method is presented that only updates the coefficient $k1$. Using a polynomial fit function it is possible to find all the updated coefficients $k1, k2, k3$ and $k4$, based on the reference data. The following steps need to be undertaken in order the confirm the conclusions of this research:

1. Update ship type specific coefficients of the *D'Almeida* method; k1, k2, k3 and k4
2. Determine R^2-scores of the updated *D'Almeida* method
3. Plot predictions against true values to get A better understanding of the actual performance
4. Compare the update white box model results with the results of the black box model and parallel hybrid model for different experiments
5. Determine which model should be used in which situation based on these results

An advantage of this method is that, as the *D'Almeida* method uses a generic formula, it can be used for all types of ships. Only the coefficients have to be derived.

A second recommendation is that in order to determine the performance of a model, one should use more performance metrics than only one, namely; the R^2-score. As mentioned in Section 5.2, the generic method performed better than the *D'Almeida* method, for a general cargo ship. However, by plotting the estimated values of the lightship weight against the true values, it became clear that the generic method was not the appropriate model for this type of ship. Thus, the scatter plot estimated values - true values provides a different insight in the performance of a model, which should be taken into account.

Thereafter, it is recommended to use a different machine learning model, as a black box model, for interpolation and extrapolation problems. First, further research is required on the extrapolation power of other machine learning models, in order to recommend a certain machine learning model.

A
Answers to research questions

Research question 1

How does a
naval architect make well-advised design decisions in an early stage of C-Job's current design process?

As mentioned by *Duchateau (2016)*, the goal of preliminary ship design is to find a balance between customer ambition (needs), available budget and possible design solutions. In order to do this the designer needs an understanding of the requirements and their effect on a design solution. As requirements are often conflicting, it is this understanding which aids the designer in making appropriate trade-offs. Based on the requirements and the personal experience of the naval architect a design strategy is chosen that best fits a certain design project. Four methods have been described that are used to determine a starting point of the design process:

1. Perform reference study and determine trend lines
2. Start with requirements and perform own calculations
3. Study literature and select relevant equations
4. Adapt a convenient reference vessel

These different method all represent a different level of how wide and thorough the initial search for a design solution is. As is concluded after some interviews, most of the naval architects at C-Job start their design process with an already focused initial search, which is directed towards the most promising area's of a design space. By performing own calculations the naval architects gain insight into the design process and the design itself. Literature describes a lot of empirical and parametric calculation methods that can be used in this stage of the design process. Next to gaining insight, based on these calculations and by comparing the results with reference vessels, the naval architects can determine if a design solution is feasible or not. In literature some requirements have been described to determine the feasibility of a design solution. To conclude, every naval architect determines his own design strategy. This will lead to different design processes and different designs, depending on who is designing. Also the limited time and budget contribute to this.

Research question 2

How can a naval architect make well-advised design decisions in C-Job's future reference based design approach?

Based on chapter 2, a list of requirements is determined for the new tool. In this requirements, the current design approaches have been taken into account as well as the challenges to lie ahead for the future design approach. These requirements are the following

1. Ability to deal with data-sparse and data-abundant regions of a design space
2. Results should comply with the laws of physics and other governing (basic) ship design rules
3. The new method should provide insight
4. The new method should be a fast method
5. Ability to deal with feedback

Based on these requirements some potential solutions have been explored and evaluated. It shows that parallel hybrid modelling is the most promising solution for the problems defined in this thesis. The advantage of this type of modelling is the available knowledge can be included in a white box model. Using this white box model an estimation can be done about the magnitude of certain parameters. Based on data of reference vessels a machine learning model can be trained to learn about the differences between this estimation and the 'true' values (i.e. the data of reference vessels). This knowledge can then be used to correct the first-principle white box model.

An advantage of this method is that the white box model is constructed by a naval architect. In this model the naval architect can define all the relevant equations and rules. This provides the insight in the design process that is required. Next to that, by including these equations and rules, the naval architect can ensure the feasibility of the proposed design solution.

By requiring that the new method must be a fast method, the naval architect is able to run the optimisation problem, explore the results, redefine and re-run the optimisation problem. Using this method the first step in the current design method, which is finding a starting point, is done by the MIT. Therefore it is possible to explore a significant amount of feasible design solutions in a reasonable amount of time, which increases the possibility of finding better design solutions compared to the current method.

Research question 3

What are the important gaps in the design space?

In the thesis, the important gaps in a design space are the interpolation and the extrapolation gaps in a design space. Interpolation gaps are often determined by limitations for ship size, due to locks, canals and bridges for example. As a result, one will find a lot of reference vessels with PANAMAX dimensions, and less reference vessels with a size a little larger than PANAMAX. However, for Neo-PANAMAX sized vessels, which are larger than PANAMAX, another concentration of reference data can be found again. In between PANAMAX and Neo-PANAMAX there is an interpolation gap. The proposed model must be able to make prediction in this gap, based on the PANAMAX and Neo-PANAMAX sized reference vessels for example.

As mentioned in Section 1.4 it is expected that novel and innovative design solutions are located at the boundaries of a design space. Ship designs located in an extrapolation gap are expected to be larger, or have a higher cargo capacity than other reference vessels in the data base. Therefor these areas of th design space might be interesting for ship owners.

Other gaps in a design space can for example be in the breadth of different vessels. A reason for this can be that there a mono hulls and catamarans in the database. These gaps will not be considered. The first reason is that in this thesis the goal is to search for incremental innovative design solutions. The difference between a mono hull and a catamaran is part of radical innovation, as these designs are totally different. A second reason is that the first step in the maritime intelligence tool is the selection of reference data by the naval architect. By selecting either mono hull vessels or catamarans, a gap in between these types of ships will not be present in the design space, generated by reference data.

Research question 4

What are leading principles in designing a model that deals with these gaps??

This research question has already been answered by the research question 2. Leading principles in designing a new model is that the tool requirements are satisfied. These requirements are

1. Ability to deal with data-sparse and data-abundant regions of a design space
2. Results should comply with the laws of physics and other governing (basic) ship design rules
3. The new method should provide insight
4. The new method should be a fast method
5. Ability to deal with feedback

Based on these requirements the parallel hybrid model has been selected as the most promising solution. There are two requirements for parallel hybrid modelling, namely:

1. Estimation methods for a design parameter
2. True values of the same design parameter

These requirements reflect the principle of parallel hybrid modelling. Using the estimation methods the value of a design parameters is estimated. This works for every region of the design space where the estimation method is applicable. Thus, there is no need for reference data. The true values of a design parameter are used to correct for any error that is made by an estimation method. For this, sufficient reference data is required in order for a black box model to perform well. To be clear, this black box model is a sub-model of the parallel hybrid model.

Unfortunately, only the lightship weight design parameter satisfied these requirements. Therefore, only the prediction of the lightship weight will be assessed in this thesis, in both data-sparse and data-abundant regions of a design space.

Research question 5

How can these principles be converted into a model, that improves the quality of the design solutions produced by C-Job's current Maritime Intelligence Tool and using its database as a design space?

As is shown in Chapter 5, the quality of the predictions of the lightship weight has been improved, by using a parallel hybrid model. When less reference data is available this model has proven to perform better than the white box model and the black box model. As mentioned in Section 5.7, based on a training data set size of 50 reference or less, the parallel hybrid model outperforms the black box model and the white box model. Thus, in this situation the parallel hybrid model is an improvement of both the current Maritime Intelligence Tool (MIT) prediction and the current estimation methods.

With a training data set size larger than 50 reference vessels, the black box model and parallel hybrid model perform similarly. This was also expected because the black box model performs better when there is more reference data available.

For interpolation and extrapolation problems another black box model should be found in order to improve the current MIT. In this thesis, the random forest regression model is used as the black box model. Unfortunately, this model turns out to be an inappropriate model for interpolation and extrapolation problems. Thus, according to the results of Chapter 5 a white box model should be chosen for these problems.

Besides the proposed parallel hybrid model, another method is presented in Chapter 5. This method is the method of D'Almeida, but now the $k1$ coefficient is corrected. It is shown that with a simple correction it is possible to obtain high performance scores without using a black box model or a parallel hybrid model. Therefore it is expected that

---

Research question 6

How can one determine if a design solution is feasible and optimal?

Papanikolaou (2019) [29] describes several typical ranges for certain weight ratios. These are given per type of ship. These ratios are lightship weight : displacement, but also steel weight : lightship weight, machinery weight : lightship weight and outfitting&equip weight : lightship weight.

First of all, the feasibility of the predicted ratio lightship weight : displacement is assessed based on the typical range of this ratio, provided by *Papanikolaou.* As can be seen in Section 4.3.4, most of the predicted ratios are within the typical range. However, also some predicted ratios are outside this typical range.

As can be seen, this is also the case for some actual ratios of lightship weight : displacement. Therefore, a second feasibility check was done by comparing the predicted ratios with actual ratios. It can be concluded that these were quite similar. Therefor most of the predicted ratios lightship weight : displacement are considered feasible.

Only for the smaller general cargo vessels the predicted ratios had different values compared to the actual ratios lightship weight : displacement. Thus, the feasibility of these results should be question. To solve this problem, a different white box model should be chosen, namely the corrected D'Almeida method. This is discussed in Section 5.6.

Unfortunately, in this thesis only the lightship weight design parameter has been discussed. The optimal design solution is therefore a design solution with a small value for the lightship weight, or a low lightship weight : displacement ratio. This means that the deadweight : displacement is high. As most of the deadweight consists of the payload, a high deadweight : displacement is beneficial for a ship owner.

Bibliography

[1] "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II". In: *IEEE Transactions on Evolutionary Computation* 6.2 (2002), pp. 182–197. ISSN: 1089778X.

[2] Gonzalo Acuña et al. "Comparison of methods for training grey-box neural network models". In: *Computers & Chemical Engineering* 23 (1999). European Symposium on Computer Aided Process Engineering, S561–S564. ISSN: 0098-1354.

[3] D.J. Andrews. "A Comprehensive Methodology for the Design of Ships (and other Complex Systems)". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 454.1968 (1998), pp. 187–211. ISSN: 13645021.

[4] D.J. Andrews. "A Creative Approach To Ship Architecture". In: *The International Journal of Maritime Engineering* 145 (Jan. 2003).

[5] C.C. Pantelides B. Duarte P.M. Saraiva. "Combined Mechanistic and Empirical Modelling". In: *International Journal of Chemical Reactor Engineering* 2.April (2004). ISSN: 15426580.

[6] W.E. Jacobsen B. Sohlberg. "Grey Box Modelling - Branches and Experiences". In: *Proceedings of the 17th World Congress The International Federation of Automatic Control Seoul, Korea, July 6-11, 2008* (2008), pp. 11415–11420.

[7] L. Breiman. "Random forests". In: *Random Forests* (2001), pp. 1–122.

[8] L. Breiman et al. *Classification and Regression Trees.* Monterey, CA: Wadsworth and Brooks, 1984.

[9] Betts C.V. "'Developments in Ship Design from a Naval Perspective". In: *Seventh international marine design conference, Kyonoiu, Korea, May 21-24, 2000* (2000).

[10] H. B. Clausen et al. "Bayesian and Neural Networks for Preliminary Ship Design". In: *Marine Technology and SNAME News* 38.4 (2001), pp. 268–277. ISSN: 19453582.

[11] D. DeLaurentis D. Mavris. "Methodology for Examining the Simultaneous Impact of Requirements, Vehicle Characteristics, and Technologies on Military Aircraft Design". In: *Icas* (2000), pp. 1–10. ISSN: 1096-1186.

[12] J. d'Almeida. "Arquitectura Naval – O Dimensionamento do Navio". In: (2009). Ed. by Prime Books.

[13] T.W. DeNucci. "Capturing Design". PhD Thesis. TU Delft, 2012. ISBN: 9789065622945.

[14] E.A.E. Duchateau. "Interactive Evolutionary Concept Exploration in Preliminary Ship Design." PhD thesis. TU Delft, 2016, p. 289. ISBN: 9789055841745.

[15] S. Ove Erikstad. "System Based Design of Offshore Support Vessels". In: January 2012 (2012).

[16] S. Estrada-Flores et al. "Development and Validation of "Grey-Box" Models for Refrigeration Applications: A Review of Key Concepts". In: *International Journal of Refrigeration* 29.6 (2006), pp. 931–946. ISSN: 01407007.

[17] J.H. Evans. "Basic Design Concepts". In: *Journal of the American Society for Naval Engineers* 71.4 (1959), pp. 671–678.

[18] V. Bertram H. Schneekluth. "Ship Design for Efficiency and Economy". In: 1993 (1998), pp. 112–148.

[19] P.A. Minderman & T.J. McAvoy Hong-Te Su N. Bhat. "Intergrating Neural networks with First Principles Models for Dynamic Modelling". In: IFAC Symposia Series (1993). Ed. by J.G. Balchen, pp. 327–332.

[20] *IBM - AI, Machine learning and Deep learning*. https://www.ibm.com/blogs/systems/ai-machine-learning-and-deep-learning-whats-the-difference/. Accessed: 2020-09-08.

[21] "Isolation forest". In: *Proceedings - IEEE International Conference on Data Mining, ICDM* (2008), pp. 413–422. ISSN: 15504786.

[22] Scikit Learn. *Random Forest Classifier*. 2020. URL: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html.

[23] Leifur Þ. Leifsson et al. "Grey-box Modeling of an Ocean Vessel for Operational Optimization". In: *Simul. Model. Pract. Theory* 16 (2008), pp. 923–932.

[24] K. Johnson M. Kuhn. *Applied Predictive Modeling*. Springer New York, 2013. ISBN: 9781461468493.

[25] M.A. Kramer M.L. Thompson. "Modeling Chemical Processes Using Prior Knowledge and Neural Networks". In: *AIChE Journal* 40.8 (1994), pp. 1328–1340.

[26] IHS Markit. *Sea-web™ Ships*. 2020. URL: https://ihsmarkit.com/products/sea-web-vessel-search.html.

[27] T.P. McDonald. "A Library Based Approach for Exploring Style in Preliminary Ship Design". In: *Department of Mechanical Engineering* December (2010), p. 375.

[28] "Chapter 9 - Ship Design, Construction and Operation". In: *The Maritime Engineering Reference Book*. Ed. by A.F. Molland. Oxford: Butterworth-Heinemann, 2008, pp. 636–727. ISBN: 978-0-7506-8987-8.

[29] A. Papanikolaou. *Ship Design*. Springer, 2019, pp. 127–127. ISBN: 9789401787505.

[30] M.G. Parsons. "Applications of Optimization in Early Stage Ship Design". In: *Ship Science & Technology* 3.5 (2009), pp. 9–32.

[31] B. Thomas R. de Winter J. Furustam. "Optimizing Ships Using the Holistic Accelerated Concept Design Methodology". In: 2019 (), pp. 1–13.

[32] R. Calantone R. Garcia. "A Critical Look at Technological Innovation Typology and Innovativeness Terminology: a Literature Review". In: *Journal of Product Innovation Management* 19.2 (2002), pp. 110–132.

[33] D.R. Myers S.K. Shum. "Intelligent System Applications: a Technology Licensor's Perspective". In: (1996). International Conference on Intelligent Systems in Process Engineering AIChE Symposium Series, 1996, p. 311.

[34] "Mean Absolute Error". In: *Encyclopedia of Machine Learning*. Ed. by Claude Sammut and Geoffrey I. Webb. Boston, MA: Springer US, 2010, pp. 652–652. ISBN: 978-0-387-30164-8.

[35] Jörg Schubert et al. "Hybrid modelling of yeast production processes – combination of a priori knowledge on different levels of sophistication". In: *Chemical Engineering & Technology* 17.1 (1994), pp. 10–20.

[36] J.H. Friedman T. Hastie R. Tibshirani. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. Springer, 2009, pp. 587–588. ISBN: 9780387848846.

[37] F. Smulders T. Van Bruinessen J.J. Hopman. "Towards a Different View on ship design: The Development of Ships Observed through a Social-Technological Perspective". In: *Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE* 1 (2013).

[38] Henricus J. L. Van Can et al. "Strategy for dynamic process modeling based on neural networks in macroscopic balances". In: *AIChE Journal* 42.12 (1996), pp. 3403–3418.

[39] Indian Maritime University Visakhapatnam. *Preliminary Ship Design Parameter Estimation*. 2017. URL: https://edisciplinas.usp.br/pluginfile.php/5690724/mod_resource/content/1/Preliminary%5C%20Ship%5C%20Design%5C%20Parameter%5C20Estimation.pdf.

[40] D.G.M. Watson. *Practical Ship Design*. Ed. by David G.M. Watson. Vol. 1. C. Elsevier Ltd, 1998, pp. 1–33. ISBN: 0080429998.

[41] R. de Winter et al. "Designing Ships Using Constrained Multi-objective Efficient Global Optimization". In: *Machine Learning, Optimization, and Data Science*. Ed. by Giuseppe Nicosia et al. Cham: Springer International Publishing, 2019, pp. 191–203. ISBN: 978-3-030-13709-0.

[42] G. Dreyfus Y. Oussar. "How to be a Gray box: Dynamic Semi-Physical Modeling". In: *Neural Networks* 14.9 (2001), pp. 1161–1172. ISSN: 08936080.

[43] S.F. Zalek. "Multicriterion Evolutionary Optimization of Ship Hull Forms for Propulsion and Seakeeping". English. PhD thesis. 2007, p. 194. ISBN: 978-1-109-91753-6.

[44] R.H. Zwart. "Trim Optimization for Ships in Service". Master's Thesis. Delft: TU Delft, 2020.

Personal Reflection

2020, it turned out to be a strange year. In January 2020 I contacted Thijs Müller from C-Job Naval Architect to ask if there would be room at C-Job for a graduate intern. An enthusiastic meeting followed quite quickly and the foundation was laid for this thesis. As I knew C-Job and its Research and Development team already from my time as a working student, I looked forward to the internship very much.

Then COVID arrived and everything changed. From one moment to the other the country went into lockdown, shops closed, sport clubs closed, events were cancelled and everybody had to stay at home as much as possible. As a freelancer in the event sector, this had quite some consequences for me. Long story short, it was challenging for me personal to start a graduation internship from home in this new situation and to stay focused.

I think that especially when you start with a new project, a lot of 5 minute chats, coffee conversations and talking with many different colleagues, will help you to kick start this project. I missed this in the initial phase of my project. As everyone was struggling to adapt to the COVID situation, people, including myself, forgot how important these conversations are. Not only to help you gain new insights in your project, also to help you to NOT think about your project for a moment. I made the mistake that at some moments, the only thing I did was working on my thesis. There was nothing else to do and I experienced quite some pressure to finish it as soon as possible.

During the summer, I was allowed to come to the office in Hoofddorp five days a week. From the first day at the office my productivity increased significantly. A good office lighting, an ARBO-responsible chair, air conditioning, a large desk with the correct height, meeting new colleagues, good coffee, not working in the same room where you sleep and a different environment. Yes, you suddenly realise how important these aspects all are. By the end of the summer I finished the literature review and it was rewarded with an 8. This meant a lot to me, especially with the COVID-situation in mind and how challenging the start of my project has been.

Also during the summer, I took quite a radical decision to leave my student room in Delft and move to Texel. It turned out to be the best decision of the past year. Literally and figuratively, I had more space which helped me a lot to not think about the thesis for a moment and to start the new week with a fresh head again.

In the last couple of months of the project I was able to finish quite some work. The more I started to understand my project, the easier it was to work from home. And the more I did besides working on the thesis, the more efficient my working days became.

Now that graduation is near and the project is nearly finished, I must say that I am proud of the final result. I think that this report gives a nice overview of the possibilities of solving ship design problems with data science techniques. Also, it describes the challenges that arise from both a ship design perspective and a data science perspective. Therefore, I believe that this report is very useful when steps are undertaken to further implement black box or parallel hybrid modelling in the ship design process.

To conclude, I have some tips based on my experiences for students who are about to start their thesis:

1. Take your time to read literature and don't worry too much about the end product. A good literature review will save you time when writing your final report.
2. Organise meetings with all the stakeholders. These are your supervisors of the company and TU Delft. Especially in the starting phase of your project it will help you to get everyone pointed in the same direction.

3. When your working days become less efficient, don't spend more time on your project, spend less. Turn off your computer and do something else that you enjoy.
4. Create a planning with sufficient detail. At least determine one task per day. It can be a small task, but it will help you to move forward. To be able to check off tasks feels good.