MASTER’S DEGREE PROJECT IN ENGINEERING MATHEMATICS

Stability Analysis of Hydrodynamic Performance Indicator Based on Historic Data Sets

by

Canan Özel Kennedy

MAA044 — Master’s Degree Project in Mathematics
MAA044 — Master's Degree Project in Mathematics

Date of presentation:
01 June 2022

Project name:
Stability Analysis of Hydrodynamic Performance Indicator Based on Historic Data Sets

Author(s):
Canan Özel Kennedy

Version:
8th June 2022

Supervisor(s):
Christopher Engström
Martin Ericson Borgh

Examiner:
Linus Carlsson

Comprising:
30 ECTS credits
Abstract

This paper presents a stability analysis of the sensor data which is collected by QTAGG from a large ocean going ship and using the stability results, introduces some information about how the measurements that come from the sensors can be improved and how reliable they are. In the theoretical part, some background information is given mainly based on British Standard(BS) ISO 19030 which was published in November, 2016. This source basically includes some information about the measurement of changes in hull and propeller performance of a vessel. Using the theoretical information, in the implementation part, the necessary methods are implemented in python programming language on a real life data set of a vessel which is given from QTAGG company. To measure the stability of the parameters in the data, we loosen the filters of the parameters and observe how they respond to the technical changes. In order to understand how loosen the filters can be made, a reference speed-power curve is created by using a curve fitting method, and after creating a performance indicator by utilizing the reference curve, Anderson-Darling and Shapiro-Wilk tests are used to measure the stability of the performance indicator. Besides these numerical tests, some visual methods such as Q-Q plot and histogram plot are also used in this process. Finally, we could provide stability results by using both our theoretical knowledge and the practical implementation.

Keywords: Stability analysis, Anderson-Darling test, Shapiro-Wilk test, curve fitting, Q-Q plot, British Standard(BS) ISO 19030.
Acknowledgements

I would like to express my deep gratitude to Dr. Christopher Engström for inspiring and encouraging me by sharing his expertise in mathematics and statistics and valuable support during this study as well as during my master programme study years at MDU. Moreover, I would like to express my sincere gratitude to my supervisor at QTAGG, Martin Ericson Borgh for sharing his precious knowledge in ships and marine technology with me, and also for his continuous support and motivation during the period of time that I have spent on my master thesis. Furthermore, I would like to thank Tomas Lindqvist for giving me this big opportunity of being able to a part of QTAGG, and having a great work experience in the very welcome environment. Last but not least, I would like to thank my family, my friends, and especially my husband for providing me an unprecedented support, continuous encouragement and their unfailing inspiration throughout my study years at MDU, especially during the challenging time period of the preparation of this master thesis.
## Contents

**Acknowledgements**

**Glossary**

1 **Introduction**
   - 1.1 Literature review ..................................................... 8
   - 1.2 Problem formulation ................................................ 9
   - 1.3 Method .................................................................. 9
   - 1.4 Data description ...................................................... 10

2 **Theoretical background**
   - 2.1 Introduction ................................................................ 12
   - 2.2 Data preparation ......................................................... 12
     - 2.2.1 Outlier elimination and data validation ..................... 12
       - 2.2.1.1 Outlier elimination ............................................. 12
       - 2.2.1.2 Data validation ............................................... 16
       - 2.2.1.3 Correction for environmental factors .................. 16
       - 2.2.1.4 Calculation of performance values ..................... 17
   - 2.3 Performance Indicators .............................................. 17
     - 2.3.1 General information for performance indicators .......... 17
     - 2.3.2 Calculation of Performance Indicators .................... 18
       - 2.3.2.1 Determination of reference conditions ............... 18
       - 2.3.2.2 Establishment of reference period and evaluation period . 22
       - 2.3.2.3 Calculation of Performance indicator .................. 22
       - 2.3.2.4 Accuracy of PI ................................................. 23
   - 2.4 Reference speed-power curve ..................................... 23
     - 2.4.1 Creating reference speed-power function ............... 24
       - 2.4.1.1 Curve fitting .................................................. 24
       - 2.4.1.2 Polynomial regression analysis ......................... 25
       - 2.4.1.3 Ordinary least squares method ......................... 25
   - 2.5 Speed loss analysis .................................................. 26
   - 2.6 Stability analysis ..................................................... 26
     - 2.6.1 Anderson-Darling test ......................................... 27
     - 2.6.2 Shapiro-Wilk test .............................................. 28
Glossary

**Data point**  The combination of a unique identifier (UI) and a complete set of data from all signals shall be referred to as a “data point”.

**Dry docking**  Bringing the ship onto dry land to maintain, repair and/or retrofit the parts of the hull that are submerged while the ship is in service.

**Dry docking performance**  Period between two consecutive dry-dockings.

**Evaluation period**  Period in time after the reference period used for comparing the baseline.

**Reference conditions**  Set of comparable conditions of environmental and/or operational factors or a set of ranges of such conditions.

**Reference period**  Period in time of a certain length used to establish a baseline.

**Retrieved data set**  The complete set of retrieved data points shall be referred to as the “retrieved data set”.

List of Figures

2.1 Propulsion system layout .................................................. 15
2.2 Derivation of Wind Speed with Heading .................................. 19
2.3 Derivation of the Total Wind Speed with Heading and Wind Relative Angle . 20

3.1 Histogram of the available metadata ...................................... 32
3.2 Sensor error in speed through water between June 2012 and August 2012 .... 33
3.3 Water depth calculation parameters ....................................... 34
3.4 Reference speed-power curve .............................................. 36
3.5 Reference speed-power function ........................................... 38
3.6 Speed loss performance .................................................... 40
3.7 Speed loss cumulative sum daily .......................................... 41
3.8 Average speed loss difference for the original filters .................... 42
3.9 Q-Q plot for the original filters .......................................... 43
3.10 Q-Q plot for outlier elimination parameters ............................ 44
3.11 Histogram plot for outlier elimination parameters ........................ 45
3.12 Q-Q plot for true wind speed ............................................ 46
3.13 Q-Q plot for rudder position ............................................. 48
3.14 Histogram plot for rudder position ...................................... 48
3.15 Q-Q plot for engine speed ................................................ 50
3.16 Histogram plot for speed over ground ................................... 51
3.17 Q-Q plot for speed through water ........................................ 52

4.1 Loosening result ................................................................... 53
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>The parameters from sensor data streams</td>
<td>11</td>
</tr>
<tr>
<td>2.1</td>
<td>Anderson-Darling test</td>
<td>28</td>
</tr>
<tr>
<td>2.2</td>
<td>Shapiro-Wilk test</td>
<td>29</td>
</tr>
<tr>
<td>3.1</td>
<td>Anderson-Darling test for original filters</td>
<td>42</td>
</tr>
<tr>
<td>3.2</td>
<td>Shapiro-Wilk test for original filters</td>
<td>43</td>
</tr>
<tr>
<td>3.3</td>
<td>Anderson-Darling test for outlier elimination parameters</td>
<td>44</td>
</tr>
<tr>
<td>3.4</td>
<td>Shapiro-Wilk test for outlier elimination parameters</td>
<td>44</td>
</tr>
<tr>
<td>3.5</td>
<td>Anderson-Darling test for true wind speed</td>
<td>45</td>
</tr>
<tr>
<td>3.6</td>
<td>Shapiro-Wilk test for true wind speed</td>
<td>46</td>
</tr>
<tr>
<td>3.7</td>
<td>Anderson-Darling test for rudder position</td>
<td>47</td>
</tr>
<tr>
<td>3.8</td>
<td>Shapiro-Wilk test for rudder position</td>
<td>47</td>
</tr>
<tr>
<td>3.9</td>
<td>Anderson-Darling test for engine speed</td>
<td>49</td>
</tr>
<tr>
<td>3.10</td>
<td>Shapiro-Wilk test for engine speed</td>
<td>49</td>
</tr>
<tr>
<td>3.11</td>
<td>Anderson-Darling test for speed over ground</td>
<td>50</td>
</tr>
<tr>
<td>3.12</td>
<td>Shapiro-Wilk test for speed over ground</td>
<td>51</td>
</tr>
<tr>
<td>3.13</td>
<td>Anderson-Darling test for speed through water</td>
<td>51</td>
</tr>
<tr>
<td>3.14</td>
<td>Shapiro-Wilk test for speed through water</td>
<td>52</td>
</tr>
<tr>
<td>4.1</td>
<td>Analysis on the modified filters</td>
<td>53</td>
</tr>
<tr>
<td>4.2</td>
<td>Analysis on the data</td>
<td>54</td>
</tr>
<tr>
<td>1</td>
<td>An example of Shapiro-Wilk test in a table</td>
<td>63</td>
</tr>
<tr>
<td>2</td>
<td>An example of Anderson-Darling test in a table</td>
<td>64</td>
</tr>
<tr>
<td>3</td>
<td>Detailed of an example of Anderson-Darling test in a table</td>
<td>64</td>
</tr>
<tr>
<td>4</td>
<td>Summary of statistical hypothesis testing against Gaussian Distribution</td>
<td>65</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

During the operational use of a vessel, some deterioration occurs in the ship hull, propeller, and rudder over time and this affects the hull and propeller performance of the ship. Additionally, during each journey, a vessel is exposed to varying resistance additions to the calm weather resistance. These effects include air resistance (wind), wave making and hull movement resistance (waves) and differences in displacement (draught of the ship). All these factors cause a ship to lose some speed during the voyage, compared to the conditions of ideal weather and a well maintained hull and propeller. In this study, the speed loss of the vessel in question is calculated daily and monthly basis, and this speed loss is used as a performance indicator of the vessel. Moreover, due to the fact that the parameter values that we obtain from the sensors cannot be correct all the time as the ship can be affected some technical changes and environmental factors, some filtering methods are applied on the original data to obtain a data that meets the requirements in the ISO standards. Furthermore, to understand how well the signal quality of the current used sensor data streams work for the calculation of hydrodynamic performance indicators, we change the filter limits which are required by ISO standards, and we observe that how the parameters that are available in our data respond to these filter changes, and thus how loosen they can be made is investigated.

1.1 Literature review

According to ISO 19030-1 [6], the hull and propeller performance is the relationship between the power which is required to move the ship through water at a given speed and the condition of the vessel underwater hull and propeller. This performance has a directly impact on fuel consumption of the ship. This is because some fouling occurs in ship hull due to some roughness and degrading in the hull. Therefore, the fouling causes more drag in the water and thus the ship resistance increases over time. Due to the increased resistance, the vessel loses some speed and more power must be delivered to the propeller to sail the ship at the desired speed. This means that more fuel is needed to propel the ship, and thus the fuel consumption increases. There are many ways to increase the energy efficiency of the ship such as increasing shaft efficiency, rotative efficiency, engine speed efficiency, and propulsive efficiency or hull and propeller efficiency. In this study, we have worked on the hull and propeller efficiency. For
example, the improvement of engine speed efficiency is covered to obtain an energy-efficient shipping \[22\]. In order to understand how the hull and propeller performance can be increased, thus the fuel consumption of the ship can be decreased, we investigate how our sensor values are affected by the technical changes and environmental factors that the vessel encounters during the voyage. However, in this study we are only interested in understanding the hull and propeller performance by observing the reaction of the sensor values against the technical and environmental factors, and understanding the ways of decreasing fuel consumption is not studied in this paper. To be able to evaluate the performance, we create a performance indicator (PI) of the vessel depending on the speed loss performance of the ship and it is evaluated by conducting a stability analysis in this study. The ship PI is evaluated by many researchers in the literature such as \[1\] and \[10\]. To evaluate the PI, one must follow the ISO 19030-2 \[7\], and in the Section \[2.3.2\] the calculation of the PIs is covered by following the produce in the ISO 19030-2 \[7\]. Moreover, a stability analysis is conducted on the PI by using various sensor values from the data, and to calculate desired parameters for this analysis, some specific formulas and constants are used in this thesis, which are provided by the ISO 19030-1 \[6\] and the ISO 19030-2 \[7\]. However, some alternative ways are also given in the ISO 19030-3 \[8\] for these calculations and measurements, but we do not use this source in this study due to lack of data in the data sets that we have used. In the literature, stability analysis is conducted for various purposes for various types of vessels. For example, a stability analysis is made for a super-large vacuum vessel by using the finite element method \[18\] and a stability analysis is performed on AHTS vessel electrical systems by using dynamic positioning system \[17\]. In this study, a stability analysis is performed by using some statistical methods such as Anderson-Darling and Shapiro-Wilk test as well as creating a Quantile-Quantile plot for a large container vessel. The implementation of these methods on the data that is collected for this large ocean going ship is given in Section \[3.7\] in this report.

1.2 Problem formulation

In this study, we investigate how reliable our sensor values are against the technical changes, and evaluate efficiency enhancement of these changes. There are certain reference conditions as well as outlier elimination and validation methods in ISO 19030-2 \[7\]. By changing the filter limits of these conditions and the methods, we observe how our stability results change, and thus how good the signal quality of the current used sensor data streams work. The main research question is how loose the filters can be made while still maintaining the stability on the remaining data. To achieve this goal, the necessary elimination/filtering and validation methods are implemented in Chapter \[3\].

1.3 Method

Based on the definitions in Chapter \[2\], the filtering methods as well as the speed loss calculation are implemented in python programming language using a real life data set which is described in Section \[1.4\]. First, we have the data points pass through the filters and then a reference speed-
power curve is created by using a curve fitting method, Moreover, the daily and monthly speed loss of the vessel are calculated based on the calculation in ISO 19030-2 [7]. Furthermore, the distribution of the daily speed loss data points in relation to the distribution of the monthly speed loss data line is visualized to see how stable the distribution is when the filters are loosen as well as when the original filters are used. In order to loosen the filters Quantile-Quantile plot and histogram plot are used to see how the data points are distributed visually, and how similar they look with Gaussian distribution. In addition to obtaining some numerical results to assess the stability, Anderson-Darling and Shapiro-Wilk tests are used.

1.4 Data description

The data are gathered with the QTAGG’s systems that already use the existing sensors on the ship. These can be the sensors used for alarm and control systems, or for the navigation of the ship. We use two different data sets in this study, which are sensor logs and metadata respectively. In each sensor data log file, we have roughly 3400 rows and 31 columns. We have the metadata between August 2011 and January 2014, and the sensor data between March 2012 and January 2017. Hence, we can use the data between March 2012 and January 2014. Each log file contains 15 min of data. So that, we have 3400x4 data points for one hour, and it means that we have 3400x4x24 data points for one day, which is equal to 326,400 data points. The other kind of data set is metadata. We have three different metadata files available, and unlike the log files, the metadata is gathered daily instead of 15 minutes time intervals. However, we do not have data for each day consecutively in the metadata. For example, in one of the metadata, we do not have any data points for almost 1.5 months. Thus, in total we have roughly 429 days of available data in the metadata. Therefore, if we multiply this number with the number of available data points in the log files for one day, we can yield 429x326,400, which is equal to 140,033,855 data points.

Moreover, from the metadata which is collected at a ‘daily rate’, or in other words when ’the loading conditions change’, we only use the parameter of mean draught(m) of the vessel. Furthermore, the sensor data streams are collected at 4Hz, and the Table 1.1 is given to show which parameters that we use from this data.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wind relative speed</td>
<td>m/s</td>
</tr>
<tr>
<td>Shaft power</td>
<td>kW</td>
</tr>
<tr>
<td>Main engine power</td>
<td>MW</td>
</tr>
<tr>
<td>Speed through water</td>
<td>kn</td>
</tr>
<tr>
<td>Speed over ground</td>
<td>kn</td>
</tr>
<tr>
<td>Depth below keel</td>
<td>m</td>
</tr>
<tr>
<td>Rudder position</td>
<td>°</td>
</tr>
<tr>
<td>Engine speed</td>
<td>min$^{-1}$</td>
</tr>
<tr>
<td>Wind relative angle</td>
<td>°</td>
</tr>
<tr>
<td>Heading</td>
<td>°</td>
</tr>
</tbody>
</table>

Table 1.1: The parameters from sensor data streams
Chapter 2

Theoretical background

2.1 Introduction

In this section, some background information is given for the methods that we apply in Chapter 3. The main source that we use in this part is British Standards (BS) ISO 19030 [6, 7].

2.2 Data preparation

The data is collected automatically and continuously with QTAGG’s data acquisition system for each parameter 4 times every second. The collected data is stored as raw data along with time stamps. In this study, the collected and stored data is prepared to evaluate the hull and propeller performance of the ship in question. Data preparation consists of retrieving, compiling, outlier filtering and data validation processes. The retrieved and compiled data is given to use in outlier filtering and data validation processes by QTAGG company for this study. After combining the two data set which is described in Section 1.4, the retrieved data set is filtered and validated for the data points in the data, as they are explained in following section. When this process is done, the correction for environmental factors must be made according to ISO-19030-2 [7]. However, due to lack of necessary parameters, this process is ignored in this study.

2.2.1 Outlier elimination and data validation

According to ISO 19030-2 [7], the outliers and missing values must be marked invalid for the retrieved data set. In the following subsections, both the outlier elimination and data validation processes are explained with the methods that are used.

2.2.1.1 Outlier elimination

We conduct an outlier elimination process to eliminate the values that are unrelated and spurious to the others. To eliminate outliers, non-overlapping blocks are created for 10 minutes periods, and every parameter in the data is filtered according to Chauvenet’s criterion which
is explained in the next section. However, we only use 'speed through water' and 'delivered power' parameters in the outlier elimination process, as we only need to use these parameters for creating the reference curve, which is a curve used for finding the expected speed through water at the corrected delivered power and at the measured displacement and trim [7], and thus, calculating the speed loss. If the data for one parameter in a 10 minutes block is detected as an outlier, then the complete data point is marked as invalid. If the data is missing for one or more parameters, then the data point must be marked as invalid as well [7].

**Chauvenet’s criterion**
Chauvenet’s criterion is a statistical test that is used to detect outliers in a given data. If the expected number of measurements is at least as bad as the suspected measurement, which is less than 0.5, in that case, the suspected measurement is rejected [14]. As we can see in [14], a 50% of chance is given to survive for each value. Moreover, we can also say that the criteria assumes that there must be as many points closer to the mean as there are further away. So, we must use the threshold of 0.5 in this study, as the outlier elimination method uses this criterion and this criterion gives the threshold as 0.5.

For the $N$ data in a data block with each row $d_i$, the probability of the occurrence of any value $d_i$ is computed with the following formula:

$$P(d_i) = \text{erfc} \left( \frac{\Delta_i}{\sigma \cdot \sqrt{2}} \right)$$

where,

- $P(d_i)$ is the probability occurrence of $d_i$.
- erfc is the complementary error function.
- $\Delta_i$ is the distance between the values $d_i$ and the mean $\mu$.
- $\sigma$ is the standard error of the mean.

**Complementary error function**
The complementary error function, erfc, is the residual area under the tails of a distribution [12]. The main idea is to understand that how close the values are to the center. The further the values away from the center, the smaller values the function gets. Thus, it is clear that if it goes to infinity, the value of the function is always 0. Note that complementary error function assumes a normality in the data. If the data points are widely separated, all data points are rejected which means that erfc is useless, when the data is not normally distributed.

**Calculation of the mean**
For the data that are not measured as angles, the mean $\mu$ for the $N$ data in a data block with values $d_i$ are computed according to the following formula:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} d_i$$
For the data that are measured as angles:

\[ \mu = \text{atan2} \left( \frac{\sum_{i=1}^{N} \sin(d_i)}{N}, \frac{\sum_{i=1}^{N} \cos(d_i)}{N} \right) \]

where, atan2 is defined as the angle in the Euclidean plane, given in degrees, between the positive x axis and the ray from the origin to the point \((x, y)\), e.g. see in [7].

**Computation of standard error of the mean**

\( \Delta_i \), the distance between the values \( d_i \) and the mean \( \mu \) is calculated, if data is not measured as angles:

\[ \Delta_i = |(d_i - \mu)| \]

When the data is measured in angles, we let \( r_i = |d_i - \mu| \mod(360^\circ) \) and define,

\[ \Delta_i = \begin{cases} 
360^\circ - r_i, & \text{if } r_i > 180^\circ, \\
r_i, & \text{otherwise.}
\end{cases} \]

where, \( \mod(k, l) \) is the modulo of two entries \( k \) and \( l \).

The standard error of the mean is computed with the following:

\[ \sqrt{\frac{1}{N} \sum_{i}^{N} \Delta_i^2} \] (2.2)

Using [Equation 2.1] a datum is considered outlier, if the following is fulfilled:

\[ \text{erfc} \left( \frac{\Delta_i}{\sigma \cdot \sqrt{2}} \right) \cdot N < 0.5 \] (2.3)

where, \( N \) is the available data in a data block.
We use the following two parameters in outlier elimination process:

**Speed through water**
Ship’s speed through water for a given set of service and loading condition, it is denoted as $V$. It is retrieved directly from sensor log files, thus there is no any extra calculations for this parameter.

**Delivered power**
Delivered power is the power which is delivered to the propeller of the ship thus, it is also called propeller power, and denoted by $P_d$. It is calculated based on where the shaft is placed. For the ship in question, the following formula is given to calculate the delivered power.

$$P_d = P_m \cdot 1000 - P_s / 0.96 \quad (2.4)$$

$P_d$ is delivered power in kW.
$P_m$ is main engine power in MW.
$P_s$ is shaft power in kW.

In the Figure 2.1 the propulsion system layout of the ship in question is shown. The reason why we need to subtract the shaft power ($P_s$) from the main engine power ($P_m$) values is that as we can observe in the Figure 2.1 the shaft power meter is placed on the shaft line before the shaft generator, and thus measures the main engine power ($P_m$) and not the power delivered to the propeller ($P_d$). When the shaft generator is delivering electric power ($P_s$) to the ships electric systems, this power is added to the power delivered to the propeller ($P_d$). Since we are only interested in the power delivered to the propeller ($P_d$), we must subtract the main engine power ($P_m$) with the mechanical power needed to produce the electrical power ($P_s$).

If the ship was not equipped with a shaft generator, this step would not be needed and we would instead use the power measured by the shaft power meter directly, i.e. $P_d = P_m$.

Figure 2.1: Propulsion system layout
2.2.1.2 Data validation

The data set must split in non-overlapping, consecutive 10 minutes blocks for data validation. After calculating the statistical parameters as they are calculated for outlier elimination in Section 2.2.1.1, the validation elimination process is held. Unlike the outlier elimination process, for invalid blocks, all data points in a 10 minutes block are marked as invalid in the data validation process. To clarify, 'If the standard error of the mean in Equation 2.2 of any of the parameters is larger than the thresholds which are specified for each parameter, the 10 min block is invalid for all parameters' [7]. The related parameters with the threshold values are given in the following:

- rpm(engine speed): $3 \text{min}^{-1}$;
- speed through water: $0.5 \text{kn}$;
- speed over ground: $0.5 \text{kn}$;
- rudder angle: $1^\circ$.

The reason why we validate the data by choosing the parameters above is that we expect the vessel ideally in steady state condition, which means that there is no wave resistance and wind resistance, and to be able to fulfil the steady state conditions, we have the strict threshold values for the specified parameters. Furthermore, since we expect the steady state conditions, we want to compare the available data with the ideal weather and well-maintained hull and propeller conditions. Moreover, the reason why we must drop a block of data instead of filtering the invalid data point is that if the validation test fails the whole ten-minute block, it can be suspected to not to be run in enough steady state conditions in terms of ship manoeuvring and sea state – therefore the full time block must be rejected. If the standard error of the mean is larger than the thresholds, then it means that it is not in the steady state during that period.

2.2.1.3 Correction for environmental factors

The delivered power entries must be corrected for wind resistance for the data points of the validated data set and the new data set is referred as the 'corrected data set' [7]. However, due to lack of parameters for this correction, we cannot correct the data set for wind resistance. The parameters that we lack for wind correction is given in the following:

- $A$ is the transverse projected area in current loading condition in $\text{m}^2$;
- $C_{rw}$ is the wind resistance coefficient, dependent on wind direction of relative wind $\psi_{wr,ref} [-]$;
- $C_{0w}$ is the wind resistance coefficient for head wind ($0^\circ$ wind direction[-]).

Note that if the ship had been a RoPax ship, the effect could have been calculated from the actual data. However, for the ship in question, this is not possible due to all the parameters above vary significantly with loading conditions.
2.2.1.4 Calculation of performance values

Performance values (PVs) are defined as the speed loss compared with a reference speed-power relation. A PV must be calculated for every data point in the corrected data set. Thus, the 'prepared data set' which is the union of performance values and the corrected data set are created [7]. The following formula is given to calculate the speed loss:

\[ V_s = V_m - V_e \]

where,

- \( V_s \): speed loss.
- \( V_m \): measured vessel speed through water.
- \( V_e \): expected speed through water.

**Expected speed**
The expected speed through water is obtained by reading the speed-power reference curve at the measured trim and displacement and at the corrected delivered power.

**Measured speed**
The measured speed is the speed which is obtained by the admiralty formula, and it is given in Equation 2.15

2.3 Performance Indicators

In this section, the different performance indicators are given, and based on the PVs from the prepared data set, the calculation of these indicators are explained.

2.3.1 General information for performance indicators

Performance Indicators (PIs) are the indicators to determine the effectiveness of hull and propeller repair, retrofit and maintenance activities of the vessel in question [7]. The measurements of ship-specific changes in hull and propeller performance are used to determine the specified effectiveness. There are 4 kinds of PIs, and they are given in the following:

- Dry-docking performance;
- In service performance;
- Maintenance trigger;
- Maintenance effect.
However, in this study due to lack of information of what happened to the ship at a specific time period, such as when the hull was cleaned and when the docking was happened, we only use In service performance, which is the effectiveness of the underwater hull and propeller solution over the full dry docking interval [6], and this is used as a PI.

2.3.2 Calculation of Performance Indicators

The following steps are applied for the calculation of PIs:

- determination of reference conditions;
- establishment of reference and evaluation period;
- extracting the Performance Values(PVs) from the values that meet reference conditions for the determined reference and evaluation period;
- calculation of PI;
- assessing the accuracy of the PI.

2.3.2.1 Determination of reference conditions

To be able to calculate the speed loss of the vessel over a period of time, there are certain reference conditions to be met for all 'Performance Indicators(PIs)'. These conditions are for the parameters of 'true wind speed', 'water depth' and 'water temperature', and they must be met simultaneously.

**True wind speed**

Since the parameters that are given Section 2.2.1.3 do not exist in the data to correct the wind resistance, and according to ISO standards, calculation of the true wind speed parameter is directly related to these parameters, we cannot calculate the true wind speed in the way that ISO standards suggest. Therefore, we use an alternative way to make this calculation. The alternative way is given in the following:

**Derivation of the wind speed of the ship**

\( w_{sx} \) and \( w_{sy} \) are the wind course as they are illustrated below.

\[
\begin{align*}
   w_{sx} &= -v_g \cdot \sin(\psi_0) \\
   w_{sy} &= -v_g \cdot \cos(\psi_0)
\end{align*}
\]

where,

- \( w_{sx} \) is the wind speed of the ship in x direction (m/s);
- \( w_{sy} \) is the wind speed of the ship in y direction (m/s);
- \( v_g \) is the speed over ground (m/s);
- \( \psi_0 \) is the vessel heading (°).
Derivation of the vessel total wind speed

\[
\begin{align*}
    w_x &= v_{wr} \cdot \cos(270^\circ - \psi_0 - \psi_{wr}) \quad (2.6a) \\
    w_y &= v_{wr} \cdot \sin(270^\circ - \psi_0 - \psi_{wr}) \quad (2.6b)
\end{align*}
\]

where,

- \( w_x \) is the total wind speed in x direction (m/s);
- \( w_y \) is the total wind speed in y direction (m/s);
- \( v_{wr} \) is wind relative speed (m/s);
- \( \psi_0 \) is vessel heading (°);
- \( \psi_{wr} \) is the direction of the relative wind (°).

Calculation of true wind speed

\[
\begin{align*}
    w_x &= w_{x x} + w_{wx} \\
    w_{wx} &= w_x - w_{x x}
\end{align*}
\]

where,
$w_x$ is the total wind speed in x direction (m/s);
$w_{sx}$ is the wind speed of the ship in x direction (m/s);
$w_{wx}$ is the wind speed in x direction (m/s).

From [Equation 2.5a](#) and [Equation 2.6a](#), we have:

$$w_{wx} = v_{wt} \cdot \cos(270^\circ - \psi_0 - \psi_{wr}) + v_g \cdot \sin(\psi_0) \tag{2.7}$$

And, in the same way, we can obtain $w_{wy}$,

$$w_y = w_{sy} + w_{wy}$$

$$w_{wy} = w_y - w_{sy}$$

where,

$w_y$ is the total wind speed in x direction (m/s);
$w_{sy}$ is the wind speed of the ship in x direction (m/s);
$w_{wy}$ is the wind speed in x direction (m/s).

From [Equation 2.5b](#) and [Equation 2.6b](#), we have:

$$w_{wy} = v_{wt} \cdot \sin(270^\circ - \psi_0 - \psi_{wr}) + v_g \cdot \cos(\psi_0) \tag{2.8}$$
Therefore, by utilizing from Figure 2.2 and Figure 2.3, and using Equation 2.7 and Equation 2.8, we can obtain the following equation:

\[ v_{wt} = \sqrt{w_{wx}^2 + w_{wy}^2} \]  

(2.9)

where, \( v_{wt} \) is the true wind speed (m/s).

To meet the related condition in ISO 19030-2 [7], the true wind speed must be between 0 m/s and 7.9 m/s.

**Water depth**

We take 'mean draught’ parameter from the metadata, and 'depth below keel’ parameter from the sensor log files and calculate the water depth value by

\[ h_w = T_M + h_b \]  

(2.10)

where,

- \( h_w \) is water depth (m);
- \( T_M \) is the draught at midship or mean draught (m);
- \( h_b \) is the depth below keel (m).

The water depth is valid, when water depth, \( h_w \) is greater than the larger of the values obtained from the Equation 2.11 and Equation 2.12, see in [7].

\[ h = 3 \sqrt{B \cdot T_M} \]  

(2.11)

\[ h = 2.75 \sqrt{\frac{V_s^2}{g}} \]  

(2.12)

where,

- \( h \) is the water depth (m);
- \( B \) is the ship breadth (m);
- \( T_M \) is the mean draught (m);
- \( V_s \) is the ship speed (m/s);
- \( g \) is gravitational acceleration (9.80665 m/s²), see in [7].

**Water temperature**

Due to absence of 'water temperature’ knowledge in the files we have, the reference condition for the related parameter is ignored. The required condition is that the water temperature must be above +2°C, and if there is no indication that the vessel is trading in ice’ [7]. However, we already know that the water temperature cannot be lower than +2°C, in the seas that this vessel visited. The coldest sea that the vessel passed during its journey was Mediterranean sea, because the other ones are Arabic sea and Red sea which are warmer than Mediterranean sea.
on average. In Mediterranean sea, the lowest surface temperature can be found in the north of Adriatic, where the mean water temperature falls to 5°C in the sea. Ice can be observed there during the winter [2]. However, the ship did not pass by the north of Adriatic. Therefore, we can conclude that we eliminate no data with this filter, and to understand which seas that the vessel visit, we check the GPS information of the vessel.

Moreover, ISO standards require that ‘Delivered power has to be within the range of power values covered by the available speed-power reference curves’ [7]. In this study, this requirement is automatically fulfilled, since the speed-power reference curve is created from the actual measurement data.

2.3.2.2 Establishment of reference period and evaluation period

Due to the lack of the information of the vessel activities such as when the hull is cleaned and when the docking is happened, we are not able to establish a reference period. Nevertheless, the evaluation period is taken as one month in this study.

2.3.2.3 Calculation of Performance indicator

We first calculate the daily speed loss by the following formula:

\[ V_d = \sum_{i=1}^{N} (V_m - V_e)_i \]  \hspace{1cm} (2.13)

where,

\( V_d \) is daily speed loss;
\( V_m \) is measured speed loss;
\( V_e \) is expected speed loss;
\( N \) is the number of data available for the day.

Note that we calculate the daily speed loss value, \( V_d \), based on the performance value in Section 2.2.1.4. Then, the monthly speed loss is calculated by taking the average of the daily speed loss values. The formula for this calculation is given in the following.

\[ V_p = \sum_{j=1}^{m} (V_d)_j \]  \hspace{1cm} (2.14)

where,

\( V_p \) is monthly speed loss;
\( V_d \) is daily speed loss;
\( m \) is the number of data points for daily speed loss in a month period.

We consider the average of the \( V_p \) values from Equation 2.14 as our PI in the speed loss analysis.
2.3.2.4 Accuracy of PI

In order to find out on what accuracy can be achieved while meeting all the filtering/elimination and data validation requirements, the accuracy of PI is estimated. To calculate the accuracy, uncertainty quantifications shall be calculated by using Monte Carlo simulation method according to ISO 19030-1 [6]. To apply this simulation method on the data, we need speed-power model and the simplified digital twin model of the ship for each of the following:

- Simulated ship specifications,
- Simulated operational profile,
- Simulated true propulsion power.

Due to the fact that we lack the speed-power model and the simplified digital twin model of the ship, we cannot obtain the specified simulations. Thus we cannot create the pdf(probability density function) which represents the uncertainty of speed loss performance indicator(%) that is provided for each time step(daily or 1/15s). Thus, we are unable to produce the combined effect of each source of uncertainty with Monte Carlo simulation method and we cannot estimate the performance value uncertainty.

Note that there are two other methods to make the uncertainty analysis which are GUM uncertainty framework and analytical methods. However, the analytical method is only valid in the simplest of cases, whereas the GUM framework is valid if the model is linearized and the input of pdfs are Gaussian. Therefore, we are not able to apply none of these methods to measure the accuracy of the PI.

2.4 Reference speed-power curve

During the voyage of the vessel, some deterioration occurs in the hull and the propeller of the vessel, and this leads some decrease in the obtained speed for a certain power. In order to obtain a certain speed, an increase in power will be needed. Therefore, the speed-power curve will be shifted upward in a certain time period. A reference speed-power curve is created to analyse at what speed the ship has in reality and at what power should be delivered from the propeller of the ship to compensate the speed loss due to the environmental factors.

The curve is created by following the steps below:

- After all filtering process is done, sorting bins are created with each bin centered on 0.1 speed resolution, starting at the minimum reference speed and ending at the maximum reference speed we have in the data. So, for example, if we have minimum 5 kn and maximum 30 kn in the data, we have the following speed bins:
  - bin$_1$: 4.95 - 5.05 kn: marked as 5 kn
  - bin$_2$: 5.05 - 5.15 kn: marked as 5.1 kn
- bin3: 5.15 - 5.25 kn: marked as 5.2 kn

- bin251: 29.95 - 30.05 kn: marked as 30 kn

- For all the speed bins, if they are available, 1000 lowest delivered power values are chosen for the corresponding reference speed. However, if there is lower than 1000 delivered power values, then we take that available number of delivered power values into the calculation. The reason why we choose the lowest 1000 delivered power values instead of choosing the mean or the median of delivered power values is that the vessel is able to achieve the desired speed through water with the minimum delivered power and we accept these as a part of reference speed.

Admiralty formula is a standard formula for estimating the speed-power curves of the ships [16]. We calculate the reference speed in the curve with this formula. It is defined as

\[ V_2 = V_1 \left( \frac{\Delta_1^{2/3}}{\Delta_2^{2/3}} \right)^{1/3} \]

where,

- \( V_1 \) is the speed at measured displacement;
- \( V_2 \) is the speed at reference displacement;
- \( \Delta_1 \) is the measured displacement;
- \( \Delta_2 \) is the reference displacement.

Note that this formula is used to normalise the measured speeds to take the varying loading conditions (draught/displacement) for different crossings into account.

2.4.1 Creating reference speed-power function

In order to create a reference speed-power function, the minimum delivered power values are taken for each corresponding bin. Since we have chosen the 1000 lowest delivered power values if they are available, taking the minimum for each bin is the most reasonable way to create such series in terms of having a very similar shape with what we have in speed-power curve.

The function is created in a curve fitting process. The curve fitting uses regression analysis, and particularly in this study polynomial regression analysis is used. Moreover the polynomial regression model is fit by using ordinary least squares method. These methods are covered in the following sections.

2.4.1.1 Curve fitting

Curve fitting is the process of generating a mathematical function, or a curve which fits to a series of data points the best [4]. The process involves either regression analysis or
interpolation. In this study, to fit the curve, polynomial regression analysis is used. The process is applied as fitting a polynomial, \( p(x) \) which is given by Equation 2.16 to points \((x, y)\). This returns a vector of coefficient 'p' which minimizes the squared error in the order of \(d, d-1, ..., 0\).

\[
p(x) = p[0]x^d + ... + p[d]
\]

where, \(d\) is the degree of the polynomial.

### 2.4.1.2 Polynomial regression analysis

Polynomial regression analysis is a type of regression analysis where the relationship between the dependent variable \(y\) and the independent variable \(x\) is modelled as a \(n^{th}\) degree polynomial in \(x\). The regression function is denoted by \(E(y|x)\), where the polynomial regression fits a nonlinear relationship between the value of independent variable \(x\), and the corresponding conditional mean of dependent variable, \(y\). The polynomial regression model is given by

\[
y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + ... \beta_n x^n + \epsilon
\]

where, \(\epsilon\) is unobserved random error with mean zero conditioned on a scalar variable \(x\). \(\beta_0, \beta_1, ..., \beta_n\) are the unknown parameters.

This model is fit by using the ordinary least squares method. The method is explained in the following section.

### 2.4.1.3 Ordinary least squares method

Ordinary least squares method (OLS) is a type of linear least squares method in a linear regression model. This method is used to estimate unknown parameters in the regression. The least squares method minimizes the variance of unbiased estimators of coefficients and this process is handled using the conditions of the Gauss-Markov theorem. The best fit minimizes the sum of squared of residuals. A residual can be defined as the difference between an observed value, and the fitted value. The method is an approach in regression analysis to approximate the solution of the systems where there are more equations than unknowns in the set of equations.

Assume that we have a data consisting of \(n\) observations \(\{x_i, y_i\}\)\(^n\). Each observation \(i\) includes a scalar response \(y_i\), and a column vector \(x_i\) of \(p\) parameters(regressors). If we have \(x_i = [x_{i1}, x_{i2}, ..., x_{ip}]^T\), and in a linear regression model, the response variable \(y_i\) can be written a linear function of regressors, which is given in the following:

\[
y_i = x_i^T \beta + \epsilon_i
\]

where,
\( x_i \) is a column vector of \( i^{th} \) observation of all explanatory variables. 
\( \beta \) is a \( p \times 1 \) vector of unknown parameters.
\( \epsilon_i \) represents the error of the i-th observation.

The model can also be written in matrix notation, which is given by

\[
y = X\beta + \epsilon,
\]

where,

\( y \) and \( \epsilon \) are \( n \times 1 \) vectors of the response variables and the errors of the n observations.
\( X \) is a \( n \times p \) matrix of regressors, and is known also design matrix.

Note that here, we emphasize having linear regression model in ordinary least squares method. However, there can be any desired relationship between regressors. For example, if the response depends linearly both on a value and its cube, where we have one regressor whose value is the cube of the another. In this case, we have a cubic value in the second regressor. Nevertheless, the regression can still be considered linear, since it is still linear in the \( \beta \) parameters. For this reason, we can use this method in polynomial regression as well.

### 2.5 Speed loss analysis

To achieve this analysis, firstly, the reference conditions that are defined in Section 2.3.2.1 are fulfilled, and the filtering process is held, as they are explained in Section 2.2.1. Furthermore, a PI is calculated to evaluate the hull and propeller performance of the vessel. In this speed loss analysis, we take the average of all daily speed loss values, \( \bar{V}_d \) in Equation 2.13, and compare these values with our PI, which is calculated in Equation 2.14. The speed loss analysis is made to see how much speed the ship lose over the whole time period of the voyage and how the decrease in speed is performed during a specific time period. We interpret this analysis as the more speed the vessel loose, the less clean hull and the worse performance the hull and propeller have overall.

### 2.6 Stability analysis

Once we make the speed loss analysis, we investigate how stable our results are both when we comply with the filter limits for each parameter in ISO standards, and when we loosen and tighten the filters of the parameters. We conduct this analysis in three ways. Two of them give visual results, while one of them gives some numerical results. We use Quantile-Quantile(Q-Q) plot and a histogram plot to visualize the distribution of speed loss values and compare them with Gaussian distribution. If the distribution is close to Gaussian distribution or roughly symmetric, then we consider it stable, if not, we say that the distribution is not stable and we cannot loosen or tighten the filters to the limit that we have updated, since this cause our PI to
be unstable.
The reason why we compare the distribution of the speed loss values with Gaussian distribution can be explained as if we can see that our errors are roughly Gaussian distributed this means that we can use standard methods for estimation of significance level and critical value or p value. If we end up having some values that can make the standard methods significantly inaccurate, this means that these values can cause overestimation or underestimation of the real values.

2.6.1 Anderson-Darling test

Anderson-Darling test is a statistical test which is used to see if a sample of data is drawn from a population with a specific distribution. The test is based on a comparison of the empirical distribution function of a given sample with the theoretical distribution to be tested [15]. Thus, Anderson-Darling test is a member of group of Goodness-of-Fit statistics which is known as EDF statistics [20]. Goodness-of-fit(GoF) test is used to test if sample data fits a distribution from a certain population, and it implies a comparison of the observed data with the data expected under the model using some fit statistic, or discrepancy measure models, such as residuals, Chi-square or deviance. Due to this reason both Anderson-Darling test and Shapiro-Wilk test is a member of GoF.

There are many distributions that can be tested with this test such as Gaussian distribution, exponential distribution, logistic distribution and Weibull distribution. In this study, we use Anderson-Darling test for Gaussian distribution. Since it assumes that random variable X has continuous cumulative distribution \( F(x; \theta) \), Cumulative Distribution Function(CDF) is used in calculation. Here, \( \theta \) is a vector of one or more parameters which describes the distribution. \( \theta = (\mu, \sigma^2) \) is used for Gaussian distribution [15].

Assume that we have \( n \) observations, \( X_i \), for \( i = 1, ..., n \), of the variable \( X \), and \( X \) is sorted in ascending order as \( X_1 \leq X_2 \leq ... \leq X_n \). To test normality the following formula is given:

The values \( X_i \) are standardized to create new values \( Y_i \), given by

\[
Y_i = \frac{X_i - \mu}{\sigma}
\]

where, \( \mu \) is the mean of the observations and \( \sigma \) is the standard deviation of the observations.

The test statistic is calculated with the standard normal CDF, \( \Phi \) by:

\[
A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1)(\ln \Phi(Y_i) + \ln (1 - \Phi(Y_{n+1-i})))
\]

Anderson-Darling test is defined in the following:

- \( H_0 \): Null hypothesis, the data that follows the specified distribution,
- \( H_a \): Alternative hypothesis, the data do not follow the specified distribution,
- \( A^2 \): Test statistic.
The critical values of the test differs depending on the distribution that is used. If the test statistic is larger than the critical value at the specific significance level, the null hypothesis is rejected. The following example is given to explain how this method works:

<table>
<thead>
<tr>
<th>Test statistic</th>
<th>Critical values</th>
<th>Significance level((\alpha))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.192</td>
<td>[0.538, 0.613, 0.736, 0.858, 1.021]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
</tbody>
</table>

Table 2.1: Anderson-Darling test

In literature, the significance level of an event is the probability that the event can occur by chance. We can consider significance level as the opposite of confidence level, which indicates the degree of confidence that the result does not occur by any sampling errors or by chance [9]. For example, a significance level of 0.15 would mean that even if the data was Gaussian, we would still reject the null hypothesis that it was Gaussian 15% of the time. To support this interpretation, [9] can be useful to review. Furthermore, a significant result which is obtained having a low significance level would indicate a strong argument for the given data to not to come from a Gaussian distribution. Moreover, since the test is designed to reject the null hypothesis which is the data comes from Gaussian distribution in our case, we are interested in the opposite one that the test to give a non-significant result at the largest possible significance level. When we find an interval such that we do not reject the hypothesis, we use an \(\alpha\) value to say that we only need to use 1 - \(\alpha\) of our total data. This is explained in the textbook [5], as in the following:

- The \(\alpha\)-error is referred to as the probability of error, statistical uncertainty, risk measure or safety threshold. The probability of \(\alpha\) error and confidence level \(S\) add up to a value of 1. Hence, the following applies:
  - \(S + \alpha = 1\) or alternatively \(S + \alpha = 100\%\). So, for a normal distribution, \(S + \alpha\) corresponds to the area below the curve of the Gaussian distribution.
  - In this thesis, we assume that the \(\alpha\), which is our significance level, implies that we can reduce the data set with the same percentage that we have in \(\alpha\).
  - In this example, since the test statistic is greater than the critical values at all significance levels, we have to reject the null hypothesis. In other words, the results are significant for all significance levels. Thus, we can say that we have sufficient evidence to reject the null hypothesis. It means that our data is not Gaussian distributed.

### 2.6.2 Shapiro-Wilk test

The Shapiro-Wilk test is a member of group of Goodness-of-Fit statistics which is known as EDF statistics [19]. This test is design to see if a random sample \(x_i, i = 1, 2, ..., n\) comes from a Gaussian distribution. The test statistic is given by

\[
W = \frac{(\sum_{i=1}^{n} a_i x(i))^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}
\]

(2.20)

where,
\( x_{(i)} \) is the ith order statistic.
\( \bar{x} \) is the sample mean.

The coefficients \( a_i \) is given with the following:

\[
(a_1, \ldots, a_n) = \frac{m^T V^{-1}}{C}
\]

where,

\( C \), is a vector norm:
\[
C = ||V^{-1}m|| = (m^T V^{-1} V^{-1} m)^{0.5}
\]

and, the vector \( m \):
\[
m = (m_1, \ldots, m_n)^T
\]
is the expected values of the ordered statistics that are independent and identically distributed random variable which follow the standard normal, \( N(0, 1) \).

\( V \) is the covariance matrix of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution. To interpret this test the following is given:

- \( H_0 \): Null hypothesis, the data is normally distributed,
- \( H_a \): Alternative hypothesis, is not normally distributed,
- \( W \): Test statistic, which lies between zero and one.

Theoretically, one way of deciding whether a sample data comes from a normally distribution with this test is comparing the \( W \) values, where one them is calculated with Equation 2.20, and the other one is obtained from a given table for this test. If the calculated \( W \) value is greater than the \( W \) value in the table, then we do not reject the null hypothesis. The other one is obtaining a \( p \) value for the corresponding test statistical value, \( W \). Here, under the assumption that \( H_0 \) is correct, the \( p \) value is the probability of obtaining test results at least as extreme as the result that is observed \([19]\). If the \( p \) value is less than the \( \alpha \) level that is chosen, then the null hypothesis is rejected, and it means that the sampling data does not come from the normal distribution. We use this approach in this study, as the programming language we use, calculate this \( p \) value automatically.

The Table 2.2 is given as an example result from Shapiro-Wilk test:

<table>
<thead>
<tr>
<th>Test statistic</th>
<th>( p ) value</th>
<th>Significance level (( \alpha ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( W = 0.9683 )</td>
<td>0.1551</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 2.2: Shapiro-Wilk test
Since the $p = 0.1551$ is larger than the chosen $\alpha$ level, 0.05, we do not reject the null hypothesis. Therefore, it is obvious that the sample data follows the normal distribution. In other words, we can be confident with 95% of the whole data that our data comes from the Gaussian distribution.

### 2.6.3 Quantile-Quantile plot

Quantile-quantile (Q-Q) plot is a graphical method to find out whether two data sets come from the same distribution. The plot allows us to compare the sample data with many distributions, such as Gaussian distribution, exponential distribution and Weibull distribution. In this study, we compare our sample data with Gaussian distribution. Q-Q plot compares two probability distributions by plotting their quantiles against each other [21]. In this graphical method, $x$ coordinate represents the quantiles from the first distribution, whereas $y$ coordinate represents the corresponding quantiles from the second distribution. Quantiles can also be referred as ‘percentiles’, which are the values dividing a probability distribution into equal intervals. For instance, for Normal distribution, the mean is zero. In this case, the 0.5 quantile is 0, which means that the half of the data lies below 0. If the quantile value is 0.4 (or 40%), the quantile is the point where 40% of the data fall above, and 60% of data fall below. The plot consists of a reference linear line and some points from the second distribution, which we want to assess. If the points approximately fall along this line, then we assume that two sets of data come from the same distribution. In general, we can assume that the farther the points from the reference line, the less likely that the two data sets come from the same distribution.
Chapter 3

Implementation

3.1 Introduction

In this study, numpy and pandas libraries from python programming language are used to implement both filtering and outlier elimination, and data validation process as well as to analyze the stability of the filters. Before performing the implementation, the duplicates and the 'nan' values are removed from the data in order to have a clean data. Moreover, when some sensor errors are detected in some specific data points, these data points are also removed from the data. In this part, first, an overview of this process is given, and then, the sensor errors are covered with an example figure. Moreover, how the sections that are covered in Chapter 2 are implemented is explained with necessary plots, and finally, the results are drawn from these plots and from the analyses.

3.2 Overview and workflow

As it is mentioned in the Section [1.4] we have metadata between August 2011 and January 2014, and sensor data between March 2012 and January 2017. We find the corresponding data points in the available metadata for each data points in the sensor log files. Before finding the corresponding data points in the metadata, the sensor data log files are converted to csv files to make it easier for pandas library to read them. The reason why the two data are combined is we could only have an access to all the parameters that we need by combining them. The available metadata histogram is illustrated in [Figure 3.1]. We can observe from [Figure 3.1] that for some days, we have more than one data. However, for some days we do not have any data available. For example, between September 15, 2013 and November 7, 2013, there is a gap in the figure, which indicates that there is no available metadata between these days. After filtering the unsatisfied data points with outlier elimination and data validation process, a reference speed-power function is created, and by using this function, daily and monthly speed loss are calculated. The goal of creating these speed loss figures is to see how stable our existing filters as well as the modified filters against to technical changes of the vessel are. Therefore, depending on the stability results, the filters are loosen or tighten to have more valid data.
3.3 Sensor error

In the available data, some data points are detected that they cause some instability and distortion in the results. Getting such data points is named as sensor error in this study. To prevent us from having some unstable results, these data points are removed from the beginning. For example, we have some parameters such as 'delivered power', 'engine speed', 'main engine power', and 'shaft power', whose values are zero at some data points. We consider these data points as sensor error, since we can neither have a power value nor a speed value as zero under normal conditions. This is likely to happen, because either the sensor did not work at a specific time period, or it did not work accurately. We can understand when the sensors have worked accurately or not by checking whether the corresponding parameter values are zero or not. So, if they are not zero, we consider these values as the values that we do not observe any error in the obtained sensor values. However, this is not the only way to detect the sensor errors. For instance, when the 'average speed loss' figure is plotted, which is Figure 3.6a, it has been observed that some data points behave very unstable and unusual in comparison to the others. For this reason, it is investigated that whether a sensor error occurs for these data points. In Figure 3.2, it can be observed that the 'speed through water' parameter is 0.3 kn, whereas 'speed over ground' parameter is -0.4 kn for more than one month. Therefore, it is considered that these data points are measured incorrectly, and they are also removed from the data. In Table 4.2, these errors are denoted by STW+SOG, as STW is the specified error in the speed through water, and SOG is again the same error in the speed over ground parameter.

3.4 Filtering methods

In this study, unlike the ISO standards suggest, first, the reference conditions are met, and then outlier elimination and data validation process is handled. The necessary reference conditions
as well as the outlier elimination and validation processes are given with the implementation methods in this section.

### 3.4.1 Reference conditions

The reference conditions for ‘true wind speed’ and ‘water depth’ parameters are met, according to the filters that ISO-19030-2 suggests.

#### True wind speed

If Equation 2.9 is between 0 m/s and 7.9 m/s, then the data points in this interval are valid. The following python code is given to apply this reference condition.

**Code Listing 3.1: True wind speed**

```python
true_wind_speed = df['True Wind Speed']
return (true_wind_speed >= minimum_true_wind_speed_threshold) & (true_wind_speed < maximum_true_wind_speed_threshold)
```

#### Water depth

Water depth parameter is calculated with Equation 2.10, and the procedure that is explained in Section 2.3.2.1 is followed to filter the unsatisfied water depth values. According to the procedure, Equation 2.10 must be greater than the larger of the values, which are calculated
with Equation 2.11 and Equation 2.12. However, for the metadata we have, the water depth parameter is always significantly greater than the two calculated water depth values, ‘h’ in Section 2.3.2.1. So that, we do not have any hint to loosen or tighten this filter, and thus the water depth values are kept as they are. Figure 3.3 illustrates this situation for the metadata which is between June 2013 and January 2014.

![Figure 3.3: Water depth calculation parameters](image)

### 3.4.2 Outlier elimination and data validation

#### 3.4.2.1 Outlier elimination

Based on the formulas in Section 2.2.1.1 the mean and standard error of the mean are calculated, and outliers are detected with Chauvenet’s criterion. This criterion is fulfilled on the non-overlapping 10 minute blocks, and if Equation 2.3 is fulfilled, then the data point is considered outlier. To create these ten minute blocks, we first create ‘date times’, and to tag these date times, we use pandas `pd.cut` built-in function, and these tags are added to a new column in the available pandas dataframe. Finally, by using the pandas `groupby` built-in function, we create the ten minute blocks. To create the binning tags the Code Listing 3.2 is used.
Code Listing 3.2: Creating and adding binning tags

```
datetimes = pd.DataFrame(index=pd.date_range(first_time, last_time, freq=frequency)).index.tolist()
binned_series = pd.cut(df['Time'], bins=datetimes, include_lowest=True)
df[tag_name] = binned_series.dropna().astype(str)
```

The 'delivered power' parameter in [Equation 2.4] and the 'speed through water' parameter from the sensor log files are used in outlier elimination process.

### 3.4.2.2 Data validation

To validate the data for the parameters of rpm, rudder position, speed through water, and speed over ground, the non-overlapping, consecutive blocks of 10 min are created with the same procedure as in the outlier elimination section.

Code Listing 3.3: Data validation

```
is_valid_series = stats['standard_error'] <= parameters[column_name]['threshold']
original_df.loc[one_block_df.index[:,], f'Is {column_name} Valid'] =
    is_valid_series
```

Note that unlike the outlier elimination process, we use Code Listing 3.3 to remove the whole block that we have created, if it is flagged as invalid.

### 3.5 Reference speed-power curve

The reference speed for creating this curve is calculated with [Equation 2.15] and the power values are calculated with [Equation 2.4]. The minimum 1000 delivered power values are taken, with the corresponding reference speed values and to create the speed bins, which are explained in Section 2.4, first, the date times are created with pandas `pd.date_range` built-in function, and then they are tagged with pandas `pd.cut` function, and named as binning tags, then they are added to a new column, and finally to create the speed bins, we group these columns with pandas `groupby` function. To visualize the relation between the reference speed values and the corresponding delivered power values, pandas `plot.box` built-in function is used. The python code to create this curve is given in Code Listing 3.4 and the reference speed-power curve is given in Figure 3.4. In the figure, there is a gap between the speed values 23.2 and 23.6. That means that there is no available value for these speed values in the data. The values are calculated not only for the minimum of delivered power values, but also for the mean and the median of these values. In the figure, we can observe that the three of the box plots lie approximately on the same curve, which means that no matter which statistical calculation we choose, the plot gives the stable result for all three of them.
Code Listing 3.4: Reference speed-power curve

```python
lowest_N = 1000
reference_df =
create_reference_speed_water_binning_tags(self.minimum_reference_speed,
self.maximum_reference_speed)

group_by_reference_speed_bins = limited_df.groupby(['Binned Reference Speed tag'])
for reference_speed_range, one_block_df in group_by_reference_speed_bins:
    reference_speed_float = np.around(one_block_df['Reference Speed'].iloc[0], 1)
    expected_column = str(reference_speed_float)
    local_lowest = one_block_df['Delivered Power'].nsmallest(lowest_N).reset_index(drop=True)
    reference_df.loc[:, expected_column] = local_lowest
return reference_df
```

Figure 3.4: Reference speed-power curve

3.5.1 Reference speed-power function

To estimate the corresponding reference speed at a specific power value, a reference speed-
power function is created by utilizing the speed-power values in the Figure 3.4. Using `polyfit`
built-in function from the `numpy` library, a delivered power vs reference speed function is
generated. The `polyfit` function uses ordinary least squares estimate method which is discussed
in Section 2.4.1.3 The code that is used for this part is given in Code Listing 3.5
However, the function that is created with Section 2.4.1.3 does not exactly give what we want, because we want a reference speed vs delivered power function to find the speed that we investigate which corresponds to the delivered power value we have. To generate this function we find the roots of the previous function and take the inverse of it. Since the function that we generated before is power of three function, here, we obtain three roots which are both complex and real roots. The python code for this part is given in Code Listing 3.6.

Note that we use the power of three function, since there is a relationship in between the reference speed ($V$) and the power values ($P_d$) as $P_d \approx V^3$. This is due to the fact that the relationship between the ship resistance ($F_{ship}$) and the speed is $F_{ship} \approx V^2$, and to calculate the effective power ($P_E$), we must multiply the ship resistance by the speed ($V$) and other two unrelated parameters, thus we obtain the cube of $V$. Therefore, we must use the third power of polynomial function to obtain the reference speed-power function [13].

The corresponding function is illustrated in the Figure 3.5, and the three functions are given for minimum, mean and median of delivered power values from the reference speed-power curve. Here the blue dots above the curve represent the data points that we lose speed, and the ones below the curve represent the data points that we gain speed.

3.6 Daily and monthly speed loss

The reference speed values come from the reference-speed power function. First binned daily time tags are created, and then with pandas `pd.cut` built-in function, the tags are found, then added as another column and finally, the blocks are created with `groupby`. First, the daily speed loss is calculated by subtracting the reference curve speed which is obtained from the reference speed-power function from the measured speed which is calculated with the Equation 2.15. Therefore, we create speed blocks for the corresponding binning tags. Note that, here the number of values in each block is also calculated in the count variable in Code Listing 3.7. Then, the average speed loss values are calculated by summing all the values in a block and dividing them by the count.
Figure 3.5: Reference speed-power function

Code Listing 3.7: daily speed loss calculation

```python
reference_curve_speed = self.vector_reference_speed_vs_delivered_power_func(one_block_df['Delivered Power'])
count, speed_loss = len(one_block_df), np.sum(measured_speed - reference_curve_speed)
new_datetime = pd.to_datetime(time_range.split(',')[0][1:], format='%Y-%m-%d %H:%M:%S')
average_speed_loss = speed_loss / count
```

Furthermore, with the calculated daily average speed loss values, monthly average speed loss values are calculated as it is shown in Code Listing 3.8. Here, the main difference between daily and monthly speed loss calculation is that in daily loss, we take the length of the one block that we create, whereas in monthly loss, we sum all the values in the one block. Figure 3.6 illustrates the daily and monthly average speed loss performance over the time period in the given data. In this figure, the speed loss is calculated with three different functions where one of them is calculated with minimum polyfit, and the other ones are mean and median polyfit functions.

Code Listing 3.8: mothly speed loss calculation

```python
count, monthly_speed_loss = np.sum(one_block_df['Count']), np.sum(one_block_df['Speed Loss'])
average_speed_loss = monthly_speed_loss / count
diff_speed_loss = one_block_df['Average Speed Loss'] - average_speed_loss
```

38
From the Figure 3.6a, Figure 3.6b, and Figure 3.6c, we can observe that there is no significant difference between the plots. So that, we can use one of the three statistical values to calculate this speed loss performance and visualize them. In this study, we choose to continue with the minimum polyfit function. To show that the vessel is losing speed over the time, the cumulative speed loss figure is created based on the average daily speed loss values in Figure 3.7. In this figure the y-axis represents the total daily average speed loss and the x-axis represents the corresponding days for these speed loss values and in the same figure, we can observe at some points the vessel does not lose any speed. This is because we do not have any data for the corresponding dates. There are 154 valid days in total from approximately 429 days. However, since for some days there are not enough data to evaluate the speed loss, these days are removed from the data to get more reliable result. Therefore, if a day includes less than 100 data points, these days are removed from the data, and we obtain 138 valid days in this case. The total average speed loss for this ship with the valid days per day is -1.467 kn, whereas the total average speed loss per year is -1.341 kn. We can approximately obtain the same value for the speed loss per day by dividing the cumulative sum which is ~200 by the number of days which is 138. Note that we obtain the approximate value instead of the exact number, -1.467 kn. This is because we exclude the days that include less than 100 data points from the calculation.

### 3.7 Stability analysis

The stability analysis is covered with three methods. Firstly, the speed loss is visualized with a histogram and a Q-Q plot, and then some numerical values are obtain with Anderson-Darling and Shapiro-Wilk tests. In all three methods, the main idea is comparing the distribution of existing data with the normal distribution.

#### 3.7.1 Distribution visualization with histogram

In this part, we look at the distribution of the blue points in relation to the red line in Figure 3.6a. The histogram that illustrates the distribution is created by calculating the distance from the each average daily speed loss points to the corresponding average monthly speed loss line. If the distribution is roughly Gaussian or at least symmetric, then we conclude that the speed loss performance, which is our performance indicator, is stable, and if not then it is unstable. The histogram is given in Figure 3.8. From Figure 3.8 we can observe that there is a roughly Gaussian distribution in the plot. It means that the data points are distributed in a balance and there is a symmetric behaviour. So, we consider our speed loss performance stable.

#### 3.7.2 Distribution visualization with Q-Q plot

In this part, we again take the average speed loss distribution, but illustrate it in a Q-Q plot. In this plot, we compare the distribution of the speed loss values with Gaussian distribution. In Q-Q plot, there is usually a line which represents the Gaussian distribution, and some points which represent the distribution of a sample data. In our case the line is red and the points are
Figure 3.6: Speed loss performance

(a) Speed loss performance with minimum polyfit

(b) Speed loss performance with mean polyfit

(c) Speed loss performance with median polyfit
blue. The main idea here is that, the closer the blue points to the red linear line is the closer the sample data distribution is to Gaussian distribution. The Q-Q plot for our sample data is given in Figure 3.9. In this figure, since the blue dots roughly lie on the red line or approximately close to the line, we consider the distribution Gaussian. This means that our PI is stable with the original filters that ISO standards is required.

3.7.3 Anderson-Darling and Shapiro-Wilk tests

To get some numerical result to make sure that the distribution is roughly Gaussian, the Anderson-Darling and Shapiro-Wilk tests are performed. Both statistical tests use scipy.stats built-in function from scipy statistical library in python. Note that for Shapiro-Wilk test, scipy library takes the significance level as 0.05 by default, and in this study, we always use this significance level for Shapiro-Wilk test. This means that we always investigate with Shapiro-Wilk test that whether we obtain 95% confidence level or not. In other words, here we obtain 5% significance level, which corresponds to 95% confidence level. The following two examples are given to clarify how we use the test results and how we interpret them.

In Table 3.1, since the test statistic is less than the critical values of 0.766, 0.893, 1.063, we can conclude that with 95% of the data, we can explain that our data comes from Gaussian distribution. Observing Table 3.2, we can conclude that since the $p$ value is greater than the $\alpha$ value which is 0.05, the distribution that we test comes from Gaussian distribution.
3.8 Loosening Filters

The filters are loosened by observing the Q-Q plot and the histogram distribution plot and also by comparing the different results from Anderson-Darling and Shapiro-Wilk tests. In this section, the parameters that we loosen the filters for are given with their test results and the distribution plots. We analyse how much the stability of PI is affected against the modified filters. Note that the test results are in the given tables, and each table includes the result with the original filter limits, which is denoted by 0% for the comparison.

3.8.1 Outlier elimination parameters

We use delivered power and speed through water parameters in outlier elimination process. We both tighten and loosen the filter limit to see in which direction there is more effect in the results and at what percentage level the stability is distorted. Note that since we use the same threshold limit for both parameters, we cannot separate the results in terms of the different parameters.

![Average speed loss difference for the original filters](image)

**Figure 3.8: Average speed loss difference for the original filters**

<table>
<thead>
<tr>
<th>Test statistic</th>
<th>Critical values</th>
<th>Significance level($\alpha$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.731</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
</tbody>
</table>

**Table 3.1: Anderson-Darling test for original filters**
In the Table 3.3 and Table 3.4, the filter is loosened 30%, 40%, 50%, and 60%, whereas also tightened 50%. From these tables, we can say that there is no major effect of tightening the filter limit in the results, since even the filter limit is tightened 50%, we can still explain with 95% of the data that the distribution that we test comes from Gaussian distribution. In the Table 3.3, we can observe that 30% loosening filter limit ends up giving 1% significance level, which corresponds to 99% confidence level. This means that with 99% of the whole data, we can show that the distribution comes from Gaussian distribution. However, when the filter is loosened 40% and more, we can see in the Table 3.3 that the test statistic values are greater than the all critical values. On the contrary, we expect these values smaller than the critical values. Moreover, in the Table 3.4, we can observe that loosening the filter limit 40% and more gives a $p$ value which is less than the significance level, 0.05. However, we expect a $p$ value which is greater than 0.05 to obtain a confidence level 95%. Therefore, we can conclude that the filter threshold of outlier elimination parameters can be at most loosen 30%.

By tightening we mean that if a certain threshold has the value $\kappa$ and we tighten with 50%, then the new threshold is $\kappa \cdot (1-50\%)$, whereas if we loosen with 50%, then the new threshold is $\kappa \cdot (1+50\%)$.
<table>
<thead>
<tr>
<th>Percentage</th>
<th>Test statistic</th>
<th>Critical values</th>
<th>Significance level((\alpha))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0% Loosening</td>
<td>0.731</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>30% Loosening</td>
<td>1.033</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>40% Loosening</td>
<td>1.088</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>50% Loosening</td>
<td>1.088</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>50% Tightening</td>
<td>0.727</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>60% Loosening</td>
<td>1.103</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
</tbody>
</table>

Table 3.3: Anderson-Darling test for outlier elimination parameters

<table>
<thead>
<tr>
<th>Loosening percentage</th>
<th>(p) value</th>
<th>Significance level((\alpha))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0% Loosening</td>
<td>0.052</td>
<td>0.05</td>
</tr>
<tr>
<td>30% Loosening</td>
<td>0.015</td>
<td>0.05</td>
</tr>
<tr>
<td>40% Loosening</td>
<td>0.006</td>
<td>0.05</td>
</tr>
<tr>
<td>50% Loosening</td>
<td>0.006</td>
<td>0.05</td>
</tr>
<tr>
<td>50% Tightening</td>
<td>0.053</td>
<td>0.05</td>
</tr>
<tr>
<td>60% Loosening</td>
<td>0.005</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 3.4: Shapiro-Wilk test for outlier elimination parameters

Furthermore, we can also observe from the Figure 3.10 that 40% loosening of the filter introduces outliers, as we can see it in the upper right corner, whereas tightening the limit does not affect the distribution as the blue dots are fairly close the red line. In contrast, in the Figure 3.11b, we obtain a distribution which does not look like Gaussian or symmetric, whereas Figure 3.11a looks fairly symmetric.

3.8.2 Reference condition parameters

There are three parameters that we discuss for the reference conditions, but since for the water temperature parameter, the temperature never went below 2\(^{\circ}\)C during the journey of the vessel,
we do not eliminate any data points with this parameter. Moreover, for water depth parameter, we cannot loosen the filter limit for the reason that is explained in Section 3.4.1. So that, we are only interested in ‘true wind speed’ parameter in this section.

### 3.8.2.1 True wind speed

<table>
<thead>
<tr>
<th>Loosening percentage</th>
<th>Test statistic</th>
<th>Critical values</th>
<th>Significance level((\alpha))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>0.731</td>
<td>([0.56, 0.638, 0.766, 0.893, 1.063])</td>
<td>([0.15, 0.10, 0.05, 0.025, 0.01])</td>
</tr>
<tr>
<td>3%</td>
<td>0.701</td>
<td>([0.56, 0.638, 0.766, 0.893, 1.063])</td>
<td>([0.15, 0.10, 0.05, 0.025, 0.01])</td>
</tr>
<tr>
<td>5%</td>
<td>0.681</td>
<td>([0.561, 0.638, 0.766, 0.893, 1.063])</td>
<td>([0.15, 0.10, 0.05, 0.025, 0.01])</td>
</tr>
<tr>
<td>7%</td>
<td>1.794</td>
<td>([0.561, 0.638, 0.766, 0.893, 1.063])</td>
<td>([0.15, 0.10, 0.05, 0.025, 0.01])</td>
</tr>
<tr>
<td>10%</td>
<td>2.474</td>
<td>([0.561, 0.639, 0.766, 0.894, 1.063])</td>
<td>([0.15, 0.10, 0.05, 0.025, 0.01])</td>
</tr>
<tr>
<td>20%</td>
<td>2.546</td>
<td>([0.56, 0.638, 0.766, 0.893, 1.062])</td>
<td>([0.15, 0.10, 0.05, 0.025, 0.01])</td>
</tr>
<tr>
<td>25%</td>
<td>15.32</td>
<td>([0.56, 0.638, 0.766, 0.893, 1.063])</td>
<td>([0.15, 0.10, 0.05, 0.025, 0.01])</td>
</tr>
</tbody>
</table>

Table 3.5: Anderson-Darling test for true wind speed

In the Table 3.5, we can see that loosening the filter up to 5% gives a fairly good result in terms of having a Gaussian distributed data. Here, we can observe a 95% confidence level, if the filter is loosened 5%. However, when the filter is loosened 7% and more, we do not observe any Gaussian distributed data, as the test statistic values are significantly higher than the critical values. Moreover, we can also observe that there is a significantly large leap in the test statistic values between 20% and 25% loosening limits. From the Table 3.6, up to 5% loosening on the filter limit gives a Gaussian distributed data, but for the rest, the \(p\) values are fairly small to come from a Gaussian distribution, as we want these \(p\) values to be greater than 0.05 for this test for a 95% confidence level. So, according to Shairo Wilk test results, 5% loosening gives 95% confidence level for our data which is safe enough for us to loosen the filter up to this limit. We can observe in the Figure 3.12 that the Q-Q plots are getting worse as we loosen the filter. In Figure 3.12a, we do not have any outliers. However, in the Figure 3.12b, we have one outlier in the right upper corner, and one potential outlier in the
<table>
<thead>
<tr>
<th>Loosening percentage</th>
<th>$p$ value</th>
<th>Significance level ($\alpha$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>0.052</td>
<td>0.05</td>
</tr>
<tr>
<td>3%</td>
<td>0.099</td>
<td>0.05</td>
</tr>
<tr>
<td>5%</td>
<td>0.104</td>
<td>0.05</td>
</tr>
<tr>
<td>7%</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>10%</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>20%</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>25%</td>
<td>0</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 3.6: Shapiro-Wilk test for true wind speed

![Diagram](image1)

(a) 5% loosening  
(b) 7% loosening  
(c) 10% loosening

Figure 3.12: Q-Q plot for true wind speed

left bottom corner, and in the Figure 3.12c, we have two outliers in the upper right corner, and some potential outliers in the bottom left corner. So, this means that we just get more of these outliers as we loosen the filter to a higher limit. Under these conditions, it is safer to loosen the filter up to 5%, but to loosen it even more, we need additional outlier detection methods to remove these unsatisfied data points.

### 3.8.3 Data validation parameters

There are four data validation parameters that are discussed in Section 3.4.2.2, which are 'rudder position', 'engine speed', speed through water' and speed over ground', and we cover all four of them in this section. However, only 'rudder position' parameter eliminates noticeable number of data points, which is 20%, and for the others, each of them eliminates only 3% of the data points from the data, as we can see in the Table 4.2. Consequently, for rudder position parameter, it is more likely to see more noticeable changes against the updated filter limits.

#### 3.8.3.1 Rudder position

In the Table 3.7, we can observe a Gaussian distributed data up to 30% filter loosening. When the filter is loosened 10%, the test statistic gives a 5% significance level, which means that a 95% confidence level is obtained, by loosening the filter up to 10%. However, when the limit is 20%, the confidence level becomes 97.5%, whereas when it is 30%, the confidence level is still 97.5%. This means that we can be confident with 97.5% of the data that our distribution
<table>
<thead>
<tr>
<th>Loosening percentage</th>
<th>Test statistic</th>
<th>Critical values</th>
<th>Significance level(α)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>0.731</td>
<td>[0.56 , 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>10%</td>
<td>0.759</td>
<td>[0.56 , 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>20%</td>
<td>0.836</td>
<td>[0.561, 0.639, 0.766, 0.894, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>30%</td>
<td>0.859</td>
<td>[0.561, 0.639, 0.766, 0.894, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>35%</td>
<td>16.88</td>
<td>[0.561, 0.639, 0.766, 0.894, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>40%</td>
<td>14.42</td>
<td>[0.561, 0.639, 0.766, 0.894, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
</tbody>
</table>

Table 3.7: Anderson-Darling test for rudder position

comes from a Gaussian distribution, whereas when the filter is loosened 10%, 95% of the data is enough to indicate that the distribution is Gaussian distribution. However, when we make the loosening percentage 35%, we have a remarkable leap in the test statistic value, and it is far above the corresponding critical values. This means that we have sufficient evidence to reject the null hypothesis, meaning that the distribution of the sample data does not come from Gaussian distribution.

We observe a significantly similar result with Shapiro-Wilk test, as we obtain with Anderson-Darling test. As the loosening percentage becomes 35%, the p value is far below than the significance level we have chosen which is 0.05. On the contrary, we expect a p value above than 0.05. So, this gives a verification that we can loosen the filter of 'rudder position' parameter up to 30%.

We can verify our findings from the statistical tests with the visual stability analyses. In Figure 3.13a, we observe no outliers, and the blue dots are close enough to perform a distribution, quite similar to Gaussian distribution. However, when we look at the Figure 3.13b the blue dots are quite distorted, and do not lie on the red line which represents the Gaussian distribution in our case. In contrast, in Figure 3.14 the two histograms are fairly different to each other. Figure 3.14a looks symmetric, while Figure 3.14b is quite far from representing neither a Gaussian nor a symmetric distribution.

### 3.8.3.2 Engine Speed

For this parameter, we do not remove remarkable number of data points from the data. Due to this fact, the analysis is made by both loosening and tightening the filter limit, in case we could...
(a) 30% loosening for rudder position  
(b) 35% loosening for rudder position

Figure 3.13: Q-Q plot for rudder position

(a) 30% loosening for rudder position  
(b) 35% loosening for rudder position

Figure 3.14: Histogram plot for rudder position
find any interesting results for different filter limits. However, neither loosening nor tightening
the filter limit gives any significant change in the results of the stability analyses. Since small
changes do not affect the stability, the loosening/tightening percentages are given higher than
the rudder position and true wind speed parameters.

In Table 3.9 and Table 3.10, the filter is both loosen and tighten to 50%, and only loosen to

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Test statistic</th>
<th>Critical values</th>
<th>Significance level((\alpha))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0% Loosening</td>
<td>0.731</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>40% Loosening</td>
<td>0.731</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>50% Loosening</td>
<td>0.731</td>
<td>[0.561, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>50% Tightening</td>
<td>0.852</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>60% Loosening</td>
<td>0.731</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>70% Loosening</td>
<td>0.731</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.062]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
</tbody>
</table>

Table 3.9: Anderson-Darling test for engine speed

<table>
<thead>
<tr>
<th>Percentage</th>
<th>(p) value</th>
<th>Significance level((\alpha))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0% Loosening</td>
<td>0.052</td>
<td>0.05</td>
</tr>
<tr>
<td>40% Loosening</td>
<td>0.052</td>
<td>0.05</td>
</tr>
<tr>
<td>50% Loosening</td>
<td>0.052</td>
<td>0.05</td>
</tr>
<tr>
<td>50% Tightening</td>
<td>0.014</td>
<td>0.05</td>
</tr>
<tr>
<td>60% Loosening</td>
<td>0.052</td>
<td>0.05</td>
</tr>
<tr>
<td>70% Loosening</td>
<td>0.052</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 3.10: Shapiro-Wilk test for engine speed

40%, 60%, and 70%. Even though the loosening/tightening percentages are significantly high,
we do not observe any pattern that the stability is distorted with any filter limits in the tables.
We can see in Table 3.9 that 50% loosening on the filter still provides us 95% confidence
level, whereas tightening the filter on the same percentage gives 97.5% confidence interval.
Moreover, for the both tests, the results for all the loosening limits are the same. They all give
95% confidence level in both Anderson-Darling and Shapiro-Wilk tests. This means that 95% of
the sample data is enough to show that our data comes from Gaussian distribution.

To show that there is no any noticeable difference between 40% loosening and 70% loosening
of the filter, Figure 3.15 is given. We can conclude from these figures that we do not change a
lot on how much data we get back by increasing the filter limit. Either we loosen 40% or 70%,
the distribution does not change much, and only a few of the data points are clearly incorrect.
This gives an indication that even we loosen the filters 100%, they will still be removed from
the data. Furthermore, It is obvious that the plots in Figure 3.15 are quite similar with the plot
that we create with original filters in Figure 3.9.
3.8.3.3 Speed over ground

In the both directions, where we both loosen and tighten the filters for speed over ground parameter, there is no any significant change in the stability of the PI. We choose the filter limits again very high to see if we can observe any distortion in the stability. In the Table 3.11 and in the Table 3.12, the filters are loosen and tighten 50%, and loosen 30% and 60% for the ‘speed over ground’ parameters. We do not obtain any remarkable change for different filter limits. In the Table 3.11, all loosening percentages as well as the tightening limit gives a 95% confidence level. This means that we can be confident with 95% of the data that the data comes from the Gaussian distribution. In Table 3.12 we can observe that 50% tightening and 60% loosening give 95% confidence, whereas 30% and 50% loosening still give a fairly good result to loosen the filters to these limits.

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Test statistic</th>
<th>Critical values</th>
<th>Significance level((\alpha))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0% Loosening</td>
<td>0.731</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>30% Loosening</td>
<td>0.754</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>50% Loosening</td>
<td>0.745</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>50% Tightening</td>
<td>0.759</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
<tr>
<td>60% Loosening</td>
<td>0.743</td>
<td>[0.56, 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
</tbody>
</table>

Table 3.11: Anderson-Darling test for speed over ground

In Figure 3.16, there are three different histogram plots for the corresponding loosening percentages. In Figure 3.16a, there is a slightly better symmetry than the other two figures have. Nevertheless, they all look symmetric.

3.8.3.4 Speed through water

The stability results are not different than ‘engine speed’ and ‘speed over ground’ parameters for ‘speed through water’ parameter, as they are still stable against the updated filters. Here, again we both loosen and tighten the filters to see if there is any effect on the stability of the updated filter limits. In Table 3.13 and Table 3.14 the filter of speed through water parameter
We can observe in the Table 3.13 and the Table 3.14 that no matter how much percent we loosen the filter, we obtain the same test statistic result, and they are all in the same confidence level, which is 90% in Table 3.13 while tightening the filter limit gives a 95% confidence level. When we tighten the filter 50%, the $p$ value for Shapiro-Wilk test is smaller the other $p$ values in Table 3.14. Although it does not give a 95% confidence level, we can still treat the data that it comes from Gaussian distribution as it is still greater than 0.01. In Figure 3.17a and Figure 3.17b, we can observe that there is no significant difference between the two plots. The right tail of Figure 3.17a looks slightly better than the right tail of Figure 3.17b, as the blue dots are closer to the red line. Nevertheless, they both look fairly good, and we do not see any outliers in the plots.

We can conclude that 'engine speed', 'speed over ground', and 'speed through water' parameters do not have a major effect on the results. So, we can say that the filters of these parameters are sort of optimal. These parameters removes a few outliers which we want to remove, but it does not really matter where we set the filter limits. Therefore, we do not include these parameters.
<table>
<thead>
<tr>
<th>Percentage</th>
<th>p value</th>
<th>Significance level((\alpha))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0% Loosening</td>
<td>0.052</td>
<td>0.05</td>
</tr>
<tr>
<td>30% Loosening</td>
<td>0.088</td>
<td>0.05</td>
</tr>
<tr>
<td>50% Loosening</td>
<td>0.088</td>
<td>0.05</td>
</tr>
<tr>
<td>50% Tightening</td>
<td>0.022</td>
<td>0.05</td>
</tr>
<tr>
<td>60% Loosening</td>
<td>0.088</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 3.14: Shapiro-Wilk test for speed through water

![Q-Q plot for speed through water](image)

(a) 50% loosening  
(b) 50% tightening

Figure 3.17: Q-Q plot for speed through water

in [Table 4.1](#) as we cannot do anything with the filter limits of these parameters. Nevertheless, it is worth to emphasize that for these parameters, tightening the filter limits give slightly worse result in comparison to loosening the filters.
Chapter 4

Results

In this chapter, the results of the implementation process is given in the Table 4.1 and Table 4.2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Percentage</th>
<th>Original threshold</th>
<th>Updated threshold</th>
<th>Total increase</th>
<th>Total increase(%)</th>
<th>Updated valid days</th>
</tr>
</thead>
<tbody>
<tr>
<td>TWS²</td>
<td>5%</td>
<td>[0, 7.9]</td>
<td>[0, 8.295]</td>
<td>209,500</td>
<td>6%</td>
<td>139</td>
</tr>
<tr>
<td>RPs</td>
<td>30%</td>
<td>1°</td>
<td>1.3°</td>
<td>253,166</td>
<td>7%</td>
<td>140</td>
</tr>
<tr>
<td>STW² and Pds</td>
<td>30%</td>
<td>0.5</td>
<td>0.65</td>
<td>1350</td>
<td>0.04%</td>
<td>138</td>
</tr>
</tbody>
</table>

1:True wind speed, 2:Rudder position, 3:Speed through water, 4:Delivered power.

Table 4.1: Analysis on the modified filters

Figure 4.1: Loosening result

From the Table 4.1, we can obtain the following results. Note that in the last row, the outlier elimination parameters are given together as we cannot separate them due to having the same
• We are able to loosen the original threshold limits of true wind speed (TWS), rudder position (RP), and outlier elimination parameters, which are speed through water (STW) and delivered power ($P_d$).

• Even though all the data validation parameters threshold limits are tried to loosen, we are only able to loosen the threshold limit of the rudder position parameter.

• Therefore, we can say that other data validation parameters are significantly stable to the technical changes of the vessel.

• Even though we could only loosen the filter limit of the true wind speed parameter up to 5%, we gain significantly more data in the end in comparison to the outlier elimination parameters which can be loosened up to 30%.

• This means that even small changes affect the stability of the true wind speed parameter significantly.

• Rudder position parameter allows us to gain 7% more data, even though it is loosened 30%, which is remarkably greater than the true wind speed parameter filter loosening percentage.

• We can say that rudder position parameter is much more stable than true wind speed parameter. However, it is not as stable as engine speed, speed through water, and speed over ground parameters.

• Loosening the limit of outlier elimination parameters do not have any major effect in the result, as we only gain 0.04% more data in comparison to having the original filter limit.

In order to visualize the results from the Table 4.1, the Figure 4.1 is given.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Parameter</th>
<th>Valid</th>
<th>Valid(%)</th>
<th>Invalid</th>
<th>Invalid(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensor error</td>
<td>Delivered power</td>
<td>122877191</td>
<td>88%</td>
<td>1715664</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>Engine speed</td>
<td>125831256</td>
<td>90%</td>
<td>14202599</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>Main engine power</td>
<td>122877191</td>
<td>88%</td>
<td>1715664</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>Shaft power</td>
<td>73255833</td>
<td>55%</td>
<td>62757982</td>
<td>45%</td>
</tr>
<tr>
<td></td>
<td>STW + SOG²</td>
<td>12758318</td>
<td>91%</td>
<td>12175537</td>
<td>9%</td>
</tr>
<tr>
<td>Reference conditions</td>
<td>True wind speed</td>
<td>79520822</td>
<td>57%</td>
<td>60513033</td>
<td>43%</td>
</tr>
<tr>
<td></td>
<td>Water depth</td>
<td>120790634</td>
<td>94%</td>
<td>122903223</td>
<td>6%</td>
</tr>
<tr>
<td>Outlier elimination</td>
<td>Delivered power</td>
<td>120703649</td>
<td>86%</td>
<td>19380206</td>
<td>14%</td>
</tr>
<tr>
<td></td>
<td>STW¹</td>
<td>128244455</td>
<td>92%</td>
<td>11789400</td>
<td>8%</td>
</tr>
<tr>
<td></td>
<td>STW²</td>
<td>135993930</td>
<td>97%</td>
<td>4039925</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>SOG²</td>
<td>138165864</td>
<td>97%</td>
<td>3870191</td>
<td>3%</td>
</tr>
</tbody>
</table>

1: Speed through water, 2: Speed over ground.

Table 4.2: Analysis on the data
In the Table 4.2, an analysis is given in terms of valid and invalid data point numbers, and their percentages:

- Due to sensor errors that we have in the specified parameters, significant number of data points are removed from the data. Especially, we lose almost a half of the shaft power parameter values, since a measurement error occurs.

- Due to having some sensor error in speed through water and speed over ground parameters as it is explained in Section 3.3, we lose 12 million data points, which is 9% of the data.

- Reference condition parameters are the parameters which eliminate the most data points in comparison to the other methods. Only 12% of the water depth parameter is valid, while almost half of the ‘true wind speed’ parameter is invalid.

- Even though we eliminate the delivered power values which are ‘0’ in the sensor error elimination part, we lose 14% of the delivered power parameter values in outlier elimination process, whereas only 8% of the speed through water parameter values is outlier.

- In the data validation process, we remove significant number of data points for rudder position parameter. The other parameters in this process eliminate only 3% of the data points for each of the parameter.

- In addition to all these result, we can say that we have 140 million data points before all the filtering processes and the sensor elimination, and we end up having 3 million data points after these processes are done.

- In other words, we only have 2% of the data after the sensor error elimination and the filtering processes are done.
Chapter 5

Conclusion

In this paper, it is investigated that how stable our sensor values in the given data are against the technical changes of the vessel. To understand the stability of the parameters in the data, first, the data is prepared by meeting the reference conditions in the ISO standards and also by filtering outliers and validating the data. After obtaining this prepared data, with the help of the reference speed-power curve which is created by a curve fitting method, the speed loss of the vessel during its journey is calculated in daily and monthly basis. Then, these speed loss values are considered Performance indicator (PI). To clarify, the speed loss performance of the vessel, which is shown in Figure 3.6, is used as our PI. Furthermore, an analysis is made on this figure by looking at the distribution of the daily speed loss values in relation to the monthly speed loss values. In order to figure out how our sensors react when the original filter limits change, some loosening techniques are used on the threshold limit of the different filters. Moreover, we observe how stable our PI is by observing some numerical test results as well as some plots. Then, we figure out up to what percentage the filters can be loosen while maintaining the stability of the PI. When all the implementation processes are done, we obtain the results that we have in Chapter 4. In light of these findings, we can say that the algorithm that is created works consistently and accurately. There is no any particular result that introduces any errors or affects any other result in a way that making it less accurate or less reliable. However, in this study, we could not assess the accuracy of our PI due to the reasons that are explained in Section 2.3.2.4. Therefore, we do not have any knowledge on what accuracy is achieved.

In a further study, it is suggested to assess the accuracy of the PI on the available data by creating the speed-power model of the ship as well as the simplified digital twin model. Moreover, since the reference speed power curve is the one which fits the available data the best, if the curve can be obtained from the ship owner in a further study, the results that we obtain can be even more accurate. Furthermore, for the parameters of engine speed, speed through water, and speed over ground, we obtain a slightly worse results by tightening the filter limits than loosening them. A research can be made by investigating the reasons of this manner in the future as well. In addition to these, in the metadata, the data is given according to the local time, whereas in the sensor files, the data is given in the GMT zone. This can also introduce some unreliable sensor values, as we have combined these two data in the beginning of the study. Therefore, for a further study, it is also recommended to collect these two different
data in the same time zone.
Bibliography


Appendices

A Least Square Method of Polynomial Regression

In the study, we use third degree polynomial function to have a reference curve by using curve fitting method. We show how we derive the least square method for the polynomial regression here. Since we use a third degree polynomial in this study, we show the derivation for the power of three polynomial function.

A third degree polynomial is defined as

$$\hat{y} = ax^3 + bx^2 + cx + d$$ (1)

In the least squares method, we want to minimize a difference between the measured data, $y$, and our third degree polynomial function, $\hat{y} = ax^3 + bx^2 + cx + d$. In this way, we obtain the difference between the measured data and the estimated data provided by our model, and we name this difference as residual and denote it by $r$. Therefore we obtain,

$$r = y - \hat{y}$$ (2)

With the coefficients $a, b, c,$ and $d$ and the residual for a data point at $i$ is defined as

$$r_i = y_i - (ax_i^3 + bx_i^2 + cx_i + d)$$ (3)

The sum of all residuals is defined as

$$\sum_{i=0}^{N} r_i = \sum_{i=0}^{N} (y_i - (ax_i^3 + bx_i^2 + cx_i + d))$$ (4)

Hence the sum of squares of residuals (RSS) is defined as

$$RSS = \sum_{i=0}^{N} r_i^2 = \sum_{i=0}^{N} (y_i - (ax_i^3 + bx_i^2 + cx_i + d))^2$$ (5)

where $N$ is the number of known data points.

Equation 5 has a value $\geq 0$ due to the square and geometrically, there can be no critical points that are maximum. Hence, the critical point for RSS has to be minimum, which can be tested
by checking the RSS value with any of $a, b, c, d$ being sufficiently large positive or negative number, and we can always find a bigger value. Therefore, we can take the partial derivative for each coefficient and equal to 0 to find the critical point, in this case minimum point, as in the following:

\[
\frac{\partial \text{RSS}}{\partial a} = \frac{\partial \left( \sum_{i=0}^{N} (y_i - (ax_i^3 + bx_i^2 + cx_i + d))^2 \right)}{\partial a} = 0 \tag{6a}
\]

\[
\frac{\partial \text{RSS}}{\partial b} = \frac{\partial \left( \sum_{i=0}^{N} (y_i - (ax_i^3 + bx_i^2 + cx_i + d))^2 \right)}{\partial b} = 0 \tag{6b}
\]

\[
\frac{\partial \text{RSS}}{\partial c} = \frac{\partial \left( \sum_{i=0}^{N} (y_i - (ax_i^3 + bx_i^2 + cx_i + d))^2 \right)}{\partial c} = 0 \tag{6c}
\]

\[
\frac{\partial \text{RSS}}{\partial d} = \frac{\partial \left( \sum_{i=0}^{N} (y_i - (ax_i^3 + bx_i^2 + cx_i + d))^2 \right)}{\partial d} = 0 \tag{6d}
\]

For the Equation 6a, we show how we take the derivation for the coefficient $a$ in the following:

\[
\frac{\partial \text{RSS}}{\partial a} = \left( \sum_{i=0}^{N} 2(y_i - (ax_i^3 + bx_i^2 + cx_i + d)) \right)(-x_i^3) = 0 \\
\Leftrightarrow \sum_{i=0}^{N} 2(-x_i^3)(y_i - (ax_i^3 + bx_i^2 + cx_i + d)) = 0 \\
\Leftrightarrow \sum_{i=0}^{N} 2(ax_i^6 + bx_i^5 + cx_i^4 + dx_i^3 - y_ix_i^3) = 0 \\
\Leftrightarrow \sum_{i=0}^{N} (ax_i^6 + bx_i^5 + cx_i^4 + dx_i^3 - y_ix_i^3) = 0
\]

We can take the derivation for the coefficients $b, c, d$ in the same way as it is shown for the coefficient $a$, and we obtain the following equations:

\[
\sum_{i=0}^{N} (ax_i^6 + bx_i^5 + cx_i^4 + dx_i^3 - y_ix_i^3) = 0 \tag{7a}
\]

\[
\sum_{i=0}^{N} (ax_i^5 + bx_i^4 + cx_i^3 + dx_i^2 - y_ix_i^2) = 0 \tag{7b}
\]

\[
\sum_{i=0}^{N} (ax_i^4 + bx_i^3 + cx_i^2 + dx_i - y_ix_i) = 0 \tag{7c}
\]

\[
\sum_{i=0}^{N} (ax_i^3 + bx_i^2 + cx_i + d - y_i) = 0 \tag{7d}
\]
which can also be shown in the matrix formulation by moving the minus variables to the right side of the equations:

\[
\begin{bmatrix}
\sum_{i=0}^{N} x_i^6 & \sum_{i=0}^{N} x_i^5 & \sum_{i=0}^{N} x_i^4 & \sum_{i=0}^{N} x_i^3 \\
\sum_{i=0}^{N} x_i^5 & \sum_{i=0}^{N} x_i^4 & \sum_{i=0}^{N} x_i^3 & \sum_{i=0}^{N} x_i^2 \\
\sum_{i=0}^{N} x_i^4 & \sum_{i=0}^{N} x_i^3 & \sum_{i=0}^{N} x_i^2 & \sum_{i=0}^{N} x_i \\
\sum_{i=0}^{N} x_i^3 & \sum_{i=0}^{N} x_i^2 & \sum_{i=0}^{N} x_i & N
\end{bmatrix}
\begin{bmatrix}
a \\
b \\
c \\
d
\end{bmatrix}
= \begin{bmatrix}
\sum_{i=0}^{N} y_i x_i^3 \\
\sum_{i=0}^{N} y_i x_i^2 \\
\sum_{i=0}^{N} y_i x_i \\
\sum_{i=0}^{N} y_i
\end{bmatrix}
\]  
(8)

When we solve this linear system equation, unless it is inconsistent, we find the values for the linear estimator

\[
\hat{\beta} = \begin{bmatrix}
a \\
b \\
c \\
d
\end{bmatrix}
\]  
(9)

The python method `numpy.polyfit` runs a similar linear system equation considering numerical calculation limitations with 140 million data, which is \( N \), in our study.
B Statistical Hypothesis Testing against Gaussian Distribution

In our study, we test if the our results come from Gaussian distribution and this appendix explains how we use statistical hypothesis testing. Moreover, as it is explained in Section 2.6.1 both Anderson-Darling and Shapiro-Wilk tests are a member of group of Goodness-of-Fit (GoF) statistics, and Shapiro-Wilk test is design to test if a sample data comes from Gaussian distribution, whereas Anderson-Darling test can be used as a test of normality as well as testing a sample of data for logistic, exponential, Weibull distributions as well [15]. Anderson-Darling test is an improved version of Shapiro-Wilk test, and they are both known as sufficiently good at testing a sample of data for Gaussian distribution, and for this reason we use these two tests from the group of GoF. To see the necessary explanations and definitions, as well as examples which can be helped to understand these tests, Section 2.6.1 and Section 2.6.2 are given in this paper.

In these tests, we use a composite hypothesis, since testing a distribution can have more than two states and we define the null hypothesis, \( H_0 \), as the distribution comes from Gaussian distribution and the alternative hypothesis, \( H_a \), as the distribution does not come from Gaussian distribution. We also define the significance level, \( \alpha \), as the probability of false positive, which is rejecting the null hypothesis while in fact it is Gaussian distribution, and the confidence interval is the probability that we do not reject the null hypothesis when it is in fact Gaussian distribution, which is \( 1 - \alpha \) and one-sided. We define \( p \) value as the probability of obtaining a test result at least as extreme as the test result was actually observed by assuming that the null hypothesis is true. For detailed information, the definitions and explanations for significance level, confidence level, and \( p \) value can be seen in the textbook [9]. Note that both Anderson-Darling and Shapiro-Wilk tests have the same hypothesis and our goal is not to reject the null hypothesis. However, they have different ways of testing the values and rejecting the null hypothesis.

Shapiro-Wilk test calculates both the test statistic and \( p \) values. The test checks if the \( p \) value is less than the default significance level value, 0.05 [1] and if it is less than the significance level, the null hypothesis is rejected and we say that it does not come from Gaussian distribution. However, in our case we change the significance value as 0.01 as our goal is to include as much data as possible. During this paper, the statistic test value is omitted while keeping the \( p \) value and significance value as in Table 1.

<table>
<thead>
<tr>
<th>Loosening percentage</th>
<th>( p ) value</th>
<th>Significance level(( \alpha ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>0.052</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 1: An example of Shapiro-Wilk test in a table

One should interpret Table 1 with the following items:

\[ ^1 \text{This is the default significance value that the python programming language provides, but this significance value can go down to 0.01 [5].} \]
• This is valid only for 0% loosening.

• Our default \( \alpha \) is 0.05, which also means our confidence interval is 95%.

• \( p \) value is greater than the significance level value, \( \alpha = 0.01 \), hence we cannot reject the null hypothesis, \( H_0 \), and this indicates that the distribution comes from Gaussian distribution. Since the \( p \) value is even greater than 0.05, we can say that with 95% confidence interval, the distribution comes from Gaussian distribution.

Anderson-Darling test, on the other hand, calculates only the test statistic value and checks if the test statistic value is greater than the critical value from the paper [20] for the given corresponding significance level values. If the test statistic value is greater than the critical value for significance level, \( \alpha = 0.01 \), then the null hypothesis is rejected and we say that the distribution does not come from Gaussian distribution. To interpret this test, an example is given in Table 2.

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Test statistic</th>
<th>Critical values</th>
<th>Significance level(( \alpha ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0% Loosening</td>
<td>0.731</td>
<td>[0.56 , 0.638, 0.766, 0.893, 1.063]</td>
<td>[0.15, 0.10, 0.05, 0.025, 0.01]</td>
</tr>
</tbody>
</table>

Table 2: An example of Anderson-Darling test in a table

One can interpret Table 2 with the Table 3.

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Test statistic</th>
<th>Critical value</th>
<th>Significance level(( \alpha ))</th>
<th>Confidence Interval</th>
<th>Reject ( H_0 )?</th>
</tr>
</thead>
<tbody>
<tr>
<td>0% Loosening</td>
<td>0.731</td>
<td>0.56</td>
<td>0.15</td>
<td>85%</td>
<td>Yes</td>
</tr>
<tr>
<td>0% Loosening</td>
<td>0.731</td>
<td>0.638</td>
<td>0.10</td>
<td>90%</td>
<td>Yes</td>
</tr>
<tr>
<td>0% Loosening</td>
<td>0.731</td>
<td>0.766</td>
<td>0.05</td>
<td>95%</td>
<td>No</td>
</tr>
<tr>
<td>0% Loosening</td>
<td>0.731</td>
<td>0.893</td>
<td>0.025</td>
<td>97.5%</td>
<td>No</td>
</tr>
<tr>
<td>0% Loosening</td>
<td>0.731</td>
<td>1.063</td>
<td>0.01</td>
<td>99%</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 3: Detailed of an example of Anderson-Darling test in a table

and in the first two rows we reject the null hypothesis, because the test statistic value is greater than the critical value. In our study, we reject the null hypothesis when the test statistic value is greater than the one for \( \alpha = 0.01 \) and in this case, we do not reject the null hypothesis. Note that it is valid only for 0% loosening.

Table 4 summarises the interpretation of both Shapiro-Wilk and Anderson-Darling tests.
<table>
<thead>
<tr>
<th>Test type</th>
<th>Null Hypothesis ((H_0))</th>
<th>Alternative Hypothesis ((H_a))</th>
<th>Way of testing</th>
<th>When to reject (H_0)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shapiro-Wilk</td>
<td>The distribution is Gaussian</td>
<td>The distribution is not Gaussian</td>
<td>Calculates test statistic value, then its corresponding (p) value and compares with the significance value (\alpha)</td>
<td>When (p) value is less than (\alpha = 0.01).</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>The distribution is Gaussian</td>
<td>The distribution is not Gaussian</td>
<td>Calculates the test statistic value and the five significance level values’ corresponding critical values and compares the test statistics with the given significance level’s critical value</td>
<td>When the test statistic value for (\alpha = 0.01) is less than the corresponding critical value.</td>
</tr>
</tbody>
</table>

Table 4: Summary of statistical hypothesis testing against Gaussian Distribution