WORLD MARITIME UNIVERSITY

Malmö, Sweden

SHIP OPERATIONAL PERFORMANCE MODELLING FOR VOYAGE OPTIMIZATION THROUGH FUEL CONSUMPTION MINIMIZATION

By

MERIAM CHAAL
Tunisia

A dissertation submitted to the World Maritime University in partial
Fulfilment of the requirements for the award of the degree of

MASTER OF SCIENCE
In
MARITIME AFFAIRS

(MARITIME SAFETY AND ENVIRONMENTAL ADMINISTRATION)

2018

Declaration

I certify that all the material in this dissertation that is not my own work has been identified, and that no material is included for which a degree has previously been conferred on me.

The contents of this dissertation reflect my own personal views, and are not necessarily endorsed by the University.

Meriam CHAAL

18 September 2018

Supervised by: **Professor Aykut I. Ölcer**

Head of Maritime Energy Management Specialization at the World Maritime University

Abstract

Increasingly stringent maritime regulations and high fuel prices are placing more pressure on shipping companies to find ways to improve their ships' fuel efficiency in order to reduce costs and comply with the new rules. As fuel counts for the largest portion of the voyage cost, small fuel savings could achieve significant voyage cost savings. Moreover, fuel consumption reduction contributes considerably to the greenhouse gas emissions reduction, which became a global concern.

Traditionally, optimizing ship speed is known to be effective in minimizing fuel consumption, and numerous ship operational optimizations focused on this method. Even though it is an effective method, it is often difficult to implement as ships have their schedules to respect in addition to the ports' logistical constraints, which limit the speed optimization scope.

Other ship controllable variables, such as the trim i.e. the difference between the aft draft and the foreword draft, and the ship course are worthy of attention when seeking to minimize fuel consumption while the vessel is cruising. The trim can be controlled by simple ballast arrangement, which may also be cost-free in case of a gravity assisted ballast water system. Often, the vessel is badly trimmed such that it generates additional fuel usage that can be saved by an optimal trim configuration. On the other hand, optimizing the ship's route by changing the ship's course, with the aim of avoiding harsh weather or benefiting from the wind and current directions, to decrease the ship's resistance, can significantly reduce the voyage fuel consumption. This method can be implemented while respecting the ship schedule by assessing the different options available and deciding accordingly.

In this thesis, different black box models are compared to predict ship fuel consumption, which depends on the ship specific and the navigational input parameters. The objective is to find the best predictive model to use in a decision support system (DSS) for energy efficient ship operation. The best prediction methodology is identified based on the comparative analysis, which yields to employing Artificial Neural Network (ANN). In the DSS, the Genetic Algorithm, an evolutionary optimization algorithm, is employed with the help of ANN in order to find the optimal set of input parameters that give the least fuel consumption.

The investigation is based on numerical data of a VLCC case ship under normal operation. It is common that this type of operational data cannot include all the input variable values ranges and in some cases the range can be quite narrow, which limits the accuracy of ship operational performance prediction models. Special attention must also be assigned to the pre-processing of this type of data. It is demonstrated how to address this aspect and build models with high predictive accuracy that, when employed with the GA for trim and route optimization, result in potential fuel savings.

Based on the successful results of VLCC, it can be confidently concluded that the developed methodology is a promising direction, which has been tried for the first time in the academic literature on ship performance modelling and its optimization. The

developed method can be adapted and applied to other merchant ship types to become part of a comprehensive on-board energy management system as long as proper tailoring is performed.

Keywords: Machine Learning, Black Box Models, K-Nearest Neighbours, AdaBoost Decision Tree, Artificial Neural Network, Genetic Algorithm, Energy Efficiency, Ship Operational Performance Modelling, Performance Optimization, Fuel Savings.

Acknowledgment

It is a pleasant duty for me to express my deep gratitude to my sponsor, the International Maritime Organization Women's Association for the great opportunity they offered me to study in this prestigious university.

I owe my deepest gratitude to Mr. Moncef BEN ROMDHAN, Director General of the Tunisian Merchant Marine and Mr. Mourad GHORBEL, Director of Tunisian Ports, who encouraged me to join the university and supported me from the application procedure until today.

To my Professor and supervisor Dr Aykut Ölcer, who through his guidance and his enlightened advice directed my thesis with great interest and attention, may you find here my sincere thanks and the testimony of my deep gratitude Professor.

My most distinguished thanks are addressed to Dr. Anita Mäkinen Chief Adviser to the Director General of Maritime Sector at the Finnish Transport Safety Agency for her great support during our field study to Finland. Without her help, I would not have been able to collect the data to conduct this research.

I am also grateful to NAPA for delivering their valuable data to make this study possible.

I have been given the opportunity to be under the guidance of Professor Jens-Uwe Schröder-Hinrichs head of MSEA specialization during my academic studies. He has been responsible for the success of this intense learning year. I want to express my sincere gratitude to him for his support, constructive criticism and the knowledge he shared during the year.

I would like also to thank Professor Elif Bal Besikci for her continuous encouragement and valuable advice that made this work a successful one.

I want to express my deepest gratitude to Dr Mahdi Abed Salmen for his patience and precious advice and excellent resources he shared with me.

I would like also to thank Mrs Inger Battista and Mrs Anne Pazaver for improving the text in numerous ways.

Finally, my most distinguished thanks are extended to my family for the support that each of them has given me throughout this year. My deepest gratitude is addressed to my best friends Mona ANTAR and Adel Ali Desher for the great patience and moral support that has encouraged me to reach my academic and personal goals.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

IMO	International Maritime Organization
UNFCCC	United Nation Framework Convention on Climate Change
MARPOL	International Convention for the Prevention of Pollution from Ships
CO2	Carbon Dioxide
GHG	Green House Gases
SEEMP	Ship Energy Efficiency Management Plan
EEDI	Energy Efficiency Design Index
MEPC	Marine Environment Protection Committee
DSS	Decision Support System
BAU	Business As Usual
KNN	K-Nearest Neighbours
CSV	Comma Separated Values
GP	Gaussian Process
ANN	Artificial Neural Networks
GA	Genetic Algorithm
SVR	Support Vector Regression
MSE	Mean Squared Error
MAE	Mean Absolute Error
CART	Classification And Regression Trees
BPA	Back-Propagation Algorithm
VLCC	Very Large Crude Carrier
GPS	Global Positioning System

RPM	Revolution Per Minute
RMSE	Root Mean Square Error
ME	Main Engine

Chapter I: Introduction

1.1.Background

Maritime transport, a vector of globalisation, today represents a competitive mode of transportation at a lower cost compared to other types of transportation. Allowing for economies of scale and low cost of transportation, maritime transport has become the flagship mode given the large capacity it can transport over long distances with an 80% share of world commercial transit (UNCTAD, 2017). In addition, the modernisation of logistics through containerization and advanced technology allowing direct monitoring, as well as easy routing, have contributed to the fast growth of maritime transport. Thus, the number of tonnes transported by sea increased by more than 200% between 1970 and 2000 and increased by 60% between 2000 and 2013 (Vigarié, 2016). Furthermore, Vessels have evolved and developed to fit profitability and competitiveness needs, which has resulted in building of ships with larger sizes and higher speeds. However, place of shipping in sustainable development and the possible alternatives to allow the evolution of the sector in respect of the environment still have to be shaped (IMO, 2014). Maritime transport development is happening in parallel with the increase of its CO2 emission, which represents a negative externality that affects the entire environment. Figure1-1 shows that the emission rate of CO2 is increasing and the estimates established by the International Maritime Organization (IMO) are close to reality (green columns in the figure), (IMO, 2009). The figure indicates that the current IMO studies and expected risks should be highly considered.

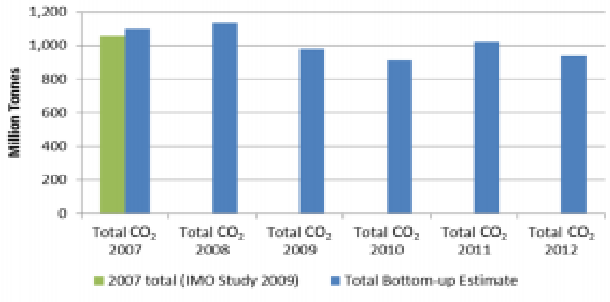


Figure 1-1 Evolution of CO2 emissions and estimations of IMO (IMO, 2009)

The increasing GHG concentration in the atmosphere and the associated warming effects are considered as a major cause of climate change (WMO, 2013). For international maritime transport, according to the third IMO GHG study as illustrated in Figure 1-2, in the absence of corrective measures, it is estimated that the rate of CO2 emissions of the sector will increase between 150% and 250% by 2050 (IMO, 2014).

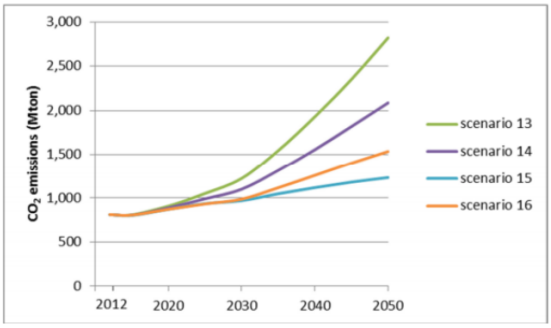


Figure 1-2 Business as usual projection of CO2 emissions from international shipping 2012-2050, (IMO, 2014)

As a result, GHG emissions from international shipping are receiving increasing attention, and possible mitigation measures are being considered, both at the regulatory and sectoral levels.

During the United Nations Framework Convention on Climate Change (UNFCCC) in 1992, the GHG emissions from all sectors were discussed and the engagement of States to reduce it was highlighted. International shipping was assigned to the IMO as a specialized United Nations (UN) body to regulate the GHG emissions from ships. Consequently, the IMO marked the entry into force of Chapter IV of

MARPOL Annex VI on 1 January 2013. This Annex represents a real turning point in the maritime sector as it represents the first mandatory global regime for control of GHG emission from the maritime sector. It introduced the Energy Efficiency Design Index (EEDI) as an energy efficiency benchmark to be respected by new ships, and the Ship Energy Efficiency Management Plan (SEEMP) to improve the operational energy efficiency of existing ships. Later in 2015, the Paris agreement came to specify a clear and precise target to limit Climate Change effects by keeping global warming well below 2° compared to 2008 levels, which put more pressure on the IMO to reduce GHG emissions from ships. The Data Collection System (DCS) was then adopted to enter into force in January 2019 as a system for recording and collecting data on the fuel consumption of ships engaged in international voyages, which is proportional to their GHG emissions. The data collected will provide a solid basis to decide on additional regulations that will complement or amend the regulations already adopted by IMO on its way toward environmentally sound maritime transport (IMO, 2016). In addition, the IMO recently adopted a new strategy specifying its commitment to reducing GHG emissions from international shipping by at least 50% by 2050 compared to 2008, which is the first quantified target fixed by IMO (IMO, 2018).

To sum up, in view of environmental degradation, maritime regulation is moving toward the development of strict measures to improve the ships' energy efficiency. The higher authorities in the maritime community and the leaders in the shipping industry are now focusing on the development of measures favourable to both, the economies and the environment. Therefore, shipping companies are under increasing regulatory pressure, requiring them to adopt appropriate solutions that achieve an economic objective in response to the rising cost of fuel and especially to the international regulations.

1.2. Problem statement

The EEDI regulation has set different time phases to achieve energy efficiency targets while building more and more energy efficient ships. The targets are increasingly strict allowing for innovation in ship design to find solutions for compliance with the more stringent rules. On the other hand, the SEEMP has been introduced to operate the ship in an energy efficient way by applying energy efficiency operational measures and monitoring the effect of any changes in ship operation. While the energy efficiency at design stage improves in parallel with technological innovations, the operational energy efficiency improvement depends on the ship operators' motivation to apply the operational measures in addition to the complexity of applying these measures. Technological innovation in ship design has been highly reliant on Artificial Intelligence (AI) for a long time to improve the hull forms and structural arrangements (Amarel & Steinberg, 1990). The development of AI has been a key factor in the development of not only ship design but also many other sectors including, medicine, biology and different engineering fields. Recently, the shipping industry at the operational level also started to benefit from this fast development of AI tools as many studies conducted by IMO working groups have proven that CO2 emissions could significantly decrease through appropriate implementation of operational measures (IMO, 2014). Traditionally these measures were implemented solely through speed optimization as it is an easy way to reduce ship fuel consumption and does not require deep knowledge of the ship operation. However, in order to implement other energy efficiency operational measures, such as trim or route optimization, ship operational performance changes in different voyage conditions should be deeply examined and monitored. It is for this purpose that the AI tools have been employed since the classical tools, using the physical relations developed at the ship design stage, are not able to precisely describe all the operational conditions, as they are often different from the limited shipyard test conditions. Many studies are currently seeking to employ AI tools to further improve ship operational energy efficiency, which still has great potential to decrease the CO2 emissions from international shipping (IMO, 2014). This could be achieved by using AI to facilitate the implementation of the operational energy efficiency measures through on-board instruments equipped with advanced and effective Decision Support Systems (DSS). Such research and developments will need to analyse the ship performance from historical operational data and examine the available AI tools in order to predict the ship performance in various conditions with the minimum possible error. Such an approach will provide the optimal decisions to take in order to improve ship performance with a minimum risk of error.

1.3. Aim and Objectives

The aim of this dissertation is to employ the most suitable AI tools to optimize ship operational performance and effectively contribute to the global efforts towards energy efficient maritime transport. This will be possible by achieving the following objectives:

- Explain the crucial steps of ship performance modelling as a data science process
- Conduct an appropriate ship operational data mining in order to avoid over fitness and high prediction errors
- Examine the applicability of a set of popular currently available machine learning algorithms to the ship performance modelling problem
- Use the most appropriate ship fuel consumption prediction model to make future predictions of ship performance
- Combine the prediction model with a successful evolutionary algorithm to solve the ship fuel consumption minimization problem
- Change the optimization constraints to conduct different ship voyage optimization scenarios and validate the models effectiveness to improve the ship's energy efficiency
- Explain the applicability of the built prediction and optimization models as a DSS for ship energy efficient operation

1.4. Dissertation outline

Chapter 2 describes the operational energy efficiency measures and gives an idea about their potential in saving fuel usage and GHG emission reduction. This chapter introduces the machine learning tools and the black box models with their application to build the ship operational performance prediction models in addition to the optimization for energy efficient ship operation. Chapter 3 explains the developed methodology of preparing the dataset, building a fuel consumption prediction model and using the GA as the optimization model. Chapter 4 merges the steps previously explained with an application of all models to the operational historical data of a case study ship (VLCC). In this chapter, different optimization scenarios are tested to validate the developed methodology. The results are presented in Chapter 5 with interpretation and assessment of the developed method through the level of success in meeting the objective of the optimization scenarios. The chapter concludes with remarks on the applicability of this DSS on-board ships. Finally, conclusions and further research are presented in Chapter 6.

Chapter II: Ship performance modelling and optimization

In this dissertation, the focus is the relationship between ship fuel consumption and the different influential variables that describe the ship's condition and the external conditions. The aim is to use this relationship to optimize the ship voyage by minimizing fuel consumption. This chapter will briefly introduce the operational energy efficiency measures and their potential to reduce ship fuel consumption, which have been a strong motivation to use modern tools for ship performance modelling and optimization. The chapter will then describe the usage of AI tools for ship performance modelling and optimization. It will conclude with a comprehensive literature review to present different studies that aimed to model, and optimize ship operational performance and the current gap in this research field.

2.1. Ship operational energy efficiency measures

Operational energy efficiency measures can have a significant effect on reducing the GHG emission from ships as they aim to reduce the consumed energy and, thus, the consumed fuel. In the years following the oil crises, shipping companies have put in place measures to reduce their fuel consumption, as fuel cost is a major part of the ship voyage cost. These traditional measures such as slow streaming i.e. reducing the voyage speed significantly compared to the design speed, have proven their effectiveness and are now back on the agenda. Thus, according to the IMO, operational energy efficiency optimization is the main vehicle for reducing emissions from a ship (IMO, 2014). Figure 2-1 illustrates the different scenarios and demonstrates that operational measures can mitigate the growth of CO2 emissions. Bold lines in the figure are Business As Usual (BAU) scenarios and thin lines represent either greater efficiency improvement than BAU or additional emission controls or both. The best scenario shows that it is even possible to reduce the emission rate and bring it down below the rate achieved in 2012 (IMO, 2014).

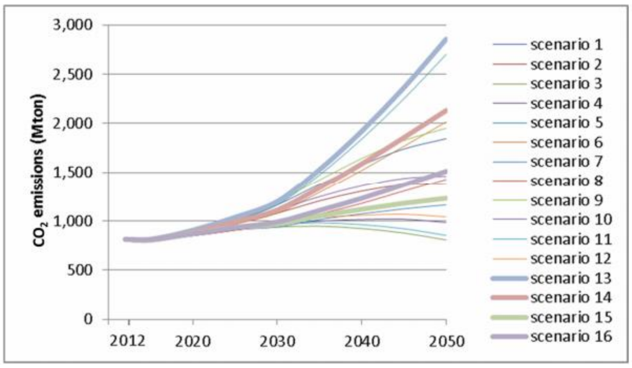


Figure 2-1 Projection of CO2 emissions from international maritime transport with different energy efficiency scenarios (IMO,2014)

2.1.1. Speed optimization

One of the most applied and oldest measures was to reduce the ship speed; this parameter has the greatest effect on ship fuel consumption, which is a cubic function of ship speed (Lindstada & Eskeland, 2015). Therefore, reducing ship speed is an effective way to reduce fuel consumption (Wartsila, 2009). As shown in Table 2-1, speed reduction can result in energy savings of up to 23% and decreasing the ship’s speed by only 1kn could save more than 5% of the energy consumption.

Speed reduction	Saving energy consumption
-0.5 kn	-7%
-1 kn	-11%
-2 kn	-17%
-3 kn	-23%

Table 2-1 Reduction of consumption according to the decrease in speed (Wartsila, 2009).

As a result, sailing at the optimal speed for each ship condition is an extremely effective energy efficiency operational measure.

2.1.2. Trim optimization

Shipyards aim to build ships and propulsion systems with the highest possible efficiency. However, once commercialized, these systems usually do not operate as efficiently as planned. The trim is one of the parameters that is often badly configured while the vessel is cruising because the trim is set during harbour time, which is different from cruising conditions considering the squat effect (Rocchi, 1994). The latter is the phenomenon of increased immersion and trim of the ship when it is cruising compared to calm water (Varyani, 2005). It has been proven that the ship consumed energy profile can vary significantly when changing the trim configurations (Journé, Rijke, & Verleg, 1987; Journée, 2003). Optimizing the ship's trim is one of the easiest and least expensive energy efficiency measures that requires simple ballast distribution modification. It has been shown that a well trimmed vessel can make important energy savings (Ziylan & Nas, 2016).

2.1.3. Weather routing

Meteorological routing represents the determination of the optimum ship route with regard to weather conditions in order to promote energy savings (Padhy, Sen, & Bhaskaran, 2007). The objective of weather routing is to offer a route with the minimum fuel consumption, while considering the safety of the ship and remaining competitive with the earliest arrival time (Lin, Fang & Yeung, 2013). It is then a matter of taking advantage of weather conditions to facilitate the route by decreasing the total ship resistance, for example, by footing downwind lanes in order to take advantage of their speed and reduce fuel consumption. Weather-based route planning allows up to 3% fuel savings (Armstrong, 2013).

2.1.4. Autopilot adjustment

The autopilot is an auxiliary deck equipment that replaces the helmsman in the bridge and insures that the ship is following the planned route while sailing in open sea or out of the high traffic areas. However, in case of poor directional stability, the frequent or large course alterations of the autopilot can increase ship fuel consumption (Amerongen, Duetz, & Okawa, 2017). Therefore, a well-adjusted autopilot has a great influence on the ability to stay on the course, reduce the use of the rudder optimize the angles and, therefore, generate significant fuel economies. It is sufficient then to find the appropriate and precise parameters of the autopilot according to the route and different criteria to allow a considerable reduction in the use of the rudder and, thus, reduce the drag and the fuel consumption. It is estimated that high accuracy Autopilot operation would reduce fuel consumption by 0.5 to 3% (IMO, 2009).

2.1.5. Propeller and hull maintenance monitoring

This measure consists of improved hull and propeller condition management in order to maintain smooth submerged surfaces. It allows appropriate polishing intervals and the choice of adequate antifouling treatment in order to decrease the hull water resistance (Demirel, Turan, & Incecik, 2016). This preventive measure can provide up to 10% improvement in hull performance compared to a fouled hull (ABS, 2015).

Operational energy efficiency has the potential to reduce CO2 emissions from ships considerably and the operational energy efficiency measures are not limited to the list above.

2.2. Operational Ship performance modelling

2.2.1. White, Grey and Black Box modelling

The concept of ship performance has different interpretations, but in the scientific papers and discussions, it means the relationship between ship speed and the corresponding energy or fuel consumption (Haranen, Salo, Pakkanen, & Kariranta, 2016). Modelling this performance consists of building a model that describes this relationship based on mathematical formulas, on statistics or on both at same time. These three approaches are known as White, Black and Grey-Box modelling, respectively. All of these approaches take into consideration the possible factors (Red in the Figure 2-2) that affect the ship's technical performance and employ them in the model.

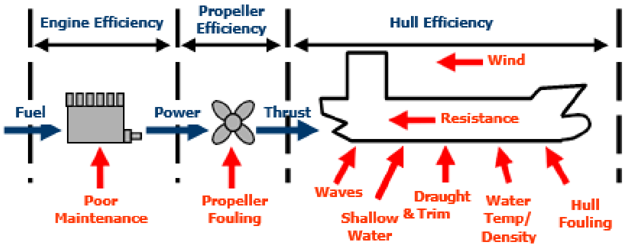


Figure 2-2 Variables affecting the ship performance (Pedersen & Larsen, 2009)

Ship performance modelling usually starts with the design process, where the interaction between the different ship systems and their interaction with the external environment are investigated and described in one model. However, at design stage the data collection tools, such as sensors, are limited and operational external conditions are not all available for testing (Logan, 2011). Therefore, the white box approach, which is based on physical laws, such as the ship's power as a function of its total resistance and its speed, is usually the only way for technical performance modelling at design stage.

Conversely, during the operation of the ship, more and more sophisticated data collection and storage systems based on thousands of sensors are available. This has allowed access to large datasets from ships, which require control with modern computing tools and AI techniques to predict the ship's future performance from its past (Solonen, 2018). Therefore, black box models such as Artificial Neural Networks based on statistical data are currently the only approach to deal with ship operational performance modelling.

Accordingly, grey box modelling, as its name signifies, is something between both preceding approaches. This means that the physical laws are employed in the model and corrected by statistical data in order to decrease the error margin and give better prediction accuracy (Haranen, Salo, Pakkanen, & Kariranta, 2016). In practice, when the grey model relies more on the physical laws than the volume of the historical data, it requires deep initial information about the ship's physical characteristics to obtain good results. On the other hand, when the model employs less physical assumptions, and relies more on historical data, it requires a broader set of data describing as many as possible of the operational conditions. This approach is still under research and development and needs the ship design information (Haranen, Salo, Pakkanen, & Kariranta, 2016).

2.2.2. Black Box for ship operational performance model

As previously mentioned, ship operational performance can be modelled based on statistical data. The historical data of the ship's voyages with different loading and weather conditions is first collected, then, using the powerful current machine learning tools, many steps are conducted to finally build a prediction model. Depending on the available data, the model will have input and output variables selected from the dataset. Generally, the consumed energy or fuel is the output variable to predict as a function of the speed, the other ship related input variables (trim, displacement, hull condition…), along with environment related (wind, current, wave…) input variables (Perera & Mo, 2016).

2.2.2.1. Machine learning tools

Nowadays, highly developed soft computing techniques, combined with machine learning, open-up the possibility to build more and more accurate ship performance prediction models from the data collected through ship sensors and weather forecast companies. In this big data era, the abundance of data from ships has made ship operational energy efficiency research a data-oriented one.

Machine learning is the scientific field that studies how a machine learns from its experience with the objective to build computer systems that are able to adapt and learn from their experience (Wilson & Keil, 1999). Therefore, a machine-learning tool is a computer program whose performance for a certain task improves with its experience on that same task (Mitchell, 1997). Machine learning has three main tasks:

2.2.2.1.1. Supervised learning

Supervised learning is the task where the machine learns a target function 'Output = f (Inputs)', which is represented by a model based on these variable values in the dataset called training data. The function is then used to predict the new output variable value from the new input variables value (Knight & Michelle, 2018). In supervised learning, there are two categories of model based on their tasks: Regression or Classification. The regression models predict a numerical output value, while the classification models predict the output class, such as a vessel type or a plant family.

2.2.2.1.2. Unsupervised learning

For unsupervised learning systems, the task is first to find the hidden relationship between the variables inside the dataset, which is known as ‘association rule mining’, and, second, to be able to assign a good structure to the main dataset through dividing it into different groups or clusters. For these systems, the training dataset is a number of instances of unlabelled variables, which will need to be divided and assigned to the created labelled groups.

2.2.2.1.3. Reinforcement Learning

In reinforcement learning, the system is directly interacting with the environment without previous knowledge about it and trying to learn through trial and error until becoming able to make sequential decisions (Knight & Michelle, 2018). It is applied in self-driving cars, dialog systems, adaptive medical treatments and others.

In the case of ship operational performance modelling, the model should represent a function that describes the relationship of the fuel consumption or the shaft power, as the dependent variable to predict, with the independent input variables. Both the target and input variables have numerical and not categorical values. Therefore, for this field of modelling, we apply a **supervised learning, regression model.**

2.2.2.2. Datasets and data pre-processing

In order to build a statistical regression model, it is necessary to collect a number of instances of historical data of the ship performance. This is called a dataset. Generally, the raw dataset is a two dimensional matrix, a CSV or an excel sheet where each column is a feature or variable and each row is an instance or one reading, and vice versa. The larger the dataset is, the better the work that can be done is. However, the real world ship data and navigational data are often incomplete, inconsistent and noisy because of some errors in the data collection tools (Markov, n,d). The sensors on-board ships may encounter functioning problems and crewmembers can make many errors when collecting the information, which results in incoherent data samples. Data pre-processing is a crucial process that consists of data cleaning, integration, transformation and reduction (Chouvarda et al., 2017). All these steps have the same objective of preparing the data in a consistent, clean and structured form to build a performant model, while reducing the computational time. However, this should not affect the integrity of the original data (Chouvarda et al., 2017).

2.2.2.3. Algorithm selection

As already explained in the machine learning tasks, based on the dataset and the objective of modelling, the model type is selected, which is supervised learning regression for the previously defined operational ship performance modelling. In the regression models, many algorithms can be applied to the same dataset to accomplish the same objective. However, the best algorithm should be selected based on the model performance in predicting future outputs. The model's evaluation is usually measured by calculating the errors in prediction of unseen data in order to make sure that the model will be accurate once deployed (Raschka, 2016). The main challenge when applying different algorithms to the dataset or searching for novel algorithms is making accurate predictions with future data. The algorithm should learn the target function that relates the output variable to the input variables. However, no single standard algorithm can apply to all datasets and give the same performance even for the same task and objective, such as regression. Therefore, algorithm selection is very important to the data analyst in order to come up with best prediction results.

2.3. Operational ship performance optimization

2.3.1. Definition

In general, optimization is simply about obtaining an optimum value of a function (minimum or maximum) by selecting its variables values from a defined set (Thomas & Mahapatrab, 2016). Optimizing the operational ship performance deals with the minimization of the operational cost of the voyage, which is mainly the fuel cost. Once the best algorithm is chosen and trained, as explained previously, it is used with new data for making predictions of ship fuel oil consumption or other proportional outputs such as the propulsion power. Minimizing the fuel consumption is called the objective of the optimization. Applying the prediction model to the new data variables will give the usual ship fuel consumption. However, using machine learning for modelling the ship's performance should not stop at the predictive modelling step (Hamm, n d). It always has the utmost objective of optimization because these complex nonlinear functions with multiple variables cannot be optimized by the traditional analytical methods (Ghanshyam, Mirjalili, Patel & J.Savsani, 2018). In order to meet the optimization objective, which is minimizing the model function, the optimal input variable values should be found (Ghanshyam, Mirjalili, Patel & J.Savsani, 2018). The dataset for operational ship performance modelling usually contains ship data and navigational data. The input variables from navigational data are external to the ship; they rather describe what happens around the ship during its voyage, such as the weather forecast. Therefore, these variables' values cannot be controlled for the optimization. In contrast, the ship specific input variables, such as speed or course are manageable and the ship operators can decide to change them, depending on their schedules and targets. These variables are called Decision Variables (Bal Besikci, Arslan, Turan, &Olcer, 2016). Thus, optimization deals specifically with finding the ship decision variables' values to minimize the fuel consumption function of the set of variables. In many applications, these optimization results have been employed in Decision Support Systems (Petersen & Winther, 2011).

There are two main categories in optimization:

2.3.2. Unconstrained optimization

The unconstrained optimization, as its name indicates, is finding the best variables to minimize or maximize a function without any constraint on the variables values.

2.3.3. Constrained optimization

Ship operators make decisions to reduce fuel consumption by reducing the speed with the specified constraint to arrive on time, in order to avoid extra fuel usage if the ship stays at the waiting area. This is a minimization problem subject to a constraint. The decision variables are not chosen among an infinite number of values. They are rather limited by one or more constraints. The system will then choose the values that meet the defined objective while respecting the specified constraints.

2.4. Literature review

Considerable research has been conducted trying to accurately model operational ship performance. Some of them pursued ship performance optimization, while some others had the objective of building a good predictive model and validating its accuracy. Petursson (2009), Petersen & Winther (2011) and Soner, Akyuz & Celik (2018) are the only studies found that employed non-parametric algorithms to predict ship operational performance. Non-parametric algorithms are simple and effective machine learning prediction algorithms, where the prediction function parameters are unspecified. They rather rely on the similarity between the training data and the data for prediction, where the similarity is simply assumed to be equivalent to the distance between the training data and prediction data (ISS-AS, 2005). Petursson (2009) used data from noon reports and highlighted the importance of the data pre-processing and its effect on the model's accuracy. The algorithms used to predict the shaft power of a car-ferry case ship were K-Nearest Neighbours (KNN) and Support Vector Regression (SVR). The different models' accuracies were ranked and the KNN had the best prediction performance. The model was then used to detect the bad trim by comparing the shaft power profile as a function of the pitch at the optimal conditions. Petersen & Winther (2011) used high quality ship historical data and conducted a complete modelling process with focus on the importance of understanding the variables' effects on ship performance and the data mining on the model's performance. It compared Gaussian Process (GP) to ANN algorithms in fuel consumption prediction modelling and has found that GP has high accuracy that was very satisfying and, in some cases, better compared to ANN, which had the best accuracy in general. Soner, Akyuz & Celik (2018) also conducted a coherent modelling process and showed the importance of selecting the best set of input variables to end up with a good model prediction performance. Soner, Akyuz & Celik (2018) has employed a non-parametric popular algorithm, the Decision Trees, and compared its performance to the Artificial Neural Networks (ANN), which was previously applied by Pedersen and Larsen (2009) to the same ship case, using the same publicly available dataset. Decision Tree provided satisfying predictive performance compared to ANN. All of this researches has significantly contributed to clarifying the role of data mining in the modelling process and effectiveness of many machine-learning algorithms to predict ship performance in different conditions. However, they are very few and far from being sufficient to validate and generalize their conclusions for the different ship types and operational conditions (Aldous, 2016; Petersen and Winther, 2011). There is still a huge gap in ship operational performance modelling and statistical models with high prediction accuracy are still needed (Aldous, 2016; Petersen and Winther, 2011). In addition, this research has only validated the models' accuracy without exploring their potential to be employed to solve a ship performance optimization problem and serve as DSS.

Few studies have employed the ANN as one of today's most performant machine-learning algorithms for non-linear regression problems. (Pedersen and Larsen, 2009; Leifsson, Sævarsdóttir, Sigurðsson, & Vésteinsson, 2008) have built ship fuel prediction models based on the ANN algorithm, which demonstrated a good predictive accuracy. Bal Besikci, Arslan, Turan, & Olcer (2016) also employed ANN to predict ship fuel consumption and highlighted its high predictive performance. In this paper, the model was further used in a DSS to minimize ship fuel consumption. However, both studies cited above were based on noon report data, which is not as reliable as the high quality data obtained from acquisition system and sensor technologies. Petersen and Winther (2011) used high quality data and demonstrated again the success of ANN as a non-linear regression algorithm to predict ship fuel consumption. However, in this research, the weather input variables were not considered and the ship performance optimization problem was not solved.

Examining the literature to find research areas to explore, Genetic Algorithm (GA) was not employed in any research to optimize general ship operational performance. It was employed in many studies for ship route planning by (Hinnenthal, 2008; Marie and Courtielle, 2009; Wang, Li, Li, Veremey, & Sotnikova, 2018; Al-Hamad, Al-Ibrahim, & Al-Enezy, 2012). In addition, it has been found that GAs are broadly used to solve optimization problems at early stage ship design, when the ship hull form has to be optimized to reduce the resistance and find the optimal propulsion power (Olcer, 2007; Bagheri & Ghassem, 2014; Hirayama & Ando, 2007; Guha & Falzaranoa, 2015).

In order to remedy the mentioned gaps, this research will apply a set of available machine learning algorithms to a high quality ship historical dataset to build a good ship operational performance model and combine it with the GA for optimal energy efficient ship operation.

2.5. Summary

This chapter gives an idea about the importance of operational energy efficiency measures to reduce ship fuel consumption and, by extension, GHG emissions from ships. Since the classical programming techniques cannot identify the complex non-linear relationship between fuel consumption and all other variables, the black box models are introduced to solve this problem, considering their ability to learn from historical ship data. The ship data nature justifies the need to prepare it to be used in black box models. Today’s non-classical methods and mainly evolutionary algorithms have been the main solution to ship operational performance optimization problems. In this chapter, the literature review showed a strong need for further research on the applicability of the available non-classical tools from AI for energy efficient ship operation.

Chapter III: The development of a ship performance modelling and optimization methodology

Generally, ship performance is a measure of the propulsion power or the fuel consumption at a certain state described by the loading condition, the ship speed and external conditions. As each problem has more than one prediction model with the existent machine learning algorithms, this chapter will describe the black box models used in this study to estimate ship fuel consumption. The ship statistical data pre-processing methods used to prepare the data will first be described. A general description of Decision Tree and Adaptive Boosting (AdaBoost) will be presented, followed by a description of the KNN and ANN models. The chapter concludes by presenting the GA as the optimization model used in this study. Finally, the steps of the developed methodology will be presented.

3.1.Dataset pre-processing

The ship dataset available to conduct this research was first examined to find the input and output variables to consider for the attainment of the research objectives. Fuel consumption was selected as the only output variable to predict with the model because it is the main indicator of ship energy efficiency and GHG emission volume. Even though it was a time-consuming process, data pre-processing was well-studied because later on, the programming step with the clean data did not need the same time as data mining. Then the raw data was globally checked in order to detect general observations such as missing and inconsistent values and to clean it. Variables with a large number of missing values were excluded and the data samples with outliers were deleted, which were defined depending on the domain knowledge and the information delivered as below;

*For all variables, the values out of the ordinary known ranges were excluded, such as angles out of 0-359,9°.

*Variables with values out of the given ranges from the dataset' owner were excluded, such as cargo mass greater than ship displacement.

*Negative values for all variables that should only be positive were also considered as outliers, such as negative flow rate or negative mass.

These outlier' considerations were translated into conditions to be treated automatically in excel, and other missing or inconsistent data samples were detected through data visualization (histograms and plots).

For data integration, only one company delivered the ship specific data and the weather hind cast dataset together, so there was no need to collect different datasets and integrate them.

In data transformation, the unit scales were standardized as illustrated in Table 3-1

Distances	Meter (m)
Time	Second (s)
Speeds	Meter per second (m/s)
Directions	Degree (deg)
Displacement	Ton
Flow rate	Kg/h

Table 3-1 Data variables' unit scales

The data was also normalized with the most common method, where each column is scaled from its min-max range to 0-1 range i.e. each value V became V' such as in equation (1).

$$V' = V\text{-}min/(max\text{-}min) \qquad (1)$$

Continuing the data transformation, a first feature selection based on the domain knowledge resulted in the elimination of variables that do not have any relation with the fuel consumption or are out of the scope of the thesis. Furthermore, in the cases of similar variables, such as ship course and ship heading, only one was kept in order to avoid redundancy.

An advanced feature selection technique was conducted based on a linear correlation matrix and data visualization to check the effect of input variables on the fuel consumption and to detect other redundancy of similar variables. The correlation analysis between the set of the selected variables was made with a simple code of the free version of the programming software Python 3.6.

In the data transformation process, new variables were calculated from a combination of existent variables that did not have linear correlation with fuel consumption. The new variables replaced the old ones, which reduces the number of variables and thereby reduces the model complexity to improve its predictive performance. In this respect, the apparent wind angle (α_{AW}) and Apparent wind speed (S_{AW}) were combined to better show their effect on fuel consumption and were replaced by only one important variable: the apparent wind speed on ship direction (S_{AWS}) calculated in excel with equation (2);

$$S_{AWS} = S_{AW} * \cos(\alpha_{AW} * (\pi/180)) \ \text{(m/s)} \qquad (2)$$

(Wright, Colling, & Park, 1999)

In addition, some variables were transformed in order to show their hidden effect on ship fuel consumption, such as the current direction (α_{C}), which was transformed into relative current direction (α_{RC}) to the ship heading ($\alpha_{Heading}$) with the equation (3) where ABS stands for absolute value;

$$\alpha_{RC} = \text{ABS}(\alpha_{Heading} - \alpha_{C}) \qquad (3)$$

Finally, the dataset was ready to use with one of the available programming languages to build the expected model.

3.2.Modelling and programming

In order to build a model that describes the fuel mass flow as a function of the input variables, a machine learning programming language is needed to code the different algorithms, apply them to the dataset and then evaluate their prediction performance. For this thesis, Python programming language version 3.6 was used to do the major work. In addition, the Artificial Neural Network toolbox in MATLAB Software version 2015a was used to apply ANN algorithm as it is a successful and fast tool for modelling, which fits neural networks to solve regression problems.

In recent years, non-parametric techniques of machine learning have been increasingly used in solving regression problems. Non-parametric algorithms such as, K-Nearest Neighbours and Decision Trees are well-known for classification tasks as they take simple assumptions on the form of the learned function (output =f(inputs). They are performant algorithms that have proven their success as classifiers and they are now taking more and more interest as regression algorithms in the machine learning community (Soner, Akyuz & Celik, 2018). They learn the underlying function with the training dataset and then predict the output values of a testing dataset by applying the function to the new input values.

In this research, a selection of most common non-parametric algorithms were fitted to the fuel consumption prediction problem. The ship dataset was loaded to Python and divided into train dataset and test dataset in order to train the models and then predict the output of the test dataset and evaluate the model performance. The split of the data was made randomly with the common ratio of 70% for train and 30% for test. Each dataset was divided into inputs as X-train and X-test and output as Y-train and Y-test. For training, the regressors (predictive regression models) were applied to the X-train and Y-train and the prediction was conducted by applying the trained algorithm to the X-test to find the Y-Predicted.

The performance of the models was quantified by the difference between the real output values Y-test and the predicted ones. The Mean Squared Error (MSE), Mean Absolute Error (MAE) and R^2 are statistically established metrics to represent this difference (Pedregosa et al., 2011). The MSE and MAE are expressed by their respective equations (4) and (5), where n is the number of test data samples and e_t (Y_{test} - $Y_{predicted}$) is the error between real output and predicted output.

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \quad \textbf{(4)}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t| \quad \textbf{(5)}$$

R^2 is the coefficient of determination, which describes how well the predictions fit the real data. It ranges from 0 to 1 with perfect fitness when R^2 is equal to 1 when 100% of the predicted outputs are equal to the real values (Pedregosa et al., 2011).

3.2.1. K-Nearest Neighbours (KNN)

KNN is a lazy non-parametric algorithm that when trained with data instances does not take any assumptions on the distribution of the data to find the input-output relation. It is able to quickly learn complex underlying functions while saving all the information in the data (Yu et al., 2016). In order to predict the output Y for a given input X, it finds the K instances in the training set with Xi in the proximity of X as shown in Figure 3-1, and then computes Y with one of two methods. The first method is to consider Yi of the closest point Xi to X. The second calculates the average of Yi responses of the K nearest neighbours to X, (referring to equation (6)), where, $N_k(x)$ is the ensemble of the K nearest points to X in the training dataset.

$$\hat{y}(x) = \frac{1}{k}\sum_{xi \in Nk(x)} Yi \quad \textbf{(6)}$$

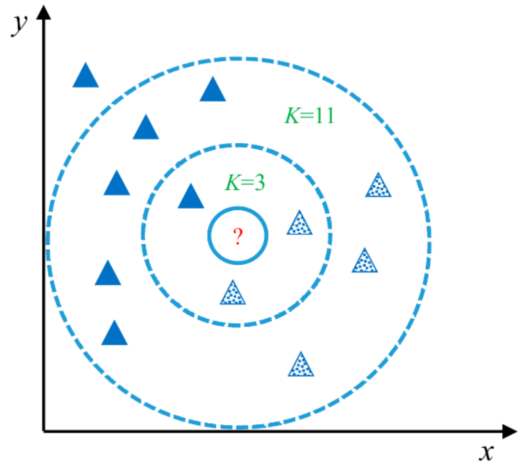


Figure 3-1 Demonstration of the K-nearest neighbours (KNN) method, (Wang et al, 2017)

The second method is more accurate as it takes many neighbours to the point, which confirms the similarity between them rather than associating the similarity to only the closest point that may result in high errors. With this method, computing the response Y can be also weighted, which means that the contribution of each neighbour is weighted by its distance from the target input X; it is called Distance Weighted KNN (Shahin, Jaksa & Maier, 2008). This method is more effective as it selects the K neighbours that are closer to the point to predict, which makes the output estimation more accurate. The distance is transformed into weight by one of the Kernel equations (Gauss, Cosine…) (Shahin, Jaksa & Maier, 2008). In order to measure the proximity to X, many formulas are used but the most common one is the Euclidean Distance, which is expressed by the formula (7) (Yu et al., 2016), where p is the number of variables in the input vector X.

$$\boldsymbol{d}(\boldsymbol{Xi}-\boldsymbol{X}) = \sqrt{(\textstyle\sum_{\boldsymbol{s}=\boldsymbol{1}}^{\boldsymbol{p}}(\boldsymbol{Xis}-\boldsymbol{Xs})^2)} \quad \textbf{(7)}$$

In the case of this thesis, the model is a distance-weighted KNN and the weight is represented by the simple Kernel inversion (8) (Shahin, Jaksa, & Maier, 2008), calculated from distance (d) between X and the neighbour Xi. It means that the closer Xi to X is, the higher the weight is or the greater influence the neighbour Xi has on the calculation of the response (Y).

$$\mathbf{Wi = 1/d} \qquad \textbf{(8)}$$

This choice has been based on the success of this algorithm in many classification problems and in some regression applications such as in Petursson (2009). Due to time constraint, deeper research on the model parameters with more complex methods were not tested.

The selection of the number of neighbours K is generally based on experiments and the first K number to start testing does not follow a fixed norm. Depending on the computation method of the response, K could start by a big or a small number. If the response is simply the closest neighbour, K could be the small one; generally, five neighbours are sufficient. In the case of the average computation, a large number of neighbours is preferred to increase precision. Then, to choose K empirically, training starts with a small K, computing the accuracy, and increasing K by three for example. If good accuracy is recorded, a test with larger K is conducted and accuracy is calculated again. Otherwise K is decreased by one until best accuracy is reached. The model with best accuracy (with optimum K) is selected.

3.2.2. Decision tree and AdaBoost

3.2.2.1. Decision tree

A decision tree is a non-parametric algorithm used in building prediction models that can handle both categorical and numerical data. The model defines a prediction rule that applies a hierarchical binary partition of the data into a number of subsets, which together form a tree. The average of the outcome of the elements in each subset represents the predicted outcome of the subset (Venkatasubramaniam et al., 2017). The objective is to set a prediction rule with a minimum error between the predicted value and the target value. The tree is composed of a number of nodes, which contains the subsets of the data observations (Breiman Friedman, Olshen & Stone, 1984). The topmost node in the tree, which is the best predictor from the input variables, is called “root node” and contains all the data observations. The most important step in building decision trees is the splitting step, which defines where and how to split the subsets below the root node (Venkatasubramaniam et al., 2017). The binary split of each node are the branches of the tree. Each node is split into two other nodes until the stopping rule is satisfied. The stopping node is called a terminal node or a leaf. All the leaves together represent a sample from the original dataset where each leaf is an observation. In order to predict a new output, the leaves into which belong the new input observations are determined and the existing outcomes of the corresponding leaves are combined to predict the new output. Decision tree has long been a good and simple classifier, but recently it has been employed for regression problem. The most common algorithm to build regression trees is CART (Classification And Regression Trees) (Breiman Friedman, Olshen & Stone, 1984), which was employed in this research. The covariate to split and the split point are the two dimensions in splitting a node. Therefore, CART search for best splits dimensions, with the objective of minimizing the relative sum of squared errors (Breiman Friedman, Olshen & Stone, 1984). The best split is sought across all possible splits.

3.2.2.2. Adaptive Boosting (AdaBoost)

Decision tree is considered as a simple and weak prediction algorithm that can be improved by an ensemble technique, which is a combination of more than one weak learner (machine-learning prediction algorithm). AdaBoost is an ensemble technique known to be the best classifier when boosting decision tree algorithms ((Breiman Friedman, Olshen & Stone, 1984). It is a technique that combines subsequent weak learners to get the optimum prediction model. This means that the same prediction algorithm learns from its previous mistakes in prediction and improves itself. The output of the boosted algorithm is a weighted combination of the outputs of all previous weak learners where the weights are assigned based on the error in prediction of each learning algorithm (University of Oxford, 2015). For decision trees as weak learners, AdaBoost improves the training process to build a good decision tree, by minimizing the sum of errors in prediction from each decision tree. In each iteration during training, each decision tree has a hypothesis of the value of output to predict. The hypothesis from each decision has a weight based on the previous error to predict the same output sample with that decision tree. These weights are also used next to improve the splits when building the subsequent decision tree. As this boosting method was very successful in building strong classification models (University of Oxford, 2015), it is employed in this study to test its performance and fitness with ship performance modelling as a regression problem.

3.2.3. Artificial Neural Networks ANN

Using the ANN to solve a prediction problem is more complicated than using the lazy non-parametric algorithms. The ANN way to learn a function is inspired by human and animal brains, which follow a complex method to train themselves in information processing (Park, 2011). As illustrated in Figure 3-2, these brains are composed of neural networks, which contain neurons that collect information from each other using their dendrites. The neuron sends out electrical signals to the synapse through the axon, which will allow or inhibit the activity. Therefore, when a neuron receives an input electrical signal higher than the inhibition level, it resends lower signals through its axon and the process is repeated with many neurons until the brain learns how to process the information ((Shahin, Jaksa & Maier, 2008).

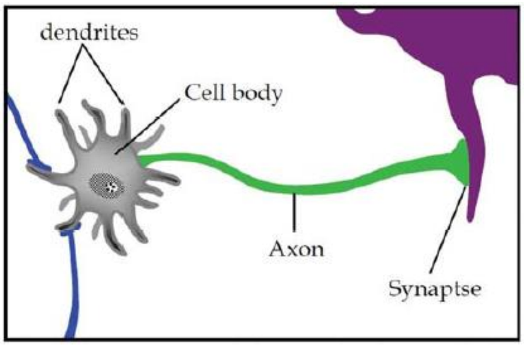


Figure 3-2 The functioning of human neurons (Park, 2011)

The artificial neural networks are sets of neurons called nodes or processing elements (PE) arranged in an input layer, an output layer and one or more intermediate hidden layers. Each node is related to the nodes in the other layers by the weighted links (Rakhshandehroo, Vaghefi, Aghbolaghi, 2012). Each node's input is multiplied by its weight and the summation of all weighted inputs with the biases represents the whole neuron network activity (Ij), which is illustrated in equation (9), where Xi is the input of the layer i, Wji is the weight of the layer i and n is the number of neurons. Biases (θj) are constant non-zero additional weights (Shahin, Jaksa & Maier, 2008).

$$I_j = \theta_j + \sum W_{ji} X_i \quad (9) \quad \text{(Shahin, Jaksa \& Maier, 2008)}$$

Next, the consequence of this summation is processed through a transfer function to find the output Yi. The type of function depends on the type the problem to solve by the network (sigmoidal, linear functions or else…) (Park, 2011). Yi is expressed in equation (10).

$$Y_i = f(I_i) \quad (10) \quad \text{(Shahin, Jaksa \& Maier, 2008)}$$

Different studies have specified that the number of hidden layers depends on the complexity of the problem and can be improved by experiment while training the network (Flood and Kartam, 1994; Ripley, 1996; Sarle, 1994). Each layer's outputs are the inputs of the next layer, which are processed with a transfer function as described previously. This network is called a multilayer feed forward neural network (see Figure 3-3).

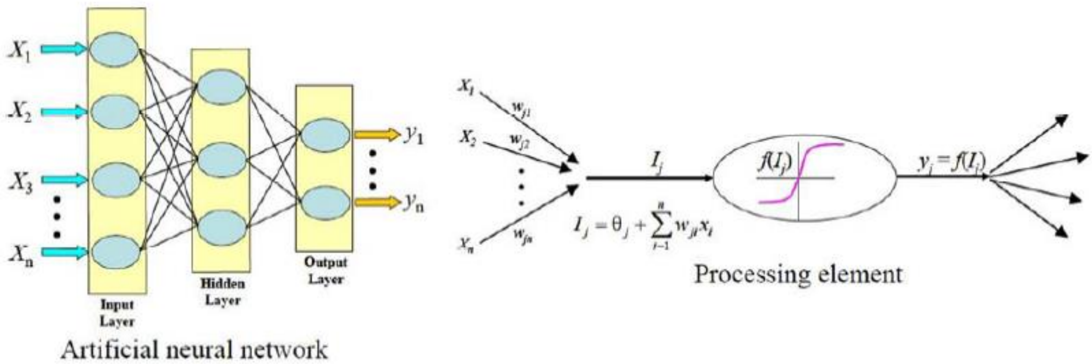


Figure 3-3 A model structure and modus operandi of Artificial Neural Networks (Shahin, Jaksa & Maier, 2008)

The training process determines the optimised set of weights of the layers. The most common and efficient training process for these networks is the Back-Propagation Algorithm (BPA) (Basheer & Hajmeer, 2000). It is a multiple iteration, or "Epoch", training process. In each epoch, the prediction error is calculated and used as a benchmark to assess the current set of weights. The training is repeated until the error stops improving or the epoch finishes (Basheer & Hajmeer, 2000). The numbers of input and output nodes are the numbers of input and output variables of the problem.

The BPA has proven its efficiency in modelling different functions and has become the most popular in training multilayer ANN (Samarasinghe, 2007). ANN with BP training Algorithms has been successfully applied in ship performance modelling in different studies (Bal Besikci, Arslan, Turan, & Olcer, 2016; Pedersen & Larsen, 2009).

For these reasons, this research employs a feed forward neural network with BPA for training in MATLAB (2015a) ANN toolbox. The number of hidden layers started with two and was improved by experiment. After each training the number of hidden layers was increased by one until no more improvement in error was recorded.

3.3. Optimization with Genetic Algorithm (GA)

Genetic Algorithm is one of the first population based meta-heuristic algorithms that has been successfully applied to numerical optimization problems and one of the most powerful optimization tools actually available (Goldberg, 1989; McGookin, Murray-Smith & Li, 1996). It is inspired by the main biology evolution principles; reproduction, crossover, mutation and selection (Holland, 1975). Each solution of the given optimization problem is encoded as a chromosome composed of a number of genes. The population evolves in each iteration and preserves the important information in the chromosomes. A new generation of chromosomes (problem's possible solutions) is produced by combining the old good chromosomes and discarding the bad ones (D.Vose, 2003). While the role of crossover and selection is to build a relationship between the old and the newly produced chromosomes and transferring the acquired information from old to new generation, the mutation is responsible for keeping diversity in the population by randomly introducing new information (De¸bski, 2010). Normally, the GA encodes each solution as a string of bites, but for numerical problems, it is rather a vector of values of the function's variables. Each numerical vector represents a feasible solution to the optimization problem and the whole population is a group of the chromosomes candidates for the optimal solution (D.Vose, 2003). For the selection step to proceed to mating for a new generation, parents (best solutions) are selected based on their fitness to solve the optimization problem. The fitness is defined by a function, which if maximized, the chromosomes are considered as best fitted ones and should be kept to become the parents that build the new generation. The whole cycle is summarized in Figure 3-4.

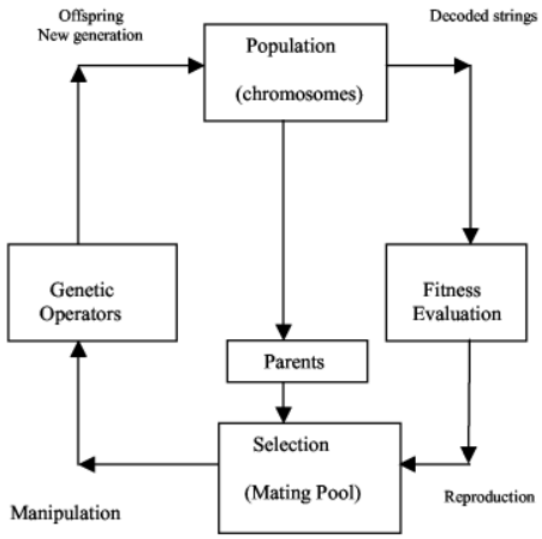


Figure 3-4 The cycle of Genetic Algorithms (Konar, 2000)

The fitness function is a translation of the optimization objective. If the optimization objective is to minimize a known function, the fitness function could be **1 / (objective function +1)** as an example. When the objective is to maximize a certain function, then the fitness function could be the objective function itself.

In the case of ship performance optimization, the model function to minimize is usually a complex one with multiple variables, which make the GA a highly fitted method to apply to find the optimal ship-related input variable values that minimize fuel consumption. The Global Optimization Toolbox with GA in MATLAB 2015a was used to solve a single objective constrained optimization problem of ship operational performance. The best prediction model selected from the previous step was employed to define the fitness function. Therefore, the selected prediction model function was called in the GA. As the objective of optimization is to minimize fuel consumption, the fitness function was defined as a fraction with the model output underline in the equation, as illustrated in equation 12. The smaller the fuel consumption function (model prediction) is, the better the chromosome is.

$$\textbf{Fitness} = \mathbf{1 / (Model\ output + 0.1)} \qquad (12)$$

In order to start the search in the space of possible solutions, an initial population of chromosomes is created with a certain size. The space of feasible solutions is limited in the case of constrained optimization. In this work, the ranges of the decision ship variables with a potential of fuel savings define the boundaries of the GA search as constraints.

From the initial population, the fitness of each chromosome is calculated (in this work by equation 12) and the best ones are selected to form the next generation. The reproduction is then carried out, which defines how many individuals from the "elites" are kept for next population and how many from the discarded will be replaced by crossover and mutation (Goldberg, 1989). The crossover recombines randomly selected genes (crossover points) of any two chromosomes (parents) to obtain new chromosomes called children or offspring (Figure 3-5). It can be a single point or two points' crossover operation.

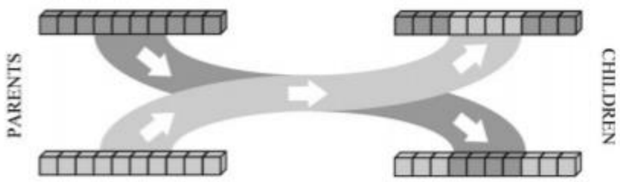


Figure 3-5 GA Crossover (Goldberg, 1989)

The mutation process then replaces the remaining poor solutions by randomly making small changes in the chromosome genes (Goldberg, 1989; McGookin, Murray-Smith & Li, 1996).

After the new population is formed, the fitness of each individual is evaluated again and the same steps are repeated until the stopping criteria is reached. The stopping criteria could be the maximum number of iterations (called generations in GA), a defined value of fitness or non-improvement from one generation to the next.

When the search in the bounded space ends, the result shows the optimal set of variables' values that minimizes fuel consumption. Next, the values are decoded from the range 0-1 into the real variables numbers with the inverse of equation (1).

The flow chart below explains the whole developed methodology in this study, starting from the raw dataset delivery to the ship performance optimization results deployment.

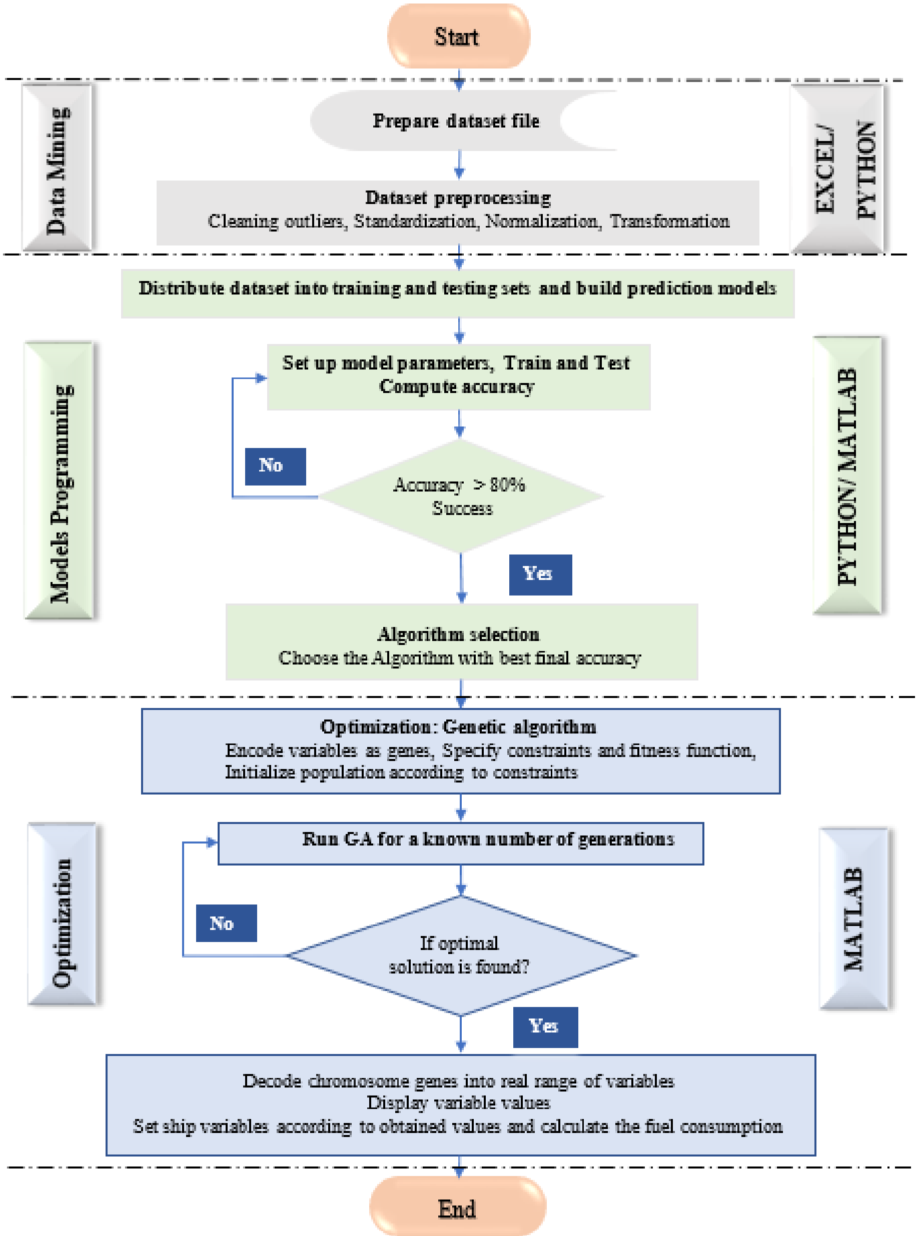


Figure 3-6 The developed methodology

3.4. Summary

This chapter presents the pre-processing methods used to prepare the dataset to be used in statistical models. The chapter then briefly describes the machine learning algorithms employed in this study. The KNN consists of estimating the mean of the given output values of the K closest data samples to a given point. AbaBoost Decision tree combines the weighted predictions of many Decision Trees to have a better accuracy. ANN is composed of sets of nodes arranged in an input layer, an output layer and intermediate hidden layers, which after the training process defines the relationship between the inputs and the output. The presented models are then evaluated, and the best is used with the GA for ship performance optimization. GA is an evolutionary algorithm that defines an initial population of the numerical vectors of the ship input variables solution to the problem of minimizing fuel consumption. After running the optimization, the optimal input variable values from a constrained search space are displayed.

In the next chapter, all these models will be applied to a high frequency ship dataset.

Chapter IV: Case study

This chapter will present the experimental application of the previously developed ship performance and optimization methodology to a dataset of a case study ship. The dataset will first be described and pre-processed to be employed with the different machine learning algorithms used in this study. The parameters of each predictive model will be given. The chapter will also explain the GA parameters and the specified constraints to test different ship voyage optimization scenarios.

4.1. The ship and the dataset

4.1.1. Data type and sources

NAPA Group, a Finnish software house that provides solutions for ship design and operation to improve safety and energy efficiency in the maritime industry, delivered the dataset for this research. The dataset is a Comma Separated Values (CSV) document, with the names of the different measured variables in column headings and data samples in the rows. The variables include the speed, loading conditions, weather conditions and different fuel rates of a VLCC ship of 320000DWT during a three months period of operation between 15/01/2017 and 31/03/2017. The sample size is ten minutes and the whole data size is 9188 samples of 28 variables. For confidentiality reasons, NAPA did not give the ship characteristics or locations in the dataset.

The variables are as below:

Real date	Displacement (t)	Speed over ground (m/s)	Course over ground (deg)	Current direction (deg)	Current speed (m/s)	Depth below keel (m)
Heading (deg)	Draft aft (m)	Draft fwd (m)	Fuel mass flow aux total (kg/h)	Fuel mass flow boiler total (kg/h)	Fuel mass flow ME total (kg/h)	Fuel temp boilers (C)
Trim (m)	Propulsion power (kW)	Propulsion RPM	Speed through water (m/s)	Combined Wave Height (m)	Wind Wave Height (m)	Wind Wave Period (s)
Swell Height (m)	Swell Crossing Period (s)	Swell Direction (deg)	True wind direction (deg)	True wind speed (m/s)	Distance travelled	Cargo mass (t)

Table 4-1 Raw data variables

Data collection tools, as provided by NAPA, are the ship’s existing sensors;

*Fuel consumption: volumetric flowmeter, which improves the measurement accuracy compared to tank sounding.

*Ship speed over ground: GPS.

*Ship heading: GPS.

*Propeller revolution: tachometer.

*Propulsion power: combining the RPM measurements and torque measurements with torsiometer.

*Displacement: calculations with hydrostatic information.

*Trim: on-board dynamic trim monitoring system where positive trim refers to aft draft > forward draft.

The above data acquisition sensors have less error probability and margin compared to noon reports, which gives the data a high quality rank that will result in a reliable prediction model.

NAPA does not have information on exact models of the sensors or uncertainty levels of the newscast-data (waves, swell and other) because an external weather service provider gave it.

4.1.2. Scope and assumptions:

In this work, the hull and propeller were clean starting from January 2017 as the NAPA Group specified that the ship was in dry-dock just before this sailing period. In addition, it was assumed that their condition was unchanged during the three months period of the data as the typical effect of the hull and propeller fouling on the propulsion performance is only few percent. Therefore, the hull and propeller conditions were excluded from the ship performance model.

The fuel consumption to predict is limited to main engines fuel mass flow as it represents the rate used for ship propulsion.

This research considers the ship’s performance during the sailing period, out of time at port.

4.1.3. Data analysis results

From all of the data variables, the fuel mass flow of auxiliary engines and fuel mass flow of the boilers were not considered, respecting the scope of the research. In addition, the draft aft and the draft forward had many missing values and were represented by the trim and the displacement, which together are sufficient to describe the loading conditions. The real date and the distance travelled were not considered as they are only a reference of sampling points and do not have any effect on fuel consumption.

In Figure 4-1, the green correlation factors represent positive correlations greater than 0.2 and the red ones are negative correlations with absolute value greater than 0.2. The analysis was conducted to remove the highly correlated input variables and to check the effect of the variables on the output (fuel mass flow). As shown in the correlation matrix, the wind relative angle and the wave encounter angle are highly correlated, which was already demonstrated in the fluid dynamics (Wright, Colling, & Park, 1999). As a result, only the wind relative angle will be considered as an input variable to the model. The correlation factor of cargo mass and displacement is more than 0.99, which was expected and displacement only was kept.

	Heading (deg)	Speed over ground (m/s)	Displacement (ton)	Cargo mass (mt)	Trim (m)	True wind speed (m/s)	True wind direction (deg)	Wind relative angle	Apparent wind angle (deg)	Apparent wind speed (m/s)	App wind speed on ship direction	Wave encounter angle (deg)	Wind Wave Height (m)	Wind Wave Period (s)	Swell Height (m)	Swell Crossing Period	True swell angle (deg)	Current speed (m/s)	Relative current angle (deg)	Fuel mass flow ME (kg/h)
Heading (deg)	1																			
Speed over ground (m/s)	0,49296	1																		
Displacement (ton)	-0,8028	-0,5703	1																	
Cargo mass (mt)	-0,8067	-0,5733	0,99972	1																
Trim (m)	0,78162	0,5923	-0,9583	-0,9625	1															
True wind speed (m/s)	0,10902	-0,074	-0,1153	-0,1116	0,08418	1														
True wind direction (deg)	-0,3092	-0,3125	0,42704	0,42275	-0,3776	-0,0671	1													
Wind relative angle (deg)	-0,0258	-0,1224	-0,0379	-0,0368	0,05447	0,15335	0,10005	1												
Apparent wind angle (deg)	0,0632	0,06939	-0,0681	-0,0652	0,04011	-0,0395	-0,0688	0,07927	1											
Apparent wind speed (m/s)	0,13079	0,04072	-0,115	-0,1163	0,12469	0,05796	-0,142	-0,0367	-0,0873	1										
App wind speed on ship direction (m/s)	0,00607	0,10785	0,01965	0,01537	0,01554	-0,342	-0,0721	-0,1257	-0,0632	0,81611	1									
Wave encounter angle (deg)	-0,0258	-0,1224	-0,0379	-0,0368	0,05447	0,15335	0,10005	1	0,07927	-0,0367	-0,12572	1								
Wind Wave Height (m)	0,13757	-0,2738	-0,0882	-0,0839	0,0526	0,52917	-0,069	0,09713	-0,1173	0,27474	-0,166752	0,09713	1							
Wind Wave Period (s)	0,11052	-0,2804	-0,0691	-0,0646	0,03613	0,43921	-0,0809	0,07418	-0,1276	0,32966	-0,044924	0,07418	0,95501	1						
Swell Height (m)	-0,2487	-0,2758	0,16108	0,16843	-0,1643	0,16718	-0,0601	0,12742	-0,0583	0,1677	0,021334	0,12742	0,40727	0,5106	1					
Swell Crossing Period (s)	-0,2255	-0,1846	0,12019	0,12416	-0,086	0,19265	-0,0957	0,19974	-0,0375	0,08857	-0,048623	0,19974	0,32906	0,35351	0,7896	1				
True swell angle (deg)	-0,2696	-0,1673	0,09664	0,10217	-0,1203	0,14861	0,09166	0,19955	-0,0817	-0,3127	-0,319463	0,19955	-0,043	-0,0737	-0,1247	-0,1267	1			
Current speed (m/s)	0,26424	0,2845	-0,3668	-0,3683	0,37631	0,15611	-0,1606	0,0086	-0,043	0,07654	-0,03337	0,0086	0,10262	0,06311	0,00079	-0,023	0,00026	1		
Relative current angle (deg)	-0,3411	-0,4046	0,28435	0,28631	-0,2828	0,03891	0,22486	0,1738	-0,0491	-0,0856	-0,093165	0,1738	0,0784	0,07882	0,14081	0,15072	0,21464	-0,1309	1	
Fuel mass flow ME (kg/h)	-0,0877	0,42356	0,14175	0,13788	-0,0861	-0,2824	0,00637	-0,2876	0,02214	0,1758	0,338127	-0,2876	-0,2812	-0,2479	-0,0487	-0,1035	-0,4232	0,00215	-0,1087	1

Figure 4-1 Correlation Matrix for all relevant variables

For wind variables, the relative wind angle has higher correlation with fuel consumption than true and apparent wind directions, it was, thus, kept to replace them. The swell and current variables do not have redundancy, so they were all considered as input variables

With further data analysis, Figure 4-2 shows that the ship has made three voyages with three different displacement values. Generally, the ship speed decreases when the ship is loaded and increases in case of ballast voyage and vice versa. In addition, the trim trend is negative when the ship is loaded and positive when it is a light ship. A displacement histogram is relevant in this case to visualize the three voyages with the different loading conditions.

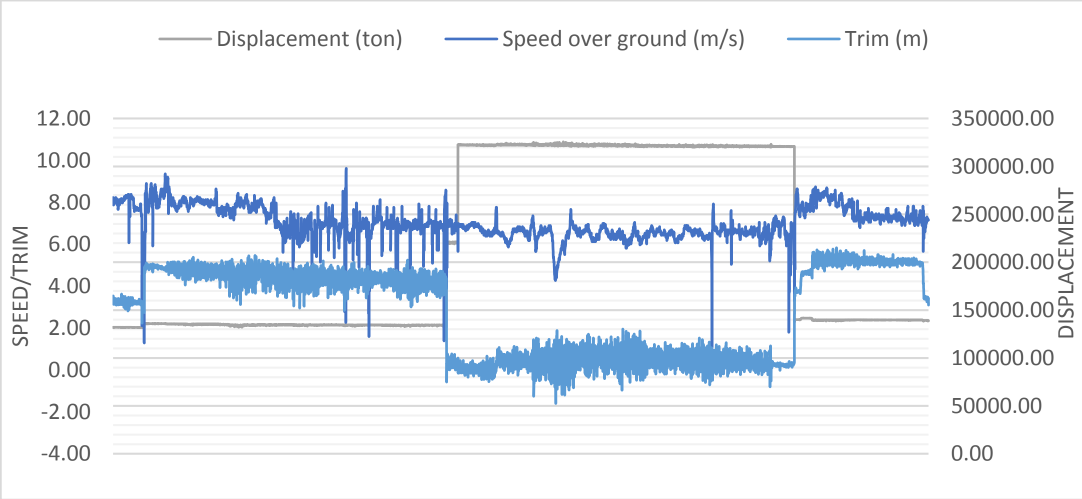


Figure 4-2 Speed, Displacement and Trim distribution for all data samples

The histogram (Figure 4-3) shows that displacement during the three voyages is between 132086-145086 tons for ballast voyages and 314086-327086 tons for loaded ship. The other values in between were considered as outliers and were eliminated as they were only a few samples that will be noisy in the model training. In addition, Figure 4-2 allowed detecting the periods in harbour where the speed was almost zero and followed by a change in the displacement. The whole fuel consumption profile during the ship operation changes with the loading conditions.

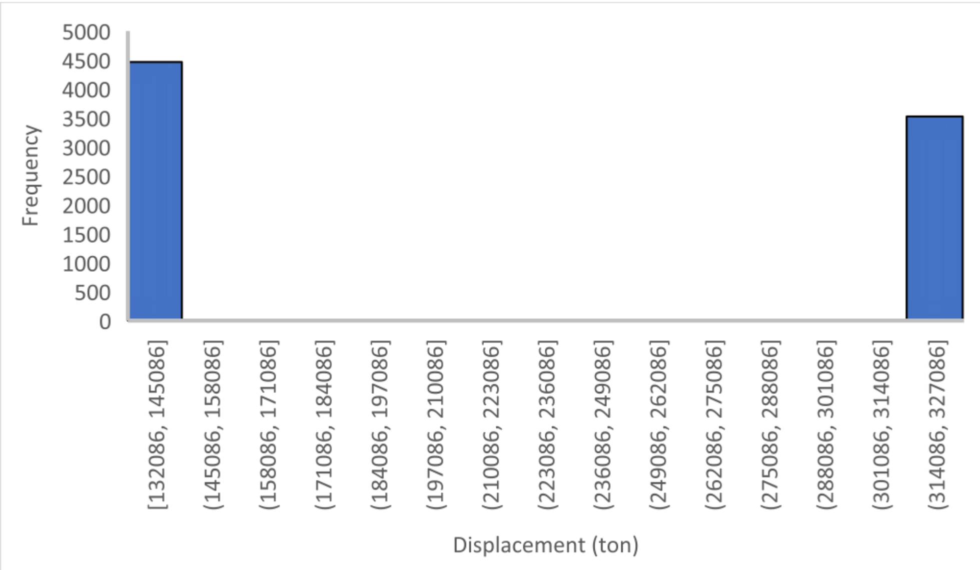


Figure 4-3 Displacement histogram for all data samples

First voyage: From 19/01/2017 until 10/02/2017. The displacement is between 132086 and 136086 tons, ballast voyage.

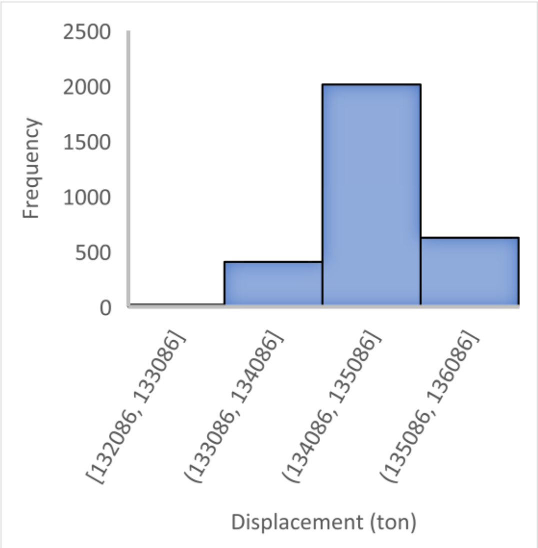


Figure 4-4 Displacement histogram first voyage

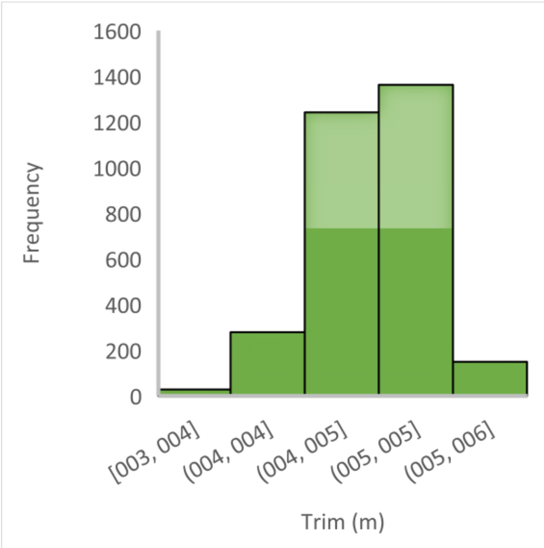


Figure 4-5 Trim histogram first voyage

Second voyage: From 18/02/2017 until 15/03/2017. The displacement is between 320000 and 325000 tons, loaded ship.

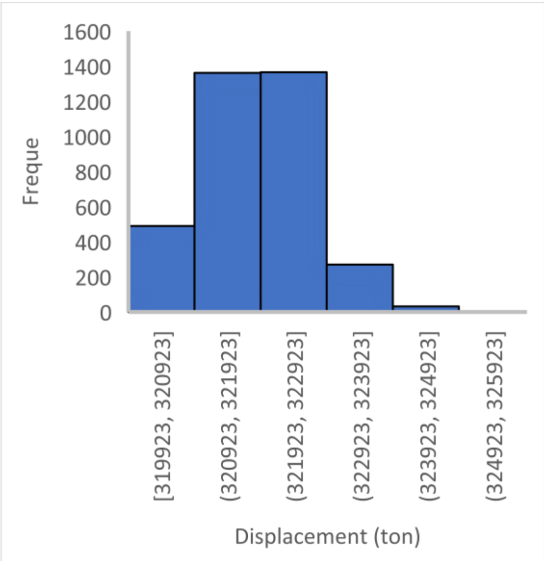


Figure 4-6 Displacement histogram second voyage

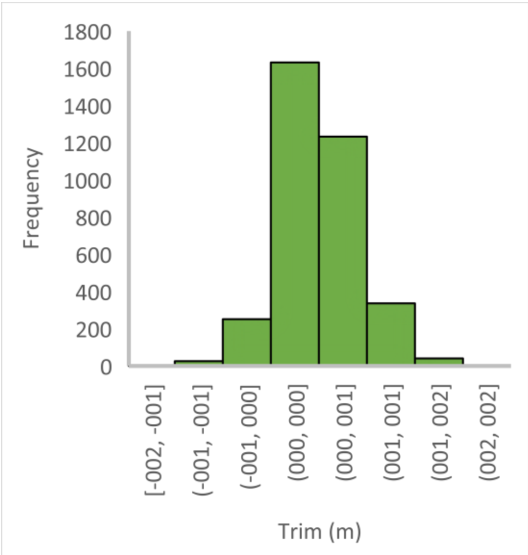


Figure 4-7 Trim histogram second voyage

Third voyage: From 22/03/2017 until 30/03/2017. The displacement is between 138443 and 140443 tons, ballast voyage.

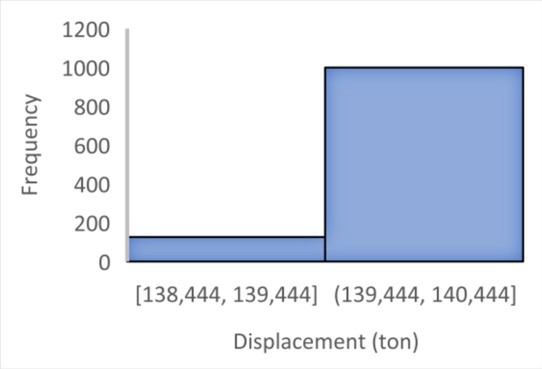


Figure 4-8 Displacement histogram third voyage

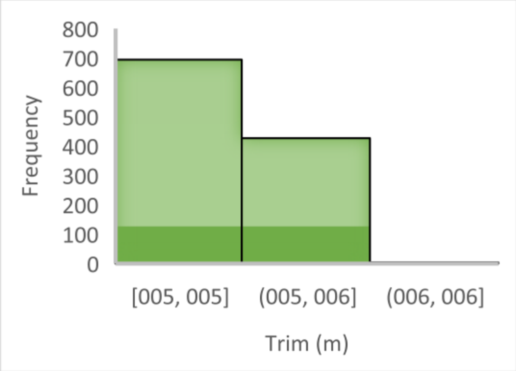


Figure 4-9 Trim histogram third Voyage

Figure 4-10 shows that there is a strong relationship between the trim distribution and the ship operational performance, which was not clear in the linear correlation analysis. The fuel consumption has different ranges corresponding to different trim ranges, which is the classical behaviour of ship operators who always set the same trim configuration for each loading condition, without considering other variables.

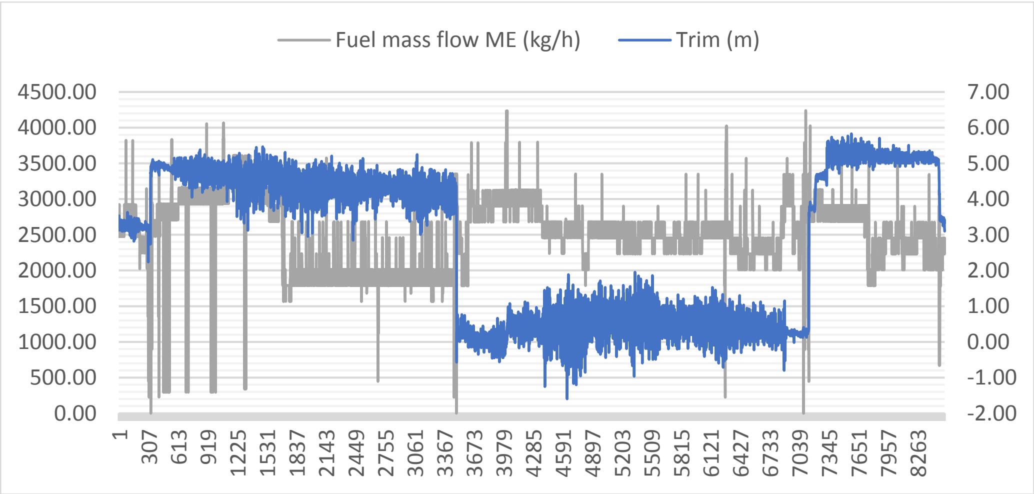


Figure 4-10 Fuel mass flow and trim data distribution, all dataset

During the different voyages, the ship encountered different weather conditions that influenced its operational performance. Apparent wind speed on ship direction is a strong variable to describe the weather encountered by the ship and its effect on the ship's performance. The correlation matrix (Figure 4-1) confirms also this relationship with the high correlation factor that the apparent wind speed had with fuel consumption. The figures 4-11, 4-12, 4-13 show the frequency of the different encountered wind speeds during each voyage. It can be seen that during the third voyage, high wind speed on ship direction were more frequent than on the other voyages. This could be considered to examine the potential of route optimization.

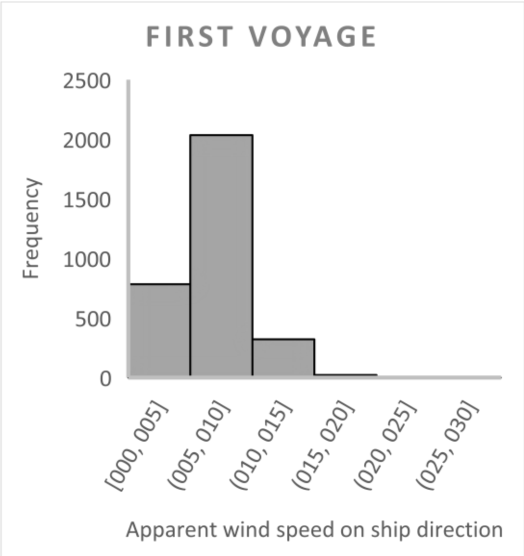


Figure 4-11 Histogram of apparent wind speed on ship direction first voyage

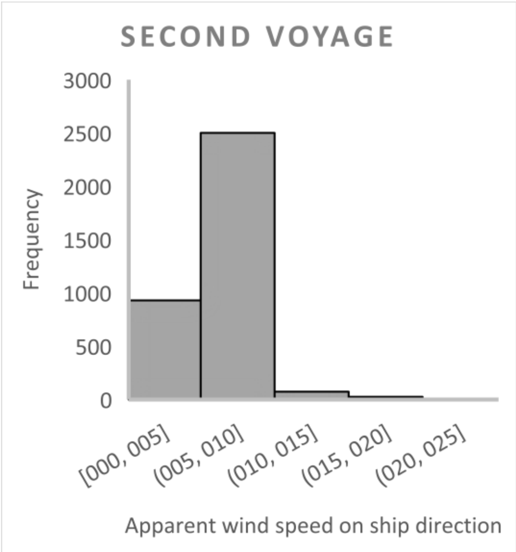


Figure 4-12 Histogram of apparent wind speed on ship direction second voyage

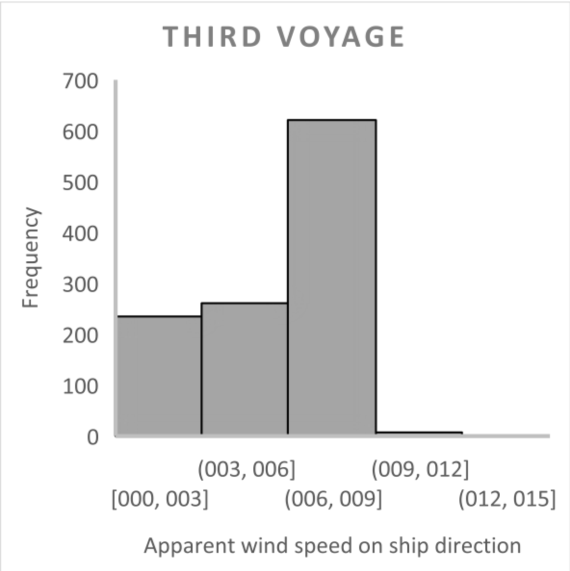


Figure 4-13 Histogram of apparent wind speed on ship direction third voyage

The data examination can further demonstrate that the fuel consumption is influenced by both the loading conditions and the weather conditions. The general trend of the fuel mass flow during the third voyage is compared to the true swell angle and swell height in Figures 4-14 and 4-15. Fuel consumption is at its maximum while the swell angle is at its minimum for the same voyage and vice versa. In addition, during most of the voyage samples, the fuel consumption profile decreases with an increased true swell angle. In addition, Figure 4-15 shows that the swell height when increasing affects the fuel consumption profile considerably. Both the swell angle and height are responsible for additional ship resistance, which could be avoided by possible course alteration in order to reduce fuel consumption.

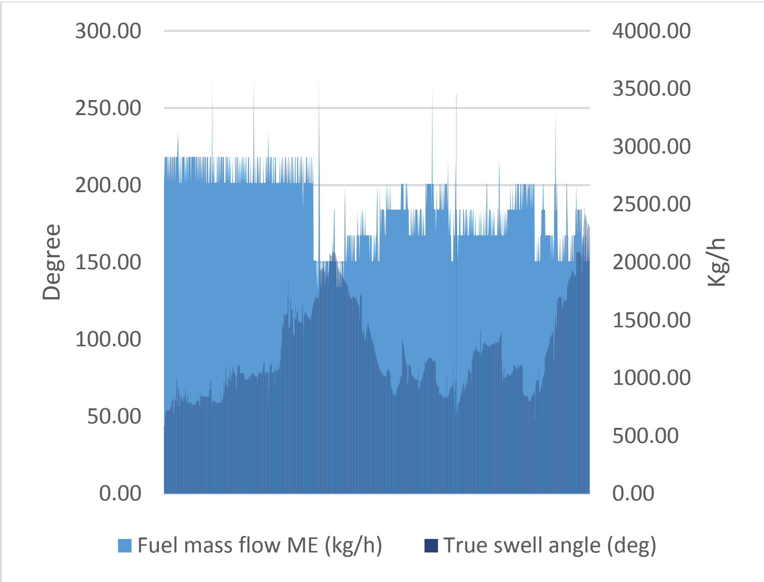


Figure 4-14 Fuel mass flow and true swell angle trend for third voyage

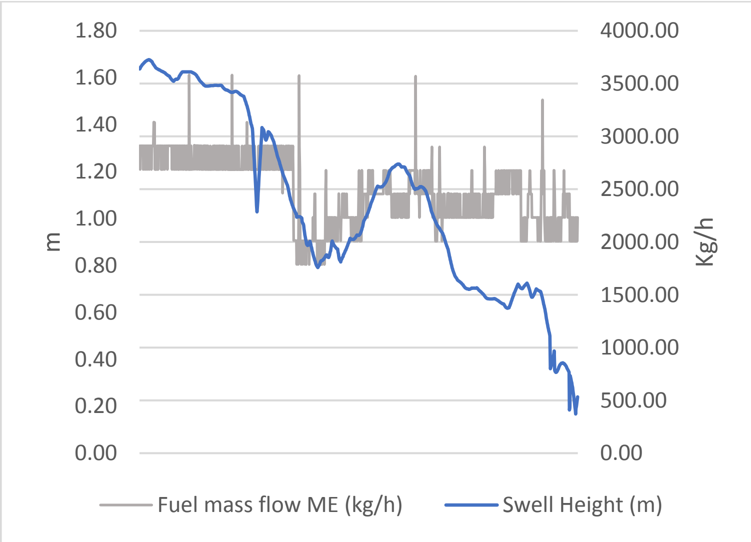


Figure 4-15 Fuel mass flow and swell height trend for third voyage

Finally, the dataset variables used for modelling are as follow;

<table>
<tr><td rowspan="2">Input variables</td><td>Heading (deg)</td><td>Speed over ground (m/s)</td><td>Displacement (t)</td><td>Trim (m)</td><td>App wind speed on ship direction (m/s)</td><td>Wind relative angle</td><td>Wind Wave Height (m)</td></tr>
<tr><td>Wind Wave Period (s)</td><td>Swell Height (m)</td><td>Swell Crossing Period (s)</td><td>Swell Direction (deg)</td><td>Current speed (m/s)</td><td>Current relative angle (deg)</td><td></td></tr>
<tr><td>Output variable</td><td colspan="7">Fuel mass flow ME total (kg/h)</td></tr>
</table>

Table 4-2 Final dataset variables for modelling

Total number of inputs: Thirteen variables.

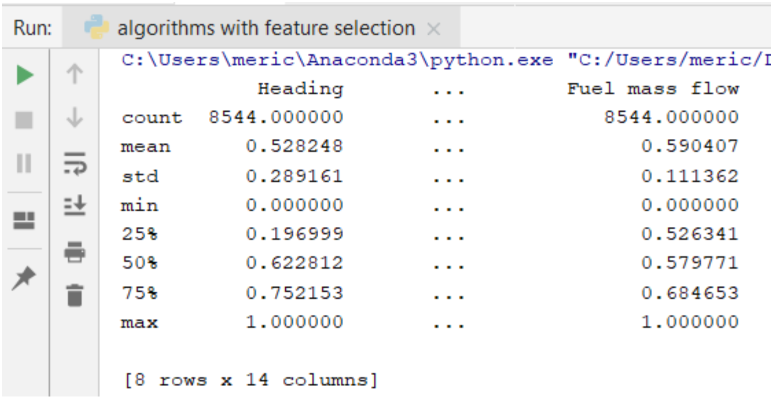
Ship specific inputs: Four (Heading, Speed, displacement and trim).

Weather inputs: two for wind, two for wave, two for current and three for swell

Output variable: one variable (fuel mass flow ME).

4.2.Models parameters

The dataset after cleaning resulted in 8544 samples instead of 9188 and the general description is presented in Figure 4-16 as loaded into Python Software.


Run: algorithms with feature selection

```
C:\Users\meric\Anaconda3\python.exe "C:/Users/meric/I
                 Heading       ...      Fuel mass flow
count   8544.000000       ...         8544.000000
mean       0.528248       ...            0.590407
std        0.289161       ...            0.111362
min        0.000000       ...            0.000000
25%        0.196999       ...            0.526341
50%        0.622812       ...            0.579771
75%        0.752153       ...            0.684653
max        1.000000       ...            1.000000

[8 rows x 14 columns]
```

Figure 4-16 The pre-processed dataset description in Python Software

4.2.1. Decision Tree and AdaBoost

During the training step, the max decision tree depth, which is the length from the root node to the leaf, should be defined and optimized in order to avoid overfitting. The optimum depth can be found only by experiment and model accuracy calculation. The tree giving best accuracy is selected, which is the one that has the best depth value. The depth of the tree for testing started with the random value four. After conducting many experiments, best accuracy was found at depth six. It was then used with AdaBoost to improve the weak learner.

4.2.2. K Nearest Neighbours (KNN)

For KNN, during training the parameter, K was first fixed at five and after many experiments, the model with four nearest neighbours was the best according to its prediction performance with the test dataset. The weight was represented by Kernel inversion as a function of the distance from neighbours as explained in Chapter 3. Due to time constraints, a deep search for the best weighting factors was not conducted and only a distance weighted algorithm with Kernel Inversion was used.

4.2.3. Artificial Neural Network

In order to build ANN model, two hidden layers may be always sufficient to solve any problem, with the first layer to save the local characteristics of the inputs and the second to extract the general features of the input patterns (Shahin, Jaksa & Maier, 2008). The same dataset was loaded to MATLAB 2015a to start the training of ANN with two hidden layers. After many training cycles, best results were recorded with 21 hidden layers. The number of input nodes is 13, which is the number of input variables and the output node is only one as presented in Figures 4-17 and 4-18.

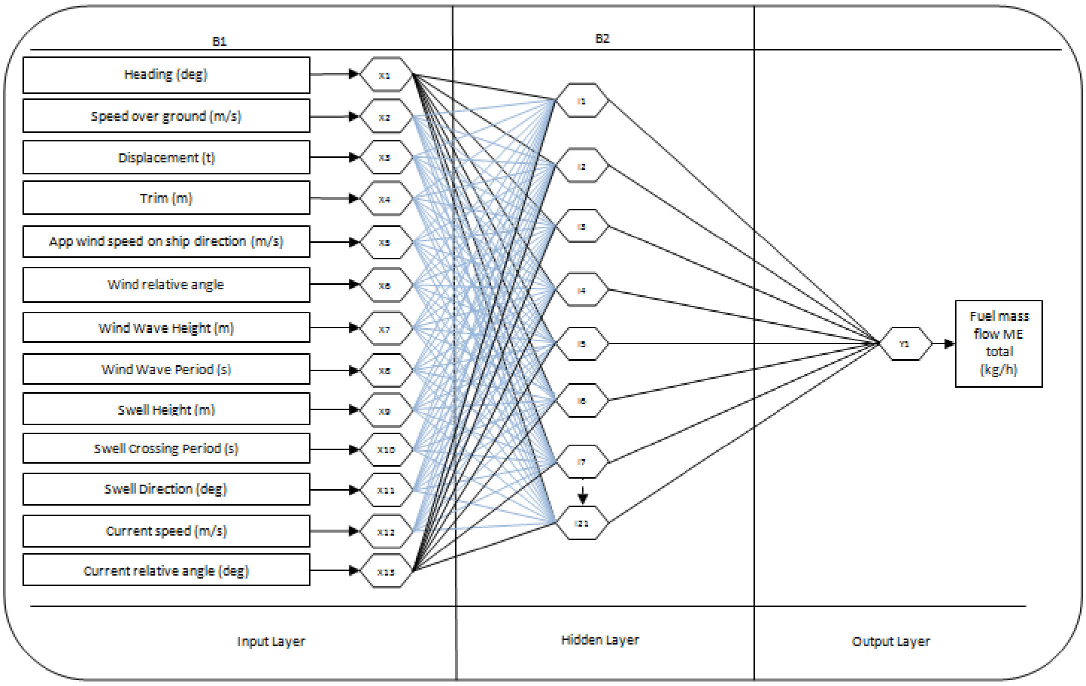


Figure 4-17 ANN model structure

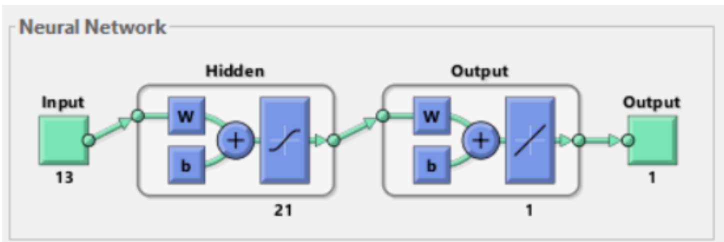


Figure 4-18 ANN Model structure, screenshot MATLAB2015a

In order to better understand and show the nature of the machine learning algorithms, which learn from their experience, and the ANN especially, a model with only the first voyage dataset (3060 samples) and same structure was tested, and all the results were compared to the main ANN model.

4.3. Optimization

4.3.1. Genetic Algorithm parameters

In this work, the population chromosomes ranking considers the fittest according to the fitness function. Ten per cent of the ranked elites are kept and 90% (poor solutions) will be replaced by crossover and mutation.

The two points' crossover was conducted with a percentage of 80% replacement of the poor chromosomes in order to have a high opportunity to find an optimal trim, which is in a very limited range.

The mutation replaces 10% of the population and does not follow any rule except respecting the defined constraints.

A solution is considered as optimum when the fitness value stops improvement and it gets worse for a certain number of iterations (200 iterations for this case). In order to have a logical time of processing, 120 seconds was defined as a time limit, which, if reached before the first condition, the search stops and shows the current optimal set of variable values that minimizes fuel consumption.

4.3.2. Scenarios

The speed, heading and trim are the variables that the ship operator can control and which have a potential to minimize fuel consumption and, thereby, optimize the ship voyage. However, the ship operator is able to vary the speed, heading and trim in a certain feasible interval that optimizes the voyage, while respecting the safety of the ship and its commercial engagement. Due to the lack of information on the ship times of arrivals and locations, the speed optimization, which is known for being highly successful in reducing fuel consumption, was not conducted. In this case, other scenarios of voyage optimization were tested.

4.3.2.1. Real time trim optimization for ballast voyage:

In order to show the potential of fuel savings of the VLCC case study ship during its voyage, the dataset of the third voyage with a displacement of 140000 tons was used to conduct a first trim optimization scenario. As shown in Figure 4-9, the trim during the third voyage was almost all the time between 4,72m and 5,72m with the displacement of a light ship. In order to be able to conduct the optimization with the information illustrated in the dataset, an hour of continuous sailing period was picked. The same data sample period of ten minutes was respected. The GA will be employed for the sample at each 10mn with the current trim value in order to find the optimal value from the voyage range for the same sample (same displacement and sea conditions).

Assumptions and constraints:

The other input values were assumed to change after each 10mn period of time as it was the frequency of data collection tools.

The trim range of 4,72-5,72m for the third voyage was assumed to be respecting the safety and stability conditions.

The constraint specified in the GA was the specified trim range.

4.3.2.2. Real time trim optimization for loaded ship

For this optimization scenario, the voyage considered is the one with the loaded ship. The displacement is 321000 tons. During this voyage, the histogram (Figure 4-7) shows the general range of the trim between –0,60m and 1,40m. The GA is employed again to find the optimal trim for each sample during one hour in order to examine the potential of savings for a loaded ship.

Assumptions and constraints:

The same assumptions were taken for this scenario with the new range for trim -0,60 and 1,40, which became the new constraint in the GA configuration.

4.3.2.3. Real time trim and route optimization for ballast voyage

In this case, the third voyage samples considered in the first trim optimization scenario are considered. The weather encountered by the ship during this voyage, as previously shown in Figures 4-13, 4-14 and 4-15, could be routed to allow for fuel savings with possible course alterations. The trim margin is 4,72-5,72 m. In addition, it is proposed to have a margin of course alteration of 20° East and 20° West to allow for possible route optimization considering the weather conditions. Twenty degrees was chosen randomly as a small range in order to give an idea of the ability of the optimization model to find optimal input variables values that would reduce the voyage CO2 emissions and fuel cost.

Assumptions and constraints:

The other input values were assumed to be unchanged during the 10mn period of time as it was the frequency of data collection tools.

The trim range of 4,72-5,72m for the third voyage was assumed to be respecting the ship safety and stability conditions.

The margin for course alteration of 20° East-West is assumed to be respecting the safety and stability conditions.

It is assumed that the new optimized heading will not result on additional travelled distance.

This scenario is informative, as some input variables such as relative current direction, are dependent on the ship heading and should be re-calculated to have the real values with the optimized heading. However, the GA while running will give an idea about the optimal heading that should have been set in the actual weather conditions to have less ship resistance and, thereby, lower fuel consumption. Due to time constraints, it was not possible to load the interdependent input variables as other functions of the independent ones.

The constraints specified in the GA were the specified trim and heading ranges.

4.4.Summary

In this chapter, the data is cleaned, transformed and filtered to include only sailing periods. The data visualization allows understanding of the hidden relations between variables and detecting of different voyages to separate them. The models parameters were set and predictive performance was quantified. The chapter presented the different optimization scenarios tested with the best predictive model associated with the GA.

All the results will be displayed in the next chapter and the fuel savings from each voyage optimization scenario will be assessed.

Chapter V: Discussion of results

In order to validate the developed methodology in this thesis, the results from the case study are illustrated in this chapter. Interpretation of the results will allow the tested models and the methodology's applicability as a decision support system (DSS) for energy efficient ship operation to be assessed. The chapter will conclude by presenting and explaining the proposed DSS.

5.1.Predictive results

The performance evaluation metrics of all models are shown in Figures 5-1 and 5-2 in the Software in normalized values. They were converted into real values of fuel mass flow (kg/h) and summarized in Table 5-1;

Model	MSE	RMSE	MAE	R^2
Decision Tree	12,99	234,76 kg/h	124,92 kg/h	0,74
AdaBoost DT	6,94	171,5 kg/h	111,55 kg/h	0,86
KNN	9,11	196,51 kg/h	119,31 kg/h	0,82
ANN	5,55	153,4 kg/h	/	0,96

Table 5-1 Models predictive performance

The Figure 5-1 shows a screenshot of the prediction results of the three models built in Python. The results are the error metrics normalized values that were converted and summarized above.

In Figure 5-2, a screenshot from MATLAB shows the minimum MSE found with the ANN structure of 21 hidden layers.

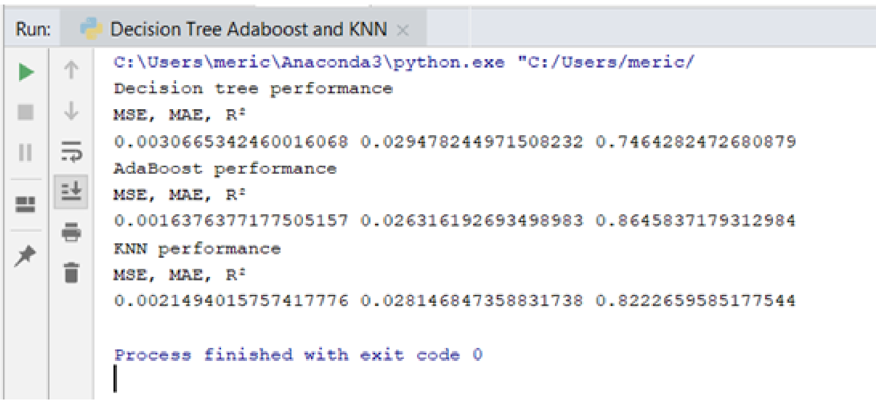


Figure 5-1 Decision Tree, AdaBoost and KNN Predictive performances, Python

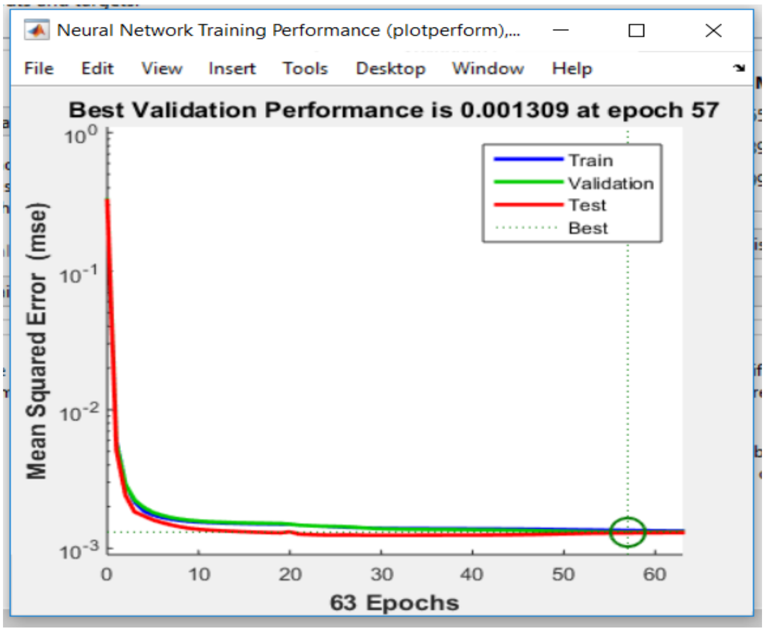


Figure 5-2 ANN Predictive performance, MATLAB

Not all the tested models performed adequately to predict the ship fuel consumption as the prediction accuracy of KNN and AdaBoost are only 82% and 86%. These values could be improved by a deeper search of the hyper-parameters of the algorithms with training. In addition, testing different weights for the input variables according to their detected effect on the fuel consumption, such a high weight for the speed having a great influence on the fuel consumption is a good technique to improve the KNN and AdaBoost accuracy.

ANN has the best prediction accuracy with 96% of correct predictions on test data (unseen data).

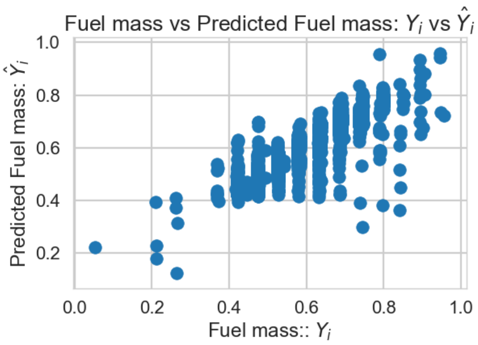


Figure 5-3 Plot of predicted fuel mass flow and the real fuel mass flow with KNN

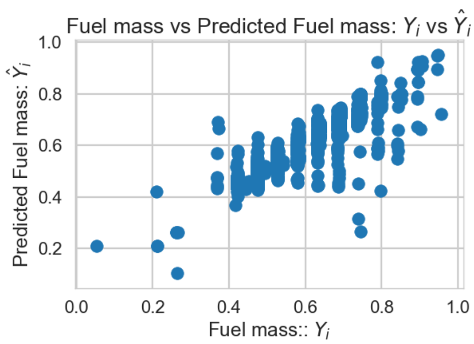


Figure 5-5 Plot of predicted fuel mass flow and the real fuel mass flow with AdaBoost

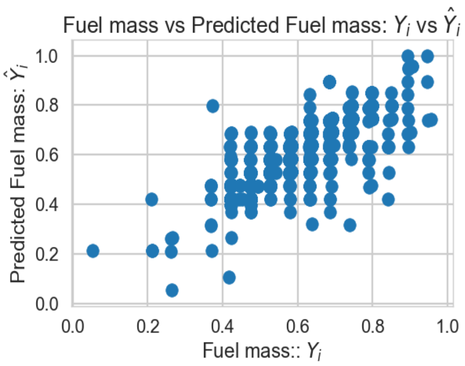


Figure 5-4 Plot of predicted fuel mass flow and the real fuel mass flow with Decision Tree

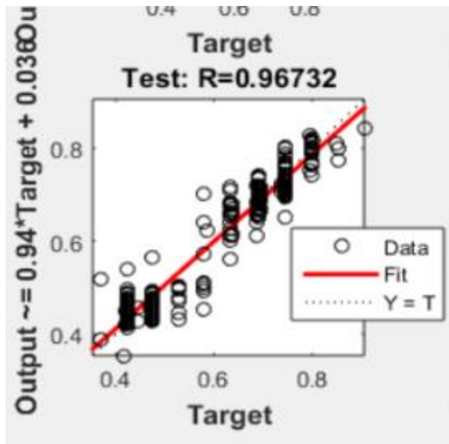


Figure 5-6 Plot of predicted fuel mass flow and the real fuel mass flow with ANN

In these figures (5-3 to 5-6), the predicted fuel mass flow of the test dataset was plotted against the real values in order to better see the regression between both values that represents the model accuracy. The figures show a significant improvement of the prediction between Decision Tree and AdaBoost Tree, which was the objective of Adaboosting. At low fuel mass flow values, many samples are far from the perfect regression line in all figures including ANN. This can be explained by unstable sailing time, where the ship is reducing its speed for a certain reason. Those samples were not sufficient for good model learning and could be improved by more data samples for these speeds. ANN was the best prediction model compared to all the other models. Moreover, in the prediction of high fuel consumption points, where the KNN and AdaBoost had noisy points with high error, ANN has had more accurate predictions.

It is difficult to directly compare the current models to the ones presented in the literature, as they do not have the same data and the same output variables. However, an indication of the performance of the built models can be presented. Petersen & Winther (2011) reported ANN error results with RMSE equal to 47L/h, which was better than the ANN model on shaft power prediction of Pedersen and Larsen (2009). The same research presented an RMSE of fuel consumption with Gaussian process as 52L/h. SVR, and KNN algorithms were used in Pétursson (2009) research to predict the shaft power and have not been used before for fuel consumption prediction. The model accuracy was presented on RMSE, which has not allowed a comparison. Bagging, Random Forest and Bootstrap were employed recently by Soner, Akyuz & Celik (2018) with the same data used by Petersen and Winther (2011) for a ferry ship and RMSE for fuel consumption were 45,2L/h, 43,5L/h and 41,3L/h respectively. Bootstrap performance was then better than ANN when used with the same dataset. ANN prediction results for fuel consumption were given by Bal Besikci, Arslan, Turan, & Olcer (2016) with ANN having better performance than Multiple Regression. The MSE and RMSE were 0,037 and 0,193mt/h. The ANN model results in this thesis are better with MSE and RMSE of 5,5 and 153kg/h for fuel consumption. In general, they are good results compared to all the cited research results. However, as already explained a direct comparison is not possible because the ship case for this study is a VLCC and the data samples are very different as the voyages are not short and similar to the ones made by a car ferry.

In order to confirm this conclusion, the results of the model built using only the first voyage with 3060 samples of same displacement and trim range are presented in

Figures 5-7 and 5-8. ANN model was over-fitted having a very high accuracy with this dataset.

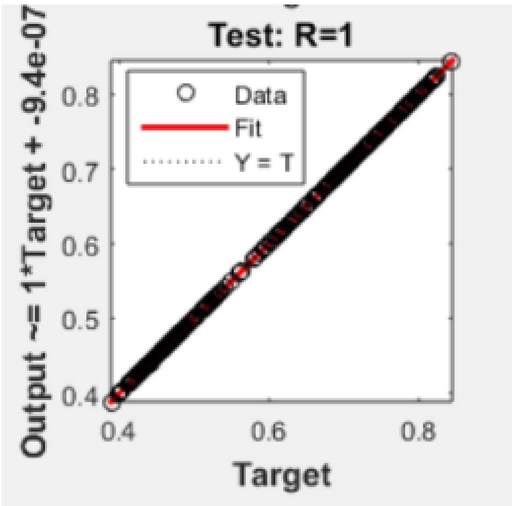


Figure 5-7 First voyage ANN model predictions

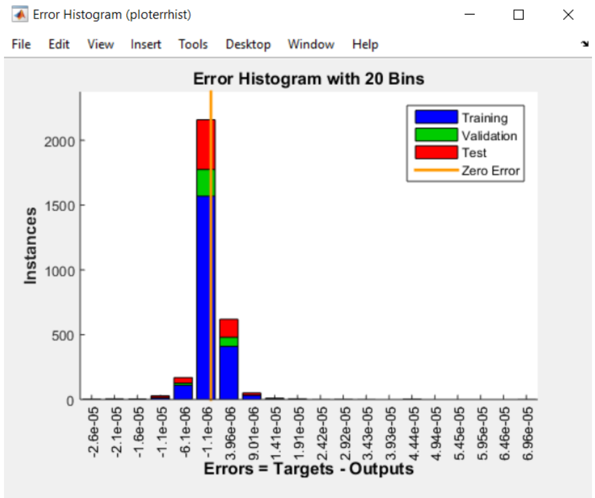


Figure 5-8 First voyage ANN error histogram

This gives a strong justification for the current main model error margin. The whole dataset is for a period of three months with the ship having different loading conditions and weather status. However, the number of samples on ship operation when it is loaded may not have been sufficient to train the model on all operational conditions in order to learn the performance pattern and be able to predict it with new data. A larger dataset for a longer period, including many data samples for different operational conditions, is always preferred to build black box models for ship performance prediction.

As a result, the ANN fuel consumption prediction model is considered a reliable model for future ship fuel consumption prediction with high accuracy compared to what was presented in the literature. In addition, as it is the best among the tested models in this study, its predictions are used in order to optimize normal ship operational performance.

5.2. Optimization results

Figure 5-9 shows a screenshot from the MATLAB Global Optimization Toolbox while the search for optimal solution to minimize fuel consumption is running with GA. The fitness value and the best individual vector are selected to be displayed while running.

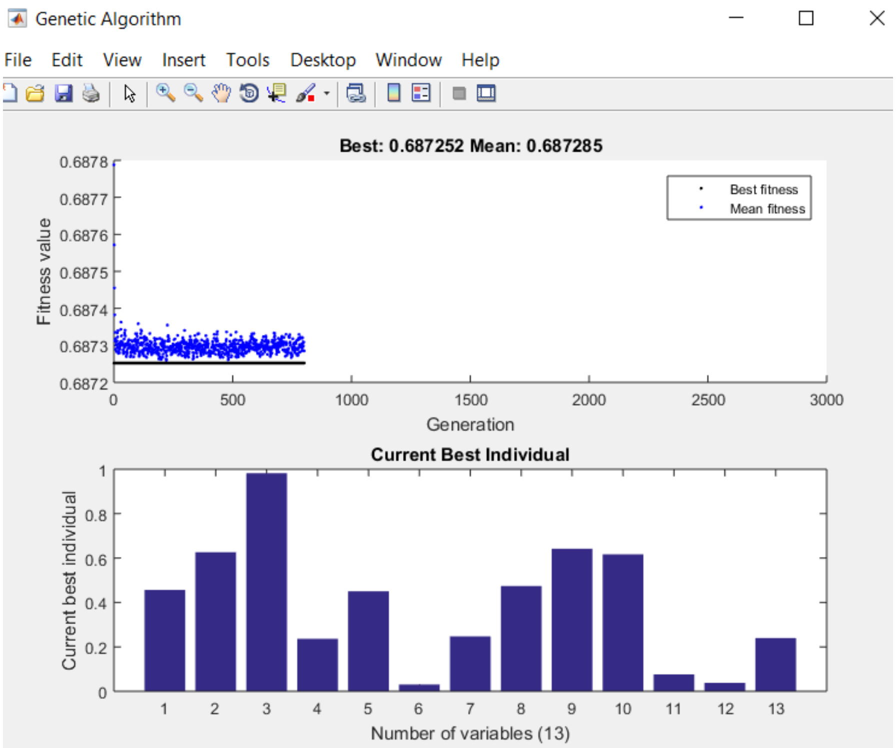


Figure 5-9 Screenshot while GA optimization is running

The tables 5-2, 5-3, 5-4 include the reference values of the trim, fuel consumption and ship heading for the samples of the different voyage optimization scenarios introduced to the GA optimization model. The new decision variable values are the optimal solution for the ship fuel minimization problems.

Displacement 140000 ton									
TRIM optimization scenario for 24 hours									
Date	Base case trim (m)	Optimized Trim (m)	Trim change (cm)	Base case fuel mass flow (kg/h)	Optimized fuel mass flow (kg/h)	Sample fuel saving (kg)	Total fuel saving (kg/day) and percentage	Fuel cost savings per day	CO2 emission reduction (kg CO2/day)
26/03/2017 14:16	5,35	4,71	64	2454,21	2430,92	3,88	747,65 kg/day (1,27%)	525$/day	2317,73
26/03/2017 14:26	5,00	4,70	30	2455,19	2420,59	5,77			
26/03/2017 14:36	5,31	4,70	61	2455,19	2427,27	4,65			
26/03/2017 14:46	5,20	4,70	50	2455,19	2429,05	4,36			
26/03/2017 14:56	5,11	4,70	41	2455,19	2437,46	2,96			
26/03/2017 15:06	5,19	4,70	49	2455,19	2397,96	9,54			

Table 5-2 First ship voyage optimization scenario, trim optimization for ballast voyage

Displacement 321000 ton									
TRIM optimization scenario for 24 hours									
Date	Base case trim (m)	Optimized Trim (m)	Change in the Trim (cm)	Base case fuel mass flow (kg/h)	Optimized fuel mass flow (kg/h)	Sample fuel saving (kg)	Total fuel saving (kg/day) and percentage	Fuel cost savings per day ($)	CO2 emission reduction (kg CO2/day)
13/03/2017 14:15	0,28	1,02	74,02	3124,71	3116,07	1,44	2347 kg/day (3,13%)	1647$/day	7278,60
13/03/2017 14:25	0,30	1,10	80,32	2901,87	2860,94	6,82			
13/03/2017 14:35	0,33	1,10	76,59	3125,70	3102,63	3,85			
13/03/2017 14:45	0,31	1,10	79,45	3571,38	3121,00	75,06			
13/03/2017 14:55	0,28	-0,22	49,82	2901,87	2895,27	1,12			
13/03/2017 15:05	0,22	-0,18	39,59	3124,71	3063,11	10,27			

Table 5-3 Second ship voyage optimization scenario, trim optimization for loaded ship

		Displacement 140000 ton								
TRIM and ROUTE optimization scenario for 24 hours										
Date	Base case trim (m)	Optimized Trim (m)	Base case course (deg)	Optimized course (deg)	Base case fuel mass flow (kg/h)	Optimized fuel mass flow (kg/h)	Sample fuel saving (kg)	Total fuel saving (kg/day) and percentage	Fuel cost savings per day	CO2 emission reduction (kg CO2/day)
26/03/2017 14:16	5,35	4,74	313,90	295,00	2454,21	2401,83	8,73	1328 kg/day (2,25%)	930$/day	4118,18
26/03/2017 14:26	5,00	4,73	316,40	295,00	2455,19	2392,56	10,44			
26/03/2017 14:36	5,31	4,73	315,70	295,00	2455,19	2403,77	8,57			
26/03/2017 14:46	5,20	4,73	315,80	295,00	2455,19	2405,83	8,23			
26/03/2017 14:56	5,11	4,71	316,20	295,00	2455,19	2420,39	5,80			
26/03/2017 15:06	5,19	4,73	316,00	295,00	2455,19	2373,69	13,58			

Table 5-4 Third ship voyage optimization scenario, trim and route optimization for ballast voyage

After applying the optimization model to the first scenario, it has been found that the vessel did not have the optimal trim during the first voyage samples. The adjustment of the trim by 30cm to 64cm could result in fuel savings of 747kg/day. Generally, for a ballast voyage, the GA real time trim optimization saved 1,27% of the fuel cost and the CO2 emission of the ship during 24hours. This perfectly meets the aim of this research, which contributes to reducing CO2 emissions from ships.

In the second case, the ship was loaded and it is known that the ship operators usually have a small margin to change the trim in this case. However, considering the assumptions made for this scenario, it has been demonstrated that changing the actual trim by 39cm to 80 cm could save 3,13% of the fuel consumed per day. This is a greater economy than the scenario of ballast voyage, which is justified by the greater resistance that the ship has when the underwater hull is bigger.

The third scenario results confirmed the same optimal trim as the first scenario, which had the same samples for testing. It has almost the same trim changes between 30cm and 65cm. In this case, the optimal trim was combined with the optimal route to save 2,25% of the fuel consumed per day, compared to 1,27% for the same case with only trim optimization.

It should be mentioned that the possible fuel savings with trim optimization presented above might not be a net gain if the ballast exchange process is not a gravity assisted one because transferring the ballast water in this case, will be associated with additional consumed energy.

5.3. The prediction and optimization models as a Decision Support System for Energy Efficient Ship Operation

The ANN model has shown good predictive results and while called into GA for optimization, both methods employed together have shown their success to optimize ship voyage performance. These machine-learning algorithms, combined can perform as a Decision Support System (DSS) for energy efficient ship operation. The input variables could be fed to the system from the different ship navigation equipment and weather forecast receivers and the build system will, first, predict the fuel consumption from this input data, and then run an optimization cycle. The results are the optimal decision variables values for minimum fuel consumption, such as optimal trim and heading. This helps the ship operators to technically define their constraints and decide on their voyage plan based on the results displayed in the system. Figure 5-10 explains the information flow in the proposed DSS.

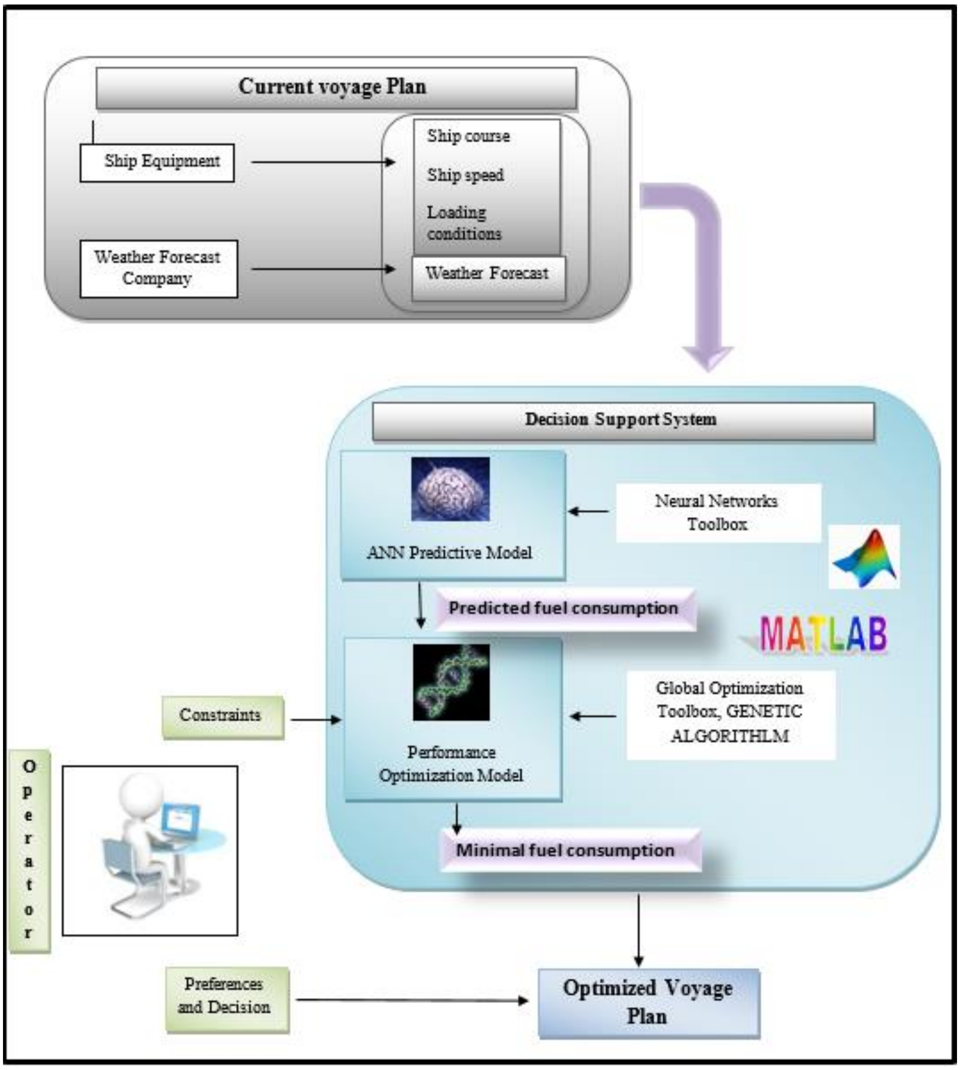


Figure 5-10 The Decision Support System

5.4.Summary

The tested models with the case study performed well with the unseen data. The ANN model was the best among all presented models, confirming that it is viable in estimating ship fuel consumption for future ship operation. The model then employed with the GA to form the DSS performed very well to make significant fuel savings in all the tested scenarios. As a result, the whole methodology developed in this thesis is considered as very successful in making fuel consumption predictions and optimizing ship operation with respect to the fuel usage.

Chapter VI: Conclusion and future research

Ship fuel consumption can represent more than 50% of ship voyage cost. With the new IMO strategy that has fixed an objective of reducing GHG emissions from ships by 50% before 2050, the future regulations may include market-based measures that will make the fuel voyage cost much greater than now. Furthermore, the environmental impact of the mass of the carbon emitted by ships is more than three times the mass of the fuel they burn, which has a major impact on the ecosystem. Therefore, from both, financial and environmental points of view, voyage optimization is extremely needed nowadays and ship operators are going to be more and more interested in reducing ship fuel consumption in order to be able to survive in a very competitive market. With increasingly stringent regulations for the shipping industry, the ship operators' incentives have become not only economic but also environment and rule-driven. This has resulted in intensive research to test new voyage optimization methods using the modern machine learning tools to build DSS for energy efficient ship operation.

This research examined the applicability of black box models to predict the ship fuel consumption of a vessel for different ship and weather conditions and combined them with the Genetic Algorithm as an optimization model in order to find an optimal trim configuration and route options with respect to fuel usage. The aim was to build a combined model to be used as a DSS for potential fuel savings and CO_2 emission reduction.

The investigation has been possible with the use of numerical data sampled from a VLCC with 320000Dwt.

The data pre-processing was given special attention. The main dataset was segmented corresponding to the voyages of the ship. The harbour periods were excluded. Data series with many missing instances were removed. Inconsistent data were filtered and outliers were excluded. Finally, all the data were normalized and standardized in order to make them well formed to be ready to use in black box modelling.

The number of input variables selected for programming was a large one, including thirteen input variables from ship conditions (speed, trim, displacement…) and weather conditions (wind, current, swell…). This is a highlight of this research comparing to what appeared in the literature with lower number of variables or with a set of variables excluding the ship loading conditions or other external conditions.

The output variable was the fuel mass flow of the main engines as it represents the fuel consumed by the propulsive system, which was the subject of the optimization.

Four statistical models were tested with the dataset, Decision Tree, AdaBoosted Decision Tree, K-Nearest Neighbours and Artificial Neural Network. Another highlight of this research is the comparative analysis between these models, conducted based on their accuracy in order to decide on the best model to use in the DSS.

The non-parametric black box models, AdaBoost and KNN had not been used before to predict ship fuel consumption, which made their investigation in this research very interesting. Their performance to estimate ship fuel consumption for unseen data (different from training data) was not sufficiently adequate to be employed for ship voyage optimization and, due to time constraints, the selection of their hyper-parameters was not deeply investigated. ANN performance was the best among the tested models and turned out to be adequate for making fuel consumption predictions on test data (unseen data samples). ANN was also performant when compared to the black box models used to predict the ship operational performance found in the literature. Moreover, a similar ANN model was built using only one voyage data subset in order to justify the main model's error margin and explain the importance of data understanding and its role in avoiding over-fitness when using statistical models. The ANN prediction model has an acceptable error in estimating fuel consumption, which is related to the nature of the ship data, which was not large enough in some conditions to allow a complete accurate predictive model to be built. Collecting large datasets with a complete range of operational conditions is not evident as the weather and sea conditions are not controllable. However, testing the black box models with broader datasets could generate highly performant models.

Subsequently, the main ANN model has been employed to predict ship fuel consumption as a first step in the DSS.

The genetic Algorithm, one of the most popular and successful evolutionary algorithms to solve optimization problems, was employed for the first time in the operational performance optimization. The objective was to minimize the ship fuel consumption by finding the optimal set of decision variables from a defined space of search. Fuel consumption was predicted by the ANN model and optimized by GA, which is a novel combination proposed in this thesis. The space to search the vector of input variables values solution to this optimization problem was defined by different constraints on decision variables. The proposed DSS was tested in three different scenarios and showed its success in making important fuel savings up to 2,25% by real time trim optimization, and 3,13% by real time trim and route optimization. It is an effective real time optimization with ten minutes time sampling, which is the same as the sensors feedback. The DSS can then be used on-board the vessel to assist the shipmaster to make real time decisions that may result in potential fuel savings by cost-free and simple actions.

The method presented in this research was applied to a VLCC and can likewise be tailored and applied to other ship types. The objective is to introduce this DSS to a global energy management system on-board a merchant ship in order to reduce GHG emissions from shipping.

6.1. Future Research

Due to time constraints, the weights for the KNN model were not further improved, which should be considered in future work. Weighting the different input variables could also be deeply examined in order to make more accurate output prediction with different models. This could be achieved by testing different sets of input variables from the main dataset and comparing the prediction performance of the different models with different hyper-parameters.

Investigating grey box models in modelling ship operational performance should be considered in future research in order to improve the model's accuracy while using less historical data.

Some input parameters such as ship speed could be introduced as a function of the independent parameters and a multi-objective ship performance optimization by the GA could be considered, such as minimizing fuel consumption while maximizing ship speed.

The number of thirteen input variables could be extended and the hull condition could be considered to improve the model's accuracy.

Applying this DSS on-board ships will require an autonomous method of retraining the model as the ship's physical data is continuously changing and thereby changing its total resistance.

Employing the same method to other ship types will further contribute to understanding the differences in the trim optimal configuration of different ship types.

A graphical user interface (GUI) for the DSS has to be built in order to facilitate the implementation of the system and take advantage of easily displayed advice messages.

References

ABS. (2015). Ship Energy Efficiency Measures Status and Guidance. Retrieved from https://www.eagle.org/eagleExternalPortalWEB/ShowProperty/BEA Repository/References/Capability Brochures/ShipEnergyEfficiency

Aldous, Lucy Gemma. (2016). Ship Operational Efficiency: Performance Models and Uncertainty Analysis. 2016, discovery.ucl.ac.uk/1477486/1/Lucy%20Aldous%20Thesis_correctedv2.pdf. Accessed 30 Aug. 2018.

Al-Hamad, K., Al-Ibrahim, M., & Al-Enezy, E. (2012). A Genetic Algorithm for Ship Routing and Scheduling Problem with Time Window. American Journal of Operations Research, 02(03), 417-429. doi:10.4236/ajor.2012.23050 Amarel, S., & Steinberg, L. (1990, November 1). Artificial Intelligence and Marine Design. Retrieved from https://pdfs.semanticscholar.org/fb9f/f720744a12c9cc61cff544c46a4ab79ebe86.pdf

Amarel, S., & Steinberg, L. (1990, November 1). Artificial Intelligence and Marine Design. Retrieved from https://pdfs.semanticscholar.org/fb9f/f720744a12c9cc61cff544c46a4ab79ebe86.pdf

Amerongen, J., Duetz, H., & Okawa, T. (2017, June 21). A Low Cost Adaptive Autopilot for Inland Ships. Retrieved from https://www.sciencedirect.com/science/article/pii/S1474667017595695

Armstrong, V. (2013, August 09). Vessel optimisation for low carbon shipping. Retrieved from https://www.sciencedirect.com/science/article/pii/S0029801813002643

Bagheri, H., & Ghassem, H. (2014). GENETIC ALGORITHM APPLIED TO OPTIMIZATION OF THE SHIP HULL ... Retrieved from https://hrcak.srce.hr/file/191513

Bal Besikci, E., Arslan, O., Turan, O., & Olcer, A. I. (2016). An artificial neural network based decision support system for energy efficient ship operations. Computers & Operations Research, 66, 393-401. doi:https://doi.org/10.1016/j.cor.2015.04.004

Basheer, I., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. Microbiological Methods, 43(1), 3-31. doi:https://doi.org/10.1016/S0167-7012(00)00201-3

Breiman L, Friedman J, Olshen R, Stone C. (1984). Classifcation and regression trees. Boca Raton: CRC Press; 1984.

Chouvarda, I., Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., & Vlahavasa, I. (2017, January 08). Machine Learning and Data Mining Methods in Diabetes Research. Retrieved July 29, 2018, from https://www.sciencedirect.com/science/article/pii/S2001037016300733

D.Vose, M. (2003, February 19). Generalizing the notion of schema in genetic algorithms. Retrieved July 31, 2018, from https://www.sciencedirect.com/science/article/pii/000437029190019G

De¸bski, W. (2010). Advances in Geophysics. Retrieved July 31, 2018, from https://www.sciencedirect.com/bookseries/advances-in-geophysics

Demirel, Y., Turan, O., & Incecik, A. (2016, December 13). Predicting the effect of biofouling on ship resistance using CFD. Retrieved from https://www.sciencedirect.com/science/article/pii/S0141118716305685

Flood, I., & Kartam, N. (1994). Neural Networks in Civil Engineering. I: Principles and Understanding. Journal of Computing in Civil Engineering, 8(2), 131-148. doi:10.1061/(asce)0887-3801(1994)8:2(131)

Ghanshyam, G., Mirjalili, S., K.Patel, V., & J.Savsani, V. (2018, April 13). An improved heat transfer search algorithm for unconstrained optimization problems. Retrieved July 31, 2018, from https://www.sciencedirect.com/science/article/pii/S2288430017302099

Goldberg, D. (1989). Genetic algorithms in searching, optimisation and machine learning. Reading, MA: Addison Wesley

Guha, Amitava, and Jeffrey Falzarano. (2015) "Application of Multi Objective Genetic Algorithm in Ship Hull Optimization." Egyptian Journal of Medical Human Genetics, Elsevier, 10 June 2015, www.researchgate.net/profile/Amitava_Guha2/publication/279198614_Application_of_Multi_Objective_Genetic_Algorithm_in_Ship_Hull_Optimization/links/55e4e80808aecb1a7ccb8e28/Application-of-Multi-Objective-Genetic-Algorithm-in-Ship-Hull-Optimization.pdf.

Hamm, Jihun. Optimization for Machine Learning (in a Nutshell). mr-pc.org/t/cse5526/pdf/07b-optimization.pdf. Accessed 31 July 2018.

Haranen, M., Salo, J., Pakkanen, P., & Kariranta, R. (2016, April). White, Grey and Black-Box Modelling in Ship Performance Evaluation. Retrieved July 27, 2018, from https://www.researchgate.net/publication/301355727_White_Grey_and_Black-Box_Modelling_in_Ship_Performance_Evaluation

Hechenbichler, K., & Schliep, K. (2004, October 13). Weighted k-Nearest-Neighbor Techniques and Ordinal Classification. Retrieved August 17, 2018, from https://epub.ub.uni-muenchen.de/1769/1/paper_399.pdf

Hinnenthal, J. (2008). Robust Pareto – Optimum Routing of Ships utilizing Deterministic and Ensemble Weather Forecasts. Retrieved from https://www.dms.tuberlin.de/fileadmin/fg3/Publikationen/Diss_und_Habilitation/Dissertationen_hinnenthal_joern.pdf

Holland, J. H. (1975) Adaptation in Natural and Artificial Systems. University of Michigan Press.

Hirayama, A., & Ando, J. (2007). Study on Multiobjective Hull Optimization for Reducing Wave-making Resistance and Wave-breaking. Journal of the Japan Society of Naval Architects and Ocean Engineers, 5(0), 185-193. doi:10.2534/jjasnaoe.5.185

IMO. (2014)“34-Mepc-67-Emissions // IMO’s MEPC Progresses Work on Air Pollution and Energy Efficiency.” International Convention for the Prevention of Pollution from Ships (MARPOL), 23 Oct. 2014, www.imo.org/en/mediacentre/pressbriefings/pages/34-mepc-67 emissions.aspx.

IMO. (2009). Second-IMO-GHG-Study-2009. Retrieved August 11, 2018, from http://www.imo.org/en/OurWork/Environment/PollutionPrevention/AirPollution/Pages/Second-IMO-GHG-Study-2009.aspx

IMO. (2016, October 28). 28-MEPC-data-collection-- New requirements for international shipping as UN body continues to address greenhouse gas emissions. Retrieved August 11, 2018, from http://www.imo.org/en/MediaCentre/PressBriefings/Pages/28-MEPC-data-collection--.aspx

IMO. (2018). Low carbon shipping and air pollution control. Retrieved August 11, 2018, from http://www.imo.org/en/MediaCentre/HotTopics/GHG/Pages/default.aspx

ISS-AS. (2005, November 18). CHAPTER 8: NONPARAMETRIC METHODS. Retrieved from http://www.cs.wichita.edu/~sinha/teaching/fall17/cs697AB/slide/ch8.pdf

Journée, J. (2003, August). Review of the 1985 Full-Scale Calm Water Performance Tests Onboard m.v. Mighty Servant 3. Retrieved from http://mararchief.tudelft.nl/file/1455/

Journé, J., Rijke, R., & Verleg, G. (1987). Marine Performance Surveillance with a Personal Computer. Retrieved August 20, 2018, from http://mararchief.tudelft.nl/catalogue/entries/8709/

Knight, Michelle. (2018). What Is Machine Learning? 26 Feb. 2018, www.dataversity.net/what-is-machine-learning/. Accessed 29 July 2018.

Konar, A. (2000). Artificial intelligence and soft computing behavioral and cognitive modeling of the human brain. Boca Raton, FL: CRC Press.

Leifsson, L., Sævarsdóttir, H., Sigurðsson, S., & Vésteinsson, A. (2008, April 04). Grey-box modeling of an ocean vessel for operational optimization. Retrieved from https://www.sciencedirect.com/science/article/pii/S1569190X08000488

Lin, Y., Fang, M., & Yeung, R. (2013, October 18). The optimization of ship weather-routing algorithm based on the composite influence of multi-dynamic elements. Retrieved from https://www.sciencedirect.com/science/article/pii/S0141118713000679

Lindstad, H., & Eskeland, G. (2015, November 06). Low carbon maritime transport: How speed, size and slenderness amounts to substantial capital energy substitution. Retrieved from https://www.sciencedirect.com/science/article/pii/S1361920915001583

Logan, K. (2011, May 25). Using a Ships Propeller for Hull Condition Monitoring Mkt1. Retrieved July 28, 2018, from http://macsea.com/wp-content/uploads/Using-a-Ships-Propeller-for-Hull-Condition-Monitoring-Mkt1.pdf

Marie, S., & Courteille, E. (2009, January). Multi -Objective Optimization of Motor Vessel Route. Retrieved from https://www.researchgate.net/profile/Eric_Courteille/publication/265919086_MultiObjective_Optimization_of_Motor_Vessel_Route/links/54216a710cf203f155c66e98/Multi-Objective-Optimization-of-Motor-Vessel-Route.pdf

Markov, Zdravko. (n.d). Data Preprocessing. www.cs.ccsu.edu/~markov/ccsu_courses/datamining-3.html. Accessed 29 July 2018.

McGookin, E. W., Murray-Smith, D. J. & Li, Y. (1996). Submarine sliding mode controller optimisation using genetic algorithms. International conference on control '96, Vol. 1 (pp. 424}429). UK; Exeter Mitchell.T. Machine learning. 07-042807-7, McGraw Hill (1997), p. 2

Mitchell, T. (1997, March 1). Machine Learning. Retrieved from https://www.cs.ubbcluj.ro/~gabis/ml/ml-books/McGrawHill - Machine Learning -Tom Mitchell.pdf

Olcer, A. (2007, February 01). A hybrid approach for multi-objective combinatorial optimisation problems in ship design and shipping. Retrieved from https://www.sciencedirect.com/science/article/pii/S0305054806003145

Padhy, C. P., Sen, D., & Bhaskaran, P. K. (2007, May 17). Application of wave model for weather routing of ships in the North Indian Ocean. Retrieved from https://s3.amazonaws.com/academia.edu.documents/46700435/Application_of_wave_model_for_weather_ro20160622-9744-1obgz80.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1537204067&Signature=OiE7mtdxrbp lrT1qikij5CbPko=&response-content-disposition=inline; filename=Application_of_wave_model_for_weather_ro.pdf

Park H. I. (2011), Study for application ofartificial neural networks in geotechnical problems, INTECH Open Access Publisher, Rijeka, Croatia.

Pedregosa, F., Grisel, O., Thirion, B., Michel, V., Gramfort, A., & Varoquaux, G. (2011). Scikit-learn, Machine learning in Python. Retrieved August 11, 2018, from http://scikit-learn.org/stable/modules/model_evaluation.html#regression-metrics

Pedersen BP, Larsen J (2009) Prediction of full-scale propulsion power using artificial neural networks. In: Proceedings of the 8th international conference on computer and IT applications in the maritime industries (COMPIT'09), Budapest, Hungary May 10–12, pp 537–550. http://www2.imm.dtu.dk/pubdb/p.php?5840

Perera, L., & Mo, B. (2016, July). Machine Intelligence for Energy Efficient Ships: A Big Data Solution. Retrieved July 31, 2018, from https://www.researchgate.net/publication/290438979_Machine_Intelligence_for_Energy_Efficient_Ships_A_Big_Data_Solution

Petersen, J. P., & Winther, O. (2011). Mining of Ship Operation Data for Energy Conservation. Kgs. Lyngby, Denmark: Technical University of Denmark (DTU). (IMM-PHD-2011; No. 264). Retrieved from http://orbit.dtu.dk/files/10015226/phd264_jppe_2_.pdf

Pétursson , S. (2009). Predicting Optimal Trim Configuration of Marine Vessels with Respect to Fuel Usag. 2009, skemman.is/bitstream/1946/3058/1/thesis_960_fixed.pdf. Processes. Gulf Professional Publishing

Rakhshandehroo G. R., Vaghefi M., Aghbolaghi M. A. (2012), Forecasting groundwater level in Shiraz plain using artificial neural networks. Arabian Journal for Science and Engineering 37(7): 1871-1883.

Raschka, Sebastian. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. 11 June 2016, sebastianraschka.com/blog/2016/model-evaluation-selection-part1.html. Accessed 30 July 2018.

Ripley, B. (1996). Pattern Recognition via Neural Networks. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.51.1817&rep=rep1&type=pdf

Rocchi, R. (1994). Delivered Power - Trim And Sinkage - Ship Hull Forms: An Analysis Of The Effects Of The Trim Variations On The Power Performance Of A Class Of Modern Container Ships. Retrieved from https://www.witpress.com/elibrary/wit-transactions-on-the-built-environment/5/11251

Samarasinghe, S. (2007, January). Neural networks for applied sciences and engineering. From fundamentals to complex pattern recognition. Retrieved from https://www.researchgate.net/publication/266189755_Neural_networks_for_applied_sciences_and_engineering_From_fundamentals_to_complex_pattern_recognition

Sarle, W. (1994, April). Neural Networks and Statistical Models. Retrieved from https://people.orie.cornell.edu/davidr/or474/nn_sas.pdf

Shahin M. A., Jaksa M. B., Maier H. R. (2008), State of the art of artificial neural networks in geotechnical engineering, Electronic Journal of Geotechnical Engineering 8: 1-26.

Solonen, A. (2018, July 05). Just what is a digital twin. Retrieved July 28, 2018, from https://www.eniram.fi/just-digital-twin-can-help/

Soner, O., Akyuz, E., & Celik, M. (2018, June 21). Use of tree based methods in ship performance monitoring under operating conditions. Retrieved from https://www.sciencedirect.com/science/article/pii/S0029801818314446#bib17

Thomas, J., & Mahapatra, S. (2016, April 23). Improved simple optimization (SOPT) algorithm for unconstrained non-linear optimization problems. Retrieved July 31, 2018, from https://www.sciencedirect.com/science/article/pii/S2213020916300362

UNCTAD. (2017). Review of Maritime Transport 2017. 2017, unctad.org/en/pages/PublicationWebflyer.aspx?publicationid=1890. Accessed 27 July 2018.

University of Oxford. (2015). AdaBoost. Retrieved from http://www.robots.ox.ac.uk/~az/lectures/cv/adaboost_matas.pdf

Varyani, K. (2005, August 08). Squat effects on high speed craft in restricted waterways. Retrieved August 20, 2018, from https://www.sciencedirect.com/science/article/pii/S0029801805001514

Venkatasubramaniam, Ashwini, et al. Decision Tree and Random Forest Models for Outcome Prediction in Antibody Incompatible Kidney Transplantation. 9 Feb. 2017, www.sciencedirect.com/science/article/pii/S1746809417300204. Accessed 17 Aug. 2018.

Vigarié, André. (2016) .Les Tendances D'évolution Des Transports Maritimes (1955-1985).[Trends in Maritime Transport]. 23 Mar. 2016, www.persee.fr/doc/geo_0003-4010_1983_num_92_509_20168. Accessed 27 July 2018.

Wang, H., Li, X., Li, P., Veremey, E., & Sotnikova, M. (2018). Application of Real-Coded Genetic Algorithm in Ship Weather Routing. Journal of Navigation, 71(4), 989-1010. doi:10.1017/S0373463318000048

Wang, Shengzheng, et al. (2017). “Predicting Ship Fuel Consumption Based on LASSO Regression.” Egyptian Journal of Medical Human Genetics, Elsevier, 28 Oct. 2017, www.sciencedirect.com/science/article/pii/S1361920917302109.

Wartsila. (2009, February 03). Boosting Energy Efficiency. Retrieved from http://www.shippingtech.it/PDF/convegni 2010/2tecnologie1/Baan.pdf

Wilson R.A, Keil F.C. The MIT encyclopaedia of the cognitive sciences MIT Press (1999)

Wright, J., Colling, A., & Park, D. (1999). Waves, Tides, and Shallow-water

WMO. (2013). WMO statement on the status of the global climate in 2012. Retrieved August 20, 2018, from http://www.wmo.int/pages/prog/wcp/wcdmp/documents/WMO_1108.pdf

Yu, et al. (2016). k -Nearest Neighbor Model for Multiple-Time-Step Prediction of Short-Term Traffic Condition. 31 May 2016, trid.trb.org/view/1398211. Accessed 17 Aug. 2018.

Ziylan, K., & Nas, S. (2016, October). A STUDY ON THE EFFECTS OF TRIM OPTIMISATION ON SHIP RESISTANCE OF A SUB-PANAMAX TYPE CONTAINER VESSEL. Retrieved from https://www.researchgate.net/publication/311576476_A_STUDY_ON_THE_EFFECTS_OF_TRIM_OPTIMISATION_ON_SHIP_RESISTANCE_OF_A_SUB-PANAMAX_TYPE_CONTAINER_VESSEL