Use of machine learning algorithms in vessel performance monitoring

Adeline Crystal John Suresh Kumar

Master’s Thesis
USE OF MACHINE LEARNING ALGORITHMS IN VESSEL PERFORMANCE MONITORING

MACHINE LEARNING ALGORITHMS TO PREDICT HULL FOULING PATTERNS

SUBMITTED BY ADELINE CRYSTAL JOHN SURESH KUMAR
# Contents

1 **Introduction** 8
   - Motivation and Scope ................................................................. 9
   - Structure of the Report .............................................................. 9
   - Background Studies ................................................................. 10

2 **Theory - Resistance Calculations** 11
   - Still Water Resistance ............................................................... 12
   - Added Resistance in Waves ....................................................... 13
   - Wind Resistance ........................................................................... 14
   - Fouling Resistance ....................................................................... 14
   - Ballast Power Correction ............................................................ 15
   - Hull Fouling .................................................................................. 16

3 **Theory - Machine Learning Algorithms** 21
   - Outline ............................................................................................ 22
   - Models ............................................................................................ 23
   - Target Data Error and Model Fit .................................................... 31

4 **Methodology** 33
   - The Data .......................................................................................... 34
   - KVLCC2 - Simulated Data .............................................................. 36
   - Vessel 1 - Recorded Data .............................................................. 40
   - Machine Learning Algorithms ....................................................... 54
   - Feature Selection ............................................................................ 54

5 **Results** 59
   - KVLCC2 .......................................................................................... 59
   - Vessel 1 ........................................................................................... 62

6 **Inferences** 84
   - The Data .......................................................................................... 84
   - The Effectiveness of Hull Cleaning .................................................. 85
   - The Machine Learning Algorithms .................................................. 86
   - Linear Regression vs LSTM Models ................................................. 86
   - The Anti Fouling Paint Effectiveness Factor, CEFF ............................ 88
The Training -Testing Data split ................................................................. 88
The different input feature combinations ................................................. 89
Choosing a suitable Model ....................................................................... 89
Future Improvements ............................................................................. 90

7 Conclusion .............................................................................................. 92

8 Appendix .................................................................................................. 96
  Constant Values Used ............................................................................... 96
  List of Variables used from Simulated Data of KVLCC2 ......................... 96
  List of Variables used from Recorded Data of Vessel 1 ......................... 97
  Lookup Table for Wake Fraction for Vessel 1 ........................................ 98
  Lookup Table for Thrust Deduction Factor for Vessel 1 ......................... 99
  Open Water Efficiency Propeller Curve for Vessel 1 ............................ 100
  Inputs to generate RAO to calculate the Added Resistance in Waves\textsuperscript{15} 100
  Tabulated Values of MAE and $R^2$ for Vessel 1 - Linear Regression ........ 101
  Tabulated Values of MAE and $R^2$ for Vessel 1 - LSTM ......................... 102
Abstract

The use of machine learning algorithms to model vessel hull fouling is investigated. Using a simulated dataset for the concept hull KVLCC2 the fouling resistance is evaluated using expressions given by Malone, Little, & Allman\textsuperscript{1} in their paper; and the results are modelled using Linear Regression and Long-Short Term Memory Networks. Long Term recorded data from a vessel is also evaluated and the observed fouling resistance is calculated then modelled using Linear Regression and Long-Short Term Memory Networks. A possible modification to be considered while evaluating the hull roughness and fouling resistance when using the expressions from the International Towing Tank Conference\textsuperscript{2} and Holtrop-Mennon’s\textsuperscript{3} expressions is studied. The advantages and disadvantages of the different algorithms are considered. It is concluded that a time series model of hull fouling resistance can be constructed, with a reasonable level of certainty - and can be used to predict the hull fouling resistance over a given period as well. However the model’s nature and its predictions depend greatly on the nature of its training data, and needs further improvements till it can be considered widely applicable.
Preface

This thesis is submitted as a part of the Final Year’s Thesis for my two-year Nordic Master’s in Maritime Engineering Programme at Chalmer’s University of Technology, Gothenburg and the Technical University of Denmark, Lyngby. This project was carried out by Adeline Crystal John Suresh Kumar during the Spring Semester of 2020 at the Technical University of Denmark (DTU), Lyngby, in collaboration with an anonymous ship owner. This thesis was supervised by Prof. Poul Anderson from the Department of Mechanical Engineering at DTU with Assoc. Prof. Wengang Mao from Chalmer’s University of Technology as a secondary supervisor.

Adeline Crystal John Suresh Kumar (s192134)
Technical University of Denmark.
11th July 2020.
Acknowledgements

This Master’s Thesis is written for my double Master’s degree in Maritime Engineering as a part of the Nordic Master’s at Chalmers University of Technology, Gothenburg and the Technical University of Denmark, Lyngby (DTU). All the work carried out as a part of this thesis is a because of all the knowledge imparted to me by my teachers and professors through out my life - and particularly my Master’s degree. I would like to express my gratitude to Poul Anderson - my professor and guide for this project, Wengang Mao - my guide at Chalmer’s University, Ole Winther - professor at DTU Compute, and Bushan Taskar - Post Doctoral student at DTU Mechanical Engineering, for their help in various parts of this thesis. I would also like to thank all my friends who’ve kept in touch with me through this quarantine time, who’ve helped make this time easier.

Dedicated to my mom and my father. Thank you for supporting me through everything.

For my Grandmother.
1 Introduction

The Maritime Industry is constantly exploring options to curb vessel fuel consumption in order to reduce emissions and work towards achieving their sustainability goals. Additionally, as mandated by the International Maritime Organization, vessels are required to report their fuel oil consumption data as part of the organization’s strategy to reduce greenhouse gas emissions. (IMO, 2018)\textsuperscript{4} Increased resistance due to Hull fouling is expected as a part of normal vessel operations, and gives rise to increased fuel consumption. The condition of the vessel’s underwater hull has a significant effect on the resistance experienced. Hull cleaning contributes to a very small percentage of the overall operating cost of the vessel; however, the scheduling of this hull cleaning is important. Better scheduling of hull cleaning can result in improved reductions in fuel consumption - which is of both environmental and economic significance. Nevertheless, hull fouling is difficult to predict as it is a seemingly random process where there is no specific trend that reflects the exact timing of required hull cleaning or repainting. Additionally, the use of harmful aggressive anti-fouling paints containing TriBuTylin (TBT) are banned by the International Maritime Organization, therefore hull fouling is a factor that is expected to play a role in vessel performance till suitable paints are developed, and is a common denominator to all ships irrespective of their type or operating profile.

With the growth of the use of Big Data techniques in several fields, Machine Learning Algorithms have been used for various aspects of Maritime Operations and evaluating Vessel Performance - predominantly for vessel propulsive power predictions - and even real time power predictions. Machine learning algorithms can be used to study the previous records of vessel performance; and accordingly predict the hull fouling over a period of time. In this thesis, different methods of estimating the fouling resistance are used, and suitable machine learning algorithms are modelled to predict the fouling based on the recorded data. Models can be developed that can to an extent model the random hull fouling process and subsequently predict the hull fouling as well.

It will be observed that hull fouling though a random process, can be considered to be a function of the vessel’s operating profile, the hull’s paint properties and time. Using all these parameters - machine learning algorithms can be used to model the vessel hull fouling based on long term recorded data, to a reasonable level of accuracy. This can be considered the first step in deriving expressions for hull fouling over time. The randomness and uncertainty of different features have been considered in this paper wherever possible - and in other cases suitable assumptions have been made.

In this paper two datasets are considered - the first is a simulated dataset for the concept vessel KVLCC\textsubscript{2}, where an ideal case of fouling resistance estimation is considered and then modelled using the Machine Learning Algorithms. The second dataset is long-term data recorded and collected from a vessel operator that is to remain anonymous in this report. The dataset from this vessel referred to as Vessel 1 is collected over several years, and
is used to determine the actual fouling resistance the vessel encountered over its operation, and is then modelled using Machine Learning Algorithms. The practicality of the different models obtained are then tested by using the models to predict the expected fouling resistance over 180 days - and in three different case scenarios.

1. When the hull is not cleaned nor repainted
2. When the hull is cleaned but not repainted
3. When the hull is cleaned and repainted.

The key differences between the two datasets will be considered, and various other inferences made in the report aim to improve the overall estimation of hull fouling resistance in the future.

**Motivation and Scope**

The application Machine Learning in recent times has been thought of as a panacea to a diverse range of problems, which has driven several industries to focus on data collection and analysis. The author believes that the problem of estimating fouling resistance is something that can be modelled to a certain level of accuracy using Machine Learning Algorithms; which to date has not been considered. Additionally, with the increasing need to reduce needless fuel consumption to curb harmful emissions, the author believes that studying hull fouling can contribute to considerable environmental and economic advancements. Hull fouling is a part of all ships and their performance - irrespective of its operating regime and propulsive mechanism. A data model which can predict the possible future fouling resistance can be of great benefit to vessel operators - particularly in scheduling their hull cleaning and vessel dry docking, as it has a direct environmental and economic advantage. A data model that can predict the future time at which the cost of the additional fuel consumption due to hull fouling is greater than the cost of cleaning the hull, is the optimum case application of the various conclusions arrived at in this report. The conclusions from this report can be used in Digital Twins of Vessels as well, which can help improve their results. Extracting the resistance values from the data recorded accordingly, feeding them into a suitable machine learning algorithms and evaluating the obtained results is the premise of this thesis.

**Structure of the Report**

The various sections in this paper are as follows:

1. **Introduction**
   
   This is a brief section on the motivation and scope of this thesis paper.

2. **Theory - Resistance Calculations**
   
   All the necessary expressions to calculate the vessel resistance, and a brief section on Hull Fouling is discussed in this section.
3. **Theory - Machine Learning Algorithms**
   This section consists of an overview on Machine Learning and the different algorithms used in this paper.

4. **Methodology**
   The data analysis, implementation of the various expressions and important assumptions considered for using the different Machine Learning Algorithms are discussed for the two different vessel datasets.

5. **Results**
   This section includes the observed results from the different Machine Learning algorithms, using different data sets and input feature combinations. The results from the KVLCC2 only consider relatively simple models, and the results for Vessel 1 consider several possible models constructed using different combinations of input features and training data ranges, and the generated models are then used to create forecast future values.

6. **Inferences**
   A discussion on the results obtained and the various assumptions that can affect possible results, along with suggestions for further improvements are detailed in this section.

**Background Studies**
This paper predominantly uses a hull fouling model proposed by Malone, Little, & Allman in their paper *Effects of Hull Foulants and Cleaning/Coating on Ship Performance and Economics* which was published in the SNAME Transaction 88. The expressions given in this paper were used alongside expressions from the *Performance Prediction Method* given by the International Towing Tank Conference and expressions given by Holtrop & Mennen (Holtrop & Mennen, 1982) to evaluate the data records and calculate the various resistance components. Additionally the Basics of Ship Propulsion by MAN Energy Solutions (MAN Energy Solutions, 2018) was also used a reference. For the various Machine Learning Algorithms, several sources have been used to provide a extensive overview of the different procedures used. The description and examples given in *Hands-on Machine Learning with Scikit-Learn and Tensor Flow* by A.Géron were helpful in understanding the basics of Machine Learning problems, and the example codes from Machine Learning Mastery (Machine Learning Mastery, 2017b) were used. In order to research how Machine Learning Algorithms have been used in the Maritime Industry - the papers *Data-driven Vessel Performance Monitoring* by Pedersen, B. P. (Pedersen, 2014) and *Big Data Techniques for Ship Performance study* by Anagnostopoulos, A (Anagnostopoulos, 2017) were studied. The paper *Ship technical and economic performance as function of hull cleaning and coating practices* by M.D.Helland (Helland, 2018) was helpful in the research of hull fouling and cleaning practices.
Section: 2  Theory - Resistance Calculations

2 Theory - Resistance Calculations

In this section, the various expressions required to evaluate the resistance of the hull are detailed. The total resistance $R_T$ of the Hull at a given speed $U$ is given by

$$R_T = R_{\text{Still Water}} + R_{\text{Wind}} + R_{\text{Added Resistance in waves}} + R_{\text{Fouling}}$$  \hspace{1cm} (1)

The calculation of the individual resistances are elaborated in succeeding sections. The effective power $P_E$ required is then calculated as

$$P_E = R_T U$$  \hspace{1cm} (2)

The Brake Power at the Engine flywheel $P_B$ is estimated by

$$P_B = \frac{P_E}{\eta_{\text{Prop}}}$$  \hspace{1cm} (3)

where the total propulsive efficiency $\eta_{\text{Prop}}$ given by

$$\eta_{\text{Prop}} = \eta_H \eta_O \eta_R \eta_S$$  \hspace{1cm} (4)

$\eta_H$  Hull Efficiency

$\eta_O$  Propeller Open Water Efficiency

$\eta_R$  Propeller Relative Rotative Efficiency

$\eta_S$  Shaft Transmission Efficiency. The hull efficiency $\eta_H$ is given by

$$\eta_H = \frac{1 - t}{1 - w}$$  \hspace{1cm} (5)

$t$  Thrust deduction factor and

$w$  Wake Fraction.

The wake fraction is to be considered because the advance velocity of the water at the propeller is smaller than the vessel speed, and the thrust deduction factor is required as the thrust required to propel the vessel is greater than the total resistance $R_T$. The Propeller Open Water Efficiency $\eta_O$ is the efficiency of the propeller given by the Open Water Curves of the Propeller from the Wageningen B-Series, or from the specified propeller curves for the propeller if available. The relative rotative efficiency of the propeller $\eta_R$ is the efficiency of the propeller operating in the non-uniform wake field at the back of the hull. $\eta_S$ is the shaft-line transmission efficiency, which has a considerable effect when the engine is not directly coupled to the propeller.


Section: 2  Theory - Resistance Calculations

**Still Water Resistance**

The Still Water resistance of the vessel is calculated using the Holtrop-Mennen Method\(^3\) (Holtrop & Mennen, 1982), which determines the overall still water resistance as

\[
R_{Still \ Water} = (1 + k)R_F + R_{tr} + R_a + R_w + R_b + R_{AAS}
\]

\(k\)  Form factor, estimated by relations given in\(^3\)

\(R_F\)  Frictional Resistance of the hull.

\(R_{tr}\)  Resistance due to immersed transom area

\(R_a\)  Resistance due to Model-ship correlation

\(R_w\)  Resistance due to Wave making and wave breaking

\(R_b\)  Additional Resistance due to bulbous bow

\(R_{AAS}\)  Air Resistance of the Hull

Expressions given in\(^3\) are used to evaluate the value of \(k, R_{tr}, R_w\) and \(R_b\). The frictional resistance is calculated as

\[
R_F = \frac{1}{2}C_F \rho S U^2
\]

For the frictional resistance, the frictional resistance coefficient \(C_f\) is given by ITTC-78 (ITTC 1978 Performance Prediction Method, 1978)\(^2\) as

\[
C_F = \frac{0.075}{(\log_{10}Re - 2)^2}
\]

The air resistance of the hull is calculated from \(C_{AAS}\) given by ITTC-78

\[
C_{AAS} = C_{DA} \frac{\rho_A \cdot A_{VS}}{\rho \cdot S}
\]

\(C_{DA}\)  Air Resistance Coefficient, assumed as 0.8

\(\rho_A\)  Air Density, 1 kg/m\(^3\).

\(A_{VS}\)  Transverse Projected Area of the Hull above the waterline

The value of \(R_{AAS}\) is calculated using the same expression as (7) replacing \(C_F\) with \(C_{AAS}\). The Correlation Allowance as per ITTC-78\(^2\) is given by

\[
C_A = (5.68 - 0.61 \log Re) \times 10^{-3}
\]

The value of \(R_{Fouling}\) is calculated using the same expression as (7). However, the use of this \(C_A\) value (and the fouling resistance coefficient \(\Delta C_F\)) will be explained in succeeding sections.
Section: 2  Theory - Resistance Calculations

Added Resistance in Waves

The Added Resistance in Waves can be estimated by different methods such as the simplified Gerritsma and Beukelman’s method (Martin Alexandersson, 2009),\textsuperscript{13} or using panel methods like as TiMIT (H. B. Bingham, 2019).\textsuperscript{14} For this project, it is estimated by the DTU method (Calculate added resistance RAO using DTU method, 2020),\textsuperscript{15} where the Response Amplitude Operator is generated for different heading of the waves. It is to be noted that for waves with heading less than 90° and greater than 180°, the added resistance in waves is 0. The non-dimensional added resistance $\frac{R_w L_{pp}}{\rho g A^2 B^2}$ is given as a function of the non-dimensional wavelength $\frac{\lambda}{L_{pp}}$. The mean added resistance is then calculated as

$$\bar{R}_w = 2 \int_0^\infty \frac{R_w(\omega)}{A^2} S(\omega) d\omega$$

where

$S(\omega)$ is the wave spectrum

$A$ Wave Amplitude

![Sample Added Resistance in Waves RAO](image)

Figure 1: Sample Added Resistance in Waves RAO\textsuperscript{15}
In this project, the Pierson-Moskovitz spectrum is used. The Pierson-Moskovitz\textsuperscript{16} spectrum is generated as:

\[ S(\omega) = \frac{\alpha g^2}{\omega^5} \exp\left(-\beta \left(\frac{\omega_0}{\omega}\right)^4\right) \]  

\[ \alpha = 8.1 \times 10^{-3}; \quad \beta = 0.74; \quad \omega_0 = \frac{g}{U_{19.5}} ; \quad U_{19.5} \approx 1.026 U_{10} \]  

where \( U_{10} \) is the wind speed recorded at 10m. In order to generate the spectrum, the range of \( \omega \) values is obtained from the values for \( \bar{\lambda} \) as

\[ \omega = \sqrt{\frac{2\pi g}{\bar{\lambda} L_{pp}}} \]  

The spectrum is then calculated from the obtained frequency values, and then multiplied with the respective dimensionalized Response Amplitude Operator and the obtained values are integrated over the frequency range to obtain the mean added resistance in waves.

**Wind Resistance**

The Wind resistance is estimated using expressions given in the Manoeuvring Technical Manual (Brix, J, 1987),\textsuperscript{11} and the resistance due to the Wind is given by

\[ R_{wind} = \frac{1}{2} C_X \rho_{air} A_{XV} U_{RWind}^2 \]  

\( \rho_{air} \) Density of Air, 1\( kg/m^3 \)  
\( C_X \) Coefficient for Wind Resistance based on Relative Wind Direction  
\( A_{XV} \) Transverse Projected Area of the Hull Above the Waterline  
\( U_{RWind} \) Relative Wind Speed

The value of \( C_X \) can be taken from the Manoeuvring Technical Manual (Brix, J, 1987)\textsuperscript{11} for different type of vessels, or can be evaluated using models given for the specific vessel, if available.

**Fouling Resistance**

The co-efficient for fouling resistance given by the ITTC-78\textsuperscript{2} method is calculated as

\[ \Delta C_F = 0.044 \left( \left( \frac{k_s}{L_{WL}} \right)^{\frac{1}{3}} - 10 \cdot Re^{\frac{1}{3}} \right) + 0.000125 \]  

\( \Delta C_F \) Fouling Resistance Coefficient
\( k_s \) Fouling Coefficient
\( L_{WL} \) Length Waterline
\( Re \) Reynolds Number
Where \( k_s \) is in metres. The minimum value of \( k_s \) to be considered stated to be 150\( \mu m \) The corresponding fouling resistance is given by

\[
R_{\text{Fouling}} = \frac{1}{2} \Delta C_f \rho S U^2
\]  

Further information regarding the use of these expressions is given in succeeding sections.

**Ballast Power Correction**

The effective power is calculated by Eqn: (2), and then to correct for the ballast condition, the following expression from Moor-Molland. (Molland, 2008)\textsuperscript{25} is utilized.

\[
\frac{P_{E_{\text{Ballast}}}}{P_{E_{\text{Loaded}}}} = 1 + [(T_R - 1)\{(0.789 - 0.270(T_R - 1) + 0.529C_B(L/T)^{0.5})
\]

\[
+ \frac{V}{\sqrt{gL}}(2.336 + 1.439(T_R - 1) - 4.605C_B(L/T)^{0.5})
\]

\[
+(\frac{V}{\sqrt{gL}})^2(-2.056 - 1.485(T_R - 1) + 3.798C_B(L/T)^{0.5})\}
\]

where \((T_R)\) is \(T_{\text{Ballast}}/T_{\text{Load}}\). The wake fraction correction is

\[
(1 - w_T)_R = 1 + [(T_R - 1)(0.2882 + 0.1054\theta)]
\]  

(18)

The thrust correction is calculated as

\[
(1 - t)_R = 1 + [(T_R - 1)(0.4322 - 0.4880C_B)]
\]  

(19)

where

\[
(1 - w_T)_R = \frac{(1 - w_T)_{\text{Ballast}}}{(1 - w_T)_{\text{Loaded}}}; \quad (1 - t)_R = \frac{(1 - t)_{\text{Ballast}}}{(1 - t)_{\text{Loaded}}}; \quad (T_R) = \frac{T_{\text{Ballast}}}{T_{\text{Load}}}
\]  

(20)

For simplification, based on (2), \( \frac{P_{E_{\text{Ballast}}}}{P_{E_{\text{Loaded}}}} \) is reduced to \( \frac{R_{T_{\text{Ballast}}}}{R_{T_{\text{Loaded}}}} \). The variation of \( R_{\text{Added Resistance in wave}} \) and \( R_{\text{Fouling}} \) with loading condition is neglected, and in the case of the \( R_{\text{Wind}} \) the lateral and transverse area variation with draft is considered in the resistance calculation. Therefore the correction is applied only to the Still Water Resistance \( R_{\text{Still Water}} \); and the wake fraction \( w \) and the thrust deduction factor as well.

It is known that vessels in ballast condition generally operate with a trim, and \( \theta \) in Eq 18 represents the trim angle, however when the correction factor was applied with the trim angle, several outliers of data were found, which is why for the purpose of this project the trim angle in ballast condition is set as 0.
**Hull Fouling**

Hull fouling is defined as the accumulation of marine growth on the underwater hull. Hull fouling is a complex process - that depends on the characteristics of the operating waters of the vessel, the salinity of the seawater, temperature, currents, the condition of the vessel hull and other factors. Hull fouling is predominantly a random process, and it is difficult to investigate its mechanisms. Biological growth occurs on the surface of the hull, and is typically thought to be of two stages - Microfouling and Macrofouling. (Little & Depalma, 1988).\(^{19}\)

1. **Microfouling**
   Also known as slime, it is the beginning stage of hull fouling which occurs in 4 stages according to Little & Depalma , - Conditioning, Colonization by bacterial species, colonization by other microorganisms and accumulation . It has been observed that sometimes the anti-fouling paints inhibit macrofouling such as weed and shell but not slime. It is easy to clean off either by chemical treatment or mechanical cleaning - and it has been shown to contribute to increase in resistance. (Townsin, 2003)\(^{18}\)

2. **Macrofouling**
   Macrofouling is caused by macro-organisms such as Barnacles, mussels, shells, sponges and algae. It was found that shells with height 14mm covering only 5% of the wetted surface area contributed to a significant increase in drag to about 66%. In addition to increasing the hull resistance, macrofouling increases the weight of the submerged area of the hull. Macrofouling can cleaned by chemical and mechanical methods. (Townsin, 2003)\(^{18}\)

**Surface Roughness\( k_s \)**

An important factor that affects the fouling resistance is the hull surface roughness. As per ITTC-78\(^2\) while predicting the still water resistance, the surface roughness is considered to be \( 150 \times 10^{-6} m \). However while calculating the fouling resistance due to hull growth, the value calculated from Eqn:(21) given by Malone’s expressions, is converted to metres and then used. This is a crucial feature that decides the fouling resistance, and therefore a basic understanding of this parameter is essential.

Surface roughness is typically a metrology parameter, which defines the deviations of a surface in the direction normal to the surface. Surface Roughness normally refers to Average Surface Roughness \( R_a \), which is the arithmetic mean of the absolute value of the profile deviations \( y_i \) from the mean line of roughness profile over a sample length. The roughness profile is obtained by isolating the surface waviness from the overall surface profile. It is calculated as \( R_a = \frac{1}{n} \sum |y_i| \). Therefore it can be seen that surface roughness is a parameter can not take a negative value. This is a crucial deduction that is relevant to the various conclusions arrived at in this paper.
Hull Painting and Repainting

Hull paints applied to the underwater hull typically consist of different layers of paints that inhibit hull steel corrosion as well as stop marine growth on the underwater hull. Typically, hull paints have a life of 5-7 years, and hulls are usually repainted when the vessel is dry docked for underwater hull surveys and other machinery maintenance. According to Malone, as the paints lose effectiveness, hull growth begins, thereby increasing the surface roughness of the hull till the hull is repainted or cleaned. In accordance with Malone’s expressions, two paints are considered to cover the vessel underwater hull: Anti-corrosion paints and the outermost Anti Fouling paint; each of different thickness and number of layers. Over the operation of the vessel, it is required to be dry docked as per class requirements or machinery repair/maintenance. Vessel dry docking is typically a significant decision, made in order that several maintenance activities and class surveys can be carried out. Every time the vessel is docked, the underwater hull is sandblasted before repainting, however the roughness of the sandblasted hull will not be as low as the roughness of the new vessel.

Underwater Hull Cleaning and Propeller Polish

Hull cleaning is done between dry docking to remove the marine growth on the underwater hull. Underwater hull cleaning typically uses divers to clean the hull using robotic cleaners to remove the hull fouling. Hull cleaning does show improvements in the vessel performance, as it will be observed in the data recorded from Vessel 1. Hull cleaning can either be carried out only on the sides of the underwater hull or both the sides and the bottom of the hull - based on the cost and operator requirements. It is to be noted that hull fouling is generally greater on the sides of the vessel than on the bottom, due to penetrating sunlight - and Malone’s expression take this factor into consideration. Removing the marine bio-growth also removes the additional weight due to the growth while reducing the hull roughness. (Global Maritime Energy Efficiency Partnerships, n.d.) Hull cleaning can be scheduled at ports where divers are available with suitable machinery, and the decision to clean the hull is normally taken when the vessel’s fuel consumption increases. The improvement in hull roughness can not be measured directly after underwater hull cleaning, and it can be predominantly perceived by subsequent better fuel consumption by the vessel. It is understood that every time the vessel is docked, the underwater hull and propeller

Hull Painting and Repainting

Hull paints applied to the underwater hull typically consist of different layers of paints that inhibit hull steel corrosion as well as stop marine growth on the underwater hull. Typically, hull paints have a life of 5-7 years, and hulls are usually repainted when the vessel is dry docked for underwater hull surveys and other machinery maintenance. According to Malone, as the paints lose effectiveness, hull growth begins, thereby increasing the surface roughness of the hull till the hull is repainted or cleaned. In accordance with Malone’s expressions, two paints are considered to cover the vessel underwater hull: Anti-corrosion paints and the outermost Anti Fouling paint; each of different thickness and number of layers. Over the operation of the vessel, it is required to be dry docked as per class requirements or machinery repair/maintenance. Vessel dry docking is typically a significant decision, made in order that several maintenance activities and class surveys can be carried out. Every time the vessel is docked, the underwater hull is sandblasted before repainting, however the roughness of the sandblasted hull will not be as low as the roughness of the new vessel.

Underwater Hull Cleaning and Propeller Polish

Hull cleaning is done between dry docking to remove the marine growth on the underwater hull. Underwater hull cleaning typically uses divers to clean the hull using robotic cleaners to remove the hull fouling. Hull cleaning does show improvements in the vessel performance, as it will be observed in the data recorded from Vessel 1. Hull cleaning can either be carried out only on the sides of the underwater hull or both the sides and the bottom of the hull - based on the cost and operator requirements. It is to be noted that hull fouling is generally greater on the sides of the vessel than on the bottom, due to penetrating sunlight - and Malone’s expression take this factor into consideration. Removing the marine bio-growth also removes the additional weight due to the growth while reducing the hull roughness. (Global Maritime Energy Efficiency Partnerships, n.d.) Hull cleaning can be scheduled at ports where divers are available with suitable machinery, and the decision to clean the hull is normally taken when the vessel’s fuel consumption increases. The improvement in hull roughness can not be measured directly after underwater hull cleaning, and it can be predominantly perceived by subsequent better fuel consumption by the vessel. It is understood that every time the vessel is docked, the underwater hull and propeller
are cleaned as well.

Figure 3: Underwater Hull Cleaning

In addition to hull cleaning, propeller polishing is also often carried out in conjunction with hull cleaning - as it also significantly contributes to improvements in vessel resistance. Vessel performance improvements due to propeller cleaning can be attributed to

1. Removal of Marine bio-growth thereby improving the propeller surface roughness
2. Removing mineral deposit on the propeller surface
3. Reducing propeller cavitation
4. Reducing surface roughness due to metal corrosion. (Bevaldia, 2020)
Malone Method

[All values in this section and relevant references to this section are in the Foot-Pound-Second unit system.]

In their paper Malone, Little, & Allman, the deterioration of Hull Roughness given by the Mean Apparent Amplitude $MAA$, is calculated as a function of the failure of the Hull paint and the accumulated fouling of the hull due to operations at slow speeds. As per this reference - there is no hull growth when the vessel speed is greater than 3 kn. The total $MAA$ in mils is given by

$$MAA_{Total} = MAA_{Steel\ Plate} + MAA_{Coating\ System} + MAA_{Corrosion} + MAA_{Fouling}$$  \hspace{1cm} (21)

The value of the $MAA_{Steel\ Plate}$ is due to the roughness of the bare hull, and is normally assumed to be a constant value. The roughness due to the paint, $MAA_{Coating\ System}$ is calculated as

$$MAA_{Coating\ System} = (BLR_{AC})(NC_{AC}) + (BLR_{AF})(NC_{AF})$$  \hspace{1cm} (22)

$BLR_{AC}$ Base Line Roughness, Anti corrosion coating [mils]
$NC_{AC}$ Number of coats, Anti-corrosion coating
$BLR_{AF}$ Base Line Roughness, Anti Fouling Painting [mils]
$NC_{AC}$ Number of coats, Anti-fouling paint.

The $MAA_{Corrosion}$ is given as a function of the percentage failure of the effectiveness of the hull paint, $PCF$

$$PCF = 1.8203 \times 10^{-3}(X)^{3.332}$$  \hspace{1cm} (23)

$X$ Vessel Paint Age in Days

The final roughness factor, $MAA_{Fouling}$ is a function of the Hull Roughness Factor, $HRF$, the cumulative number of days the vessel has spent at port $PT$ and the Anti-fouling Coating Effectiveness Factor $CEFF$.

$$MAA_{Fouling\ Sides} = (HRF)(PT)(CEFF)$$  \hspace{1cm} (24)

$$MAA_{Fouling\ Bottom} = 0.75(HRF)(PT)(CEFF)$$  \hspace{1cm} (25)

$$CEFF = 1 - [2.72/e^{\frac{X}{Z}} - 0.240(Z - 1.0)^{0.263}]$$  \hspace{1cm} (26)

$Z$ Ratio of Vessel Paint Age to Vessel Paint Effective life
Section: 2  Theory - Resistance Calculations

It is considered that the fouling on the sides is greater than that on the bottom due to light available for marine growth on the sides. The values of the \( HRF \) is based on the location of the vessel - the salinity of the water, the corrosiveness of the seawater at the port it is situated at and other factors; and the values of \( HRF \) are given on qualitative scale as follows:

<table>
<thead>
<tr>
<th>Qualitative Fouling Scale</th>
<th>Fouling Severity</th>
<th>HRF Value / Port Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>Clean</td>
<td>0</td>
</tr>
<tr>
<td>2.0</td>
<td>Trace</td>
<td>( 2.1 \times 10^{-3} )</td>
</tr>
<tr>
<td>4.0</td>
<td>Trace to Light</td>
<td>( 3.09 \times 10^{-4} )</td>
</tr>
<tr>
<td>6.0</td>
<td>Light</td>
<td>( 1.507 \times 10^{-3} )</td>
</tr>
<tr>
<td>8.0</td>
<td>Light to Moderate</td>
<td>( 4.64 \times 10^{-5} )</td>
</tr>
<tr>
<td>10</td>
<td>Moderate</td>
<td>( 1.111 \times 10^{-2} )</td>
</tr>
<tr>
<td>12</td>
<td>Moderate to Severe</td>
<td>( 2.266 \times 10^{-2} )</td>
</tr>
<tr>
<td>14</td>
<td>Severe</td>
<td>( 4.140 \times 10^{-2} )</td>
</tr>
</tbody>
</table>

Table 1: \textit{Hull Roughness Factor values given by Malone}

Additionally, the seasons were considered to have an effect on the hull growth - and hull growth is considered to be worst in summer and the least in the winter, factors described in below table.

<table>
<thead>
<tr>
<th>HRF Correction Factor for Season</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter</td>
<td>0.1</td>
</tr>
<tr>
<td>Spring</td>
<td>0.55</td>
</tr>
<tr>
<td>Summer</td>
<td>1</td>
</tr>
<tr>
<td>Fall</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Table 2: \textit{Hull Roughness Factor Corrections for the different Season}

Although the paper has received criticism for its over simplistic model of hull fouling - particularly given that surface roughness is not usually a sum of individual layer roughness but mainly the outer-most layer roughness, the expressions do give results which do have relevance. In the calculations used in this paper - the hull roughness as per Malone is calculated as given, and additionally the roughness due to Hull fouling \( MAA_{Fouling} \) is considered to accumulate over time till a Hull cleaning is carried out. It will be seen that the calculated total hull roughness \( MAA_{Total} \) and the vessel anti-fouling paint effectiveness factor \( CEFF \) are important features that can be used to help model the hull fouling.
3 Theory - Machine Learning Algorithms

This section compiles information and algorithms from various sources, in order to provide a comprehensive overview of the various procedures used in this thesis. Machine Learning mainly refers to "programming computers to learn from data." (Géron, 2017).\(^6\) Machine Learning Algorithms have wide applications - from email filters, image classification, speech recognition, text prediction - to cancer detection, housing rent prediction and stock price prediction. Machine Learning problems are typically categorized to Supervised Learning or Unsupervised Learning.

1. **Supervised Learning**
   
   Supervised Learning is where the programme is explicitly coded to give an output. These are of two types.

   (a) **Classification Algorithms**
   
   This is used in programmes where the output predicted is one of ’n’ categories, or a discrete variable. For example classification of Spam-Not Spam emails, or Spam-Not Spam- Important-Advertisement-Updates emails. Examples of this includes Logistic Regression and Random Forests.

   (b) **Regression Algorithms**
   
   These are used when the predicted output is a continuous variable, such as housing rent or stock prices. Linear Regression and Artificial Neural Networks are some algorithms that can be used for Regression problems.

2. **Unsupervised Learning**

   These are used where the programme is made to recognize patterns without any particular target output. This includes Clustering and Anomaly detection.
Outline

Inputs variables are called features and the desired output feature is termed as the target variable. In this thesis a form of a Supervised Regression Algorithm called Long-Short Term Memory Networks is mainly used, alongside Linear Regression. Python is a powerful open source language with various available modules that can be used for Machine Learning. In this thesis, the Keras module is used along side the Sci-kit module.

Notations used are as follows:

\( x \)  Data feature or attribute
\( m \)  Number of features
\( y \)  Target Value or ‘True Value’
\( \hat{y} \)  Predicted Output of the Machine Learning Algorithm.
\( n \)  Number of Data samples

A basic schematic of solving a Machine Learning Problem is as follows:

![Figure 4: Schematic of a Typical Machine Learning Problem](image)

A typical machine learning problem takes a dataset with several features and records and splits it into training and test data. The training data is used to build or ‘train’ the model, and the test data is used to check how well the model’s outputs are in-line with the expected result and based on this, the model can be used to forecast future values.
Models

Linear Regression

The simplest form of Machine Learning, where the target $y$ is approximated to be a linear function of the features, which gives a predicted output $\hat{y}$ given as

$$\hat{y}(w, x) = w_0 + w_1 x_1 + \ldots w_m x_m$$

(27)

$x_i$ $i^{th}$ Feature

$w_i$ Coefficient or 'Weight' of $i^{th}$ feature

$m$ Number of features

The Ordinary Least Squares (OLS) method is a non-iterative method used to evaluate the value of the various coefficients, or weights. In matrix form, it is given by

$$\hat{y} = w^T x$$

(28)

$w = [w_0, w_1, w_2 \ldots, w_m]$ and $x = [x_0, x_1, x_2 \ldots, x_m]$

The coefficients are determined by minimizing the squared error term, given as

$$\varepsilon = (y - \hat{y}); \quad \text{minimize} \quad \Sigma \varepsilon^2$$

(29)

For a model with one feature, two coefficient values, $w_0$ and $w_1$ evaluated by the OLS method are

$$w_1 = \frac{\text{Covariance}}{\text{Variance}} = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (x-\bar{x})(x-\bar{x})}; \quad w_0 = \bar{y} - w_1 \bar{x}$$

(30)

For multivariate problems with more than one feature, the weights matrix $w$ is given by

$$w = (X^T X)^{-1} X^T Y$$

(31)
Artificial Neural Networks (ANN)

Neural Networks are powerful machine learning algorithms, that were first designed to work in the same fashion as neurons in the biological brain. The neurons in the brain work by receiving electrical impulse signals and on analysing these signals - sends out its own signal to another neuron. Several consecutive layers of neurons form a cohesive system where each neuron receives and sends out signals in a particular direction and finally achieves a desired output. Artificial Neural Networks are built to emulate this mode of functioning.

Figure 5: Schematic of a Basic Neural Network

Artificial Neural Networks have an input layer, one or more hidden layers and a final output layer. Each layer consists of several 'neurons' termed as units, \{x_i| x_1, x_2, ..., x_m\}. Each unit in the input layer is first initialized based on the input features. Each unit in the hidden layer is given an 'activity' value based on weights (a matrix) and the input features, then 'activated' by applying a non linear activation function. Bias units (constants) are also added in the hidden layers. Several such hidden layers can be stacked and finally the output is read from the output layer. The optimizing of the various weights and bias is done over several iterations by evaluating the loss at various iterations. The loss value to validate the output against can be set to the Mean Squared Error, Mean Absolute Error or other available loss functions. The optimization of the weights is done by various available optimizers.
Activation Functions

The role of the activation is to give an non-linear output \( f(x) \) between a desired range, based on the input \( x \). It is referred to as \( a \) or \( h \). They are sometimes referred to as squishing functions. Several activation functions exist, in this thesis the following are predominantly used.

**tanh Activation Function**

The tanh activation function gives an output between (-1,1) for all input values. An important feature to note is that the tanh activation for an input of 0 is 0.

\[
f(x) = \frac{\cosh x}{\sinh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}
\]  

**Sigmoid Activation Function**

The sigmoid \( \sigma \) activation function, sometimes called the logistic regression function, this function gives an output between (0,1) for all input values. The sigmoid activation for an input of 0 is 0.5

\[
f(x) = \frac{1}{1 + e^{-x}}
\]

**ReLU Activation Function**

The Rectified Linear Unit activation ReLU function, gives the maximum value from a set of inputs greater than 0 as the output, else it gives the output as 0. This activation is typically used in the final layer of a model to prevent negative values from being predicted (for ex. in weather predictions to prevent predicting negative rainfall.)

\[
f(x) = \max(0, x)
\]

![Diagram of tanh, Sigmoid, and ReLU Activation Functions](image)
Loss Function

Sometimes termed cost function, the loss function is used to determine the error between the target variable and the model output. In this thesis, the Mean Absolute Error (MAE) is used, which is defined as

\[
MAE = \frac{|y - \hat{y}|}{n}
\]  

(35)

This particular loss function is chosen as it works better with Gaussian distributed data with large outliers. (Machine Learning Mastery, 2019)\(^2\)

Optimizers

Optimizers aim to minimize the loss value by finding the global minimum of the loss and then the weights are updated based on the value of the gradient of the loss function, by taking a step in the direction opposite to the loss function gradient. For example, in the Gradient Descent Optimizer, the weights of the network are updated as

\[
w = w + \Delta w; \quad \Delta w_j = -\alpha \frac{\partial J}{\partial w_j}
\]

(36)

where \(J\) is the loss function and \(\alpha\) is the learning rate of the algorithm. Setting a too large value of \(\alpha\) can result in the algorithm oscillating between two points, and a very low value of \(\alpha\) can result in a large computation time.

Adam

Adam refers to Adaptive Moment Estimation, and combines the advantages of two existing algorithms - AdaGrad and RMSProp. It differs from the Gradient Descent optimizer where instead of a constant learning rate throughout the iterations, the learning rate is adapted every iteration based on the magnitude of the gradients. Adam sets the learning rates for the different parameters by using both the mean and the variance of the features, making it an effective and fast learning algorithm. In the case of Adam, the network weights are updated as

\[
w_{t+1} = w_t - \frac{\alpha_t}{\sqrt{v_t} + \varepsilon} \hat{m}_t; \quad \alpha_t = \frac{\alpha \cdot \sqrt{(1 - \beta_2^t)}}{(1 - \beta_1^t)}
\]

(37)

\(m_t\) Exponential Moving Average of the first moment (mean) of the gradient or the \(G\) \n\(v_t\) Exponential Moving Average of the second moment(variance) of the gradient at iteration \(t\) \n\(\beta_1, \beta_2\) Exponential Decay Rates for the mean and variance values, set to 0.9 and 0.999 by default. \n\(\varepsilon\) A hyper-parameter which is very small to prevent the algorithm from dividing by 0.

An exponential moving average is "a type of moving average that places a greater weight and significance on
the recent data points." (Adam Hayes, 2020)\textsuperscript{31} As per Kingma & Ba, 2015\textsuperscript{29}, "The moving averages themselves are estimates of the 1st moment (the mean) and the 2nd raw moment (the uncentered variance) of the gradient. " This allows for the model to converge to a minimum quickly as the change in weights are can be interpreted to be proportional to the gradient value - i.e. larger changes in weights when the gradient value is higher and smaller changes in weights when the gradient is lower.

**Points to be noted:**

A Neural Network can be used to model a non-linear model in this fashion. However a few points to be kept in mind while using Neural Networks are

1. All features should be scaled to the same range values, this helps in quicker convergence. Having a feature in 1000 and another in 0.1 ranges will require a lot of iterations to reach the global minimum.

2. The selection of the number of layers, number of neurons in each layer, learning rate (i.e. hyper-parameters) is an iterative process.

3. The loss functions generally have more than one local minimum, and as the network is randomly initialized - the minimum value reached each time can differ. This can lead to different errors values for same configuration of the Network.

4. A Neural Network can not work on time-based inputs, as for a trained neural network - the same inputs will always generate the same output. In order to overcome this drawback - Recurrent Neural Networks are implemented.
Long-Short Term Memory Networks

Recurrent Neural Networks:
In order to overcome the drawback of Artificial Neural Networks, Recurrent Neural Network are used, which are networks that allow previous information in the individual units to persist. Recurrent Neural Networks are used widely used - from text prediction software to stock pricing predictors. A Recurrent Neural Network cell, in addition to giving weights and activations to the next layer, also retains information from previous inputs in the following fashion:

![Recurrent Network Module Schematic](image)

The information that get stored is selected by using activation functions on the inputs. A drawback of Recurrent Neural Networks is that mostly the recent information was stored in the model, and data from earlier time steps was lost. This can be overcome by using a specific type of the Recurrent Neural Network - the Long-Short Term Memory Network or LSTM.

Long-Short Term Memory Networks:

A Long-Short Term Memory Network is a type of Recurrent Network where current activation can have 'long-term dependencies,' influenced by earlier inputs and not just recent ones.
The information that gets retained in an LSTM network cell is determined by different activation functions and expressions, as given below.

1. The cell state, which transfers the unchanged cell-state data from the previous input.

2. The **Forget gate** which applies the Sigmoid activation function on the input which itself a combination of the current input and the activation from the previous state.
Section: 3 Theory - Machine Learning Algorithms

3. The Input gate - which consists of two activation functions applied separately to the inputs. The Sigmoid function $\sigma$ decides which features to update, and the tanh function decides the new values for the next state.

4. The outputs from the two functions of the input gate are combined to give current state of the cell, which is obtained by combining the output from the Forget Gate and the Input Gate.

5. To predict the activation of the next layer in the network, a combination of the sigmoid and a tanh of the input is used.

6. The current output is fed to the next input of the cell, along with the current state of the cell. (Github, 2015)\textsuperscript{33}

In this fashion, the LSTM networks retains information from previous states to the next one. However being a Neural Network - it is still prone of the points mentioned earlier such as feature scaling and hyper-parameter selection.

**Implementation**

Training the LSTM model includes splitting the dataset into smaller batches of a given batch size, and training it over a set number of iterations termed epochs. When using LSTM models from Keras, the data to be given to the model must be re-shaped in the following format: \((\text{Batch size, Time Step, Number of Features})\); and the output of the format \((\text{Batch size, No. of Neurons})\). (MC.AI, 2019)\textsuperscript{32} The output from \(n\) neurons is then combined to a single neuron with the help of a basic Neural-Network layer with one cell. Batch sizes can be defined manually or by default it is set to 32. For bigger datasets it is best to set bigger batch sizes to improve the error over iterations. In this thesis, the time step is set to 1 - with every record taken as a single time step. It is important to ensure that the model does not shuffle the training data, or go backwards. (Keras)\textsuperscript{34} In this thesis the validation dataset is set as the input dataset itself - in order to make sure the model learns the dataset as a whole.

**Scaling Input Features**

The input features given to the LSTM network need to be scaled between 0 and 1, and this is done using Scikit learn’s MinMaxScaler function, which scales the individual features based on its range as follows:

$$X_{SC} = \frac{X - X_{Min}}{X_{Max} - X_{Min}}$$ \hspace{1cm} (38)

**Label Encoders**

In order to feed non-numerical features to the Machine Learning Algorithm, in this case inputs such as "Event Type" which takes the value of either "No Event", "Hull Cleaned", "Hull Repainted" - each of these need to be encoded as a numerical value. This is done using One Hot Encoding - which defines input features as a...
combination of binary values which together read the input feature. Label encoding can also be done as cardinal encoding, however it is not preferred as sometimes higher assigned values - which can occur when encoding several features can cause the model to not behave as expected. For example, for this project the "Event Type" is encoded as

<table>
<thead>
<tr>
<th>Event</th>
<th>Encoded Event Variable 1</th>
<th>Encoded Event Variable 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Event</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hull Cleaning</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Propeller Cleaning</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dry Docking</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3: *Label Encoding for Variable Event Type*

**Target Data Error and Model Fit**

In order to verify the the overall accuracy of the algorithms, the **Mean Absolute Error** and the $R^2$ score is used. These values give an impression of how well the model fits the input data and predicts values close to the observed values, however it alone can not be used to judge how good or sensible the model.

**Mean Absolute Error, MAE**

The Mean Absolute Error is evaluated as the absolute difference between the predicted value and the actual target variable value, as

$$ MAE = \frac{|y - \hat{y}|}{n} \quad (39) $$

**The Coefficient of Determination, $R^2$ Score**

The coefficient of Determination or $R^2$ score, is a regression coefficient, which determines the correlation between the target data and the predicted output variables. The value lies between 0 and 1, 1 implying that the correlation between the two is maximum or the variance between the target data and the predicted output variables is minimal. It is calculated as

$$ R^2 = 1 - \frac{\sum_i(y_i - \hat{y}_i)^2}{\sum_i(y_i - \bar{y})^2} \quad (40) $$
Points to be considered

In this project, a few points that differ from the typical Machine Learning problem are the following:

1. An important aspect to note is as this is a time-series prediction, the input records cannot be shuffled, as opposed to traditional Machine Learning problems. This makes the testing and validation of the algorithms sensitive to input batches and the sequence of the data records.

2. Ideally Machine Learning problems use separate Training and Validation sets, to ensure that the model is not biased towards either data set, but in this project - the validation and train data sets are the same, in order to observe how well the model considers the total dataset. This is to ensure the model is attuned to the input data given.

3. The values "Mean Absolute Error" and "Fit" are typically taken for the test data, however in this problem these values are taken based on the entire dataset mentioned and not just a test data set alone, as the model is expected to mirror the behaviour of the whole dataset and not just the training/test dataset.

4. The accuracy of the fit and the relevance of the predicted targets will be studied based on the sequence of inputs given to the model, as well as the features selected, however it will demonstrated that a better fit and accuracy does not necessarily imply the model is good at logical predictions. Adding features that reduce the accuracy but improve predictability will be discussed.

5. Over-fitting of the dataset is to be avoided, as a model that is over-fit will give poor predictions.
4 Methodology

As the data obtained are in terms of parameters such as vessel speed, draft, wind speed and direction - first the resistance must be calculated before it can be given to the Machine Learning Algorithm. For Data Analysis and manipulation, Python’s Pandas library is used, alongside necessary mathematical computational libraries. Modular Scripts to calculate Still Water Resistance, Wind Resistance and the Fouling Resistance by Malone’s equations were programmed, and for the Added Resistance in Waves, script was written to retrieve the RAO from the Ship Simulation Workbench.\textsuperscript{15}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{data_analysis_flow_chart.png}
\caption{Data Analysis Flow Chart}
\end{figure}
1. **Data Pre-Processing**
   This stage deals with checking the received data and filtering in incomplete or erroneous data, deleting unnecessary data attributes, combining relevant data files to make a comprehensive input file and making the input file complete in terms of time series data records.

2. **Data Processing**
   This stage predominantly included calculating the resistance and power values, and thereby obtaining the required input features from the data. Based on the calculated values - erroneous values and outliers are identified and removed. Feature Selection is also an important step, and it will be seen that selecting relevant features helps in improving the Algorithm’s accuracy.

3. **Machine Learning Algorithm**
   The Machine Learning Algorithm is then trained by feeding it the scaled input features and run for different hyper-parameters and configurations to obtain a final trained model, with a relevant accuracy scores.

4. **Results and Evaluation**
   The network’s output are then studied and future predictions are obtained using simulated data.

Between the Data Processing and the Machine Learning Stages - there were few iterations till the results were suitable. After observing discrepancies in the values obtained, the process had to be re-evaluated and then the programs had to be modified accordingly in order to arrive at the results.

**The Data**

**Vessels Parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Notation</th>
<th>KVLCC2</th>
<th>Vessel 1</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length between Perpendiculars</td>
<td>$L_{pp}$</td>
<td>320</td>
<td>234.07</td>
<td>m</td>
</tr>
<tr>
<td>Moulded Breadth</td>
<td>$B$</td>
<td>58</td>
<td>42.03</td>
<td>m</td>
</tr>
<tr>
<td>Design Draft</td>
<td>$T$</td>
<td>20.8</td>
<td>12</td>
<td>m</td>
</tr>
<tr>
<td>Design Speed</td>
<td>$U$</td>
<td>7.97</td>
<td>13</td>
<td>kn</td>
</tr>
<tr>
<td>Block Co-efficient</td>
<td>$C_b$</td>
<td>0.81</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>Displacement</td>
<td>$\Delta$</td>
<td>312622</td>
<td>93985</td>
<td>m$^3$</td>
</tr>
<tr>
<td>Propeller Diameter</td>
<td>$D_{prop}$</td>
<td>9.86</td>
<td>7.7</td>
<td>m</td>
</tr>
<tr>
<td>Propeller Expanded Area Ratio</td>
<td>$EAR$</td>
<td>0.431</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>Vessel Type</td>
<td></td>
<td>Tanker</td>
<td>Tanker</td>
<td></td>
</tr>
</tbody>
</table>

*Table 4: Table of Vessel Particulars*
In order to use the Holtrop-Mennen method to estimate the Still water resistance, the following parameters are also required. Suitable values were assumed wherever applicable, as highlighted below.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>KVLCC2</th>
<th>Vessel 1</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midship Coefficient</td>
<td>$C_M$</td>
<td>0.998</td>
<td>0.998</td>
</tr>
<tr>
<td>Shape of Stern</td>
<td>$C_{stern}$</td>
<td>-3</td>
<td>-3</td>
</tr>
<tr>
<td>Position of LCB</td>
<td>$LCB$</td>
<td>3.48</td>
<td>3 %</td>
</tr>
<tr>
<td>Depth</td>
<td></td>
<td>30</td>
<td>21.03</td>
</tr>
<tr>
<td>Bulbous Bow Transverse Area</td>
<td>$A_{BT}$</td>
<td>30</td>
<td>15</td>
</tr>
<tr>
<td>Height of Bulbous Bow</td>
<td>$H_{BT}$</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Immersed Transom Area</td>
<td>$A_{Transom}$</td>
<td>75</td>
<td>20</td>
</tr>
<tr>
<td>Lateral Windage Area</td>
<td>$A_L$</td>
<td>3737.5</td>
<td>2506.86</td>
</tr>
<tr>
<td>Transverse Windage Area</td>
<td>$A_T$</td>
<td>900</td>
<td>771.78</td>
</tr>
</tbody>
</table>

Table 5: *Additional Assumed Parameters*
KVLCC2 - Simulated Data

The KVLCC2 is a concept hull design based on SIMMAN, 2008, and a simple simulated dataset for the KVLCC2 was used; the aim of using this data set was predominantly to observe the functionality of LSTM Networks, particularly with modelling Malone’s Expressions, as the simulated data set reflects a very simplistic dataset - assuming ideal conditions. As there are no recorded power values - the total power is calculated based on the sum of the total resistances based on (1), where $R_{Fouling}$ is estimated using Malone’s Expressions. The simulated data set consisted of different features, given in Table. 20.

As the dataset was simulated - it allowed for several assumptions. In order to make the dataset complete, the following assumptions where made:

1. Each data record was assumed to be data recorded every 1/10th of a day.
2. The voyage distance was assumed to be constant between two ports, with a value of 16716.64 km.
3. At the end of each voyage, the vessel spends 2.5 days at port.
4. The vessel undertakes alternating ballast and laden voyages.
5. The hull is cleaned every 240 days.
6. The vessel is docked and repainted every 5 years.

Additionally to make use of Malone’ Expressions, the following assumptions are also considered:

1. **Hull Roughness Factor $HRF$**
   In order to evaluate the fouling roughness, the value of Hull Roughness Factor needs to be selected. For this problem, is it randomly selected for each of the records from Tab: 1.

2. **Fouling Roughness; $MAA_{Fouling}$**
   As per Malone’s expressions, the $MAA_{Total}$ is the sum of the individual roughness of the various layers. In this thesis, however it is assumed that the roughness due to fouling $MAA_{Fouling}$ accumulates over the each record, and the accumulation of the fouling in mils is given by

   $MAA_{FoulingT} = MAA_{FoulingT} + MAA_{FoulingT-1}$  \hspace{1cm} (41)

3. **Paint Specifications**
   The paint was assumed to have the following properties, based on sample values given in Malone’s paper.
### Added Features

The following features that can be easily estimated, and that are important factors in predicting the Fouling Resistance by Malone’s expressions are added as follows:

1. **Vessel Paint Age**
   
   An important factor that can controls the fouling resistance value given is the Vessel Paint Age, a factor that is calculated in days. Every time the hull is repainted, the Vessel Paint Age is reset to 0.

2. **Cumulative Port Days**

   Every day spend at harbour, the cumulative port day is incremented, and given in days, and is set to 0 every time the hull is cleaned or repainted. As it is considered that hull fouling occurs mostly at lower speeds - the time spend at port is a crucial factor to be considered as well while evaluating hull fouling.

3. **Days since Last Hull Cleaning (Cleaned Days)**

   This value is a counter that keeps track of the number of days since the last hull cleaning.

### Still Water Resistance and Propulsive Efficiency Estimation

In order to calculate the different wetted surface area $S$ at different drafts, expressions given in Holtrop & Mennen\(^3\) were used. Keeping in mind the conclusions arrived at previously, the following are the values used to evaluate the total resistance of the KVLCC2

1. $\eta_H$ is calculated using Thrust deduction factor and Wake Fraction values calculated using expressions given in Holtrop & Mennen.\(^3\)

2. $\eta_O$ is assumed to be constant at 0.5, a value chosen based on the expected loading condition of the propeller.

3. $\eta_R$ is calculated using expressions given in Holtrop & Mennen.\(^3\)

4. $\eta_S$ is assumed to be constant at 0.99, considering direct coupling between the propeller shaft and the engine flywheel without a gearbox.

---

**Table 6: Values of Parameters considered for the Vessel Paint for using Malone’s Expressions for KVLCC2**

<table>
<thead>
<tr>
<th>Anti Fouling Paint Thickness</th>
<th>Anti Fouling Paint Number of Coats</th>
<th>Anti Corrosion Paint Thickness</th>
<th>Anti Corrosion Paint Number of Coats</th>
<th>Anti Fouling Paint Effective Life</th>
</tr>
</thead>
<tbody>
<tr>
<td>$BLR_{AF}$ mils</td>
<td>$NC_{AF}$ mils</td>
<td>$BLR_{AC}$ mils</td>
<td>$NC_{AC}$ mils</td>
<td>Years</td>
</tr>
<tr>
<td>0.48</td>
<td>2</td>
<td>0.55</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

---

\(^3\) Holtrop & Mennen, 2000.
Resistance Plot

The Still water resistance is calculated using Holtrop and Mennen’s expressions as mentioned previously. In order to evaluate the Wind Resistance, the ship heading was set at 0, as the relative wind direction was given. The resultant wind direction and speed was then calculated. The value of $C_X$ was taken from the Manoeuvring Technical Manual (Brix, J, 1987)\textsuperscript{11} for tankers in ballast and loaded conditions respectively. To calculate the Fouling Resistance from the hull roughness given by Malone’s equations, expressions from Holtrop and Mennon are used, over expressions given in ITTC - 78 - this will be explained in succeeding sections. The resistances are evaluated and found as follows:

![Resistence vs Speed Plot for KVLCC2](image)

**Figure 11: Plot of Different Resistances vs Speed, KVLCC2**

The two different distinct trends in the still water resistance reflects the still water resistance in the laden and ballast conditions. It was observed that the Added Resistance in Waves was quite high at lower speeds and reduced as the speed increased and eventually became negative. However as for the KVLCC2 this value was given as a part of the simulated dataset, this pattern not considered and the added resistance was not included as an input feature. The wind resistance value is quite low comparing to the other resistances. - as well as the calculated Fouling Resistance - around the range of 30kN. The plot of hull roughness is obtained as follows:
This plot demonstrates Malone’s assumptions, there is no hull fouling as long as the anti-fouling paint is effective - upto about 3 years. After 3 years as the hull paint begins to fail - hull growth begins as the CEFF value increases till the hull is repainted at 5 years. From this plot - it can be seen that even after the life of the anti-fouling paint - it is still somewhat effective and is never completely ineffective. (A CEFF value of 0 indicates no fouling at all, 1 denotes maximum fouling). The hull roughness itself is quite low - 230µm - even when the hull fouling is assumed to accumulate every 1/10 of a day. After every hull cleaning - the hull roughness resets to 150µm (the considered minimum value of hull roughness). As there are a lot of data points, while using the Machine Learning Algorithms - it is easier to split the dataset into smaller segments.
**Vessel 1 - Recorded Data**

For **Vessel 1**, comprehensive daily noon reports for about 13.5 years of the vessel operations is used, complete with details of history of vessel hull cleaning, propeller cleaning and vessel dry docking. The records are mostly evenly spaced with an average interval 24 hours, with variations. Out of the various data features recorded, a select subset of the attributes were used for the data analysis, as given in Table 21.

**Engine Power**

For **Vessel 1**, the Main Engine power values was recorded only for a subset of records. However the fuel consumed in the time period of the record is given for most records - and these values can be used to calculate the average engine power over the record duration, from the engine specific fuel oil consumption - SFOC.

As per the OEM Manual for the installed engine (MAN B& W S70MC6), the engine SFOC was given as 171 g/kWhr for a fuel with a calorific value of 42,700 kJ/kg. In case of **Vessel 1** where the predominantly used fuel was HFO with a calorific value of 40,300 kJ/kg, the SFOC value is linearly approximated to 181 g/kWhr. Based on this, the average power in kW is calculated as

\[
P_B = \frac{\text{Fuel consumed}[\text{ton}] \times 10^6}{\text{Duration}[\text{hr}] \times \text{SFOC}[\text{g/kWhr}]} \tag{42}
\]

For a subset of records, the run duration of the Main Engine was recorded - which was used for the value of duration. In cases where the values were not recorded, the report duration was assumed as the engine run duration. In order to consider for variations of the SFOC with engine loading, a normal distribution of values was generated, with a mean value of 181 g/kWhr and a standard deviation of ±4 g/kWhr, and a random value from this distribution was selected to evaluate the Power for each record.

\[
SFOC_{MGO} \sim N(181, 16) \tag{43}
\]

In very few records where MGO was used, the above procedure was repeated with the respective SFOC value. For a thorough analysis of power, ideally the values of measured power would be recorded with high frequency and then integrated over the 24 hours duration in order to obtain the average instantaneous power the engine delivered over the time of the vessel record. However this rigorous process can be easily replaced the above calculation with little effect on overall accuracy. Additionally the randomly selected SFOC value considers the variation in the SFOC based on engine loading, RPM and other factors.
Added Resistance in Waves Calculation

The waves coming from the North is recorded as 180°. Both the relative and Meteostation data in the global notation are recorded, and as the Meteostation data is considered more accurate, it was used wherever available. In order to correct for the heading and the global notation, and to bring the wave heading to the notation as given in above figure, the following transformation is used:

\[
W_{\text{dir}} = \begin{cases} 
W_{\text{dir Relative}}, & \text{if } W_{\text{dir Relative}} \text{ available} \\
W_{\text{dir global}} - \psi - 360, & \text{if } W_{\text{dir global}} - \psi > 180 \\
W_{\text{dir global}} - \psi + 360, & \text{if } W_{\text{dir global}} - \psi < -180
\end{cases}
\]

Once the corrected \(W_{\text{dir}}\), the RAO is generated from\(^{15}\) with the inputs as given in Tab: 24.
Wind Resistance Calculation

The wind coming from the North is recorded as 0°. For the recorded data, both the relative wind direction and the Meteostation data are recorded. It was observed that the Meteostation data was more accurate, therefore wherever available, was used. In records where the Meteostation data was not recorded, the relative wind direction and wind speed was used. The program calculates the resistance based on the heading and wind direction in the global scale.

In order to correct the recorded values where the relative wind direction values were used, the following relation was used

$$W_{\text{global}}, \psi = \begin{cases} W_{\text{dir Recorded}} ; & \psi = \psi_{\text{Recorded}} \quad \text{if } W_{\text{dir recorded}} = W_{\text{dir Meteostation}} \\ W_{\text{dir Recorded}} ; & \psi = 0, \quad \text{if } W_{\text{dir recorded}} = W_{\text{dir Relative}} \end{cases}$$

Using these expressions and based on the wind speed due to the forward motion of the vessel, the relative wind speed $R_{\text{wind}}$ and the resultant angle $\varepsilon$ is evaluated. The value of $C_X$ was provided for the Vessel 1 for various values of $\varepsilon$ and was approximated to the following polynomial.

$$C_X = \left(8.09E - 07\right)e^3 + 0.000223e^2 - 0.00673e - 0.7045 \quad (44)$$
Inversely Calculated Fouling Resistance

When evaluating the data of Vessel 1, the fouling resistance was evaluated with the following approach. The total resistance of the hull is calculated based on the average power as

$$R_T = \frac{P_B \eta_{Prop}}{U}$$ \hspace{1cm} (45)

Where $P_B$ is calculated from (42). The fouling resistance is then assumed to be

$$R_{Fouling Inv} = R_T - (R_{Still Water} + R_{Wind} + R_{Added Resistance in Waves})$$ \hspace{1cm} (46)

Figure 15: Plot of Resistance components vs Speed (Left), Plot of Power vs Speed (Right);

Calculated Total Power $P_{Calculated Total} = (R_{Still Water} + R_{Wind} + R_{Added Resistance in Waves}) \frac{\eta_{Prop}}{U}$, $P_B$, given by (42)

Using the expressions given by ITTC-78 during the data analysis, it was found that using the equations yielded negative fouling resistance values for a significant number of records, with no discernible pattern.
This however cannot be wholly attributed to outlier data, because as per (16,17), a positive value of Hull Roughness can give rise to a **negative** resistance coefficient, particularly at lower speeds. This is because it was considered the hulls which were smooth when newly built would experience a negative resistance. This can be observed in the following plot:

![Inversely Obtained Fouling Resistance - Using ITTC-78 Expressions](image)

**Figure 16: Inversely Obtained Fouling Resistance - Using ITTC-78 Expressions**

![Plot of $\Delta C_f$ vs $k_s$; for different Vessel Speeds from 1-8 m/s (Left) Inversely Obtained Hull Roughness $k_{s_{reverse}}$ - Using ITTC - 78 (Right)](image)

**Figure 17: Plot of $\Delta C_f$ vs $k_s$: for different Vessel Speeds from 1-8 m/s (Left) Inversely Obtained Hull Roughness $k_{s_{inverse}}$ - Using ITTC - 78 (Right)**
Therefore, the fouling resistance coefficient can be inversely obtained by rewriting Eqn (17) as

\[
\Delta C_{f,\text{inverse}} = \frac{R_{\text{Fouling Inv}}}{0.5 \rho SU^2};
\]

(47)

and thereby the hull roughness can be obtained by rewriting (16) as

\[
k_{s,\text{Inverse}} = L_{\text{vol}} \left\{ \frac{(\Delta C_{f,\text{inverse}} - 0.00125)}{0.044} + 10Rn^{-1} \right\}^3
\]

(48)

Using the above relations, the \(k_{s,\text{Inverse}}\) is obtained as follows shown in Figure: 17. This yields several outliers of data, as Surface Roughness \(k_s\) is value that cannot be negative. Therefore a different method to obtain the \(R_{\text{Fouling Inv}}\) is sought. In order to overcome this, the expression for correlation allowance and frictional resistance coefficients used in calculating the Still Water Resistance were replaced with the expression in (Holtrop & Mennen, 1982)\(^3\) which combines the model correlation coefficient and frictional resistance coefficients with the following relation.

\[
C_A = \frac{(0.105k_s^{\frac{1}{3}} - 0.005579)}{L^{\frac{1}{3}}}
\]

(49)

Using the above expression when evaluating the Still Water Resistance, and then using (46) to calculate \(R_{\text{Fouling Inv}}\), the following plot is obtained.

![Inversely Obtained Fouling Resistance - Using Holtrop-Mennen Expressions](image)

**Figure 18: Inversely Obtained Fouling Resistance - Using Holtrop-Mennen Expressions**

The above expression yields better results for the fouling resistance values calculated, and therefore for **Vessel 1** the fouling resistance \(R_{\text{Fouling Inv}}\) is evaluated by (49), and the \(C_A\) value while evaluating the \(R_{\text{Still Water}}\) is set to
0. As seen from the above plots, the fouling resistance yet sometimes is calculated to be **negative**. This occurs yet again as a positive value of Hull Roughness can give rise to a **negative** resistance coefficient when using the Holtrop and Mennen expressions as well - though it is decoupled from the vessel speed, as seen below:

![Plot of Fouling Resistance Coefficient $C_A$ vs Hull Roughness $k_s$](plot1)

**Figure 19: Plot of $C_A$ vs $k_s$ (Left)
Inversely Obtained Fouling Resistance - Holtrop-Mennen Expressions (Right)**

Therefore negative fouling resistance values recorded can not be treated as outlier values. Now, calculating the $k_{s_{\text{inverse}}}$ by rewriting per (49) as

$$k_{s_{\text{inverse}}} = \left( \frac{(\Delta C_{f_{\text{inverse}}}L^{\frac{1}{3}}_{\text{ad}} + 0.005579)}{0.105} \right)^3$$  \hspace{1cm} (50)

The hull roughness values are obtained as shown in Figure: 19, which in accordance to expected values of hull roughness. This value of $R_{\text{fouling Inv}}$ and $k_{s_{\text{inverse}}}$ can be considered as the fouling resistance and the hull roughness experienced by the vessel at the particular time, considering all assumptions previously stated.

**Total Resistance Estimation**

In order to calculate the different vessel displacement $\Delta$ at different drafts, the values given by the vessel owner at different drafts were fitted to a polynomial as follows:

$$\Delta = (-0.1974T^3) + (57.122T^2) + (7288.6T) - 1322.3$$  \hspace{1cm} (51)

In order to calculate the different wetted surface area $S$ at different drafts, the expression given in Holtrop & Mennen$^3$ was used. Keeping in mind the conclusions arrived at previously, the following are the values used to evaluate the total resistance of Vessel 1.
1. $\eta_H$ is calculated using Thrust deduction factor and Wake Fraction values provided for Vessel 1, at different speeds and drafts, given in the Appendix at Table: 22 and 23.

2. $\eta_O$ is given by the Open Water Curve generated for a Standard Wageningen B-Series propeller, given in the Appendix at Fig. 53

3. $\eta_R$ is calculated using expressions given in Holtrop & Mennen.\(^3\)

4. $\eta_S$ is assumed to be constant at 0.99, considering direct coupling between the propeller shaft and the engine flywheel without a gearbox.

5. $C_A$ is set to 0 while calculating the Still Water resistance, as it is included in the Fouling Resistance when it evaluated using expressions given in Holtrop & Mennen.\(^3\)

**Added Features - Day Counters**

The following features that can be easily estimated, and that are important factors in predicting the Fouling Resistance are added as follows:

1. **Vessel Paint Age**
   An important factor that can controls the fouling resistance value given is the Vessel Paint Age, a factor that is calculated in days. Every time the hull is repainted, the Vessel Paint Age is reset to 0.

2. **Cumulative Port Days**
   The cumulative ports days are calculated based on the 'Report Type' recorded, and every day spend at harbour, the cumulative port day is incremented. This value is also given in days, and is set to 0 every time the hull is cleaned or repainted, as it is considered that hull fouling occurs mostly at lower speeds - the time spent at port is a crucial factor to be considered as well while evaluating hull fouling.

3. **Days since Last Hull Cleaning**
   This value is a counter that keeps track of the number of days since the last hull cleaning.

4. **Days since Last Propeller Cleaning**
   This value is a counter that keeps track of the number of days since the last propeller cleaning.
Time-line of Hull Cleaning Events

From the data received for Vessel 1, it can be seen that the vessel has been docked 4 times over 13 years, with several hull and propeller cleaning events interspersed. The following are some relevant conclusions to be drawn from this plot.

1. It is to be noted that in Period 2 - there have been several hull and propeller cleaning events, in spite of a recent vessel docking. It was confirmed that every time the vessel was docked the underwater hull water repainted, which suggested that this particular Anti-fouling Paint failed much earlier than expected.

2. All Hull cleaning events (except for one occurrence) are accompanied by a propeller cleaning. Therefore the effect of hull cleaning alone can not be properly estimated. In Period 3 and 4 - only propeller cleaning activities have been carried out. It is difficult to isolate the resistance improvements due to a hull cleaning and propeller cleaning independently, and in this paper the effect of hull cleaning alone is studied, while using propeller cleaning events as an added input feature.

Malone Fouling Resistance

It will be seen that for the Machine Learning Algorithm to predict the roughness with better accuracy, features obtained from the expressions given by Malone, Little, & Allman,\(^1\) can be used. As not all the values required to calculate the Fouling Resistance given by Malone are available, they are assumed such a manner that the Inversely calculated Fouling resistance \(R_{Fouling Inv}\) matches the Malone Fouling Resistance. Some of these values
are considerably high, considering the sample values given in their paper, however the assumptions that best match the inverse fouling data are made. The following are the factors considered and assumptions made.

1. **Hull Roughness Factor \( HRF \)**

   In order to evaluate the fouling roughness, the value of Hull Roughness Factor needs to be selected. For this problem, is it **randomly selected** for each of the records from Tab: 1.

2. **Fouling Roughness; \( MAA_{Fouling} \)**

   As per Malone’s expressions, the \( MAA_{Total} \) is the sum of the individual roughness of the various layers. In this thesis, however it is assumed that the roughness due to fouling \( MAA_{Fouling} \) accumulates over the days, and the accumulation of the fouling in mils is given by

   \[
   MAA_{Fouling_T} = MAA_{Fouling_T} + 0.1 * MAA_{Fouling_{T-1}} \tag{52}
   \]

3. **Steel Plate Roughness after repainted; \( MAA_{Steel\ Plate} \)**

   It is also noted that every time the hull is repainted, the \( MAA_{Steel\ Plate} \) is considered to increase as given in Table: 7.

4. **Painting Scheme**

   Different painting schemes were assumed during the different dry docking intervals- as given in Table: 7.

5. **Vessel Paint Effective Life** This is a critical value given in years, and as it is not given, it is assumed as given in Table: 7.

<table>
<thead>
<tr>
<th>Painting Scheme</th>
<th>Steel Base Roughness after Dry Docking</th>
<th>Anti Fouling Paint Thickness</th>
<th>Anti Corrosion Paint Thickness</th>
<th>Anti Fouling Paint Effective Life</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( mils )</td>
<td>( mils )</td>
<td>( BLR_{AF} )</td>
<td>( NC_{AF} )</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>0.55</td>
<td>3</td>
<td>0.55</td>
</tr>
<tr>
<td>1</td>
<td>150</td>
<td>0.55</td>
<td>3</td>
<td>0.55</td>
</tr>
<tr>
<td>2</td>
<td>125</td>
<td>0.55</td>
<td>3</td>
<td>0.55</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>0.55</td>
<td>3</td>
<td>0.55</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>0.55</td>
<td>3</td>
<td>0.55</td>
</tr>
</tbody>
</table>

**Table 7: Assumed Values of Parameters to estimate Fouling Resistance using Malone’s method for Vessel 1**

It is to be noted that Paint Life given by Paint suppliers in the product data sheets will differ from this Vessel Paint Failure Age value, and it is a value purely assumed in order that the CEFF value and subsequently the resistance values obtained match the observed resistance values.
6. **Anti-Fouling Coating Effectiveness Factor - CEFF**

The anti-fouling coating effectiveness factor is a value that is to take a value between 0 and 1 based on the Vessel Paint Age and the Effective Life of the Paint. The reasoning behind this parameter is that it is 0 when the paint is effective, thereby there is no increase in the fouling roughness, and it exponentially increases to 1 as the paint losses its effectiveness and is 1 when the paint is no longer effective. However it can be seen that using the expressions given, \( CEFF \) can take a value greater than 1. Therefore wherever it was calculated to be greater than 1, it was reset to 1.

![AntiFouling Coat Effectiveness Factor - CEFF](image)

**Figure 21: Plot of Anti-Fouling Coating Effectiveness Factor - CEFF vs. Ratio of Vessel Paint Age to Vessel Paint Effective Life, Z**

On making these assumptions and calculating the Fouling Resistance using Malone’s method, the values are obtained and compared to the \( R_{Fouling\ Inv} \) and \( k_{Inv} \), and though the values do not match exactly - the values do correlate reasonably well.
As it can be seen in the above plot, it is **not an exact match** of values - particularly as the observed hull roughness has several spikes, which are difficult to replicate using Malone’s expressions. It can be seen that using Malone’s expressions - the hull roughness increases till the hull is repainted or cleaned, but in the observed data this pattern can not be observed, the hull roughness any point can be higher or lower than previous hull roughness values, irrespective of Hull cleaning or repainting. Additionally the fouling resistance given using Malone’s expressions are never negative, as a minimum hull roughness of 150µm is assumed.

**Removing Outliers and Final Data**

Values recorded during extreme weather conditions need to be filtered out, therefore all data samples recorded when the Beaufort value was greater than 5 are filtered out. Additionally records where the inversely calculated hull roughness was calculated to be greater than 15mm are also removed, which is reflected in Fig.23 and Fig.17. Also records where the inversely calculated hull roughness was found to be negative are also removed. After filtering the data, there are 4157 records to use for the Machine Learning Algorithms. *The aim is to predict the actual fouling resistance the hull can encounter based on previous recorded values, as opposed to fouling resistances given by Malone.*
Trend-line of Inversely Calculated Fouling Resistance

![Graph showing trend of fouling resistance and hull roughness](image)

**Figure 23: Trend of Fouling Resistance, evaluated between hull cleaning and dry docking intervals**

On studying the trend of the fouling resistance $R_{Fouling\ Inv}$ and hull roughness $k_{5\text{inverse}}$, it can be seen that dry docking results in a significant drop in values, whereas as hull cleaning does not always reflect a drop in fouling resistance values. This can be attributed to several causes - perhaps erroneous recording of hull cleaning events, hull cleaning carried out due to available vessel port-time as opposed to necessary hull cleaning, or partial hull cleaning, different vessel operating profiles at different times. For most of the cases, hull cleaning does result in an slight improvement in $R_{Fouling\ Inv}$ and $k_{5\text{inverse}}$, though it is not as high an improvement when the vessel is repainted. The effect due to propeller cleaning can not be isolated - as all but 3 propeller cleaning events are not accompanied by hull cleaning. In this paper the independent effect of hull cleaning and propeller cleaning is not considered, but the feature 'Days since last propeller cleaning' is added to the machine learning algorithm to improve its predictions. On analysing the trends of the inversely calculated fouling resistance and the hull roughness, the following comments can be made.

1. **Period 0**

   Here there is no hull cleaning or propeller polish carried out, and based on data received - the value of the
variables like Vessel Paint Age, Days since Last Cleaning, Days since last Prop Cleaning, are significantly higher than the rest of the dataset. (about 10 years.) It will be seen that sometimes excluding this period from the model can help the overall fit of the model - due to the comparatively high numbers observed in this range.

2. Period 1

A comparatively long period of operation between dry docking of about 5 years - it can be seen there have been considerable gaps in vessel operation. The fouling resistance values experienced in this period are comparatively high - of the range 400-600 kN. There are 5 events in this period - 4 Hull cleaning along with propeller polishing, and 1 hull cleaning event alone. It can be in noted the case of the 4 hull cleaning events - 2 of the cleaning events has resulted in shifting the trend line lower, which is considered to imply the hull cleaning did have an impact on the fouling resistance the vessel experienced after it was cleaned. In the case of the hull cleaning event along with the propeller polish quickly followed by a hull cleaning - the trend line is broken because of the short time interval between the events. In the case of the final hull cleaning and propeller polish before the dry docking - it shows a downward trend, which is contrary to expectations - but could be attributed to the small time gap before the dry docking as well as the lower values of resistance observed after the dry docking.

3. Period 2

A eventful period of 3 years - with 5 hull cleaning events along with propeller polishing, and 1 propeller polish event. The trend lines do not consider the stand alone propeller polish events. The values of the fouling resistance in this period are lower compared to the previous period - mostly in the range of 200-400 kN, occasionally exceeding 400 kN. It can be seen however in this period - the hull cleaning and propeller polish events do not always result in a downward shift in the trend line, contrary to expectations. In some cases a descending trend is yet again observed, opposite to expectations. Due to the many recorded cleaning events recorded in this period - it is considered that the the hull paint in this period has failed quite early.

4. Period 3

A short period of approximately 2 years - there are only 2 propeller polish events recorded in this period, and in this period it can be seen that the fouling resistance is steady - neither increasing nor decreasing - and is consistently around 400 kN.

5. Period 4

The time from the previous dry docking to final data record received, the fouling resistance can be observed to be increasing - but the resistance values itself are within the 200-400 kN range.
Machine Learning Algorithms

Target Variable Selection

For this problem, one of two target variables can be used.

1. The Fouling Resistance $R_{Fouling}$ which can be directly predicted or

2. The Hull Roughness $k_s$, from which the Fouling Resistance can be subsequently calculated.

It was found over the course of working with the different algorithms that it was better to predict the hull roughness and calculate the fouling resistance accordingly. This is likely caused by the foul resistance being a function of other input features such as speed and draft. Predicting the hull roughness and thereby calculating the fouling resistance improved the logic of the future predictions. It is to be noted that for some models the MAE and the $R^2$ score for the Fouling resistance is better than the hull Roughness, and this is because while evaluating the Fouling Resistance, negatively predicted Hull Roughness values are set to 0.

Feature Selection

Feature selection is an important step of solving any Machine Learning Problem, as too many input features can make the program slow and use a lot of computing power; selecting few input features can make the error large. Additionally, selecting features which have good correlation to the target variable is important to avoid using unnecessary features. For Neural Networks - Permutation Feature Importance is to be carried out, where different input variables are given random numbers and the effect on the output is studied. However in this project this is not explored. An important facet to keep in mind is that variables like Power, Fuel Consumption cannot be used as input features as for future predictions if these inputs were available the fouling resistance can be calculated as opposed to predicted. Relevant features that are assumed to have an influence on hull roughness are selected using the heat map of input features correlation.
KVLCC2

Figure 24: Heat-map showing correlation of various features for KVLCC2

It can be seen that the fouling resistance correlates well with the Vessel Paint Age, the days since the last hull cleaning (Cleaned Days), the paint CEFF value and the hull roughness, all factors in Malone’s expressions. The still water resistance, loading condition, propulsive efficiency and other resistances have a low correlation, however they are still used in the algorithm when modelling the data set.
### Vessel 1

**Figure 25:** Heat-map showing correlation of various features for Vessel 1

Variables like the Event Type (Encoded Event_1, Encoded Event_2, as mentioned in Tab:3), Days since Last Cleaning, Days since last Prop Cleaning and MAA Total have comparatively low correlation with the Hull Roughness $k_s$. The Paint CEFF Value has a slightly better correlation - however the need for these variables will be demonstrated in succeeding sections.

### Structure of the LSTM Networks Model

The LSTM network is built as explained in the previous sections, with 750 Neurons in the bottom layer with tanh and a Dense layer with a single unit with a basic Neural network to give a single output. The tanh activation is used, as it is desired that input values of 0 give an activation of 0, as illustrated in the previous section. After several iterations - the following configuration was seen to work the best with the dataset. The model is optimized with the Adam Optimizer, and the MAE loss, and trained with a batch size of 180 days (In the case of the KVLCC2 this is 1800 records and in the case of Vessel 1, this is 180 records.) and when training the dataset, the training data set itself is set as the validation dataset. The fit of the LSTM network with Test data can be observed, however it is not considered of predominant importance in this project - as much as the logic of future predictions in the case of different cases. This will be elaborated further in succeeding sections.
Future Predictions for Vessel 1

In order to evaluate the future predictions of the different algorithms, a random sample of 180 recorded is selected from the available dataset of Vessel 1, and it is assumed this dataset continues from the last record of the total dataset. It is to be noted that the sequence of this dataset will be slightly erratic (For example, the loading condition can switch between loaded and ballast), however this is ignored. For the selected input features, the following assumptions are made.

1. **Speed**
   The speed is selected from a normal distribution of 1000 numbers, generated from the randomly selected record speed as
   \[ U_{Future} \sim N(U_{Random\ Record}, 0.0025) \]  
   (53)
   (Or with mean as the randomly selected record’s speed, between extreme values between ±0.5 m/s).

2. **Draft**
   The draft is evaluated in the same fashion;
   \[ T_{Future} \sim N(T_{Random\ Record}, 0.0025) \]  
   (54)
   (Or with mean as the selected record’s draft, between extreme values between ±0.5 m).

3. **Still Water Resistance and Propulsive Efficiency**
   The still water resistance and propulsive Efficiency \( \eta_{Prop} \) is then calculated based on the randomised speed and draft values using Holtrop-Mennen Expressions.

4. **Wind Resistance**
   The wind resistance of the future record is generated from the randomly selected record, by generating a normal distribution of 1000 numbers in the following fashion
   \[ R_{Wind\ Future} \sim N(R_{Wind\ Random\ Record}, 111, 111) \]  
   (55)
   (Or with mean value as the wind resistance of the randomly selected record, with extreme values between ±1 kN.)

5. **Added Resistance Resistance**
   The added resistance in waves of the future record is generated from the randomly selected record in the same fashion from a generated normal distribution of 1000 numbers as
   \[ R_{Added\ Resistance\ in\ Waves\ Future} \sim N(R_{Added\ Resistance\ in\ Waves\ Random\ Record}, 44.44 \times 10^6) \]  
   (56)
(Or with mean value as the wind resistance of the randomly selected record, with extreme values between
\( \pm 20kN \).)

In this process, the future predictions are ensured to have different values from the original dataset. The values of Vessel Paint Age, Days since Last Cleaning, Days since Prop Cleaning, Cumulative Port Day are all set to the values based on the last record of the data received. The future predictions of the algorithms are evaluated in three different cases.

(a) **Case 1: No Hull Cleaning**
In this case, the hull is not cleaned, and the values of 'Vessel Paint Age, Days since Last Cleaning, Days since Prop Cleaning, Cumulative Port Day' increment through the 180 records.

(b) **Case 2: Hull Cleaned and Propeller Polished on the 90\(^{th}\) Day**
In this case, it is assumed that the Hull is cleaned after 90 days - therefore the values of the features 'Days since Last Cleaning, Days since Prop Cleaning, Cumulative Port Day' are reset to 0 at Day 90 and then increase. The Vessel Paint Age is **not** reset.

(c) **Case 3: Vessel Dry Docking on the 90\(^{th}\) Day**
In this case, it is assumed that the vessel is docked after 90 days - therefore the values of 'Vessel Paint Age, Days since Last Cleaning, Days since Prop Cleaning, Cumulative Port Day' are all reset to 0 at Day 90 and then increase. A single day’s dry docking is not feasible in reality, but for the sake of evaluating the algorithm’s output predictions it is considered.

In all cases mentioned, correspondingly the values of Encoded Event Type are also modified.

6. **Anti Fouling Paint Effectiveness Factor CEFF**
This parameter is evaluated using Malone’s expressions, and according to the painting scheme in Period 4, this value is mostly 0 as the paint hasn’t failed in this period. After the hull is repainted in Case 3, the paint is assumed to have the same properties as the values in Painting Scheme 4, given in Tab: 7.

7. **Total Fouling Roughness** \( MAA_{Total} \)
This is calculated using Malone’s expressions and the assumptions given in Tab: 7. After the hull is re-painted in Case 3, the paint is assumed to have the same properties as the values in Painting Scheme 4, given in Tab: 7.
5 Results

KVLCC2

Linear Regression

Linear Regression using the Ordinary Least Squares method as explained in the previous sections is carried out. While fitting the data, the Mean Absolute Error and $R^2$ is evaluated based on the complete respective data set given to the model. The following are the input features used model the data.

<table>
<thead>
<tr>
<th>Data Used</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>[m/s]</td>
</tr>
<tr>
<td>Loading Condition</td>
<td></td>
</tr>
<tr>
<td>Still Water Resistance</td>
<td>[N]</td>
</tr>
<tr>
<td>Propulsive Efficiency</td>
<td></td>
</tr>
<tr>
<td>Wind Resistance</td>
<td>[N]</td>
</tr>
<tr>
<td>Vessel Paint Age in Days</td>
<td>[days]</td>
</tr>
<tr>
<td>Cleaned Days</td>
<td>[days]</td>
</tr>
<tr>
<td>Cumulative Port days</td>
<td>[days]</td>
</tr>
<tr>
<td>CEFF Value</td>
<td></td>
</tr>
</tbody>
</table>

Table 8: Input Features for Linear Regression Model - KVLCC2

From the following plot - it can be seen that linear regression does not give a very good fit to the input data, rather it predicts a slight increase in hull roughness till the hull is cleaned (as per given inputs this is every 240 days).
Though the plot for the data shows a relatively poor fit, the MAE and $R^2$ score of linear regression are good.

<table>
<thead>
<tr>
<th>OLS Model - Fit and Error</th>
<th>MAE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface Roughness [µm]</td>
<td>6.50</td>
<td>0.59</td>
</tr>
<tr>
<td>Fouling Resistance [kN]</td>
<td>2.27</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Table 9: Linear Regression - MAE and $R^2$ Values, KVLCC2
LSTM

In order to first train the LSTM Network - it was first set up with the total dataset - however as it consisted of 25000 data points, it took a considerable amount of time to compute. As it was of import to model the data after the paint failed, it was considered to model the data in the time between when the hull paint fails (Year 3) and when the hull is repainted (Year 5). After removing the rest of the data points - the model was trained using 80% of the data and tested on the remaining data. As it can be observed - the LSTM network does gives a much better fit to the data, and though it doesn’t not match with the test data exactly - it still does give reasonable predictions.

![Figure 27: LSTM - Plot of Model Predictions vs Target Variable, KVLCC2](image)

It can be observed that the LSTM still has MAE and $R^2$ values comparable to that of the Linear Regression Model, but it is to be kept in mind that if the MAE and $R^2$ for the Linear Regression model were calculated only for the data points between Year 3-5, the error value would be greater and the fit would be poorer. It can be seen that the LSTM model is able to model Malone’s expressions better than the Linear Regression model.

<table>
<thead>
<tr>
<th>LSTM - Fit and Error</th>
<th>MAE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface Roughness [μm]</td>
<td>6.66</td>
<td>0.61</td>
</tr>
<tr>
<td>Fouling Resistance [kN]</td>
<td>2.38</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Table 10: LSTM - MAE and $R^2$ Values, KVLCC2
Vessel 1

Linear Regression

Linear Regression using OLS as explained in the preceding sections is carried out. While fitting the data, the Mean Absolute Error and $R^2$ is evaluated based on the complete respective data set given to the model, and for this reason, it is not necessary to split the data into training and testing data sets for linear regression. Adding different input features at different stages, the MAE, $R^2$ and logic of the predicted fouling resistance is studied.

1st Model - Using only the available features

Using the following features, different models are generated using different segments of the input data, and their accuracy and $R^2$ value are observed.

<table>
<thead>
<tr>
<th>Data Used</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encoded Report Type</td>
<td></td>
</tr>
<tr>
<td>Encoded Cleaning Event</td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>[m/s]</td>
</tr>
<tr>
<td>Encoded Loading Condition</td>
<td>[m]</td>
</tr>
<tr>
<td>Draft</td>
<td>[m]</td>
</tr>
<tr>
<td>Still Water Resistance</td>
<td>[N]</td>
</tr>
<tr>
<td>Propulsive Efficiency</td>
<td>[N]</td>
</tr>
<tr>
<td>Wind Resistance</td>
<td>[N]</td>
</tr>
<tr>
<td>Added Resistance in Waves</td>
<td>[N]</td>
</tr>
<tr>
<td><strong>Target: Surface Roughness</strong></td>
<td>[µm]</td>
</tr>
</tbody>
</table>

Table 11: Inputs for the 1st OLS Model Type

Two types of models are considered - one where the a single linear regression model is generated based on the complete dataset termed Universal Model, and one where individual models are created for the different periods between the dry docking, termed Separated Models.
Figure 28: **Linear Regression - 1\textsuperscript{st} Model Type - Plot of Model Predictions vs Target Variable for different models with different training data ranges, Vessel 1**

<table>
<thead>
<tr>
<th>Universal Model</th>
<th>MAE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface Roughness [(\mu m)]</td>
<td>571.64</td>
<td>0.58</td>
</tr>
<tr>
<td>Fouling Resistance [kN]</td>
<td>49.89</td>
<td>0.73</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Separated Models</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface Roughness [(\mu m)]</td>
<td>397.41</td>
<td>0.74</td>
</tr>
<tr>
<td>Fouling Resistance [kN]</td>
<td>37.93</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Table 12: **Linear Regression - 1\textsuperscript{st} Model Type - Tabulated values of MAE and $R^2$ for the different models with different training data ranges, Vessel 1**

As expected, using separate models for the different periods gives a lower Mean Absolute Error and a better fit to the dataset. However to evaluate the three prediction cases as discussed earlier, the behaviour of a dry docking is also to be considered by the model. Keeping this in mind, a new intermediate model is built **The Chosen model** - using the data from the 1\textsuperscript{st} period to midway between the 2\textsuperscript{nd} period, and the output given in the 3 cases is plotted as follows:
It can be observed this model does not predict any changes in hull roughness and thereby fouling resistance in the three cases, as there is only one feature that changes between the three cases, the "Event Type" (Encoded Event Type). Therefore though this model has a good fit and reasonable errors values, this model can not be used to predict any fouling resistance improvements in the case of hull cleaning or repainting.
2\textsuperscript{nd} Model - Adding the Day Counter features to improve Predictability

In order to improve the future predictions of the model, the variables 'Vessel Paint Age, Days since Last Cleaning, Days since Prop Cleaning, Cumulative Port Day' are added to the input features as mentioned previously, and the same process is repeated. It is to be kept in mind that from the Heat-map that these parameters have comparatively low correlation with the hull roughness parameter and the fouling resistance as well.

![Linear Regression 2\textsuperscript{nd} Model - Plot of Predicted vs Target Values](image)

Figure 30: Linear Regression - 2\textsuperscript{nd} Model Type - Plot of Model Predictions vs Target Variable for different models with different training data ranges, Vessel 1

Note that the chosen model is built on data from Period 1 to midway between Period 2, and because of this - this model predicts extremely high values in Period 0, as the values of the different Day counter parameters in this period are higher than those in the rest of the data set.
Section: 5 Results

<table>
<thead>
<tr>
<th>Universal Model</th>
<th>MAE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface Roughness [$\mu m$]</td>
<td>629.27</td>
<td>0.64</td>
</tr>
<tr>
<td>Fouling Resistance [$kN$]</td>
<td>43.71</td>
<td>0.77</td>
</tr>
</tbody>
</table>

**Separated Models**

| Surface Roughness [$\mu m$] | 429.17 | 0.75 |
| Fouling Resistance [$kN$] | 35.23 | 0.83 |

Table 13: Linear Regression - $2^{nd}$ Model Type - Tabulated values of MAE and $R^2$ for the different models with different training data ranges, Vessel 1

It can be seen that adding these variables does improve the MAE and the $R^2$ value of the two cases, and as observed from the $1^{st}$ model - separate models give a better fit to the data versus a universal model from the complete data set. Building a model based the data from the $1^{st}$ period to midway between the $2^{nd}$ period, the output given in the 3 cases is plotted as follows:

![Predictions by the $2^{nd}$ Linear Regression Model (Including Day Counters)](image)

**Figure 31: Linear Regression - $2^{nd}$ Model Type - Future Predictions for the Three Cases (Model built from Selected Range)**

A difference can be observed in the hull roughness predicted by the model, in Case 1 the hull roughness is the highest, in Case 2 the hull roughness values are slightly lower and in Case 3 the hull roughness values are the lowest, and this is accordance to expectations. However it can be seen that predicted values merely shifted down in the three cases.
Adding features from Malone’s Expressions to improve the Results

For the final model, two parameters from Malone’s expressions - \( MAA_{Total} \) and \( CEFF \) are added to the above model, and then again the fit of the model with the data is studied.

![Graphs showing Linear Regression 3rd Model - Plot of Predicted vs Target Values for different models with different training data ranges, Vessel 1](image)

Figure 32: Linear Regression 3rd Model Type - Plot of Model Predictions vs Target Variable for different models with different training data ranges, Vessel 1

Again it is to be noted that the chosen model is built on data from Period 1 to midway between Period 2, and because of this - this model predicts extremely high values in Period 0, as the values of the Day counter parameters in this period are higher than those in the rest of the data set. Adding these parameters does not improve the MAE or the \( R^2 \), but a appreciable difference can be seen in the future predictions of this model, as can be seen in the following plot.
## Section: 5 Results

<table>
<thead>
<tr>
<th></th>
<th>Universal Model</th>
<th>MAE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface Roughness [µm]</td>
<td></td>
<td>628.15</td>
<td>0.64</td>
</tr>
<tr>
<td>Fouling Resistance [kN]</td>
<td></td>
<td>44.16</td>
<td>0.77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Separated Models</th>
<th>MAE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface Roughness [µm]</td>
<td></td>
<td>428.63</td>
<td>0.75</td>
</tr>
<tr>
<td>Fouling Resistance [kN]</td>
<td></td>
<td>35.13</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Table 14: *Linear Regression - 3rd Model Type* - Tabulated values of MAE and $R^2$ for the different models with different training data ranges, Vessel 1

![Predictions by the 3rd Linear Regression Model](image)

**Figure 33:** *Linear Regression - 3rd Model Type* - Future Predictions for the Three Cases (Model built from Selected Range)

A clear improvement can be observed between the three cases, though the model still predicts negative hull roughness values. Therefore adding these parameters does improve the model, even though it does not improve the fit of the model itself greatly.
The choice of the 'Chosen Model’ Data range

As seen, the chosen model which uses only the data from Period 1 to midway between period 2 is used to give predictions, as opposed to other possible period combinations. The reason why this particular range is chosen is - as observed earlier this is the range where the data follows expectations (i.e. hull cleaning improves the roughness etc). However, this period range also observes relatively higher fouling resistance values, which can be observed in the predicted values - where the predictions are all among the range of 600 kN and higher - quite high values of fouling resistance. Let us consider the predictions based on a model based on the total dataset.

![Predictions by the 3rd Linear Regression Model - Trained using Total Dataset](image)

**Figure 34: Linear Regression - 3rd Model Type - Future Predictions for the Three Cases(Model built from Total Dataset)**

Here something interesting can be observed - where in the case of the hull being cleaned the hull roughness is predicted to be higher than when it is not cleaned. A prediction that is contrary to expectations, but still in line with the majority of the observed data. The applicability of this prediction needs to be studied further. Additionally the fouling resistance values predicted are still on the higher end. Let’s consider a model made from the data from Period 2-4.
The same pattern can be observed, though the overall resistance values given are more aligned to the recently observed resistance values. In order to build a model which can make predictions which follow the expected values, the range of data between Period 1 to midway between Period 2 is considered.

For all the models, it can be observed that not all of them are able to capture the sudden ‘spikes’ in the hull roughness (except for some extent the separated models), and this is to be expected as these spikes themselves are quite random.

**Drawbacks**

1. As it can be observed, the predictions do not necessarily predict a trend but rather just a downward shift in the roughness/resistance values.

2. The model does predict negative hull roughness values, which is not desirable.

In order to overcome the above drawbacks, and the variation in predicted trends based on the input data, LSTM networks are considered.
LSTM

The LSTM Networks are built with the configuration previously specified. The network is made to train over a few epochs in order to stop the network from overfitting the data, but at the same time it is ensured that the loss value does reduce over the iterations. The obtained models are saved and reloaded to make predictions in the different cases, and the input file in the case of the different predictions are modified accordingly as well. After evaluating the results from the different models, the MAE and the $R^2$ value of the model over the training data is calculated. In the following sections, the input dataset is not split into training or testing data, but is given to the model as a whole to appraise how well the algorithm models the total dataset. Doing this training-test data split can reduce the bias of the model (if any) but it is not considered of importance in this paper, as the logic of the model’s predictions were of more significance.

1st Model - Using only the available features

An LSTM network is built as explained in the previous sections, using only the inputs given in Tab: 11, to observe if the Algorithm is capable of modelling the data without the day counters.

It appears that the LSTM model is not able to model the hull roughness without the day counter features, though the calculated fouling resistance do fit considerably well. Therefore this model can model the inversely obtained fouling resistance well, as seen from the MAE and $R^2$ score of the model.

Figure 36: LSTM - 1st Model Type - Plot of Model Predictions vs Target Variable for model trained with Total Dataset, Vessel 1
Table 15: LSTM - 1\textsuperscript{st} Model Type - Tabulated Values of MAE and $R^2$, Vessel 1

<table>
<thead>
<tr>
<th></th>
<th>MAE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface Roughness [$\mu m$]</td>
<td>480.95</td>
<td>0.50</td>
</tr>
<tr>
<td>Fouling Resistance [$kN$]</td>
<td>43.98</td>
<td>0.81</td>
</tr>
</tbody>
</table>

When using this model to predict improvements in the hull roughness in the future cases as described earlier, the following output was observed.

It can be seen that there is no change in the outputs given by the model, therefore though this model can to an extent predict the future fouling resistances, it can not predict an improvement in resistances in the event of a hull cleaning or hull repainting event, though the predicted fouling resistances are on the higher side.
2nd Model - Adding the Day Counter Features

In the next step, an LSTM model is built by adding the day counter features such as the 'Vessel Paint Age, Days since Last Cleaning, Days since Prop Cleaning, Cumulative Port Day' to the features mentioned in Tab: 11, and the following plot was observed.

![LSTM - 2nd Model Type - Plot of Model Predictions vs Target Variable for model trained with Total Dataset, Vessel 1](image)

This model clearly gives a superior fit to the hull roughness data than the previous model. Let's consider two more models, one built with the data from Period 2 to Period 4, and another built with the data from Period 3 to Period 4.
Figure 39: LSTM - 2nd Model Type - Plot of Model Predictions vs Target Variable for model trained with Data from Period 2-4, Vessel 1

Figure 40: LSTM - 2nd Model Type - Plot of Model Predictions vs Target Variable for model trained with Data from Period 3-4, Vessel 1
It can be easily observed that these models fit the data from their relevant periods better than the rest of the data set, and this can also be observed in the case of the MAE and $R^2$ values as well.

<table>
<thead>
<tr>
<th>LSTM - Fit and Error - 2nd Model</th>
<th>MAE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Universal Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Roughness [$\mu m$]</td>
<td>421.74</td>
<td>0.64</td>
</tr>
<tr>
<td>Fouling Resistance [kN]</td>
<td>39.26</td>
<td>0.85</td>
</tr>
<tr>
<td>From Period 2-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Roughness [$\mu m$]</td>
<td>478.66</td>
<td>0.60</td>
</tr>
<tr>
<td>Fouling Resistance [kN]</td>
<td>43.52</td>
<td>0.74</td>
</tr>
<tr>
<td>From Period 3-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Roughness [$\mu m$]</td>
<td>587.15</td>
<td>0.32</td>
</tr>
<tr>
<td>Fouling Resistance [kN]</td>
<td>64.31</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Table 16: LSTM - 2nd Model Type - Tabulated Values of MAE and $R^2$, Vessel 1 for the different Models trained with different data sets

Note: When evaluating the error and $R^2$ for the Models From Period 2-4 and the model Model 3-4, the data from Period 0 is removed, as the data in this period has high values in comparison to the rest of the dataset, thereby giving inaccurate error and fit values.

Now to evaluate the future predictions from these generated models.

Figure 41: LSTM - 2nd Model Type - Future Predictions for the Three Cases(Model built from Total Dataset)
In the case of the model built from the total dataset, the same difficulty as before arises, where the model predicts a better hull roughness when the hull is not cleaned as opposed to cleaned, which is contrary to expectations. For the second model based on data from Period 2 - 4, this same issue can be observed.

Figure 42: **LSTM - 2nd Model Type - Future Predictions for the Three Cases(Model built from Data from Period 2-4)**

For the predictions from the model built with the data from Period 3-4, the predictions are in accordance to expectations, however the model predicts almost marginal improvements for the different cases, as the input training data has no individual hull cleaning events, except for during the vessel dry docking. The values of the resistance itself are more in-line with the fouling resistance values observed in Period 4 - of the range 400 -600 kN.
Figure 43: LSTM - 2\textsuperscript{nd} Model Type - Future Predictions for the Three Cases(Model built from Data from Period 3-4)
3\textsuperscript{rd} Model - Adding the features from Malone’s Expressions

A final model is considered, with both the days counter features and adding two features from Malone’s expressions - the $\text{MAA}_{\text{Total}}$ calculated, and the Anti fouling paint effectiveness factor $\text{CEFF}$.

Figure 44: LSTM - 3\textsuperscript{rd} Model Type - Plot of Model Predictions vs Target Variable for model trained with Total Dataset, Vessel 1

The overall fit of the data with the target values does improve. Let us consider the 2 models in the same fashion seen earlier - one trained with the data from Period 2-4 and another trained with data from Period 3-4.
Figure 45: LSTM - 3rd Model Type - Plot of Model Predictions vs Target Variable for model trained with Data from Period 2-4, Vessel 1

Figure 46: LSTM - 3rd Model Type - Plot of Model Predictions vs Target Variable for model trained with Data from Period 3-4, Vessel 1
Table 17: LSTM - 3rd Model Type - Tabulated Values of MAE and $R^2$, Vessel 1 for the different Models trained with different data sets

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Surface Roughness [µm]</th>
<th>MAE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Universal Model</td>
<td>374.95</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td>Fouling Resistance [kN]</td>
<td>35.01</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>From Period 2-4</td>
<td>409.53</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>Surface Roughness [µm]</td>
<td>36.34</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>Fouling Resistance [kN]</td>
<td>445.66</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>From Period 3-4</td>
<td>38.59</td>
<td>0.81</td>
<td></td>
</tr>
</tbody>
</table>

Note: When evaluating the error and $R^2$ for the Models From Period 2-4 and the model Model 3-4, the data from Period 0 is removed, as the data in this period has high values in comparison to the rest of the dataset, thereby giving inaccurate error and fit values.

As it can be seen from the MAE and the $R^2$ values, these models do provide a slightly improved fit than the previous case, and as observed earlier - these models fit the input data better in their relevant data ranges than the rest of the dataset. Now the future predictions from the different models are considered.

Figure 47: LSTM - 3rd Model Type - Future Predictions for the Three Cases(Model built from Total Dataset)
Using the total dataset again predicts an increased hull roughness in the event of hull cleaning, contrary to logical expectations.

Figure 48: LSTM - 3rd Model Type - Future Predictions for the Three Cases(Model built from Data from Period 2-4)

An interesting development from the previous similar model is that now the model trained with data from Period 2-3 predicts values in accordance with expectations. It is to be kept in mind the CEFF value for the paint in Period 2 would be high, which could be aiding this particular model in giving this kind of a prediction.
As observed in the previous similar model, the predictions from this model are in accordance to expectations, but the differences between the three different cases are only slight, which is again because the input training data itself has no individual hull cleaning events, except for during the vessel dry docking. The values of the resistance itself are around the range of the values observed in Period 4, of around 400 kN.
Malone’s Predictions

Finally, as the various parameters to use Malone’s Expressions are already inversely arrived at, the hull fouling resistance given by Malone’s expressions are calculated.

As it can be seen, Malone’s expressions do not give any improvement in the fouling resistance, though the hull roughness improves when the hull is repainted. This is because it is considered that the hull is repainted before it fails, therefore the only improvement is the roughness due to the MAA Coating, which is marginal. No improvement is observed when the hull is cleaned, because as the hull paint hasn’t failed there is no fouling observed, both before and after the hull cleaning. The values of the resistance are quite high compared to the resistance values observed in Period 4.
The Data

Keeping in mind the assumptions made, the differences in the fouling resistance given by Malone’s expressions for fouling resistance the KVLCC2 and the observed fouling resistance in Vessel 1 are significant. As it can be seen that Malone’s expressions reflect a simplistic nature of hull fouling, where hull fouling does not occur at all till the hull anti fouling paint fails, and even after the paint fails - it still does not loose its total effectiveness immediately and repainting the hull returns the hull to its new ship condition. Working with the data from Vessel 1, it can be seen that this is not the case - the hull does experience additional resistances even after being repainted - which in this project is considered to be the fouling resistance. The hull roughness can be observed to be low after the vessel is repainted, but the baseline roughness is never as low as the new hull, presumably due to sandblasting the underwater hull. Therefore it can be concluded that the hull roughness in reality and consequently the fouling resistance are far from Malone’s idealistic expressions. This is not to say Malone’s expressions can be disregarded as it has been shown that the various parameters from Malone’s expression can be added to available features to improve modelling of the fouling resistance.

The use of ITTC-78 vs Holtrop & Mennen Expressions

As seen in the Methodology section, when using equations from Holtrop & Mennon\(^3\) to evaluate the still water resistance and calculating the Inversely calculated fouling resistance, it can be seen that often the inversely calculated fouling resistance is found to be negative. It can be argued that Holtrop & Mennon’s expressions give a conservative value of still water resistance, in comparison to other methods of evaluating the still water resistance such as Guldhammer and Harvald or Hollenbach. Understandably the value of the Inversely Calculated Fouling Resistance and consequently the Inversely calculated hull roughness will be different when using the different methods, the value of the inversely calculated hull roughness obtained must be investigated before proceeding with the results, because as shown - hull roughness can not take a negative value. In the expressions for the fouling resistance coefficient used in this paper (given by Holtrop & Mennen) the equation does not consider the vessel speed. ITTC-78’s expressions do consider the vessel speed in the evaluation of the fouling resistance coefficient (in the form of the Reynold’s number) and the inter-relatedness between these two features can additionally add another level of understanding the nature of vessel hull fouling. However as discussed - negative fouling resistances can be observed in both cases - but in the case of the ITTC -78 expressions (refer Fig: 17), at lower speeds - substantially higher hull roughness values can still give a negative fouling resistance coefficient, which is to be kept in mind if this method is used. It can be considered that the ITTC-78’s expressions are better suited for evaluating the vessel resistance during the vessel design stages, and the use of these expressions to evaluate recorded data to study hull fouling resistances values is to be done with care.
The Fouling Resistance

In the case of the KVLCC2, it is given by Malone’s expressions, and for a 320m ship is calculated to be of the range of only 30kN. In the case of Vessel 1, it can be seen that the observed values of fouling resistance are in the range of 200-600kN, with occasionally spikes to higher numbers as well. With Vessel 1, where it is inversely calculated using the vessel’s fuel consumption, it can be seen that the long term data is contrary to expectations in some areas. There are periods where a hull cleaning shows an increase in the trend of fouling resistance, there are periods where the trend line of the fouling resistance is decreasing as opposed to increasing. These discrepancies can be seen to be absorbed by the model as well, as shown in the previous section. Keeping in mind the assumptions made - the recorded data does appear to sometimes defy expectations. This is to be kept in mind while choosing the model and its training data, and it can be observed that sometimes using the total dataset is erroneous.

It is to be noted that the Wake Fraction and the Thrust deduction factor for Vessel 1 has a few irregularities - where the wake fraction at specific drafts differs greatly from the nearby values, specifically around 11.1m. This discrepancy could give rise to some irregular values in the propulsive efficiency $\eta_{prop}$ values.

On average, the fouling resistance makes up approximately 10% of the total resistance in the case of Vessel 1, and it is understood and that any procedures that can result in the improvements of this number can lead to significant reduction in vessel fuel consumption.

The Effectiveness of Hull Cleaning

Based on the dataset from Vessel 1, it would be specious to conclude that Hull cleaning does not contribute to improvements in the fouling resistance. The data at different time periods differs in operating area and profile, as well as several other factors. It can also be argued that the fouling resistance could have been worse were the hull not cleaned. In the same fashion that no two hulls can have the exact same hull fouling, no two hull cleaning events can be said to have the same effect because of these intricate differences. In order to truly evaluate the effectiveness of hull cleaning, it would be required to study vessels with the same operating profile and operating area between the same type of vessel, or evaluate several datasets from different vessels, and normalize the data to a common value based on different reasonable assumptions.
The Machine Learning Algorithms

KVLCC2

For the KVLCC2, the dataset is quite simplified and the Machine Learning algorithms are able to essentially model Malone’s expressions relatively well. It can be seen that the LSTM model is much better than the Linear Regression model, though the latter can be improved on if more input features are considered.

Vessel 1

It can be observed that Machine Learning Algorithms can model the Fouling resistance observed by the hull. Linear Regression can give a quick and simple model of the fouling resistance, and using features from Malone’s expressions helps improve the model quite well.

The LSTM network shows marked improvements over the Linear Regression, in that the model uses the memory of the hull roughness (and consequently the fouling resistance) in the last period of the data set and predicts values in accordance to the recent data, something the linear regression model can not do. However it can be seen that the choice of the input features and the training data range of the model affects the outputs markedly.

Linear Regression vs LSTM Models

It can be observed that the Linear Regression models’ prediction are more of a shift in the values as opposed to a predicted trend. Considering the results from the most suitable Linear Model and LSTM Networks for Vessel 1, the following can be seen:
Linear regression gives predictions along the same range as its training data (between Period 1- to midway between Period 2.) where the resistance values are one the higher side. Consequently this model predicts fouling resistances around 600-800 kN, which is higher than the values seen in Period 4. The improvement itself predicted by the different cases is of the order of about 100kN between each of the three cases, at its peak value. The LSTM model predicts resistances in the range of 200-400kN - as it has modelled the patterns overall and the thereby predicts roughness and thereby resistance values in a more expected range, based on more recent values. The improvement itself predicted by the different cases is of the order of about 50kN between each of the three cases, at its peak value. The author strongly feels that the LSTM network’s predictions are more in line with expected values of fouling resistance.
The Anti Fouling Paint Effectiveness Factor, CEFF

In the case of the KVLCC2, the effect of this factor can be clearly seen - there is no hull growth up unto the point the hull anti fouling paint fails at 3 years, and after the paints - the paint still contributes marginally to preventing hull growth till the hull is repainted at 5 years. However in the case of Vessel 1, where the values of the Effective Paint Life had to be assumed in order to match the observed fouling resistance - it can be seen that the value of the Effective Paint Life is quite low. Compared to the paint Product Data sheet values of product life of 7.5 years (JOTUN, 2018) - it can be seen that the inversely arrived-at values are comparative low. It is understood that the product data sheet value refers to the overall lifetime of the paint considering several other factors; however similar factor such as the CEFF can be included by paint company manufacturers for different paints - which considers the time period over which the hull anti fouling paint is most effective.

<table>
<thead>
<tr>
<th>Painting Scheme for different Periods</th>
<th>Anti Fouling Paint Effective Life (Assumed) Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 18: Values of Effective Paint Life assumed for Vessel 1, in order for Malone’s Expression to yield fouling resistance values to the same extent as the inversely calculated fouling resistance.

The Training -Testing Data split

The conventional training-testing data split is not dealt with in detail in this project. An aim of the project was to consider if the machine algorithms were capable of modelling the input data as a whole, as well as give good predictions. For example, consider this case where the model is trained with 80% of the dataset and tested on the remaining 20% of the dataset, a typical train-test process. The model predictions on the test data are observed as follows:
The model does work well - except for predicting the random spikes, however when using this model to make future predictions the same discrepancy was observed, where the predictions when the hull is cleaned is greater than when the hull is not cleaned. In order to evaluate the logic of the model rather than overall fit with unseen test data, the training -testing data split was not considered it this project, as considering this would only result in marginal changes in the overall model.

The different input feature combinations

As it has been demonstrated, the models can effectively model the observed fouling resistance - even without the Day counter variables, and adding more features which have little correlation with the hull roughness numerically - has been shown to improve the logic and the practicality of the model’s predictions.

Choosing a suitable Model

It can be observed in some cases that selecting a model which has a low $R^2$ value and high MAE over the entire dataset gives more reasonable results than models which have better $R^2$ values and lower errors. This suggests that a model should not be assessed purely on these metrics alone- but also by the practicality of its results as well. As the long term recorded data sometimes is contrary to expectations, the decision to chose a model that makes expected predictions in accordance with expectations versus a model that gives predictions based on
observed discrepancies is a decision that is to be made. As it can be seen - the algorithms are capable of modelling both cases, and choosing an appropriate model is the next step. Adding relevant features, particularly ones that reflect the paint functionality would be extremely valuable in making the models more balanced = as observed. A single dataset can be used to generate several models - using different combinations of input features and training datasets, therefore it can be stated that the next step would involving choosing a suitable model - one that captures both the nuances of the observed dataset as well as logical expectations.

**Future Improvements**

The following are various improvements that can be done in order to obtain better understand of hull fouling.

1. **Include the Effects of Shallow Water, Trim, Steering Resistances in the total resistance estimation**
   Though these resistance values are neglected in this project, including these features can improve the dataset and consequently the model - as some of the spikes observed could be attributed to these effects.

2. **Split the datasets in the different periods as Training and Test Data**
   Splitting the dataset into training and test data can improve the uncertainty of the model, or remove any implicit bias the model develops when being trained, however as stated previously it is expected that this will result in marginal changes.

3. **Permutation Feature Importance**
   As mentioned earlier, this is a method of calculating the feature importance of a neural network - carrying this out on the various possible input features can help in identifying features that though have low correlation with the target variable - aid in building a better model.

4. **Isolating the effect of propeller cleaning if possible data sets are available**
   It can be argued that a propeller polish can in some cases provide better improvements in hull resistance at certain times as opposed to a complete hull cleaning, and in order to evaluate this - datasets where the hull alone is cleaned without a propeller polish and then where the propeller is polished without the hull being cleaned can be studied. The latter would be more easily attainable than the former - but studying datasets can help study the improvements due to these activities individually.

5. **Evaluating the Hull Still Water resistance by different methods**
   As it was observed using Holtrop and Mennon’s expressions results in Inversely calculated fouling resistances values to be negative for a significant number of the data records. However using different methods to evaluate the Still Water Resistance such as Guldhammer and Harvald, or Hollenbach could give rise to different values of Still Water Resistance, and thereby Inversely Calculated Fouling Resistance.
6. **Use Different Machine Learning Algorithms**
   There are several other kind of machine learning algorithms available - and regression models such as Gaussian Process Regression can be considered.

7. **Using the LSTM Network with Different Time Steps**
   Generating different models using different time steps of the training data set can generate Networks which give different outputs, and this can also be studied further.

8. **Using Features from other Methods to Evaluate the Fouling Resistance**
   As it has been observed, adding features from Malone’s expressions improve the model, this can be further extended to using features from other methods to calculate the Fouling Resistance.
7 Conclusion

Hull fouling is expected to play a part in vessel performance until environment-friendly and effective anti-fouling paints are developed. Hull fouling will play a factor in all vessels’ performances - be it wind-assisted, battery-operated, biofuel or conventional fuel ships. There is a considerable potential for improving vessel performance if hull fouling can be determined at different points in time. The ultimate objective of this endeavour would be to pinpoint the exact future time at which the cost of the additional fuel consumption due to increased hull fouling would exceed the cost of cleaning the hull. With increasingly stringent emission regulations and focus on data analysis by the International Maritime Organization - evaluating hull fouling and the fouling resistance is an undertaking where even a little clarity can bring about sizeable benefits.

Hull fouling can be considered a function of the vessel operating profile, the hull paint and its properties and time - and using these parameters to generate models from existent vessel data can improve the understanding of the hull fouling process. As it has been illustrated, few alterations need to be made to the expressions while evaluating the data records of vessels, as well as a critical study of the dataset. Several models can be built from a single dataset as demonstrated, based on different combinations of input features and training datasets; and studying carefully which model best fits the future data is the next step.

Generating models from the data recorded from different vessels with different operating profiles could help in developing empirical relations that can predict the fouling resistance over time, with a given level of accuracy, as after all hull fouling is still a random process. Models that fit the data well can be used for this purpose, and other parameters which are expected to influence the hull fouling can be included as well. However, care is to be taken while generating these models as the model will reflect its training data. Reducing the level of uncertainty in its input data will improve the model immensely. The results from these algorithms can be compared with the vessel fuel consumption in real time and corrected, improved and validated.

The requirement for parameters like ’Effective Paint Life’ and ’Hull Cleaning Effectiveness’ is also evident, as these attributes can both improve modelling of vessel fouling, as well as give a better comparison and benchmark the different aspects of hull paints and underwater hull cleaning.

Hull fouling and fouling resistance can no longer be considered to be a completely random variable, but can be modelled to a certain degree of accuracy - even individually for different vessels based on its own recorded data, with improvements to the procedures mentioned in this paper and further investigation.
References


8 Appendix

Constant Values Used

<table>
<thead>
<tr>
<th>Constant Values used</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceleration due to gravity ( g )</td>
<td>9.81</td>
<td>( m/s )</td>
</tr>
<tr>
<td>Density of Seawater ( \rho )</td>
<td>1025</td>
<td>( kg/m^3 )</td>
</tr>
<tr>
<td>Dynamic Viscosity of Water ( \eta )</td>
<td>0.000001139</td>
<td>( Pa \cdot s )</td>
</tr>
</tbody>
</table>

Table 19: Constant Values Used

List of Variables used from Simulated Data of KVLCC2

<table>
<thead>
<tr>
<th>List of Variables used from Simulated Data of KVLCC2</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velocity</td>
<td>( m/s )</td>
</tr>
<tr>
<td>Added Resistance in Waves</td>
<td>( N )</td>
</tr>
<tr>
<td>Calm Water Resistance</td>
<td>( N )</td>
</tr>
<tr>
<td>Wave Direction</td>
<td>( deg )</td>
</tr>
<tr>
<td>Wind Velocity</td>
<td>( m/s )</td>
</tr>
<tr>
<td>Wind Direction</td>
<td>( deg )</td>
</tr>
</tbody>
</table>

Table 20: List of Variables used from Simulated Data of KVLCC2
List of Variables used from Recorded Data of Vessel 1

<table>
<thead>
<tr>
<th>Data Used</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report Date</td>
<td>Date</td>
</tr>
<tr>
<td>Report Duration</td>
<td>[h]</td>
</tr>
<tr>
<td>Report Type</td>
<td></td>
</tr>
<tr>
<td>Draft Fore</td>
<td>[m]</td>
</tr>
<tr>
<td>Draft Aft</td>
<td>[m]</td>
</tr>
<tr>
<td>Trim</td>
<td>[m]</td>
</tr>
<tr>
<td>Mean Draft</td>
<td>[m]</td>
</tr>
<tr>
<td>Ship Heading</td>
<td>[deg]</td>
</tr>
<tr>
<td>Observed Miles</td>
<td>[nm]</td>
</tr>
<tr>
<td>Logged Miles</td>
<td>[nm]</td>
</tr>
<tr>
<td>Weather-Wind Speed at 10m</td>
<td>[m/s]</td>
</tr>
<tr>
<td>Weather-Wind Direction</td>
<td>[deg]</td>
</tr>
<tr>
<td>Weather-Wave Direction</td>
<td>[deg]</td>
</tr>
<tr>
<td>Meteostation Weather-Wind Speed at 10m</td>
<td>[m/s]</td>
</tr>
<tr>
<td>Meteostation Weather-Wind Direction</td>
<td>[deg]</td>
</tr>
<tr>
<td>Meteostation Weather-Wave Direction</td>
<td>[deg]</td>
</tr>
<tr>
<td>Beaufort</td>
<td>[BF]</td>
</tr>
<tr>
<td>ME1 Power</td>
<td>[kW]</td>
</tr>
<tr>
<td>ME1 Run Duration</td>
<td>[h]</td>
</tr>
<tr>
<td>ME1-LS MGO-Consumption</td>
<td>[t]</td>
</tr>
<tr>
<td>ME1-LS MGO-LCV</td>
<td>[kJ/kg]</td>
</tr>
<tr>
<td>ME1-LS HFO-Consumption</td>
<td>[t]</td>
</tr>
<tr>
<td>ME1-LS HFO-LCV</td>
<td>[kJ/kg]</td>
</tr>
<tr>
<td>ME1-HS HFO-Consumption</td>
<td>[t]</td>
</tr>
<tr>
<td>ME1-HS HFO-LCV</td>
<td>[kJ/kg]</td>
</tr>
<tr>
<td>ME1-HS MGO-Consumption</td>
<td>[t]</td>
</tr>
<tr>
<td>ME1-HS MGO-LCV</td>
<td>[kJ/kg]</td>
</tr>
</tbody>
</table>

Table 21: List of Variables used from Recorded Data of Vessel 1
## Lookup Table for Wake Fraction for Vessel 1

<table>
<thead>
<tr>
<th>Draft [m]</th>
<th>Speed [m/s]</th>
<th>6.66</th>
<th>7.4</th>
<th>7.7</th>
<th>8</th>
<th>8.92</th>
<th>11.1</th>
<th>12.3</th>
<th>13.6</th>
<th>14.82</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.833</td>
<td>0.415 0.415</td>
<td>0.415</td>
<td>0.415</td>
<td>0.415</td>
<td>0.404</td>
<td>0.339</td>
<td>0.311</td>
<td>0.307</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>4.312</td>
<td>0.415 0.415</td>
<td>0.415</td>
<td>0.415</td>
<td>0.415</td>
<td>0.404</td>
<td>0.339</td>
<td>0.311</td>
<td>0.307</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>5.318</td>
<td>0.415 0.415</td>
<td>0.415</td>
<td>0.415</td>
<td>0.415</td>
<td>0.404</td>
<td>0.339</td>
<td>0.311</td>
<td>0.307</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>5.654</td>
<td>0.415 0.415</td>
<td>0.415</td>
<td>0.415</td>
<td>0.415</td>
<td>0.404</td>
<td>0.339</td>
<td>0.311</td>
<td>0.307</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>5.932</td>
<td>0.415 0.415</td>
<td>0.415</td>
<td>0.415</td>
<td>0.415</td>
<td>0.404</td>
<td>0.339</td>
<td>0.311</td>
<td>0.307</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>6.229</td>
<td>0.415 0.415</td>
<td>0.408</td>
<td>0.408</td>
<td>0.336</td>
<td>0.311</td>
<td>0.305</td>
<td>0.288</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.525</td>
<td>0.408 0.408</td>
<td>0.402</td>
<td>0.402</td>
<td>0.395</td>
<td>0.333</td>
<td>0.311</td>
<td>0.302</td>
<td>0.286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.822</td>
<td>0.402 0.402</td>
<td>0.395</td>
<td>0.395</td>
<td>0.391</td>
<td>0.33</td>
<td>0.311</td>
<td>0.3</td>
<td>0.284</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.119</td>
<td>0.395 0.395</td>
<td>0.388</td>
<td>0.388</td>
<td>0.386</td>
<td>0.327</td>
<td>0.31</td>
<td>0.297</td>
<td>0.282</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.415</td>
<td>0.388 0.388</td>
<td>0.382</td>
<td>0.382</td>
<td>0.382</td>
<td>0.324</td>
<td>0.31</td>
<td>0.295</td>
<td>0.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.712</td>
<td>0.382 0.382</td>
<td>0.375</td>
<td>0.375</td>
<td>0.377</td>
<td>0.321</td>
<td>0.31</td>
<td>0.293</td>
<td>0.277</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.009</td>
<td>0.375 0.375</td>
<td>0.369</td>
<td>0.369</td>
<td>0.373</td>
<td>0.317</td>
<td>0.307</td>
<td>0.29</td>
<td>0.275</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.305</td>
<td>0.369 0.369</td>
<td>0.362</td>
<td>0.362</td>
<td>0.362</td>
<td>0.368</td>
<td>0.314</td>
<td>0.304</td>
<td>0.288</td>
<td>0.273</td>
<td></td>
</tr>
<tr>
<td>8.602</td>
<td>0.362 0.362</td>
<td>0.355</td>
<td>0.355</td>
<td>0.355</td>
<td>0.364</td>
<td>0.311</td>
<td>0.301</td>
<td>0.286</td>
<td>0.271</td>
<td></td>
</tr>
<tr>
<td>8.899</td>
<td>0.355 0.355</td>
<td>0.355</td>
<td>0.355</td>
<td>0.355</td>
<td>0.364</td>
<td>0.311</td>
<td>0.301</td>
<td>0.286</td>
<td>0.271</td>
<td></td>
</tr>
<tr>
<td>9.195</td>
<td>0.355 0.355</td>
<td>0.355</td>
<td>0.355</td>
<td>0.355</td>
<td>0.364</td>
<td>0.311</td>
<td>0.301</td>
<td>0.286</td>
<td>0.271</td>
<td></td>
</tr>
<tr>
<td>9.492</td>
<td>0.355 0.355</td>
<td>0.355</td>
<td>0.355</td>
<td>0.355</td>
<td>0.364</td>
<td>0.311</td>
<td>0.301</td>
<td>0.286</td>
<td>0.271</td>
<td></td>
</tr>
<tr>
<td>9.788</td>
<td>0.355 0.355</td>
<td>0.355</td>
<td>0.355</td>
<td>0.355</td>
<td>0.364</td>
<td>0.311</td>
<td>0.301</td>
<td>0.286</td>
<td>0.271</td>
<td></td>
</tr>
<tr>
<td>10.085</td>
<td>0.355 0.355</td>
<td>0.355</td>
<td>0.355</td>
<td>0.355</td>
<td>0.364</td>
<td>0.311</td>
<td>0.301</td>
<td>0.286</td>
<td>0.271</td>
<td></td>
</tr>
</tbody>
</table>

Table 22: Lookup Table for Wake Fraction for Vessel 1
## Lookup Table for Thrust Deduction Factor for Vessel 1

<table>
<thead>
<tr>
<th>Draft [m]</th>
<th>6.66</th>
<th>7.4</th>
<th>7.7</th>
<th>8</th>
<th>8.92</th>
<th>11.1</th>
<th>12.3</th>
<th>13.6</th>
<th>14.82</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.833</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.158</td>
<td>0.179</td>
<td>0.168</td>
<td>0.18</td>
<td>0.199</td>
</tr>
<tr>
<td>4.312</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.158</td>
<td>0.179</td>
<td>0.168</td>
<td>0.18</td>
<td>0.199</td>
</tr>
<tr>
<td>5.318</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.158</td>
<td>0.179</td>
<td>0.168</td>
<td>0.18</td>
<td>0.199</td>
</tr>
<tr>
<td>5.654</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.158</td>
<td>0.179</td>
<td>0.168</td>
<td>0.18</td>
<td>0.199</td>
</tr>
<tr>
<td>5.932</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.158</td>
<td>0.179</td>
<td>0.168</td>
<td>0.18</td>
<td>0.199</td>
</tr>
<tr>
<td>6.229</td>
<td>0.186</td>
<td>0.186</td>
<td>0.184</td>
<td>0.184</td>
<td>0.158</td>
<td>0.178</td>
<td>0.168</td>
<td>0.18</td>
<td>0.197</td>
</tr>
<tr>
<td>6.525</td>
<td>0.184</td>
<td>0.184</td>
<td>0.181</td>
<td>0.181</td>
<td>0.158</td>
<td>0.177</td>
<td>0.168</td>
<td>0.18</td>
<td>0.195</td>
</tr>
<tr>
<td>6.822</td>
<td>0.181</td>
<td>0.181</td>
<td>0.178</td>
<td>0.178</td>
<td>0.158</td>
<td>0.175</td>
<td>0.168</td>
<td>0.18</td>
<td>0.192</td>
</tr>
<tr>
<td>7.119</td>
<td>0.178</td>
<td>0.178</td>
<td>0.175</td>
<td>0.175</td>
<td>0.158</td>
<td>0.174</td>
<td>0.168</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>7.415</td>
<td>0.175</td>
<td>0.175</td>
<td>0.172</td>
<td>0.172</td>
<td>0.158</td>
<td>0.173</td>
<td>0.168</td>
<td>0.18</td>
<td>0.188</td>
</tr>
<tr>
<td>7.712</td>
<td>0.172</td>
<td>0.172</td>
<td>0.17</td>
<td>0.17</td>
<td>0.158</td>
<td>0.172</td>
<td>0.172</td>
<td>0.19</td>
<td>0.186</td>
</tr>
<tr>
<td>8.009</td>
<td>0.17</td>
<td>0.17</td>
<td>0.167</td>
<td>0.167</td>
<td>0.158</td>
<td>0.171</td>
<td>0.173</td>
<td>0.192</td>
<td>0.184</td>
</tr>
<tr>
<td>8.305</td>
<td>0.167</td>
<td>0.167</td>
<td>0.164</td>
<td>0.164</td>
<td>0.158</td>
<td>0.169</td>
<td>0.174</td>
<td>0.194</td>
<td>0.182</td>
</tr>
<tr>
<td>8.602</td>
<td>0.164</td>
<td>0.164</td>
<td>0.161</td>
<td>0.161</td>
<td>0.158</td>
<td>0.168</td>
<td>0.175</td>
<td>0.196</td>
<td>0.179</td>
</tr>
<tr>
<td>8.899</td>
<td>0.161</td>
<td>0.161</td>
<td>0.161</td>
<td>0.161</td>
<td>0.158</td>
<td>0.168</td>
<td>0.175</td>
<td>0.196</td>
<td>0.179</td>
</tr>
<tr>
<td>9.195</td>
<td>0.161</td>
<td>0.161</td>
<td>0.161</td>
<td>0.161</td>
<td>0.158</td>
<td>0.168</td>
<td>0.175</td>
<td>0.196</td>
<td>0.179</td>
</tr>
<tr>
<td>9.492</td>
<td>0.161</td>
<td>0.161</td>
<td>0.161</td>
<td>0.161</td>
<td>0.158</td>
<td>0.168</td>
<td>0.175</td>
<td>0.196</td>
<td>0.179</td>
</tr>
<tr>
<td>9.788</td>
<td>0.161</td>
<td>0.161</td>
<td>0.161</td>
<td>0.161</td>
<td>0.158</td>
<td>0.168</td>
<td>0.175</td>
<td>0.196</td>
<td>0.179</td>
</tr>
<tr>
<td>10.085</td>
<td>0.161</td>
<td>0.161</td>
<td>0.161</td>
<td>0.161</td>
<td>0.158</td>
<td>0.168</td>
<td>0.175</td>
<td>0.196</td>
<td>0.179</td>
</tr>
</tbody>
</table>

Table 23: Lookup Table for Thrust Deduction Factor for Vessel 1
Open Water Efficiency Propeller Curve for Vessel 1

Figure 53: $\eta_O$ Propeller Open Water Efficiency curve for Propeller of Vessel 1, Generated as Standard Wageningen B-Series Propeller

Inputs to generate RAO to calculate the Added Resistance in Waves$^{15}$

<table>
<thead>
<tr>
<th>Data Used</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>[m]</td>
</tr>
<tr>
<td>Breadth</td>
<td>[m]</td>
</tr>
<tr>
<td>Draft</td>
<td>[deg]</td>
</tr>
<tr>
<td>Block Coefficient</td>
<td></td>
</tr>
<tr>
<td>Wave Direction</td>
<td>[deg]</td>
</tr>
<tr>
<td>Ship Speed</td>
<td>[m/s]</td>
</tr>
</tbody>
</table>

Table 24: Inputs to generate RAO to calculate the Added Resistance in Waves$^{15}$
## Tabulated Values of MAE and $R^2$ for Vessel 1 - Linear Regression

<table>
<thead>
<tr>
<th></th>
<th>1st Model</th>
<th></th>
<th>2nd Model</th>
<th></th>
<th>3rd Model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Universal Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Roughness $[\mu m]$</td>
<td>571.64</td>
<td>0.58</td>
<td>629</td>
<td>0.64</td>
<td>628.15</td>
<td>0.64</td>
</tr>
<tr>
<td>Fouling Resistance $[kN]$</td>
<td>49.89</td>
<td>0.73</td>
<td>43.71</td>
<td>0.77</td>
<td>44.16</td>
<td>0.77</td>
</tr>
<tr>
<td><strong>Separated Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Roughness $[\mu m]$</td>
<td>397.41</td>
<td>0.74</td>
<td>429.17</td>
<td>0.75</td>
<td>428.63</td>
<td>0.75</td>
</tr>
<tr>
<td>Fouling Resistance $[kN]$</td>
<td>37.93</td>
<td>0.82</td>
<td>35.23</td>
<td>0.83</td>
<td>35.13</td>
<td>0.84</td>
</tr>
<tr>
<td><strong>Chosen Data Range</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(From Period 1 to 2 mid)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Roughness $[\mu m]$</td>
<td>496.43</td>
<td>0.67</td>
<td>640.89</td>
<td>0.70</td>
<td>605.14</td>
<td>0.71</td>
</tr>
<tr>
<td>Fouling Resistance $[kN]$</td>
<td>44.52</td>
<td>0.74</td>
<td>41.52</td>
<td>0.76</td>
<td>39.87</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Table 25: *Tabulated Values of MAE and $R^2$ for Vessel 1 - Linear Regression*
Tabulated Values of MAE and $R^2$ for Vessel 1 - LSTM

<table>
<thead>
<tr>
<th>LSTM - Fit and Error - For Vessel 1</th>
<th>MAE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1st Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universal Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Roughness [$\mu m$]</td>
<td>480.95</td>
<td>0.50</td>
</tr>
<tr>
<td>Fouling Resistance [$kN$]</td>
<td>43.98</td>
<td>0.81</td>
</tr>
<tr>
<td><strong>2nd Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universal Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Roughness [$\mu m$]</td>
<td>421.74</td>
<td>0.64</td>
</tr>
<tr>
<td>Fouling Resistance [$kN$]</td>
<td>39.26</td>
<td>0.85</td>
</tr>
<tr>
<td>From Period 2-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Roughness [$\mu m$]</td>
<td>478.66</td>
<td>0.60</td>
</tr>
<tr>
<td>Fouling Resistance [$kN$]</td>
<td>43.52</td>
<td>0.74</td>
</tr>
<tr>
<td>From Period 3-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Roughness [$\mu m$]</td>
<td>587.15</td>
<td>0.32</td>
</tr>
<tr>
<td>Fouling Resistance [$kN$]</td>
<td>64.31</td>
<td>0.47</td>
</tr>
<tr>
<td><strong>3rd Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universal Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Roughness [$\mu m$]</td>
<td>374.95</td>
<td>0.70</td>
</tr>
<tr>
<td>Fouling Resistance [$kN$]</td>
<td>35.01</td>
<td>0.87</td>
</tr>
<tr>
<td>From Period 2-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Roughness [$\mu m$]</td>
<td>409.53</td>
<td>0.65</td>
</tr>
<tr>
<td>Fouling Resistance [$kN$]</td>
<td>36.34</td>
<td>0.82</td>
</tr>
<tr>
<td>From Period 3-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Roughness [$\mu m$]</td>
<td>445.66</td>
<td>0.54</td>
</tr>
<tr>
<td>Fouling Resistance [$kN$]</td>
<td>38.59</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Table 26: Tabulated Values of MAE and $R^2$ for Vessel 1 - LSTM