



UNIVERSITY OF GENOVA
MASTER THESIS IN COMPUTER ENGINEERING

Mechanically and Biologically informed and explainable Machine Learning for Biofouling Growth

by
Edaordo Pozzi

MASTER DEGREE IN COMPUTER ENGINEERING ARTIFICIAL INTELLIGENCE
AND HUMAN-CENTERED COMPUTING

December 2024

Luca Oneto	Supervisor
Andrea Cordaddu	Supervisor
Harm Jan Khamphof	Company supervisor
Thomas Van Terwisga	Company supervisor

Dibris

Department of Informatics, Bioengineering, Robotics and Systems Engineering

I would like to dedicate this thesis to my family, girlfriend and friends that supported me through this journey, without the help of each one of you it would have been impossible to make.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Your name
November 2024

Acknowledgements

I would like to express my gratitude to all those who provided support, guidance, and encouragement throughout this journey. This work would not have been possible without the assistance of mentors, colleagues, family, friends, and institutions that contributed resources and expertise.

Abstract

In 2023, the global fleet consists of 118,928 merchant vessels, each carrying an unwanted and non-paying passenger that has been causing inefficiencies since the 4th millennium BCE, when the first ships sailed the seas: biofouling. Biofouling in the maritime industry refers to the unwanted growth of flora and fauna on the wet surfaces of ships. It poses challenges related to component degradation, changes in hydrodynamic resistance, and environmental issues such as increased greenhouse gas emissions and the invasion of alien species. The shipping industry accounts for 90% of all goods transported worldwide and is known for its efficiency and reliability. However, it is also responsible for 3% of global GHG emissions, and this share is expected to rise due to the increasing number of ships operating each year. There is a logarithmic increase in emissions percentage with rising biofouling rates. Estimates suggest that fouling can impact fuel consumption and GHG emissions by up to 55%, even with relatively small percentages of hull coverage. The International Maritime Organization aims to address this with a new directive focused on reducing the impact of biofouling, but estimating its growth remains a challenging task due to the high amount of different variables involved. This project aims to develop two machine-learning-based models to study the growth and speed loss due to biofouling on ships. The proposed models are:

- A machine-learning model based on instantaneous endogenous variables such as shaft power, torque, propeller RPM, and exogenous variables like wind speed and direction to predict speed and calculate speed loss.
- A machine-learning model based on exogenous data derived from the ship's operational profile, including features such as the distribution of chlorophyll in water, sea surface temperature, speed, and the amount of time the ship spent stationary, to predict the increase of the speed loss.

These two models tackle the biofouling problem from different perspectives and time horizons.

The one based on mechanical variables works with instantaneous values and provides an estimation of the speed loss at that exact moment. This is vital information for ship owners, as they can derive the increase in fuel consumption and, consequently, the increase in operational costs for each ship at any given time.

In contrast, the model based on biological variables operates over a longer period. It takes into account all the variables and their distributions that influence growth. As stated later in this report, these variables follow seasonal patterns and are thus predictable. Information about the operational zone and profile of a ship can be fed into the model to estimate the speed loss after a certain amount of time.

Based on the previous statements, we learned that if a model is trained on a ship right after a cleaning event, the speed loss can be calculated not only for that ship but also for other ships of similar size and shape, providing insights into every ship a company operates. Additionally, if the future operational profile of a ship is known, the increase in speed loss can be predicted using the other model. This enables the ship owner to forecast the increase in costs and optimize maintenance and cleaning schedules in advance.

For the first part of this project, a dataset from MARIN will be explored and analysed to identify its potential and optimize its use. For the second part, data from the Copernicus project, funded by the EU, will be used to create a dataset composed of biological variables, aiming to analyse the problem from a different perspective. The final section will provide an in-depth analysis of the results from both models and discuss potential future improvements.

Table of contents

List of figures	ix
List of tables	xi
Nomenclature	xii
I Section One	1
1 Introduction	2
1.1 General context	2
1.2 Biofouling overview	2
1.2.1 Growth stages	3
1.2.2 Growth variables	4
1.2.3 Water based variables	4
1.2.4 Ship variables	6
1.2.5 Environmental variables	6
1.2.6 Effects on ship hydrodynamic resistance	8
1.2.7 Effects on ship component	9
1.2.8 GHG emission	9
1.2.9 Alien species invasion	10
1.2.10 Antifouling techniques	10
1.3 What has been done	13
1.4 Proposed solution	13
2 State of the art	15
2.0.1 Previous on biofouling state prediction	16
2.0.2 Biofouling growth models	17

3 Available data	18
3.1 Context	18
3.1.1 Ship description	18
3.1.2 Ship operational profile	19
3.2 Data acquisition	20
3.2.1 Speed sensors	21
3.2.2 Anemometer	22
3.2.3 Shaft sensors	22
3.2.4 GPS	23
3.3 Power trend identification	23
3.4 Adopted simplifications	25
3.4.1 Speed sensors	25
3.4.2 Speed range	26
3.4.3 Years considered	28
3.4.4 Dataset averaging	29
3.5 Dataset enrichment with ERA-5 database	29
3.5.1 Sea surface temperature	29
3.5.2 Chlorophyll concentration	31
3.6 Dataset creation for the biological model	34
3.6.1 MARIN, SST and chlorophyll merging process	34
3.6.2 Imputation of missing values	35
3.6.3 Final biological dataset	35
4 Proposed methodology	36
4.1 Speed prediction based on mechanical data	36
4.1.1 Features and correlation	36
4.1.2 Data preparation	37
4.1.3 Models test	40
4.2 Speed-loss prediction based on biological data	44
4.2.1 Feature correlation	44
4.2.2 Data preparation	48
4.2.3 Models test	50
5 Results	51
5.1 Mechanical model	51
5.1.1 Speed-loss identification	52

Table of contents	viii
5.2 Biological model	54
6 Future development	57
6.1 Discussions	57
6.2 Suggestions for improvement	57
7 Conclusions	59
References	60

List of figures

1.1	Ship hull with heavy presence of slime and weeds	3
1.2	Growth stages (MARIN internal report by Harm Jan Kamphof)	4
1.3	SST mean variation based on latitude (12)	5
1.4	Climate zones based on (12)	7
1.5	Typical resistance components of a tanker container ship and fishing vessel. The roughness component is for a ship without fouling (MARIN internal report)	8
1.6	Carbon dioxide estimated reductions per type of intervention (10)	9
1.7	GHG emission, IMO estimate	10
1.8	Antifouling techniques (MARIN internal report)	12
3.1	Ship of similar size and characteristics	18
3.2	Recorded location for various years of the dataset	19
3.3	Percentage of time the ship spent cruising each year	20
3.4	Schematic figure of the location of each sensors	23
3.5	Power values over 2014	25
3.6	Power values over 2014	25
3.7	Power values various years	26
3.8	Comparison of an entire day worth of recorded speeds for each speed water flow sensors	27
3.9	Doppler speed distribution over the years	28
3.10	Comparison between winter and summer SST in the two port where the ship docks.	30
3.11	Chlorophyll sample coverage comparison, each point is a recorded sample in a selected week	32
3.12	Comparison between winter and summer chlorophyll level in the two port where the ship docks.	33

4.1	Mechanical variable model blocks scheme	36
4.2	Input feature correlation map, warm colors for positive value, cold for the negative one	38
4.3	Number of sample per rounded speed in knots	39
4.4	Performance comparison with different metrics of each model in train and validation set	43
4.5	Mechanical variable model block	44
4.6	Mechanical variable model block	45
4.7	Pairwise correlation of all the input variables of the model	46
4.8	Pairwise correlation of the speed-loss value with every other variable	47
4.9	Visual representation of the folds generated during cross-validation (cro) . .	48
4.10	Number of sample per speed-loss value	49
4.11	Performance comparison with different metrics of each model in train and validation set	50
5.1	Error during validation, dividend in a scatter plot and a density plot	52
5.2	Comparison of scatter and density error plot for selected weeks in 2013 . .	53
5.3	Comparison of scatter and density error plot for selected weeks in 2014 . .	54
5.4	Weekly speed-loss predicted by the mechanical model	55
5.5	Comparison between the scatter plots of training and validation sets for the KRR	55
5.6	Mean weekly speed-loss comparison between the real and predicted one, the last 4 sample are predicted by the model after training and validation	56

List of tables

3.1	Speed columns description	22
3.2	Anemometer columns description	22
3.3	Shaft colums description	23
4.1	Distribution of samples before and after applying resampling techniques . .	38
4.2	Hyperparameters for Kernel Ridge	40
4.3	Hyperparameters for Lasso	40
4.4	Hyperparameters for SVR	41
4.5	Hyperparameters for XGBRegressor	41
4.6	Hyperparameters for RandomForestRegressor	41

Nomenclature

Acronyms / Abbreviations

CDF Computational fluid dynamics

CSV Comma separated values

DDM Data driven model

DF Pandas dataframe

GHG Green houses gases

HM Hybrid model

IMO International maritime organization

KRR Kernel ridge regressor

MAE Mean absolute error

MAPE Mean absolute percentage error

MARIN Maritime research institute Netherlands

ML Machine learning

MSE Mean squared error

NaN Not a Number

PM Physical model

SST Sea surface temperature

TEU Twenty-foot equivalent unit

Part I

Section One

Chapter 1

Introduction

1.1 General context

In 2023, the global fleet consists of 118,928 merchant vessels, each carrying an unwanted and non-paying passenger that has been causing inefficiencies since the 4th millennium BCE, when the first ships sailed the seas: biofouling. Biofouling in the maritime industry refers to the unwanted growth of flora and fauna on the wet surfaces of ships. It poses challenges related to component degradation, changes in hydrodynamic resistance, and environmental issues such as increased greenhouse gas emissions and the invasion of alien species. The shipping industry accounts for 90% of all goods transported worldwide and is known for its efficiency and reliability. However, it is also responsible for 3% of global GHG emissions, and this share is expected to rise due to the increasing number of ships operating each year. There is a logarithmic increase in emissions percentage with rising biofouling rates. Estimates suggest that fouling can impact fuel consumption and GHG emissions by up to 55%, even with relatively small percentages of hull coverage. The International Maritime Organization aims to address this with a new directive focused on reducing the impact of biofouling, but estimating its growth remains a challenging task due to the high amount of different variables involved. The International Maritime Organization (IMO) aims to address this issue with a new directive focused on reducing inefficiencies caused by biofouling(IMO).

1.2 Biofouling overview

Biofouling consist of an undesirable accumulation and growth of organic matter, such as weeds and small animals, in water immersed structures, the term fouling includes also the

accumulation of inorganic matter. The most common fouling organisms are algae, sea squirts, barnacles, bryozoans, hydroids, mussels, and tubeworms (38). These organisms are placed into two categories based purely on the size that the fouling organism can reach when is fully developed: micro fouling, consisting of bacteria and diatoms that create the a slime layer and macro fouling, that include weeds, algae, barnacle, and mussels.

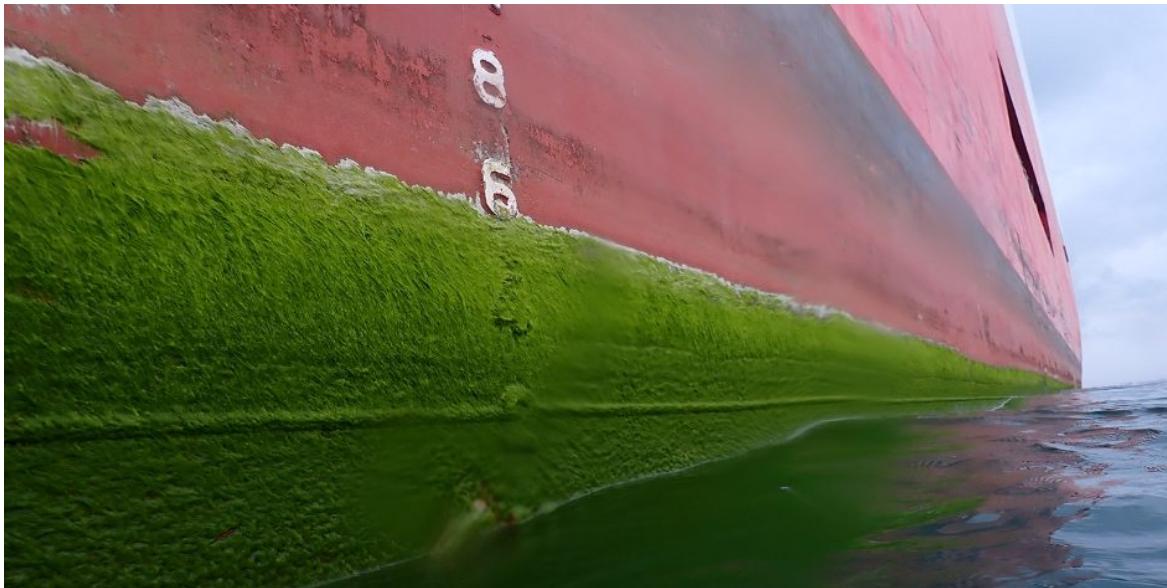


Figure 1.1 Ship hull with heavy presence of slime and weeds

1.2.1 Growth stages

As soon as an object is submerged in water the fouling process starts. It is shown that, within the first minutes, a thin organic film attaches to the surface and completely covers it in a matter of hours, making the surface wettable (24; 29). Wettability creates conditions that allow bacteria, the primary colonizers, to settle. They are attracted by physical forces into the film, and when in contact with it, a chemical bond begins to form. In this initial phase, the bond between the bacteria and the film is still loose, and bacteria can be removed simply by the water flowing over them. Along with bacteria, diatoms¹ begin to attach to the film and grow, creating a symbiotic commensal relationship² that causes both to form and grow the slime layer (29). Growth rate, structure and morphology depends on the external condition provided by the environment, such as nutrient availability, temperature and water

¹a group of micro algae

²Commensalism is a type of symbiotic relationship between different species of organisms in which one species benefits from the relationship while the other species is neither harmed nor benefits from the relationship.

flow velocity. The slime layer serves as the colonizing ground for secondary colonizers: protozoa and algae spores that begin to settle within the first week. After 2-3 weeks, the secondary colonizers have grown enough to allow larvae of major macro foulers to attach. While this formulation represents the ideal scenario, in real-world conditions, even if the general order is mostly followed, not all parts of the surface will be affected in the same way, even under identical conditions (2).

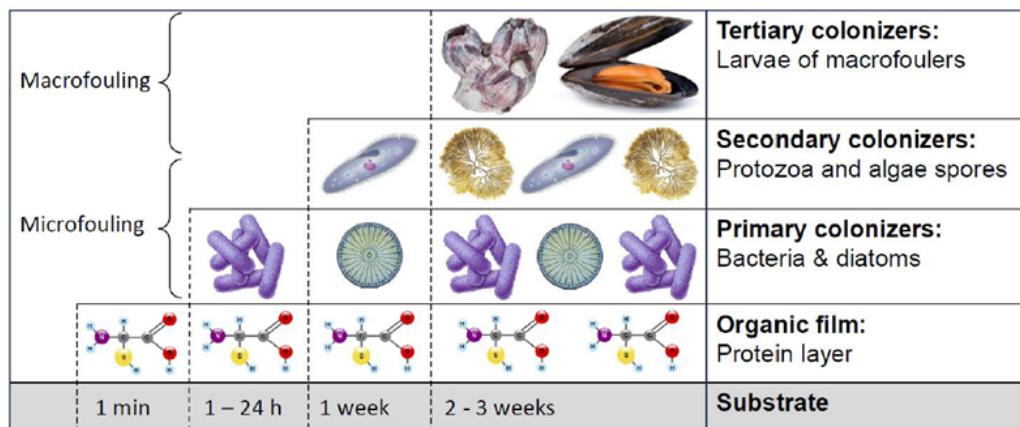


Figure 1.2 Growth stages (MARIN internal report by Harm Jan Kamphof)

1.2.2 Growth variables

Globally, there are more than 66 different marine ecosystems , each with its own characteristics (GEF). Each fouling species has its own shape, hardness, growth rate, and tolerance to changes in water factors, making it difficult, if not impossible, to generalize a precise growth rate on a global scale.

1.2.3 Water based variables

In this section all the variable that influence fouling that are related to the sea water are analyzed.

Sea surface temperature

Sea surface temperature (SST) is a key factor in quantifying biofouling growth. In some cases, such as in (11), it is the only variable considered. SST primarily varies based on latitude and season, which makes it relatively easy to classify the world into different climate

zones. When ships travel from one region to another, changes in temperature can have two outcomes: either killing the existing growth if the temperature exceeds the maximum tolerance of the species, or enhancing growth if the temperature gradually increases and stays in the optimal range for the species. The first case occurs with a strong and rapid increase or decrease in SST, while the latter occurs with a gradual increase(40).

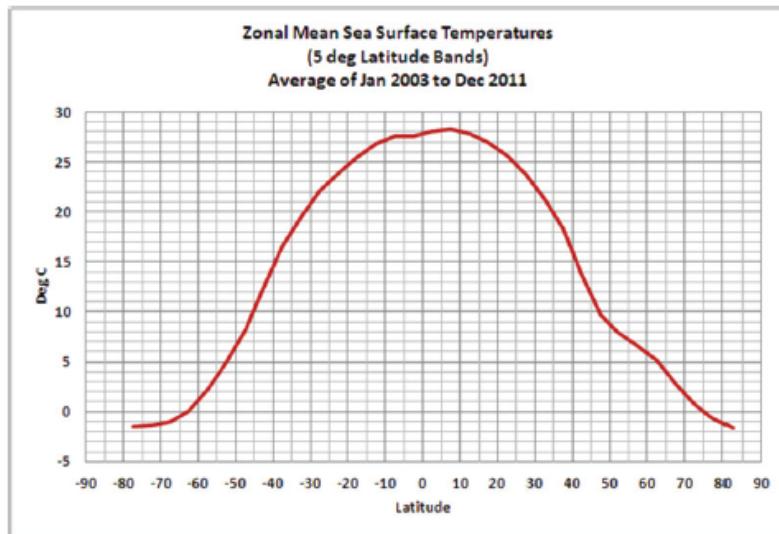


Figure 1.3 SST mean variation based on latitude (12)

Nutrient density

Nutrient density refers to the concentration of chemical particles needed by biofouling species to live and grow, dissolved in water. This amount heavily influences growth, as a higher nutrient density can support a greater number of species. The most important nutrients are nitrogen, phosphorus, cobalt, and iron (23; 31). The presence of these substances depends on several factors, such as proximity to the shore, the presence of rivers that flow into the sea carrying both natural and human-made fertilizers, and the geographic region (11).

Various

Conductivity, pH, and salinity can be considered secondary variables (11). Their presence influences the types of fouling species that grow more than the growth itself. Absolute differences in their concentrations and values across different parts of the world are also negligible (39) (32).

1.2.4 Ship variables

In this section all the variable that influence fouling that are related to the ship are analyzed.

Ship speed

Ship speed determines the amount of water that flows in contact with the ship hull, which influences which species can grow and how (29). For example, in the case of the slime layer, the latest growth is not yet strongly attached and can detach with a strong enough force. If the operational profile³ consists of numerous short trips, it causes repeated detachment over time. As a result, the slime growth is slower but the consistency is thicker. The same principle applies to macrofouling species: the larger they are, the greater the resistance they create, and thus the greater the impact on the ship (34).

Antifouling coating

Antifouling coatings—along with their quality, quantity, and characteristics—make a significant difference. Higher-quality, long-lasting paint will perform better than lower-end options. More details on this topic can be found in section 1.2.10 where a more detailed explanation is provided.

Surface disposition

Surface disposition is important as the shape of a ship's hull influences where and how fouling grows (11). A flat hull will receive less light compared to a V-shaped hull and, since algae and diatoms need light to survive and thrive, light attenuation in water is a factor that must be considered as it affects all species (20; 31).

Various

Hull roughness, hull conductivity, surface color, and energy are minor factors that are often mentioned but rarely taken into consideration. This also applies to this work (11).

1.2.5 Environmental variables

In this section all the variable that influence fouling that are related to the environment are analyzed.

³series of parameters and characteristics that define typical operations, including how it is used, where it is used, the conditions under which it operates, and its performance metrics

Location

Ship location determines where the ship is positioned. Based on this information, all the other previously mentioned variables can be derived . For example, knowing that the ship was stationary and located in a tropical zone provides a better insight of which species can grow and how fast, giving a better understanding of the entire phenomenon(11).

Seasons

Seasons play a role for marine species just as they do for their terrestrial counterparts, affecting their growth and reproductive cycles. For ships that remain in the same region, this factor must be considered. During reproductive periods, more spores and eggs will be present in the water, accelerating the fouling attachment process, compared to dormant periods (11; 27).

Weather

Weather influences the amount of light exposure that fouling species receive, thus accelerating or decreasing their growth rate. Additionally, foul removal can occur due to inclement weather, such as high waves slamming against the side of the ship, or even more severely, the hull of the ship slamming into the water if the waves approach from the bow. Also, as already said, the climate zone in which the ship is located greatly influences fouling.

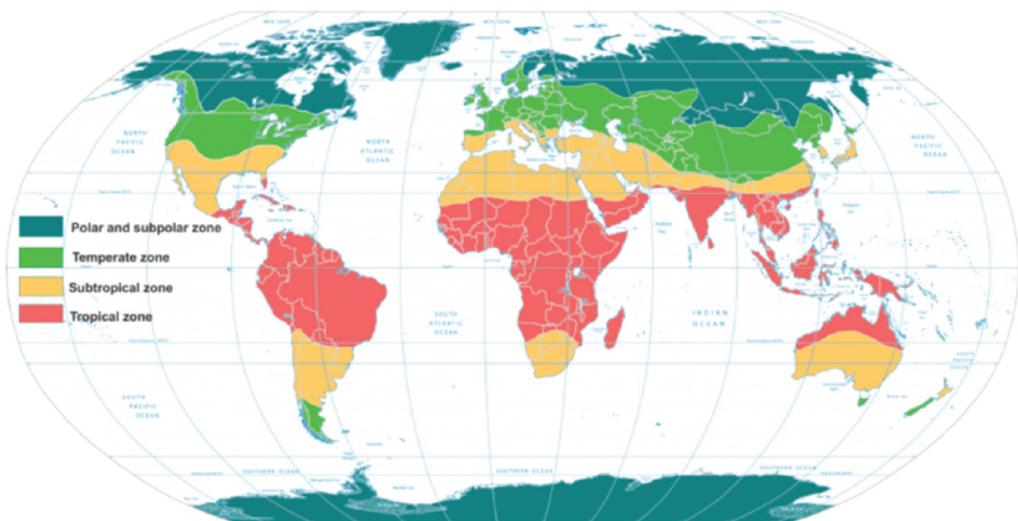


Figure 1.4 Climate zones based on (12)

1.2.6 Effects on ship hydrodynamic resistance

Every object moving through a liquid encounters resistance, which is the opposition offered by a fluidic element to the flow itself. In hydrodynamics, resistance can be divided into two components: viscous resistance and wave resistance. Wave resistance is caused by the energy required to push the liquid away from the hull, while viscous resistance is due to the viscosity of the fluid. Biofouling impacts the latter (35); the roughness of the hull increases fouling growth, leading to an increase in drag (18; Kane). The increase in resistance is non-linear: slime is porous, seaweeds are swayed by the water flow, and barnacles are more static. Each of these types of fouling has a different effect, and the total impact is generally obtained by combining their individual effects. Schultz et al.(35) reported that the presence of slime alone required a 21% increase in shaft power, compared to an otherwise identical slime-free frigate, whereas heavy calcareous fouling led to an 86% increase in shaft power requirements at 15 knots. These data can vary depending on the type of ship, Demirel et al. (13) showed an increase in hull resistance of about 25% for light slime and about 175% for heavy calcareous fouling. Biofouling grows differently in various parts of the ship. Where the water flow is faster, fouling growth is less and slower, in contrast to areas with slower flow, creating a sort of asymmetry in the hull (Kane). This further complicates research efforts.

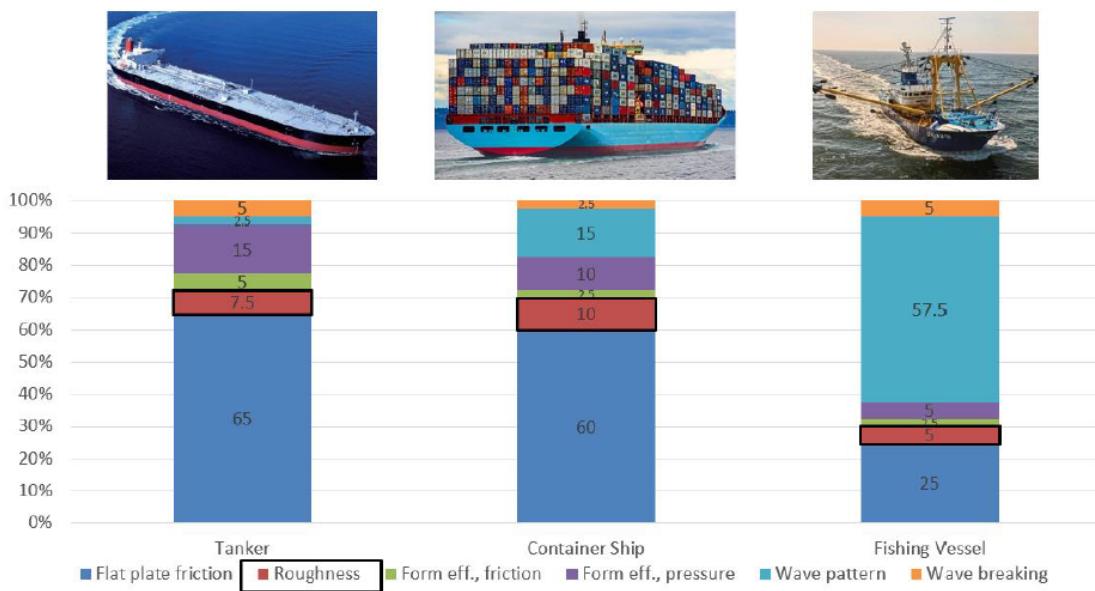


Figure 1.5 Typical resistance components of a tanker container ship and fishing vessel. The roughness component is for a ship without fouling (MARIN internal report)

1.2.7 Effects on ship component

Ships are designed based on their operational profile, with operational speed being a key factor around which every component is optimized (4). The gradual increase in viscous resistance due to biofouling formation reduces the ship's speed if the shaft torque remains constant. To maintain the specified cruise speed, the power must be increased to counteract the added resistance. This results in greater stress on each component and faster deterioration. An increase of 21% in shaft power is required in the presence of slime alone (13; 35), and this requirement can rise to 175% in the case of heavy calcareous fouling.

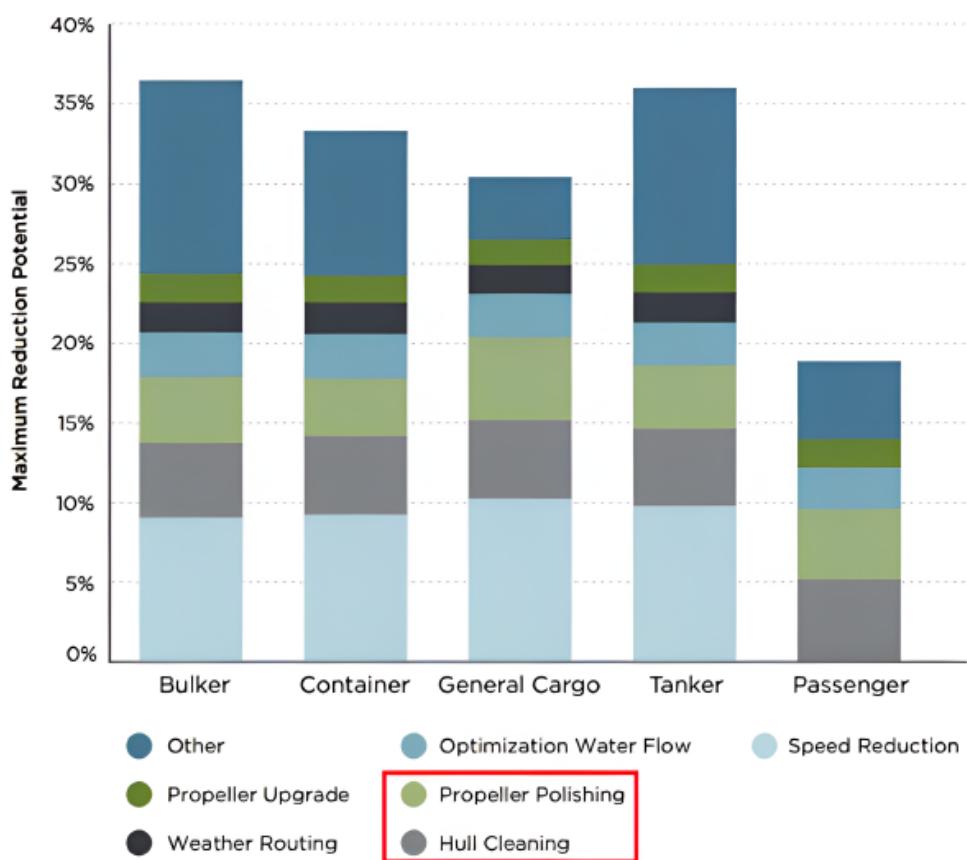


Figure 1.6 Carbon dioxide estimated reductions per type of intervention (10)

1.2.8 GHG emission

The maritime sector is responsible for 3% of GHG emissions, and this share is increasing due to the growing number of ships that travel each year. In an already efficient sector, making

improvements is a challenging task. The International Maritime Organization aims to address this issue with a new directive focused on reducing inefficiencies caused by biofouling(IMO).

As shown in the figure 1.7, there is a logarithmic increase in emissions percentage with the rise in biofouling rate. Estimates suggest that fouling can impact fuel consumption and GHG emissions by up to 55%, even with a relatively small percentage of hull coverage (IMO).

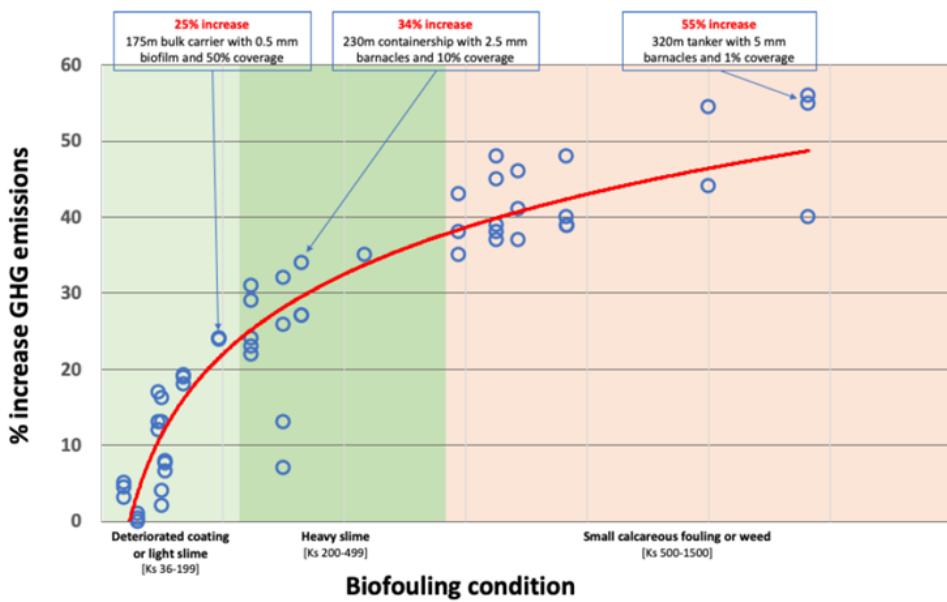


Figure 1.7 GHG emission, IMO estimate

1.2.9 Alien species invasion

More than four thousand species contribute to biofouling, each originating from specific maritime zones. Shipping is known to cause the translocation of non-native marine species and is the dominant global vector for unintentional translocation of these species, both in ballast tanks and attached to hulls as fouling organisms (14; 30; 42). One example is the invasion by the golden mussel from the Amazon River (33), which is currently threatening fragile ecosystems.

1.2.10 Antifouling techniques

The most effective technique for preventing the effects of biofouling on ships involves hull and propeller cleaning in a dry dock and re-coating all surfaces. However, this method

is not the most economically viable due to the high cost and time required for cleaning (37). Shipping companies aim for an optimized strategy that balances the increased costs associated with biofouling and the expenses of removing it. Antifouling coating technology has evolved over the years, with various options available, divided into three categories: biocidal, foul-release, and passive (43).

Biocidal paints aim to create a toxic layer on the hull to kill fouling organisms. Tributyltin Self-Polishing Copolymers (TBT-SPC) were once considered state-of-the-art antifouling paints. They worked by releasing small quantities of a highly toxic compound to prevent growth. However, TBT-SPC was banned in 2008 due to its harmful effects on non-fouling marine life (22). Today, tin and copper-based self-polishing coatings are used, as their local toxicity is sufficient to kill or limit the growth of organisms on the ship without causing excessive environmental damage.

Non-biocidal paints, on the other hand, aim to create a surface that fouling organisms cannot adhere to, or if they do adhere, the bond is not strong enough to withstand the force created by the water flow when the ship is sailing (28).

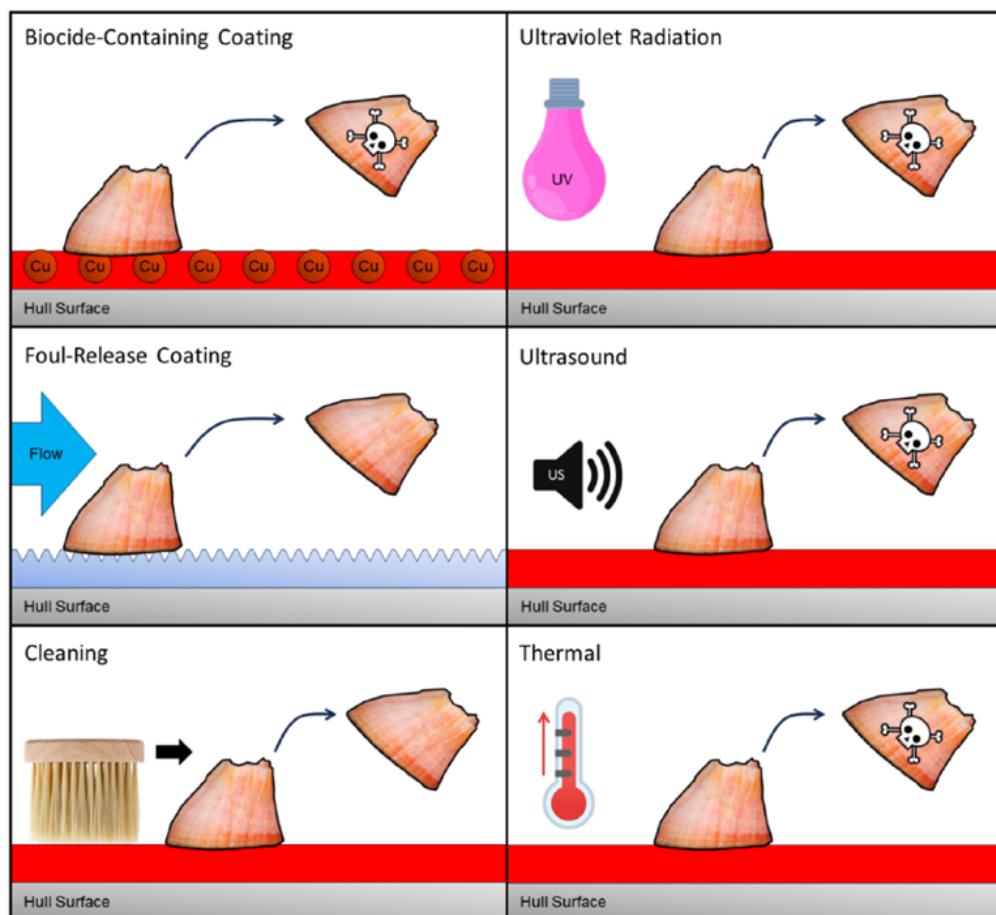


Figure 1.8 Antifouling techniques (MARIN internal report)

1.3 What has been done

Machine learning techniques have already made accomplishment in the maritime sectors, including efforts to address the growth of fouling. The main indicator of a possible accumulation of fouling species on the wet surface are:

- **Speed-loss** defined as the difference in speed between the ship with a clean hull and the same ship after a defined time period in which fouling species grow on the hull
- **Increase in fuel consumption** due to an increase in the hydrodynamic resistance
- **Increase in power output** required to keep the same speed in similar condition

Various ML models have been developed to predict the fouling state based on ship endogenous variable, such the one previously mentioned. Other models instead tried to predict the added resistance value has on the image of the fouling attached to the hull.

1.4 Proposed solution

Speed-loss is defined as the difference between the speed of a ship with a clean wet surface and the one with the presence of fouling, in equal weather conditions and engine power. The general ML approach to identify this phenomenon involves training a regression model on data collected immediately after a cleaning event. The model learns a function where the inputs include the ship's endogenous variables —such as shaft power, fuel consumption, and propeller RPM— under various exogenous conditions like weather. The output of the model is the ship's speed (7; 9; 41).

Subsequently, the model is tested with a new set of data. The hydrodynamic resistance is lower immediately after cleaning, and, since it can only increase over time, under similar weather and engine conditions, the predicted speed will be higher than the actual measurements due to the accumulated fouling. The discrepancy between the model's predictions and the real values from test samples is the speed-loss. If component degradation can be disregarded —typically valid if the test period is relatively close to the training period- the observed speed loss in terms of knots can be attributed mostly to the increase in resistance due to fouling. As said, several factors influence a ship's speed, including draught, component degradation, wind profile, cargo weight and positioning on the deck. While these factors cannot be entirely ignored, with sufficient data, the model can account for these variables as part of its training.

Even in case of absence of information regarding cleaning events, the model can still be developed. The training data will be treated as representing the cleaned state, and the trend of increasing error will indicate the probable accumulation of fouling. If the trend levels off or if the difference between predicted and actual values becomes negative, it signifies that a cleaning event has occurred, and the predicted speed is now lower than the actual speed.

All the above considerations are valid when the available data are instantaneous measurement. In such cases, predictions can be made for each sample and averaged over the desired time period to obtain the results.

However, this approach does not account for the biological aspect of fouling growth, excluding all the variables mentioned in section 1.2.2 since, if taken as instantaneous measurement, they do not provide meaningful insights into the problem. A water temperature of 27°C does not directly influence the ship's speed, but it affects the growth rate of fouling. Since fouling growth is a gradual process, it must be characterized over a period where changes in speed become noticeable. This time discrepancy results in missing significant information that could be valuable for a comprehensive analysis.

The idea to overcome the time discrepancy problem is as follows: obtain the distribution of variables influencing growth rate from the time when the ship was cleaned. By using the model that depends on mechanical variables, the average speed-loss over a week can be calculated. This value can then be associated with the distribution of all biological variables from the time the hull was cleaned until the decided period. With sufficient data, a model can be created that takes biological variables as input and outputs the speed loss, and that will be the main focus of this work since has not been done before.

Chapter 2

State of the art

Machine learning techniques have already been used in the maritime sector, and the biofouling problem is no exception. As highlighted in (41), three different approaches can be identified, categorized according to the numerical models used in existing literature: Physical Models (PMs), Data-Driven Models (DDMs), and Hybrid Models (HMs).

PMs are developed by experts in the field and require in-depth mechanistic knowledge of the real system to accurately characterize the phenomenon being simulated (5; 26). The main advantage of this methodology is that, when implemented correctly, these models offer a high degree of reliability since they are based on the physical laws governing real-life phenomena, thus always generating physically plausible predictions. However, this type of model comes with significant computational demands, with simulations that can take days or even weeks depending on the complexity and desired level of accuracy, making them difficult to use without adequate infrastructure.

Data-Driven Models (DDMs) are machine learning techniques built on existing datasets, created by recording observations of a phenomenon—such as a vessel during normal operational activities. These datasets are then used to train machine learning models to predict behavior. This means that deep knowledge of the physical system is not required, making these models accessible even to non-experts.

To effectively generalize a complex system like a cruising ship, where many variables influence its behavior (7), a large amount of data is required to make accurate predictions (8). The computational resources needed vary depending on the model and the size of the dataset used for training. This is one of the main disadvantages of this approach. However, once the model is trained, making predictions becomes computationally inexpensive and can be done quickly.

The major disadvantage, though, is that these models perform well statistically (on average, with good standard deviation) only if the dataset adequately characterizes and generalizes the system. If it fails to do so, particularly in unique cases, the model may produce physically implausible predictions (3).

Hybrid Models (HMs) aim to combine the best aspects of both PMs and DDMs to overcome the limitations of each. The idea is to leverage the physical knowledge from PMs to prevent implausible predictions, while also reducing computational demands by incorporating historical data that partially characterizes the phenomenon. This results in models capable of auto correcting their predictions if they deviate from plausible outcomes.

2.0.1 Previous on biofouling state prediction

For this part will be focused on DDMs and how they are build. To estimate the impact of biofouling on the ship different parameters can be used, the most chosen one are:

- Fuel consumption: in similar weather and engine conditions an increase can be attribute to the increase in hydrodynamic resistance due to biofouling
- Propulsion power: engine power, shaft torque and engine RPM can be analyzed to catch possible increasing trend
- Speed through water: again, if in similar weather and engine conditions the speed of the ship is less than before the decrease can be attribute to fouling growing on the ship hull and propeller.

That said, discerning if a change in one of the variable is due to biofouling or other factors, like human decision, component degradation, different operational profile is itself a challenge. A combination of the above mentioned variables is usually used to determine the cause of the shift.

In this paper (7), Coraddu et al. aim to predict the condition of the hull and propeller based on operational data collected from onboard sensors. They demonstrate that it is possible to use unsupervised machine learning techniques, which do not require labeled data, to achieve a reliable estimate of biofouling growth. Sea trials were conducted to assess the ship's condition in varying sea states and weather conditions to mitigate the disturbances these factors introduce. A total of 41 variables were recorded, with a focus on the ship's power plant and its components.

In this paper (6), the problem of performing Condition-Based Maintenance through the use of Data driven models is explored with a total of 25 variables used as input. In order

to be able to schedule a maintenance the future state of the component analyzed must be predicted, a bad prediction can cost millions of euros, if a component is replaced when it's still usable part of the cost of it is wasted, in contrast if it's not replaced it can break down during operation and more expensive maintenance is required.

In (41) a review of all the past and recent model is proposed. As described in the paper, a growth model based on environmental data that influence the species responsible for fouling does not exist. To estimate the fouling state operational data from the ship such as speed loss, increase in shaft power, additional fuel consumption and change in resistance are used. All these models are good in their estimate but have the limitation of a reference: the calm water resistance after dry docking with a clean hull. If those data are not available, the model cannot compute the difference.

Senteris et al. (36) used artificial neural network to estimate the shaft power required to maintain the same speed at any operational, environmental and loading state. Starting from the clean hull and propeller state, a comparison between the predicted values and the real one was made to model the amount of speed-loss after a given period. No more information regarding the chosen model or the final accuracy were given.

In the paper by Erol et al. (15) two datasets, collected 9 months apart, are used to predict the speed loss. The analysis was made using Curve Fitting and Detrended Fluctuation Analysis (DFA) to detect the impact of biofouling on the degradation of the performance of the ship under continuous monitoring. The final results showed that after 9 month a the average speed loss was around 6%.

2.0.2 Biofouling growth models

Numerous attempts have been made to predict biofouling growth and the amount that accumulates on a ship's wet surface. The two most notable and accurate models can be found in these papers (11; 12). The goal was to determine a growth rate based on biological variables (a more in-depth analysis will follow later) in order to predict the fouling state after a period of immersion and its subsequent increase in hydrodynamic resistance. The results were satisfying enough and modeled the actual growth although numerous simplification were made.

Chapter 3

Available data

3.1 Context

Data were acquired from the ship during normal operations based on the needs of the shipowner without the influence of MARIN. Speed sensors were installed on the hull in 2010 and a foul release coating was applied. Reports state that was also cleaned by divers at the end of 2012.

3.1.1 Ship description

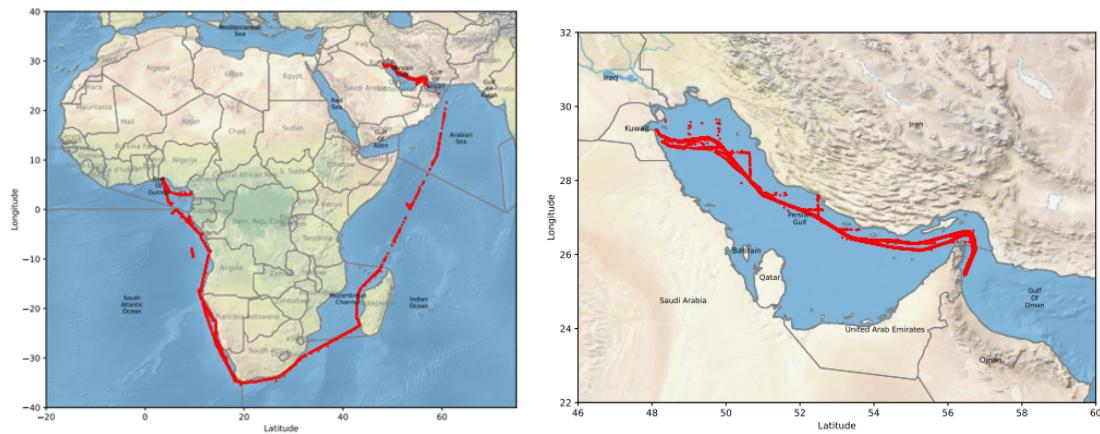
The ship under analysis is a small container ship, with a length overall of 180 meters and beam length of 28 meters, capable of carrying up to 1800 TEU with a dead weight at design draught of 20'000 tons.



Figure 3.1 Ship of similar size and characteristics

3.1.2 Ship operational profile

From 2010 to the end of 2011 the voyages were concentrated between Chinese sea and the Guinean gulf. From 2012 to first semester of 2013, as shown in figure 3.2a, the route extended from the Guinean gulf, in the Atlantic ocean, to the Persian gulf near the Arabian peninsula. It's a journey that spans for over 9'000 nautical miles, given a cruising speed ranging from 15 to 18 knots, it takes approximately 24 days with favourable weather conditions to complete the entire voyage. From the second semester of 2013 to the end of 2016, all the voyages were concentrated in the Persian gulf, from Kuwait to Oman and occasionally United Arab Emirates, for a total of 700 nautical miles. In this region the cruise speed was set between 13 and 14 knots, making a one-way trip last 2-3 days.



(a) Ship route from 2012 to first half of 2013 (b) Ship route from second half of 2013 to 12/2016

Figure 3.2 Recorded location for various years of the dataset

As previously mentioned in section 1.2.2 fouling grows at a faster rate when a ship is stationary, knowing the amount of time spent in that condition helps in generalizing a growth rate and discover, or validate, possible trends. As shown in figure 3.3, in 2010 and 2011 stationary time accounts only for one third of the total due to the longer distance that the ship has to travel. When the operational zone shifted to focus on the Persian gulf, the percentage increases, from 50% in 2014, up to 66% in 2015. This increment can be attributed to the shortest distance between ports that reduce the voyage to 2-3 days and to the amount of time required for loading and unloading operations that, for the ship under analysis is generally a couple of days.

Since the ship was coated with a foul release coating, speed plays a key role in the process of fouling management. It has been shown in (43) and in section 1.2.10 that the performances depend on the specific type and quality of the coating itself, but generally with a speed

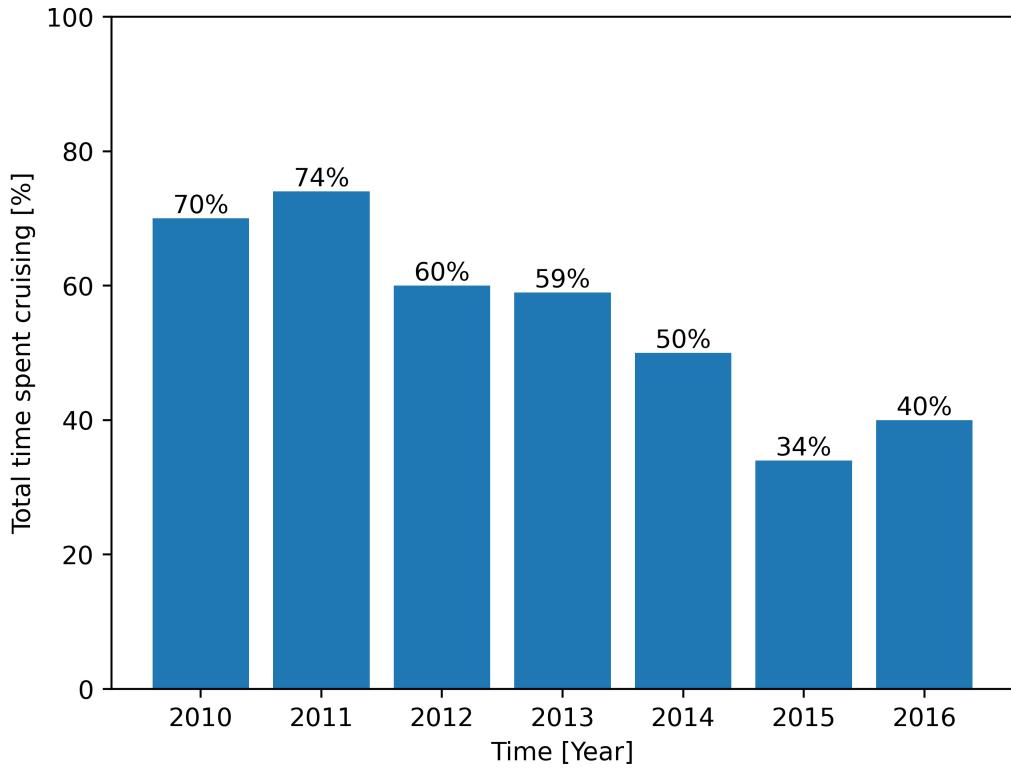


Figure 3.3 Percentage of time the ship spent cruising each year

higher than 14 knots the shear force caused by the water flow on the hull and propeller is sufficient to cause foulings detachment. This cleaning process is more efficient if the speed is maintained for a longer period, and at higher speed this works even better. In this study case, the applied coating is not the optimal choice, cruise speed rarely exceed the ideal range for fouling release and this particular type doesn't prevent the accumulation of fouling during stationary period, that accounts for half of the total time, in contrast with biocide-containing coating.

3.2 Data acquisition

A total of 7 years worth of data were collected from 2010 to 2017, during that time various improvements were made to the recording system leading to a heterogeneous dataset, from 21 recorded variables in 2011 up to 125 starting from 2013 and after. For each measurement

the sampling rate was set to 1 Hz, capturing a sample each second for a total of 86400 of records each day, creating a total of 2.5 million records that require 60 GB of disk space.

3.2.1 Speed sensors

Speed over ground

Speed Over Ground (SOG) is calculated using the GPS position, measuring the time required to move between two locations along with the distance traveled. This speed reflects the actual speed of the ship, as it accounts for environmental factors such as wind, currents, and waves, which can affect how the ship moves through the water since in relation to a fixed point on the Earth's surface.

Doppler speed

Doppler speed is measured by analyzing the speed and direction of water currents relative to the ship's motion and is placed below the bow. This technique involves transmitting an ultrasonic signal from the ship's hull into the water and measuring the frequency shift of the signal as it bounces off particles in the water. It is possible for the Doppler speed to register a certain speed corresponding to the current speed even if the ship is anchored and has a speed over ground of zero, due to the effect of the water current.

Water flow sensors

Five sensors were installed on different positions of the ship's hull to measure water flow. Two sensors were placed on each side of the hull at different heights: one at a lower position to avoid turbulence caused by drift and one at a higher position to account for wave height. The remaining sensor is positioned at the bow of the ship. It is worth noting that each water flow sensor operates based on pressure and is placed within the boundary layer¹. If turbulence caused by drift occurs near the sensor, it can lead to chaotic pressure levels, resulting in highly imprecise measurements that are unsuitable for analysis, more in section 3.4.1.

¹The boundary layer is the thin layer of fluid in the immediate vicinity of a bounding surface, formed by the fluid flowing along the surface.

Table 3.1 Speed columns description

	Type	Measurement unit
Speed over ground	Float	Knots
Doppler speed	Float	Knots
Water flow sensors 1-5	Float	Knots

3.2.2 Anemometer

An anemometer is a device used to measure wind direction, quantified in degrees with zero corresponding to the ship's heading, and wind speed in knots. Aerodynamics are not a primary focus during the design phase of a ship due to their relatively low impact on total resistance compared to other metrics, and the high variability in aerodynamic profiles caused by the different positioning of containers on the deck, which creates a unique aerodynamic profile for each trip. The precision of an anemometer's measurements depends on its installation position, and often more than one anemometer is required for accurate readings. In this case, the anemometer was installed on top of the bridge that is the best possible position and can be considered precise.

Table 3.2 Anemometer columns description

	Type	Measurement unit
Mean wind speed	Float	Knots
Mean wind direction	Float	Degrees

3.2.3 Shaft sensors

Three measurement were recorded to accurately characterize the engine load: shaft power, torque and propeller revolution per minute (RPM). Torque refers to the measure of the rotational force exerted by the engine. Shaft power, on the other hand, is the rate at which work is done by the engine or the rate at which energy is converted from one form to another and represents the total power output of the engine. Revolutions per minute (RPM) measures the number of complete rotations the engine's shaft, and thus the propeller makes in one minute.

Table 3.3 Shaft columns description

	Type	Measurement unit
Shaft power	Float	Kilowatt
Shaft torque	Float	kNm
Shaft RPM	Float	RPM

3.2.4 GPS

The Global Positioning System (GPS) is a satellite-based navigation system that provides location and time information anywhere on Earth. It used to calculate the speed over ground and to get the precise position of the ship in terms of latitude and longitude.

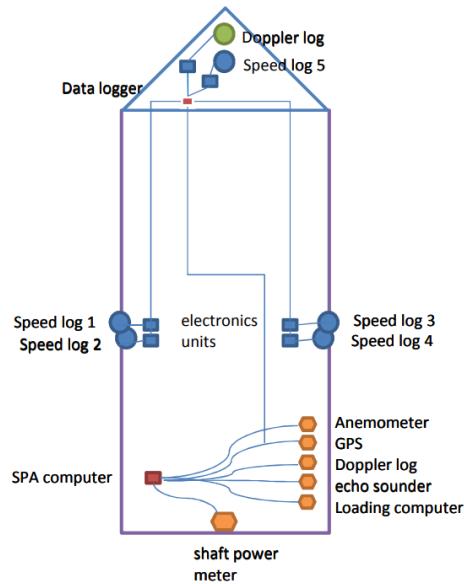


Figure 3.4 Schematic figure of the location of each sensors

3.3 Power trend identification

The primary indicator of a potential speed loss is a slow but constant increase in engine power required to maintain a constant speed in similar weather conditions. Many variables influence the power and can be divided into two categories depending on whether they are local phenomena or trends:

Short term variables like:

- Trim²
- Draught³
- Sea State⁴
- Wind
- Currents
- Fuel Quality
- Human factors

can vary from trip to trip. Fuel may have a higher calorific value⁵, weather conditions change, and different quantities of cargo can be loaded. While useful for understanding local variations these factors do not generalize over time.

Long term trends:

- Component Degradation
- Fouling

Mechanical components within the propulsion system, even with the recommended maintenance, lose efficiency over time due to degradation. Similarly, fouling, if left untreated, tends to worsen over time. The goal is to identify a steady increase trend in power usage to detect potential speed loss due to fouling. Figure 3.5 displays the power supplied for each data sample. It is evident that a clear trend cannot be discerned from this graph. For 2014, the speed decided cruise speed and also mode is 13 knots 3.9c, and to smooth out local changes, the plot can be done using a moving average at 5'000 sample instead of the raw values. The result is shown in figure 3.6. From January onwards, a clear trend in increasing power is observed, even when considering peaks and valleys associated with local factors or reasons discussed in section 3.4.2. The sudden drop in mid-August could indicate a cleaning event, given that one and a half years had passed since the previous cleaning in December 2012, although no reports were available at the time of analysis to confirm this theory.

²Ship trim refers to the distribution of weight on a vessel that affects its balance and stability. It involves adjusting the position of cargo, fuel, and other weights to ensure the ship sits level in the water, optimizing performance and safety.

³Ship draught (or draft) is the vertical distance from the waterline to the lowest point of the ship's hull.

⁴Sea state refers to the general condition of the free surface on a large body of water, considering wind waves and swell at a specific location and time.

⁵Calorific value is the energy content per unit mass of the fuel.

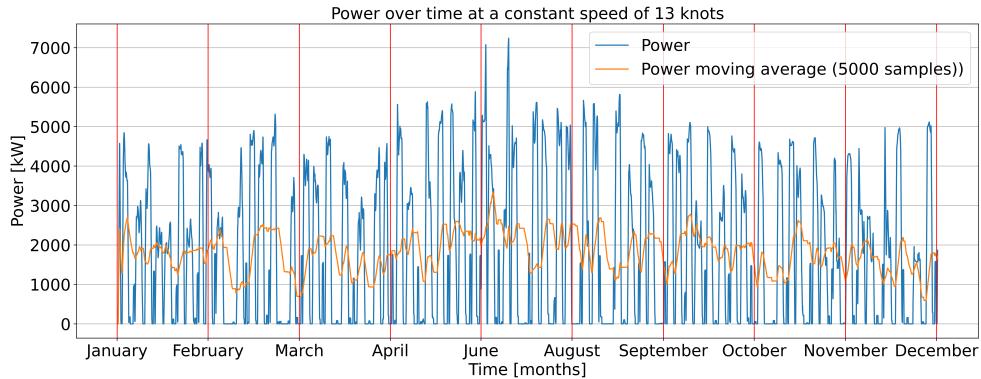


Figure 3.5 Power values over 2014

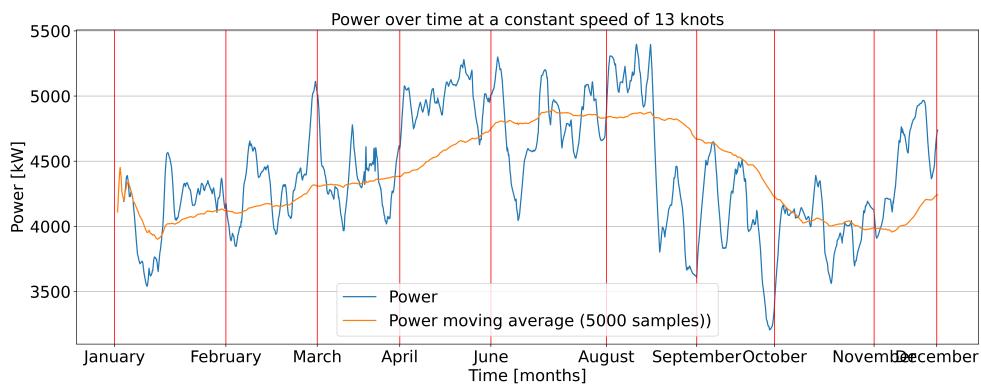


Figure 3.6 Power values over 2014

Figure 3.7 illustrates torque values from the end of 2012 to the end of 2016. A slow increase is noticeable, it begins right after the ship's cleaning in early 2013, continuing until a minor cleaning event in mid-2014. For all the reasons previously mentioned , this period is selected for the machine learning models as a potential indicator.

3.4 Adopted simplifications

3.4.1 Speed sensors

The choice of speed sensor is the most critical phase, as the ML model will rely on this measurement. Ideally, the sensor should be precise and have a minimal percentage of NaN values. The Speed Over Ground (SOG) sensor contains NaN value for more than 25% , thus necessitating a high amount of imputation and row elimination, making it less valuable for accurate analysis. As shown in figure 3.8, pressure sensors, placed within the boundary

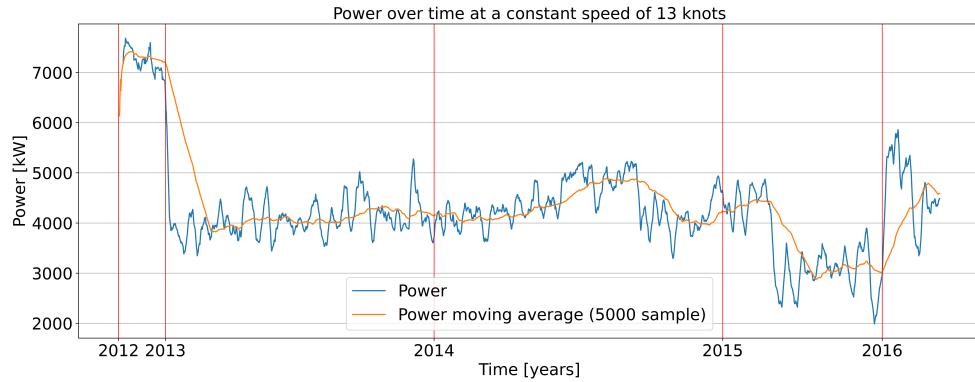


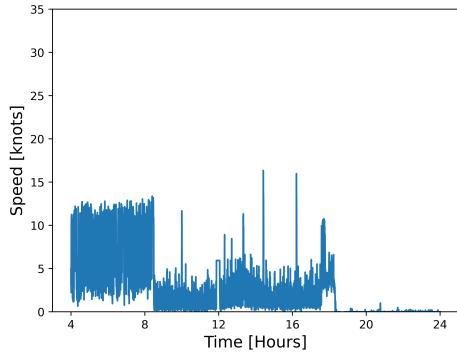
Figure 3.7 Power values various years

layer, suffer from imprecise and discordant measurements. This results in significant outliers and speed variations that are physically implausible, such as speed in figure 3.8b, sudden increases/decreases in speed 3.8d and large fluctuation due to the closeness of the hull to the dock 3.8c. Additionally, there was a malfunction in the recording system during a certain period, leading to a data void that cannot be ignored. Given the previously mentioned criticality, the optimal choice is to adopt Doppler measurements as the reference.

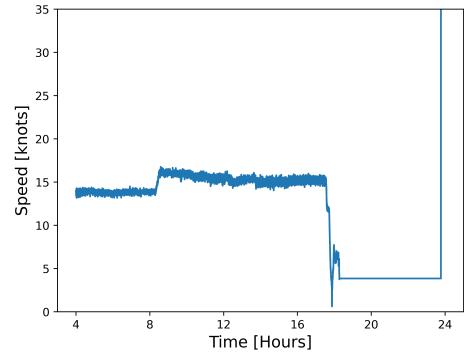
3.4.2 Speed range

The speed range considered for the ML model prediction needs to be close to cruise speed. To achieve a desired speed, engine power must be gradually increased until the target is reached. For instance, if the speed target is set to 18 knots, the power recorded for speeds below 18 knots will reflect the increased power needed during the acceleration phase. This results in a higher power reading than what is required to maintain the target speed, creating a false power pattern for that speed. The same issue occurs during deceleration phases where the power is gradually decreased.

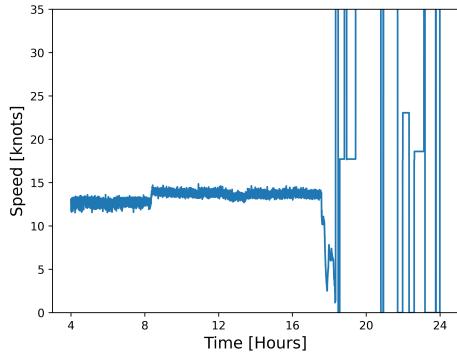
The ship under analysis experienced various operational profiles, resulting in changes to its cruise speed over the years. From 2010 to the first semester of 2013, the cruise speed ranged between 15 and 17 knots. From the second semester of 2013 until 2016, the cruise speed was set at 14 knots. Consequently, power values recorded at 14 knots during this period reflect the same acceleration and deceleration principles previously mentioned, which introduces additional errors into the model.



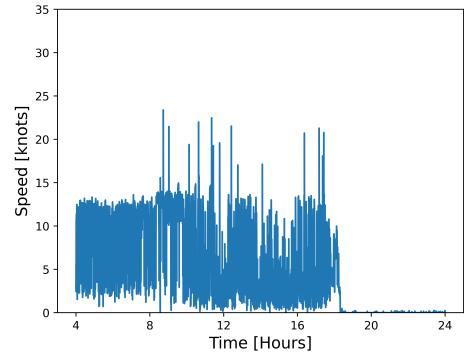
(a) Water flow 1 sensor recorded speed during a day at water line port side



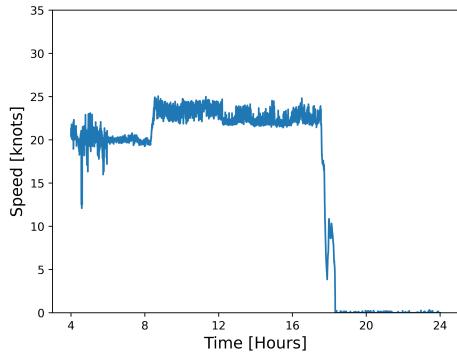
(b) Water flow 2 sensor recorded speed during a day at keel line port side



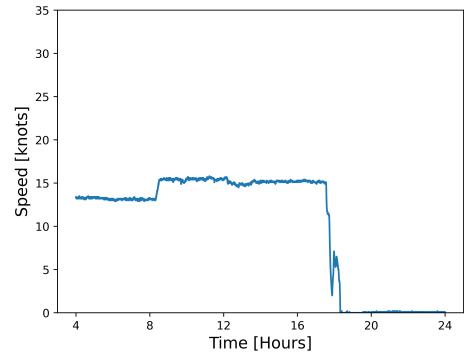
(c) Water flow 3 sensor recorded speed during a day at water line starboard side



(d) Water flow 4 sensor recorded speed during a day at keel line starboard side



(e) Water flow 5 sensor recorded speed during a day at bow keel line



(f) Doppler speed recorded speed during a day at bow keel line

Figure 3.8 Comparison of an entire day worth of recorded speeds for each speed water flow sensors

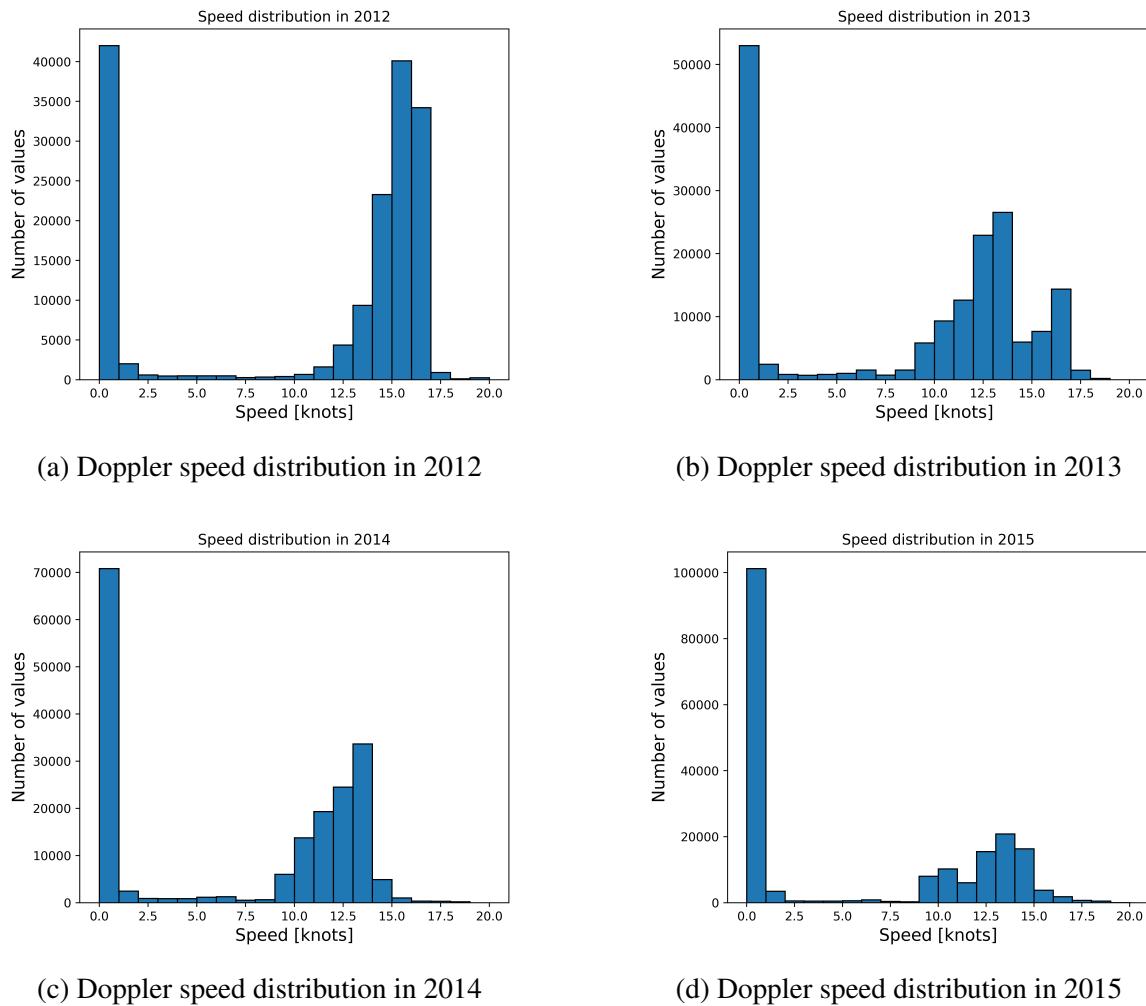


Figure 3.9 Doppler speed distribution over the years

3.4.3 Years considered

As mentioned in section 3.4.2, the ship under analysis went through a range of operational profiles. While this variation can provide a more comprehensive and varied dataset, it also introduces numerous challenges. The only weather data acquired is from the anemometer. For other variables such as wave direction, height, period, and frequency, and lastly swell⁶, extensive imputation would be required. The results of such imputation may be unsatisfactory due to the high variability of maritime zones and the absence of data for certain period or zones. Given the above reasons, not all recorded years were used in this project. Only the data from 2013 to 2014 were considered, as the operational profile was in the same zone and

⁶Regular wave motion caused by large meteorological disturbances operating at a distance

the trips were similar, allowing for a clearer identification of fouling growth and its impact on the increase in shaft power required.

3.4.4 Dataset averaging

The original frequency of data acquisition by the monitoring system is one sample per second. To facilitate data handling, reduce computational load, and shorten training time, the data are processed to produce 2-minute averages, or 120 samples without losing significant information. This approach is feasible because cruising is a relatively stable process, the only variables that can vary frequently are the ones related to wind, but even in that case the changes are slow enough to be captured in a 2 minute averaged interval. Also, the desired speed is typically maintained for entire days, as in the case being analyzed, and a lower amount of sample reduces the number of identical values in the dataset. Averaging also reduces the impact of outliers by spreading their effect across all data points. Since the mean takes into account every value, extreme outliers have less influence on the overall average compared to the central values in the dataset.

3.5 Dataset enrichment with ERA-5 database

ERA-5 is one of the largest databases for the global climate and weather data. It is part of the completely free Copernicus project from the European Union (17).

3.5.1 Sea surface temperature

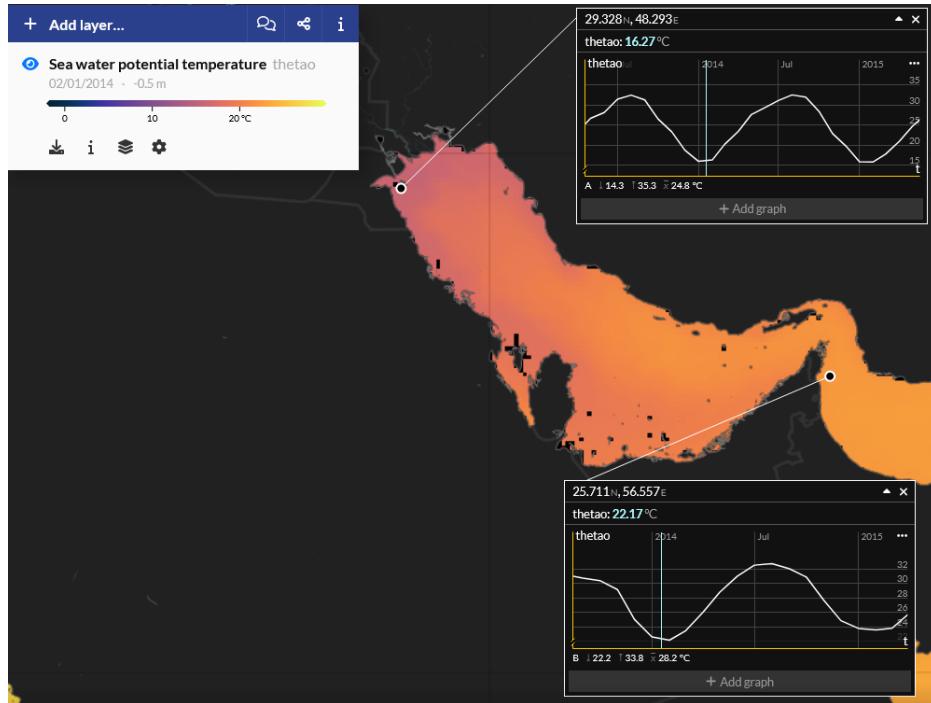
Sea surface temperature is the primary variable influencing the growth of marine algae and animals. Some models presented in the existing literature consider SST as the sole factor for formulating growth models (11). Temperature is primarily influenced by factors such as latitude and season, while longitude has a minor impact, as discussed in section 1.2.2.

General dataset analysis

The dataset consists of measurements taken every hour every 0.25° of latitude and longitude creating a squared grid of 750 squared kilometers, with Kelvins as unit of measure. In the Persian gulf zone the dataset is completed and no NaN values are present, giving a good representation of seasonal pattern of the temperature, as shown in figure 3.10.



(a) Summer SST level in Kuwait city port in the north and Fujairah in the south



(b) Winter SST level in Kuwait city port in the north and Fujairah in the south

Figure 3.10 Comparison between winter and summer SST in the two port where the ship docks.

3.5.2 Chlorophyll concentration

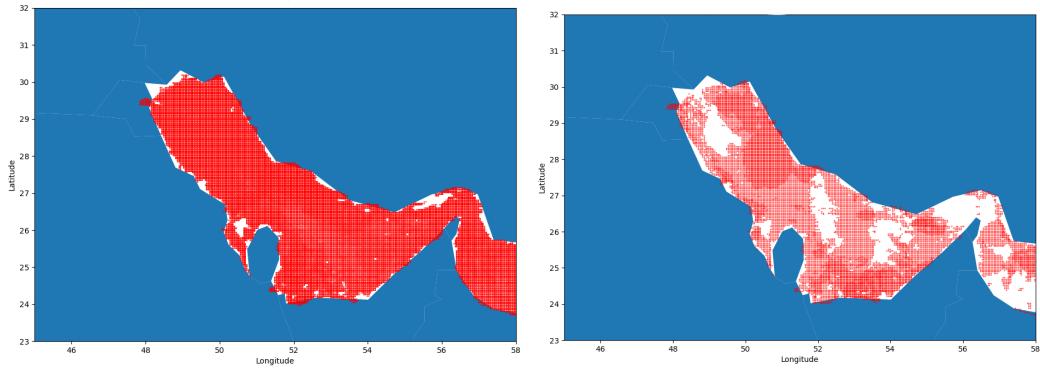
This dataset provides daily estimates of chlorophyll concentration in ocean water, measured in mg/m³. The concentration is calculated based on sea surface color and offers unique insights into the amount of algae present in an area. Chlorophyll is essential for algae as it supports the photosynthetic process, enabling plants to absorb energy from sunlight and convert it into nutrients. Thus, there is a direct correlation between chlorophyll levels and the amount of algae in a water body. Chlorophyll concentration influences the fouling process: higher density indicates that more organisms are living, growing, and reproducing in the water, which increases the amount of fouling that can attach to structure immersed in water. Additionally, it reflects the water fertility of a zone, allowing for differentiation between higher and lower fouling risk areas.

General analysis

The entire detailed process of how the dataset is obtained can be found here (16), but in short this is the procedure on their site: "Remote-sensing reflectance (or Rrs) is defined as the ratio of water-leaving radiance to downwelling irradiance and serves as the main input to algorithms used to derive other ocean colour products. Chlorophyll-a (Chl-a) is the main photosynthetic pigment found in phytoplankton, which form the base of the marine food-web and are responsible for approximately half of global photosynthesis. Chl-a can be estimated from Rrs data using different algorithms (see details in the Documentation)". The data are stored on a daily basis, but chlorophyll values are not always present because the measurements rely on reflections from the sea surface captured by a satellite, which are not always available. This limitation results in a dataset with inconsistencies. As illustrated in figure 3.11, in the best cases, multiple values are close together, even less than a couple of kilometers apart. In the worst cases, entire zones can be missing for several months at a time.

Patterns

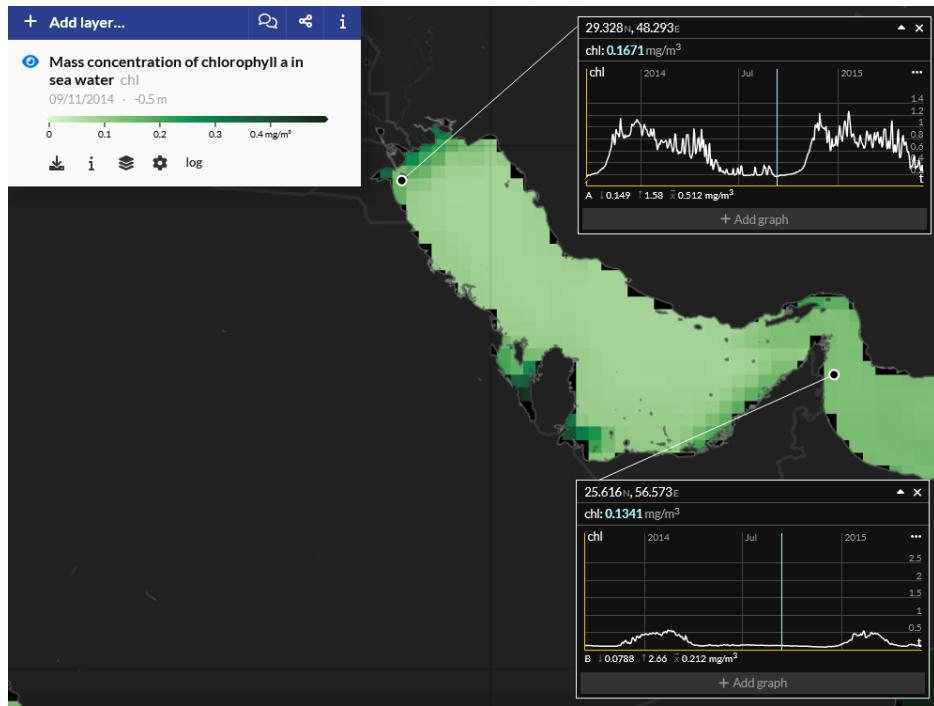
The Copernicus visualizer uses a different data scale compared to the dataset used for the ML part, but the patterns are the same, and the analysis remains valid. As shown in figure 3.12, chlorophyll levels vary based on multiple variables: location, season, temperature and the presence of rivers. For these reasons, the analysis will be divided into two parts based on where the ship stops: in Kuwait, in the north near the rivers, and in Oman, in the south. In the northern part, starting from November, an increase in the chlorophyll level can be seen, rising from 0.2 mg/m³ up to 1 mg/m³. This quantity remains steady until the start of February,



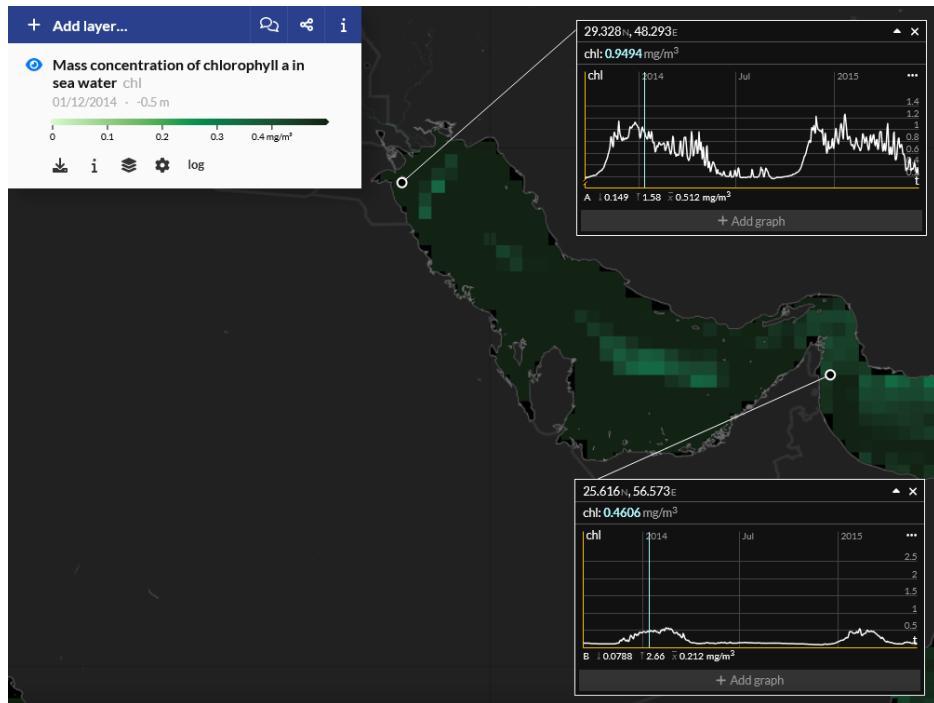
(a) Map chlorophyll samples in a selected week (b) Map chlorophyll samples in a selected week

Figure 3.11 Chlorophyll sample coverage comparison, each point is a recorded sample in a selected week

when a diminishing trend begins to appear, although it oscillates within a range of ± 0.4 . From March to October, the values are all below 0.3 mg/m^3 , with occasional spikes to 0.4 mg/m^3 . In the southern part, the concentration is always lower than in the north, although the trend is the same and follows the seasonal pattern. The maximum value never exceeds 0.5 mg/m^3 during the seasonal peak and is never more than 0.2 mg/m^3 in summer.



(a) Summer chlorophyll level in Kuwait city port in the north and Fujairah in the south



(b) Winter chlorophyll level in Kuwait city port in the north and Fujairah in the south

Figure 3.12 Comparison between winter and summer chlorophyll level in the two port where the ship docks.

3.6 Dataset creation for the biological model

A dataset composed by the previously mentioned variable was not available and for this project an ad hoc data set was created. The MARIN dataset was divided in weeks that are further split in three distinct parts: samples recorded during cruising, samples taken while stationary in the north, and samples taken while stationary in the south. This division was implemented for several reasons:

- **Fouling growth:** For this project, one week has been identified as the optimal trade-off between having a sufficient number of samples to train the model and capturing a slight increase in speed loss. It cannot be excluded that in areas with a higher fouling rate, a similar increase might be observed in a shorter time span, which could provide more samples. Additionally, if more data become available, it is possible to extend the analysis up to two weeks.
- **GPS Accuracy Issues:** At times, the GPS position recorded in the dataset was significantly inaccurate, showing the ship stationary in locations far from its actual position, such as in the middle of a desert. This inaccuracy led to a substantial number of NaN values because the merging process could not find any sufficiently close SST or chlorophyll values for those erroneous positions.
- **Port Positions:** The ship consistently docked at two specific ports. By saving the accurate latitude and longitude coordinates for these ports, it was possible to overcome the inaccurate samples problem. Merging was conducted based on these known port locations, improving the accuracy of the data for stationary periods.

This approach ensured that the dataset was more reliable and that the subsequent analyses and modeling were based on accurate and relevant data.

3.6.1 MARIN, SST and chlorophyll merging process

For the stationary data, values of chlorophyll and SST within a defined area around each port were collected. Specifically, a squared zone with a longitude range of ± 1.5 degrees and a latitude range of ± 1 degree from the port's coordinates were considered. This broader area was chosen because, at times, data were not available for locations very close to the shore. Within all the samples in this zone, the distributions of chlorophyll and SST values were calculated and stored for the period during which the ship was stationary. For the moving part of the dataset, chlorophyll and SST values were merged with a tolerance margin

of ± 0.5 degrees in both latitude and longitude. This approach allowed for the inclusion of additional samples, as chlorophyll and SST levels tend to vary less significantly in the open sea compared to coastal areas. By extending the range slightly, more data points were gathered, improving the robustness of the analysis for the ship's movement phase.

3.6.2 Imputation of missing values

The chlorophyll concentration dataset for both 2013 and 2014 lacked values for several months. To address this, imputation was performed by averaging. Specifically, missing values were filled using the moving averages of the three previous weeks. This approach preserved the diminishing trend observed in the figures 3.12. In the SST dataset, no values were missing, so no imputation was necessary.

3.6.3 Final biological dateset

The obtained dataset consist of a CSV file with 72 rows and as columns:

- Period
- Chlorophyll distribution composed of 10 bins in the range of 0-7 mg/m³
- SST distribution composed of 10 bins in range 280-310 K
- Speed distribution composed of 10 bins in range 0-20 knots
- Amount of time spent stationary

Chapter 4

Proposed methodology

4.1 Speed prediction based on mechanical data

In this section the development of the model to calculate the speed-loss based on mechanical variable and the results are discussed.

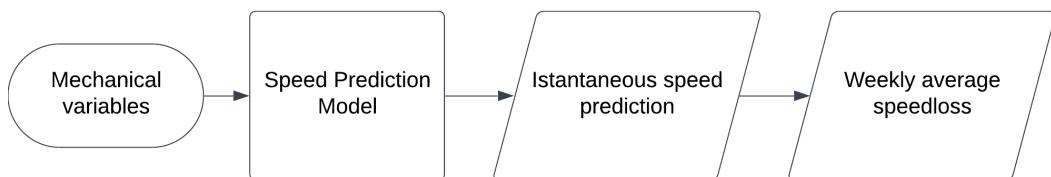


Figure 4.1 Mechanical variable model blocks scheme

4.1.1 Features and correlation

The first model implemented is based on mechanical and meteorological features. As input the model takes:

- Mean wind direction
- Mean wind speed
- Shaft power
- Shaft (engine) torque

- Shaft (propeller) RPM

and the target to predict is the speed through water measured by the Doppler sensor.

In figure 4.2 the pair wise correlation¹ between each variables is shown. Values between 0 and 1 indicate a positive correlation², between -1 and 0 instead a negative correlation³. It's important to remind that correlation doesn't imply causation, in the analysed case the real world theory suggest a causation, but it must not be taken as granted. As expected, the variables that strongly positive correlated are the shaft one, with values between 0.80 and 0.91. Wind speed instead is negatively correlated since an increase of this value means a stronger force that oppose the ship movement if it comes from the bow, or drift if it comes from the side.

4.1.2 Data preparation

Considered the simplifications discussed in section 3.4.3, the first 50'000 sample corresponding to the first semester of 2013 were discarded. As training set the samples from 50'000 to 84803 were chosen. Of this set, only the part with a speed between 10 and 14 (included) was selected, generating a training set composed of 13'616 samples.

In the first column of table 4.1 the number of sample for each speed rounded to the nearest integer⁴ is reported. It is important that the values are equally distributed in each class. This balance helps prevent the model from becoming biased towards predicting the more frequent class, which can occur when the model minimizes the overall error by favoring the majority class. To address this, two techniques are combined: over-sampling the minority class or under-sampling the majority class. Both approached were used leading to the result shown in the third column of table 4.1 .

In Figure 4.3, the distribution of the target variable across the training and testing splits is shown and the similarity is clear. It is important for these distributions to be similar in order to correctly represent the phenomenon that the model have to learn.

The training dataset was further split: 70% of the sample were used for training and 30% for validation. Data splitting is a step necessary in any ML model development mainly for two reasons:

¹Pairwise correlation refers to the statistical relationship between two variables in a dataset. It measures the strength and direction of their linear relationship

²Positive correlation means the variables tend to increase together, but not perfectly

³Negative correlation means that as one variable increases, the other tends to decrease, but not perfectly

⁴An integer is any positive or negative number that does not include decimal parts or fractions

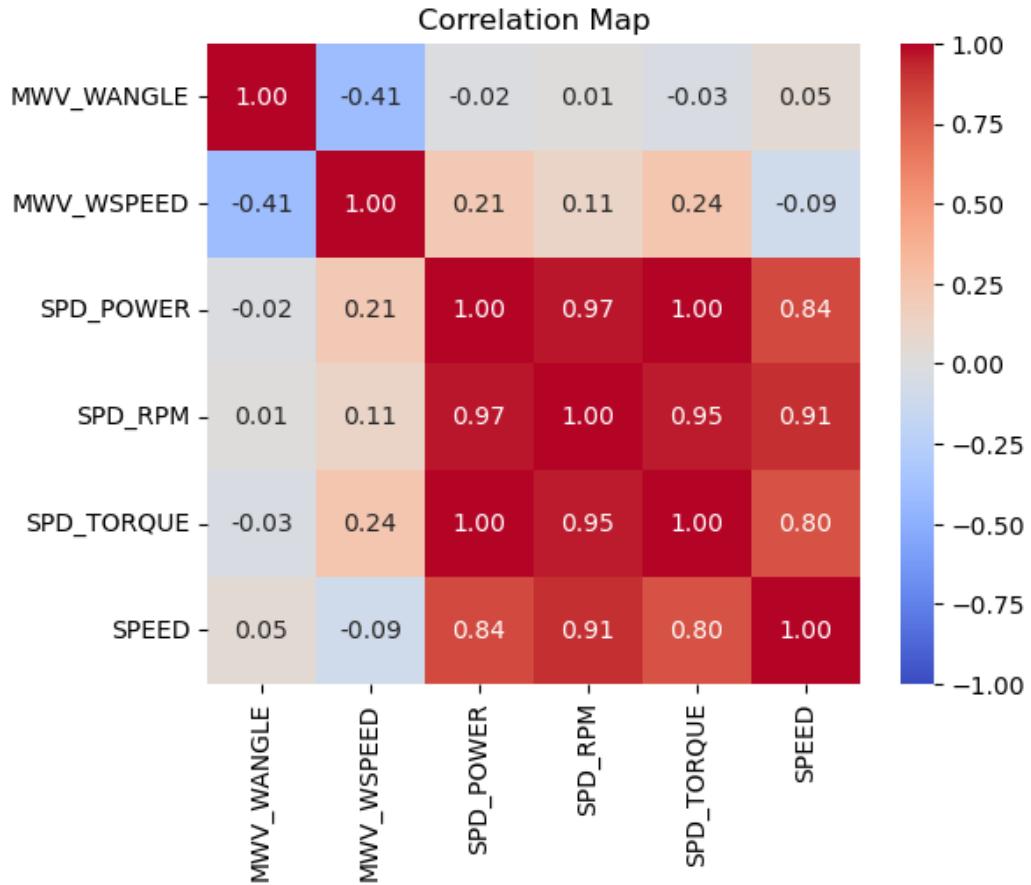


Figure 4.2 Input feature correlation map, warm colors for positive value, cold for the negative one

Rounded Speed	Original distribution	Distribution After Sampling
10	579	3500
11	1075	3500
12	3540	3500
13	3835	3500
14	4587	3500

Table 4.1 Distribution of samples before and after applying resampling techniques

- **Validation:** when the models is validated on data used during the training phase the results are not representative of the real performance due to the introduced bias caused by the fact that model already processed the samples.
- **Overfitting prevention:** overfitting occurs when a model performs well on the training data but poorly on the validation data. The goal during training is to minimize empirical

error while maintaining generalization capabilities. This is assessed using a separate validation set to ensure the model performs well on unseen data.

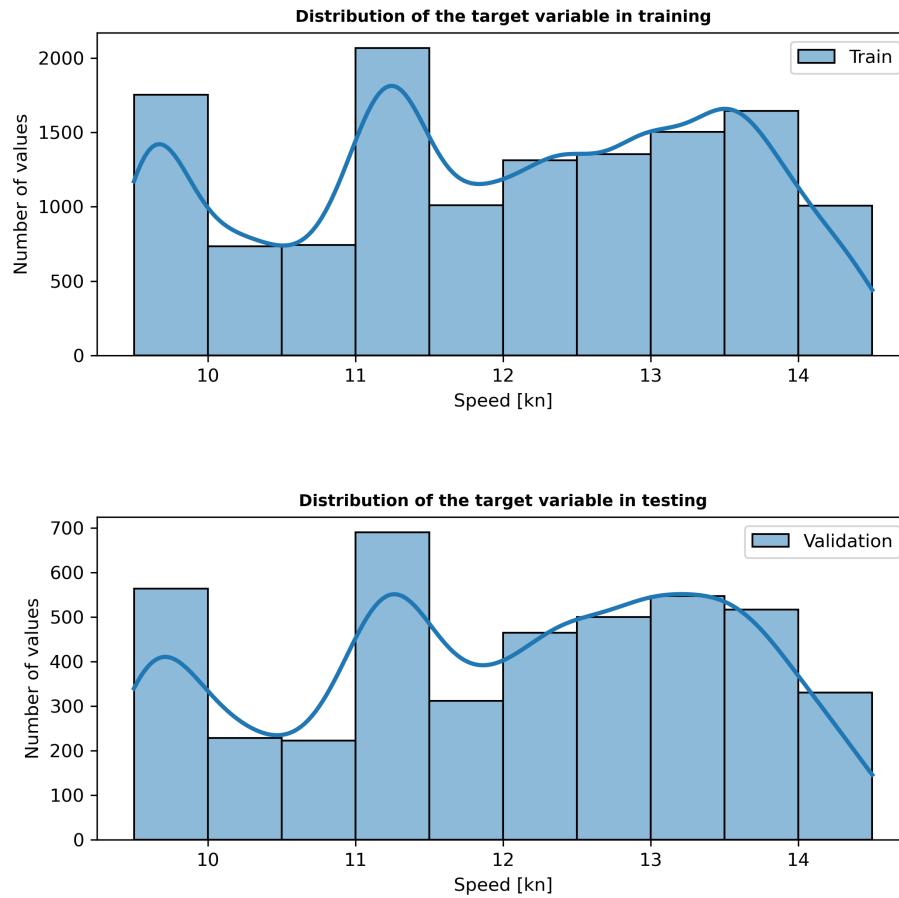


Figure 4.3 Number of sample per rounded speed in knots

After the splitting a MinMaxScaler from sklearn was applied to the input variables, excluded the target. It is important step because many machine learning algorithms are sensitive to the scale of input features, gradient descent-based models can struggle if features vary widely in scale, potentially leading to biased results or inefficient training. Scaling features to a common range ensures that each feature contributes equally to the model, preventing any single feature from dominating due to its larger numerical range. This not only improves the algorithm's performance and convergence speed but also enhances numerical stability, making the training process more efficient and reliable.

4.1.3 Models test

Having only 5 features as input simplify the choice of the ML model, since there is no need to use complex models to capture correlations between numerous of variables. Five different model were selected as candidate:

- **Kernel ridge regression:** (KRR) combines Ridge regression and classification (linear least squares with l2-norm regularization) with the kernel trick, meaning it learns a linear function in the space generated by the kernel and the data. For non-linear kernels, this corresponds to a non-linear function in the original space.

Hyperparameter	Values
alpha	0.01, 0.05, 0.1, 0.5, 1, 5
gamma	0.005, 0.1, 1, 5, 8, 10, 15, 20

Table 4.2 Hyperparameters for Kernel Ridge

- **Lasso regression:** is a type of linear regression that adds a penalty for large coefficients to the model. This penalty, called L1 regularization, encourages sparsity, meaning it drives some coefficients to zero, effectively selecting a subset of the features and helping to prevent overfitting.

Hyperparameter	Values
alpha	0.1, 1, 10, 100
max_iter	1000, 2000, 3000
tol	1e-4, 1e-3, 1e-2

Table 4.3 Hyperparameters for Lasso

- **Supported vector regressor:** (SVR) is a type of regression that uses the principles of Support Vector Machines (SVM). It aims to find a function that predicts values within a specified margin of error (epsilon) while minimizing the model complexity. SVR focuses on finding a balance between fitting the data and keeping the model simple, making it robust to outliers and capable of handling non-linear relationships with kernel functions.
- **Gradient boosting XGBoost (Extreme Gradient Boosting):** is an ensemble learning method that builds a series of decision trees, where each tree corrects the errors of the

Hyperparameter	Values
C	0.1, 1, 10, 100
gamma	1e-4, 1e-3, 1e-2, 1e-1
kernel	linear, poly, rbf, sigmoid

Table 4.4 Hyperparameters for SVR

previous ones. It uses gradient boosting to optimize the model by minimizing a loss function, and incorporates regularization to prevent overfitting.

Hyperparameter	Values
n_estimators	100, 200, 300
max_depth	3, 4, 5
learning_rate	0.01, 0.1, 0.2
subsample	0.6, 0.8, 1.0

Table 4.5 Hyperparameters for XGBRegressor

- **Random forest:** is an ensemble learning method that constructs multiple decision trees during training and outputs the average prediction of the trees. It improves accuracy and robustness by combining the predictions of many trees, each trained on a random subset of the data and features, which reduces overfitting and enhances generalization

Hyperparameter	Values
n_estimators	100, 200, 300
max_depth	None, 10, 20, 30
min_samples_split	2, 5, 10
min_samples_leaf	1, 2, 4

Table 4.6 Hyperparameters for RandomForestRegressor

For each model a randomized grid search was done to test all the previously mentioned parameters:

```
for model_name, model_info in models_param_grid.items():
    random_search = RandomizedSearchCV(
        estimator=model_info['model'],
        param_distributions=model_info['params'],
```

```
n_iter=50,  
scoring='neg_mean_absolute_error',  
cv=5,  
verbose=2,  
random_state=42,  
n_jobs=-1  
)
```

The parameters in that case are :

- `estimator`: specifies the model to be tuned.
- `param_distributions`: defines the hyperparameter grid for the search.
- `n_iter`: is set to 50, indicating the number of random combinations to try.
- `scoring`: is set to `'neg_mean_absolute_error'`, which means we are optimizing for the negative mean absolute error.
- `cv`: specifies the number of cross-validation folds, set to 5.
- `verbose`: controls the level of detail in the output, set to 2.
- `random_state`: ensures reproducibility by setting a random seed.
- `n_jobs`: is set to -1 to use all available CPU cores for computation.

The total time required to complete the grid-search is one hour and 10 minutes.

The result are shown in figure 4.4 The best-performing model on both training and validation is the Kernel Ridge Regressor, which has a Mean Absolute Error (MAE) one-third lower than the second-best model, XGB. Although there is a slight overfitting, as indicated by a small increase in error during validation compared to training, the overall increase is minimal. Therefore, it can be concluded that the Kernel Ridge Regressor is a strong model.

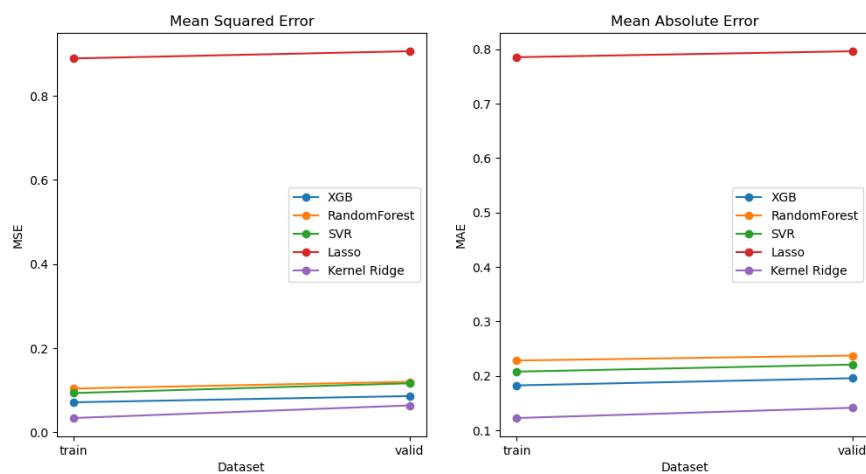


Figure 4.4 Performance comparison with different metrics of each model in train and validation set

4.2 Speed-loss prediction based on biological data

Considering the available model described in the previous section 4.1, we can now calculate the speed loss on a weekly basis. Given the historical data on the ship's locations and the additional biological variables, we can train the model with these inputs. The idea is for the model to use biological data, along with information on how long the ship remains stationary, to learn and predict speed loss under these conditions. In figure ?? the block schema of the biological model is shown, instead in figure 4.6 the schema represent how the two model were integrated together.

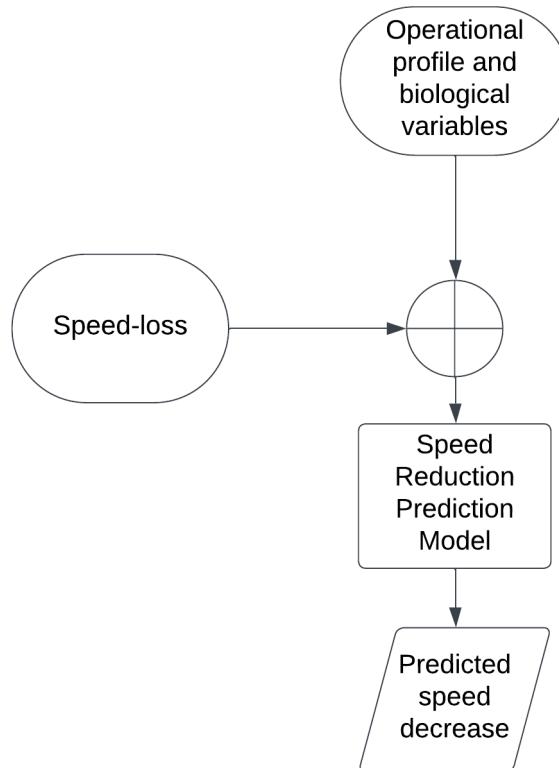


Figure 4.5 Mechanical variable model block

4.2.1 Feature correlation

This model is based on biological variables and the operational profile of the ship, from right after the cleaning to the value in the column 'week'. As input features the model takes:

- **Chlorophyll distribution:** composed by 10 bins with a fixed range of 0-7 mg/m³

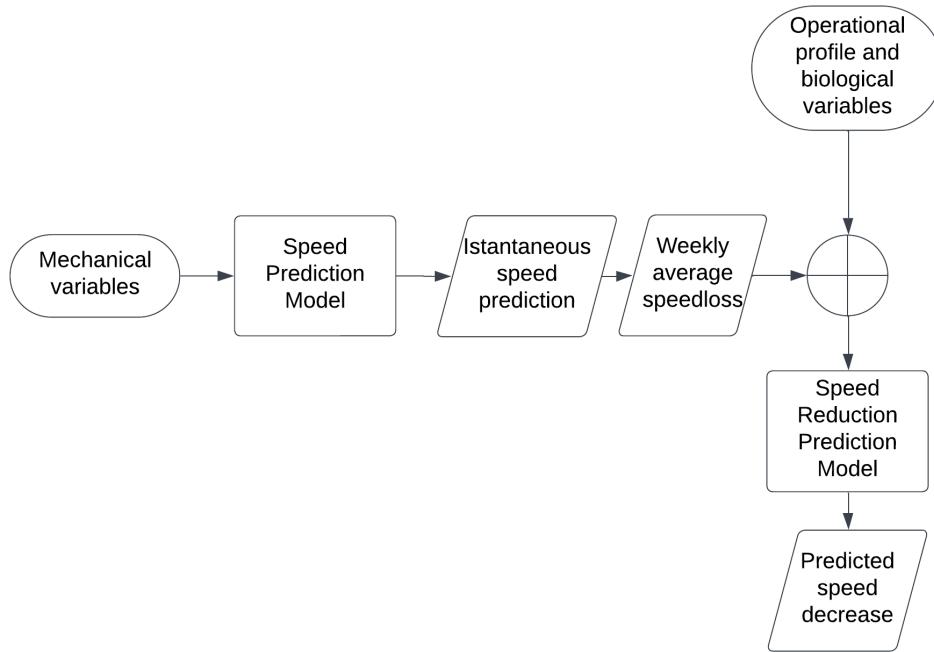


Figure 4.6 Mechanical variable model block

- **SST distribution:** composed by 8 bins with a fixed range 286-310 Kelvin
- **Speed distribution:** composed by 10 bins with a fixed range of 0-20 knots
- **Time stationary:** float value that represents the amount of hours spent stationary

The target is the speed-loss.

In figure 4.7 the pair wise correlation between each variables is shown, in figure 4.8 a detail of the correlation between the speed-loss and every other variable is presented. There is a clear positive correlation with the time spent stationary, as expected based on the literature, but also with higher values of chlorophyll level and mild water temperature. There is instead a negative correlation with the cruise speed, that checks out with the real world scenario where the fouling grows less during cruising time. Interesting to notice that also higher values of SST are shown as negative, a possible explanation can be attributed to the fact that it's right after the reproductive cycle and growth can be dormant.

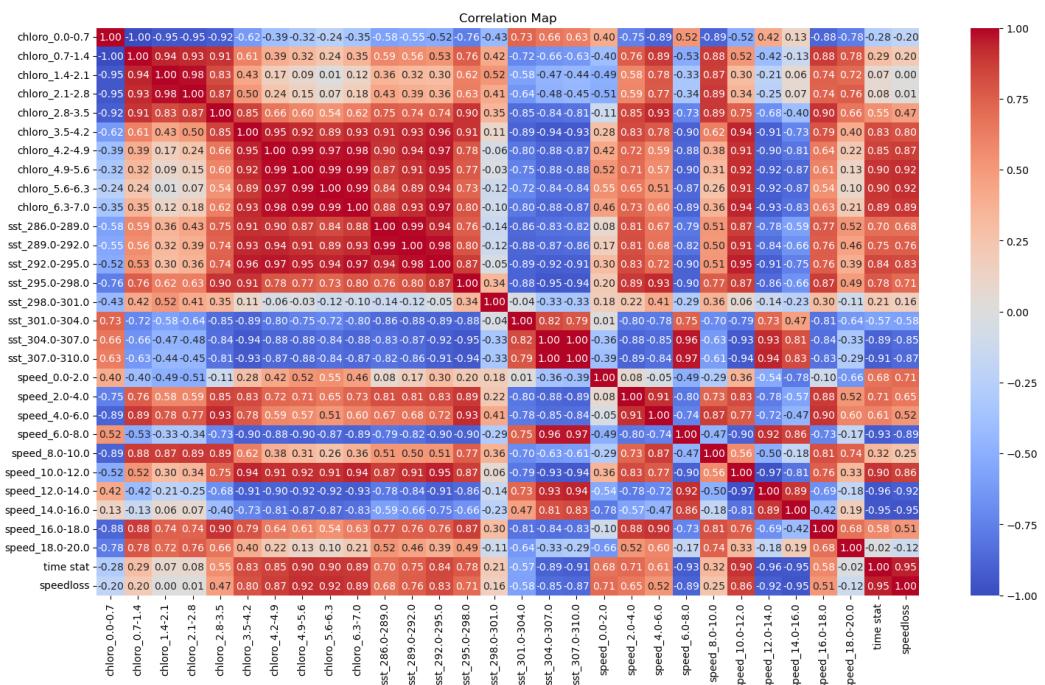


Figure 4.7 Pairwise correlation of all the input variables of the model

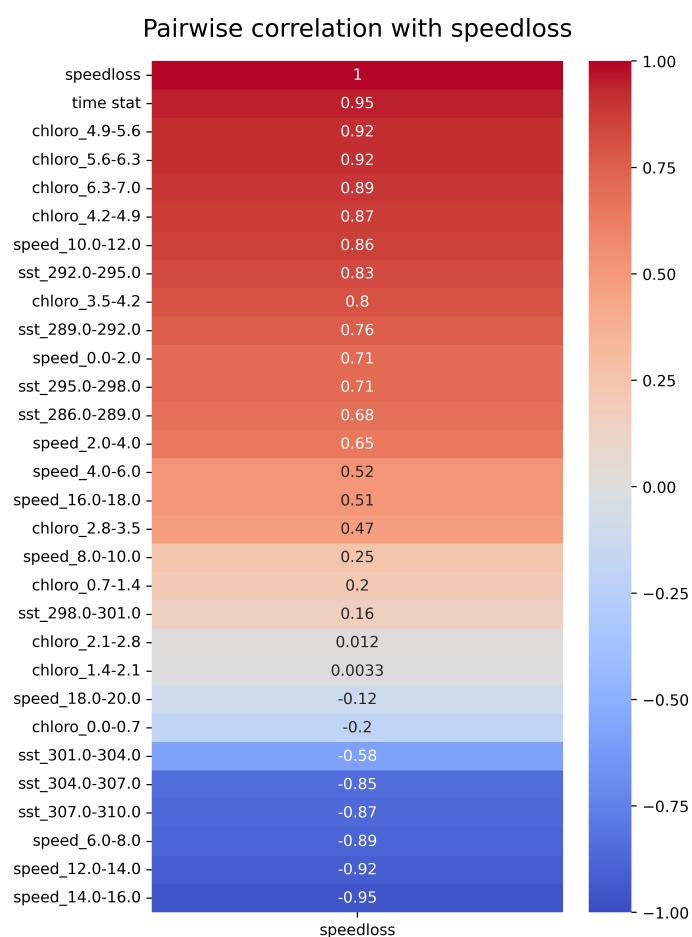


Figure 4.8 Pairwise correlation of the speed-loss value with every other variable

4.2.2 Data preparation

In this case, the available dataset size is limited to only 40 samples, leading to a small sample size problem. To address this, the approach was to use K-fold cross-validation⁵ to better validate the model, figure 4.9 shows how it works.

Validation is crucial to determine if the model is correctly trained in this type of problem. Therefore, the training dataset was randomly split into 50% for training and 50% for validation, meaning 20 samples each. The distribution of the target variable in training and validation set is shown in figure 4.10.

In that case the distribution were normalized in respect to the total amount of values to accurately describe the distribution, so a scaling process in not necessary.

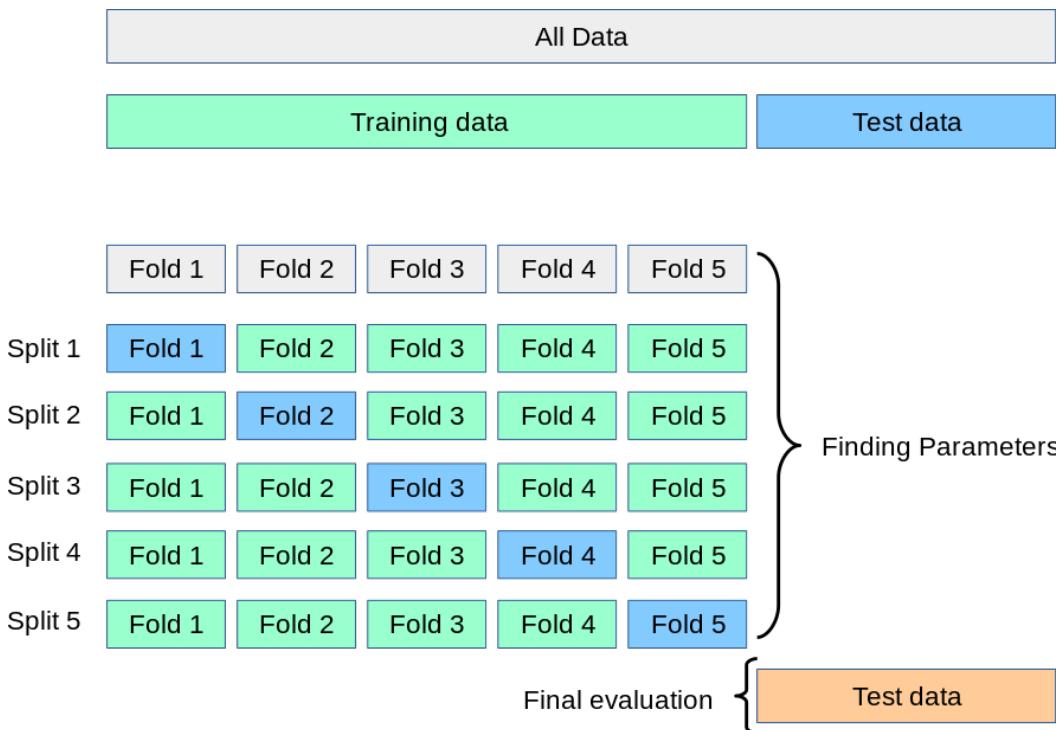


Figure 4.9 Visual representation of the folds generated during cross-validation (cro)

⁵It involves splitting the dataset into k subsets or folds, where each fold is used as the validation set in turn while the remaining $k - 1$ folds are used for training.

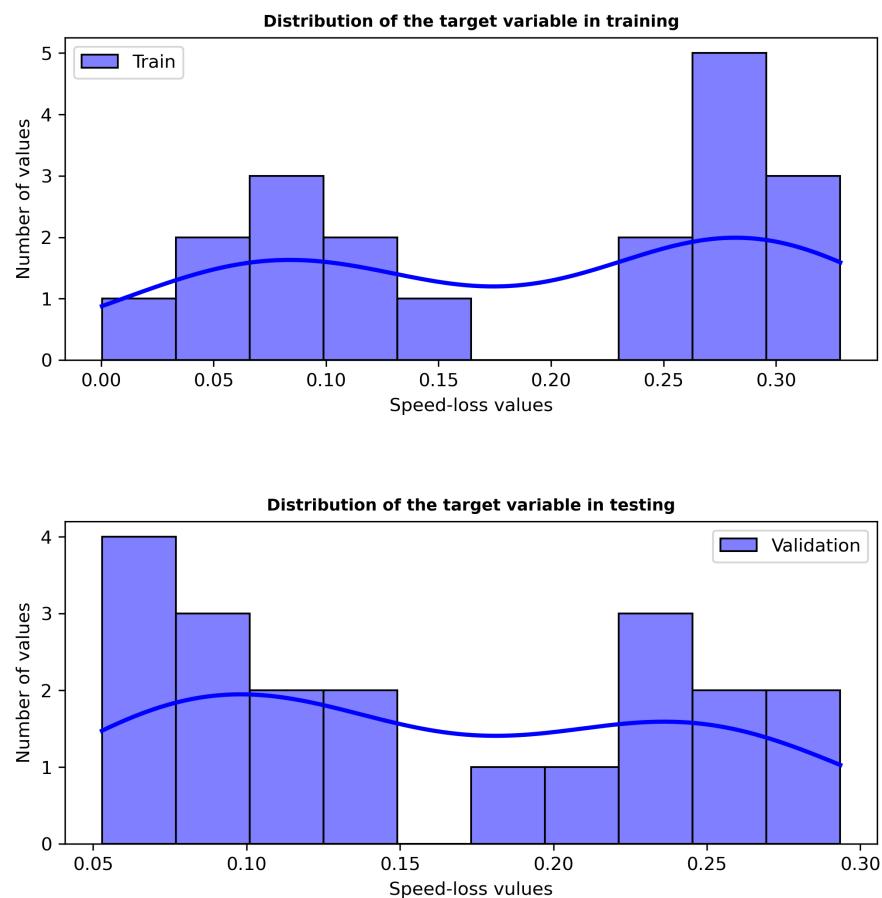


Figure 4.10 Number of sample per speed-loss value

4.2.3 Models test

The same models used for the mechanical part in section 4.1.3 were applied with the same parameters. Although the number of variables has increased, the sample size has decreased, making more complex models less effective in this scenario. The total training and validation time is two minutes, producing the results shown in figure 4.11.

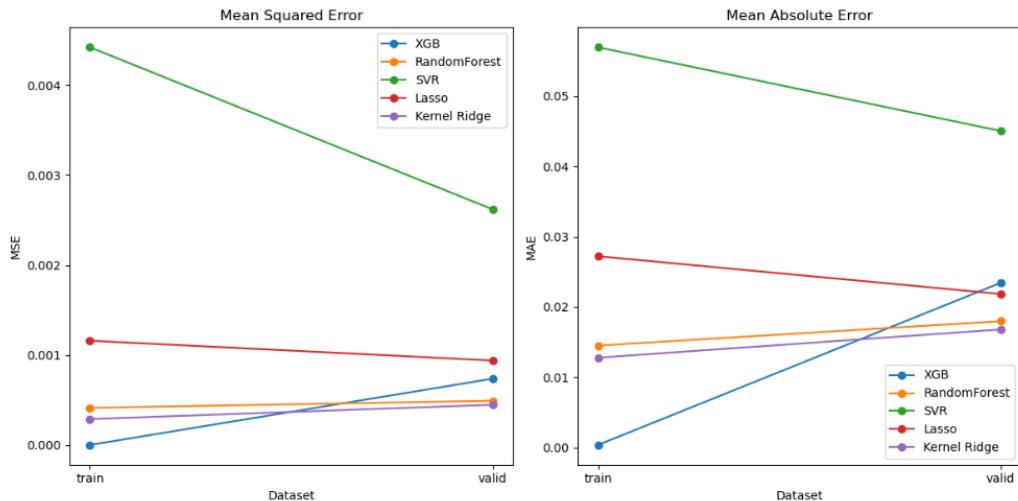


Figure 4.11 Performance comparison with different metrics of each model in train and validation set

It is clear from the figure 4.11 that the KRR is the one with the best performance in both metric. It also doesn't show sign of overfitting like the blue line of the XGB, or underfitting like the green line of the SVR.

Chapter 5

Results

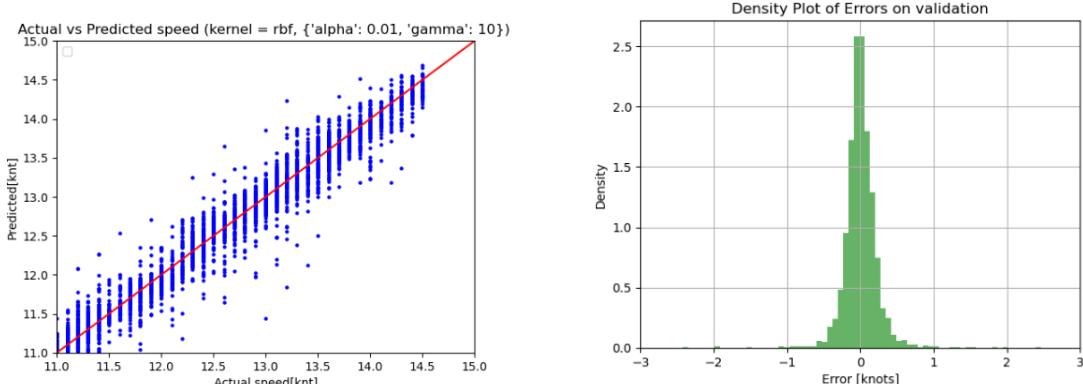
5.1 Mechanical model

As final model the KRR was chosen. A complete re-train was made with the best performing parameters from the grid search corresponding to: alpha = 0.01, gamma = 10 and the Gaussian kernel. The total time required to complete the grid-search is 20 minutes plus another minutes for the final training. Time required to make a prediction is at around a couple of seconds. The loss function to minimize was set again to 'neg_mean_absolute_error'.

Two score mechanisms were used to evaluated the performance, mean absolute error and mean squared error and their final values were:

- **MAE:** 0.13, meaning that, on average, on average, predictions deviate by 1.2% of the average predicted value.
- **MSE:** 0.06. which is less than the MAE with could be counter intuitive at first. This occurs when most of the errors are small; when squared, these small errors become even smaller, thus lowering the MSE compared to the MAE.

From the scatter in figure 5.1a is clear that, excluding outliers in which the error is above one knot, the majority of the points is close to the diagonal, that represent a perfect prediction. It is important to notice that this type of graph empathise the errors since are the only values that are actually visible. The distribution of the errors 5.1b clearly show what said before, the peak is in zero and most of the other values are between ± 0.3 .



(a) Scatter plot of the validation set predictions

(b) Density plot of the error in the validation set prediction

Figure 5.1 Error during validation, dividend in a scatter plot and a density plot

5.1.1 Speed-loss identification

Taking as reference 5.1a and 5.1b, the majority of errors are within ± 0.3 , with the center of the distribution around zero. When comparing these plots with those in Figure 5.2, it becomes evident that the scatter plot points tend to be higher relative to the diagonal over time, indicating an increase in error with the model predicting higher values than the actual ones. Similarly, the distribution of errors, which was initially centered around zero, is slightly shifted to the right as expected. In figure 5.3 the trend becomes even more clear, the predicted point tends to be all overestimated by the mode as time passes, particularly in figure 5.3c representing the scatter plot of July 2014. That is confirmed by the distributions of the error in both picture 5.2 and 5.3 where the center of is clearly offset from the zero in contrast to the one used in validation 5.1. All these pictures confirms the trend of overestimating the speed, confirm the presence of speed-loss.

In figure 5.4 the mean weekly error of the mechanical model's speed predictions is plotted as a yellow solid line. The first noticeable observation is that, after the 43rd sample, there are significant oscillations in the predictions. Prior to this point, the error showed an increasing trend, suggesting that a change in the ship's conditions occurred, but no data were available to explain it comprehensively. Consequently, this segment will be disregarded.

Additionally, in the first 10 samples, there appears to be a diminishing trend in the error. This can be attributed to the cruise speed during that period being set higher than in subsequent periods and thus discarded; further details on why that is done are provided in Section 3.4.2.

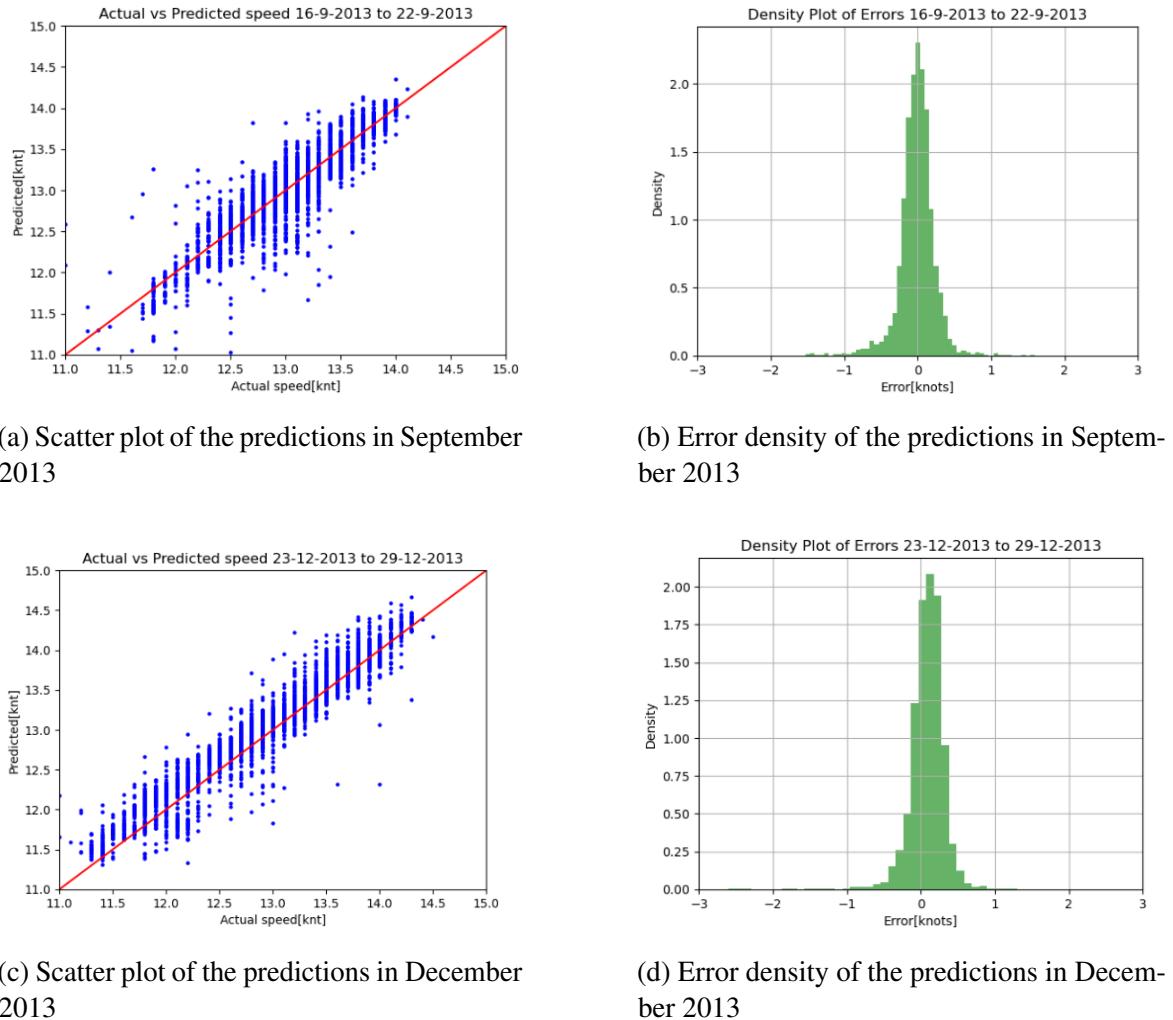


Figure 5.2 Comparison of scatter and density error plot for selected weeks in 2013

The most relevant section for analysis is from sample 12 to 43, where, apart from a small peak and subsequent decreasing, there is a steady increase in error, likely due to fouling. This segment will be the focus of our analysis.

To address the high oscillations in the error, which are a result of the limited number of variables used for prediction, the biological model will employ a moving average of the error values, as indicated by the green dotted line in Figure 5.4. This smoothing process, which involves applying a moving average over 5 samples, helps mitigate local fluctuations and emphasizes the overall trend, rather than individual values.

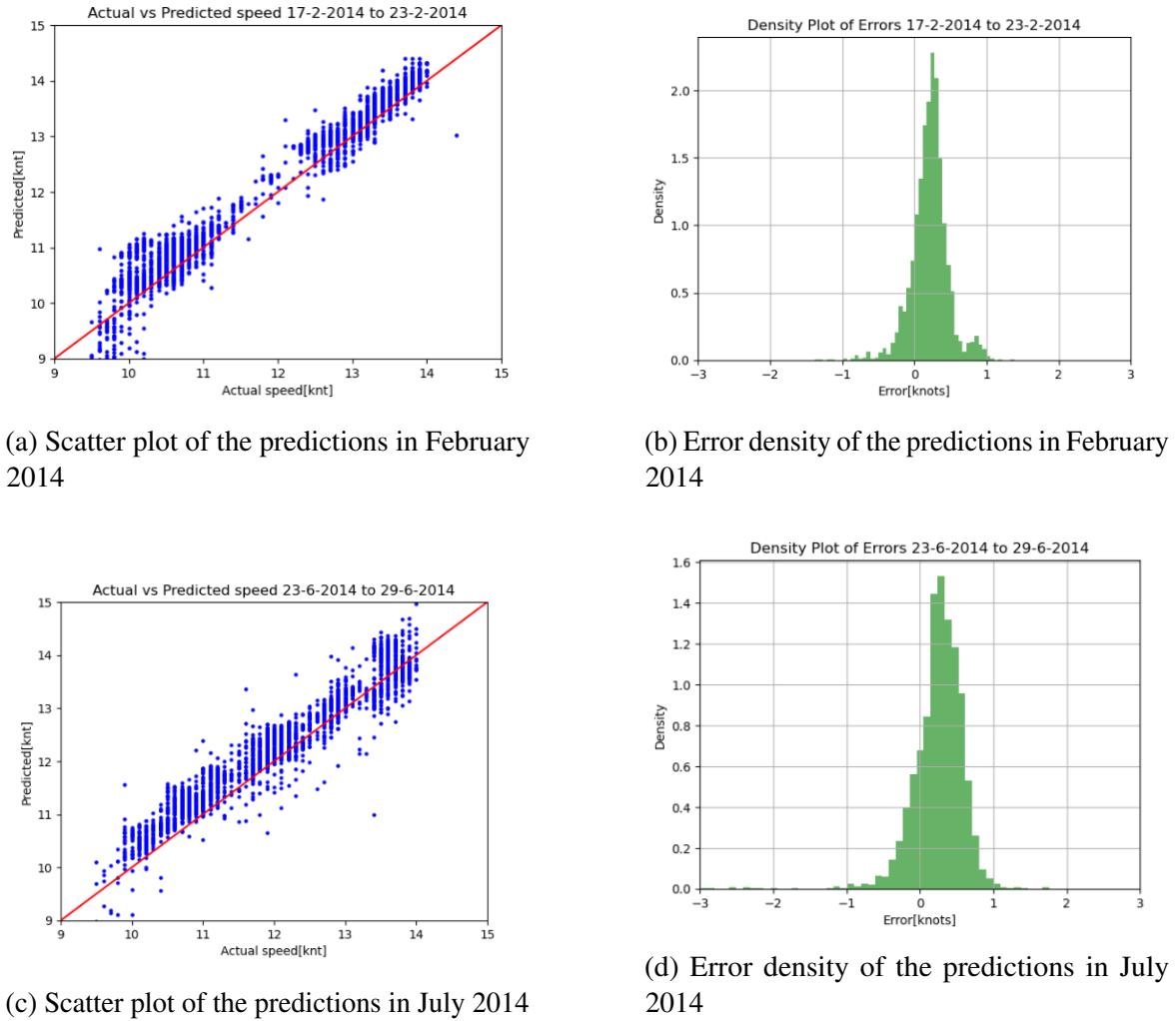


Figure 5.3 Comparison of scatter and density error plot for selected weeks in 2014

5.2 Biological model

As final model the KRR was chosen again as the best performing. A complete re-train was made with the optimal parameters from the grid search corresponding to: alpha = 1, gamma = 0.1 and the Gaussian kernel. The total time required to complete the grid-search and the final training is less than a minute. The loss function to minimize was set again to 'neg_mean_absolute_error'. Two score mechanisms were used to evaluate the performance, mean absolute error and mean squared error and their final values were:

- **MAE:** 0.016, this means that, on average, your predictions are off by about 9.7% of the average predicted value.

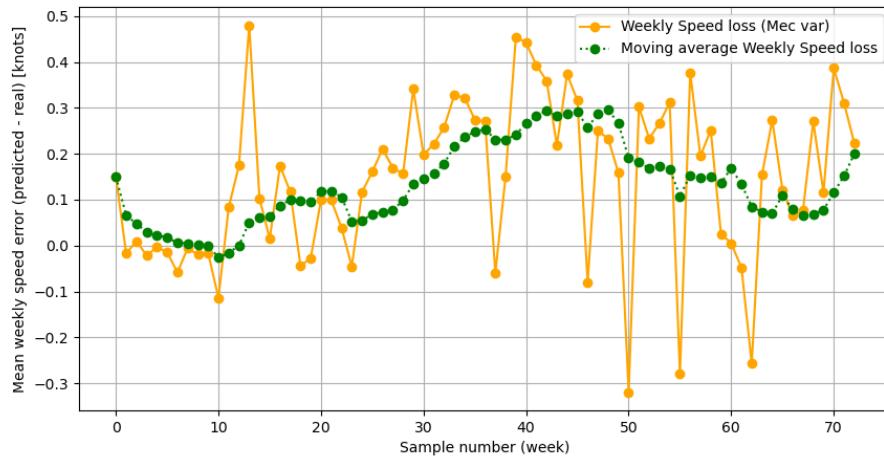
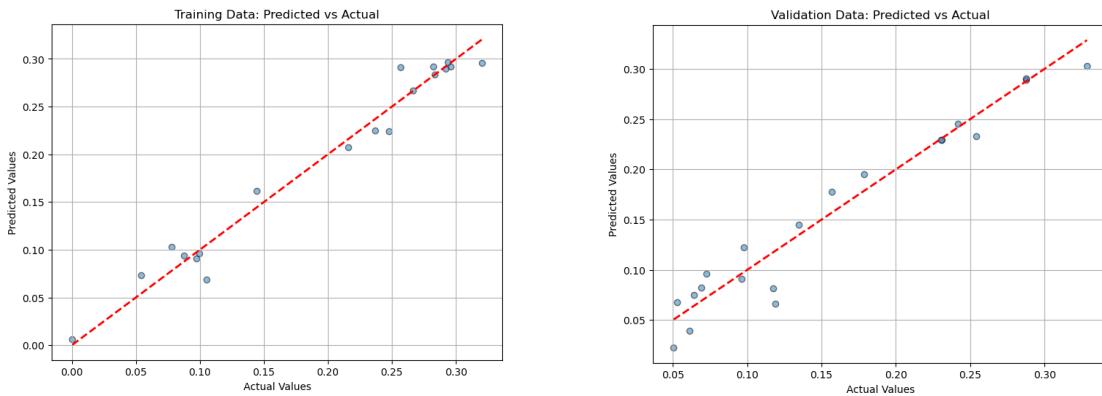


Figure 5.4 Weekly speed-loss predicted by the mechanical model

- **MSE:** 0.0004, this value is very small because all the predicted values are small, and squaring these small errors further reduces their magnitude.

The scatter plots in figure 5.5 show that model's predictions are consistent with the observed data and the absence of strong outliers contributes to the reliability of the model's performance metrics.



(a) Scatter plot of the real speed-loss on the x axis and the predicted one on the y axis on the training set

(b) Scatter plot of the real speed-loss on the x axis and the predicted one on the y axis on the validation set

Figure 5.5 Comparison between the scatter plots of training and validation sets for the KRR

The final test results are shown in figure 5.6. In this case, all the samples before the black horizontal line are used for the training and validation phase, and the last four samples were predicted. The lines corresponding to the real and training samples are close to one

another, as expected from the low sparsity shown in the scatter plots, and the final predictions accurately capture the stagnation trend that characterizes the last period. Based on this, it has been proven that the model works with a good degree of precision and generalization capabilities.

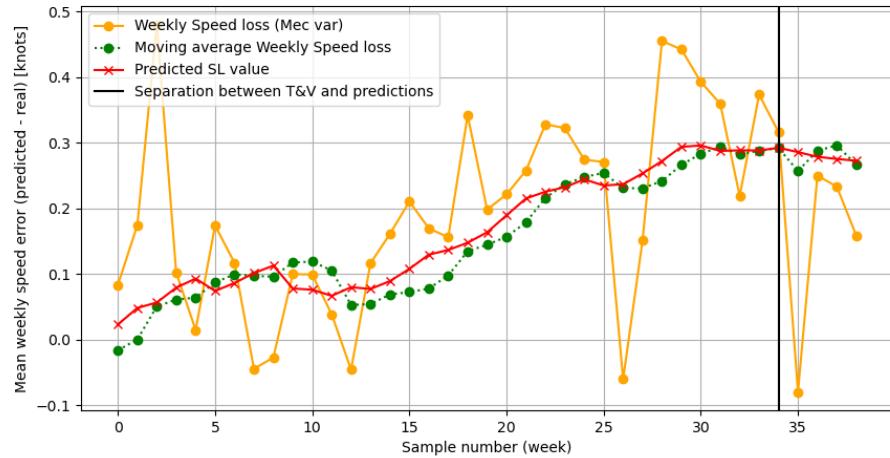


Figure 5.6 Mean weekly speed-loss comparison between the real and predicted one, the last 4 sample are predicted by the model after training and validation

Chapter 6

Future development

6.1 Discussions

To develop a functional model, several limitations were applied, including those mentioned in Section 3.4. The major limitations are as follows:

- **Moving Average:** The value predicted by the biological model is not the actual speed loss but rather the moving average of it. While this technique helps smooth the data, it is still an approximation due to the inherent imprecision of the mechanical model.
- **Dataset Size:** The dataset is relatively small, with only 40 samples, which is insufficient to fully capture the complexity of a phenomenon as intricate as biofouling.
- **Dataset Information:** The dataset lacks information on maintenance and cleaning events. Including such data would improve understanding of how and why trends change.

6.2 Suggestions for improvement

Several aspects can be improved or added to enhance the final combined model:

- **Accuracy:** Both models could be improved in terms of accuracy. For the mechanical model, incorporating additional metrics could better characterize the phenomenon. Examples include fuel consumption, fuel quality, time elapsed since the last major maintenance or component replacements, and more comprehensive weather data such as waves and swell. For the biological model, accuracy could be enhanced with more

data or by including additional samples across different cleaning events and years to capture various conditions.

- **More Biological Features:** Several parameters were excluded from this study for practical reasons but could be added in future models. Examples include water salinity, pH, and nutrient density.
- **Time Horizon:** For the biological model, if additional samples are available, the time horizon could become months instead of weeks, or, if in regions with rapid fouling growth, it could be shortened.
- **Speed-Loss changes:** The current approach calculates speed loss based on the distribution of biological variables starting from the first week not used in the training of the mechanical model until the chosen week. A more detailed analysis could involve comparing speed loss relative to the previous week or month instead of just the clean state, providing insights into growth patterns in specific zones and periods.
- **Fouling Growth:** If periodic monitoring of fouling growth on the ship is feasible, the target could shift from speed loss to estimating the growth size or equivalent surface roughness.
- **Real weekly speed-loss:** The real speed loss for each week could be used instead of the moving average if the predictions from the mechanical model are more precise.

Chapter 7

Conclusions

This work demonstrates that, even with a limited number of mechanical features —only three, all related to shaft measurements— and the wind measurement, speed loss over time can still be predicted with an acceptable margin of error. From this, a new conceptual model has been developed that is not only applicable to a single ship but can be generalized to any ship of similar size. The results allow for the derivation of additional metrics relevant to shipowners, such as increases in fuel consumption, component degradation, and optimal dry-docking schedules. As for the final speed-loss in the chosen period the values is 0.3 knots after 40 weeks and can be considered a reasonable number. After that a new cleaning event was probably scheduled and it reduced the speed-loss again to near zero values.

References

[cro]

- [2] Abarzua, S. and Jakubowski, S. (1995). Biotechnological investigation for the prevention of biofouling. i. biological and biochemical principles for the prevention of biofouling. *Marine Ecology Progress Series*, (123):301–312.
- [3] Atlar, M., Yeginbayeva, I., Turkmen, S., Demirel, Y., Carchen, A., Marino, A., and Williams, D. (2018). A rational approach to predicting the effect of fouling control systems on “in-service” ship performance. *GMO SHIPMAR*, 24(213):5 – 36. Cited by: 13.
- [4] C. Banks, O. Turan, A. I. G. T. S. I. C. S. and Tian, X. (2013). Understanding ship operating profiles with an aim to improve energy efficient ship operations. In *In Low Carbon Shipping Conferenc*.
- [5] Carchen, A., Atlar, M., Turkmen, S., Pazouki, K., and Murphy, A. J. (2019). Ship performance monitoring dedicated to biofouling analysis: Development on a small size research catamaran. *Applied Ocean Research*, 89:224–236.
- [6] Cipollini, F., Oneto, L., Coraddu, A., Murphy, A. J., and Anguita, D. (2018). Condition-based maintenance of naval propulsion systems with supervised data analysis. *Ocean Engineering*, 149:268–278.
- [7] Coraddu, A., Lim, S., Oneto, L., Pazouki, K., Norman, R., and Murphy, A. J. (2019a). A novelty detection approach to diagnosing hull and propeller fouling. *Ocean Engineering*, 176:65–73.
- [8] Coraddu, A., Oneto, L., Baldi, F., and Anguita, D. (2015). Ship efficiency forecast based on sensors data collection: Improving numerical models through data analytics. In *OCEANS 2015 - Genova*, pages 1–10.
- [9] Coraddu, A., Oneto, L., Baldi, F., Cipollini, F., Atlar, M., and Savio, S. (2019b). Data-driven ship digital twin for estimating the speed loss caused by the marine fouling. *Ocean Engineering*, 186:106063.
- [10] D. Owen, Y. K. Demirel, E. O. T. T. and Incecik, A. (2018). Investigating the effect of biofouling on propeller characteristics using cfd. *Ocean Engineering*, 159(3):505–516.
- [11] D. Uzun, Y. K. Demirel, A. C. and Turan, O. (2019). Time-dependent biofouling growth model for predicting the effects of biofouling on ship resistance and powering. *Ocean Engineering*, 191:106 432.

- [12] D. Uzun, R. Ozyurt, Y. D. and Turan, O. (2018). Time based ship added resistance prediction model for biofouling. *Marine Design XIII*, 2:971–979.
- [13] Demirel, Yigit Kemal, D. U. Y. Z. H.-C. F. A. H. D. and Turan, O. (2017). Effect of barnacle fouling on ship resistance and powering. *Biofouling*, 33:819–884.
- [14] Drake, J.M. Lodge, D. (2007). Hull fouling is a risk factor for intercontinental species exchange in aquatic ecosystems. *Aquatic Invasions*, (2):121–131.
- [15] Erol, E., Cansoy, C. E., and Özgür Aybar, O. (2020). Assessment of the impact of fouling on vessel energy efficiency by analyzing ship automation data. *Applied Ocean Research*, 105:102418.
- [16] ESA, Europea Commission, C. Era-5 chlorophyll. <https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-ocean-colour?tab=overview>. Accessed: 2024-07-01.
- [17] ESA, Europea Commission, C. Era-5 dataset. <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview>. Accessed: 2024-06-30.
- [18] et al, J. A. (2010). Drag force and surface roughness measurements on freshwater biofouled surfaces. *Biofouling*, (4):487–496.
- [GEF] GEF. Large marine ecosystems.
- [20] H. Lee, L. E. and Yang, S. (1987). Kinetics and bioenergetics of light-limited photoautotrophic growth of spirulina platensis. *Biotechnology and bioengineering*, 29(7):832–843.
- [IMO] IMO. Impact of ships' biofouling on greenhouse gas emissions.
- [22] IMO (2008). International convention on the control of harmful anti-fouling systems on ships. In *International convention on the control of harmful anti-fouling systems on ships*.
- [23] J. C. Goldman, W. J. O. and Jenkins, D. (1974). The kinetics of inorganic carbon limited algal growth,. *Journal (Water Pollution Control Federation)*, page 554–574.
- [24] J. W. Costerton, G. G. G. and Cheng, K.-J. (1978). How bacteria stick. *Scientific American*, (238):86–95.
- [Kane] Kane, D. Marine vessel environmental performance (mvep) assessment guide. energy efficiency: Hull and propeller operations and maintenance.
- [26] Karagiannidis, P. and Themelis, N. (2021). Data-driven modelling of ship propulsion and the effect of data pre-processing on the prediction of ship fuel consumption and speed loss. *Ocean Engineering*, 222:108616.
- [27] Kenneth M. Brown, D. C. S. (1998). Effects of seasonality, length of immersion, locality and predation on an intertidal fouling assemblage in the northern gulf of mexico. *Biotechnology and bioengineering*, 225:107–121.

- [28] Lagerström, Maria, A.-L. W. D. R. O.-L. G. A. I. L. and Ytreberg, E. (2022). Are silicone foul-release coatings a viable and environmentally sustainable alternative to biocidal antifouling coatings in the baltic sea region? *Marine Pollution Bulletin*, 184.
- [29] Melo, L. and Bott, T. (1997). Biofouling in water systems," experimental thermal and fluid science. *Marine fouling organisms and their use in antifouling bioassays*, 14:375–381.
- [30] Molnar, J.L., G.-R. R. C . S.-M. (2008). Frontiers in ecology and the environment. *Trends in Ecology and Evolution*, (6):485–492.
- [31] P. Darvehei, P. A. B. and Moheimani, N. R. (2018). Model development for the growth of microalgae: A review. *Renewable and Sustainable Energy Reviews*, (97):233–258.
- [32] Qiu, J.-W., Q. P.-Y. (1998). Combined effects of salinity and temperature on juvenile survival, growth and maturation in the polychaete *hydroides elegans*. *Marine Ecology Progress Series*, (168):127–134.
- [33] Reuben P. Keller^{1*}, John M. Drake², M. B. D. and Lodge, D. M. (2011). Linking environmental conditions and ship movements to estimate invasive species transport across the global shipping network. *Diversity and Distributions*, 17:93–102.
- [34] Schimanski, K., Piola, R., Goldstien, S., Floerl, O., Grandison, C., Atalah, J., and Hopkins, G. (2016). Factors influencing the en route survivorship and post-voyage growth of a common ship biofouling organism, *bugula neritina*. *Biofouling*, 32:969–978.
- [35] Schultz, M. (2007). Effects of coating roughness and biofouling on ship resistance and powering. *Biofouling*, (23):331–341.
- [36] Senteris, A., Kanellopoulou, A., and Zaraphonitis, G. (2019). *A machine learning approach to assess vessel performance based on operational profile*, pages 496–502.
- [37] Song, C. and Cui, W. (2019,). Review of underwater ship hull cleaning technologies. *Journal of marine science and application*, 19(3):93–102.
- [38] Sylvester, F. MacIsaac, H. (2010). Is vessel hull fouling an invasion threat to the great lakes? *Diversity and Distributions*, (16):132–143.
- [39] Takahashi, T., S. S. C. D. G. J. H. C. N. T. S. C. M. (2014). Climatological distributions of ph, pco₂, total co₂, alkalinity, and caco₃ saturation in the global surface ocean, and temporal changes at selected locations. *Marine Chemistry*, (164):95–125.
- [40] V. Thiagarajan, T. H. and Qian, P.-Y. (2003). Combined effects of temperature and salinity on larval development and attachment of the subtidal barnacle *balanus trigonus darwin*. *Journal of Experimental Marine Biology and Ecology*, 287:223–236.
- [41] Valchev, I., Coraddu, A., Kalikatzarakis, M., Geertsma, R., and Oneto, L. (2022). Numerical methods for monitoring and evaluating the biofouling state and effects on vessels' hull and propeller performance: A review. *Ocean Engineering*, 251:110883.
- [42] Wilson, J.R.U., D. E. P. P. L. A . R. D. (2009). Something in the way you move: dispersal pathways affect invasion success. *Trends in Ecology and Evolution*, (24):136–144.

- [43] Y. K. Demirel, M. Khorasanchi, O. T. and Incecik, A. (2013). On the importance of antifouling coatings regarding ship resistance and powering. In *International Conference on Technologies, Operations, Logistics and Modelling for Low Carbon Shipping*.