NTNU

Norwegian University of
Science and Technology

Estimation of Speed Loss due to Current, Wind and Waves

A Data-Driven Approach

Jens Christoffer Gjølme

Marine Technology
Submission date: June 2017
Supervisor: Sverre Steen, IMT
Co-supervisor: Øyvind Dalheim, IMT

Norwegian University of Science and Technology
Department of Marine Technology

NTNU Trondheim
Norwegian University of Science and Technology
Department of Marine Technology

MASTER THESIS IN MARINE TECHNOLOGY

SPRING 2017

FOR

Jens Christoffer Gjølme

Estimation of speed loss due to current, wind and waves - A data-driven approach

The amount of data collected and stored on ships during normal operation is increasing quickly, and one is looking for ways to utilize these data to improve the operation and maybe also the design of new ships. However, there are many obstacles, like data quality, that need to be dealt with in order to put the logged data to good use. Low data quality can, at least partly, be remediated by averaging a larger number of measurements than what is possible with data from dedicated measurement campaigns. The idea behind the current project is to utilize measurement data of ships that contain measurement of speed, shaft power, and environmental data to build an empirical model for the speed change (loss) or added power due to environmental conditions. In the SFI SMART Maritime participating ship owners are supplying ship in-service data to the project. These data will be a key resource for the proposed master thesis.

The objective of the master thesis is to develop an empirical prediction model for speed loss or added power due to environmental conditions (wind, waves and current), based on available in-service measurement data. The accuracy and reliability of the method should be documented, and it's degree of validity for other than the case ship should be discussed.

To reach the objective, it will be necessary to investigate the nature of speed loss and added power – what effects that are causing it, to perform data validation and error correction (as necessary), to couple the on-board measurements to hindcast estimates of wind, waves and current, and to determine which measurement variables that are appropriate to include as input variables to such an empirical model. Then, a data-driven empirical model shall be constructed, for instance (but not necessarily) using Artificial Neural Networks.

In the thesis the candidate shall present his personal contribution to the resolution of problem within the scope of the thesis work.

Theories and conclusions shall be based on mathematical derivations and/or logic reasoning identifying the various steps in the deduction.

The thesis work shall be based on the current state of knowledge in the field of study. The current state of knowledge shall be established through a thorough literature study, the results of this study shall be written into the thesis. The candidate should utilize the existing possibilities for obtaining relevant literature.

The thesis shall be organized in a rational manner to give a clear exposition of results, assessments, and conclusions. The text should be brief and to the point, with a clear language. Telegraphic language should be avoided.

NTNU Trondheim
Norwegian University of Science and Technology
Department of Marine Technology

The thesis shall contain the following elements: A text defining the scope, preface, list of contents, summary, main body of thesis, conclusions with recommendations for further work, list of symbols and acronyms, reference and (optional) appendices. All figures, tables and equations shall be numerated.

The supervisor may require that the candidate, in an early stage of the work, present a written plan for the completion of the work. The plan shall include a budget for the use of laboratory or other resources that will be charged to the department. Overruns shall be reported to the supervisor.

The original contribution of the candidate and material taken from other sources shall be clearly defined. Work from other sources shall be properly referenced using an acknowledged referencing system.

The thesis shall be submitted electronically (pdf) in DAIM:

- Signed by the candidate
- The text defining the scope (this text) (signed by the supervisor) included
- Computer code, input files, videos and other electronic appendages can be uploaded in a zip-file in DAIM. Any electronic appendages shall be listed in the main thesis.

The candidate will receive a printed copy of the thesis.

Supervisor : Professor Sverre Steen
Advisor : Øyvind Øksnes Dalheim
Start : 12.01.2017
Deadline : 11.06.2017

Trondheim, 12.01.2017

Sverre Steen
Supervisor

Preface

This thesis concludes my time as a stud.techn. at the Department of Marine Technology at the Norwegian University of Science and Technology. The work has been completed during the spring of 2017, and is the final requirement for the degree of *Master of Science*.

The work is the continuation of research performed as a summer research assistant at the Department of Marine Technology the summer of 2016, and during the project thesis completed in December 2016. The initiative behind the topic comes from Professor Sverre Steen as part of SFI Smart Maritime.

Some level of background in naval hydrodynamics and machine learning concepts is beneficial, but not necessary.

Jens Christoffer Gjølme
June 9, 2017
Trondheim

Acknowledgements

I would like to thank my supervisor Professor Sverre Steen for the opportunity and encouragement to continue the work done during the summer and fall of 2016 into this master thesis. Additionally, I would like to express my gratitude to PhD Candidate Øyvind Øksnes Dalheim for his continuous guidance as co-supervisor. An expression of gratitude is also needed for the shipping company, which will remain undisclosed, that has provided data which has been essential for the thesis. Andrew Ng's free online course on Machine Learning has also been of great value, which merits a mention. Lastly, I would like thank my fellow students from office C1.062 for creating an encouraging and ambitious atmosphere.

J.C.G

Abstract

There is an increasing demand from customers, governments and maritime authorities when it comes to emission control and energy efficiency in the maritime industry. Demands in efficiency requires a more systematic approach to the modeling of marine operations such that implementation of operational optimization is possible. A prerequisite for optimization of marine operations, is accurate calculation or reliable estimation of speed loss. In a seaway, factors such as direct wind and wave action, and indirect effects of waves linked with ship motions, may cause considerable speed loss. Over the years, substantial effort has been devoted to understanding the influence of wind and waves on a ship in sea passage, but a simple yet complete model does not exist. Coinciding with the increasing demand for efficiency, there is a trend towards digital transformation of businesses. Included in this transformation is the logging of operational data. This data, together with other data sources, such as weather hind- and forecast data presents an opportunity for a new approach for estimation and prediction of speed loss. This thesis aims to make use of the digitalization trend to develop a new method for estimation and prediction of speed loss through machine learning.

The method that has been develop solves the estimation and prediction problem by breaking the problem down into two parts. Using data from ship operations combined with weather hindcast data, a database is created. From this database, data which conforms to some calm water condition is collected in a calm water subdatabase. This calm water subdatabase is used to train a custom nonlinear regression model for the estimation of a calm water reference speed V_{cw} with trim, draft, and shaft power as inputs. V_{cw} represents an estimate of the ship speed in a calm seastate, for a specific set of draft, trim, and shaft power. V_{cw} is used, in conjunction with a measurement of actual speed through water, to calculate a speed loss quantity for the full database. A neural network regression model is then trained on the database to predict speed loss. The result is a speed loss prediction model, valid for the ship(s) that generated the data.

The custom nonlinear regression model has been formulated based on fundamental physical relationships between the input and target variables. Validation studies based on expected correlations and comparison with similar models has been performed. The quality of the speed loss estimates depends on the available calm water data. To ensure that the artificial neural network regression model generalizes well, a strategy for training has been developed. Regularization by implementation of early stopping results in available computing power being the factor that decides the number of hidden units in the hidden layer. Depending on the input variables, the neural network regression model is able to predict speed loss with with an approximate average error of between 2-5%, which corresponds to an error on knots of about 0.3-0.7 for the average calm water speed.

Sammendrag

I maritim industri er økende krav fra kunder, og statlige og maritime myndigheter hva gjelder utslippskontroll og energieffektivitet et faktum. Økte krav til effektiv drift krever en mer systematisk tilnærming til modellering av marine operasjoner slik at implementering av operativ optimalisering er mulig. En forutsetning for optimalisering av marine operasjoner, er nøyaktig beregning eller pålitelig estimering av hastighetstap. I reell sjø kan faktorer som direkte vind- og bølgepåvirkning, og indirekte påvirkning av bølger knyttet til skipsbevegelser, føre til betydelige fartstap. De senere år har mye innsats blitt nedlagt for å forstå innflytelsen av vind og bølger på et skip i seilas, men det finnes ingen enkel, fullstendig modell. Sammen med den økende etterspørselen etter effektiv drift, ser man tydelig trend mot økt digitalisering. Mer logging av operasjonelle data er en konsekvens av denne trenden. De operasjonelle dataene, sammen med historiske værdata åpner opp for en ny tilnærming til estimering og prediksjon av hastighetstap. Denne avhandlingen tar sikte på å benytte seg av økt tilgjengelighet av data for å utvikle en ny metode for estimering og prediksjon av hastighetstap gjennom maskinlæring.

Metoden som har blitt utviklet løser estimerings- og prediksjonsproblemet ved å dele problemet i to deler. Ved å bruke data fra operasjonelle data kombinert med historisk værdata, opprettes en database. Fra denne databasen hentes data som tilfredsstiller et krav som beskriver en rolig sjøtilstand. Denne avledede databasen brukes til å trene en ulineær regresjonsmodell for estimering av en referansehastighet for rolig sjø V_{cw}, med trim, dypgang og akseleffekt som inngangasvariabler. V_{cw} representerer et estimat av skipets hastighet i en rolig sjøtilstand, for et bestemt sett med dypgang, trim og akseleffekt. V_{cw} brukes sammen med måling av faktisk hastighet for å beregne mål på hastighetstapet for hele databasen. Et nevralt nettverk for regresjon blir deretter trent på de tilgjengelige dataene for å predikere hastighetstapet. Resultatet er en prediksjonsmodell for hastighetstap, gyldig for skipet eller skipene som genererte dataene.

Den ulineære regresjonsmodellen er utviklet basert på grunnleggende fysiske forhold mellom inngangs- og målvariabler. Valideringsstudier basert på forventede korrelasjoner og sammenligning med andre modeller har blitt gjennomført. Kvaliteten på estimatene avhenger av hvor mye data fra rolig sjøtilstand som er tilgjengelig. For å sikre at det nevrale nettverket er tilstrekkelig generaliserende, har det blitt utviklet en strategi for trening. At regularisering har blitt implementert ved bruk av tidlig stopping, resulterer i at tilgjengelig databehandlingskraft blir den dimensjonerende faktor for antall nevroner i det skjulte laget. Avhengig av inngangsvariablene kan den nevrale nettverksregresjonsmodellen forutsi hastighetstap med en omtrentlig gjennomsnittlig feil på mellom 2-5%, noe som tilsvarer en feil på ca. 0,3-0,7 knop for en vanlig hastighet.

Contents

List of Figures

List of Tables

1 Introduction

This thesis begins by introducing the background and motivation behind its research topic. Some aspects of previous work is then noted, before the research objectives is concretized. Lastly, the outline of the rest of the thesis is presented.

1.1 Background

In the ship building industry, the design of ships has traditionally been optimized for the lowest possible calm water resistance at design speed and draft. However, in recent years an increased focus has been established on the ship performance in an actual seaway. Deviation from the design speed and design draft occur regularly due to the operational requirements. More importantly, calm water is seldom the actual sea state a vessel experiences. Wind, waves, current, and manoeuvring will affect the vessel performance in a number of ways and it is well known that a seaway can add between 15-40% of the calm water resistance. This is a significant increase, and will affect the overall efficiency of operation substantially by lowering the speed of a ship compared to the expected speed in calm water. Therefore, increased understanding of speed reduction as a phenomenon is key to perform ship operations of high efficiency.

In recent years, increased focus on operational efficiency has emerged in the maritime industry itself, and from regulatory institutions. The environmental impact of the industry is considerable. According to estimates of the International Maritime Organization (IMO), the shipping industry emitted over 796 million tonnes of CO_2 in 2012 (Smith et al., 2015). Due to the massive amount of ship traffic, increased efficiency will be impactful. It is with this as a background that the Marine Environment Protection Committee (MEPC) developed the first industry-wide, mandatory greenhouse gas reduction regime. The result is efficiency indices for all ships that are to be incrementally tightened. Increased efficiency is also a focus area for most shipping companies to improve their profit margins. All in all, there is little downside to increasing operation efficiency.

Today, many companies utilize route optimization tools, both on a ship and fleet level. It is however, not possible to optimize routing with respect to fuel consumption criteria or operational logistics criteria (such as arrival time) without accurate estimation of speed loss due to environmental factors. A tool that can integrate operational parameters, such as speed, position, and heading, together with weather forecasting to predict speed loss, is a integral part of a better voyage optimization solution. Great work has been carried out in trying to calculate speed loss based on a theory-driven approach. However, accurate modelling of the full ship operation is demanding. Therefore, it could be beneficial to employ a different approach. This is where machine learning comes into the picture.

Coinciding with the increased focus on energy efficiency, is an increased focus in the maritime industry to utilize the large amounts of data that is gathered during operation. In recent years the belief that it is possible to extract important information about operational aspects has grown with good reason. Most maritime companies have an increased focus on collecting and analyzing operational data in hopes of improving profit margins and decreasing environmental impact. With increased computing power and

data storage possibilities, it should be possible to better our understanding of ship operation and design. There are many ways to tackle the challenge of analyzing data to learn valuable information. Several tools and methodologies exist. From computer science we have machine learning algorithms that has seen increased attention both in main stream media and the industry. They are able to perform a wide variety of tasks, from facial recognition and sorting social media feeds, to controlling autonomous cars. In this thesis, machine learning algorithms will be utilized to estimate and predict speed reduction. Different classes of regression methods will be used to develop a method that uses data to learn how different external variables impact the loss of speed in a seaway.

1.2 Previous Work

The topic of this thesis is a continuation of work performed in the fall of 2016, which resulted in a project thesis (Gjølme, 2016). This project thesis used a subset of the data found in this thesis to predict the shaft power of a ship. The work that was performed in this project thesis has been a foundation for the work performed in this thesis. The author experienced during the project thesis a lack of expertise regarding the machine learning aspects of the work. Therefore considerable effort has been devoted to general machine learning concepts.

1.3 Objective

The objective of this thesis is to develop a method for estimation and prediction of speed loss. The method shall take some input that describes the operational and environmental condition of the ship, and output a prediction of speed loss. To accomplish this feat, general machine learning regression algorithms will be used. The development of these algorithms into empirical models for speed loss is enabled by the availability of in-service measurements of operational data and hindcast weather data. To achieve this objective, the following tasks has been established:

1. Describe the general nature of ship resistance to establish an understanding of what physical processes influence the relationship between the ship speed, ship resistance and the power output of the propulsor.
2. Describe general concepts needed to utilize machine learning regression in an efficient manner.
3. Describe thoroughly the theory behind artificial neural networks.
4. Perform an investigation of the available data that includes:
   (a) Processing of the data.
   (b) Transformation of the data.
   (c) Evaluate the utility of the data.
   (d) Validation of the data.
5. Develop and test a method that uses machine learning regression algorithms to estimate and predict speed loss.

1.4 Structure of the Report

The rest of the thesis is structured as follows.

- Chapter 2 presents a review of ship resistance and propulsion, both in calm water and in a seaway.
- Chapter 3 concerns machine learning as a discipline. Furthermore, essential concepts to the utilization of machine learning is explained.
- Chapter 4 consists of a thorough review of the theory behind artificial neural networks.
- Chapter 5 presents some concepts regarding the processing of data.
- Chapter 6 is a review and discussion on the data that is used to facilitate the method that has been developed.
- Chapter 7 goes through the development of the method. Aspects and decisions regarding the method is presented and discussed. Lastly, testing procedures are presented.
- Chapter 8 presents results and discussion of the outlined tests, together with a discussion of the method.
- Chapter 9 contains the final conclusion on the work that has been performed.

2 Ship Resistance

In this chapter I will examine the basics of the fundamental physical problem of a ship moving through water. To achieve the goal of better prediction of speed loss for a ship in a seaway, a thorough understanding of the problem is needed. The goal of this chapter is to establish the relationship between the different aspects of ship resistance. This encompasses both the calm water resistance problem and the consequences of including a more realistic seaway. Some parts of the chapter is based on work from the project thesis (Gjølme, 2016).

2.1 Calm Water Resistance

We begin the examination of ship resistance by ignoring the effects of weather, and consider only a ship moving through calm water. Since the goal of this thesis is to estimate the difference between this theoretical calm water condition and the real case, it is important to understand calm water resistance.

The resistance that arise when is moving through water has several physical components. These act on the ship and produce the resistance that has to be overcome by the ship's propulsion system. From a ship moving through water, two observations can be made that illustrate two physical mechanisms that lead to ship resistance. We see that the moving ship produces changes in the surface elevation. This change in surface elevation follows a pattern, a wave pattern, and this wave pattern moves with the ship. The other other observation is that along the hull, a region of turbulent flow will develop. These phenomenon require energy, and the energy comes from the propulsion system of the moving ship. Energy is transferred to the water through forces that are acting on the ship. When the ship moves through the water, we will get distributions of these forces over the hull. The forces, and the physical properties that cause them leads to a natural breakdown of ship resistance.

Depending on how you consider the resistance problem, different breakdowns of the problem is possible. If we consider the forces that are acting on the hull we get the following two components:

1. Due to the viscous nature of water a shear force will develop tangential to the flow over the hull. The component of this shear force that acts along the direction that the hull is moving in can be summed over the hull. The result is what we refer to as the frictional resistance.
2. The pressure distribution will generate a force that acts normal to the surface of the hull. The components of this force that acts along the direction of the moving hull can be summed over the hull. This will result in the pressure resistance.

This breakdown of the resistance is the result when we consider the forces that are acting. If we instead consider the resistance as the result of the dissipation of energy from the hull, we get the following:

1. The energy that is lost in the wake of the ship comprises of energy that is lost due to viscous effects such as separation, as well as the viscous pressure resistance. The viscous pressure resistance is the result of the pressure loss over the afterbody.

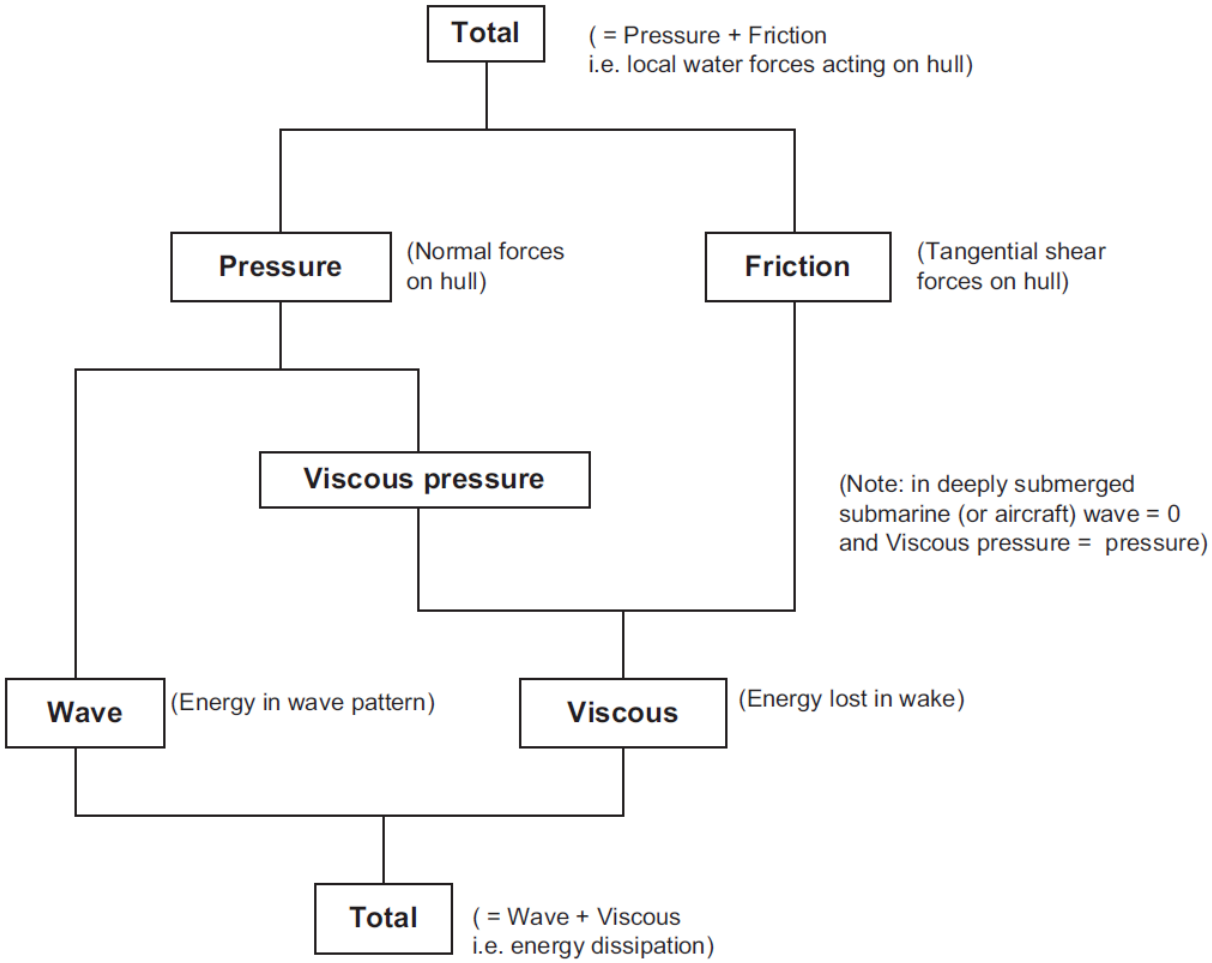


FIGURE 2.1: Components of ship resistance.

2. The other source of energy dissipation is the wave pattern.

A summary that shows how the different breakdowns of the resistance relates is shown in Figure 2.1. This breakdown of ship resistance components and figure 2.1 is based on Molland, Turnock, and Hudson (2011).

How these different forces and energy dissipation mechanisms affect the resistance is dependent on many factors. Aspects of the naked hull has considerable influence on the resistance. The hull geometry, possibly including a bulbous bow, decides the flow over the hull and thus influences the resistance. Another aspect of the hull, is the roughness of the surface. Increased roughness can lead to increased frictional drag forces. The roughness of a given hull will increase over time due to fouling, which is degradation of the hulls performance due to growth of weeds and barnacles. This effect can be mended by periodic docking and painting.

Additional contributions to the resistance comes from the appendages. Appendages can be bilge keels, rudders, shaft brackets or other equipment. In the best case, these appendage will only affect the resistance by their skin friction resistance. However it is difficult to place the appendages in such away that they do not contribute to the resistance by introducing flow separation.

The ship resistance is also dependent on the part of the hull that is above water as well as the superstructure. The size, or projected area, and the aerodynamic properties of the part of the ship that is above water

will influence the overall resistance (Molland, Turnock, and Hudson, 2011).

2.2 Calculating Calm Water Resistance

When designing a ship, the consideration of calm water resistance is a fundamental part. It sets the basis for the propulsion power needed. Thus, calm water resistance prediction has been an important area for ship designers. The calculation of calm water resistance can performed by:

- Performing model test and scale the results according to recommended procedures.
- Utilizing empirical methods developed based on previous model and trial tests.
- Numerical methods using computational fluid dynamics.

2.3 Powering

As we now have described the basic resistance of a ship, we need to describe how this resistance R , which is a force, relates to the speed and power of the ship. To do this in a reasonable way, we need a few different notions of power and efficiencies. Firs we define effective power P_E, which is the product of the resistance R and vessel speed V_S as seen in equation (2.1).

$$P_E = R \cdot V_s \tag{2.1}$$

This power is what is needed to tow a the vessel in question at the speed V_S. We then define what is called delivered power P_D

$$P_D = 2\pi n Q = \frac{P_E}{\eta_D}. \tag{2.2}$$

In (2.2), the delivered power is first defined in terms of n and Q, which is the rotational speed and torque at the propeller respectively, and then expressed in terms of the effective power P_D and efficiency η_D. This η_D is the quasi-propulsive coefficient (Molland, Turnock, and Hudson, 2011), which captures the losses that occur between the propulsive unit and the actual towing power. This coefficient is formulated as

$$\eta_D = \eta_0 \ \ \eta_H \ \ \eta_R, \tag{2.3}$$

where η_0 is the open water efficiency, η_H is the hull efficiency and η_R is the relative rotational efficiency. A more detailed description can be found in Carlton (2007). Lastly we need to describe a final relation. In this thesis it is the shaft power or brake power P_B which is used when power is mentioned. It relates to P_D as

$$P_B = \frac{P_D}{\eta_M}, \tag{2.4}$$

where η_M is the mechanical efficiency.

These methods is how one traditionally has decided on the power that is installed on a vessel. The efficiencies can usually be estimated by empirical methods (Lu et al., 2015). The author will not describe the powering of a ship further, but found it necessary to describe the relationship between the shaft power, which will be used later, the resistance, and the ship speed.

2.4 Speed Reduction

Ship resistance as described in 2.1 is an idealized look at ship resistance as a whole. When we look at the resistance in this way, we look only at a completely calm sea state. The real case is very different, and the power requirements are different. The main difference between the calm sea state and a real sea state is the presence of waves and wind. It is obvious that the waves, especially when they are of significant size, will greatly effect the relationship between power consumption and ship speed. The wind will also affect this relationship similarly to how the air resistance does. However, the wind can of course change direction and it can also reach speeds higher than the that of the air resistance. In a real sea state the motion of the vessel will also impact the relationship between speed and power. This can happen in several ways. Roll motions can alter the underwater hull geometry which can change the resistance. Pitch motion will also alter the geometry, and thus also affect the resistance. Motion in heave and pitch, together with waves can also result in a higher risk of phenomenas such as propeller ventilation and in-and-out-of-water effects. This may result in speed loss (Prpić-Oršić and Faltinsen, 2012).

2.4.1 Voluntary Speed Reduction

There is also the case where, due to environmental factors, the ship master chooses to voluntarily reduce the ship speed. In rough sea states, there are numerous situations that can occur which can cause the ship master to make such a decision. Excessive ship motions can cause unacceptable accelerations, slamming effects and propeller racing. The ship master will act, either by reducing the speed or changing the heading of the ship (Faltinsen, 1993).

2.4.2 Involuntary Speed Reduction

What we call involuntary speed loss stems from a variety of effects. It is useful to make a distinction between speed reduction due to reduced available power, and speed reduction due to increased required power. When it comes to reduced available power, the magnitude of this effects depends mainly on details regarding the ship's propulsion plant as well as the propulsor's interaction with waves and the hull. More specifically, the power plants performance will drop during overload and the ship motions excited by the seastate will reduce the propulsive efficiency (Lewis, 1989, p. 138). The reduction of propulsive efficiency is caused by propeller racing or air drawing, unsteady propeller effects, and reduced hull efficiency (Lewis, 1989, p. 118).

The increase in required power is as mentioned earlier a function of the current seastate, which results in added drag. This is called added resistance. The factors that comprise added resistance are (Lewis, 1989, p. 118)

- The direct wind and wave action.
- The indirect effect of waves linked with ship motions.
- The use of rudder.

To gain a better understanding of how the head sea, which is most critical, affects added resistance, we can consult Maruo (1957). He concluded as follows (Lewis, 1989, p. 118).

(a) The excess resistance is independent of the still water resistance.

(b) The additional resistance is proportional to the square of the wave height.

(c) The pitching motion has a dominating effect upon the resistance increase.

(d) The direct effect of the reflection of sea waves is comparatively small.

(e) The maximum increase of resistance occurs at a slightly higher speed than that for pitch synchronism, if the natural pitching period is longer than the natural heaving period.

The added resistance due to waves is an important part of the added resistance and has been widely studied. Gerritsma and Beukelman (1972), based on Maruo's work developed a far-field method (Seo et al., 2013). They equated the added resistance to the energy radiated by ship motions in pitch and heave (Gerritsma and Beukelman, 1972). In the 60s and 70s many contributed to the study of added resistance and a relationship that stood fast in all the proposed methods is that the added resistance is proportional to the wave height squared. A consequence of the relationship is that the superposition principle can be applied when investigating added resistance in irregular waves (Lewis, 1989, p. 119).

A near-field approach to the problem of added resistance can also be utilized. This approach is based on direct pressure integration. Faltinsen introduced this approach in 1980 together with an added resistance estimation formula for short waves (Faltinsen et al., 1980).

In a more recent study (Seo et al., 2013), a comparison of three different numerical approaches is performed. The strip method, Rankine panel method and Cartesian grid method are the approaches that are utilized. The near-field method (direct pressure integration method) and the far-field method (momentum conservation method and radiated energy method) are used as described in the works previously mentioned. In general all methods and approaches give good results when validated against experimental data. There exists in other words modern numerical methods of calculating the added resistance due to waves.

Wind can sometimes be an important part of vessel resistance. In a calm water situation, the air resistance as a result of the vessel speed may not of great importance. However, in a real sea state we can have high speed winds which will influence the added resistance. The effect will depend on the particulars of the superstructure and hull above water, as well as relative wind speed and direction. The recommended way to calculate the resistance increase due to wind can be found in ITTC (2014). It can also be calculated based on a decomposition of the relative wind (Steen and Minsaas, 2013).

2.5 Empirical Estimation of Speed Reduction

In the previous section we looked at how the added resistance may be calculated based on different theoretical methods. They have in common that they are fairly complicated, and therefore one can be motivated to develop a simpler method for estimation of speed reduction. This has been done by Townsin and Kwon (1983), with an extension by Kwon (2008). This method is based on previously developed methods on calculating speed loss due to wind, ship motions and wave reflection resistance. Firstly the method presents an estimate for percentage speed loss when the seaway is characterized only by the Beaufort number. This characterization is motivated by the Beaufort number's ability to capture effects of both waves and wind in a single metric. Essential to the development of this method is the thrust consistency assumption, which allows for the following relationship between resistance and speed.

$$\frac{\Delta V}{V_{cw}} = \left(1 + \frac{\Delta R}{R_{cw}}\right)^{\frac{1}{2}} - 1 \tag{2.5}$$

In Equation (2.5), ΔV is the speed reduction associated with the resistance increase ΔR. The subscript cw denotes the respective value in calm water. We assume that $R \propto V^2$. The full formula for percentage speed loss is given by the three factors in

$$\alpha \cdot \mu \cdot \left(\frac{\Delta V}{V_{cw}} 100\% \right)_{head}, \tag{2.6}$$

where α is the correction factor for block coefficient C_B and Froude number F_n, μ is the reduction factor for weather direction, and finally $\left(\frac{\Delta V}{V_{cw}} 100\% \right)_{head}$ is the percentage speed loss for head weather calculated as a function of the Beaufort number and the volume displacement.

3 Machine Learning

The following chapter will present a general review of machine learning. Increased understanding of machine learning concepts is a prerequisite for effective use. An introduction of general machine learning is given firstly, followed by an introduction to regression by use of linear regression. Several necessary concepts is then presented in a linear regression context.

In recent years there has been a surge in technology that makes use of machine learning. Although the field of machine learning has been around since the second half of the 20th century, it has been implemented more actively in the last decade due to the rapid increase in computing power. Machine learning has been successfully implemented in many applications in different environments, and has proven very powerful. Some good examples are e-mail spam filters, image classification and language processing. A machine learning approach to an engineering problem differs from the engineering approach. Typically, when faced with some engineering problem, a model is developed based on the relevant theory. However, there are many challenges that can occur with the engineering approach, and analytical solutions is often unattainable. In a machine learning approach we explore data that is measured from the process, and use this data as the basis to develop a useful model.

3.1 Introduction to Machine Learning

The goal of machine learning is to develop programs that can learn to perform a task without being explicitly programmed to perform that task. The key is to develop algorithms that are programmed to learn, rather than being programmed to suit the specifics of a task. A modern, more formal definition of this idea is offered by Mitchell (1997).

> A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

As apparent from this definition, a key part of any machine learning project is the learning, or training. To make use of a machine learning algorithm, it must be trained. The training is what makes the general algorithm mould to a specific problem and allows for useful applications of machine learning. There are two main types of learning situations which reflect the type of problem at hand. We draw a line between problems where there is a clear input-response relationship between the variables in our data. Specifically, we have some response variables $\mathbf{y}$ that we think is related to our input variables $\mathbf{x}$ in some way. The other case is when there is no specific response variable $\mathbf{y}$. They are named as follows:

- **Supervised learning**: Supervised learning is when we have some input data that relates to some output data. The machine learning algorithm is given input data together with the response variable is trained to capture the relationship between them. The goal is to predict correct responses for unknown input data.

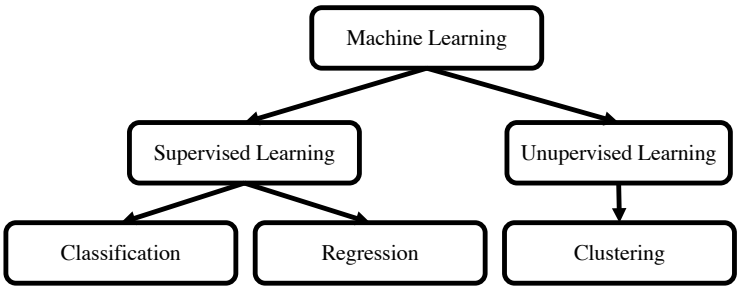


FIGURE 3.1: Illustration of the categorization of machine learning problems.

- **Unsupervised learning**: When there is no specified output data, unsupervised learning aims to discover hidden structure in the data.

Depending on the nature of the response variable in a supervised learning problem, we can further categorize these types of problems. These categories are called regression problems and classification problems. Simply stated, a problem is a regression problem if the response variable is continuous, and it is a classification problem if the response variable takes on discrete values. In a regression problem context, we aim to train a function f that is a continuous mapping from input $\mathbf{x}$ to output $\mathbf{y}$, whereas in a classification problem we aim to train a function f that is a discrete mapping from input $\mathbf{x}$ to output $\mathbf{y}$. This relationship is seen in equation (3.1).

$$y = f(\mathbf{x}) \tag{3.1}$$

In the case of unsupervised learning, there is no response variable that is related to the dataset. Instead we aim to gain knowledge about the structure of the dataset. This can be useful if we want to explore the data and have yet to define a specific goal, or if we are unsure of what information is contained in the data. In most cases of unsupervised learning we employ what is called cluster analysis. this type of analysis aims to group the samples and variables in our dataset based on some shared trait(s). The goal is to define clusters such that objects that to belong the same cluster are similar and objects of different clusters are distinct.

The categorization of the different machine learning problems is illustrated in Figure 3.1. Within each machine learning discipline, numerous algorithms are available. Some algorithms have implementations that work in more than one categories, which hints to the fact that several algorithms work on similar principles.

The rest of this chapter will focus on machine learning where regression is the relevant category of learning. This is typical in a context where the end goal is the prediction of some continuous variable the is the result of a natural process. In general, the response $\mathbf{y}$ may be multidimensional, but in this thesis, we will mostly focus on a single response variable y.

3.2 Introduction to Regression

As previously mentioned, the machine learning problem that is to be solved in this thesis is most sensibly posed as a regression problem. Thus we will review the necessary aspects of regression. To do this we will start by looking at the most basic type of regression model, namely linear regression. A lot of the concepts

that will be utilized throughout the thesis can be understood in a linear regression context. Later we will expand these concepts to more complex regression models.

Firstly we will introduce some notation. We have some training data, which consists of N samples of the D-dimensional vector $\mathbf{x}$, expressed as $\mathbf{x} = (x_1, x_2, \ldots, x_D)^T$. The complete dataset of input variables is denoted as $\mathbf{X} = (\mathbf{x^{(1)}}, \mathbf{x^{(2)}}, \ldots, \mathbf{x^{(N)}})^T$. Together with the input variables we have the output, or response variable dataset. For each sample we have an associated $y^{(n)}$, thus we have the response variable vector which is denoted $\mathbf{y} = (y^{(1)}, y^{(2)}, \ldots, y^{(N)})^T$. We assume that the response variable y is given by some deterministic function f with some Gaussian noise ϵ as seen in (3.2).

$$y = f(\mathbf{x}) + \epsilon \tag{3.2}$$

The goal is to propose a model for the function f parametrized by parameters $\mathbf{w}$. This model is denoted $h_{\mathbf{w}}(\mathbf{x})$.

3.3 Linear Regression

The idea is to form a hypothesis function h that maps the input variables to a prediction of the response variable. This hypothesis function will look different depending on the model we choose. The most basic linear model is presented in (3.3).

$$h_{\mathbf{w}}(\mathbf{x}) = w_0 + w_1 x_1 + \ldots + w_D x_D \tag{3.3}$$

We see that our hypothesis function h is a linear combination of the elements of the input vector and some parameters, or weights, $w_0, w_1, \ldots, w_D$. The weights may be arranged in the parameter vector $\mathbf{w} = (w_0, w_1, \ldots, w_D)^T$. For convenience we may define $x_0 = 1$ such that (3.3) can be expressed as seen in (3.4).

$$h_{\mathbf{w}}(\mathbf{x}) = \mathbf{w}^T \mathbf{x} \tag{3.4}$$

The simplicity of the linear model results in limited performance in many problems, but can serve as a benchmark when it comes to performance. There are many modifications that can be done to better its performance as we shall see.

In (3.3) and (3.3) we have a model that we want to predict the correct output when given the input $\mathbf{x}$. The way we do that is by tuning the parameters. It is this tuning process that we refer to as training. The goal is to train the parameters of the hypothesis function h, such that its predictions are as close as possible to the correct response. To do this we first need some kind of performance measure. In regression, the performance measure is typically based on the sum of the squared errors. Intuitively we want to tune the parameters such that we minimize the sum of squared errors. We refer to this quantity as the cost function. Thus the training aims to minimize the cost function. In the current context a typical cost function is seen in (3.5).

$$J(\mathbf{w}) = \frac{1}{2N} \sum_{i=1}^{N} \left(h_{\mathbf{w}}(\mathbf{x}^{(i)}) - y^{(i)} \right)^2 \tag{3.5}$$

It is known that this optimization problem has a closed form solution based on a least squares approach (Bishop, 2007). This solution, as seen in (3.6), is a technique that is referred to as batch training. The name comes from the fact that the full dataset is processed in one computation. This may not always be possible, and even when it is possible it may be very inefficient due to the required computational power for larger data sets.

$$\mathbf{w} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \tag{3.6}$$

Another batch learning algorithm that has a wider set of applications is gradient descent. The algorithm first initializes the parameter vector $\mathbf{w}$ to some value. Then, for some number of iterations, the algorithm updates the parameter vector according to (3.7),

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \eta\nabla J(\mathbf{w}^{(\tau)}), \tag{3.7}$$

where τ is an iteration variable, $\eta > 0$ is the learning rate parameter, and $\nabla J(\mathbf{w})$ is the gradient of J with respect to the parameters in $\mathbf{w}$. If we have the cost function defined in (3.5), (3.5) will take the form seen in (3.8).

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \frac{\eta}{N}\mathbf{X}^T(\mathbf{X}\mathbf{w}^{(\tau)} - \mathbf{y}) \tag{3.8}$$

There are several training algorithms that work on roughly the same principle as the simple gradient descent algorithm, but with significantly better performance. A discussion on training algorithms will be presented later.

3.4 Adding More Complexity

To this point we have discussed the purely linear regression model. It is apparent that such a simple model will not be able to describe the processes that we want with sufficient accuracy. Most processes need a more sophisticated model. A way to add more complexity is to allow polynomial terms. To do this while staying within the same framework, can be done by adding more feature variables that are powers or products of the original feature variables. In (3.9) we see and example where this has been done for the case of two feature variables of polynomial degree two.

$$\mathbf{x} = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2, x_1^2x_2, x_1x_2^2, x_1^2x_2^2)^T \tag{3.9}$$

This type of variable manipulation can also be done with other types of mathematical operations, not only polynomials. It is in many cases necessary to have more complex models than the purely linear, due to the fact that many processes are nonlinear. The balance between model complexity and simplicity is very relevant in application of machine learning. To illustrate the importance of the right level of complexity, and the machine learning concepts of underfitting and overfitting we will consider a data set of $N = 10$ points that is the result of the function $f(x) = \sin(2\pi x) + \epsilon$, where ϵ is some measurement noise. This data set is illustrated in Figure 3.2.

We can fit a polynomial regression model to this data set do discover the underlying function. The model is of the form

$$h_{\mathbf{w}}(x) = w_0 + w_1x + w_1x^2, \ldots, +w_Mx^M, \tag{3.10}$$

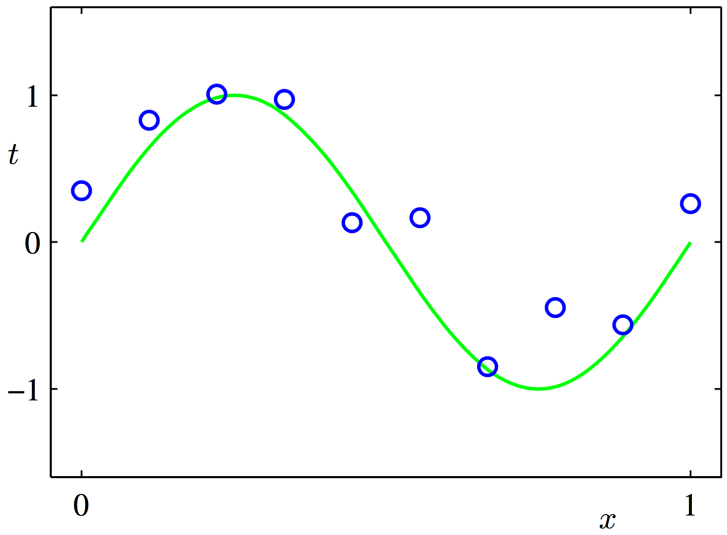


FIGURE 3.2: Sample data set of a periodic function with noise (Bishop, 2007).

where M is the order of the polynomial. It is up to the user to decide the right polynomial order such that the result is useful for predictions on new data, although there are some techniques to guide this. In Figure 3.3, we see the resulting predictions of models of different orders. The green line is the underlying function $\sin(2\pi x)$ and the blue circles are the measurements that form our dataset as we recognize from Figure 3.2. The red lines in Figure 3.3 is the resulting regression line from the model of order M. It is clear that the simplest models, where $M = 0, 1$ are too simple. When $M = 0$ the resulting regression model is simply a constant, which obviously does not describe the underlying function in a good way. For $M = 1$, the resulting regression model is a straight line, which also performs poorly. These two models does not describe neither the underlying function nor the data set in a good manner. The concept of having a regression model that is too simple is called underfitting. We also say that the model has a high bias.

For the model of polynomial degree $M = 9$ we see something entirely different. The regression is able to fit the data set perfectly, due to there being only 10 points in the data set, but it does not represent the underlying function in a good way. Thus, it seems that a model of order $M = 9$ is too complex. This is an example of overfitting, or a model of high variance. The last regression model that has been trained is a model of order $M = 3$. When compared visually, this seems like the best choice of the available models. It represents both the underlying function and the data set fairly well.

The parameters of the models in this example are the result of the minimization of the cost function (3.5) over the training data as described in 3.3. This method of minimizing some cost function over the data is common for any machine learning method, but as evident from the foregoing example, low, or even perfect performance on the training set, does not necessarily produce the model with best generalizing ability. It is necessary to evaluate the the model in some other way. To do this we need additional data for model testing.

3.5 Splitting Training Data

As we have previously shown, the training performance may not be a good indicator of a models predictive performance. To solve this problem we split the data set. We take some fraction of the data set and

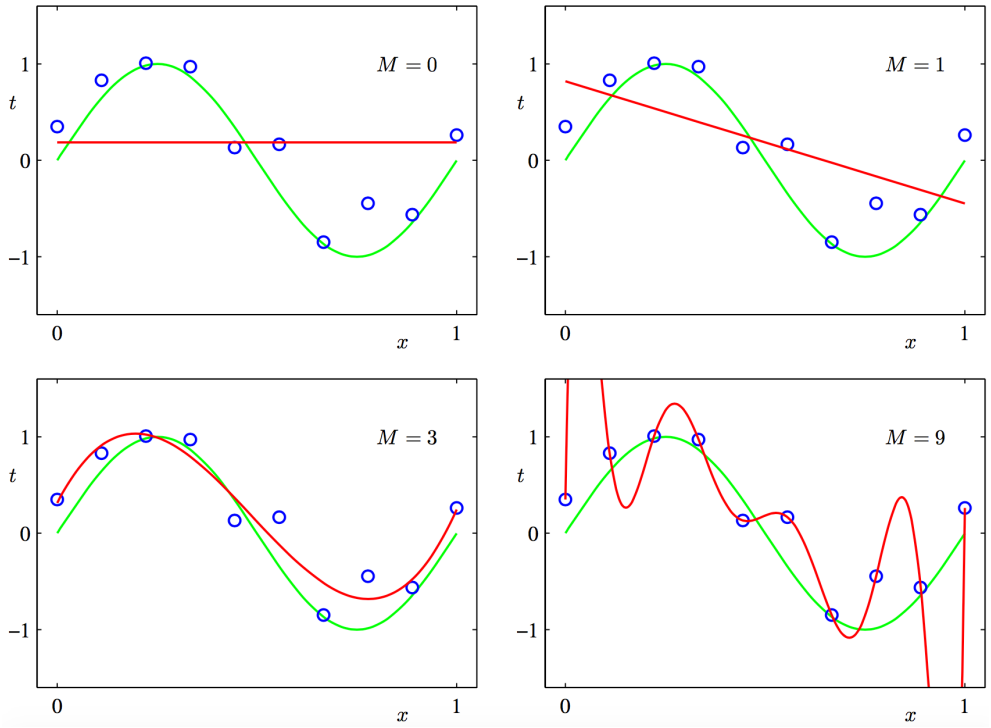


FIGURE 3.3: Plots of regression lines with varying polynomial degree (Bishop, 2007).

use it for model validation or testing. Typically we use most of the data as training data. Then we train different models, for example models of different polynomial degree, and pick the model that performs best on the validation training set. Lastly, we run this model on the test set to get the best indication of the predictive ability of the model.

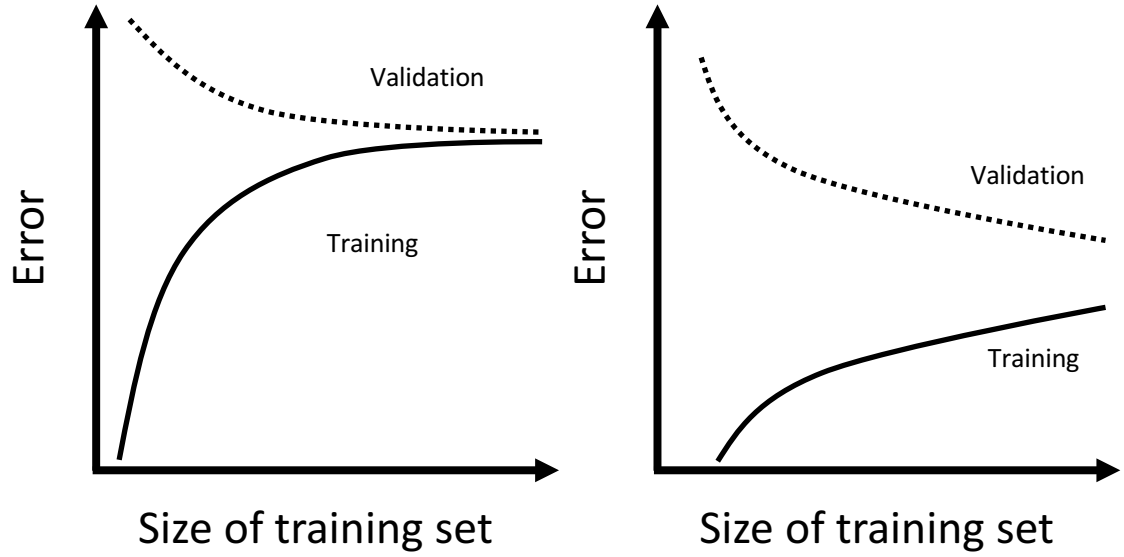
3.6 Symptoms of Over- and Underfitting.

It can be difficult to assess the different models that are available, there are however some techniques that can indicate if a model is suffering from high bias or variance. To evaluate a model we can use a graphical approach that employs what we call learning curves. Illustrative learning curves are presented in Figure 3.4

To make a learning curve we first have to split the data set into a training set and a set for model validation. We then train our model with more and more of the available training data and plot the training error. We also plot the performance of the model on the validation set.

If the model is underfitting the data, the learning curve may look like Subfigure 3.4a.The training error will initially be low, because it is easier for a model to fit a smaller data set. It will then grow asymptotically towards some relatively large error. It seems as if more data will not help. The validation error will decrease and settle at a similar error. If we see this behaviour in our learning curve, we have a model with high bias.

If the model is overfitting the data, the learning curve may look like Subfigure 3.4b. The training data error will start low. It will then grow while still maintaining a relatively low error. The validation error will start high and decrease, but remain relatively high in the case of high variance. We see that the gap



(A) Learning curve with high bias. (B) Learning curve with high variance.

FIGURE 3.4: Example illustrations of learning curves.

between the validation error and training error is decreasing. A key indication of high variance is that it looks like more data helps lowering the validation error.

Now that we may be able to identify why a model is not performing to the desired level of performance we need to know how to mend this problem.

3.7 Model Selection

To prevent our model from over- or underfitting the data we need to choose the best level of model complexity. To pick the right level, we may employ a graphical method. Similarly to how we plot the learning curves, we plot the error of a given model versus some measure of its complexity. We then see how the validation error develops as the model complexity increases. An illustration is seen in Figure 3.5.

The goal is to choose the level of model complexity that has the highest and most stable predictive ability. This translates to having lowest error on the validation data. In Figure 3.5 we see that for a model of low complexity both the training error and validation error is high. This means that our model is too simple and will underfit the data. As the complexity increases, the validation error will decrease until some point. We then have reason to believe that this is the ideal level of model complexity, and further increase in model complexity will result in higher validation error despite the continuous decrease in training error.

Even though this method seems relatively simple, this may not always be the case. The plots rarely looks as clean as in Figure 3.5. It may also be difficult to evaluate complexity. It is simple in the case of a univariate polynomial, but in other cases it can be more difficult to evaluate complexity.

3.8 Regularization

Another way to prevent overfitting is through a technique called regularization. The motivation for this technique is to be able to develop somewhat complicated models even though the training set is of limited

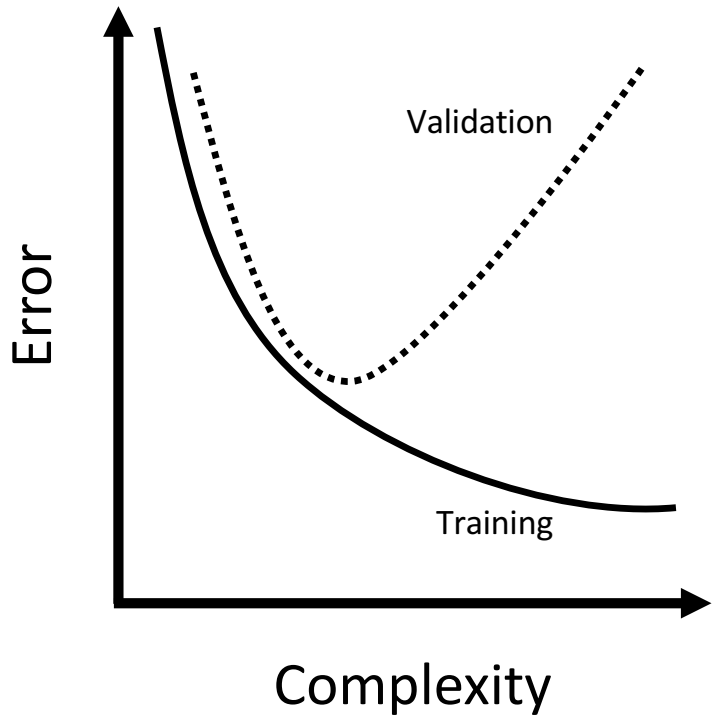


FIGURE 3.5: Illustration of method for evaluating model complexity.

size. In linear regression, it is seen that a model that suffers from very large parameters in the parameter vector $\mathbf{w}$ is probably overfitting the data (Bishop, 2007). To help this we may try to keep these parameters from growing very large. Introducing a new cost function that we minimize to find $\mathbf{w}$ can help limit the parameter size. In linear regression we can formulate a new cost function

$$J(\mathbf{w}) = \frac{1}{2N}\left[\sum_{i=1}^{N}\left(h_{\mathbf{w}}(\mathbf{x}^{(i)}) - y^{(i)}\right)^2 + \lambda\sum_{j=1}^{D} w_j^2\right] \tag{3.11}$$

which is fairly similar to the cost function (3.5), but has a term with regularization parameter λ and a squaring of the parameters in $\mathbf{w}$. This new regularization term will penalize large parameter values, and thus a new parameter vector $\mathbf{w}$ with smaller values will be the $\mathbf{w}$ that minimizes (3.11). Note that we by convection exclude the bias parameter w_0 from the cost function. The trade off between penalizing the parameter size and minimizing the squared error is governed by the regularization parameter λ. So when we want a fairly flexible model, but our training data is limited we can use regularized regression. It is then important to find the optimal regularization parameter λ. We can use a similar graphical approach as we have seen before when we wanted to find a optimal level of model complexity. An illustration is shown in Figure 3.6.

If we consider a model that if trained unregularized will overfit the data, and plot the training and validation performance for increasing regularization it may look like Figure 3.6. For low regularization the training data performance will be very low and the validation performance will be poor. This suggests a model with high variance. The validation error will decrease with increased regularization until some optimal choice of λ. This is the best choice of regularization. Further increase will worsen the validation error.

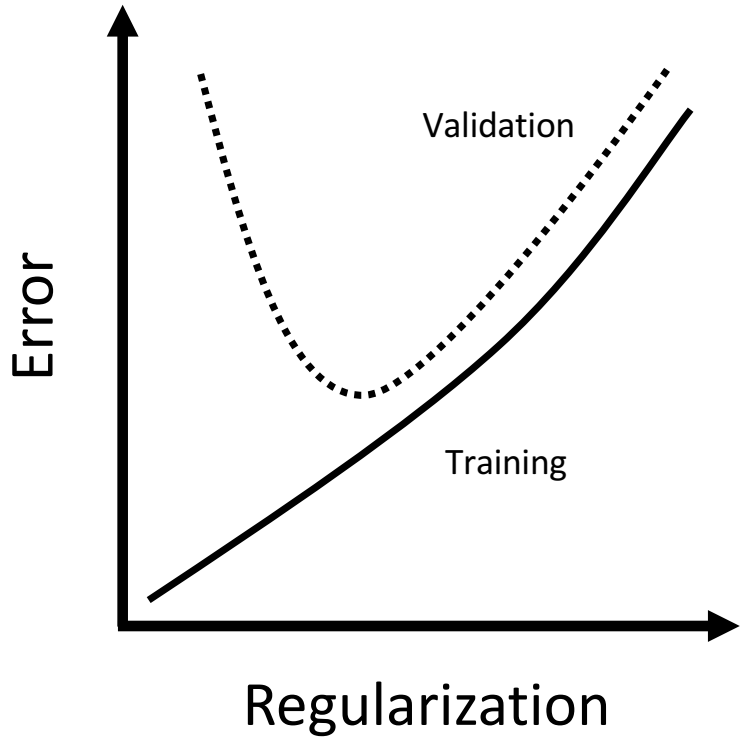


FIGURE 3.6: Illustration of method for evaluating model regularization.

3.9 Debugging the Learning Algorithm

The last part of this chapter will describe general actions that can be taken if a regression model that has been developed performs worse than what is expected or acceptable. In this case we need to figure out if the model suffers from high bias or high variance. The best way to do this is by plotting learning curves as seen in 3.6. This method will hopefully be sufficient to diagnose the learning algorithm with either high variance or high bias. For either case there are some actions that can be done.

3.9.1 High bias

To improve a model that suffers from high bias there are few possible actions. They are based on adding or allowing for more complexity. If regularization is implemented in the algorithm, decreasing the parameter controlling the regularization may be recommended. If this seems to be of little use, we need to consider the training data. As we have seen we can add more complexity by adding polynomial features and interaction features. This allows for the regression model to generalize better to more complex training data. Most applications of machine learning contain some kind of nonlinearity and this it may be necessary to add these nonlinear features. This method of expanding the features used in the model can also incorporating gathering entirely new features. We have to consider weather or not we can expect reasonable predictions based on the features available. If this not is not the case we may be forced to gather additional data to improve the predictive ability of out model.

3.9.2 High variance

It the case of a model that suffers from high variance there are a few things that can be done. In the case of a regularized model, we can decrease the regularization parameter as seen in 3.8. We can also reduce the features we use. We can try to not make use of features that does not help the predictive ability, but rather add noise to our model. Lastly, as we can see from Figure 3.4b, increasing the number of samples in our data set may improve the model. It is important to note that more data is not always better, this is only the case if the model is sufficiently complex.

4 Artificial Neural Networks

In the previous chapter we examined the foundation for developing good regression models using machine learning. Although this was done mostly in the context of linear regression, the concepts are valid for other regression techniques. We saw that going from a simple linear regression model to a model with more complexity was possible if we transform and expand the available features. However the user still has to specify the features as we see exemplified in Equation (3.9). If the relationships between the input data and the target is unknown it can be be inefficient to try to guess the form of the features. In this chapter we will introduce a nonlinear regression model known as Artificial Neural Network (ANN). The name reflects the fact that it is loosely based on a biological nervous system. It this chapter the ANN-model will be presented together with other aspects of using a neural network for regression.

4.1 Mathematical Model

The basic neural network that will be used in this thesis has the following functional building blocks. We begin with the input layer which is simply the elements of feature vector $\mathbf{x}$[1]. This means that the input layer consists of D units, which is the dimension of $\mathbf{x}$. We then construct the hidden layer. This will have M units, so we construct M linear combinations of $\mathbf{x}$. This will look like

$$a_j = \sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}, \tag{4.1}$$

where $w_{ji}^{(1)}$ denotes the weight of the ith feature in the jth unit of the 1st hidden layer. The superscript denotes the layer. $w_{j0}^{(1)}$ is the bias of unit j. The quantities a_j are known as activations and is transformed through an activation function $g(\cdot)$, which is some differentiable function. In the hidden layer this is typically the tan-sigmoid function.

$$g(\mathsf{x}) = \tanh(\mathsf{x}), \tag{4.2}$$

as seen in Figure 4.1, and the identity function (4.3) in the output layer of a regression network.

$$\sigma(\mathsf{x}) = \mathsf{x} \tag{4.3}$$

The quantities that is the result of the activations mapped through the activation function is known as the hidden units z_j.

$$z_j = g(a_j) \tag{4.4}$$

[1]The notation in this chapter is largely based on Bishop (2007).

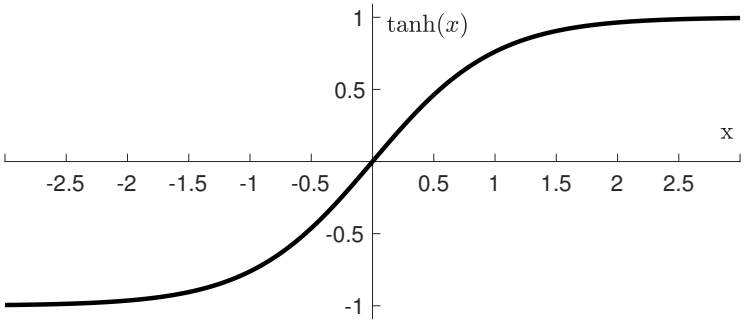


FIGURE 4.1: Plot of the tan-sigmoid transfer function as described in (4.2).

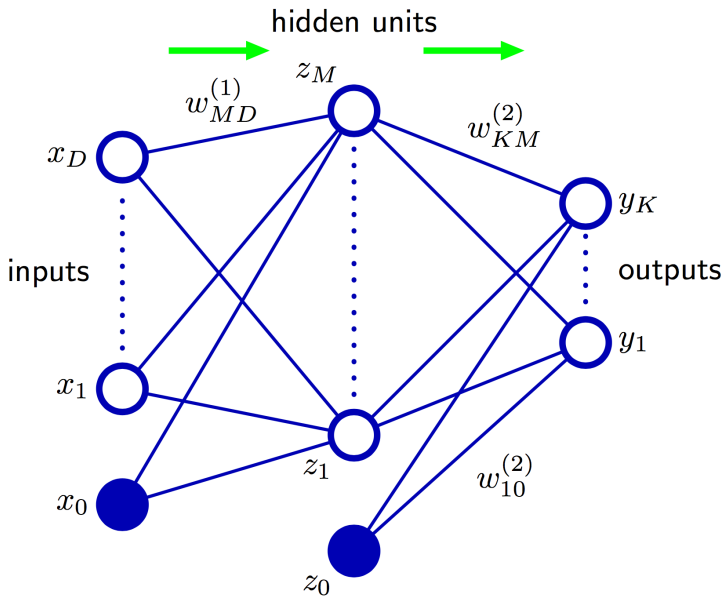


FIGURE 4.2: Neural network diagram depicting a network of two layers (Bishop, 2007).

Similarly we produce the output activations

$$a_k = \sum_{i=1}^{M} w_{kj}^{(2)} z_j + w_{k0}^{(2)}, \tag{4.5}$$

for the K outputs which in our setting is one. Thus, we get the activation

$$a_k = w_{kj}^{(2)} z_j + w_{k0}^{(2)}, \tag{4.6}$$

and network output

$$h_{\mathbf{w}}(\mathbf{x}) = \sigma(a_k) = a_k. \tag{4.7}$$

The combination of these functions can be written in the final network hypothesis function

$$h_{\mathbf{w}}(\mathbf{x}) = \sigma\left(\sum_{j=1}^{M} w_{kj}^{(2)} g\left(\sum_{i=1}^{D} w_{ji}^{(1)} + w_{j0}^{(1)}\right) + w_{k0}^{(2)}\right). \tag{4.8}$$

This construct is typically presented in a network diagram as shown in Figure 4.2.

4.2 Network Configuration

In general, a neural network can have many configurations. The difference between the different configurations is the choice hyperparameters in the network. The hyperparameters are, as opposed to the previously mentioned parameters w, the parameters that decide the structure and characteristics of the network. In a neural network there are several different hyperparameters that all influence the regressional performance of the network.

4.2.1 Hidden Layers and Units

Maybe the most important hyperparameters in a neural network is the number of hidden layers and the number of hidden units in each layer. The choice of these hyperparameters guide the networks ability to fit complex nonlinear functions. More layers and more hidden units increase the available network complexity. The choice of these parameters is very important. By using techniques that are based on the concepts presented in Chapter 3, we are able to make good choices when it comes to network complexity. In this thesis we will restrict the number of hidden layers to one. It is known that a neural network with a single hidden layer can approximate any function to an arbitrary degree of accuracy given a sufficient number of hidden units (Hornik, Stinchcombe, and White, 1989; Hornik, 1991). The use of neural networks with more than one hidden layer is used in the machine learning discipline called deep learning. In this thesis deep learning neural networks will not be used. The number of hidden units in the single hidden layer is still an important topic. There are several aspects of the problem that guide what is a good network structure. For instance, the number of data points directly guides what is a good level of network complexity (Bishop, 1995). Another point is that the number of hidden units may also be limited by the computational power at hand, since more complex networks require more computing power to train. To develop a network that generalizes well, a reasonable choice of hidden units as well as the use of other generalization methods is needed. Regularization and early stopping in the context of neural networks will be presented later in this chapter.

4.2.2 Activation Function

Another important parameter is the activation function. As mentioned in 4.1 the tan-sigmoid function is often used. Early neural networks typically used a discontinuous step function. These multilayer networks was called multilayer perceptrons. The term activation function comes from the fact that the output was binary; one or zero. An output of one, denotes an activated function. In modern neural networks we employ continuous nonlinearities in the form of sigmoidal functions. They have in common that they squish the input to some range, $[0, 1]$, as in the case of the standard logistic function

$$g(\mathrm{x}) = \frac{1}{1 + e^{-\mathrm{x}}}, \tag{4.9}$$

or to $[-1, 1]$ as in Equation (4.2).

The use of these continuous functions results in the network function being differentiable. This property is important when it comes to network training (Bishop, 2007).

4.3 Regularization

Regularization has already been mentioned in Chapter 3. Here the concept of regularization was introduced as away to handle overfitting by penalizing large model parameters. In the context of neural networks we can consider regularization as a tool for controlling model complexity, alternative to changing the number of hidden units. We can refer to the approach of choosing a large number of hidden units and control complexity by the regularization parameter. The simplest regularization techniques is called weight decay and works by introducing a term on the form

$$\tilde{J}(\mathbf{w}) = J(\mathbf{w}) + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w}. \tag{4.10}$$

In 4.10, $J(\mathbf{w})$ denotes regular squared error cost function. This form of cost function is the same as we introduced in Equation (3.11). We see that we can control the size of the parameters in $\mathbf{w}$ by changing λ. This parameter can thus be considered a hyperparameter of the network. To find the ideal choice for the regularization parameter we can consult the method illustrated in Figure 3.6.

4.4 Network Training

The training a neural network is similar to the training of a regression model as seen in 3.3. We have a model, which in this case is the network hypothesis function (4.8), that is parametrized by the weights w. To evaluate our model we need some measure of its performance. Previously we used the cost function $J(\mathbf{w})$ seen in (3.5), based on the mean of the errors squared. This is common in neural network regression too. To make our model perform as good as possible we need to solve the optimization problem of minimizing the cost function subject to the parameter weights. This is equivalent to maximizing the likelihood function. In other words maximizing the likelihood the that observations $\mathbf{t}$ came from the $\mathbf{x}$-dependent hypothesis function $h_{\mathbf{w}}(\mathbf{x})$. Unlike in the case of linear regression, the nonlinear network function results in the cost function being nonconvex. This makes the optimization problem more difficult as it facilitates the possibility of local minima. A local minimum exist when other points in the parameter space has the property that the gradient of the cost function is zero and any sufficiently small step in the parameter space increases the value of the cost function. An analytical solution to $\nabla J(\mathbf{w}) = 0$ is not possible, so iterative methods must be used. There are many techniques to solve this optimization problem and approach the global minimum of the cost function. Most techniques are based on first initializing the the parameter vector to some initial $\mathbf{w}^{(0)}$ and then iteratively updating the parameter vector as the the algorithm moves through the parameter space according to

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} + \Delta\mathbf{w}^{(\tau)}. \tag{4.11}$$

We recognize this form as the basis for the previously mentioned gradient descent algorithm (3.7) that was presented in 3.3. However, many other techniques exist that have different ways of finding an appropriate update rule $\Delta\mathbf{w}^{(\tau)}$. Gradient descent is maybe the most intuitive, but it is lacking compared to other algorithms when it comes to speed and robustness. Other algorithms do have in common with gradient descent that they too make use of the gradient of the cost function. Examples of more efficient algorithms are conjugate gradients or quasi-Newton methods (Bishop, 2007). These algorithms are batch learning algorithms, meaning they include the whole training set in all parameter updates. There are also sequential learning algorithms that process one sample at the time.

4.4.1 Weight Initialization

As mentioned, algorithms first initialize $\mathbf{w}^{(0)}$ before making the first update. Some consideration of the initialization is important because it may affect the training algorithm's ability to produce good parameters. The weights are generally initialized randomly to small values. The reason for the values being being small is to avoid saturation of the sigmoidal activation functions. This saturation leads to a small sigmoidal derivative, which subsequently leads to smaller gradients and slower learning (Bishop, 1995). On the other hand, too small initializations will result in the activation functions being approximately linear which also can lead to slow training (Bishop, 1995). The randomness of the initialization is to alleviate the consequences of network symmetries.

4.5 Backpropagation

The gradient of the cost function is, as we have seen, an important tool in the optimization problem. In gradient descent and other algorithms continuous evaluation of the gradient ∇J is needed, and thus it is important to have an efficient way of evaluating this gradient. In neural network training, the error backpropagation algorithm offers a computationally efficient way to do this evaluation.

Consider a standard feed-forward network with some number of inputs and a single output. A general activation, similar to (4.1), can be written as

$$a_j = \sum_i w_{ji} z_i \tag{4.12}$$

where z_i are the inputs to the unit in question.[2] This general activation is then transformed by the activation function $g(\cdot)$ to produce the input to the next layer (or the output). For each training sample we calculate all these unit values to create the forward propagation of information. This propagation leads to the hypothesis denoted $h_{\mathbf{w}(\mathbf{x})}$ which use to evaluate the cost by

$$J_n(\mathbf{w}) = \frac{1}{2}\left(h_{\mathbf{w}}(\mathbf{x}^{(n)}) - y^{(n)}\right)^2. \tag{4.13}$$

Now we can consider the gradient of this cost function with respect to some w_{ji}. This gradient is given as

$$\frac{\partial J_n}{\partial w_{ji}} = \left(h_{\mathbf{w}}(\mathbf{x}^{(n)}) - y^{(n)}\right) x_i^{(n)}. \tag{4.14}$$

The general gradient expression can be rewritten with the chain rule as

$$\frac{\partial J_n}{\partial w_{ji}} = \frac{\partial J_n}{\partial a_j}\frac{\partial a_j}{\partial w_{ji}}. \tag{4.15}$$

To simplify the notation it is common to define

$$\delta_j \equiv \frac{\partial J_n}{\partial a_j}. \tag{4.16}$$

such that (4.15) becomes

$$\frac{\partial J_n}{\partial w_{ji}} = \delta_j z_i \tag{4.17}$$

[2] We do not consider the biases explicitly because they can be handled by defining a $z_0 = 1$ similar to what was done in 3.3.

due to

$$\frac{\partial a_i}{\partial w_{ji}} = z_i \tag{4.18}$$

from 4.12.

We now see from Equation (4.17) that to calculate the the derivative, we must simply multiply the unit output δ by the unit input z. For the output layer this yields

$$\delta_k = \mathbf{x}^{(n)} - y^{(n)} \tag{4.19}$$

Finally for any particular unit which sends signals to some other unit indexed k, we have that

$$\delta_j = g'(a_j) \sum_k w_{kj}\delta_k \tag{4.20}$$

We summarize the algorithm as it is presented in Bishop (2007)

1. Let some input $\mathbf{x}_n$ propagate through the network and calculate activations by (4.12) and (4.4).
2. Calculate all output δ_ks by (4.19).
3. For all hidden units, use the backpropagation formula (4.20) to calculate the corresponding δs.
4. The derivatives are now possible to calculate by (4.17).

The discussion on why exactly the backpropagation algorithm is computationally efficient has been left out.

4.6 Training Algorithms

It is essential for the efficient use of neural networks to consider the choice of training algorithm. The problem that these algorithms aim to solve has been well established. We have the cost function that we want to minimize by changing the network parameters, it has been established that the gradient of the cost function is important, and that backpropagation is an efficient way to evaluate the gradient. Most algorithms operate on the basis of what has been established previously. They do however do this in different ways.

It is also worth noting that the backpropagation can also be used to calculate other derivatives. The Jacobian matrix $\mathcal{J}$ has elements of first order derivatives that are defined as

$$\mathcal{J}_{ki} \equiv \frac{\partial h_k}{\partial x_i}, \tag{4.21}$$

where h is the network output. Along with the Jacobian, we have the Hessian matrix $\mathcal{H}$. The Hessian is the second order derivatives of the cost function with respect to the weights.

$$\mathcal{H} \equiv \frac{\partial^2 J}{\partial w_{ji}\partial w_{lk}}, \tag{4.22}$$

These matrices of derivatives are mentioned because they are used, especially the Hessian, in many non-linear optimization algorithms. Conjugate gradient algorithms make implicit use of second order derivatives by a local Hessian matrix. Algorithms based on Newton's method make explicit use of the Hessian, or an iteratively developed approximation of the inverse Hessian. These algorithms are general-purpose

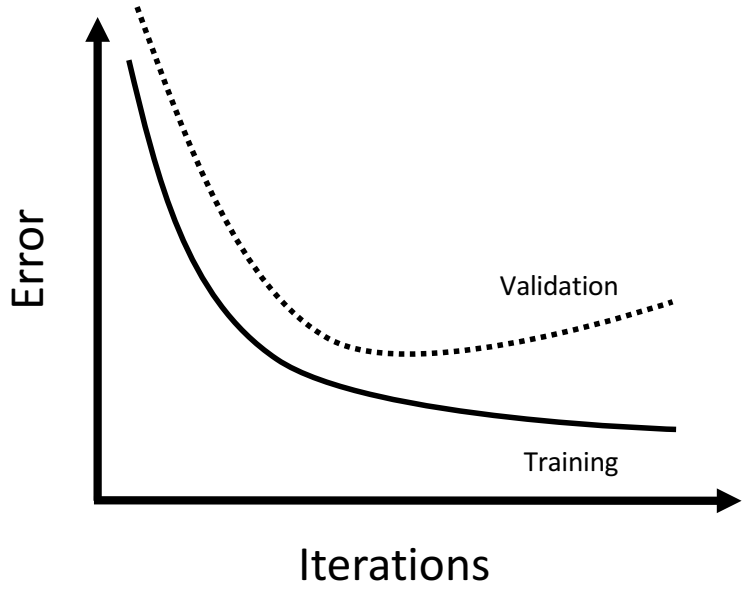


FIGURE 4.3: Illustration early stopping.

methods that can handle a wide class of cost functions. If we consider only a sum-of-squares cost, we can make use of the Levenberg-Marquardt algorithm. This algorithm has a parameter that controls the balance between regular gradient descent and Newton's method to ensure fast convergence. More specifically, this parameter governs the size of the trust region in which the linear approximation of the error function is trusted. A comprehensive review of training algorithms is found in Bishop (1995).

4.7 Early Stopping

As we have mentioned earlier, there exists a relationship between the ideal network complexity and the number of data points available in the training set. The consequence of this is that if one wants to train different, but similar data sets, it can be cumbersome to constantly work on having a good level of network complexity. An example of a case where we have similar data sets of different size, can be when we have data sets with the same input variables for different ships. It would be advantageous if there was another method to avoiding overfitting. A method that can fulfill this need is what is known as early stopping.

A property of many training algorithms is that they guarantee that the cost function does not increase from one iteration to the next. This property comes from the adaptive implementation of the learning rate η. However, the validation error does not conform to this behaviour. A common behaviour of the different errors is that after an initial rapid drop, the training error continues to decrease more slowly while the validation error eventually increases. An illustration of this is shown in Figure 4.3. Early stopping then refers to a method where training stops when the validation error starts to increase, effectively limiting network complexity.

When a squared error cost function is used, early stopping behaves similarly to regularization by weight decay. In early stopping, the product of the iteration index and the learning rate, $\tau\eta$, functions as the reciprocal of the regularization parameter λ (Bishop, 2007).

The advantage with early stopping is how it allows for more lenience when it comes to network structure. If early stopping is implemented, the network designer can employ a network design strategy where a

large number of hidden units is used, and overfitting is handled by early stopping. This is particularly useful in a case where one has data sets of varying size, since the ideal network complexity, in terms of hidden units, is highly dependent on the size of the data set. It is known that early stopping works well, and it is disproportionately more beneficial to have a too large network than it is disadvantageous to have a too small network (Caruana, Lawrence, and Giles, 2000).

5 Processing of Data

The following chapter describes some concepts pertaining to the necessary processing of data. It is largely based on work done in Gjølme (2016), although some additional details have been added.

A substantial and time consuming phase in a data driven research project is the data preprocessing or data preparation phase. This phase consists of several processes which aims to make it easier for the machine learning algorithm to work efficiently on the training data, which will subsequently increase the quality of the empirical model. A data set can have multiple problems which can lower its quality. Some of the most prevalent and general problems are stated below.

- A data set can have redundant information.
- A data set can have irrelevant information.
- Measurement can contain significant measurement noise.
- A data set can contain missing values.
- Measurements can be outliers.

There exists numerous methods and algorithms to resolve or alleviate the severity of these issues. I will present some of the most well known procedures in a general sense, and some that is important within a framework of machine learning.

5.1 Data Selection

By selecting a portion of the available data set, it is possible to have a better understanding of the results. It can be advisable to remove portions of the data that carry an increased level of uncertainty. The motivation behind selection different proportions f the data set is to have a clearer idea of what characterizes the domain where a data-driven approach is valid. This will increase understanding of the level of confidence that can be placed in the validity of a particular model. If this is done in a sensible manner, it is possible to extract information and build models with more defined applications

5.2 Outliers

An outlier is an observation that lies a considerable distance from the other observations in the data set. The observation can be an outlier due to some irregularity in the nature of what is measured. If this is the case, the observation is not wrong, but it possesses some abnormal quality. The other case is that the observation is wrong. This is the case when the observation does not represent the actual quantity that was measured due to some unforeseen error. Examples of this can be sensor faults or error in the transmission of data. It can be challenging, depending on the context, to detect what kind of outlier is present. It is however important to try make an effort, because the failure to remove a erroneous measurement will

decrease the validity of the results. Conversely, the removal of a irregular, but correct, measurement will lessen the analysis' ability to capture the nature of the system that is studied.

An extension of the outlier concept that should be noted is that if you extend an observation to contain more than one measurement, an observation can be an outlier in the multivariate sense. In the former paragraph, we discussed the concept in a univariate sense. In the multivariate sense, the various measurements may be within the expected range, but the combination of the measurements can make the observation a multivariate outlier. To deal with these challenges can be difficult in practice, but important to be aware of.

In the univariate sense we can deal with outliers on a variable-by-variable basis, and define unacceptable values that will be deemed outliers. One method is to define minimum and maximum allowed values. Another method is to define some allowed deviation in terms of the mean and the standard deviation.

It may also be useful to employ some graphical methods to get an overview of the data. Scatter plots can help create an impression of the correlations between variables. Furthermore, histograms can display the distribution of a variable and indicate outliers. In some cases, box plots may also be a viable alternative for discovering outliers.

5.3 Missing Values

In large data sets missing values are very common. They can occur for several reasons. How one chooses to handle them should depend of the context, and many sensible options are available. Some suggestions can be to substitute the missing value with the variable mean, substitute it based on some regression or simply ignoring the measurement all together.

5.4 Normalization

Often the input data varies in range and order of magnitude. It is important to normalize the data such that all the variables are on the same order of magnitude. This is important for neural networks (Kotsiantis, Kanellopoulos, and Pintelas, 2006). Failure to do this can result in the training of the network to be very slow. This has to do, as mentioned in 4.4.1, with the saturation of the sigmoid function, which yields low gradients (Beale, Hagan, and Demuth, 2016, p. 3-9). This is the reason why one often applies pre- and postprocessing to the data. A common way of doing this is to apply the normalization algorithm in Equation (5.1).

$$z = \frac{(z_{max} - z_{min})(x - x_{min})}{(x_{max} - x_{min})} + z_{min} \tag{5.1}$$

Where z_{min} and z_{max} is defines the range of the output of the precessing. Typically they will be -1 and 1 respectively. The process is the reversed on the output data.

Another normalization algorithm that is often used is the following mean-centered normalization:

$$z = \frac{x - \bar{x}}{s_x} \tag{5.2}$$

where $\bar{x}$ is the sample mean and s_x is the sample standard deviation. z will then be a transformation of x with zero mean and unit standard deviation.

6 Data

In this chapter I will present the data that is used in this thesis. Up until this point, we have been fairly general, but onward the focus will be on the actual data used. The approach in this thesis is data-driven, which requires significant care when it comes to the selection and processing of data. This is because what is done in this phase of the analysis directly impacts the results and their validity. There are several aspects of the data that has to be taken into consideration, and these aspects will be covered in this chapter. Firstly, a description of the data and the sources used is presented. Secondly, the methods of data processing is described. These methods are described in 5. Lastly, some validation studies and a discussion on the selected features is presented. Some of the presented material is based on work done in the project thesis (Gjølme, 2016).

6.1 Data Description

The data used in this thesis comes from two kinds of sources. The first is from an onboard performance monitoring system. The other type is weather hindcast data. The goal of the study is to learn how the weather reported in the hindcast data affected the ships in operation as documented by the onboard performance monitoring data.

6.1.1 Ship Monitoring Data

The ships that supply the data in this study is as previously mentioned equipped with an onboard performance monitoring system called Marorka Online. This system collects the data that is generated on a ship, and makes it possible to obtain an overview of the performance of the entire fleet. From this source, I have two data sets. One data set, hereby referred to as the fleet data set, contains data for a number of ships over approximately six months. The ship names have been anonymized. Some key aspects of the fleet data set is the following.

- The data is sampled from 24 different ships, from five different classes.
- All ships are part of an open hatch fleet.
- The time period available is from 00:15 January 1st 2016 to 08:30 June 14th 2016 (UTC).
- The sampling period is 15 minutes.
- The total amount of samples is 330 414.
- The number of different measurements is 32.
- The samples are distributed unevenly over the ships.

TABLE 6.1: Sampled measurements from the monitoring system in the fleet data set (*Mororka Online User Guide* 2016).

Name	Unit	Comment
Company Name	-	-
Ship Name	-	-
IMO No	-	Not used.
Date Time	-	Date and time of sample (UTC).
Voyage ID	-	Not used.
Number of Records	-	Not used.
Cargo	[t]	-
State	-	Operational state of ship.
Boiler Consumed	[MT/24h]	Fuel oil consumption of all boilers.
DG 1-6 Electrical Power	[kW]	Electrical power from generator #.
DG Total Consumed	[MT/24h]	Fuel oil consumption of all diesel generators.
DG Total Electrical Power	[kW]	Total electrical power.
Draft Aft	[m]	-
Draft Fore	[m]	-
Draft Mean	[m]	Calculated from fore and aft draft.
Trim	[m]	Calculated by subtracting the fore from the aft draft.
Sea Depth	[m]	-
GPS Speed	[knots]	Speed over ground (SOG).
Log Speed	[knots]	Speed through water (STW).
Latitude	[°]	In decimal.
Longitude	[°]	In decimal.
ME Consumed	[MT/24h]	Main engine fuel consumption.
Shaft Power	[kW]	-
Shaft RPM	[rpm]	-
Beaufort (BF)	-	Beaufort weather value. Calculated from wind measurements.
Wind Direction Relative	[°]	Relative wind direction (0-360).
Wind Speed Relative	[m/s]	Wind speed relative to the ship.

In addition to the fleet data set, we have what will be referred to as the single ship data set (ss-data set). This is a data set consisting of measurements from a single ship, over a period of approximately three years. Some key aspects of the ss-data set is the following.

- The time period available is from 00:15 May 1st 2014 to 17:15 April 2nd 2017 (UTC).
- The sampling period is 15 minutes.
- The ship is Ship E10, which is also part of the fleet data set.
- The total amount of samples is 101 229.
- The number of different measurements is 26.

An overview of the sampled features from the fleet data set and the ss-data set is presented in Table 6.1 and Table 6.2 respectively.

What the different variables describe varies greatly in nature. Some describe the name of the company or the ship in question, others, for instance voyage ID, are completely blank and unused. It is apparent the much of the available data will not be used. What data will be used is dictated by hydrodynamic considerations based on the theory of Chapter 2. This discussion will be presented later in this chapter.

TABLE 6.2: Sampled measurements from the monitoring system in the single ship data set (*Mororka Online User Guide* 2016).

Name	Unit	Comment
Ship Name	-	-
IMO No	-	Not used.
Date Time	-	Date and time of sample (UTC).
Cargo	[t]	-
State	-	Operational state of ship.
DG 1-3 Electrical Power	[kW]	Electrical power from generator #.
Aux consumed	[MT/15min]	Auxiliary engine(s) consumption.
Aux electrical power output	[kW]	Electrical power output of auxiliary engine(s).
Draft Aft	[m]	-
Draft Fore	[m]	-
Draft Mean	[m]	Calculated from fore and aft draft.
Trim	[m]	Calculated by subtracting the fore from the aft draft.
Sea Depth	[m]	-
GPS Speed	[knots]	Speed over ground (SOG).
Log Speed	[knots]	Speed through water (STW).
Latitude	[°]	In decimal.
Longitude	[°]	In decimal.
OG heading	[°]	Heading over ground.
Gyro heading	[°]	-
ME Consumed	[MT/15min]	Main engine fuel consumption.
ME Load measured	[%]	-
Shaft Power	[kW]	-
Shaft RPM	[rpm]	-
Shaft Torque	[kNm]	-
Beaufort (BF)	-	Beaufort weather value. Calculated from wind measurements.
Wind Direction Relative	[°]	Relative wind direction (0-360).
Wind Speed Relative	[m/s]	Wind speed relative to the ship.

In this thesis there are some uncertain aspects when it comes to the measurements of the data. An investigation of how all the different features are measured have not been done. This introduces a level of uncertainty when it comes to the general accuracy of the measurements. To gain a better understanding of the quality of the data, an in depth investigation of the measuring techniques, equipment and other aspects should have been performed. This is however out of the scope of the current investigation. The analysis will therefore be performed under the assumption that the data in general, is of reasonably high quality. Validation work will be performed to establish a level of confidence in this assumption. This work will be presented in 6.5.

6.1.2 Climate Data

The climate data that is used to augment the data gathered from the ships, originates from The European Centre for Medium-Range Weather Forecasts (ECMWF). The ECMWF provide analyses and reanalyses of weather. These (re)analyses are developed by a process where modeling and observations of many sorts combine to make a consistent best estimate of atmospheric and oceanographic parameters. They gather large amounts of data from large areas and do weather predictions and hindcast analyses of weather and other environmental conditions. The data from these sources should provide an adequate description of the sea states experienced by the ships, such that it is possible to evaluate the influence of the sea states.

ECMWF Data

The data I have used by the ECMWF are from their ERA-Interim reanalysis. This is a global atmospheric reanalysis. The reanalysis spans in time from 1979 and is continuously updated[1]. There are some key characteristics that has to to be mentioned about the ERA-Interim analysis.

- The data has a spatial resolution of roughly 80 km (T255).[2]
- The temporal resolution of the data is 6 hours.

The ECMWF offer a great variety of data from their archives. In addition to the ERA-Interim reanalysis, they have several other archives of analyzed data spanning different time periods and atmospheric and oceanographic areas. The data that is deemed important, and subsequently used in this analysis is presented in Table 6.3.

6.1.3 Conventions and Reference Frames

Due to the various sources of data, it is important to have clearly defined conventions of how the data defined in regards to direction and sign conventions. This is also important because we will operate in two different frames of reference. The first is the earth fixed reference frame, which will be the absolute reference frame. The second is the relative ship fixed reference frame. This reference frame follows the ship speed and heading. In the absolute reference frame, the position is denoted by the longitude and latitude as seen on the world map in Figure 6.1.

The weather data is expressed in the absolute frame of reference, most importantly wind (true wind), and the mean wave direction. The sign and direction conventions for these variables are indicated in Figure

[1] The reanalysis was continuously updated during the writing of this thesis as of June 2017

[2] T255 signifies the spectral truncation at which some of the fields in the analysis are saved. There is a relationship between this and the grid point resolution T255 is related to the reduced N128 Gaussian grid (Paul Berrisford, 2011).

TABLE 6.3: Weather parameters from ECMWF (ECMWF, 2016).

Name	Unit	Comment
10 metre U wind component	[m/s]	Eastward wind component at 10 m above sea level.
10 metre V wind component	[m/s]	Northward wind component at 10 m above sea level.
Mean wave period	[s]	Spectral mean wave period, computed by integrating over all directions.
Mean wave direction	[°]	Spectral mean direction over all frequencies and direction of the two-dimensional wave spectrum. Zero means 'coming from north' and 90 'coming from east'.
Significant wave height	[m]	Defined as 4 times the square root of the integral over all directions and all frequencies of the two-dimensional wave spectrum.

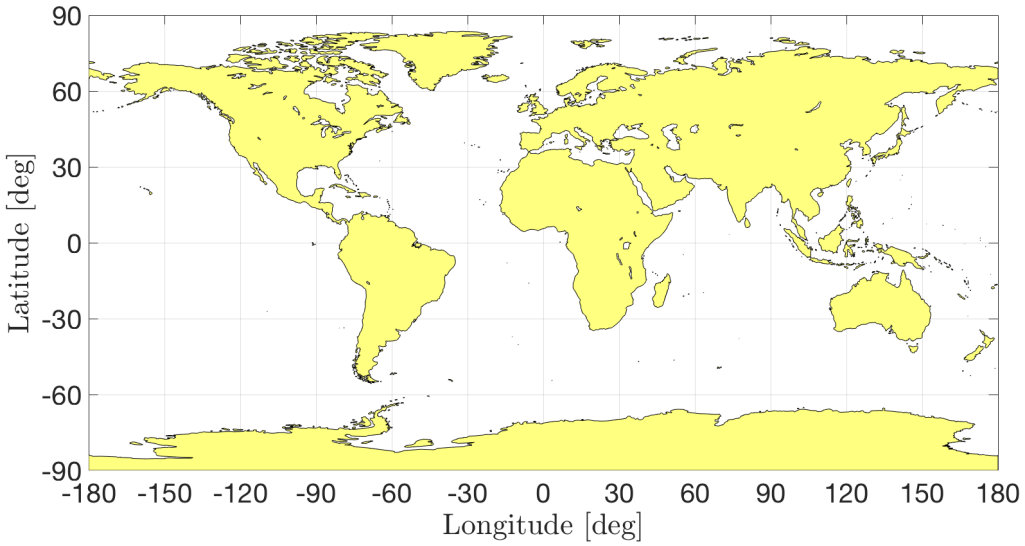


FIGURE 6.1: World map shown with the plate carrée map projection.

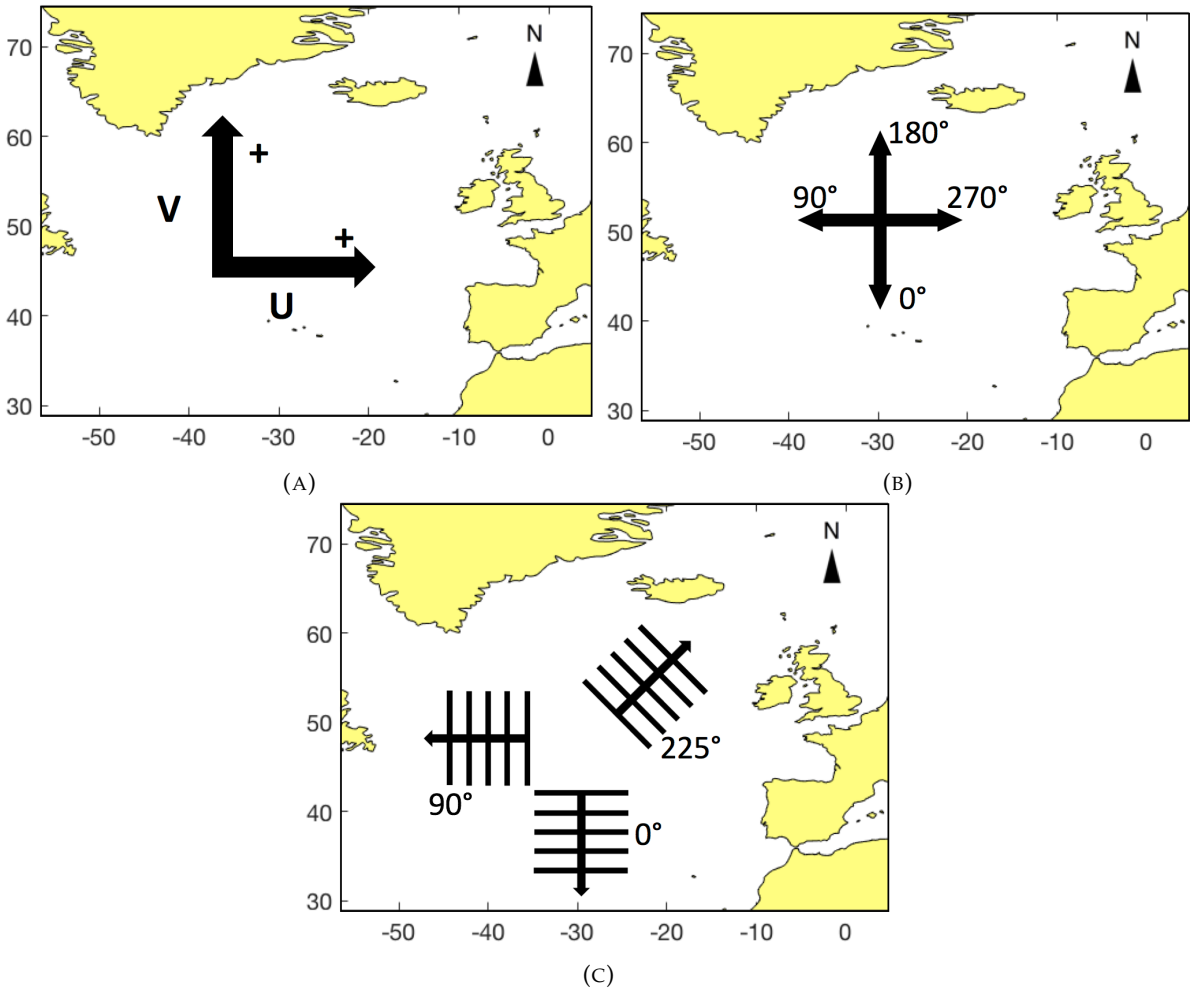


FIGURE 6.2: (A) Shows the sign convention for weather variables on component form.
(B) Shows the true wind direction convention.
(C) Shows the direction convention for waves propagating in the indicated direction.

6.2. Weather data expressed on component form has a U component that is positive for eastward motion and a V component that is positive for northward motion. This way of representing the wind can be transformed to be expressed as a speed along with a direction. In this case the direction is to be denoted 0 [°] when the wind is propagating southward ('coming from the north'), and 90 [°] when propagating westward ('coming from the east'). The direction is also to fall within the range $[0, 360]$. This is illustrated in 6.2b. The wave direction convention is defined the same as for wind is shown in Figure 6.2c.

All data that is expressed in the absolute reference frame can be transformed to the ship relative reference frame. When this is done we conform to the direction conventions illustrated in Figure 6.3. Conversions between the different reference frames can be done by formulae presented in ISO (2016, Annex E), or alternatively by a method described in 6.3.2.

6.1.4 Systematizing Data

Due to the fact that the data considered in this project stems from different sources, some work has to be done combine and systematize the data. The ship monitoring data was used as the basis for this. This data consists of samples with a sampling period of 15 minutes. To be able to consider the additional weather

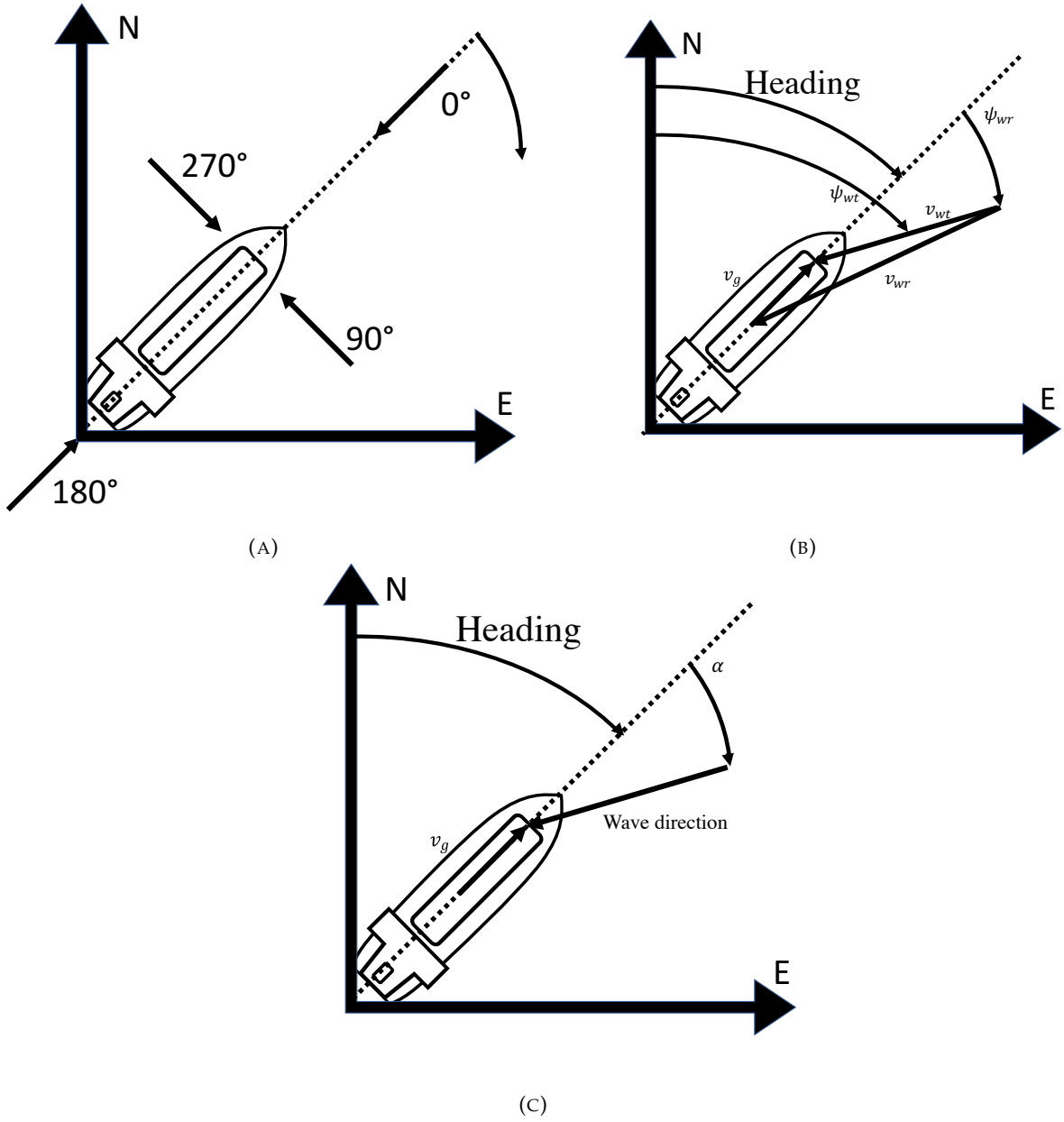


FIGURE 6.3: (A) Shows the sign conventions for directions relative to the heading ship heading.
(B) Shows the connection between the true and relative wind speed, the vessel speed and direction over ground, and the relative wind direction and heading. The figure is reproduced from ISO (2016). v_g is the speed over ground, v_{wr} is the relative wind speed, v_{wt} is the true wind speed, ψ_{wr} is the relative wind direction, and ψ_{wt} is the true wind direction.
(C) Shows the sign convention for wave direction

data we need measurements for each of the variables that we want to consider at the correct time. The data from ECMWF has a temporal resolution of six hours. In addition, these variables are only given at the grid points defined according to its spatial resolution. There are several ways to go about transforming the weather data into the reference frame of the ship monitoring data. One of the simplest is to perform linear interpolation. This interpolation will be three dimensional in nature, two spatial dimensions, and the temporal. Most of the data that is gathered about the weather data are mean values of some sort. That means that their effect should be slowly varying, and thus this linear interpolation should suffice. This assumption is confirmed by validation studies in 6.5 and by Mao et al. (2016).

6.2 Data Processing

In this section I will describe the processing applied to the available data before it is used as training data. The methods are based on the concepts described in Chapter 5.

6.2.1 Data Selection

Because the data has continuous measurements, and not only data from when it is in sea passage, a substantial amount of the data has to be disregarded. Only the samples that are logged while a ship is recorded as being in transit will be used. There are a few reasons for this. For instance, it makes no sense to discuss speed loss when the ship is at berth or anchored. We can also not expect the weather data, such as significant wave height, to be correct when the ship is maneuvering within a protected harbour. All measurements that is not stated to have the operational state 'Sea Passage' are thus removed. How the measurements are distributed over the ships and the operational sea states is displayed in Figure 6.4. We see that Sea Passage is the most commonly occurring operational state.

6.2.2 Missing Values

In 5.3, missing values and methods for handling them was discussed. The data sets used in this thesis contains significant proportions of missing values. This is the case for data from both the onboard measurement system, and the weather hindcast data. Due to the number of available measurements, the a choice has been made to exclude all samples with missing data in the relevant variables. What is meant by relevant variables will become apparent in succeeding chapters, but essentially it refers to all variables that is used in the final model, directly or indirectly. Indirect use refers to variables that is used to compute some other variable that is used directly. For example, the significant wave height will be used directly, while the mean wave direction is used indirectly, as it is used to calculate the relative mean wave direction. Missing data in either of these variables will result in the sample being excluded. This filtering will reduce the number of available measurements, but is necessary to build a reliable model. The size of the remaining data sets is illustrated Figure 6.5. The data that is left after the removal of missing values, is what will be used in to develop the models.

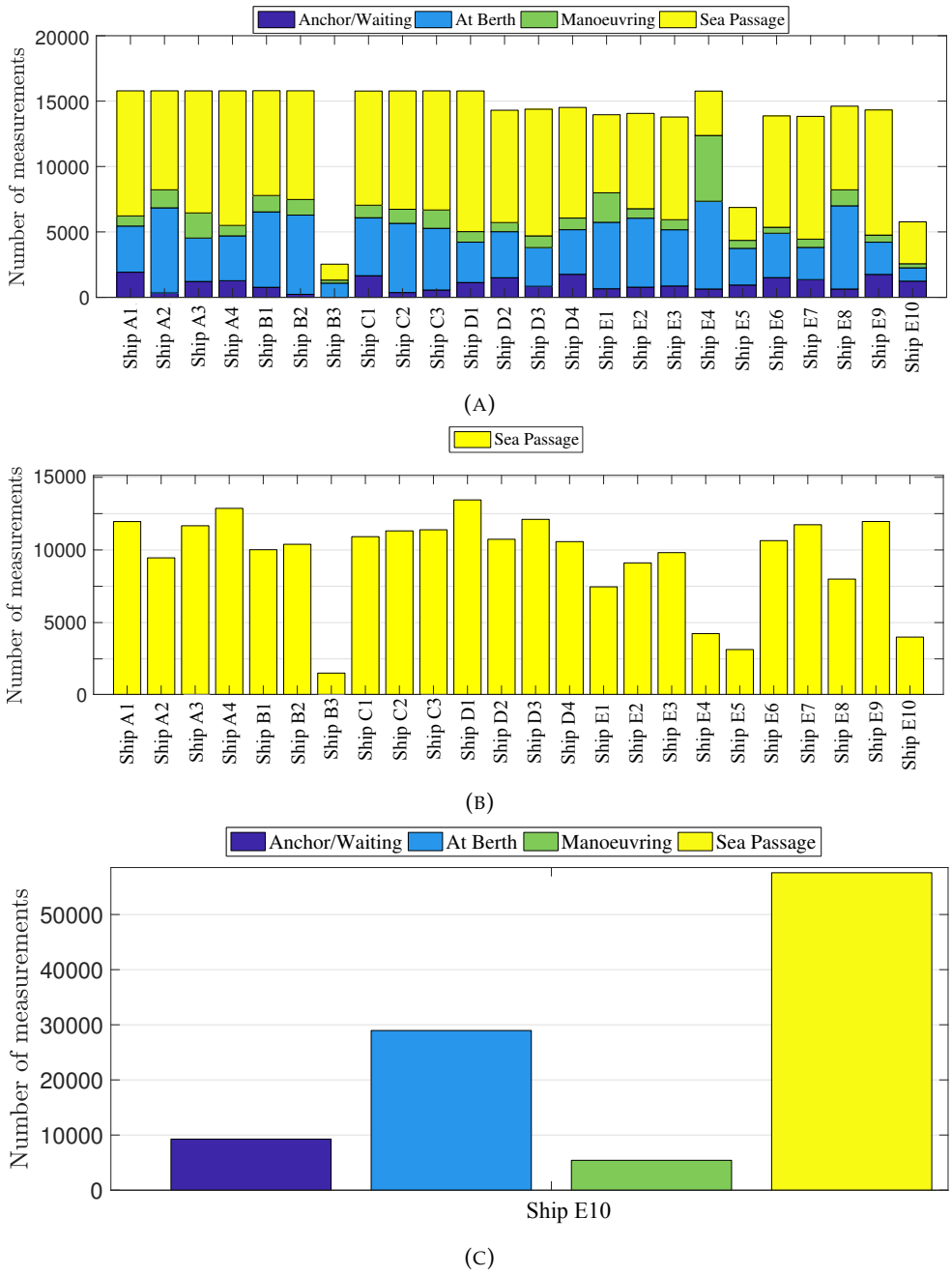


FIGURE 6.4: (A) Shows the number of measurements distributed over the ships and operational states in the fleet data set.
(B) Shows the number of measurements distributed over the ships for the operational state Sea Passage in the fleet data set.
(C) Shows the number of measurements distributed over the operational states for the ss-data set.

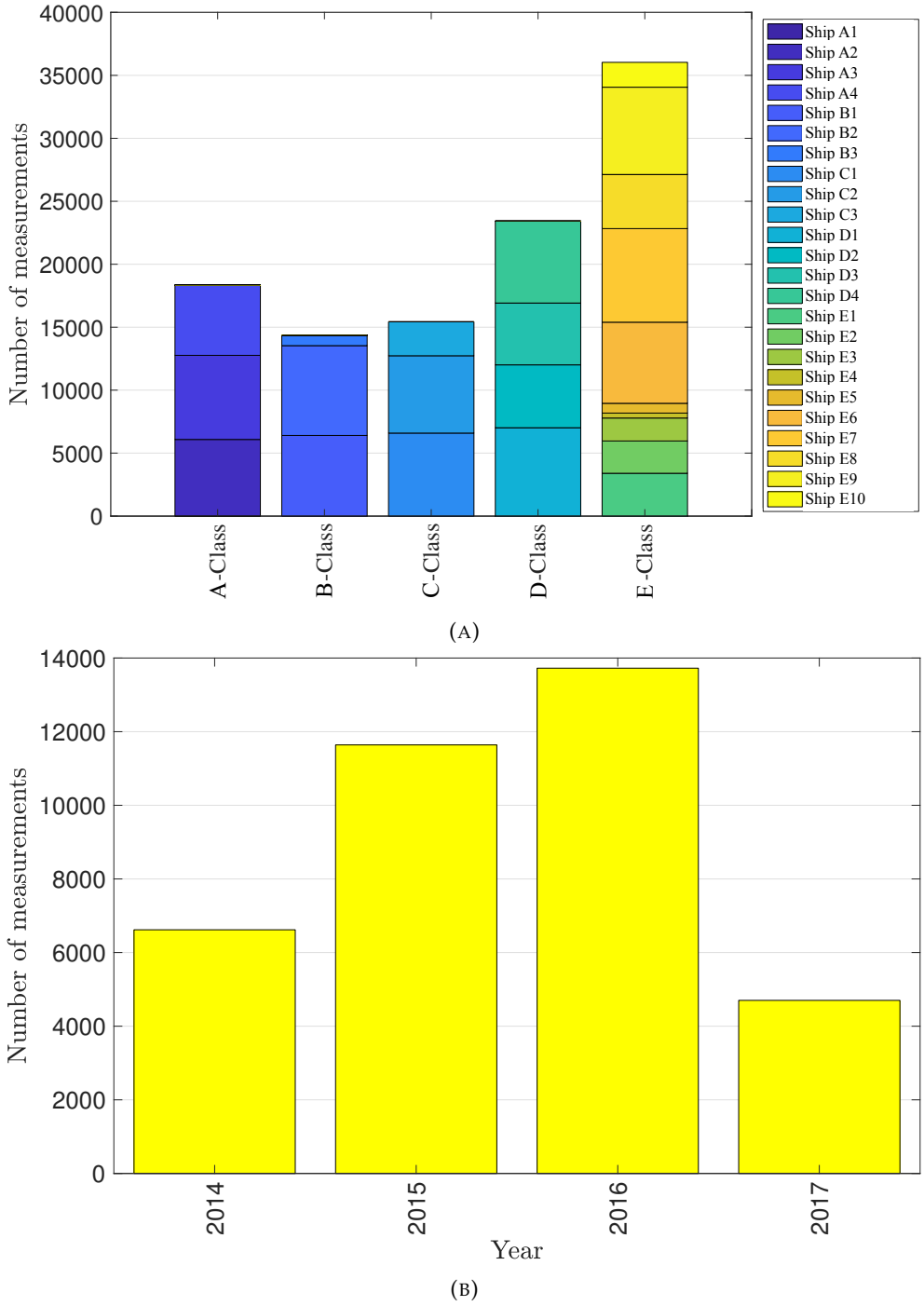


FIGURE 6.5: (A) The distribution of measurements over the ships and classes after removing samples with missing values from the fleet data set.
(B) The distribution of measurements after removing the samples with missing values in the ss-data set.

6.2.3 Outliers

The strategy for decreasing the influence of outliers has been to inspect the data sets by means of graphical methods. By looking at histograms and variable correlations some problems was discovered. The resulting actions is summarized in 6.2.5.

6.2.4 Normalization

The normalization of the data has been performed by use of the formula in Equation (5.2). The results is reversed such that the results can be seen in its original form.

6.2.5 Filtering

The following summarizes the filtering that was done on the data for various reasons.

- The measurements of relative wind speed had too many suspiciously low values for some ships. This is probably the result of a faulty wind sensor. The measurements with a relative wind speed lower than 0.3 [m/s] was thus removed for ships with a disproportionate amount of low relative wind measurements.
- In the ss-data set, measurements with a large discrepancy between calculated[3] and measured heading was removed. No samples with an error bigger than 10 [°] was used.
- The operational state of a vessel seems to be changed manually. We can therefore expect the switch between the different states display signs of human error, such as wrong or delayed switching of operational state. This problem was handled by only setting some minimal speed through water. this solves the problem under the assumption that all measurements with a speed higher than 10 knots, is in sea passage.
- For the data in the fleet data set, several ships had extreme trim measurements. This resulted in the implementation of max values for the trim being implemented. This trim was limited to ± 3 [m]. This issue was far less prevalent in the ss-data set.

6.3 Data Transformation

Machine learning algorithms for regression are more efficient the more facilitative the data set is. This means that a data set where all the variables in some sense exist in the same reference frame will yield better results, and be more efficient than a data set that is not. When the training data spans over different reference frames the ANN has to learn the relationship between the reference frames as well as learning the actual relationships that we are interested in. Therefore it is beneficial to transform the data such that it conforms to a reference frame. This is possible since we have prior knowledge of the reference frames of the data.

In this investigation, we will consider data relative to the ship. This is the hydrodynamic reference frame that follows the vessel speed and heading. Then environmental loads and effects will be expressed relative to this reference frame.

[3] See 6.3.1 and 6.5.2.

Some of the relevant weather variables are given in a decomposed form. Typically northward and eastward components. Therefore a method for transforming such data to the relative reference frame is shown in 6.3.2.

6.3.1 Heading

Although the ss-data set contains two measurements for heading (see Table 6.2), the fleet data set does not. For transformations in this section we need data on the ships orientation. Therefore we need to estimate this. We will assume that the difference between heading and course is negligible, and therefore, in the remainder of this thesis, we will refer to the calculated value as heading. The heading is calculated as follows.

$$\psi_i = \operatorname{atan2}\left(\sin\left(\Delta\lambda_i\right)\cos\left(\phi_{i+1}\right), \cos\left(\phi_i\right)\sin\left(\phi_{i+1}\right) - \sin\left(\phi_i\right)\cos\left(\phi_{i+1}\right)\cos\left(\Delta\lambda\right)\right), \tag{6.1}$$

where $\operatorname{atan2}$ is the arctangent with two arguments, $\Delta\lambda_i = \lambda_{i+1} - \lambda_i$ is the difference in longitude, and ϕ_i and λ_i is the latitude and longitude at sample i.

6.3.2 Transforming data

By considering the ship speed v_g, heading ψ, and some decomposed phenomenon, e.g. wind, we can calculate the relative direction and speed of the phenomenon. First we decompose the ship speed from v_g and ψ. This yields the decomposed northward and eastward ship speed components. By subtraction we get the relative components of the phenomenon. From this we can calculate the relative speed with

$$\nu = \sqrt{u_{rel}^2 + v_{rel}^2}. \tag{6.2}$$

To calculate the relative direction we consider the relative speeds and the heading. We calculate the direction first in a absolute sense and then subtract the heading and correct such that the angle $\alpha \in [0, 360]$ is the direction of the phenomenon. The idea is illustrated in Figure 6.6.

This transformation will be applied to all relevant features. Additionally it is possible to use this transformation to decompose the phenomenon to a component normal to, and along the center line. This may be used to accentuate the fact that an effect may be most important when considered only through the a component parallel to the center line of the vessel.

6.3.3 Wind

We have wind data from two independent sources. The ship monitoring data, as well as the ECMWF data, contains wind measurements. The wind measurements from the ships constitute relative wind speed and direction. The ECMWF data presents the wind decomposed as northward and eastward wind components. We can bring these wind components to the relative reference frame that the ship measurements are in by use of the method in 6.3.2. That means that we have a relative wind speed and relative wind direction. The wind as an input can either be given as these two quantities, or in a combination in the sense of using the relative speed and relative direction to calculate the component that is along the center line. Both are valid approaches as we saw in 2.4.2.

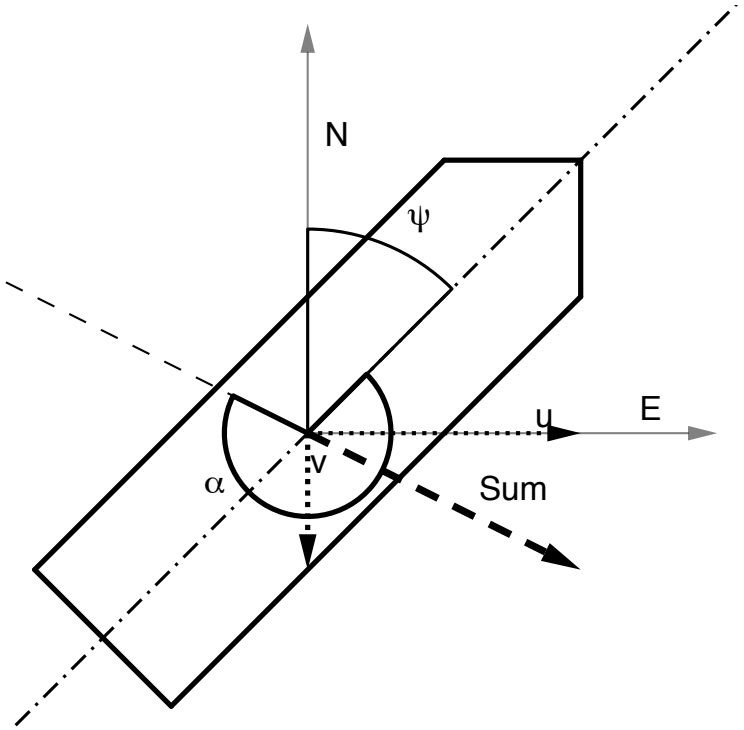


FIGURE 6.6: Illustration of relative direction.

Since we have two sets of measurements, we have to decide which one will best describe the way the wind will affect the vessel, and what is most useful in the context of the method that is developed. This will be addressed in 6.4.2.

6.3.4 Wave Direction

As for wind we assume that the direction of the waves is important. Therefore we want the network to understand where the waves are coming from relative to the ship. This is done by calculation the relative angle that the mean wave direction is coming from. The relative angle is initially presented as a relative angle between 0-360 [°], but this will be mapped to 0-180 [°] as it is assumed the effect of waves is symmetric.

6.4 Feature Selection

The performance of the model that is produced by a neural network is determined by how well the input data describe the relationship between the input and the target variable. As previously mentioned, careful consideration should be taken when evaluating what features should go into the ANN. We should use features that we expect impact the added resistance effect. From the literature we have that several of the available features impact the effect.

6.4.1 Waves

In 2.4.2 we see that the literature concludes that waves is an integral part of added resistance. Among the features that are available, several encompass some information about the wave condition. The significant wave height should be important as several sources conclude that the wave height in particular is a factor in the added resistance. As stated previously, added resistance is proportional to the square of the wave height. When it comes to added resistance due to ship motion, in particular pitch motion has a dominating effect. The frequency of the waves that interact with the ship is important as there exist a relationship between the wave frequency and wave length. The ratio of the wave length compared to ship length is suggests whether the waves will be mostly reflected or result in ship motion. The impact of ship motion is significantly more important to the added resistance compared to wave reflection. The wave direction may also help capture rolling effects or the effects of varying encounter frequency, and that is why it is included in other speed loss methods, such as in Kwon (2008). It is quite natural to include as much information about the wave condition as possible. Therefore, significant wave height, relative mean wave direction and mean wave period should be included.

6.4.2 Wind

It is intuitively clear that wind will be a part of the resistance picture, as it is in the sense of air resistance in a calm water resistance situation. As stated in 2.4.2 the relative wind speed and direction will affect the added resistance of a vessel. Among the features that are available to the neural network, both ship measurements of the wind and calculated wind speed and direction measurements are available. On one hand the ship measurements are probably a better representation of the wind condition at the the time of the measurements, and might therefore result in better performance, but it is more probable that what is available when using the developed method is weather data such as from the ECMWF. It is thus more valuable to concentrate on a method that can use a weather source such as the ECMWF.

6.4.3 Trim and Draft

A big of part ship resistance is the viscous resistance. In this resistance component, the area of the wetted surface is important. The wetted surface will change between voyages and during voyages. Between voyages, the cause is the change in loading condition. Depending on the geometry of the vessel, increased loading will increase the draft and subsequently the wetted surface. This affects the ship resistance. There are two ways the trim of the vessel will affect the ship resistance. The trim will constantly change as it is the result of pitch motion. This effect should be captured by the wave input. However, the potential steady trim offset can significantly change the calm water resistance. As previously mentioned most hulls are optimized with respect to some design condition and deviation due to improper cargo distribution may impact the vessel resistance. By including data on trim and mean draft the network may be able to learn the consequences of these effects and improve the predictions.

6.5 Data Validation

To get an idea of the validity of the data, one can investigate if the data behaves as expected, and if expected relationships are present. This does not replace a thorough investigation into the precision and

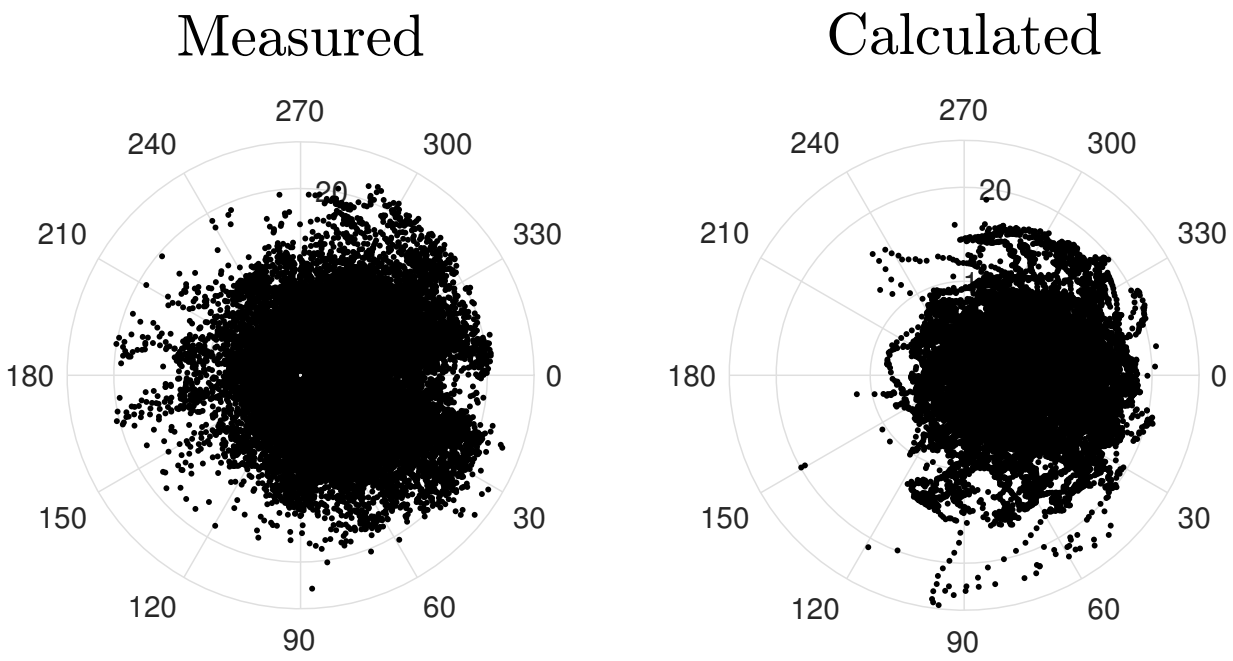


FIGURE 6.7: Polar scatter plot that shows the different measurements of relative wind speed [m/s] and direction [°] from the ss-data set.

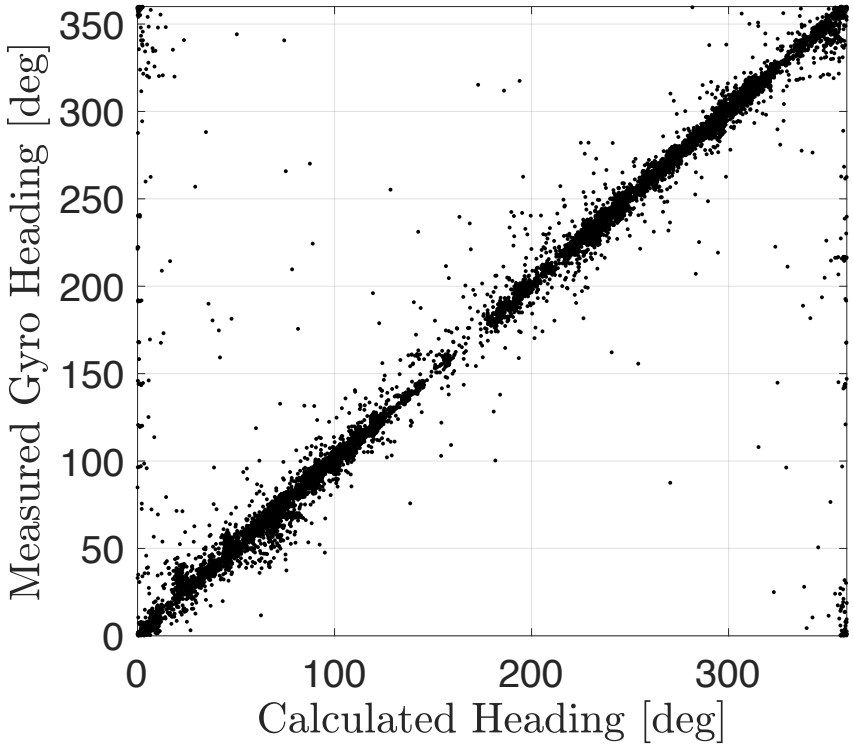
accuracy of the data, but is considered an adequate measure for the scope of this thesis. In this section I will present a few tests that look into some expected relationships.

6.5.1 Wind

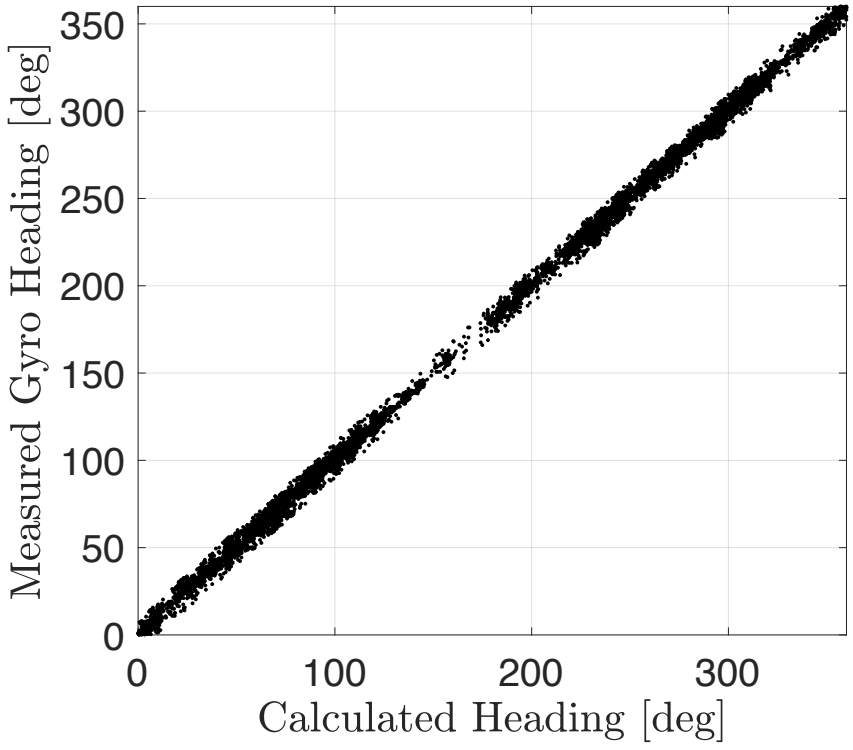
We have two different sources of wind measurements, and even though we have argued for using climate data as our source of data from the ECMWF it is still important to see how they compare. To see if they are similar we can bring the ECMWF data to the hydrodynamic reference frame and compare. In Figure 6.7 we see that the ship measurements are more randomly spread compared to the values that are calculated. This is because the calculated values are based on mean values. Wind gusts and local variations of wind is not accounted for. More relative wind speeds occur from behind in the measured data, which is probably due local variatios in relative wind.

6.5.2 Heading

In 6.3.1 we showed how the heading data is calculated when it is not available. To evaluate the assumption made here, as well as the accuracy of the resulting measurements, validation is needed. To do this we can compare the calculated headings with the available measurements. In Figure 6.8 we see that the calculated headings is close to the measured headings. This means that we can use the calculated heading with confidence. The figure depicts the data before and after the filtering described in 6.2.5.



(A)



(B)

FIGURE 6.8: Scatter plot of measured gyro heading and calculated heading, before (A) and after (B) filtering.

6.6 Comparing the Data Sets

In the following chapter I will present the method that has been developed. Throughout the process of developing the model it has become apparent that the ss-data set is the easiest to work with. It is only natural that there exist less unexplained variance when modeling a single ship than when we are modeling a class of ships. The ss-data set has also proven to have more stable measurements for draft and trim, which is essential for the first part of the modelling. Additionally, there are more available samples from the ss-data set than there are from any of the single class data sets. The aforementioned points is why the ss-data set is used for the development of the model.

7 Method Development

This chapter will describe the the process of developing the speed loss estimation method. Some context is established before the development of the modules of the method described. The first module relates to how the speed loss is estimated. The second module relates to how the method learns to predict the speed loss. Lastly some, testing procedures are presented.

7.1 Motivation for Method

Before explaining the method that has been developed, it is important to discuss context. The problem of speed loss may be investigated in several by several approaches. As exemplified in 2, it can be examined from a purely hydrodynamical point of view. In this context, added resistance calculation is done by theoretically motivated methods. Often, this consideration of added resistance is then treated as a resistance component similarly to any other resistance component. It has been mentioned that an approach for calculating speed loss following the theoretical approach is very demanding, which increases the motivation for an alternative approach.

The opposing approach is the data-driven approach. An example of an empirical method for speed loss is Kwon's approximate formula which was mentioned in 2.5. The strength of this formula is how easy it is to implement. It does however, have some notable drawbacks. For high Beaufort numbers, above 6, the method is unlikely to be accurate because the detailed calculations on which the method is based becomes inaccurate. Moreover, voluntary speed reduction and propeller racing, which the method does not account for, can be expected at high Beaufort numbers.

A method for predicting speed loss based on common operational data, that can handle challenging weather conditions seems beneficial. Where Kwon's formula aims to predict the speed loss based on detailed calculations for various types of ships and conditions, the method developed in this thesis uses operational data for the relevant class of ship or specific ship. If enough relevant data is available, this method aims to give accurate speed loss predictions when given sufficient operational data describing the state of the ship as well as the weather. A fitting application for such a method of speed loss prediction is where some ship owner, for some ship or class of ships, have operational data and want to use this data for speed loss prediction. This is a realistic application for many ship owners. As have been previously mentioned, most ship owners log operational data for performance analysis. Further continuous logging will also make additional data available and can increasingly make the prediction of speed loss better:

7.2 General Approach

In a case where data is available, such as the data presented in Chapter 6, it should, by the machine learning methods previously presented, be possible to predict speed loss. Before detailing the full approach a brief summary is presented:

1. Collect ship performance monitoring data.
2. Collect relevant weather data.
3. From the collected data, create a calm water data set.
4. Develop a calm water speed regression model.
5. Calculate speed loss for the historical data based on regression model.
6. Train a neural network regression model on the historical data to predict speed loss.

The steps are also visualized as a flowchart in Figure 7.1. The background for this step by step method comes in part from how speed loss is defined in Townsin and Kwon (1983) and Kwon (2008), and in part from the two-part consideration of total resistance in ITTC (2011). The goal of the neural network that is trained is to predict the speed loss defined similarly to how the speed loss is defined in (2.6). However, in the developed method, the percentage speed loss contains all effects, so that there are no correction factors. In ITTC (2011), a recommended approach to predicting and investigating the total resistance is to treat the calm water resistance and added resistance separately.

$$R_T = R_{cw} + R_{added} \tag{7.1}$$

With this as the background, the method in this thesis is developed to first calculate some reference speed from a calm water consideration, and then predict the difference between this and the actual speed through water.

7.3 Calm Water Reference Speed

A key element for producing good predictions of the speed loss, is the calculation of V_{cw}. It forms the basis on which the neural network is trained. This means that if the calculated V_{cw} is erroneous, there is no value in the speed loss estimation. It may however be difficult to pinpoint exactly how V_{cw} should be defined and calculated. It is supposed to represent the speed of the ship if it was completely unaffected by weather conditions.

During the development of the method, several approaches have been considered for the calculation of V_{cw}. Empirical methods, model tests and data-driven regression models are all viable options. One method for calculating calm water resistance is using one of the various empirical methods that are available. Data from model tests may be used to develop regression models based on main dimensions. Two examples are Holtrop and Mennen (1982) and Hollenbach (1998). In Kim et al. (2016), Lu et al. (2015), and Prpić-Oršić and Faltinsen (2012) Holtrop and Mennen's method is used for calculating calm water values, which indicate that it could have been valid approach.

Another method for calm water resistance prediction is the use of model tests and Froude scaling to calculate the resistance and required power. In this thesis, the model test data for the D-class and E-class ships have been available, such that it could have been possible use these tests as the reference calm water speed.

The methods for calculating the calm water reference speed that has been mentioned up to this point may have been able to give a relatively reasonable estimate for V_{cw},that could have been used for speed loss calculations. However, considering the arguments that follow, a different method have been chose. The

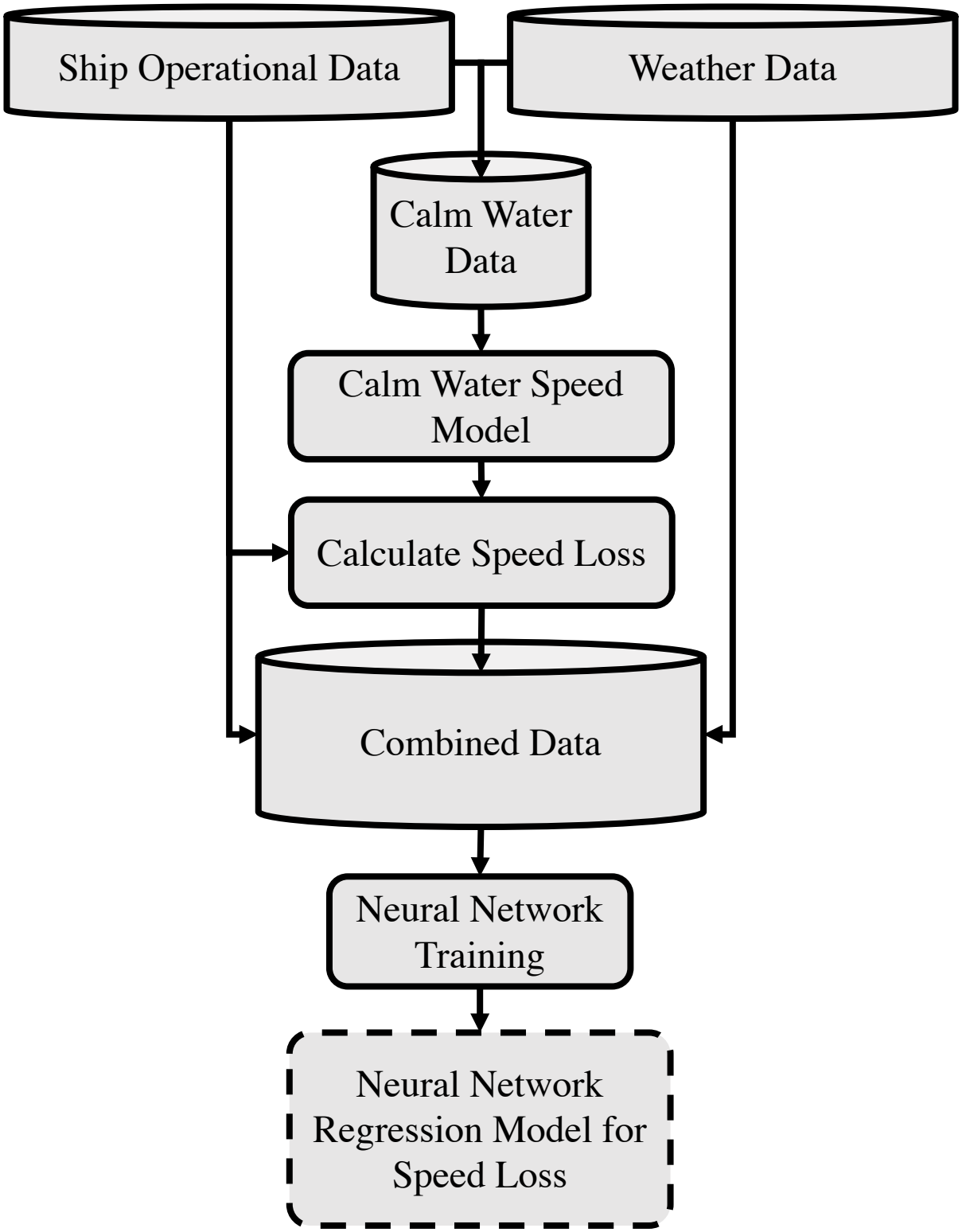


FIGURE 7.1: Flowchart outlining the method that has been developed.

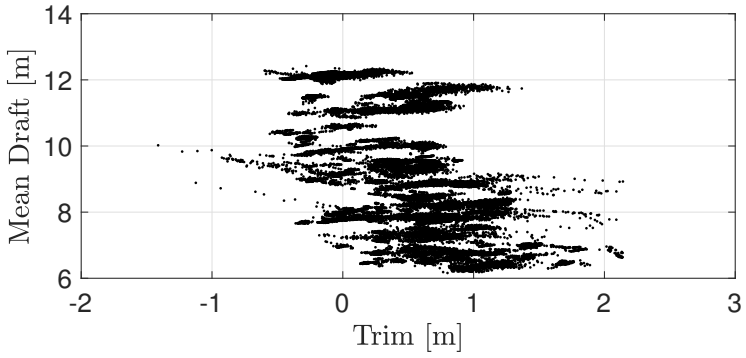


FIGURE 7.2: Scatter plot of mean draft and trim for all sea passage samples collected from Ship E10.

empirical methods that are available are limited in that they are only valid for cases that are similar to the data used to develop the regression models. An reason for why this can be a problem is that the regression models are based on a data base of model tests, and model tests are typically only performed for two or three loading conditions. A typical model test can consist of tests for the design draft condition, ballast condition and maybe some intermediate condition. These loading conditions have some mean draft and some trim, but as we see in Figure 7.2, the trim can vary significantly for a given mean draft. And since the trim can have a considerable impact on the calm water resistance due to the change in underwater geometry, model tests and the empirical regression models seem insufficient.

As the alternative to the aforementioned methods, we can instead use the data that is available. If we consider the ship speed as a function of a set of variables as seen in (7.2).

$$V = f\left(P, T, Trim, \alpha, Wind, H_S, T_w\right) \tag{7.2}$$

where P is power, T is the draft, $Trim$ is the trim, α is the relative weather direction, $Wind$ is a measure of the wind speed, H_S is the significant wave height and T_w is the wave period. From the dataset we can filter out samples with significant weather, and create a model

$$V_{cw} = f_{cw}\left(P, T, Trim, Wind \approx 0, H_S \approx 0\right) \tag{7.3}$$

which in practice will be

$$V_{cw} = f_{cw}\left(P, T, Trim, Wind < Wind_{max}, H_S < H_{S,max}\right). \tag{7.4}$$

Varying the condition that describes the calm water is possible, and the choice should be guided by the data that is available. A max value for the significant wave height set to $H_{S,max} = 0.5$ [m] has been used for most of the thesis. The impact of significant wave height is important to the added resistance as established by the literature (see Subsection 2.4.2). The strong correlation between speed loss and significant wave height is confirmed in Subsection 8.1.3 which further supports the choice that is made.

The strength of this method is that it will capture all the effects, such as fouling, that is not treated in the aforementioned methods since it is a data-driven method. It is ship specific, or hull specific depending on what data is used to develop the method, which is likely to produce useful results for the calm water speed. A challenge with this method is that there needs to be an adequate number of data points that

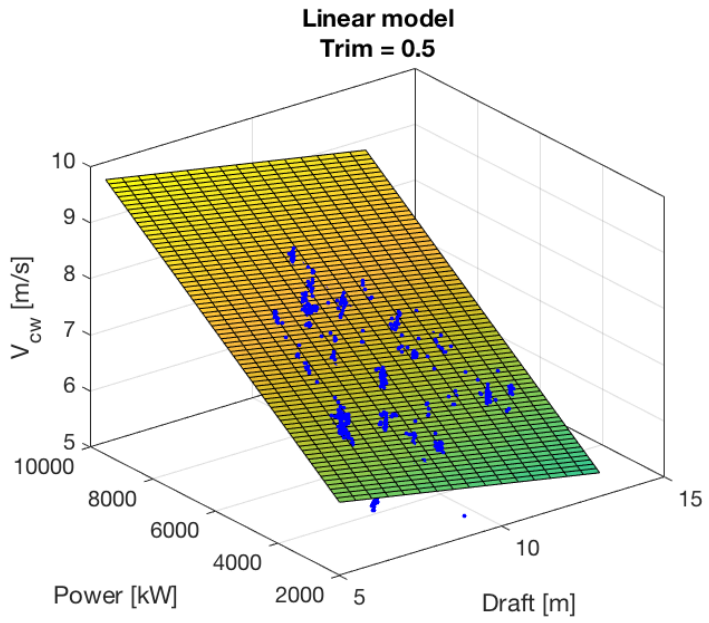


FIGURE 7.3: Illustration of a linear regression calm water model for $Draft = [2, 13]$ [m], $Power = [2000, 9500]$ [kW] and $Trim = 0.5$ [m], along with the calm water ($H_s < 0.5$)[m] training data for Ship E10.

satisfies the condition for 'calm water'. One also has to monitor this model and update it as the relationship between the input variables may change over time. This change may come from fouling of the hull and propeller, or due to maintenance work. There are many ways to develop such a model, and in the following subsections we will explore some methods.

7.3.1 Linear Regression Approach

For any problem of this type it is possible to try linear regression. In many cases the resulting model is too simple, but it is always a useful benchmark. We restate the problem in a regression context by proposing that we can develop a model

$$y = h_{cw,\mathbf{w}}(\mathbf{x}), \tag{7.5}$$

where y is the calm water speed, and $\mathbf{x}$ is the vector containing trim, draft and power. In Table 7.1 the performance of a linear regression model is expressed through the mean squared error and the R^2. To further illustrate the resulting model we can plot the surface that is the resulting model. This is seen in Figure 7.3. It is important to note that this visualization of the model has its flaws. The model is plotted for a range of values in draft and power, but only a single trim value. Additionally, it is inherently problematic to portray a four dimensional model in two dimensional, but I have chosen to do it to illustrate a point that will become clear later in the chapter.

We see that the model is linear from Figure 7.3, and since we know that the relationships we want to model are nonlinear it makes sense to try more complex models.

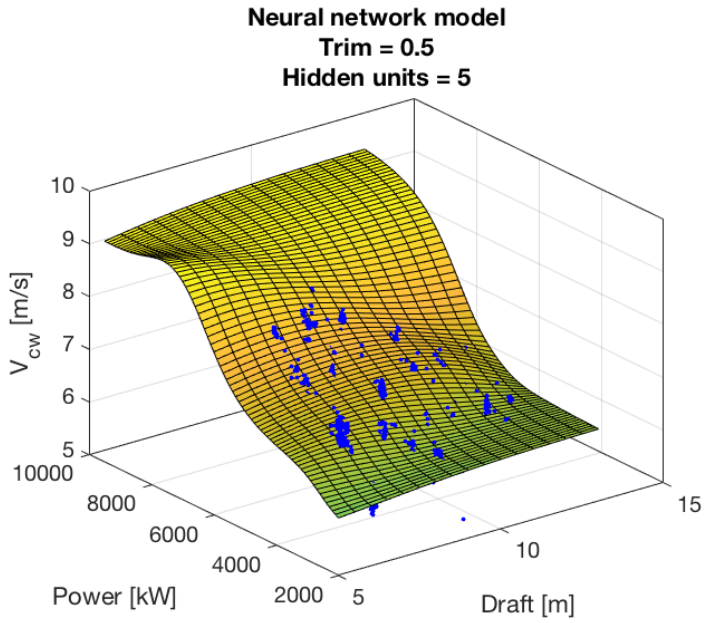


FIGURE 7.4: Illustration of a neural network regression calm water model for $Draft = [2, 13]$ [m], $Power = [2000, 9500]$ [kW] and $Trim = 0.5$ [m], along with the calm water ($H_s < 0.5$)[m] training data for Ship E10. This model has five hidden units.

7.3.2 Neural Network Regression Approach

Another approach that is natural given the use of neural networks in this thesis, is to create a neural network regression model. To evaluate this possibility two networks was trained, with five and 32 hidden units respectively. The networks was trained with the Bayesian regularization algorithm and made use of early stopping to prevent overfitting. It becomes apparent from Table 7.1 that the neural network models perform better than the linear model due to lower mean squared errors and higher coefficient of determination R^2.

The illustrations of the neural network regression models shown in Figures 7.4 and 7.5 show the nonlinear nature of the models. It does however become apparent that we may get some unphysical behaviour from the model. For instance in Figure 7.4 we observe some bumps that seem unphysical. We also see, for low power, that increasing the draft will increase the speed, which seems unlikely. There seems to be room for improvement in the calm water reference model.

7.3.3 Custom Nonlinear Regression Approach

A custom nonlinear regression model allows the model designer to specify a model more specifically. This type of regression modelling can be useful in cases where the model designer has some prior knowledge of how the system that is being modeled. A custom nonlinear model may thus be a reasonable approach for the modelling of the calm water resistance. To develop the custom nonlinear regression model we need to examine the relationships between the input variables and the calm water speed. We look at the power first.

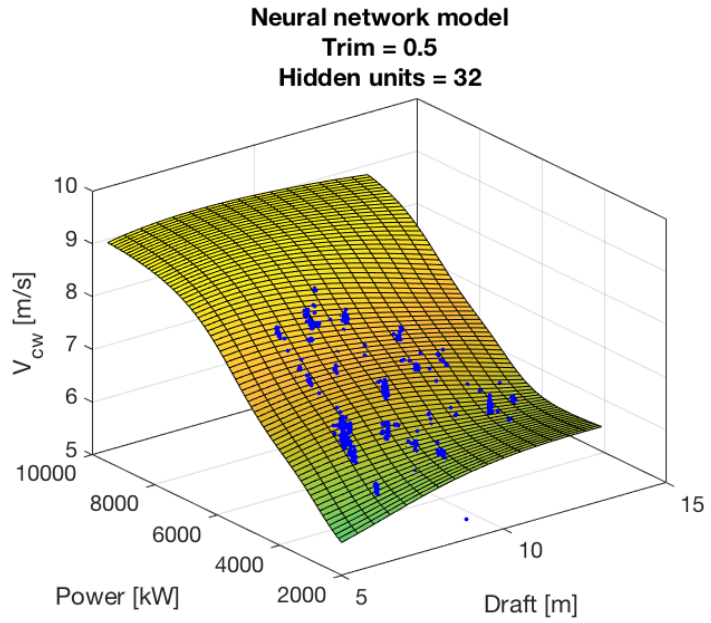


FIGURE 7.5: Illustration of a linear regression calm water model for $Draft = [2, 13]$ [m], $Power = [2000, 9500]$ [kW] and $Trim = 0.5$ [m], along with the calm water ($H_s < 0.5$)[m] training data for Ship E10. This model has 32 hidden units.

Power

The data that comes from the data set is denoted shaft power, but motivated by the linear relationship between the power quantities shown in (2.4) we will operate with a single power P in this model because the difference will be captured by the parameters in the model. In 2.5 the proportionality between ship speed and resistance $R \propto V^2$ is mentioned. This relationship is also shown in the way the resistance coefficient is defined (Harvald, 1983). We say that

$$R = C\frac{1}{2}\rho V^2 S \propto \left(V^2, S\right), \tag{7.6}$$

where C is the resistance coefficient, ρ is the density, V is the speed and S is the wetted surface area. If we multiply by the ship speed, we have that $P \propto V^3$ which suggests $V \propto P^{1/3}$. It seems reasonable to suggest a term of the form

$$w_i P^{1/3}, \tag{7.7}$$

where w_i is some numbered parameter in the model.

Draft

The draft is essential for predicting the calm water speed, and is thus an important part of the model. We can investigate the relationship between the draft by looking again to Equation (7.6). If we multiply by the ship speed we have that $P \propto SV$ which suggests that $V \propto \frac{S}{V}$. We now need to investigate the relationship between the wetted surface area and the draft. From Table B.1 we can plot Figure 7.6. Although the there is a nonlinear relationship between the draft and the surface area, due to the complex geometry of a ship hull, we see that for the drafts in question it is viable to assume linearity. We can thus say that, since

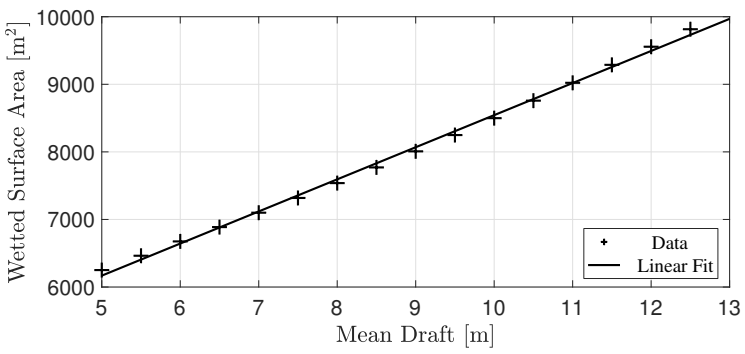


FIGURE 7.6: Illustration of the relationship between the mean draft and the wetted surface area of Ship E10.

$S \propto T$, we have $V \propto \frac{T}{P}$. This suggests that a term in our model on the form

$$w_i \frac{T}{P}, \tag{7.8}$$

is favorable.

Trim

The last variable in the custom nonlinear calm water model is the trim. How trim affects the speed can be very difficult to predict. Even the slightest change in trim may lead to significant change in the flow over the hull, and thus change the ship speed. This sensitivity is due to how the underwater geometry changes. A bulbous bow especially changes this flow field. In an instance where the bulbous bow is fully submerged, but due to a change in trim, breaks the surface it is obvious that the resistance will change. To investigate the influence of trim, ship owners may carry out trim optimization tests. In Figure 7.7 we see a figure from a trim optimization test carried out on the B-class hull. Although this is a result for a single hull geometry, it is assumed that we can use this behaviour as a guideline for the regression model. The figure suggests a relationship between speed and trim that can be described by a third order polynomial. Additionally, we see that the effect of trim becomes smaller for increasing displacement. This suggests that we can have a cubic polynomial term that is scaled by the reciprocal of the displacement. We know that $\Delta = \rho\nabla$ and from Figure 7.8 we can extract that $\nabla \propto T$. We can then scale the polynomial by the draft instead of the displacement, and suggest terms as seen in

$$\frac{w_i}{T}\left(w_{i+1}Trim^3 + w_{i+2}Trim^2 + w_{i+3}Trim\right) + w_{i+4}, \tag{7.9}$$

which simplifies in the parameters w_i to

$$\frac{1}{T}\left(w_i Trim^3 + w_{i+1}Trim^2 + w_{i+2}Trim\right) + w_{i+3}. \tag{7.10}$$

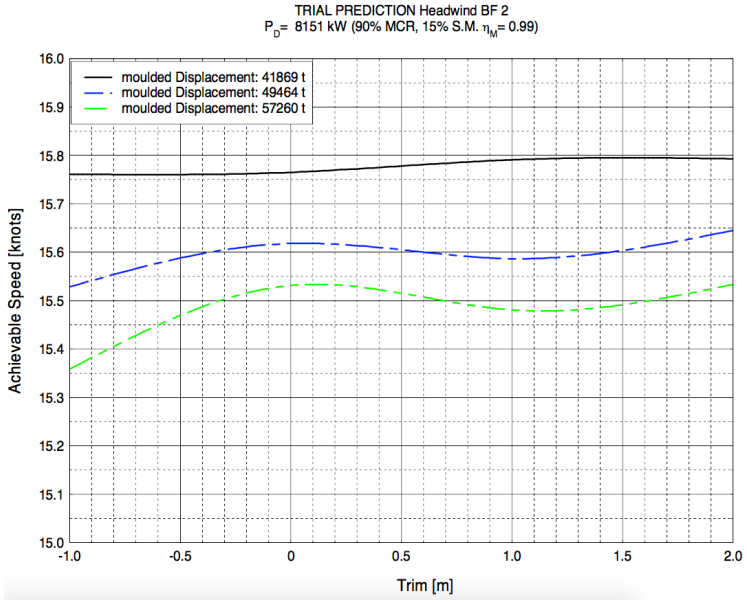


FIGURE 7.7: Result from trim optimization of the the B-class hull showing the relationship between achievable speed and trim for three loading conditions.

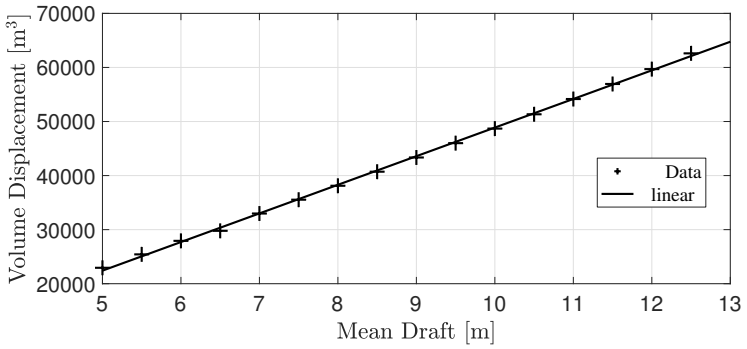


FIGURE 7.8: Illustration of the relationship between the mean draft and the volume displacement of Ship E10.

Final Model

Based on the terms suggested in (7.7), (7.8) and (7.10), we can now formulate the final custom nonlinear regression hypothesis function as

$$h_{cw,\mathbf{w}}\left(P,T,Trim\right) = w_0 + w_1 P^{1/3} + w_2 \frac{T}{P} + \frac{1}{T}\left(w_3 Trim^3 + w_4 Trim^2 + w_5 Trim\right). \tag{7.11}$$

Now that the model is established we need to find the parameters $\mathbf{w}$. The problem is essentially the same as what is discussed in Chapters 3 and Chapter 4. We need to minimize some cost function over the parameters in $\mathbf{w}$ too ensure a good fit. If we use a squared error cost function similar on the form

$$J(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{N}\left(h_{\mathbf{w}}(\mathbf{x}^{(i)}) - y^{(i)}\right)^2, \tag{7.12}$$

we can use the Levenberg-Marquardt algorithm. This is implemented in MATLAB, and can be done relatively simply. An illustration of the trained model is seen in Figure 7.9. The behaviour that lower power and higher draft results in lower speed is present. There is also no unphysical behaviour like what was seen in the neural network models portrayed in Figures 7.4 and 7.5.

If we compare the performance metrics of the nonlinear model, as seen in Table 7.1, we see that the nonlinear model lands somewhere between the neural network models and the linear model when it comes to performance. This may suggest that the neural network models may be better suited for the modeling of the calm water speed. However, there are some points about the nonlinear model that results in the final choice of model is the nonlinear custom regression model.

An essential aspect of the model is how it is generalizing, and there are a few details that suggest that the nonlinear model is the most generalized model. Firstly, if we look at the values presented in Table 7.1, we can see that the nonlinear model performs very similar on the test set compared to the training set. This is an indication that the model may be suffering from high bias. This concept is shown in Figure 3.4. This bias-like behaviour comes from the fact that the model is fairly specific. So we have a trade-off between the models ability to conform to the training data, and a model that is based on the physical properties. The strength of the nonlinear model is that it seems to be more valid outside of the training data. The unphysical behaviour that is observed in Figures 7.4 and 7.5 would probably be mended by more data. The conclusion when it comes to the calm water model is that both a neural network approach and a custom nonlinear regression approach is viable, depending on the data available.

In a later chapter we will see that it may be beneficial to modify the model slightly to increase model quality. The modification increases the model flexibility, and is recommended for applications when sufficient calm water data is available.

7.4 Calculate Speed Loss

When the regression model for the calm water reference speed loss is developed, it is used to calculate the speed loss that has occurred. This is done according to

$$\frac{\Delta V}{V_{cw}}100\% = \frac{V_{cw} - V}{V_{cw}}100\%, \tag{7.13}$$

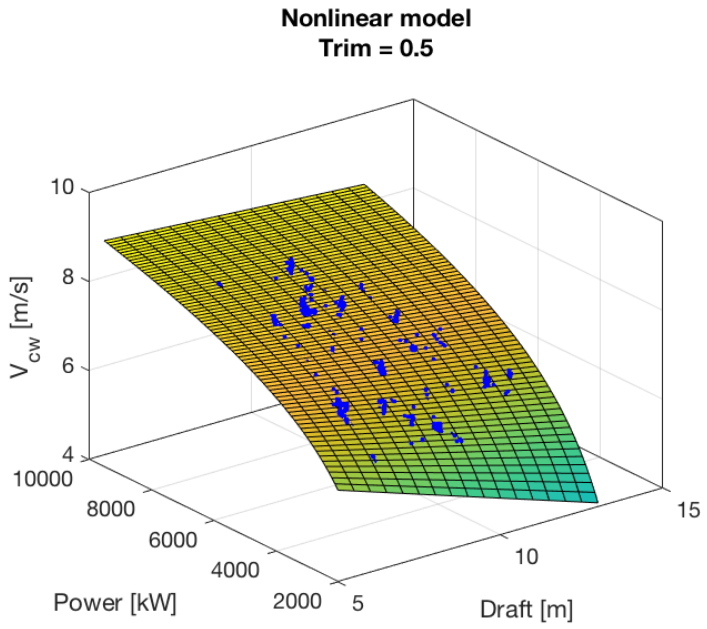


FIGURE 7.9: Illustration of the nonlinear regression calm water model for $Draft = [2, 13]$ [m], $Power = [2000, 9500]$ [kW] and $Trim = 0.5$ [m], along with the calm water ($H_s < 0.5$)[m] training data for Ship E10.

TABLE 7.1: Performance of regression models on calm water ($H_s < 0.5$)[m] Ship E10 data.

Model	Data set	MSE	R^2
Linear model	Training	0.074	0.857
	Test	0.081	0.850
Neural network model	Training	0.047	0.910
5 hidden units	Test	0.053	0.901
Neural network model	Training	0.048	0.907
32 hidden units	Test	0.055	0.898
Nonlinear model	Training	0.070	0.865
	Test	0.068	0.872

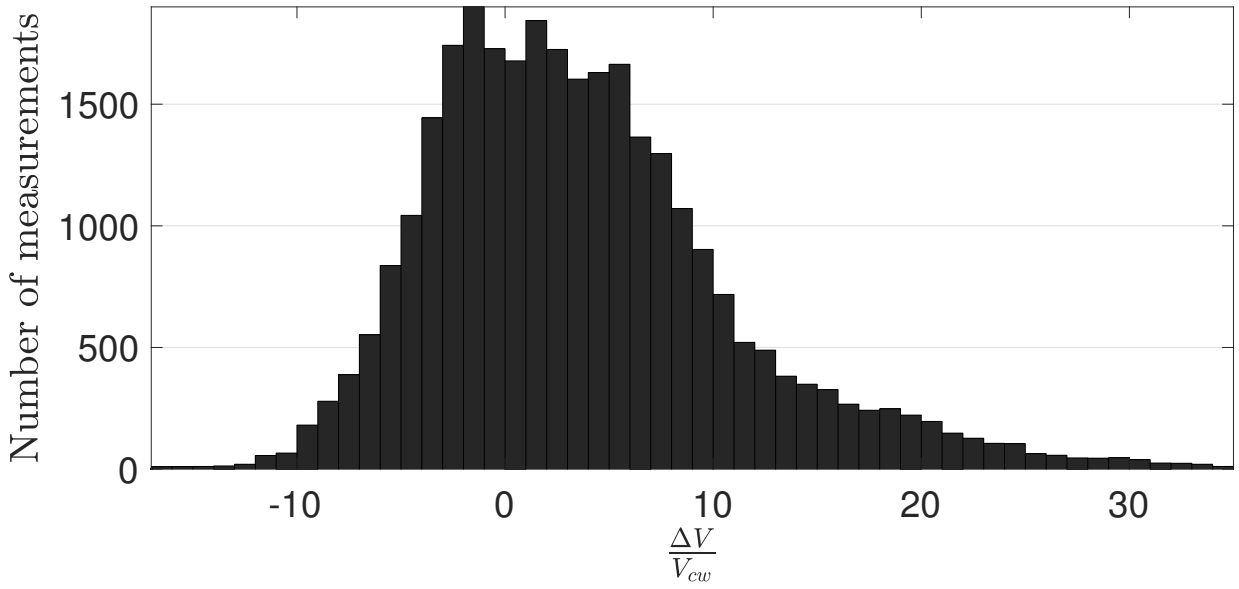


FIGURE 7.10: An example of the distribution of speed loss from the ss-data set. The mean of the metric in the distribution is 2.65%.

where V_{cw} is the calm water reference speed, V is the actual measured speed and thus ΔV is the difference as it is defined in (7.13). We note that the quantity $\Delta V > 0$ when $V_{cw} > V$. In other words if a reduction of speed occurs, the quantity ΔV is positive.

A result of using actual data to define the calm water reference speed is that the model used will average out some effects. A consequence is that the calculated value for speed loss will be negative more often than one might expect from other methods, indicating a speed increase, more than one would expect. This is illustrated in Figure 7.10. After the speed loss has be calculated it is included in the data used for training and validation.

7.5 Main Network

After the calculation of speed loss is done, all variables that is to be used as regression inputs is available. As previously mentioned, there are numerous combinations of network parameters and input variables that give different predictions of speed loss. In this section I will show how the different network configurations and input variable combinations behave

7.5.1 Network Configuration

In Chapter 4, the different aspects of neural networks was discussed. It is well established that we have to consider how we implement the neural network with great care. It was suggested in 4.7 that a reasonable approach for training a network that generalizes well, is to train a large network while making use of early stopping to prevent overfitting. In this thesis, many network input combinations, with data sets of different sizes will be tested, and to avoid having to manually optimize regularization and/or network complexity for each individual test, an approach based on early stopping is used. To validate that the early stopping is an efficient method for controlling the neural network model complexity an investigation of this is performed.

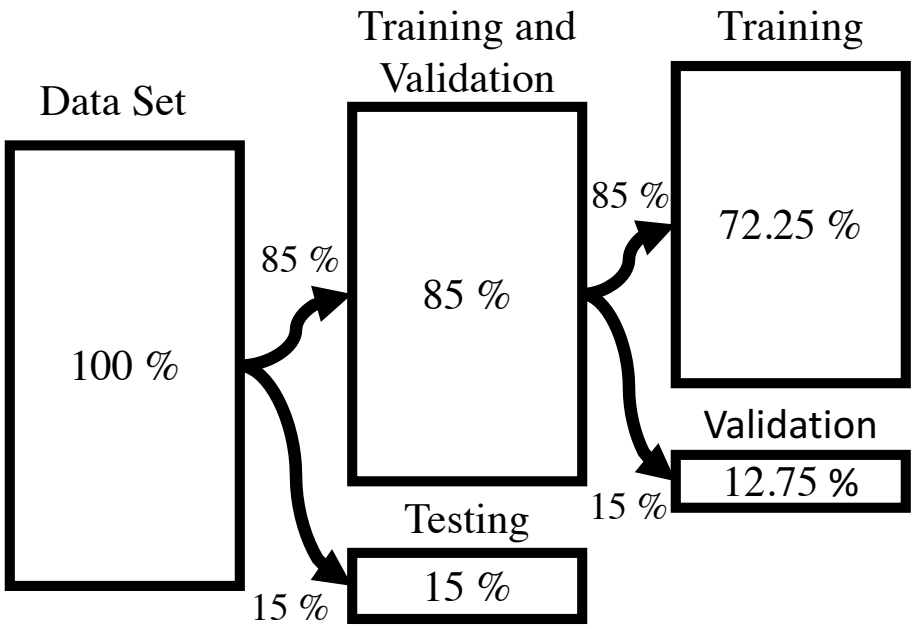


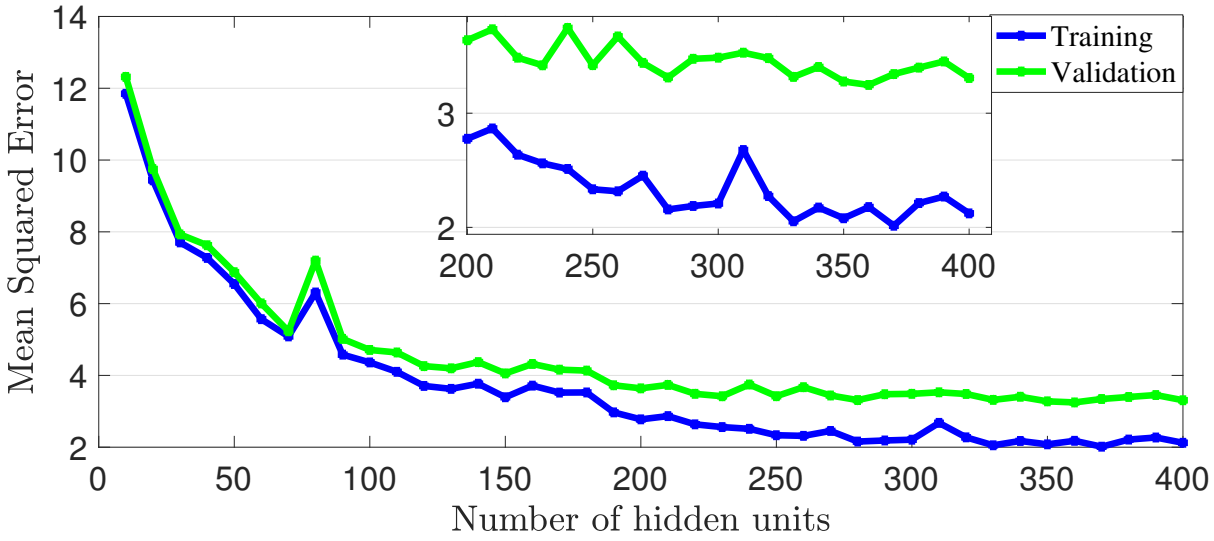
FIGURE 7.11: Illustration of the scheme for the splitting of data.

To test that early stopping works, we can train several networks with an increasing number of hidden units. In the absence of early stopping, we expect the prediction error of the network to behave as seen in Figure 3.5, namely that the training error decreases with complexity, while the validation error increases after some complexity level. However, with early stopping implemented, the preferred result is that this increase is removed. This changes why there is a need to limit complexity in the network structure. Without early stopping, this limit is to prevent overfitting, now the limit is, at least in the case of this thesis, based on computing power.

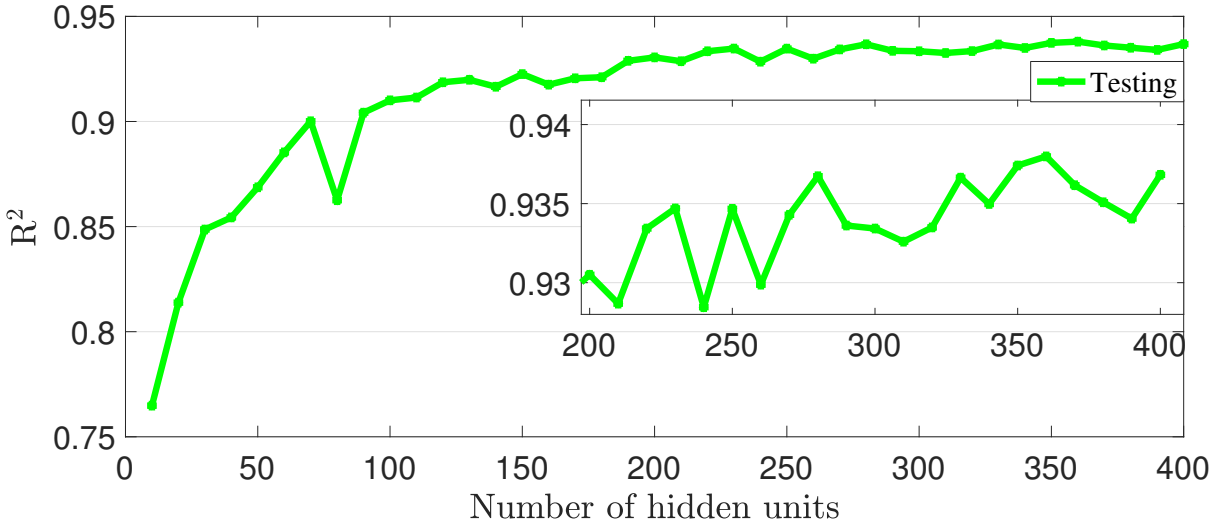
From Figure 7.12 we see that the implementation of early stopping is working. In this example we have trained a network to predict speed loss based on data in the ss-data set. We use 85% of the data for training and validation and 15% for testing. 15% of the 85% is used for early stopping validation. The resulting fractions of data is illustrated in Figure 7.11. The implementation of early stopping seems to work well based on the fact that we do not see the behaviour from Figure 3.5. The networks trained with early stopping improve as the network become more complex, but as the complexity increases, the benefit of increased complexity is marginalized. A conclusion is that the number of units should be based on computing power and available time.

7.5.2 Input Variable Form

In Chapter 6 we discussed the different input variables, and how the should put into the network. In this section we shall investigate the difference in performance when it comes to how the different variables is given. This is important to ensure good performance and stability. There is a balance between giving the network data that we know is related, and letting the network learn the relations from the data. This aspect of the network design is most relevant for wind. To test how the different ways of giving information differs we shall train a number of networks on data presented in different ways and compare the resulting performance. Ideally there will be a noticeable difference in performance that can be used to draw a conclusion and make a recommendation.



(A) Mean squared error as a function of the number of hidden units.



(B) R^2 as a function of the number of hidden units.

FIGURE 7.12: Learning curves depicting the two performance metrics as a function of network complexity. The network was trained on the ss-data set.

When it comes to wind there are two ways to let the network learn the effect of wind in its speed loss prediction, either as one wind component parallel to the ship, or as the wind direction and wind speed relative to the ship. A test was done to determine what format resulted in the best performance. By training two networks on the same data presented in different ways we can see how varying the format influences the performance. The result from this test is presented in Table 7.2. The result indicates that to introduce wind as the relative direction and relative speed is marginally better. Both the mean squared error and R^2 has a better mean value and a lower standard deviation in this case. We can therefore conclude from the test that it is recommended to have two input variables for wind.

TABLE 7.2: Performance of networks trained with wind data presented in different formats. The networks had 125 hidden units and was trained on the ss-data set ten times.

Wind data form	$\overline{MSE}$*	$\sigma(MSE)$**	$\overline{R^2}$*	$\sigma(R^2)$**
Relative direction and relative speed	4.0	0.14	0.93	0.003
Head wind speed component	4.1	0.16	0.92	0.003

*The bar denotes mean values. **$\sigma(\cdot)$ denotes the standard deviation.

7.5.3 Input Combinations

In a simulation setting, there can be a varying availability of the different input variables. An overview of what performance we can expect from the different input combinations is therefore useful. To gain knowledge on this topic we can perform an analysis by testing a number of combinations. In Table 7.3 an overview of the tested inputs and the included variables are given. The different inputs are motivated by different simulation scenarios.

- Input 1 is for a setting where all weather parameters are known, as well as the loading condition of the vessel.
- Input 2 is similar to Input 1, except that we input the speed through water, effectively predicting the attainable speed.
- Input 3 is similar to Input 1 and Input 2 except it is more closely related to the relationship between the engine speed and the speed loss.
- Input 4 is similar to Input 1, Input 2, and Input 3, except it is more closely related to the relationship between the propulsive power and the speed loss.
- Input 5 is a simpler input combination to predict attainable speed given a simpler characterization of the sea state without any notion of direction.
- Input 6 is similar to the input of Kwon's method except it has the speed as an input variable.
- Input 7 has similar input variables as Kwon's method.

To evaluate the different input combinations we will perform a test similar to how we considered the input form of wind. Ten networks will be trained on each input combination to generate a basis for comparison. The networks are trained with 125 hidden units and utilize early stopping to prevent overfitting. The results of these tests are presented in Section 8.2

TABLE 7.3: Sets of possible inputs.

Variable	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6	Input 7
Draft	x	x	x	x	x	x	x
Trim	x	x	x	x			
Shaft Power				x			
Shaft RPM			x				
Log Speed		x			x	x	
Significant wave height	x	x	x	x	x		
Mean wave period	x	x	x	x			
Relative wave direction	x	x	x	x		x	x
Relative wind speed	x	x	x	x			
Relative wind direction	x	x	x	x			
Beaufort value					x	x	x

7.6 Testing

The last part of developing the method is to find useful validation tests. We will distinguish between two areas of testing. The first relates to the fist part of the model, where the reference speed V_{cw} is calculated. It is essential that the calm water speed estimates are reasonable if they are used to calculate speed loss. A few ways to evaluate the estimates are as follows:

- Compare the estimated speed loss to what is predicted by Kwon's method
- Compare the estimated speed loss to the speed loss that is predicted when V_{cw} is determined by using available model tests.
- Investigate if the speed loss correlates as expected with the input variables to the main network.

The second part that has to be tested is how well the network is able to learn the relationship between the input variables and the estimated speed loss. This is done continuously by the use of splitting the data set into training data and test data as mentioned in 7.5.1. As a benchmark for comparison we can see how well a linear regression model is able to perform the same task.

7.6.1 Kwon's Method

Kwon's method is a method for predicting speed loss based on Beaufort value, weather direction and loading condition. We should see similar tendencies in the results of Kwon's method as in the method developed in this thesis. To investigate this we filter out data that is is fitting for Kwon's method. To be applicable for this method there is a few things that must be done. The weather direction must be defined and binned into the four categories defined in Kwon (2008). The relative wave direction is used to do this binning. After the weather direction is defined we have to extract data that is definitely in ballast or fully loaded condition. Drafts between 5.5 m and 7.0 m is used to define ballast condition, and drafts above 11 m is used to define fully loaded conditions. In his paper Kwon notes that his method is unlikely to be accurate for Beaufort values above 6, so data that conflicts with this requirement are left out. The distribution of speed losses that occur when the losses are calculated according to Kwon is presented in Figure 8.1.

7.6.2 Model Test

The result of the model test used for performance prediction for the E-class is available. In these results we find the relationship between power and speed for three loading conditions. To calculate the calm water reference speed we can make a linear interpolation model with draft and power as input, and returns the calm water reference speed $V_{cw,m}$, where the subscript m denotes that it comes from the model test. This model will only be valid for data that falls within the domain where the model was tested. The resulting domain of validity can be specified as $6.45 < T < 12.00$, and $3677 < P < 12825$, where T is mean draft in meters, and P is shaft power in kW. From the model test we can also develop a linear interpolation model from calculations where a sea margin of 15% is added to the predicted necessary power.

With the model test as the basis of the estimation V_{cw} we can calculate the speed loss and compare the new speed loss figure to the speed loss calculated based on the nonlinear regression. The distribution of speed losses calculated by the two methods is presented in 8.3.

7.6.3 Inspecting Variable Correlation

In 6.4, potential input variables was evaluated to establish a theoretical foundation for what data should be used to predict the speed loss. The discussion was founded on theory established in Chapter 2. An indication that the model is estimating reasonable speed loss values is that we find expected correlations between the input variables that was not used in the calm water speed regression model. The expected behaviour can be summarized in short as:

- Significant wave height should correlate positively with speed loss.
- Relative wave direction should correlate negatively with speed loss when relative direction is defined from 0 to 180 as shown in Subfigure 6.3c.
- Mean wave period should correlate with positively with speed loss.
- Relative wind speed should correlate positively with speed loss.
- Relative wind direction should correlate negatively with speed loss when relative direction is defined from 0 to 180 as shown in Subfigure 6.3b.

7.6.4 Linear Regression Benchmark

Similarly to how the different methods for calculating V_{cw} was performed, by initially comparing the neural network model results with results from linear regression, we will use a linear regression model as a benchmark to evaluate how beneficial the added complexity of a neural network model is. The result of this testing is shown in Subsection 8.2.2.

7.6.5 Prediction Testing

To get an impression of the method in actual application setting we can perform the some of the following tests:

- Remove the data from one voyage of the ss-data set and train a network on the remaining data. Use the network to predict speed loss for this voyage and compare with the actual speed loss.

- Train a network based on the ss-data set and test the network a ship of the E-class from the fleet data set.

The aforementioned tests should illustrate a realistic application of the method.

8 Results and Discussion

The following chapter will present the results describing the quality of the developed method based on the test procedures outlined in 7.6. Relevant discussion on the results will be communicated along with the tables and figures that illustrate the results. Two main areas of performance is evaluated. First, the speed loss quantity from this thesis is compared to the speed loss obtained from other methods. The differences are illustrated in 8.1. The second area of testing pertain to how well the network is able to learn relationship between the speed loss that is estimated and the input variables given. This is illustrated in 8.2.

It should be noted that the results in this chapter relates primarily to the ss-data set unless something else is mentioned. Results for the fleet data set is for some tests reproduced in the appendix.

8.1 Modelling Calm Water Reference Speed

The first relevant result is the resulting speed loss distributions from the different methods for estimating the calm water reference speed. Subsequently, the variable correlations is shown.

8.1.1 Comparison with Kwon's Method

In Figure 8.1, the distribution of the speed loss quantity as estimated from the developed method and Kwon's method is plotted for both loading condition and ballast condition.

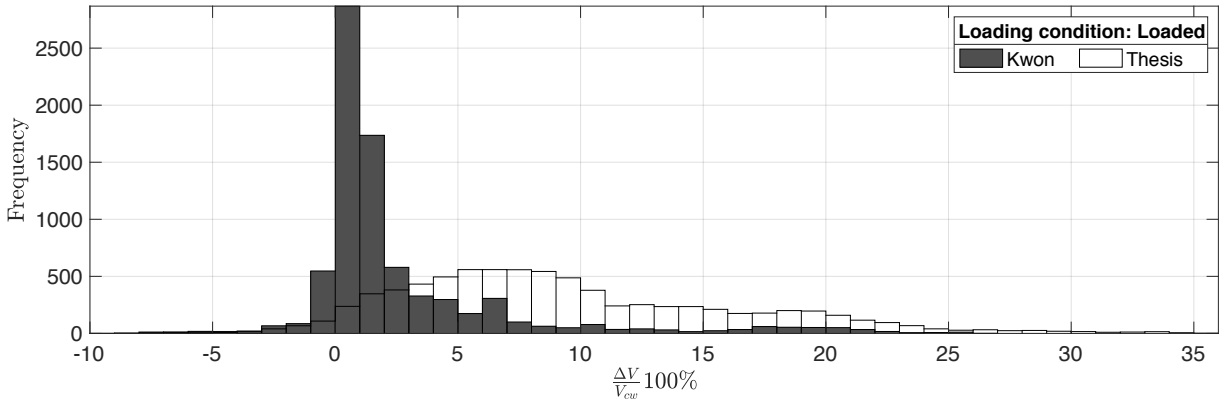
In a loading condition that can be described as fully loaded, the difference between Kwon's method and the custom nonlinear regression method is significant. Kwon's method produces estimates of the speed loss that is much more centered around the mean, and the mean value is also lower than that of the thesis method, as seen in Table 8.1.

In ballast condition, Kwon's method results only in positive speed loss, and a substantially higher mean and standard deviation than for the loaded condition. This change in behaviour from the loaded condition is opposite of what we see in the thesis method.

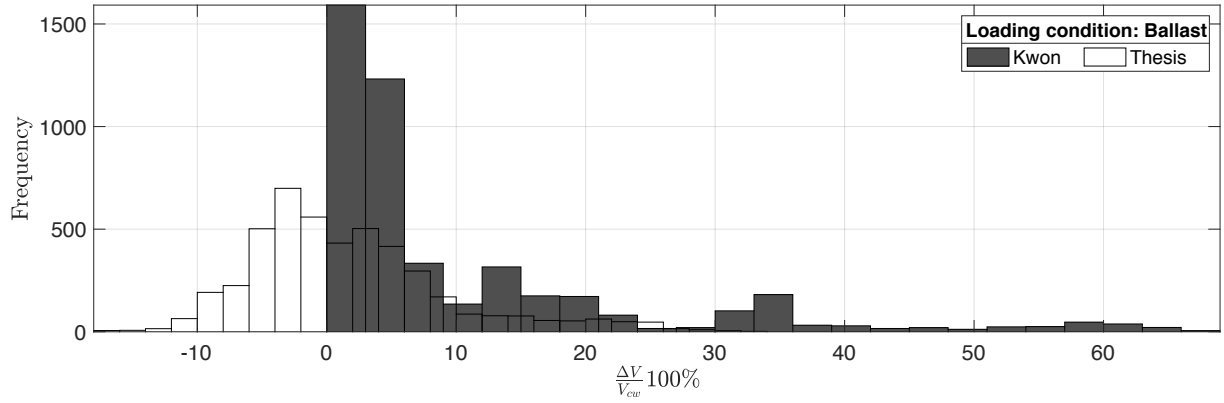
It is difficult to draw any direct conclusions from the comparison. What is seen from the comparison is that Kwon's method increase the estimated value of the speed loss metric when the draft is greater. In Kwon's

TABLE 8.1: Comparison of speed loss distributions from Kwon's method and thesis method.

Method	**Loading condition**	Average $\frac{\Delta V}{V_{cw}}100\%$	$\sigma\left(\frac{\Delta V}{V_{cw}}100\%\right)$
Kwon	Loaded	2.7	4.5
Thesis		9.5	6.8
Kwon	Ballast	10.6	13.6
Thesis		1.7	7.7



(A)



(B)

FIGURE 8.1: (A) Histogram comparing the distributions of the speed loss metric from Kwon's method and the method developed in thesis. The data is from a fully loaded loading condition.
(B) Histogram comparing the distributions of the speed loss metric from Kwon's method and the method developed in thesis. The data is from ballast condition.

TABLE 8.2: Speed loss distributions from the modified thesis method.

Method	**Loading condition**	Average $\frac{\Delta V}{V_{cw}}100\%$	$\sigma\left(\frac{\Delta V}{V_{cw}}100\%\right)$
Thesis	Loaded	6.1	7.0
Thesis	Ballast	1.7	7.5

method this stems from how the equations treat the displacement, and since the method stems from analytic considerations we may conclude that a ship in ballast condition is more vulnerable to weather effects.

The opposite effect in the thesis model does not come from a theory that a ship in ballast condition is less vulnerable to weather effects. Investigating this, it has been found that the nonlinear regression model lacks the flexibility to accurately handle calm water speed estimation for higher drafts. The reason for this stems from the assumption that V_{cw} is linearly proportional to the draft, as seen in Equation (7.8). This claim is supported by visual inspection of Subfigure 8.4a, where the distribution of speed loss appears to take on greater values for drafts between 11 and 13 meters. This may be mended by adding additional flexibility in the regression model term regarding draft. A term on the form

$$w_i \frac{T^{w_{i+1}}}{P} \tag{8.1}$$

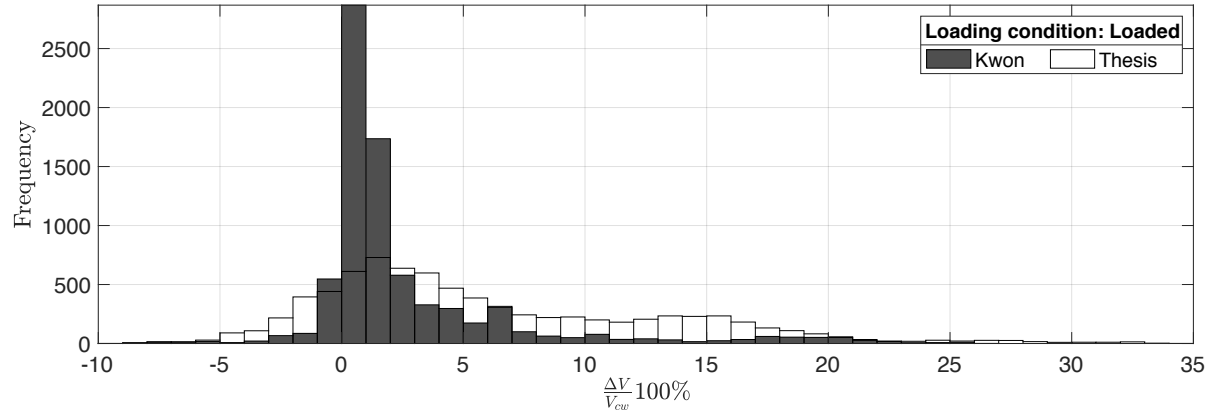
may be suggested. Such a term allows the estimated calm water speed to vary more freely with the draft term. In Figure 8.2 it becomes apparent that such a model estimates speed reductions closer to what Kwon's method does for higher drafts, and with a substantially lower mean compared to the unmodified model. The change in ballast condition, remains relatively small. That means that the added flexibility avoids the systematic overestimation of the calm water reference speed in loaded condition. When tested on the fleet data set, where there in general is less available calm water measurements it is seen that this added flexibility makes the model less robust. The conclusion is that more data of high quality allows for more flexibility, which is a general result in machine learning that is recognized from Chapter 3.

From the results we also see a larger standard deviation in loaded condition from the thesis model. This comes from the difference in how the models calculate the speed loss. Kwon's method takes some inputs and calculates the speed loss directly, while the thesis model takes some inputs and calculate V_{cw} which then is used together with the measured speed to calculate the speed loss. The thesis method ads another layer of potential variance by going about the estimation of speed loss in two steps. The measured speed may vary significantly which results in larger standard deviation.

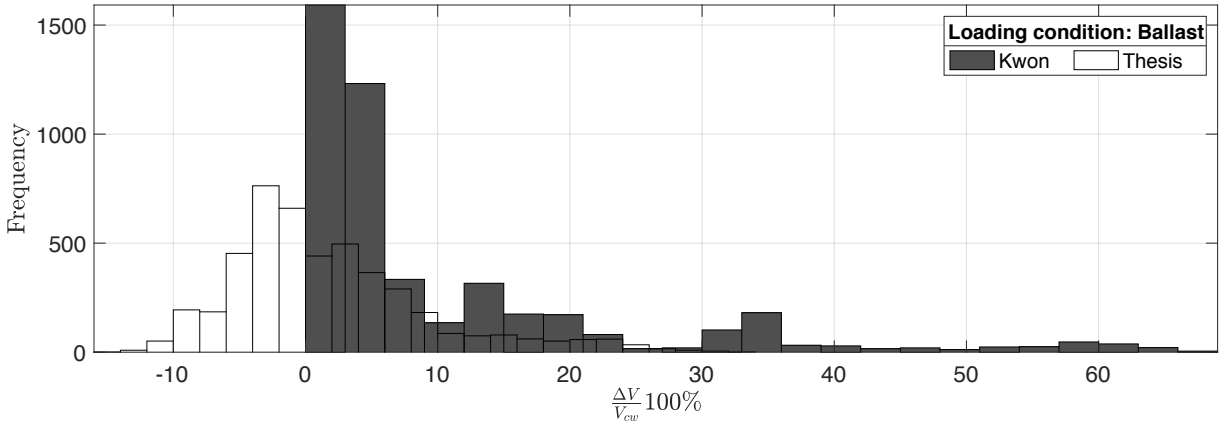
8.1.2 Comparison with Model Test Approach

Another way to calculate speed loss, is to first estimate the reference speed V_{cw} by use of the model test that is available. The resulting speed loss distributions is shown in Figure 8.3, together with the mean and standard deviations in Table 8.3.

In the case of no added sea margin we see a speed loss distribution with a notably lower standard deviation model test method. This is due to the lack of flexibility of the model in the higher loading conditions. It is difficult to draw important conclusions based on the comparison, but the speed loss distributions do look to carry some of the same properties indicated by the similarity in the form of the distributions.



(A)

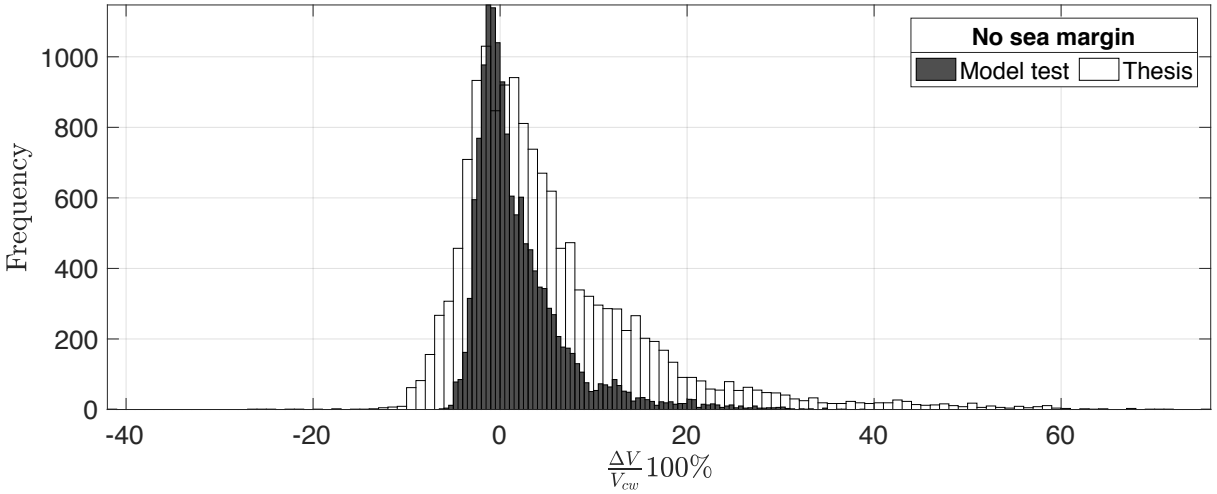


(B)

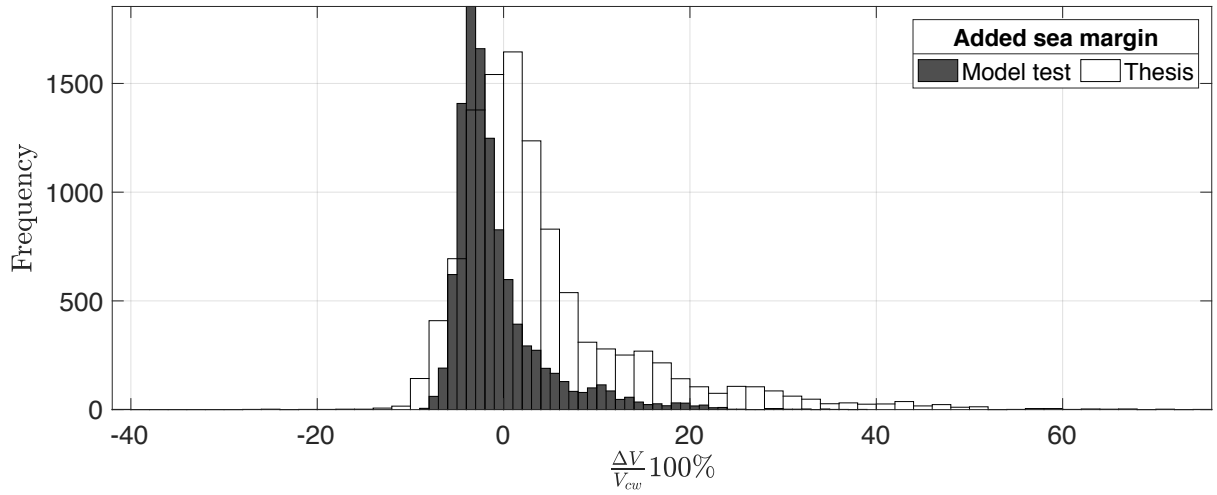
FIGURE 8.2: (A) Histogram comparing the distributions of the speed loss metric from Kwon's method and the modified method developed in thesis. The data is from a fully loaded loading condition.
(B) Histogram comparing the distributions of the speed loss metric from Kwon's method and the modified method developed in thesis. The data is from ballast condition.

TABLE 8.3: Comparison of speed loss distribution of the model test and thesis method.

Method	**Margin**	Average $\frac{\Delta V}{V_{cw}}100\%$	$\sigma\left(\frac{\Delta V}{V_{cw}}100\%\right)$
Model test	No margin	2.0	5.1
Thesis		5.3	10.4
Model test	Added margin	-0.8	5.0
Thesis		4.1	10.1



(A)



(B)

FIGURE 8.3: (A) Histogram comparing the distributions of the speed loss metric from the model test method with no added sea margin, and method developed in the thesis.
(B Histogram comparing the distributions of the speed loss metric from the model test method with added sea margin, and method developed in the thesis.

8.1.3 Correlation

The correlation between the input variables and the speed loss can indicate if the speed loss that is calculated is reasonable. In 7.6.3, a few expected correlations is mentioned. The correlations can be inspected visually by the scatter plots seen in Figures 8.4 and 8.5. The two figures represent the resulting speed losses calculated with and without the added flexibility in the draft term respectively. To represent the correlation numerically, we use a correlation coefficient[1].

As previously mentioned, there is a tendency for the draft to correlate with the speed loss due to the bias in higher drafts. The result is a positive correlation which is slightly mended when the modified model is used.

There is close to no correlation between the trim of the ship and the speed loss. The same is the case for the engine speed. The shaft power has a slight correlation with the speed loss, which makes sense because speed loss is connected to a power increase.

There is an evident structure to the relationship between the measured ship speed and the estimated speed loss. There is greater variation in the speed loss for low ship speed, than there is for the higher ship speeds, and there is an over all negative correlation. This behaviour follows directly from the calculation of speed loss. When the measured ship speed is low it is either because the ship deliberately moves at this speed, or because the ship suffers from speed loss. Thus, the corresponding speed loss may vary greatly. When the ship speed is high significant, speed loss cannot occur because that means that the ship wanted to have an even higher ship speed, which is not realistic. Negative speed loss at the higher ship speeds is also not to be expected because this means that the ship has a significantly higher speed than expected, which is very unlikely.

As expected, there is a notable positive correlation between the speed loss and the significant wave height. This is in accordance with the literature that claim a quadratic relationship between the added resistance in waves and wave height in head sea.

There seem to be weak correlation between the mean wave period and the speed loss. This is expected, as the wave period is correlated with the significant wave height. Additionally, we see that there seems to be a notably higher occurrence of large speed loss values for mean wave periods in the range 7-11 seconds. It might be tempting to conclude that this has something to do with the ship's natural period in pitch, which is important for added resistance (see section 2.4.2), and while this may be the case it may also simply be a consequence of the fact that most measured mean wave periods lie in this range (see Subfigure C.1g).

The relative wave direction seem to correlate as expected, with head sea correlating with more speed loss.

Relative wind speed is expected to correlate positively with the speed loss, and from Subfigure C.1i we see that this is the case. The relative wind measurement will affect the air resistance, and correlate with the roughness of the seastate.

Similarly to the expected behaviour from the relative mean wave direction, we expect a negative correlation between the relative wind direction and speed loss. This stems from the added resistance due to wind being more prominent when the relative wind comes from the front. The actual correlation seem to be consistent with the expectations. Additionally, we see that there is a larger spread in the speed loss values when the relative wind direction is closer to zero. This is probably because most measurements have this relative wind direction (see Subfigure C.1j).

[1]The correlation coefficient that is used is the Pearson correlation coefficient. A value of one means there is an absolute positive linear correlation, zero means no correlation, and negative one means an absolute negative linear correlation.

The last input variable that we will consider the correlation of, is the Beaufort value. This value is calculated from the true wind, and should subsequently correlate positively with the speed loss. From looking at the data we can see that it correlates weakly with the speed loss.

A difference between the results from the two different calm water speed models is the presence of some negative speed loss outliers in the unmodified model. These measurements are also extreme values in trim. The values have a trim of around negative one meter. Interestingly, the outliers does not show up when the modified model is used.

8.2 Main Network Testing

The following results describe the networks ability to learn and generalize the relationship between the input sets described in Table 7.3 and the target quantity. First, the difference in performance between the input sets will be investigated.

8.2.1 Input Set Performance

It is useful to see how the different input sets perform, as the available input variables may change depending on the context. The inputs sets are initially presented in Table 7.3, but they are repeated in Table 8.4 together with a ranking of their performance as concluded from the data in Figure 8.6

TABLE 8.4: Sets of possible inputs with relative performance ranking.

Variable	**Input 1**	**Input 2**	**Input 3**	**Input 4**	**Input 5**	**Input 6**	**Input 7**
Ranking	4	1	3	2	5	6	7
Draft	x	x	x	x	x	x	x
Trim	x	x	x	x			
Shaft Power				x			
Shaft RPM			x				
Log Speed		x			x	x	
Significant wave height	x	x	x	x	x		
Mean wave period	x	x	x	x			
Relative wave direction	x	x	x	x		x	x
Relative wind speed	x	x	x	x			
Relative wind direction	x	x	x	x			
Beaufort value					x	x	x

The seven different sets inputs can be separated based on the input variables. The first four input sets give in general more information regarding the weather and operational condition and will thus perform better than the last three that have less data. This is reflected in the results. The first four inputs perform the best in general. Input 1 is the basis of these inputs, and 2, 3, and 4, have one additional variable describing the operational condition. Out of these, the addition of the speed through water is the most useful. This is probably because of the structure in the relationship between speed loss and actual speed as seen in Figure 8.4e. The ranking of input sets 2, 3, and 4 is consistent with the level of correlation the

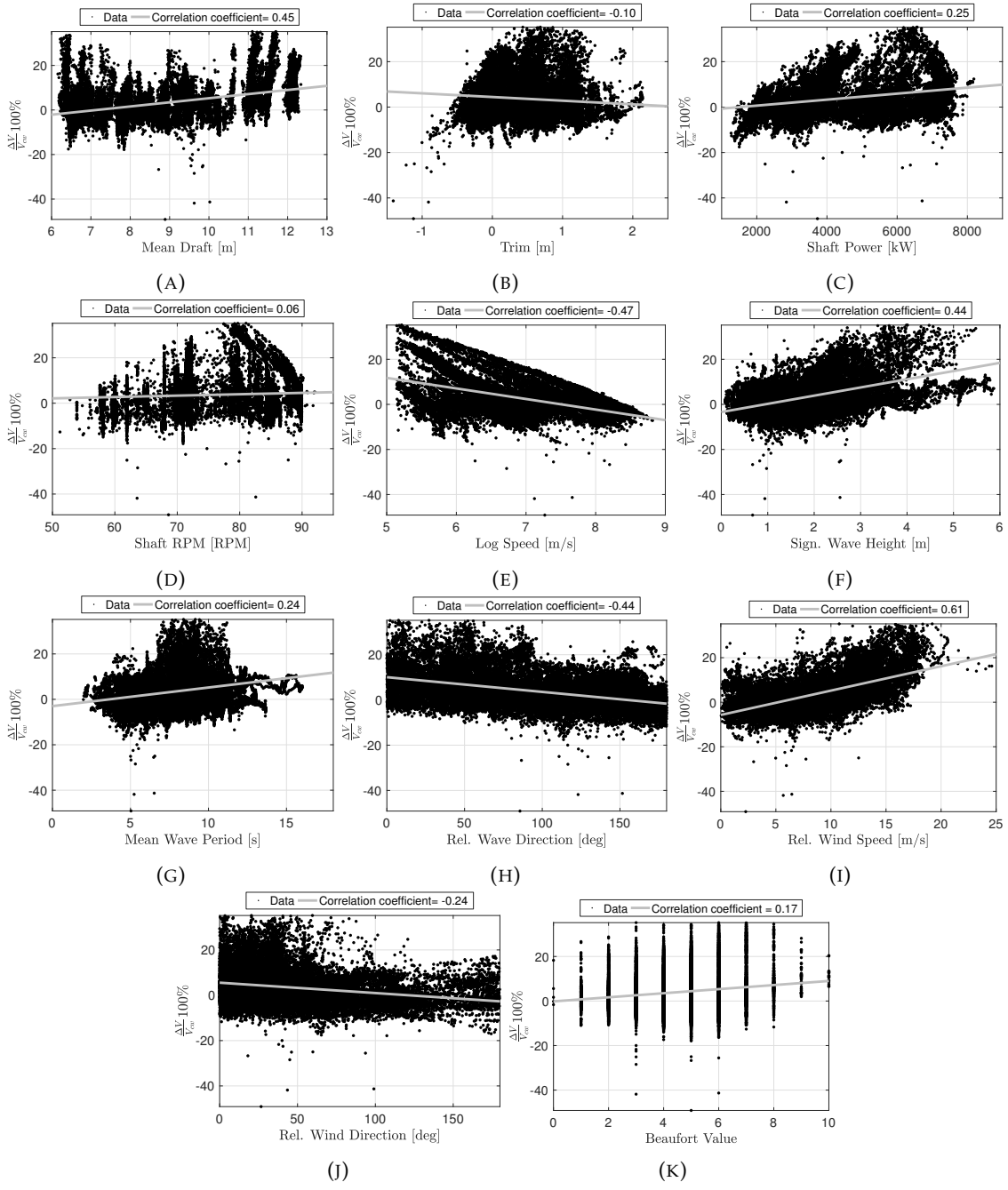


FIGURE 8.4: Scatter plots depicting the correlation between input variables and speed loss.

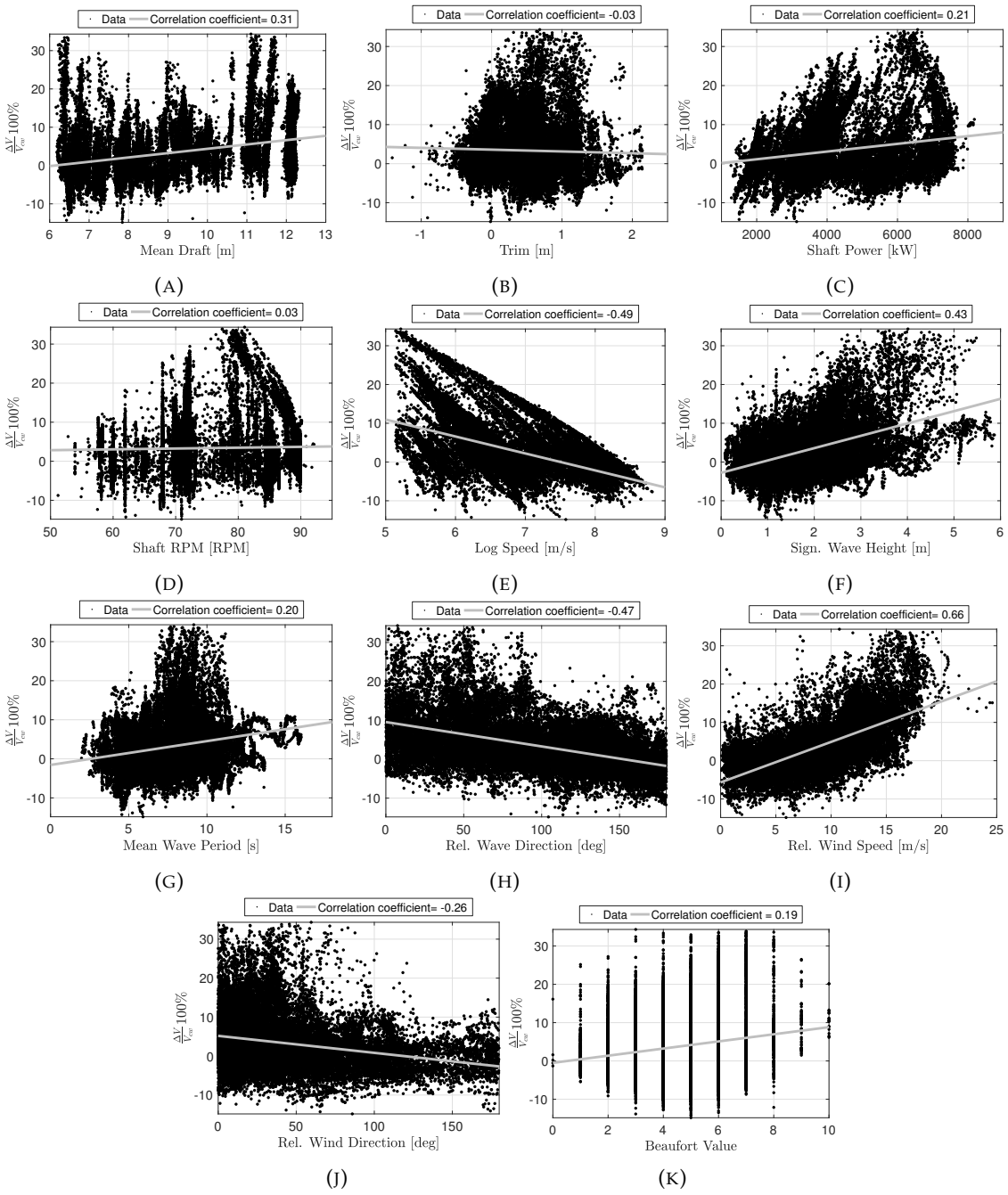


FIGURE 8.5: Scatter plots depicting the correlation between input variables and speed loss from the modified custom linear regression model.

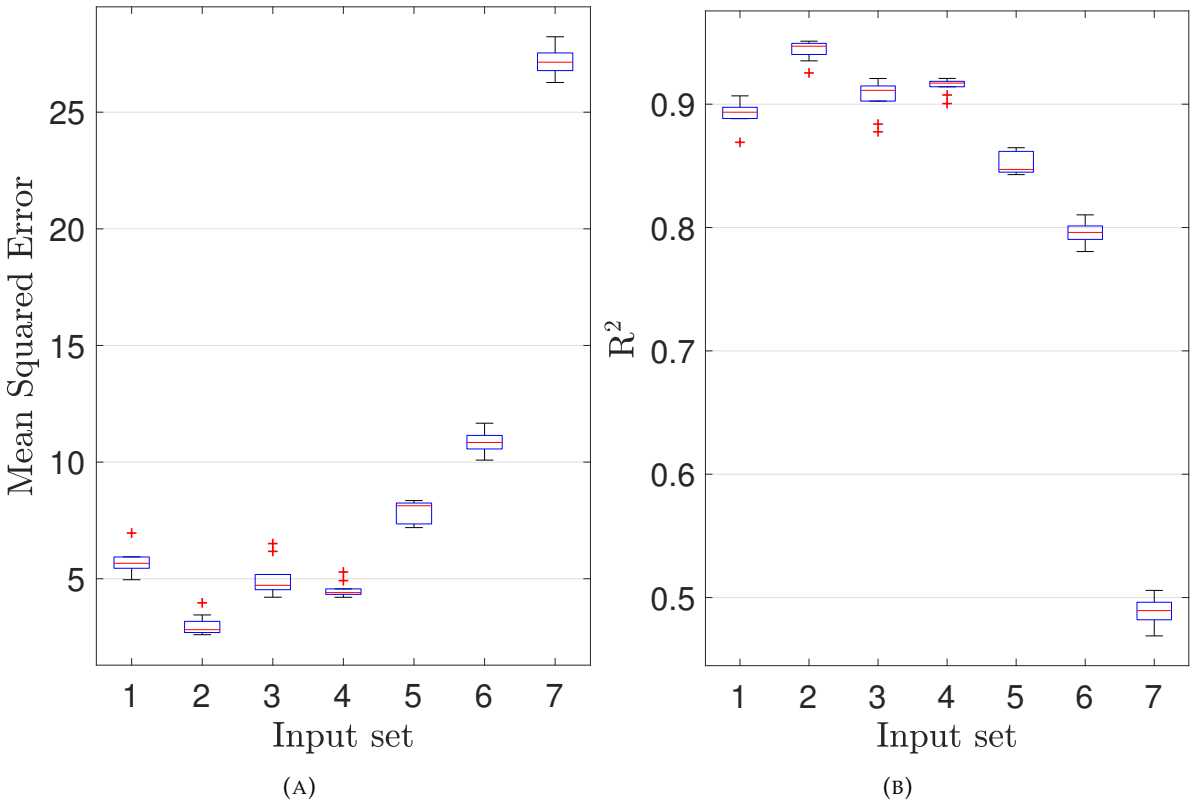


FIGURE 8.6: (A) Boxplot indicating the mean and spread of the mean squared test error of ten networks with 125 hidden units trained on the seven sets of inputs of the ss-data set. (B) Boxplot indicating the mean and spread of the test R^2 of ten networks with 125 hidden units trained on the seven sets of input of the ss-data set.

added variable has with the speed loss. It is apparent that there is value in the added variable, since input sets 2, 3, and 4 perform better than input set 1.

The last three input sets indicate again that the inclusion of the speed through water is beneficial. Furthermore, they indicate that there is more value in knowing the significant wave height than the relative mean wave direction. The last input set, which is similar to the input in Kwon's method performs significantly worse than the other input sets.

From Figure 8.6 and Table 8.5 we see that the difference in performance over the ten networks that was trained is very stable. This is indicated by the low standard deviation.

As it may be hard to relate to how good a prediction error is when measured in terms of the mean squared error of a quantity given as a percentage, the results can be transformed to units of speed by a rough calculation. If we take a mean squared error of 5, an upper bound on the mean absolute error is the root of the mean squared error, so the root can suffice as an estimate on the mean error, hence an estimate on the mean absolute error is $\sqrt{5} \approx 2.2$. The average value of the calm water speed for the ss-data set is $\overline{V_{cw}} \approx 7$ [m/s] or $\overline{V_{cw}} \approx 14$ [knots]. The mean error of the speed loss prediction is around 2 % of the mean value of V_{cw}. This turns out to be around 0.14 [m/s] or 0.28 [knots]. A similar rough calculation based on a mean squared error of 27, results in a mean error of around 0.4 [m/s] or 0.7 [knots].

TABLE 8.5: Overview of performance metrics and their variation for the seven input sets. The training was averaged over ten networks with 125 hidden units. The training was done using the ss-data set.

Measure	Data	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6	Input 7
$\overline{MSE}$*	Training	3.265	1.535	2.759	2.466	5.942	6.497	14.392
$\sigma(MSE)$**		0.084	0.062	0.092	0.145	0.161	0.325	0.448
$\overline{R^2}$*		0.940	0.972	0.949	0.955	0.891	0.881	0.736
$\sigma(R^2)$*		0.002	0.001	0.002	0.003	0.003	0.006	0.008
$\overline{MSE}$*	Testing	4.060	2.002	3.606	3.152	6.401	6.593	15.725
$\sigma(MSE)$**		0.208	0.071	0.113	0.268	0.194	0.226	0.462
$\overline{(R^2)}$*		0.925	0.963	0.933	0.941	0.881	0.878	0.708
σR^2*		0.004	0.001	0.002	0.005	0.004	0.004	0.009

*The bar denotes mean values. **$\sigma(\cdot)$ denotes the standard deviation.

TABLE 8.6: Performance measures when a linear regression model is used to predict speed loss. The model is trained on the ss-data set. 15 % is used for testing.

Measure	Data	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6	Input 7
MSE	Training	19.1933	14.8270	18.1788	19.0423	28.3486	27.1575	31.5424
R^2		0.6478	0.7279	0.6664	0.6505	0.4797	0.5016	0.4211
MSE	Testing	19.1832	14.6365	18.0775	18.9819	28.6803	26.8345	31.3338
R^2		0.6439	0.7283	0.6645	0.6477	0.4677	0.5019	0.4184

8.2.2 Linear Regression Benchmark

As expected, the linear regression model significantly underperforms compared to the neural network regression model. The neural network regression model yields mean squared error values on the test set between approximately 2 and 16, where the value of 16 is an outlier compared to the performance of the other input sets. The linear regression model have mean squared error values between 15 and 32. The R^2 goodness of fit coefficients behave similarly. The highest test set R^2-coefficient of the linear regression results, is similar in performance to the lowest of the neural network regression results.

An additional property of the performance results is that the performance ranking between the two regression models stays consistent. This yields additional weight to the qualitative remarks in 8.2.1 on what variables is most important to predict speed loss.

The conclusion from the comparison with a linear regression model is that the neural network model performs significantly better. The combination of high model flexibility and a quite large number of measurements allows the use of neural networks in an efficient way.

8.2.3 Prediction Testing

The following results depict the predictive performance of the method in a realistic setting, by the testing described in Subsection 7.6.5. In Figure 8.7 the resulting predictions for the test where one voyage is removed from the dataset is shown. Figure 8.8 illustrates the results for when the ss-data set is used to train a model for predicting the speed loss of Ship E7. The networks used in this section have the network structure that has been used earlier, of 125 hidden units. Input set 1 is the chosen input set.

The results from the case where one voyage is removed shows how the predictions compare to the actual speed and speed loss. From the figure showing the actual speed, it is shown that the predicted speed seem

to follow the measured speed. However, it appear as if the speed loss is overestimated for some samples. This behaviour is evident in the samples around number 400.

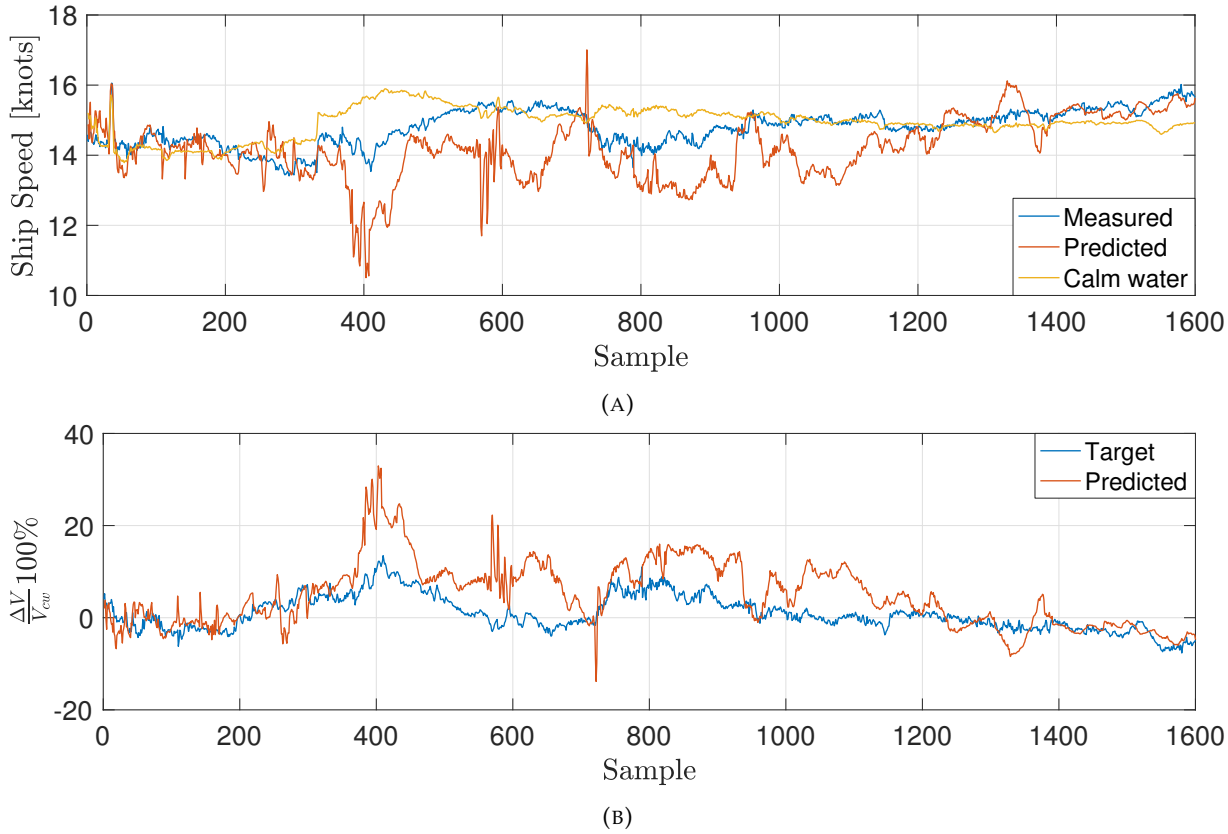


FIGURE 8.7: (A) Illustration of the predicted, measured and calm water speed when one voyage is left out of the training data.
(B) Illustration of the target and predicted speed loss when one voyage is left out of the training data.

In Figure 8.8, it appears that the network works well for some samples. From sample 1500 and onward the predictions follow the actual speed very well. In the samples before, the predicted speed loss seems to vary.

All in all, the test seem to indicate that the network struggles for some samples, while other is predicted very well. A probable explanation for the varying performance is that the samples where the network struggles the most are those that are the least similar to the training data. This is a general pitfall in using any type of regression model, and relates to the previously mentioned concept that more data, gives better generalization.

8.3 Method Summary and Discussion

The objective of this thesis has been to develop a method for predicting speed loss to aid in route optimization. The solution to this problem has been to first develop a regression framework for the estimation of a calm water reference speed V_{cw}. This regression model is optimized with respect to the particulars of a specific ship, by being trained on a subset of available operational data characterized by a calm water condition. From measurements of actual speed through water, we can define a speed loss that can be

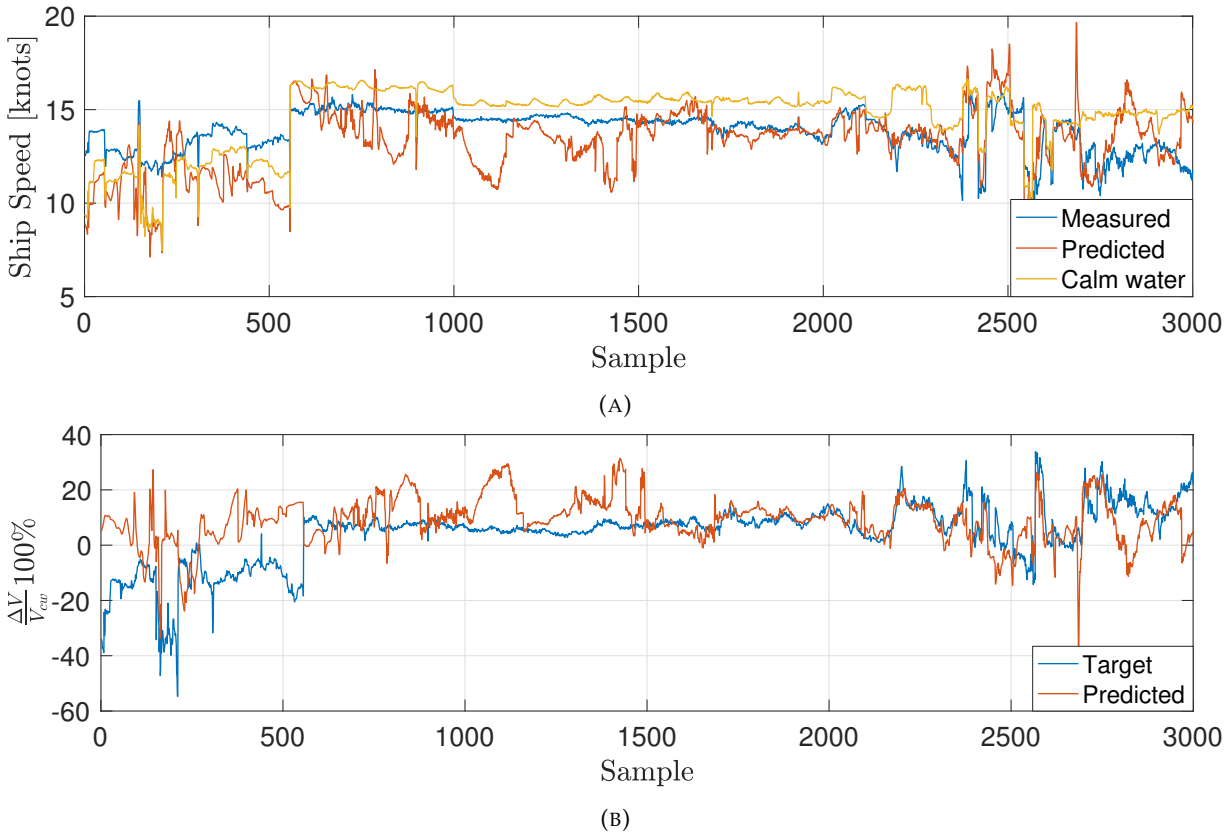


FIGURE 8.8: (A) Illustration of the predicted, measured and calm water speed the ss-data set is used for training and the predictions are made on Ship E7.
(B) Illustration of the predicted and target speed loss when the ss-data set is used for training and the predictions are made on Ship E7.

calculated for all samples in the complete database. This process yields a database which contain ship operational data, weather data, and a speed loss. The final step is to train a regression model that can predict speed loss based on input variables describing the operational state and the weather state. To develop this model, a neural network regression model approach has been employed.

The previous paragraph summarizes the process behind the solution of the model that has been developed. The success of this method as whole is dependent on a few conditions regarding the steps of the method.

Firstly, the data that is used has to be of high quality. This has been ensured by filtering and transforming data to facilitate efficient training of the regression models. As an additional measure, data validation has been performed.

The second part of the method is the estimation of the calm water reference speed. In Section 7.3, the development of the calm water regression is described. The choice of a custom nonlinear regression resulted in a robust model that could produce estimates of the calm water reference speed despite low availability of calm water measurements. A weakness was the lack of flexibility related to the term incorporating the draft. It has been shown that this can be mended by modifying this term, under the condition that enough calm water measurements is available. A key point behind the choice of this way to compute V_{cw} was a preference for the calculation to be based on data. In general the model used to calculate V_{cw} can be replaced, to facilitate speed loss predictions based on other needs or assumptions. This was done (as seen

in Section 8.1) to validate the custom nonlinear regression model. The underlying goal with this part of the method is to establish speed loss, and other methods may also be useful.

The next part of the method is the development of the regression model for predicting future speed loss. In this step it is assumed that the training data, containing ship operational data, weather data, and speed loss data, is valid. The work that has been done regarding the specifics of the neural network concerns firstly what data, and in what form the data is to be given to the network. A few tests to investigate the effectiveness of the different alternatives, and the subsequent discussion is shown in Sections 7.5.2, 7.5.3, and 7.5.3. It was for instance shown that the network performs best when the wind data is given separately, as relative wind speed and relative wind direction. Different sets of input has been tested and their relative effectiveness has been determined.

Based on the review of general machine learning and neural network theory, a strategy for ensuring good artificial neural network performance has been developed. A complex neural network configuration effectively regularized by implementation of early stopping is the recommended approach. The use of neural network as a tool was deemed to be appropriate based on the performance results on their own, and in comparison with a linear regression approach.

A final point of discussion is the applicability of the method in the context that was introduced in Chapter 1. The reliable estimation of speed loss is essential to perform route optimization regardless if the optimization objective is environmentally driven or logistically driven. The different input sets represent requirements necessary for achieving a particular prediction performance in terms of input if the developed method is to be used in such a context. A viable alternative that can possibly be implemented is the use of input set 1. A suggested algorithm is presented in Figure 8.9. The developed method, identified by the dashed border, can fit into a route optimization workflow as indicated. If a route is defined, with a starting location, starting time, destination location, and arrival time. Routing software can suggest a required speed under the assumption of calm water. Based on operational parameters and forecasted weather, the speed loss model will predict the actual speed. If this speed is lower than the required speed, a new higher speed is suggested. This iterative process is concluded when the predicted actual speed is sufficient. The corresponding calm water speed is then known as the last speed that was suggested, and required shaft power and engine speed can be predicted under calm water assumptions, which again can be used to predict fuel consumption, emissions, or other relevant quantities. This algorithm is only one of many possible ways to utilize the method that has been developed, but represents a concrete way the method achieves the objective that was outlined.

One last point to be emphasized is the modular nature of the method. It has been previously mentioned that the calm water reference speed model can be exchanged for another model depending on the needs. This is also true for the neural network regression. If another, better regression model, either for the estimation of V_{cw} the speed loss directly, is found, or another better regression model for learning based on the estimated speed loss is found, there is no problem in replacing the current modules to improve the overall method.

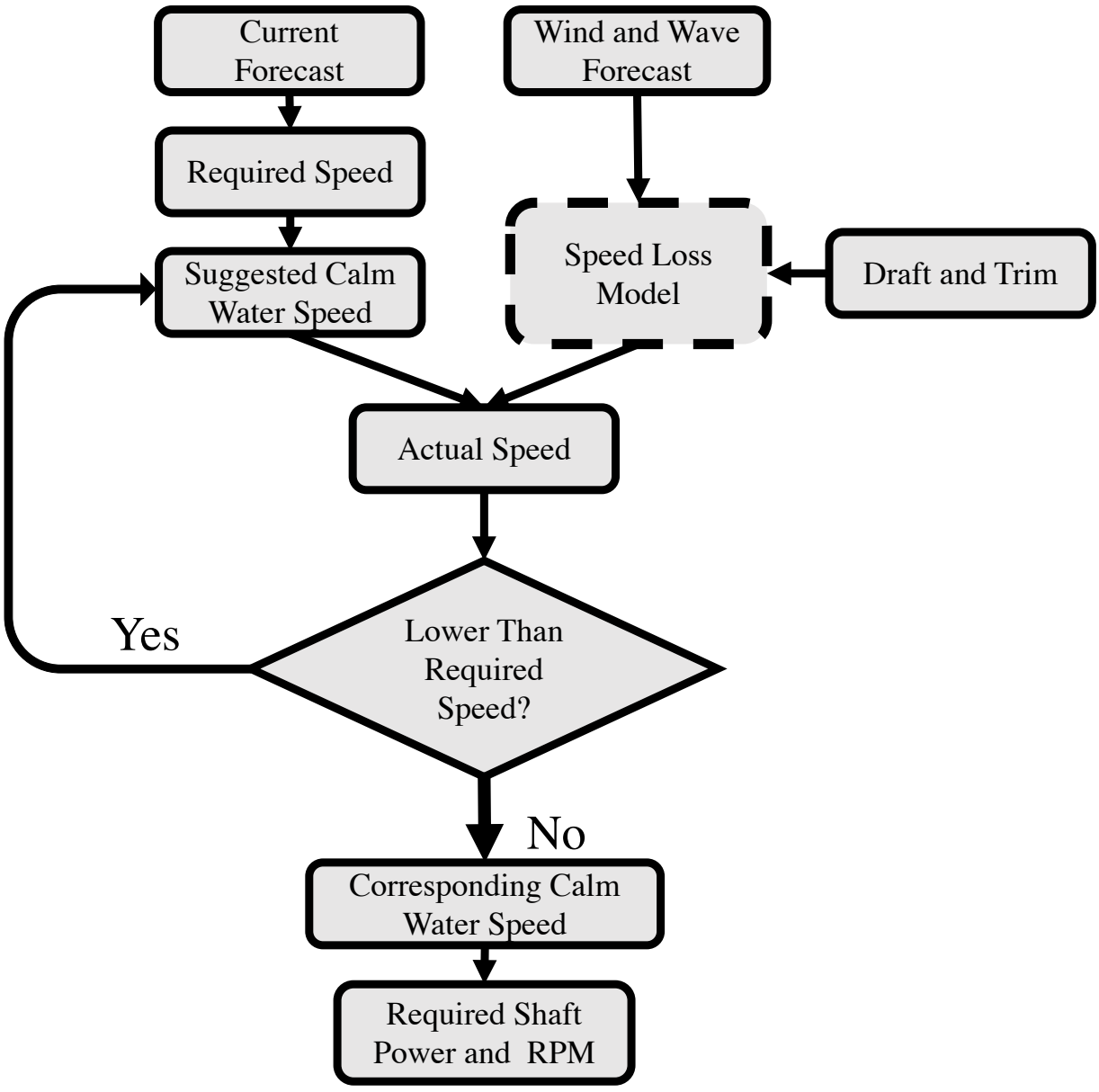


FIGURE 8.9: Flowchart outlining a suggested algorithm for application of the developed method.

9 Conclusion

The research goal of this thesis has been to develop a method that uses a data-driven approach to estimate and predict the speed loss of a ship when exposed to a seaway. A method has been proposed that has accomplished this goal. The developed method works by first combining ship operational data and weather data to a combined database. From this database a subdatabase is extracted based on a calm water condition. This subdatabase is used to estimate a calm water reference speed that is then used to calculate a speed loss. The combination of ship operational data, weather data, and the speed loss data is subsequently used to train an artificial neural network to predict speed loss. The key modules of this method is the the calm water speed regression model, and the neural network speed loss prediction model. The conclusions made on the most important aspects of the developed method can be summarized as follows:

To estimate the reference calm water speed V_{cw}, and subsequently the speed loss $\frac{\Delta V}{V_{cw}}100\%$, a custom nonlinear regression model has been developed based on physical considerations. This method uses real calm water data to establish the relationship between trim, draft, shaft power, and speed through water. This model is robust and requires smaller amounts of data than a neural network regression model to perform good estimates, and it performs better than a simple linear regression model. In a validation study based on the expected correlations between weather variables and the corresponding speed loss, the model performs satisfactory. Overestimation of the reference speed occurs for loading conditions where the draft is high, which in turn results in an overestimation of the speed loss. A modified version of the model is suggested, which is able to alleviate this problem if there exists a sufficient amount of good calm water data. Depending on the application and available resources, other methods for speed loss estimation may be advantageous.

To predict speed loss, a strategy for developing neural network regression models has been suggested. The strategy suggests the use of neural networks with a large number of hidden units, on the order of 125-400, coupled with the implementation of early stopping to prevent overfitting. Depending on the input set, mean squared errors are on the order of 2-5 [$\%^2$] on the test set, and R^2 coefficients on the order of 0.90-0.95. The equivalent prediction error in terms of speed turns out to be close to 0.14 [m/s] based on approximate calculations. The strategy has shown good behaviour regarding overfitting and the stability of the solution, which suggests that the neural network is approaching the limit of what can be achieved in terms of its predictive ability given the available data.

The effectiveness of some sets of input variables has been compared. The relative ranking of the input sets can guide the best use of the developed method in a route optimization context. One suitable algorithm is given to establish the utility of the method for one particular input set.

An additional strength of the method is its modular nature. The structure of the method allows for either regression models to be replaced if another model is better suited for specific needs, or another model is found to have better predictive performance.

Bibliography

Beale, Mark Hudson, Martin T. Hagan, and Howard B. Demuth (2016). *Neural Network Toolbox - User's Guide*. MathWorks. URL: https://www.mathworks.com/help/pdf_doc/nnet/nnet_ug.pdf.

Bishop, Christopher M. (1995). *Neural networks for pattern recognition*. Vol. 92, p. 482.

Bishop, Christopher M. (2007). *Pattern Recognition And Machine Learning*, p. 738.

Carlton, John S. (2007). *Ship Resistance and Propulsion*, pp. 299–332.

Caruana, Rich, Steve Lawrence, and Lee Giles (2000). "Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping". In: *Advances in neural information processing systems*.

ECMWF (2016). *ECMWF Parameter Database*. URL: http://apps.ecmwf.int/codes/grib/param-db.

Faltinsen, Odd Magnus (1993). *Sea Loads on Ships and Offshore Structures (Cambridge Ocean Technology Series)*. Cambridge University Press.

Faltinsen, Odd Magnus et al. (1980). "Prediction of Resistance and Propulsion of a Ship in a Seaway". In: *Proceedings of the 13th Symposium Naval Hydrodynamic*, pp. 505–529.

Gerritsma, J and W Beukelman (1972). "Analysis of the resistance increase in waves of a fast cargo ship". In: *International shipbuilding progress* 19.217.

Gjølme, Jens Christoffer (2016). "Development of an Empirical Model for Added Resistance Using Artificial Neural Networks". Project Thesis. NTNU.

Harvald, SV. AA. (1983). *Resistance and propulsion of ships*. 5-6. Lyngby: Wiley, p. 477.

Hollenbach, Klaus Uwe (1998). "Estimating Resistance and Propulsion for Single Screw and Twin Screw Ships". In: *Ship Technology Research* 45.

Holtrop, J and G G J Mennen (1982). "An approximate power prediction method". In: *International Shipbuilding Progress*, pp. 166–170.

Hornik, Kurt (1991). "Approximation capabilities of multilayer feedforward networks". In: *Neural Networks* 4.2, pp. 251–257.

Hornik, Kurt, Maxwell Stinchcombe, and Halbert White (1989). "Multilayer feedforward networks are universal approximators". In: *Neural Networks* 2.5, pp. 359–366.

ISO (2016). *ISO 19030-2*.

ITTC (2011). *ITTC - Recomended Procedures - 7.5-02 -07-02.2 - Prediction of Power Increase in Irregular Waves from Model Test*. URL: http://ittc.info/media/1311/75-02-07-022.pdf.

ITTC (2014). *ITTC - Recomended Procedures - 7.5-04 -01-01.2 - Speed and Power Trials, Part 2 Analysis of Speed/Power Trial Data*.

Kim, M. et al. (2016). "A study on ship speed loss due to added resistance in a seaway". In: *Proceedings of the International Offshore and Polar Engineering Conference* 2016-Januar.

Kotsiantis, SB, D Kanellopoulos, and PE Pintelas (2006). "Data preprocessing for supervised learning". In: *International Journal of Computer Science* 1.2, pp. 111–117.

Kwon, Y. J. (2008). "Speed loss due to added resistance in wind and waves". In: *Naval Architect* MAR. Pp. 14–16.

Lewis, Edward V. (1989). *Principles of Naval Architecture (Second Revision), Volume III - Motions in Waves and Controllability*. Society of Naval Architects and Marine Engineers (SNAME).

Lu, Ruihua et al. (2015). "A semi-empirical ship operational performance prediction model for voyage optimization towards energy efficient shipping". In: *Ocean Engineering* 110, pp. 18–28.

Mao, Wengang et al. (2016). "Statistical models for the speed prediction of a container ship". In: *Ocean Engineering* 126, pp. 152–162.

Maruo, H. (1957). "The Excess Resistance of a Ship in Rough Seas". In: *International Shipbuilding Progress* 4.

Mitchell, Tom M. (1997). *Machine Learning*. first. McGraw-Hill Education.

Molland, Anthony F., Stephen R. Turnock, and Dominic A. Hudson (2011). *Ship Resistance and Propulsion: Practical Estimation of Propulsive Power*. Cambridge University Press.

Mororka Online User Guide (2016). Marorka ehf. Borgartun 26, 206 Reykjavik, Iceland.

Paul Berrisford, Dick Dee (2011). *The ERA-Interim archive*. Tech. rep. ECMWF.

Prpić-Oršić, Jasna and Odd Magnus Faltinsen (2012). "Estimation of ship speed loss and associated {CO2} emissions in a seaway". In: *Ocean Engineering* 44, pp. 1 –10.

Seo, Min-Guk et al. (2013). "Comparative study on computation of ship added resistance in waves". In: *Ocean Engineering* 73, pp. 1 –15.

Smith, T. W. P. et al. (2015). *Third IMO GHG Study 2014*.

Steen, Sverre and Knut Minsaas (2013). *TMR4220 Naval Hydrodynamics - Ship Resistance*.

Townsin, R.L. and Y. J. Kwon (1983). "Appoximate formulae for the speed loss due to added resistance in wind and waves". In: *Transactions of the Royal Institute of Naval Architects* 125, pp. 199–207.

A Abbreviations

ANN	**A**rtificial **N**eural **N**etwork
ECMWF	**E**uropean **C**entre for **M**edium-range **W**eather **F**orecasts
IMO	**I**nternational **M**aritime **O**rganization
MEPC	**M**arine **E**nvironment **P**rotection **C**omitee
MSE	**M**ean-**S**quared **E**rror
SOG	**S**peed **O**ver **G**round
STW	**S**peed **T**hrough **W**ater
SS	**S**ingle **S**hip

B Tables

TABLE B.1: Table showing the relationship between the wetted surface area and the displaced volume as a function of the mean draft a E-class hull. The data is taken from the hull geometry files and calculated by ShipX.

Mean Draft $[m]$	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0	11.5	12.0	12.5
∇ $[m^3]$	22942	25423	27924	29780	32982	35538	38116	40717	43345	45998	48691	51337	54159	56942	59691	62606
Wetted area $[m^2]$	6251	6463	6675	6887	7100	7318	7538	7770	8008	8249	8499	8758	9022	9288	9556	9815

C Input data

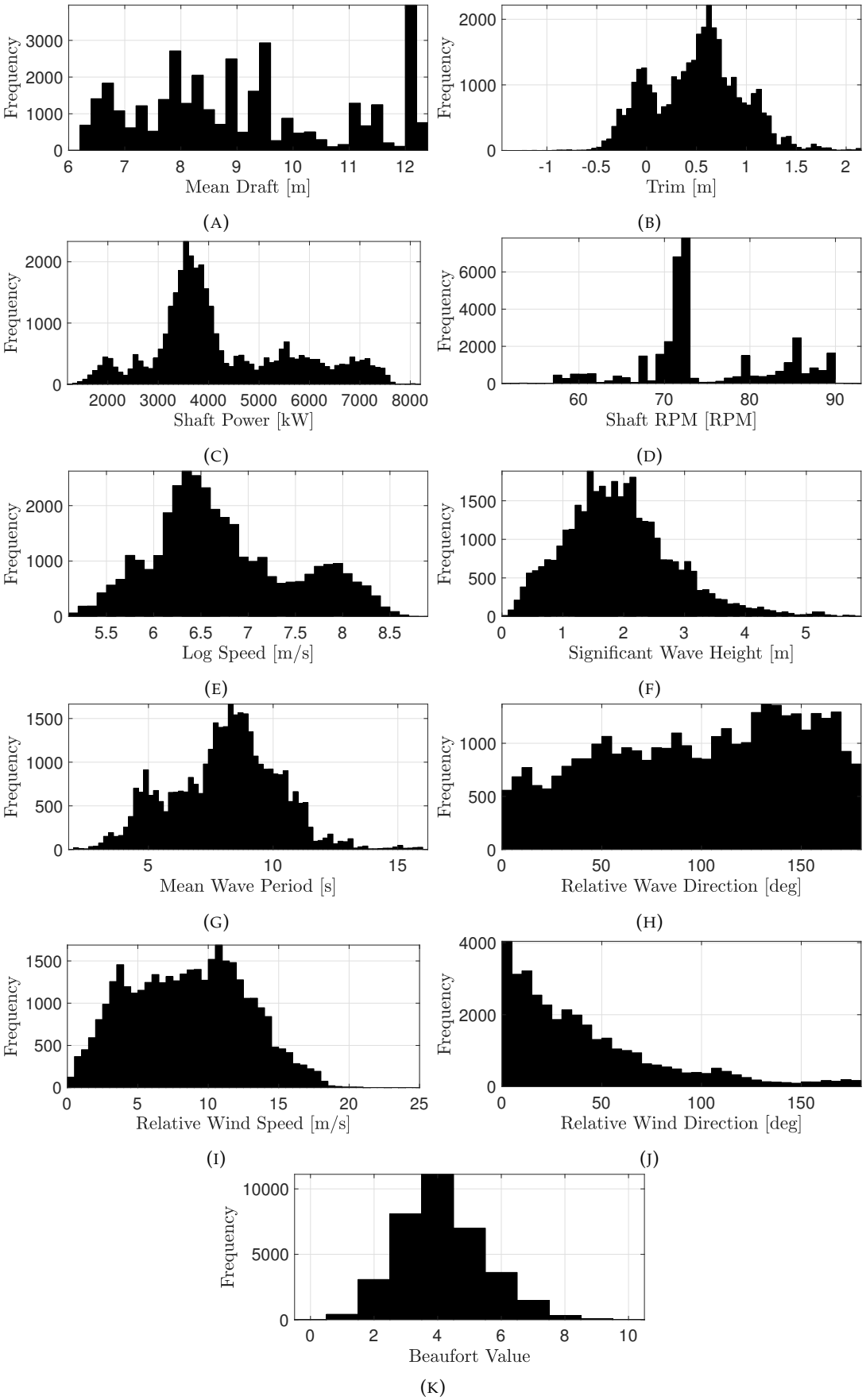


FIGURE C.1: Histograms of input data of the ss-data set.

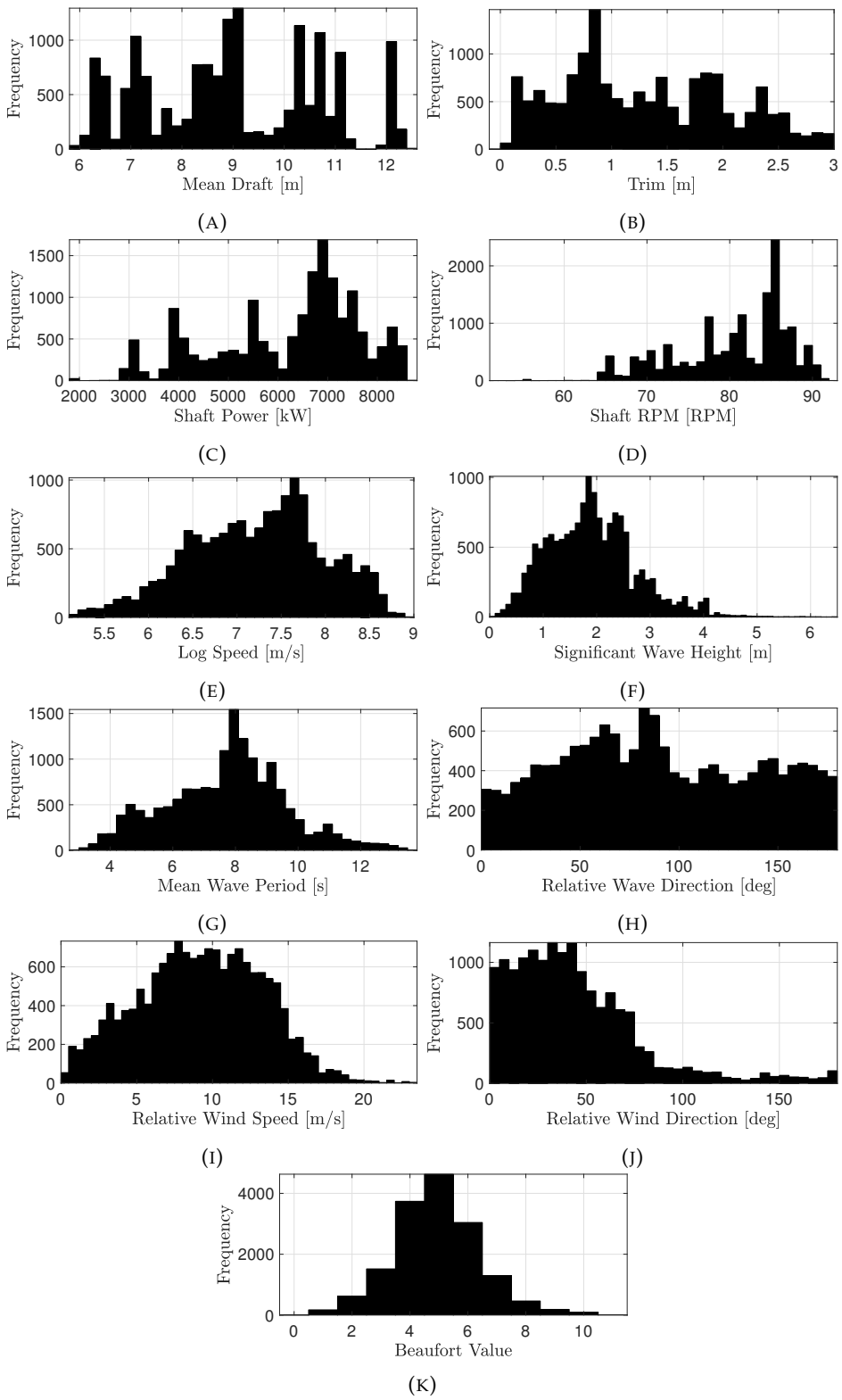


FIGURE C.2: Histograms of input data of the A-class data set.

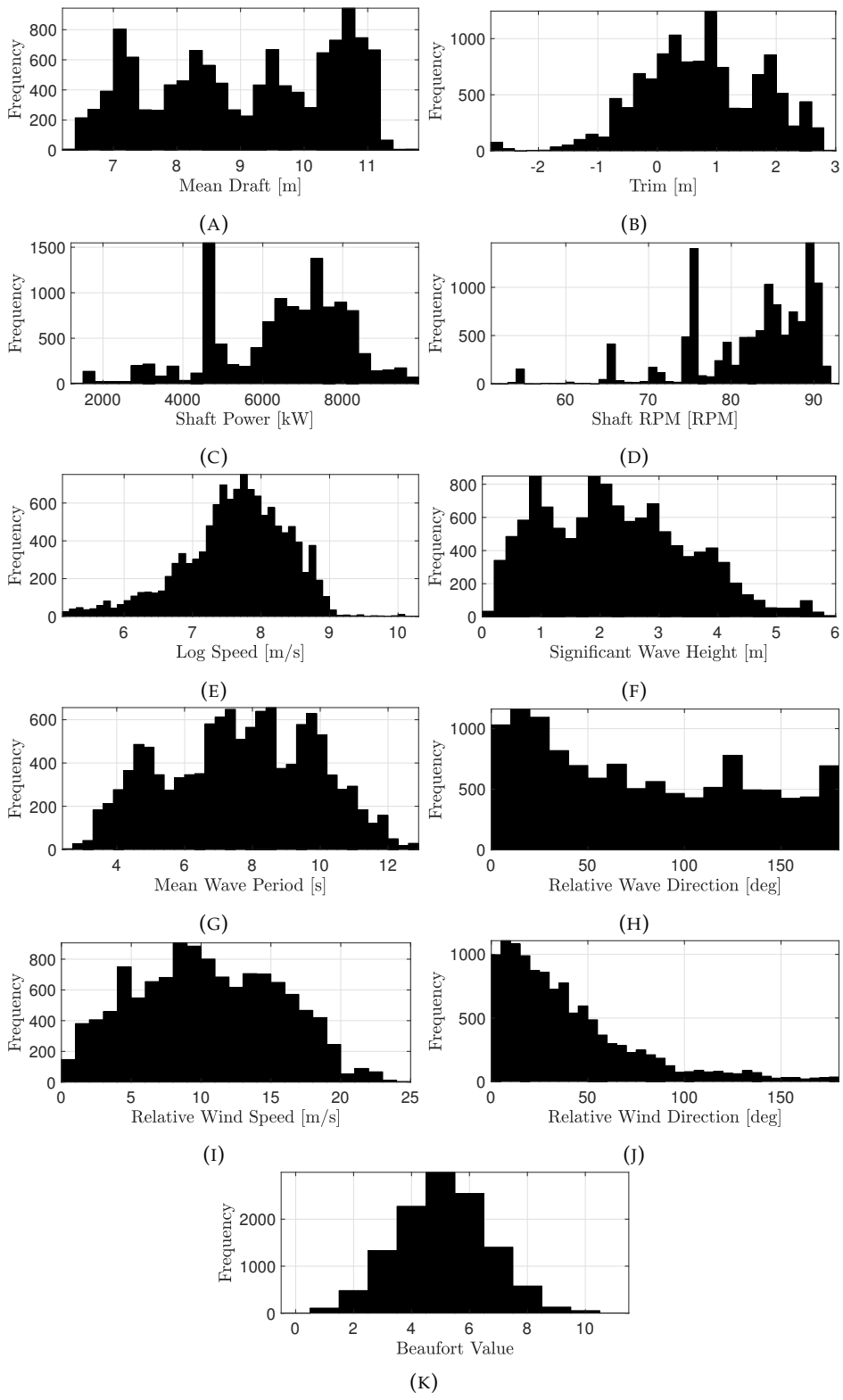


FIGURE C.3: Histograms of input data of the B-class data set.

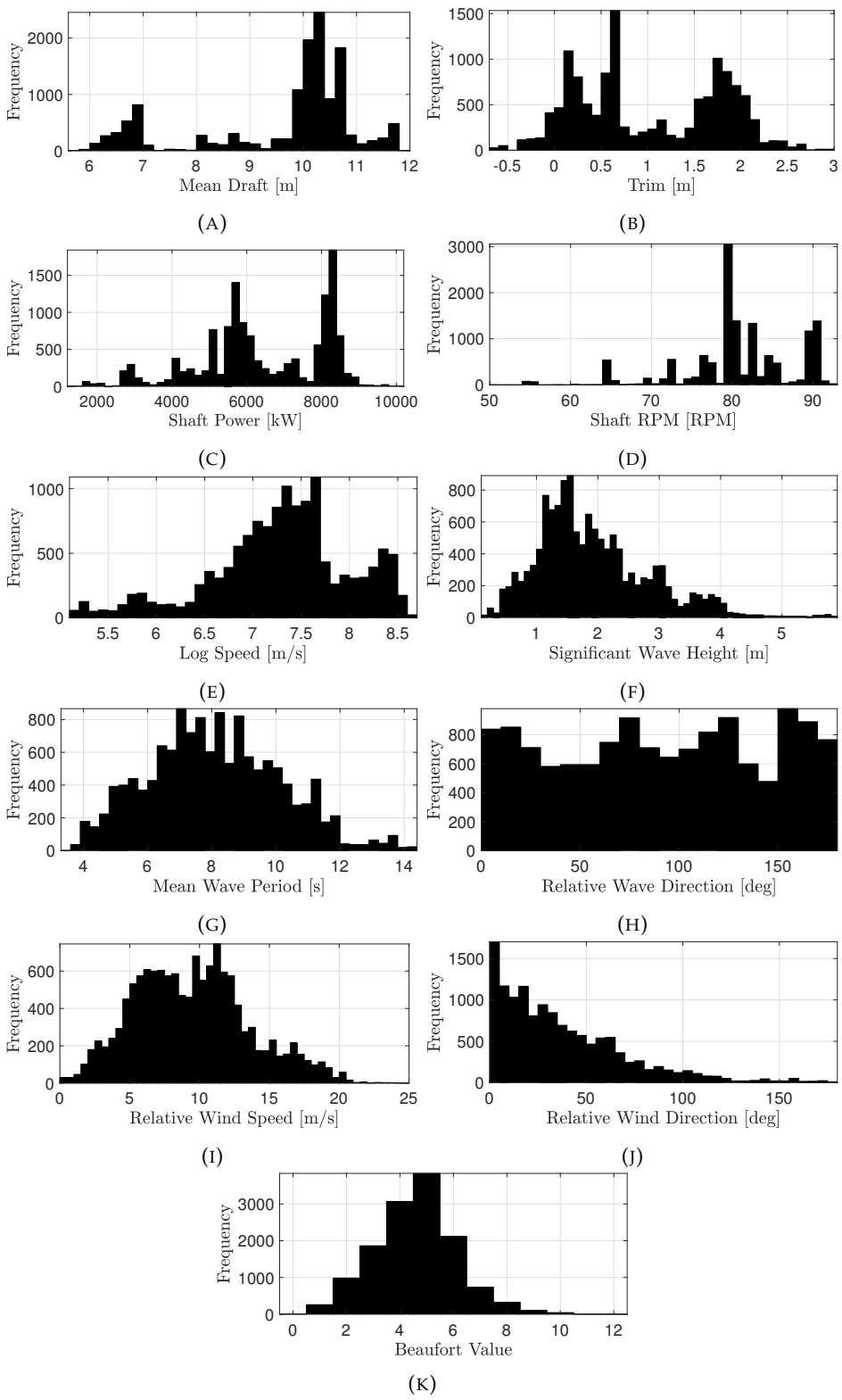


FIGURE C.4: Histograms of input data of the C-class data set.

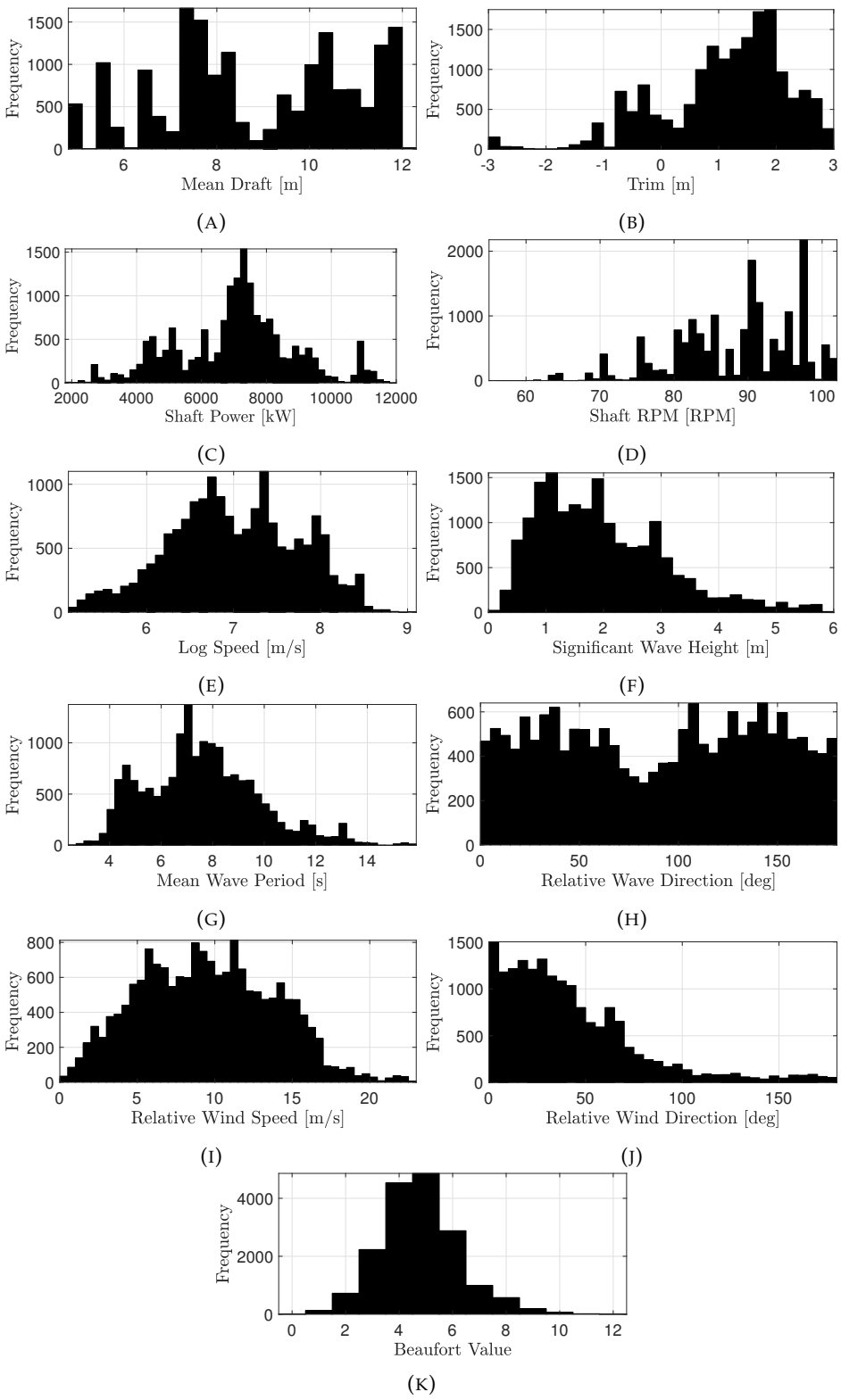


FIGURE C.5: Histograms of input data of the D-class data set.

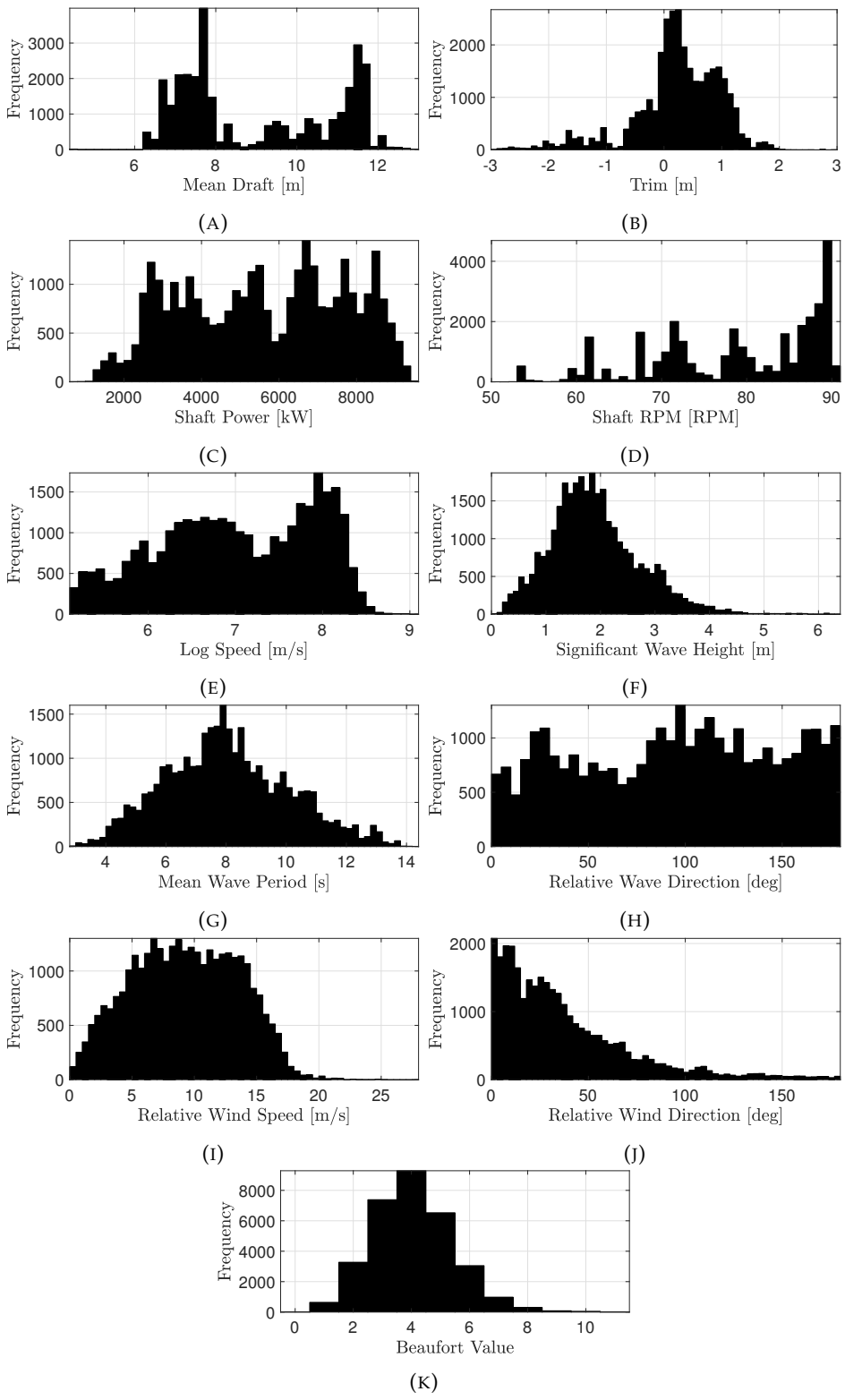


FIGURE C.6: Histograms of input data of the E-class data set.

D Result Data

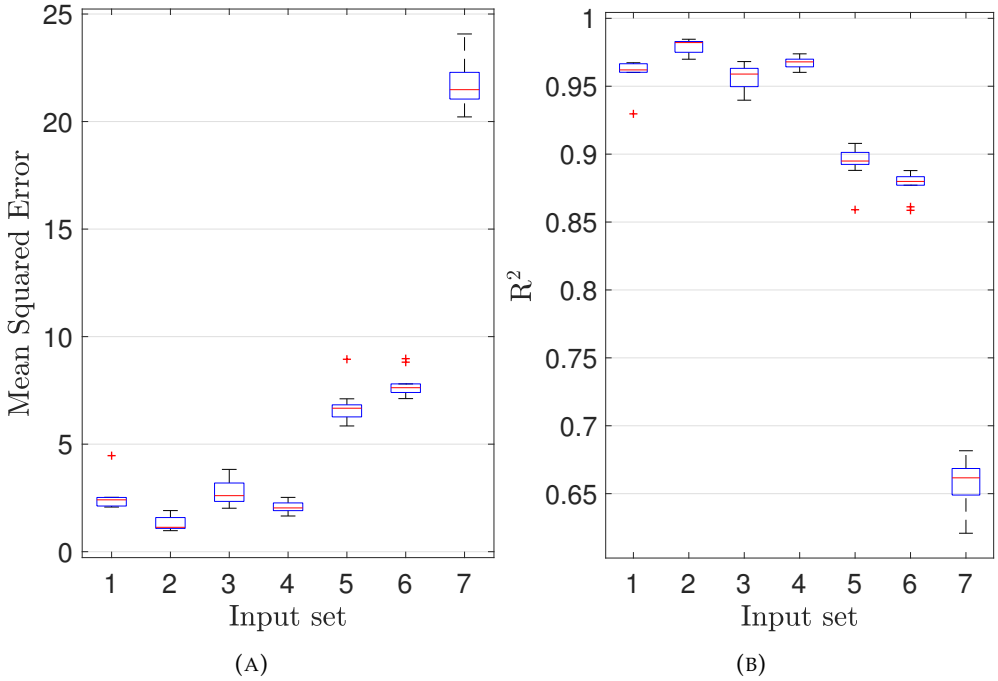


FIGURE D.1: (A) Boxplot indicating the mean and spread of the mean squared error of ten networks with 125 hidden units trained on the seven sets of input of the A-class data set. (B) Boxplot indicating the mean and spread of the R^2 of ten networks with 125 hidden units trained on the seven sets of input of the A-class data set.

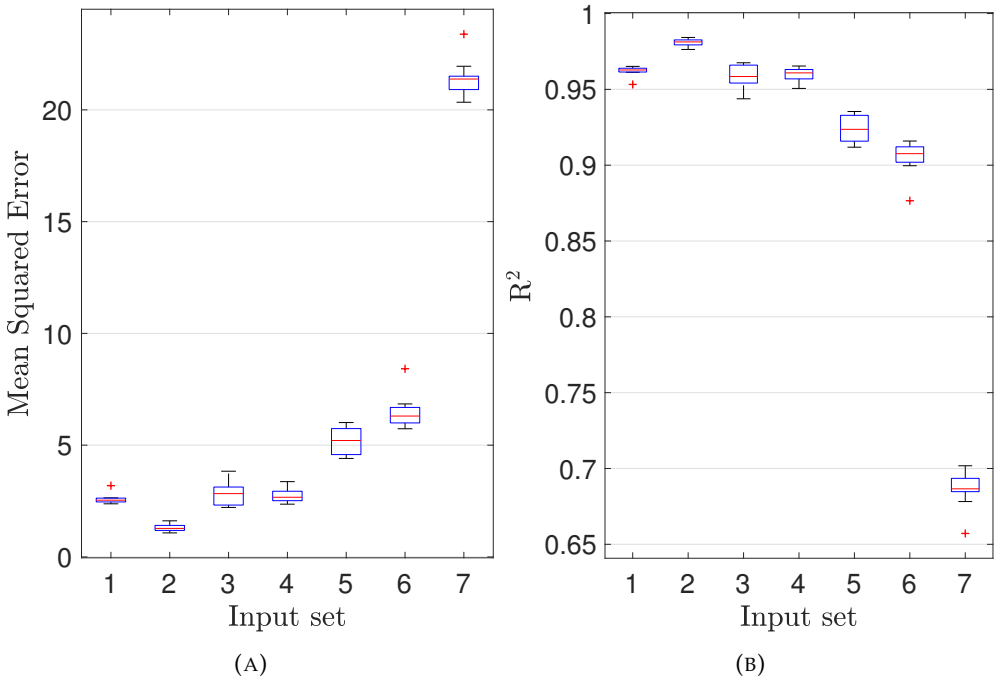


FIGURE D.2: (A) Boxplot indicating the mean and spread of the mean squared error of ten networks with 125 hidden units trained on the seven sets of input of the B-class data set. (B) Boxplot indicating the mean and spread of the R^2 of ten networks with 125 hidden units trained on the seven sets of input of the B-class data set.

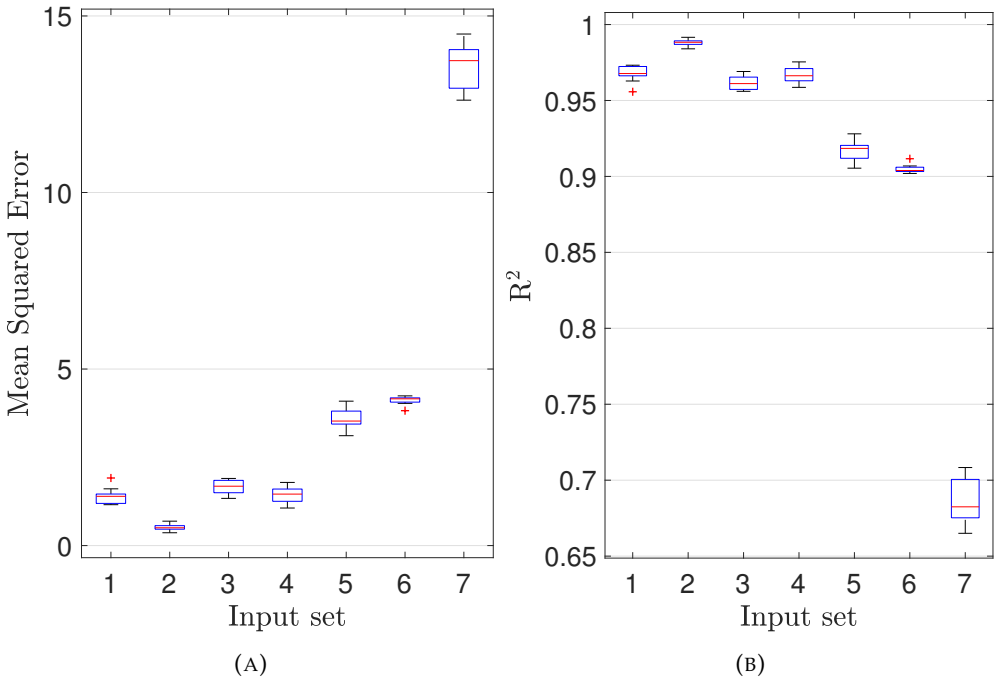


FIGURE D.3: (A) Boxplot indicating the mean and spread of the mean squared error of ten networks with 125 hidden units trained on the seven sets of input of the C-class data set. (B) Boxplot indicating the mean and spread of the R^2 of ten networks with 125 hidden units trained on the seven sets of input of the C-class data set.

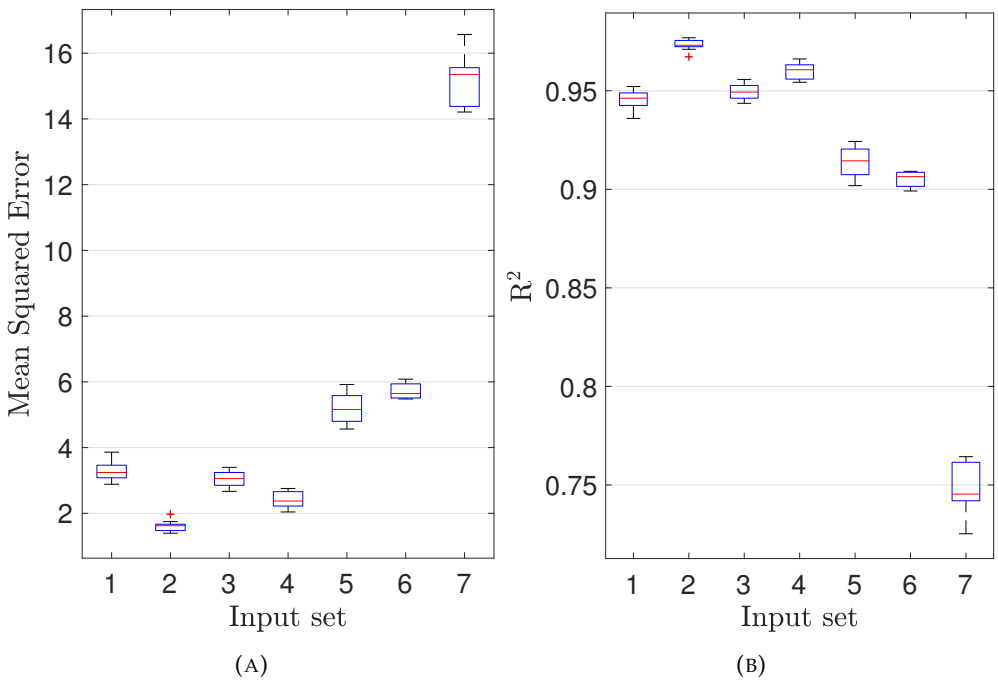


FIGURE D.4: (A) Boxplot indicating the mean and spread of the mean squared error of ten networks with 125 hidden units trained on the seven sets of input of the D-class data set. (B) Boxplot indicating the mean and spread of the R^2 of ten networks with 125 hidden units trained on the seven sets of input of the D-class data set.

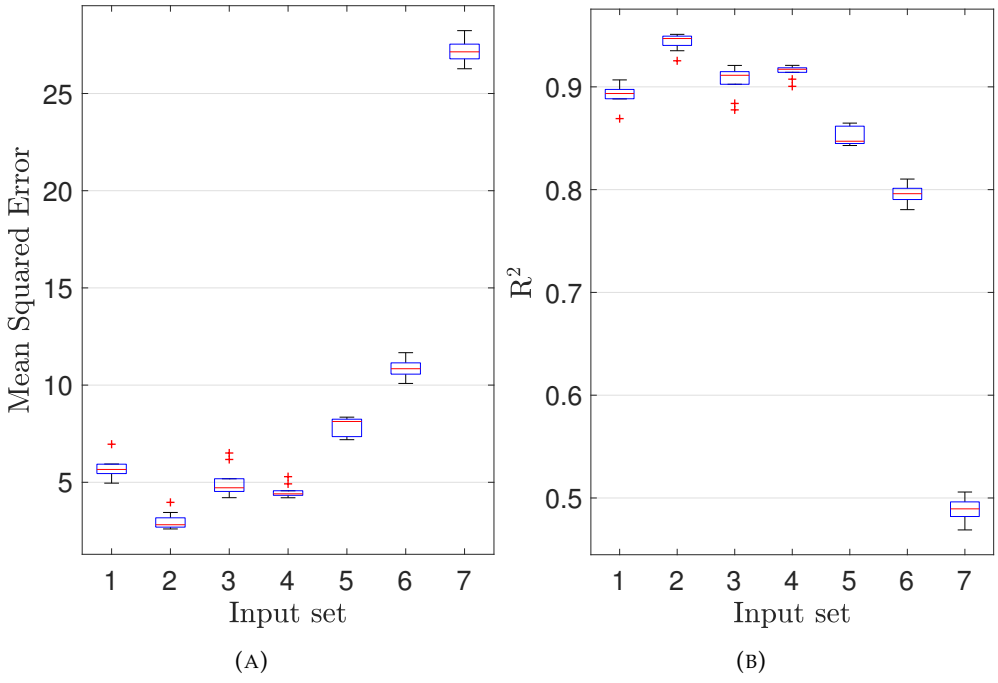


FIGURE D.5: (A) Boxplot indicating the mean and spread of the mean squared error of ten networks with 125 hidden units trained on the seven sets of input of the E-class data set. (B) Boxplot indicating the mean and spread of the R^2 of ten networks with 125 hidden units trained on the seven sets of input of the E-class data set.