Ship Power Prediction Using Machine Learning

by

Anthony Kriezis

Submitted to the Department of Mechanical Engineering in partial fulfillment of the requirements for the degree of Master of Science in Naval Architecture and Marine Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2022

© Massachusetts Institute of Technology 2022. All rights reserved.

Author

Department of Mechanical Engineering

May 6, 2022

Certified by

Themistoklis Sapsis
Professor of Mechanical Engineering
Thesis Supervisor

Accepted by

Nicolas Hadjiconstantinou
Chairman, Department Committee on Graduate Theses
Ship Power Prediction Using Machine Learning

by

Anthony Kriezis

Submitted to the Department of Mechanical Engineering on May 6, 2022, in partial fulfillment of the requirements for the degree of Master of Science in Naval Architecture and Marine Engineering

Abstract

One of the biggest challenges facing the shipping industry in the coming decades is the reduction of carbon emissions. A promising approach to this end is the use of the growing amount of data collected by vessels to optimize a voyage so as to minimize power consumption. The focus of this paper is on building and testing machine learning models that can accurately predict the shaft power of a vessel under different conditions. The models examined include pure theoretical models, pure neural network models, and combinations of the two. Using data on two car carrying vessels for 8 years it was found that neural networks incorporating some physical intuition can achieve a mean absolute percentage error of less than 5%, and an $R^2$ above 95%. This performance can be further improved by the addition of wave information, but it deteriorates when the data collection becomes less frequent.

Thesis Supervisor: Themistoklis Sapsis
Title: Professor of Mechanical Engineering
Acknowledgments

First of all, I would like to thank Professor Themistoklis Sapsis, my thesis supervisor, for his continuous support and valuable guidance during my graduate studies. I have also had the honor to work with Professor Chryssostomos Chryssostomidis who provided critical advice on all stages of this project and whose theoretical model I used to compare against the machine learning models. I also want to thank Alexis Charalampopoulos for his help on implementing neural networks, Dr. Ethan Pickering for his help on ensemble neural networks, and Dr. Samuel Rudy for his help on sparse regression. Finally, I want to thank Professor Nicholas Patrikalakis for his invaluable support and for being my advisor during my undergraduate studies at MIT.

I am grateful to the Department of Mechanical Engineering for awarding me the fellowship that funded this research.

I thank Dr. George Kriezis and Neptune Lines for providing us a large part of the data used in this study.
# Contents

1 Introduction .......................................................... 17
   1.1 Related work ...................................................... 18
   1.2 This Study’s Contribution .................................... 21

2 Background ............................................................ 23
   2.1 Multiple Regression .............................................. 23
   2.2 LASSO Regression ................................................. 24
   2.3 Neural Networks .................................................. 24
      2.3.1 Ensemble Neural Network ................................. 26
   2.4 Theoretical Model ............................................... 26

3 Data ................................................................. 27
   3.1 Car Carrier Dataset .............................................. 27
      3.1.1 Data Cleaning and Processing ........................... 28
      3.1.2 Data Presentation .......................................... 29
      3.1.3 Data Compatibility ........................................ 31
   3.2 LNG Dataset ..................................................... 32
      3.2.1 Data Cleaning and Processing ............................ 32
      3.2.2 Data Presentation .......................................... 33
      3.2.3 Data Compatibility ........................................ 33

4 Setup of Problem ................................................... 35
   4.1 Models Used ..................................................... 35
4.1.1 Theoretical Model ........................................ 35
4.1.2 Multiple Regression ................................. 36
4.1.3 LASSO Regression ................................. 36
4.1.4 Sparse Regression ................................ 37
4.1.5 Fully Connected Neural Network ............... 37
4.1.6 Multiple Neural Network ......................... 38
4.1.7 Naive Theoretical and Neural Net ............. 39
4.1.8 Combined Models .................................. 40
4.1.9 Admiralty Neural Network ....................... 40
4.1.10 Physical Neural Network ......................... 41
4.2 Validating the Models ................................ 41
  4.2.1 Training and Validation ......................... 41
  4.2.2 Metrics Used .................................. 42
  4.2.3 Plots Used .................................. 43

5 Results .................................................. 45
  5.1 Car Carrier Dataset ................................. 45
    5.1.1 Theoretical Model ............................. 45
    5.1.2 Multiple and Sparse Regression ............. 48
    5.1.3 Neural Networks ................................ 50
    5.1.4 Combined Models ............................. 55
    5.1.5 Physical Models ............................. 57
  5.2 Adding Waves ....................................... 61
  5.3 LNG Dataset ......................................... 63

6 Conclusions ................................................. 69
List of Figures

2-1 Visual schematic of a fully connected neural network architecture. Many inputs $x_i$ are multiplied by weights $W$ and added into units, which pass them through an activation function $f$ before sending the output $a$ to the next layer or the output layer. Adapted from [1]...

3-1 Histograms showing the primary car carrier vessel’s data distribution of the quantities used in our models. Most quantities seem to follow a bell shaped distribution...

3-2 Histograms showing the secondary car carrier vessel’s data distribution of the quantities used in our models. Most quantities follow a bell shaped distribution...

3-3 Plot of the vessel and wind speed as measured by the vessel and weather forecaster datasets for the primary car carrier vessel. The x-axes show the (a) vessel speed and (b) wind speed measured from the ship, while the y-axes correspond to the vessel and wind speed as obtained from the weather forecasting dataset. The data is presented normalized. Note that, although the vessel speed is measured similarly between the two datasets, the wind speed has a lot of differences between them...

3-4 Histograms showing the LNG dataset’s data distribution of the quantities used in the models. Note that most quantities follow a bell shaped distribution...
3-5 Plot of the wind speed as measured by the LNG vessel and as predicted by the weather model. The x-axis shows the wind speed measured from the ship, while the y-axis is the wind speed as obtained from the company models. The wind data do not seem to match very well. . . .

5-1 Power curves for the primary vessel, as predicted by the theoretical model. The quantities plotted are all those used as inputs to the theoretical model. Note that all the quantities are smooth and have a physically intuitive relationship, as they are derived from physical equations.

5-2 Error plots for the predictions of the theoretical model on the primary vessel. The theoretical model tends to underpredict power at low speeds.

5-3 Box plot for the primary vessel using the theoretical model. The trend is evidently linear, but there is room for improvement.

5-4 Power curves for the secondary vessel, as predicted by the theoretical model.

5-5 Error plots for the predictions of the theoretical model on the secondary vessel. The theoretical model has large errors at low and high speeds.

5-6 Box plot for the secondary vessel using the theoretical model. The predictions are not as good as for the primary vessel, and there is a lot of uncertainty.

5-7 Power curves for the primary vessel, as predicted by the regression models. Pitch and days since hull cleaning have been added since these are also inputs to the regression models. Note that for the sparse regression model the relationships are similar to those predicted by the theoretical model, with the exception of trim.

5-8 Error plots for the predictions of the regression models on the primary vessel. Both models perform well at high speeds, but are less accurate at low speeds.
5-9 Box plot for the primary vessel using the two regression models. The multiple regression plot is curved rather than linear, reflecting an inability to fit non-linear data. .......................... 50

5-10 Power curves for the secondary vessel, as predicted by the regression models. The plots for both models are similar to those for the primary vessel. .......................... 51

5-11 Error plots for the predictions of the regression models on the secondary vessel. Both models perform well at high speeds, but are less accurate at low speeds. .......................... 51

5-12 Box plot for the secondary vessel using the two regression models. The multiple regression plot is curved rather than linear, reflecting an inability to fit non-linear data. .......................... 51

5-13 Power curves for the primary vessel, as predicted by the neural network models. The curves predicted by the FCNN are not always intuitive, though they follow the right trends. The curves for the MNN are much smoother since this model is trained for fewer epochs. .......................... 52

5-14 Error plots for the predictions of the neural network models on the primary vessel. The FCNN in particular has a very good error profile, with 75% of data points having an error below 10% for most speed ranges. .......................... 53

5-15 Box plot for the primary vessel using the two neural network models. Both models have a good performance, with the FCNN in particular having much lower variation from the straight line. .......................... 53

5-16 Power curves for the secondary vessel, as predicted by the neural network models. The curves predicted by the FCNN are similar to those predicted for the primary vessel. .......................... 54

5-17 Error plots for the predictions of the neural network models on the secondary vessel. The FCNN has a very good error profile, with 75% of data points having an error below 10% for most speed ranges. .......................... 54
Box plot for the secondary vessel using the two neural network models. Both models have a good performance, with the FCNN having much lower variation from the straight line.

Power curves for the primary vessel, as predicted by the combined models. The curves for the neural network models seem to converge to those of the FCNN.

Error plots for the predictions of the combined models on the primary vessel. The error profiles are similar to those of the FCNN.

Box plot for the primary vessel using the combined models. The neural network models match the linear trend well.

Power curves for the secondary vessel, as predicted by the combined models. The curves have the same shape as those of the primary vessel.

Error plots for the predictions of the neural network models on the secondary vessel. All neural network models appear to have a low error profile.

Box plot for the secondary vessel using the combined models. All neural network models have a behavior close to a straight line.

Power curves for the primary vessel, as predicted by the physical models. The curves predicted by the admiralty model are smoother.

Error plots for the predictions of the physical models on the primary vessel. Both models have very good error profiles, with 75% of data points having an error below 7.5% for most speed ranges.

Box plot for the primary vessel using the two physical models.

Power curves for the secondary vessel, as predicted by the physical models. The predicted curves for the two models are very similar.

Error plots for the predictions of the physical models on the secondary vessel. Both models have good error profiles, with 75% of data points having an error below 10% for most speed ranges.

Box plot for the secondary vessel using the two physical models.
5-31 Power curves for the primary vessel using the wave dataset. Note that the curves for waves are predicted correctly.

5-32 Error plots for the predictions of the best models on the primary vessel using the wave dataset. The error profiles are low for the neural network models, with those for the admiralty neural net below 5% most of the time.

5-33 Box plot for the primary vessel using the wave dataset. The plot for the admiralty coefficient in particular has very low variance.

5-34 Power curves for the secondary vessel using the wave dataset. Note that the curves for waves are predicted correctly.

5-35 Error plots for the predictions of the best models on the secondary vessel using the wave dataset. The error profiles are low for the neural network models, with those for the admiralty neural net below 5% most of the time.

5-36 Box plot for the secondary vessel using the wave dataset.

5-37 Power curves for the LNG vessel. These are mostly predicted correctly with the exception of waves and hull cleaning.

5-38 Error plots for the predictions of the best models on the LNG vessel. The regression models tend to have high errors at low speeds, while the neural networks are less accurate at high speeds.

5-39 Box plot for the LNG vessel. The trend follows the linear one closely for all models.
List of Tables

4.1 All parameters investigated for the neural networks. .......................... 38
4.2 Optimal parameters for the fully connected neural network. ............... 38
4.3 Optimal parameters for the multiple neural network. ......................... 39

5.1 Numerical results for the primary vessel, using the theoretical model. Note that the first column, MAE, is normalized by dividing by the maximum power to preserve confidentiality, and so is not in the units of kW, as it would normally be. ........................................ 46
5.2 Numerical results for the secondary vessel, using the theoretical model. 47
5.3 Numerical results for the primary vessel, using the two regression models. Sparse regression is clearly the one with the best performance. . 49
5.4 Numerical results for the secondary vessel, using the two regression models. Sparse regression is the model with the best performance. . 50
5.5 Numerical results for the primary vessel, using the two neural network models. The fully connected neural network has a much better performance than the rest of the models seen up to now. ............... 52
5.6 Numerical results for the secondary vessel, using the two neural network models. The fully connected neural network has a much better performance than the rest of the models seen up to now. ............... 53
5.7 Numerical results for the primary vessel, using the combined models. All models including neural networks perform equally well, with the naive model having the lowest MAPE. ................................. 55
5.8 Numerical results for the secondary vessel, using the combined models. The combined theoretical and neural network model has the best performance for this vessel.

5.9 Numerical results for the primary vessel, using the physical models. The admiralty model performs quite well on this vessel, with a MAPE error of only 3.5%.

5.10 Numerical results for the secondary vessel, using the two physical models. Both perform equally well, with the admiralty neural network being slightly better.

5.11 Numerical results for the primary vessel on the wave dataset. The results are similar to those for the full dataset without waves.

5.12 Numerical results for the secondary vessel on the wave dataset. The performance is significantly increased, especially for the admiralty neural net.

5.13 Numerical results for the LNG vessel. The performance of the neural networks is lower than for the other vessels, but on par with the rest of the models.
Chapter 1

Introduction

International shipping is one of the leading sources of greenhouse gas (GHG) emissions, accounting for about 2.5% of total GHG emissions [17]. As part of a global effort to reduce emissions and to combat climate change, the International Maritime Organization (IMO) has pushed to reduce carbon dioxide emissions from shipping by 40% by 2030 and 70% by 2050 [9]. These ambitious reduction targets present one of the biggest challenges facing ship builders, operators, and owners in the coming decades. While many technologies need to be developed to achieve those goals, one approach is the use of data-driven methods to find optimal conditions that minimize fuel consumption and emissions. This approach is becoming more and more feasible as new laws by the IMO have imposed data collection protocols on vessel operators, and as a result many shipping companies have collected a large amount of data on their vessels [18]. One of the most promising approaches to analyze this data is using machine learning, a field that has been continuously evolving over the last years and that is now being applied to more and more problems globally.

The focus of this project is to use machine learning methods to predict the shaft power of a vessel at different operating conditions, and to compare its performance to other data-driven methods, such as regression or to other more theoretical models. The rationale behind this investigation is that once an accurate model is developed, it would be straightforward to extend it to predict fuel consumption or carbon emissions. This tool can then be used to determine the ship parameters that could minimize...
emissions, whether by changing the trim, by setting the optimal speed, by selecting the most favorable voyage path or by estimating the best time for dry-docking.

1.1 Related work

There have been various efforts at using machine learning for ship performance prediction and optimization. Most of them attempt to predict shaft power or fuel consumption at different operating conditions, and to subsequently calculate the optimal ship voyage. The vast majority of papers use some forms of neural networks. Other models have been attempted as well such as multiple regression, LASSO regression, decision trees, and Gaussian processes.

First of all are the simple data-driven approaches. Kee et al. [11] use multiple linear regression to predict fuel consumption of tugboats. Using data on speed, wind, displacement, and distance traveled, they get an $R^2$ of 90%.

Moving on to papers that use neural networks, Parkes et al. [15] used 27 months of data on speed, draft, trim, wind speed, and waves to create a neural network to predict shaft power. After optimizing, the neural network yielded a mean relative error of 7.8%, on par with multiple regression which serves as a baseline because of its wide use in the industry. The authors also provide plots of the predicted curves for each quantity, with the speed power curve being predicted relatively well by the neural network.

Kim et al. [12] used six months of data from a container ship to predict fuel efficiency given speed, draft, trim, wind, and a few other quantities. They also mostly investigated neural networks and linear regression, and found that a properly tuned neural network can achieve an $R^2$ of up to 99%.

Radonjic et al. [16] predicted towboat shaft power. As opposed to the other analyses, this was done in a river, under controlled circumstances with no wind or waves, and used different towboat characteristics like beam and draft to estimate shaft power. The two models investigated were neural networks and ensemble neural networks, with the latter performing the best and achieving a mean relative error
mostly around 1-5%.

Besikci et al. [2] used 17 months of data from an oil tanker and developed a neural network which performed better than multiple regression, with an $R^2$ of around 80%.

Yuan et al. [22] used a month of data from a cargo river inland vessel to predict fuel consumption. They developed a neural network that achieved a mean relative error of 15% and an $R^2$ of 98%. They also compared various other machine learning methods and found that the neural network had the best performance.

There is also a large collection of papers that use different machine learning approaches in addition to neural networks.

Gkerekos et al. [6] used a variety of machine learning models to predict fuel oil consumption using speed, draft, trim, wind, current, and sea state, and found that extra tree regression, random forest regression, and neural networks had the best performance, yielding an $R^2$ above 90%.

Jeon et al. [10] also estimated fuel consumption. Their data included information on draft, trim, shaft power, wind speed, and vessel speed. They found that neural networks have a much better performance than polynomial regression or support vector machines, with an $R^2$ above 90%.

Wang et al. [19] mainly investigated LASSO regression to predict fuel consumption. They used a very large dataset from about 100 ships, with quantities including ship characteristics, speed, trim, wind, and waves. They concluded that a LASSO regression has better performance than neural networks, support vector regression, and Gaussian processes.

Yuan et al. [21] also used a Gaussian process metamodel to predict fuel consumption using speed, draft, trim, wind, and waves, and found it had an acceptable performance.

Hu et al. [8] used a year of data from a container ship with information on speed, draft, trim, wind, and waves. After running both neural networks and Gaussian process regression, they found that the two models resulted in similar performance, achieving an $R^2$ about 98%, with the Gaussian process taking much longer to train.

In a series of two papers, Liang et al. [13] and Liang et al. [14] used a big dataset
from a very large number of ships with two years of data to estimate vessel propulsion power. They used a variety of quantities including among others, vessel speed, draft, trim, wind, and waves. They found that physics based models do not perform as well as data-driven ones, and that most machine learning models, such as decision trees, support vector machines, and neural networks, achieve a similar performance, with an $R^2$ of about 80%.

Finally, there are a few papers which attempt to blend data-driven models with some physical intuition.

Yang et al. [20] employed genetic optimization algorithms that incorporate some physical insights in their models to predict fuel consumption. One of these insights is the use of the admiralty coefficient as a means of relating the data in a physically relevant way. Using seven years of data from a vessel with speed, draft, and wind as inputs, they were able to obtain a mean relative error of about 7%.

Lastly, Coraddu et al. [4] used data from a tanker with a large amount of measurements to predict shaft power, shaft torque, and fuel consumption. They used three different approaches. The first, called white box model, is a theoretically derived model. The second, named black box model, is a collection of machine learning techniques, mostly multiple and LASSO regression and random forests. The final, the gray box model, combines the first two by using the results of the white box model in the black box model. In the end they found that the gray box models had a superior performance.

The above literature review leads to several conclusions. First of all, there is no consensus on which model has the best performance, with multiple regression, LASSO regression, neural networks, Gaussian processes, genetic algorithms, and gray box models all being found to perform well. Second of all, there are various ways of estimating the performance of different models, such as mean relative absolute error, mean absolute error, and $R^2$. In terms of visual results, box plots of predictions vs. measurements are used in the majority of cases. Finally, most of the papers use information on vessel speed, draft, trim, wind speed, and waves to predict fuel consumption.
1.2 This Study’s Contribution

The purpose of this thesis is to further efforts made in developing accurate models to predict vessel shaft power. As such, a variety of methods widely used in the literature, such as multiple regression, LASSO regression, neural networks, ensemble neural networks, theoretical models, and combinations of the above are all investigated on different datasets. Some of the novel contributions of this work include the addition of hull cleaning as one of the input variables, the development of a piecewise neural network that can incorporate some physical insight, use of combinations of the above models, methods to quantify their uncertainty, as well as the use of different datasets to test the robustness of models whose parameters have been selected based on a single dataset.
Chapter 2

Background

This section provides some background that will be useful to understand the models explored in this research.

2.1 Multiple Regression

The simplest approach to finding relationships from data is using linear regression, and its extension to multiple variables known as multiple regression and also known as Ordinary Least Squares (OLS) regression. This essentially tries to find the line of best fit, that is to solve the problem \( \min_\beta (Y - X\beta)^2 \), where \( Y \), the dependent variable, is a column vector containing all the data points, \( X \) is a matrix whose columns correspond to different independent quantities, and whose rows correspond to different data points, and \( \beta \) is a vector of the coefficients giving a hyperplane (or line in 2D) that fits the data best. By optimizing the above problem, it can be shown that the optimal coefficients can be found using the formula \( \beta = (X^T X)^{-1} X^T Y \). This gives the line of best fit, and once the optimal parameter \( \beta \) is found it can be used for error estimation and predictions.

One technique commonly used in linear regression is feature transformation. If we know that the dependence of \( Y \) on \( X \) is nonlinear, and is for example square, we can regress \( Y \) on \( X^2 \) and obtain a coefficient that fits the line \( Y = \beta X^2 \), which gives a quadratic fit. This is an easy modification to add non-linearity to linear regression.
2.2 LASSO Regression

Least Absolute Shrinkage and Selection Operator (LASSO) regression is similar to linear regression in that it also tries to minimize the squared error of predictions and measurements. The difference is that in LASSO regression we want to do this while keeping the matrix of coefficients $\beta$ sparse, that is with few non-zero entries. The idea is to add many different parameters to the regression, and through LASSO regression find the most important ones that fit the data. The LASSO model can thus be written as $\min_\beta (Y - X\beta)^2 + \lambda \|\beta\|_1$, where $\lambda$ is a parameter that controls how sparse we want $\beta$ to be [5].

2.3 Neural Networks

Neural networks are a more advanced version of regression. They also have parameters that they find by minimizing a function, but in contrast to regression, are nonlinear and do not have a closed form solution.

The basis of a neural network is the unit. A unit is basically a function that takes a sum of weighted inputs $\sum X_i w_i$ and passes them through an activation function, with the final output being $z = f(\sum X_i w_i)$. These $w_i$ are the parameters of the neural network, similar to the coefficients $\beta$ of linear regression, and are also the quantities that the model tries to fit to the data. The activation function used in neural networks varies. It can be linear, which would reduce it to multiple regression, or non-linear, which is used in order to fit non-linear functions. Common activation functions used are the ReLU, tanh, and sigmoid functions. A lot of units can be stacked together to form a layer, and then more layers can be added, taking as inputs the outputs of the previous layer until we reach the output layer. An example configuration of units and layers is shown in figure 2-1. This collection of units and layers creates many different weights that need to be fitted. Since there is no closed form solution to find those
weights, an iterative algorithm such as gradient descent is typically used to find the optimal values. There are many such algorithms, with the most popular being Adam and Stochastic Gradient Descent. The error metric neural networks use to ensure a good fit varies. Typically, it is mean squared error, that is the squared difference between measurements and predictions, exactly like for multiple regression.

Similar to LASSO regression, regularization can be added to the loss function, to ensure the weights found are either small when using L2 regularization, or sparse when using L1 regularization.

Another important setting of neural networks are epochs. As explained above, the parameters of a neural network are optimized iteratively, and epochs control the number of iterations. Too few iterations will not lead to convergence, while too many can lead to overfitting, both of which decrease the usefulness of the neural network.

The above describes a fully connected neural network, i.e. one in which a layer takes every unit from the previous layer as inputs. Since this can lead to too many parameters and overfitting, many times dropout layers are added that remove some of the connections between layers, potentially leading to more robust parameters [1].
2.3.1 Ensemble Neural Network

Neural networks have randomness in the way they are trained, mainly because of their random initialization. In order to remedy this problem and create smoother solutions ensemble neural networks (ENNs) are used. These basically consist of many different neural networks trained on the same data using different initialization parameters. The predictions ENNs make are just the mean of the predictions of their constituent neural networks. A good example of the use of ENNs for this problem can be found in Radonjic et al. [16], where they were found to perform better than individual neural networks.

2.4 Theoretical Model

The above models are data-driven methods, in that they derive their parameters from a best fit of the data. Theoretical models relating to our problem are much harder to obtain due to the absence of exact equations describing the motions of ships. There have been however some attempts, and this work uses the model developed by Chalfant et al. [3] to estimate the shaft power of a vessel at different conditions.

The main idea behind this theoretical model is to first compute the total resistance of the vessel, 
\[ R_{Total} = R_{friction} + R_{wave} + R_{air} \], where each resistance component is found using empirical formulas based on the hull characteristics of the vessel. The effective power can then be computed as 
\[ P_e = R_{Total} \cdot V \], where \( V \) is the vessel speed. Finally the shaft power is calculated as 
\[ P_s = \frac{P_e}{\eta_o \eta_h \eta_r \eta_s} \], where \( \eta_o, \eta_h, \eta_r, \eta_s \) are various efficiencies which can be calculated using empirical formulas from the vessel characteristics.
Chapter 3

Data

Two datasets are used in this project. The main dataset is from two car carrying sister vessels and is collected from September 2014 to February 2022. This dataset was used to develop the methodology for the project. The second dataset is from an LNG vessel collected from September 2016 to March 2022, and is used to test the models developed using the main dataset. The following sections present the full details of each dataset.

3.1 Car Carrier Dataset

The car carrier dataset is the one used for most of the project, to develop all the methods presented and to obtain the first results. It comes from car carrying sister vessels with one variable pitch single screw propeller each. One of the two vessels is used for all the model tuning, and is labeled as the primary vessel in this report, while the other is used for validation and is labeled as the secondary vessel. The data consists of 237,914 and 256,022 data points for the primary and secondary vessels respectively, collected every 15 minutes from September of 2014 to February of 2022. Various quantities are collected, with the ones used in this paper including speed from GPS, speed from the ship log, shaft power, draft fore, draft aft, relative wind speed, relative wind direction, and propeller pitch.

In addition, for the years 2021 and 2022 the following data, measured from a
weather company, are also available for both vessels: significant wave height, wind angle, wind speed, vessel course angle, and vessel speed. These data are measured every 10 minutes.

Finally, we also have information on the ships’ characteristics, such as displacement, coefficients, etc. and data on dry docking, hull cleaning, and propeller cleaning events.

3.1.1 Data Cleaning and Processing

Real world data is very noisy, both because of the many different conditions encountered by the vessel and because of sensor imperfections. As such one of the most important steps in running a successful model is the development of a thorough data cleaning protocol.

The first cleaning step is based on the sea state. Every point is described by a sea state, which is either “Sea Passage,” “Maneuvering,” or “At Birth.” The latter two were discarded, as we are only interested in operating state predictions. This is a significant step as it removed nearly half (45%) of the data points.

The next cleaning step is the removal of all data points that are missing at least one of the important quantities. Important quantities are those that are used by at least one of our models, and include the log vessel speed, draft forward and aft, wind speed and direction, shaft power, rudder angle, and propeller pitch. For the same reason, and in order to exclude conditions where the vessel is moving in ports or is stationary, speeds below 5kn were discarded.

We also performed a visual check on the resulting data by plotting speed vs. power, and removed the periods which display behavior that is unreasonable. Overall this resulted in removing two months of data showing maximum power measured at minimum speed, unlikely in the real world.

Even after the above processing, there were a lot of outliers, and also regions with few data points, such as at extreme draft conditions. Since our models would have a hard time to predict extreme conditions with limited data, we performed an additional cleaning operation by removing the top and bottom x% of the data points.
for every important quantity, where $x$ can be varied depending on the model used but is usually set at 5%. This step also cut a lot of data points depending on $x$, and for a value of 5% cut about half of the remaining points.

Apart from removing data points, we also combined some of the quantities into more physical ones before inputting them into our models. As such, we defined draft as the average between draft forward and draft aft, trim as draft aft minus draft forward, and wind speed as $\text{Wind speed} = \text{Rel wind speed} \cdot \cos(\text{Rel wind direction})$, so that the wind speed we used is the component that is in the direction of the vessel.

One important factor that is not accounted by the measurements collected is the fouling of the hull. In order to account for this, an additional quantity was created to measure the number of days that have passed since the last hull cleaning event or since the vessel was first deployed.

Finally, when using only the data from 2021 onward, we also added the wave height measure from the weather forecast dataset. Since this is collected every 10 minutes as opposed to the 15 minute frequency of the ship data, we used linear interpolation to obtain the wave height estimate at the times when the ship data is collected. This was then added to the ship data, and any missing values were discarded.

### 3.1.2 Data Presentation

Figures 3-1 and 3-2 show the histograms of the most important data for the primary and secondary car carrier vessels. The data is presented after filtering out null values and also non-operational non-sea passage data points, as explained above. Note that the $x$-axis showing the values of the quantities has been normalized by dividing by the maximum in order to preserve the confidentiality of the data.

A first thing to note is that the histograms for both sister vessels look remarkably similar, as would be expected. The second observation is that most quantities have a bell shaped distribution, with the number of data points away from the mean being small. Since we want our model to have enough data at every operating condition to make predictions, we also filter out the data points at the tails of the distributions, as discussed above.
Figure 3-1: Histograms showing the primary car carrier vessel’s data distribution of the quantities used in our models. Most quantities seem to follow a bell shaped distribution.

Figure 3-2: Histograms showing the secondary car carrier vessel’s data distribution of the quantities used in our models. Most quantities follow a bell shaped distribution.
Figure 3-3: Plot of the vessel and wind speed as measured by the vessel and weather forecaster datasets for the primary car carrier vessel. The x-axes show the (a) vessel speed and (b) wind speed measured from the ship, while the y-axes correspond to the vessel and wind speed as obtained from the weather forecasting dataset. The data is presented normalized. Note that, although the vessel speed is measured similarly between the two datasets, the wind speed has a lot of differences between them.

3.1.3 Data Compatibility

As mentioned above, for both car carrier vessels and for year 2021, we had access to an additional dataset obtained from a weather forecasting company which included information on the vessel’s speed as well as the wind and waves it is experiencing.

In order to determine the compatibility of this new dataset with that obtained from the vessel, we plotted the quantities measured by both datasets, which are the vessel speed and wind speed (in the direction of the vessel), in figure 4. If it were the case that the two datasets measure the same value at the same time, then the plot of the two data sources would be a straight line. While this is largely true for the vessel speed, the relationship is not as good for the wind speed, and although the relationship is linear, it is very noisy. This could be due to a variety of factors, such as the discrepancy in the collection rates (the ship collects data every 15 minutes while the forecaster every 10 minutes) and subsequent interpolation, or due to different sensors and methods of estimation. This decreases our faith in the weather dataset, and although it is used in some of the analysis, it is considered a low fidelity dataset.
3.2 LNG Dataset

The purpose of the LNG dataset was to test the performance of the models developed on an entirely different type of vessel.

It is taken from a liquefied natural gas (LNG) twin screw vessel. It consists of 43,621 data points measured at one hour intervals from September 2016 to March 2022. The quantities that are measured include the following: ship speed from GPS, ship speed from the ship log, draft forward, draft aft, trim, wind speed, relative wind direction, ship course, shaft 1 power, and shaft 2 power. Along with the above data, which is measured by sensors, we are provided with data calculated using company models, which importantly include the significant wave height, as well as wind speed and direction.

Apart from the above quantities, we additionally have information on the LNG vessel’s displacement, center of buoyancy, characteristic coefficients etc. at different draft values, as well as information on the ship’s two propellers. We are also provided with the date when the vessel performed hull cleaning in this time period.

3.2.1 Data Cleaning and Processing

The cleaning methods for the LNG dataset resemble those used for the car carrier dataset as much as possible. As such, any points from which some of the measured quantities were missing, as well as any points which had erroneous data and were labeled so in the dataset, were removed, and speeds below 5kn were discarded. In addition, the bottom and top 5% of the distribution for each measure were filtered out in order to make predictions in regions with enough data points.

Some of the data were then combined to create more meaningful inputs to our models. Draft is defined as the average between draft forward and draft aft. Wind speed is defined as $\text{Wind Speed} = \text{Rel wind speed} \cdot \cos(\text{Ship course} - \text{Wind direction})$, so that it is in the direction of the ship. Finally, the same measure for the hull condition used for the car carrier dataset, days since last hull cleaning, was created and added to the LNG dataset.
Figure 3-4: Histograms showing the LNG dataset’s data distribution of the quantities used in the models. Note that most quantities follow a bell shaped distribution.

3.2.2 Data Presentation

Figure 3-4 shows histograms of the quantities most widely used from the LNG dataset. They include the pre-processing step removing null values, and the transformation to physically meaningful quantities for the case of draft and wind speed. As with the car carrier dataset the x-axis showing the values of the quantities has been normalized by dividing by the maximum in order to preserve the confidentiality of the data.

As with the car carrier dataset, most quantities have a bell shaped distribution, justifying filtering out the data points at the tails of the distributions, as discussed above.

3.2.3 Data Compatibility

As was the case for the car carrier dataset, the LNG vessel does not collect data on waves, and these are obtained from company models. Since wind speed is also predicted by these models, we run a compatibility check between the wind speed measured by the vessel and that obtained from the company predictions. The resulting plot is shown in figure 3-5. We can see that the predictions vary a lot from the measurements. This could be due to the fact that the data are collected hourly, and the wind can change speed and direction a lot over an hour. The model predictions
Figure 3-5: Plot of the wind speed as measured by the LNG vessel and as predicted by the weather model. The x-axis shows the wind speed measured from the ship, while the y-axis is the wind speed as obtained from the company models. The wind data do not seem to match very well.

are thus considered of low fidelity, and the wave predictions are the only ones used because of their importance in predicting shaft power.
Chapter 4

Setup of Problem

This section describes the setup of the problem including details on all the models run, as well as the ways their performance is judged, both numerically and visually.

4.1 Models Used

4.1.1 Theoretical Model

The first model implemented is a purely theoretical model (TM), dependent on theoretical formulas, derived either rigorously, empirically, or numerically (CFD software), to make predictions, and none of its parameters are determined by the data.

More details on the origin and features of this model can be found in section 2. Once implemented, the way this model is used to make a prediction at a particular data point, is to take the vessel speed, wind speed, draft, and trim and interpolate draft and trim based on the ship’s tables to find its parameters (e.g. displacement, block coefficient, wetted surface) at these conditions, and then to use empirical and theoretical equations to make a power estimation. Note that the theoretical model does not take into account hull cleaning, propeller pitch, and waves, and is intended only for approximate calculations, so it is expected to be the least accurate among the models presented here.
4.1.2 Multiple Regression

In contrast to the theoretical model, data-driven models use data to fit all their parameters. Data-driven models were used extensively in this work, and they lie at the heart of machine learning. The first data-driven model used in this analysis is multiple regression (MR). This represents the most basic data-driven model, and only runs a linear regression on the input parameters:

\[
\text{Power} = \beta_0 + \beta_1 \text{Speed} + \beta_2 \text{Wind Speed} + \beta_3 \text{Draft} + \beta_4 \text{Trim} \\
+ \beta_5 \text{Days since hull cleaning} + \beta_6 \text{Pitch} + \beta_7 \text{Wave height} \quad (4.1)
\]

where \( \beta_i, i = 0, \ldots, 7 \) are constants to be determined by the data. Since the relationship between power and the input variables is known to be non-linear, this model is not expected to fit the data well.

4.1.3 LASSO Regression

LASSO regression (LR) is mainly used in order to find which of the quantities used are most important, as was detailed in the background section. It was run using three different input categories, and the most important parameters were then singled out and used as inputs to multiple regression. The optimal parameters of the LASSO regression were found using cross-validation on the training set. The dataset used was the full dataset on the primary and secondary car carrier vessels, so wave data was not available.

The first input set included all the six main input quantities, namely log speed, draft, trim, wind speed, hull cleaning, and pitch, raised to the power 1, 2, and 3.

The second input set included all the six input parameters raised to the power 1 and 2, and multiplied with each other. So for example, if speed is denoted \( V \) and draft \( T \), this set would include \( V, T, VT, V^2T, VT^2, V^2T^2 \).

The third input set included all the six input parameters raised to the power 0.5,
1, 1.5, 2, 2.5, and 3.

After running LASSO regression on these three sets, for the two car carrier vessels, the quantities selected the most by LASSO regression were wind speed to the power 3, draft, trim, hull cleaning, pitch to the power 3, and vessel speed to the power 2, 3, or 2.5.

These inputs were then used in multiple regression, where it was found that raising pitch to a cubic power did not change the results much, and vessel speed was found to have better results for powers of 2.5 and 3. Since speed cubed is the relationship predicted by theory, this was used in the multiple regression. As such, the final inputs to multiple regression were vessel speed cubed, wind speed cubed, draft, trim, pitch, days since hull cleaning, and significant wave height (when available), as described in the next section.

4.1.4 Sparse Regression

Sparse Regression (SR) refers to the multiple regression augmented based on the results of LASSO Regression. It includes the same inputs as MR, with the difference that vessel speed and wind speed are added to a cubic power. Sparse regression is thus expected to perform much better than multiple regression, as it transforms the inputs in a way that most closely matches the data. The setup for sparse regression is the following:

\[
\text{Power} = \beta_0 + \beta_1 \text{Speed}^3 + \beta_2 \text{Wind Speed}^3 + \beta_3 \text{Draft} + \beta_4 \text{Trim} \\
+ \beta_5 \text{Days since hull cleaning} + \beta_6 \text{Pitch} + \beta_7 \text{Wave height} \quad (4.2)
\]

4.1.5 Fully Connected Neural Network

The fully connected neural network (FCNN) is a basic neural network, as described in chapter 2. The difficulty in creating it is due to its many settings that need to be tuned to provide an adequate performance for the task at hand. In order to
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epochs</td>
<td>[1,1000]</td>
</tr>
<tr>
<td>Units</td>
<td>[1,200]</td>
</tr>
<tr>
<td>Layers</td>
<td>[1,100]</td>
</tr>
<tr>
<td>Activation function</td>
<td>ReLU, Tanh, Sigmoid, Linear, Softmax, Softsign, Elu</td>
</tr>
<tr>
<td>Regularization</td>
<td>L1,L2</td>
</tr>
<tr>
<td>Regularization parameter</td>
<td>[0,1]</td>
</tr>
<tr>
<td>Loss</td>
<td>Mean Squared Error, Mean Average Error</td>
</tr>
<tr>
<td>Optimizer</td>
<td>Adam, SGD, RMSprop, Adagrad, Adamax, Nadam</td>
</tr>
<tr>
<td>Dropout</td>
<td>[0,0.9]</td>
</tr>
</tbody>
</table>

Table 4.1: All parameters investigated for the neural networks.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epochs</td>
<td>500</td>
</tr>
<tr>
<td>Layers</td>
<td>2</td>
</tr>
<tr>
<td>Units</td>
<td>10</td>
</tr>
<tr>
<td>Activation function</td>
<td>ReLU</td>
</tr>
<tr>
<td>Regularization</td>
<td>L1</td>
</tr>
<tr>
<td>Regularization parameter</td>
<td>0.001</td>
</tr>
<tr>
<td>Loss</td>
<td>Mean Average Error</td>
</tr>
<tr>
<td>Optimizer</td>
<td>Adam</td>
</tr>
<tr>
<td>Dropout</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 4.2: Optimal parameters for the fully connected neural network.

create the optimal network, many different combinations of parameters were run and the results checked on the validation set of the primary vessel. The entire range of parameters investigated are shown in table 4.1. The parameters that yielded the best performance are shown in table 4.2. Note that before being used in the FCNN, the data was normalized by subtracting the mean and dividing by the standard deviation. In addition, to ensure results robust to different initializations an ensemble of five FCNNs was used for this model.

4.1.6 Multiple Neural Network

The multiple neural network (MNN) is a modification of the FCNN for this particular problem. The idea behind it is that the power experienced by the vessel is a function
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epochs</td>
<td>10</td>
</tr>
<tr>
<td>Layers</td>
<td>2</td>
</tr>
<tr>
<td>Units</td>
<td>20</td>
</tr>
<tr>
<td>Activation function</td>
<td>ReLU</td>
</tr>
<tr>
<td>Regularization</td>
<td>L1 for Hull Cleaning</td>
</tr>
<tr>
<td>Regularization parameter</td>
<td>0.001</td>
</tr>
<tr>
<td>Loss</td>
<td>Mean Squared Error</td>
</tr>
<tr>
<td>Optimizer</td>
<td>Adam</td>
</tr>
<tr>
<td>Dropout</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 4.3: Optimal parameters for the multiple neural network.

of many different parameters, some of which are correlated and some of which are uncorrelated. We can then split up the dependence on the different parameters into piecewise components. To give an example, suppose power is dependent on speed \( V \), draft \( T \), and wind speed \( W \), and speed and draft are correlated. We can then split up \( P = f(V, T, W) = f_1(V, T) \cdot f_2(W) \) and fit separate neural networks to \( f_1 \) and \( f_2 \). The advantage of these MNNs is that they have far fewer parameters by incorporating the knowledge that some parameters are physically uncorrelated and don’t need to be trained. This makes training easier and simplifies the model.

The parameters considered coupled were vessel speed, draft, and trim, while wind speed, days since hull cleaning, pitch, and waves (where available) were added as independent components.

As for the FCNN, the same range of parameters were investigated for the MNN, and the best ones are shown in table 4.3. For the MNN the data was normalized by dividing by the maximum. As for the FCNN, an ensemble of five MNNs was used to increase robustness.

### 4.1.7 Naive Theoretical and Neural Net

All of the above models apart from the theoretical one are pure data-driven models. While those have the advantage of fitting the data closely, they tend to produce results that are difficult to interpret. To address this, various models have been developed
that try to incorporate some physical intuition into the neural networks.

The first such model, labeled naive theoretical and neural network (NTNN) or just naive, is adapted from Coraddu et al. [4]. It works by simply using the theoretical model to make a first prediction, which is then inserted as an additional input to a FCNN.

4.1.8 Combined Models

A more elaborate way to combine the theoretical with the data-driven models is used in our so-called combined models, inspired from the grey box models of Coraddu et al. [4]. The way these models work is the following: first a model from the ones above is run on the input data and makes a first prediction, which has an error with respect to the measurement. The error is then fed into a different model from the ones above, so in a sense the second model tries to fit the errors of the first one. Different combinations are used, and the ones employed in this study include using SR as a first model and MNN as a second (SRMNN), using TM as a first model and MNN as a second (TMMNN), and using TM as a first model and SR as a second (TMSR).

4.1.9 Admiralty Neural Network

While the above combined models combine data-driven and theoretical approaches, they still do not quite add physical intuition into the neural networks, and the second model typically tends to dominate the results. Another more physically grounded model depends on using the admiralty coefficient, inspired from Yang et al. [20].

The admiralty coefficient is defined as \( A = \frac{\nabla^{2/3}V^3}{P} \), where \( \nabla \) is the displacement, \( V \) the ship speed, and \( P \) the power. By rewriting this as \( P = \frac{\nabla^{2/3}V^3}{A} \), since \( A \) is assumed constant for a vessel, we get a new quantity \( B = \nabla^{2/3}V^3 \) which combines speed, draft, and trim into a physically meaningful quantity that power depends on and that we can add as an input to our models. In particular, we can calculate \( \nabla \) by interpolation from tables that give its dependence on draft and trim for a particular
vessel. Note that this is not really a separate model, but consists of the FCNN with the speed, draft, and trim inputs replaced by the admiralty coefficient input $B$. This is then labeled the admiralty neural network (ANN) in this work.

### 4.1.10 Physical Neural Network

Finally, an additional model that aims to use some more physical insight is developed and named physical neural network (PNN). This is again a neural network, but instead of fitting the power to the inputs, it fits the resistance coefficient, defined as $C = \frac{P}{\frac{1}{2} \rho S V^3}$, where $P$ is the power, $\rho$ the density of water, $S$ the wetted surface (obtained by interpolating the draft and trim using ship tables) and $V$ the speed. Since this coefficient depends on the speed, draft, trim, hull condition, and wind speed, this is a more physically intuitive parameter for the neural network to fit. The predicted coefficients can subsequently be multiplied by $\frac{1}{2} \rho S V^3$ to obtain the final power predictions.

### 4.2 Validating the Models

#### 4.2.1 Training and Validation

Most of the models described above are data-driven, and require data to determine their parameters. However, a problem arises when these models fit the data set too closely and do not generalize to other datasets. As such, when using any of the above models, the data is always split into a training and a test set. The training set is used to train the parameters of each model, while the test set is used to determine how well it performed.

For this study the data was always split with 80% of the data kept for training and the remaining 20% used for testing. The same deterministic split, where the validation points correspond to the 40% to 50% and 80% to 90% cut of the data, is solely used in this study. Additional splits were attempted but they yielded similar results so this standard cut is used here for consistency.
Note that the way the problem is set up, the models are always trained on the training dataset, but the hyperparameters of some models, such as the neural networks, are chosen based on the performance on the test set. This is where the secondary vessel comes into play. As for the primary vessel, the data on the secondary vessel is also split into training and testing, and the models for that vessel are trained on the training set using the hyperparameters found from the primary vessel. The test set of the secondary vessel is seen for the first time by all models, and thus represents a good estimate of the performance of the models.

4.2.2 Metrics Used

Various metrics are used to judge the performance of the different models. These include the mean absolute error, mean absolute percentage error, and $R^2$. All of them are evaluated on the test set.

Mean Absolute Error

The first metric used to gauge the performance of the various models is the mean absolute error (MAE). This is computed as $\text{MAE} = \frac{\sum |P_{\text{mes},i} - P_{\text{pred},i}|}{N}$ and is preferred to mean squared error since it has the same units as power, and can thus offer an intuitive number to understand the error margin of the ship power. The downside of this metric is that it is biased to high speeds, which correspond to a higher power and thus higher absolute errors. In this paper the MAE error is always divided by the maximum power to preserve confidentiality.

Mean Absolute Percentage Error

The next metric, the mean absolute percentage error (MAPE), is defined as the relative error, i.e. $\text{MAPE} = \frac{1}{N} \sum_i \frac{|P_{\text{mes},i} - P_{\text{pred},i}|}{P_{\text{mes},i}}$. This metric describes how large the absolute error is relative to the actual power at that point, and so attempts to remedy the issue of the MAE weighting higher powers more, while at the same time it gives a good non-dimensional number that can be used to compare the models’ performance.
for different datasets. This metric is problematic for low vessel speeds, as it can result in division by very small numbers, but this is not a problem when low speeds are filtered out, as in this analysis.

R-Squared

R-squared \( (R^2) \) is a metric used to determine how close the predictions vs. measurements are to a straight line. It is calculated as \( R^2 = 1 - \frac{\sum_i (P_{\text{meas},i} - P_{\text{pred},i})^2}{\sum_i (P_{\text{meas},i} - P_{\text{ave}})^2} \), where \( P_{\text{ave}} \) is the average measured power. This measure is also used to provide a good performance metric that is not biased towards high speeds. Its disadvantage is that it is not as intuitive compared to the other metrics used, due to the more complicated calculations needed to obtain it.

4.2.3 Plots Used

One type of model verification that hasn’t received much attention is plotting. This is the most straightforward way to determine whether the results predicted by the models, some of which can be very complicated, make sense. In this work plots were used extensively to judge the performance of the different models, in addition to the above metrics. For this reason, the types of plots used are described in detail here.

Power Curves

First of all are the power curves, also used in Parkes et al. [15]. These are two dimensional plots of power vs. one of the inputs, and plot the power each model predicts throughout the range of that input, setting all the other inputs to their average value. For example, a power curve for speed would plot the power predicted at the entire speed range for every model, setting draft, trim etc. to their average value. An example of this plot is shown in figure 5-1.
Error Plot

A metric that has not received much attention in the literature for this problem is uncertainty. In this work we attempt to quantify and visualize the uncertainty of the various models through error plots.

These essentially plot the average error with error bars as a function of speed. The way this plot is created is by calculating the error (either MAE or MAPE) of every validation input, and then sorting it into bins based on the velocity of the ship at that point. The bin interval lengths can be varied, but for this study are 0.1kn intervals, which roughly give 50 points per bin for low speeds and 500 points per bin for high speeds, using the standard dataset. Then for each bin the average, 25% and 75% percentiles, and the resulting plot includes the average error at each bin with the 25% and 75% percentiles as error bars. This way we can visually check whether different models have a consistent error or one that varies greatly. An example plot is shown in figure 5-2.

Box Plot

The third type of plot is more widely used and plots power predicted vs. power measured for all the validation points, also referenced as a box plot in this study. The idea is that for good models, all points should lie on a straight line. This type of plot is thus closely related to the $R^2$ metric. An example is shown in figure 5-3.
Chapter 5

Results

This section presents the results from running all the models described in the previous sections on all the datasets. The way it is structured is first the results for all models on the car carrier dataset are presented, followed by the main dataset augmented with waves and finally the LNG dataset.

5.1 Car Carrier Dataset

We start with the car carrier dataset, which includes the primary and secondary car carrying vessels. The results for both ships are presented in this section.

5.1.1 Theoretical Model

The first model run is the theoretical model. Being theoretical, this model uses no information from the data to determine its parameters, as explained previously. As such, it is expected that this model is the least accurate. However, it can serve as a good baseline to compare the data-driven models to.

First of all, the results of the theoretical model on the primary vessel are given in table 5.1. The results are good, with an 11.9% mean average percentage error and an $R^2$ of 85%. In addition, the power curves predicted by this model, shown in figure 5-1, are of the exact shape predicted by theoretical models. These plots will thus be
Table 5.1: Numerical results for the primary vessel, using the theoretical model. Note that the first column, MAE, is normalized by dividing by the maximum power to preserve confidentiality, and so is not in the units of kW, as it would normally be.

<table>
<thead>
<tr>
<th>Theoretical</th>
<th>MAE (kW/kW)</th>
<th>MAPE (%)</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.073</td>
<td>11.9</td>
<td>85</td>
</tr>
</tbody>
</table>

Figure 5-1: Power curves for the primary vessel, as predicted by the theoretical model. The quantities plotted are all those used as inputs to the theoretical model. Note that all the quantities are smooth and have a physically intuitive relationship, as they are derived from physical equations.

considered the most “correct” ones when comparing with the rest of the models. The error plots, displayed in figure 5-2, show the main drawbacks of the theoretical model, namely that it is not calibrated for this particular ship and thus tends to have high errors at low and high speed regimes. Finally, the box plot for this model, shown in figure 5-3, shows a generally correct trend, with lots of room for improvement.

Moving on to the secondary vessel, the results of the theoretical model, shown in table 5.2, are less good, with the MAPE as high as 14% and the $R^2$ only 76%. Even though the power curves, shown in figure 5-4, remain correct, the error plots in figure 5-5 show high errors, and the box plot in figure 5-6 is worse than for the primary vessel. It is thus evident that the theoretical model is not performing as well on this vessel, and alternative models need to be investigated.
Figure 5-2: Error plots for the predictions of the theoretical model on the primary vessel. The theoretical model tends to underpredict power at low speeds.

Figure 5-3: Box plot for the primary vessel using the theoretical model. The trend is evidently linear, but there is room for improvement.

<table>
<thead>
<tr>
<th></th>
<th>MAE (kW/kW)</th>
<th>MAPE (%)</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical</td>
<td>0.087</td>
<td>14.0</td>
<td>76</td>
</tr>
</tbody>
</table>

Table 5.2: Numerical results for the secondary vessel, using the theoretical model.

Figure 5-4: Power curves for the secondary vessel, as predicted by the theoretical model.
5.1.2 Multiple and Sparse Regression

The first data-driven model attempted is a regression model. This type of model can only fit linear or simple non-linear terms, and so is expected to not be able to capture the full non-linearities of this problem. Two models are presented here. The first is the simplest multiple regression on the input data, without any feature transformations. The second is a multiple regression augmented with cubic speed and wind speed. These augmentations are obtained by LASSO regression, as explained in previous sections, and the model is thus labeled sparse regression.

First of all, the results for the primary vessel are presented in table 5.3. Sparse regression has a better performance for all metrics, as expected since it has a more advanced form. We also see that, although multiple regression has about the same performance as the theoretical model, sparse regression improves on it by about 1% in terms of MAPE. With regards to the power curves, shown in figure 5-7, both regressions produce similar plots, with the sparse regression able to capture the non-linearities in speed and wind speed better. The graphs produced by sparse regression resemble those of the theoretical model in all but trim, which seems to have the opposite direction. Finally, note that the curve for hull cleaning appears to have
Table 5.3: Numerical results for the primary vessel, using the two regression models. Sparse regression is clearly the one with the best performance.

<table>
<thead>
<tr>
<th></th>
<th>MAE (kW/kW)</th>
<th>MAPE (%)</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple Regression</td>
<td>0.045</td>
<td>7.7</td>
<td>91</td>
</tr>
<tr>
<td>Sparse Regression</td>
<td>0.040</td>
<td>6.6</td>
<td>93</td>
</tr>
</tbody>
</table>

Figure 5-7: Power curves for the primary vessel, as predicted by the regression models. Pitch and days since hull cleaning have been added since these are also inputs to the regression models. Note that for the sparse regression model the relationships are similar to those predicted by the theoretical model, with the exception of trim.

an increasing effect, as would be expected, while pitch is decreasing, again as would be expected from physical intuition. The error plots, presented in figure 5-8 show that the errors, although still high for low speeds, are much lower than those of the theoretical model throughout the speed ranges, especially for the sparse regression. Finally, the box plots in figure 5-9 show a good agreement for sparse regression, while multiple regression presents some curvature, mostly because it cannot fit the non-linear relationship between power and speed.

The results for the secondary vessel are shown in table 5.4. As for the theoretical model, the performance deteriorates on this vessel, although it remains much better than the theoretical model, with the errors for sparse regression just below 9%. The power curves in figure 5-10 are very similar to those of the primary vessel, indicating robust trends, while similar error profiles to those of the primary vessel are shown in figure 5-11. The same is true of the box plots of figure 5-12, with the box plots of the
5.1.3 Neural Networks

The regression models provided some improvement to the theoretical model by fitting their parameters to the ship characteristics, however due to their linear nature they are constrained in the complexity of the relationships they can represent. Neural networks, on the other hand, can represent any function, so they are a good next step to the regression models.

Two types of neural networks are investigated, the fully connected (FCNN) and multiple (MNN) neural networks, described in previous sections. When run on the

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE (kW/kW)</th>
<th>MAPE (%)</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple Regression</td>
<td>0.055</td>
<td>9.6</td>
<td>92</td>
</tr>
<tr>
<td>Sparse Regression</td>
<td>0.049</td>
<td>8.7</td>
<td>93</td>
</tr>
</tbody>
</table>

Table 5.4: Numerical results for the secondary vessel, using the two regression models. Sparse regression is the model with the best performance.
Figure 5-10: Power curves for the secondary vessel, as predicted by the regression models. The plots for both models are similar to those for the primary vessel.

Figure 5-11: Error plots for the predictions of the regression models on the secondary vessel. Both models perform well at high speeds, but are less accurate at low speeds.

Figure 5-12: Box plot for the secondary vessel using the two regression models. The multiple regression plot is curved rather than linear, reflecting an inability to fit non-linear data.
<table>
<thead>
<tr>
<th></th>
<th>MAE (kW/kW)</th>
<th>MAPE (%)</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCNN</td>
<td>0.024</td>
<td>3.9</td>
<td>95</td>
</tr>
<tr>
<td>MNN</td>
<td>0.040</td>
<td>6.9</td>
<td>92</td>
</tr>
</tbody>
</table>

Table 5.5: Numerical results for the primary vessel, using the two neural network models. The fully connected neural network has a much better performance than the rest of the models seen up to now.

primary vessel, the results are shown in table 5.5. Although the MNN has difficulty achieving a low error, the FCNN has a remarkably good performance, with a MAPE error below 4% and an $R^2$ of 95%, indicating a much better performance than that achieved by the regression models. The only downside, as shown in figure 5-13, is that this model tends to fit the data very closely and produces plots that have some peculiarities, such as a trend of decreasing power at very high speeds. Overall though the relationships predicted have the correct trend and the same as predicted by the regression models. A look at the error plots, in figure 5-14, shows that, apart from the very initial speed range, the 75% percentile of the MAPE errors is always below 10%, indicating a consistently good performance for the FCNN throughout the speed ranges. Finally, the box plots of figure 5-15 confirm the above observations that the FCNN has very good performance for the primary vessel.
Figure 5-14: Error plots for the predictions of the neural network models on the primary vessel. The FCNN in particular has a very good error profile, with 75% of data points having an error below 10% for most speed ranges.

Figure 5-15: Box plot for the primary vessel using the two neural network models. Both models have a good performance, with the FCNN in particular having much lower variation from the straight line.

Similar observations can be made for the secondary vessel. The results, exhibited in table 5.6, show that the FCNN performs much better than all the other models seen thus far, with a MAPE error of only 5%. The power curves, presented in figure 5-16, depict the same trends as for the primary vessel, and the error plots in figure 5-17 show that for this ship as well, 75% of data points tend to have an error below 10%. Finally, the box plots in figure 5-18 further demonstrate the good results of the FCNN.

<table>
<thead>
<tr>
<th></th>
<th>MAE (kW/kW)</th>
<th>MAPE (%)</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCNN</td>
<td>0.029</td>
<td>5.0</td>
<td>95</td>
</tr>
<tr>
<td>MNN</td>
<td>0.052</td>
<td>9.9</td>
<td>91</td>
</tr>
</tbody>
</table>

Table 5.6: Numerical results for the secondary vessel, using the two neural network models. The fully connected neural network has a much better performance than the rest of the models seen up to now.
Figure 5-16: Power curves for the secondary vessel, as predicted by the neural network models. The curves predicted by the FCNN are similar to those predicted for the primary vessel.

Figure 5-17: Error plots for the predictions of the neural network models on the secondary vessel. The FCNN has a very good error profile, with 75% of data points having an error below 10% for most speed ranges.

Figure 5-18: Box plot for the secondary vessel using the two neural network models. Both models have a good performance, with the FCNN having much lower variation from the straight line.
Theoretical and Neural Net  |  0.025  |  4.1  |  95  |
Regression and Neural Net  |  0.028  |  4.5  |  94  |
Theoretical and Regression  |  0.043  |  7.7  |  91  |

Table 5.7: Numerical results for the primary vessel, using the combined models. All models including neural networks perform equally well, with the naive model having the lowest MAPE.

5.1.4 Combined Models

Having obtained pure theoretical and pure data-driven models, combinations of them were subsequently attempted. Four main combinations were used. The first, named the naive theoretical and neural network model, feeds the outputs of the theoretical model to the neural network. The other three, combinations of the theoretical, neural network, and regression models, essentially use the first one as an initial prediction and feed its errors to a second model. For example, the theoretical and neural network model feeds the errors of the theoretical model to a neural network. The FCNN described above was the model used for the combinations given its good performance.

The results for the primary vessel are shown in table 5.7. The main observation is that the combinations that include a neural network tend to perform equally well to the simple FCNN, while the theoretical and regression models combined cannot achieve as high an accuracy. Furthermore, from the power curves in figure 5-19 we observe that the combined neural network models produce very similar curves to the simple FCNN. We can thus conclude that the combined models converge, given enough time, to the data-driven models. Finally, an examination of the error plots, in figure 5-20, and the box plots, in figure 5-21, yields similar conclusions to the simple FCNN case, in short that those combined models containing FCNNs have a high accuracy.

Moving on to the secondary vessel, the results, shown in table 5.8, further strengthen the conclusions drawn from the primary vessel that the neural network models outperform the theoretical and regression ones. For this vessel the theoretical and neural
Figure 5-19: Power curves for the primary vessel, as predicted by the combined models. The curves for the neural network models seem to converge to those of the FCNN.

Figure 5-20: Error plots for the predictions of the combined models on the primary vessel. The error profiles are similar to those of the FCNN.

Figure 5-21: Box plot for the primary vessel using the combined models. The neural network models match the linear trend well.
Table 5.8: Numerical results for the secondary vessel, using the combined models. The combined theoretical and neural network model has the best performance for this vessel.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE (kW/kW)</th>
<th>MAPE (%)</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive Th and NN</td>
<td>0.030</td>
<td>5.5</td>
<td>95</td>
</tr>
<tr>
<td>Theoretical and Neural Net</td>
<td>0.024</td>
<td>4.0</td>
<td>96</td>
</tr>
<tr>
<td>Regression and Neural Net</td>
<td>0.026</td>
<td>4.4</td>
<td>96</td>
</tr>
<tr>
<td>Theoretical and Regression</td>
<td>0.052</td>
<td>10.2</td>
<td>91</td>
</tr>
</tbody>
</table>

Figure 5-22: Power curves for the secondary vessel, as predicted by the combined models. The curves have the same shape as those of the primary vessel.

Network models remained the best, with a MAPE error of 5%. Furthermore, as can be seen by the power curves in figure 5-22, the predictions for the machine learning models are similar to those for the primary vessel. The error profiles in figure 5-23 are good, while the box plots in figure 5-24 also show good performance for the neural network models.

5.1.5 Physical Models

The above section demonstrated that combining pure data-driven and theoretical models can yield some improvements to the predictions, but this is only slight, and the combined models containing neural networks quickly converge to the pure data-driven solution. In this section we continue along the line of combining neural networks with physical intuition, and investigate the performance of the physical and admiralty
Figure 5-23: Error plots for the predictions of the neural network models on the secondary vessel. All neural network models appear to have a low error profile.

Figure 5-24: Box plot for the secondary vessel using the combined models. All neural network models have a behavior close to a straight line.
Table 5.9: Numerical results for the primary vessel, using the physical models. The admiralty model performs quite well on this vessel, with a MAPE error of only 3.5%.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE (kW/kW)</th>
<th>MAPE (%)</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Neural Net</td>
<td>0.025</td>
<td>3.9</td>
<td>95</td>
</tr>
<tr>
<td>Admiralty Neural Net</td>
<td>0.022</td>
<td>3.5</td>
<td>96</td>
</tr>
</tbody>
</table>

The results for the primary vessel are shown in table 5.9. The model using the admiralty coefficient in particular is able to improve on the performance of the previous models, increasing the $R^2$ to 96%. This model also leads to smoother plots, as seen from figure 5-25. Both models seem to have a very good error profile, with 75% of the errors for most speeds below 7.5%, as seen in figure 5-26. Finally, the box plots in figure 5-27 continue to present good matching of the data, especially for the model using the admiralty coefficient.

The results for the secondary vessel are similar. Both models perform quite well, with the admiralty neural network model having a slightly lower MAPE error of 4.1%, as shown in table 5.10. The power curves in figure 5-28 are very similar for both models, and the same is true of the error profiles in figure 5-29, as well as the box plots in figure 5-27.
Figure 5-26: Error plots for the predictions of the physical models on the primary vessel. Both models have very good error profiles, with 75% of data points having an error below 7.5% for most speed ranges.

Figure 5-27: Box plot for the primary vessel using the two physical models.

<table>
<thead>
<tr>
<th></th>
<th>MAE (kW/kW)</th>
<th>MAPE (%)</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Neural Net</td>
<td>0.026</td>
<td>4.1</td>
<td>95</td>
</tr>
<tr>
<td>Admiralty Neural Net</td>
<td>0.025</td>
<td>4.0</td>
<td>95</td>
</tr>
</tbody>
</table>

Table 5.10: Numerical results for the secondary vessel, using the two physical models. Both perform equally well, with the admiralty neural network being slightly better.

Figure 5-28: Power curves for the secondary vessel, as predicted by the physical models. The predicted curves for the two models are very similar.
5.2 Adding Waves

One of the main causes of uncertainty for the models described above was the absence of waves, which can cause a substantial increase in the resistance of a vessel. This data was only available for the year 2021. Even though this is a low fidelity dataset, as detailed in previous sections, we still run the models on this reduced dataset to gauge the performance of the models when waves are added. As such the best model for each of the four categories described above, namely the theoretical, sparse regression, FCNN, combined theoretical and neural net, and admiralty neural net models were run on both vessels using only 2021 data.

The results for the primary vessel are shown in table 5.11. As can be seen, all neural network models continue to perform well on this new dataset, however there is no notable improvement from the previous one. The wave height trend does seem to be predicted well, as seen from figure 5-31, and the error plots (figure 5-32) are quite low for most models, being mostly below 5% for most speed ranges for the admiralty neural net model. The box plots in figure 5-33 are also very good, especially for the admiralty model.
<table>
<thead>
<tr>
<th>Model</th>
<th>MAE (kW/kW)</th>
<th>MAPE (%)</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse Regression</td>
<td>0.041</td>
<td>7.4</td>
<td>88</td>
</tr>
<tr>
<td>FCNN</td>
<td>0.024</td>
<td>4.2</td>
<td>94</td>
</tr>
<tr>
<td>Theoretical and Neural Net</td>
<td>0.025</td>
<td>4.3</td>
<td>95</td>
</tr>
<tr>
<td>Admiralty Neural Net</td>
<td>0.021</td>
<td>3.7</td>
<td>95</td>
</tr>
</tbody>
</table>

Table 5.11: Numerical results for the primary vessel on the wave dataset. The results are similar to those for the full dataset without waves.

Figure 5-31: Power curves for the primary vessel using the wave dataset. Note that the curves for waves are predicted correctly.

Figure 5-32: Error plots for the predictions of the best models on the primary vessel using the wave dataset. The error profiles are low for the neural network models, with those for the admiralty neural net below 5% most of the time.
Figure 5-33: Box plot for the primary vessel using the wave dataset. The plot for the admiralty coefficient in particular has very low variance.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE (kW/kW)</th>
<th>MAPE (%)</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse Regression</td>
<td>0.041</td>
<td>5.9</td>
<td>93</td>
</tr>
<tr>
<td>FCNN</td>
<td>0.022</td>
<td>3.4</td>
<td>97</td>
</tr>
<tr>
<td>Theoretical and Neural Net</td>
<td>0.022</td>
<td>3.2</td>
<td>97</td>
</tr>
<tr>
<td>Admiralty Neural Net</td>
<td>0.018</td>
<td>2.8</td>
<td>98</td>
</tr>
</tbody>
</table>

Table 5.12: Numerical results for the secondary vessel on the wave dataset. The performance is significantly increased, especially for the admiralty neural net.

For the second vessel, on the other hand, a big improvement is observed when waves are added, with the performance of the model with the admiralty coefficient reaching a very low MAPE of 2.8% and an $R^2$ of 98%, as seen in table 5.12. The power curves for the wave height (figure 5-34) also are intuitive, and again for most speed ranges the errors (figure 5.12) of the admiralty neural net model are below 5%. Lastly the same conclusions can be drawn for the box plots, which closely follow a straight line for the neural net models (figure 5-36).

5.3 LNG Dataset

Having determined the models that work very well on the car carriers, the next and final step in this analysis was to check their performance on an entirely new type of vessel. We thus tested our best models using a dataset from an LNG vessel. These
Figure 5-34: Power curves for the secondary vessel using the wave dataset. Note that the curves for waves are predicted correctly.

Figure 5-35: Error plots for the predictions of the best models on the secondary vessel using the wave dataset. The error profiles are low for the neural network models, with those for the admiralty neural net below 5% most of the time.
five models are the best in each of the four categories, namely the theoretical, sparse regression, FCNN, theoretical and neural net, and admiralty neural net. Note that this vessel had a constant pitch propeller, so there is no variable pitch parameter. Also note that for this vessel we have data on waves obtained from company models, albeit of low fidelity.

The results for all models are shown in table 5.13. At a first glance these results seem opposite to the ones shown previously. The theoretical model performs the best, followed by sparse regression, while the neural networks are the least good. The difference however is not that large, with all of the models attaining an \( R^2 \) above 95%, and the MAPE errors are less than 2% different for the best and worst models. A look at the power curves, in figure 5-37, shows good predictions for most of the models, with the exception of the wave parameter which has unusual predictions, and the days since hull cleaning effect. And from the error plots in figure 5-38 we can see that the results are similar for most of the models, with the neural networks only breaking down at high speeds and regression at low speeds. Lastly, the box plots in figure 5-39 indicate good predictions for all models. In short, while the neural networks did not perform as well for the LNG vessel, their accuracy was similar to that of the other models, which is encouraging considering all the parameters of the models were determined on completely different ships.

Some of the reasons for the lower performance of the neural networks on this
Table 5.13: Numerical results for the LNG vessel. The performance of the neural networks is lower than for the other vessels, but on par with the rest of the models.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE (kW/kW)</th>
<th>MAPE (%)</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical</td>
<td>0.034</td>
<td>7.0</td>
<td>97</td>
</tr>
<tr>
<td>Sparse Regression</td>
<td>0.037</td>
<td>8.6</td>
<td>97</td>
</tr>
<tr>
<td>FCNN</td>
<td>0.048</td>
<td>8.8</td>
<td>96</td>
</tr>
<tr>
<td>Theoretical and Neural Net</td>
<td>0.044</td>
<td>8.0</td>
<td>96</td>
</tr>
<tr>
<td>Admiralty Neural Net</td>
<td>0.047</td>
<td>8.7</td>
<td>96</td>
</tr>
</tbody>
</table>

Figure 5-37: Power curves for the LNG vessel. These are mostly predicted correctly with the exception of waves and hull cleaning.

A vessel could be the following. First of all, as mentioned previously, the parameters of the neural networks were all selected on an entirely different vessel. Second of all, the data for the new vessel were collected at hourly intervals, as opposed to 15 minute ones, and during that time many of the quantities, such as wind speed and direction, wave height, and even vessel speed can change a lot. And finally there is the issue of different datasets, since the entire data cleaning methodology was based on a thorough investigation of the data from an entirely different vessel. The LNG vessel could have different sensors that measure data with different degrees of accuracy.
Figure 5-38: Error plots for the predictions of the best models on the LNG vessel. The regression models tend to have high errors at low speeds, while the neural networks are less accurate at high speeds.

Figure 5-39: Box plot for the LNG vessel. The trend follows the linear one closely for all models.
Chapter 6

Conclusions

The aim of this thesis was to provide a structured approach to the problem of estimating the shaft power of a vessel using open sea data. First of all, a detailed data cleaning process was described to obtain a dataset most suited for this problem. Many models were considered, starting from purely theoretical ones and purely data-driven ones and moving to simple and advanced combinations of them, and the process of obtaining the most suitable inputs and parameters was described. Finally, the models were tested on different types of vessels to determine their merits and issues. The final evaluation was done using both traditional numerical results as well as novel ways that include uncertainty estimates and physically intuitive plots.

Some general conclusions that can be drawn include the following: First of all, theoretical models are quite satisfactory at predicting physically correct relationships, since they are based on physical laws, but can perform inadequately as they do not take into account the particulars of each vessel. Regression models, when given the correct inputs, can also produce easily interpretable linear plots, but cannot match the complex relationships of the real data and thus cannot reach a high level of accuracy. Neural networks are very good at fitting complicated data points and achieving very low errors even on validation sets, but they tend to fit the data in complicated ways that are not always interpretable. This is partially fixed with the more physically intuitive neural networks developed in this study, with neural networks utilizing the admiralty coefficient having the best performance with the
most intuitive results. However, when tested on different and coarser datasets they lose their excellent performance.

In terms of numerical performances, when trained using high quality, frequent data that include the speed, draft, trim, wind, and the hull condition, physically inspired neural networks can reach an error of less than 5%, and an $R^2$ over 95%. With the addition of waves the error becomes as low as 3% with an $R^2$ as high as 98%.

There are several paths in this line of research that can be followed. One would be to further investigate the merging of theoretical and data-driven methods, such as the physical and admiralty neural nets developed here. Another could be the pursuit of better interpretability of the results, as in the power curves and error plots of this work. But the path we believe has the most potential is the more detailed investigation of the input data. This concerns the more accurate and less error-prone collection of the important quantities (such as speed). It also concerns the collection of new quantities that are vital to this problem but are not readily available, with the example discussed here being the wave characteristics, which need to be measured using onboard sensors as opposed to models. In short, we believe that with a better data collection and data processing protocol, coupled with data-driven models based on solid theoretical foundations, a very accurate system of monitoring real-time power and fuel consumption of vessels can be developed.
Bibliography


