MSc Business Analytics Thesis

CO_2 EMISSION OF INLAND SHIPPING IN THE PORT OF ROTTERDAM

June 6, 2021

Dieuwertje Kardolus
Vrije Universiteit van Amsterdam
Faculty of Science

CO_2 EMISSION OF INLAND SHIPPING IN THE PORT OF ROTTERDAM

Dieuwertje Kardolus

MSc Business Analytics thesis

June 6, 2021

Supervisors:
Dr. R. Hindriks
Dr. E.R. Dugundji (second reader)

Vrije Universiteit van Amsterdam
Faculty of Science
De Boelelaan 1085
1081 HV Amsterdam
The Netherlands

Supervisors:
Dr. A. Lobo Gomes
Mrs. M. Dalhuisen

Port of Rotterdam
World Port Center (WPC)
Wilhelminakade 909
3072 AP Rotterdam
The Netherlands

Preface

This thesis is my graduation project in order to obtain the MSc degree of Business Analytics at the Vrije Universiteit van Amsterdam. This program is an interdisciplinary degree in the fields of mathematics, computer science and economics. All the knowledge I have gained and skills I developed during this program were useful during the internship at the Port of Rotterdam. The research I conducted here focused on the CO_2 emission by inland shipping in the Port of Rotterdam. Although this internship started at the very beginning of the COVID-19 crisis, which made it more challenging, I am grateful that I was able to finish it.

I would like to thank my supervisors dr. R. Hindriks from the Vrije Universiteit van Amsterdam, dr. A. Lobo Gomes of the Port of Rotterdam Authority for all the help and support and dr. E.R. Dugundji for being the second reader. A special thanks to dr. H.J.M. van Goor-Balk, the internship coordinator, my family and friends for the great support and motivational talks that helped me through the last phase of writing this thesis.

Summary

The Green Deal on Maritime and Inland shipping and Ports [12] comprises the guidelines from the National Climate Agreement that apply to the Port of Rotterdam. The ultimate goal is to reach climate neutrality by 2050 and stop further global warming. Carbon dioxide (CO_2) is one of the greenhouse gasses that contribute to climate change the most. Although research has been done with regards to the emissions of the sea ships little insight is available about the emissions of inland ships. Therefore the main purpose of this thesis was to estimate the CO_2 emission of the inland vessels in the Port of Rotterdam by 2050.

Firstly, the current CO_2 emission was calculated for every observation in the AIS data set that occurred within the Port of Rotterdam from April until September 2020. These observations were from 122 inland vessels that operate in the segments containers, tankers, dry bulk and break bulk and visited the Port of Rotterdam 1850 times. While the observations originated from the AIS data set, the Verblijfsmonitor data set and two provided Excel sheets provided information about the vessels and the visits. The CO_2 emission was calculated by multiplying the energy consumption by the emission factor of the energy source. The energy consumption was estimated with the use of an existing model. Currently, more or less all inland vessels use diesel oil as energy source. Hence the emission factor of diesel oil was used for the calculations. Empirical distributions (histograms) of several variables were derived from the results. Thereafter Monte Carlo simulations that include these distributions were performed to obtain possible CO_2 emission values for 2020 until 2050. Besides the distributions, the results of the simulations depended also strongly on the yearly throughput and the emission factor. The emission factor was determined by the energy source. The energy source could be classified as 'diesel oil' or 'zero emission'. The latter had an emission factor of 0. The emission factor of 'diesel oil' decreased over the years due to an increasing percentage of bio fuel, which was assumed to be mixed. The assumption was that the emission factor of bio fuel was also 0. Since it is unknown how the emission factor will develop, a simple linear decreasing emission factor was chosen. The classification of an inland vessel as 'diesel oil' or 'zero emission' was determined by rules that were based on a group classification and the age of the engine. Unfortunately, the first was based on a method that turned out to not perform well and the latter was based on data of insufficient quality. The simulation algorithm was run with two different models: model 1 included dependency between two groups of variables, while all the variables were independent of each other in model 2. Lastly, to investigate if the chosen bin size of the histograms of the input variables had an impact on the results, the algorithm was run with different bin sizes.

After all the assumptions that were made, the missing values imputation and the data transformations the calculated CO_2 emission was 0.0744 megaton in the observed 6-months time period. Secondly, although the proposed algorithm for the simulations performed well, the output was not usable. Due to technical limitations the real throughput forecast could not be used and the simulations were run with a much smaller fictional throughput. As a consequence, the number of inland vessels that were required to transport this much smaller yearly throughput was very low. Therefore the energy transition from 'diesel oil' to a 'zero emission' energy source did not have an impact, while the decreasing emission factor, i.e. the percentage of bio fuel that is mixed with the diesel oil, fully determined the results. The range of output values of model 1 showed to be much smaller compared to model 2: its average value was higher, while its standard deviation was smaller. The input values of model 2 were less restricted, which leads to a wider range of possible outcomes of the algorithm and a lower average value. Nevertheless, inclusion of dependency between (some of) the variables is always preferred, since there exist significant correlations between the input variables. It will therefore lead to realistic results. The bin size of the empirical distributions did not show to have a significant impact.

The expectation is that if the values of the input variables of the model to estimate the energy consumption are more generalized and therefore less depended on the specific information of the data sets, the results will improve. The application of the model per section of the Port of Rotterdam and the use of the characteristics per vessel class instead of every individual vessel can be possible

adjustments to the proposed method. If the maximum value for all the input variables is chosen this should lead to the current maximum CO_2 emission within a section. The summation of the calculated CO_2 emission of each section leads to the total amount. For the Monte Carlo simulations the software Python should not be used, since the algorithm appeared too complex in combination with the server specifications. Additionally the advise is to reconsider the rules that determine the age of the engine and the energy source, because these were the most important factors of the results but had the lowest data quality. Also, a different method should be used to classify the vessels into groups. Lastly, the usage of characteristics per vessel class for the distributions will make a model for the simulations that takes the correlations into account, unnecessary. This will also avoid unrealistic combinations of certain input variables in the simulations and therefore improve the results.

Contents

List of Figures

List of Tables

Introduction

The world is getting more aware of the consequences of climate change. Climate change is considered to be an urgent and potentially irreversible threat to human societies and the planet [3]. As response to this problem the Paris agreement was adopted in 2015 and entered into force in 2016. The Paris Agreement is a legally binding international treaty within the United Nations Framework Convention on Climate Change (UNFCCC) [34]. Carbon dioxide, or CO_2, is one of the greenhouse gasses that contributes the most to climate change. An essential way to tackle climate change is through reduction of these global emissions. The ultimate goal is to reach climate neutrality by 2050. The Dutch government translated the agreements of the Paris Agreement into national goals. Parliament Rutte III of October 2017 stated in the National Climate Agreement that the goal is to reduce greenhouse gasses in the Netherlands by 49% with respect to 1990 before 2030 [15]. In this agreement one can find all the measures and deals between companies, civil society organisations and governments to cooperate in order to accomplish this goal [33]. An undeniable fact was that there was a lot of room for improvement of the sustainability of maritime, inland vessels and ports. Therefore guidelines for these groups were also included in this agreement.

The guidelines from the National Climate Agreement are elaborated in the Green Deal on Maritime and Inland shipping and Ports [12]. The Green Deal, in which the Port of Rotterdam was also involved, defines ambitions, goals and intended results. For Dutch inland shipping the ambitions are:

- By 2030 CO_2 emissions will be reduced by 40% to 50% with respect to 2015 and at least 150 inland vessels have a zero emission powertrain.
- By 2035 emissions of environmental pollutants are reduced by 35% to 50% with respect to 2015.
- By 2050 to have a zero emission and climate-neutral inland fleet

The first goals mentioned in [12] for Dutch inland vessels are CO_2 emission reduction with at least 20% by 2024 with respect to 2015, reduction of the emissions of environmental pollutants with 10% with respect to 2015 and the development of new European management instruments that will encourage achievement of the ambitions.

The Port of Rotterdam (PoR) is Europe's largest sea port. "The objective of the Port of Rotterdam Authority (PoRA) is to enhance the port's competitive position as a logistics hub and world-class industrial complex. Not only in terms of size, but also with regard to quality. The core tasks of the Port Authority are to develop, manage and exploit the port in a sustainable way and to render speedy and safe services for shipping." [2] The PoRA is responsible that the PoR complies with the guidelines from the Green Deal. Although research has been done with regards to the emissions of the sea ships little insight is available in the amount of emissions of the inland vessels. More knowledge should be gained to estimate how much CO_2 emission this group produces at the moment in order to determine the starting position of the road to zero emission. The PoRA is also interested to know what future developments until 2050 could be concerning this emission of inland vessels.

To measure the CO_2 emission of a specific source (e.g. inland vessel) is not easy. CO_2 mixes well with air which causes it to rise up in the atmosphere. As a consequence it is not possible to assign measured CO_2 emission levels to specific sources. The CO_2 emission is fuel depended. The literature on this topic indicate that the energy or fuel consumption is always required input. The amount of grams CO_2 emission per unit of fuel or energy, called the emission factor, can be calculated. The emission factor multiplied by the number of units fuel or energy that was used leads to the CO_2 emission. Unfortunately the PoRA has no information about the energy or fuel consumption of the inland vessels that visit the PoR.

The main purpose of this thesis is to estimate the CO_2 that will be emitted by the inland vessels in the PoR by 2050 in two steps. The first step is to quantify the current CO_2 emission of this group with the use of an existing model to calculate the energy consumption. The seconds step uses this information to calculate several distributions that make it possible to simulate the CO_2 emission from 2020 until 2050.

This thesis is structured as follows: first Chapter 1 gives background information about global warming, the Port of Rotterdam and the considered problem. Afterwards, the data sets and methodology are described in respectively Chapter 2 and Chapter 3. The method and results of several sensitivity analyses are found in Chapter 4. Two different approaches of the method for the simulations are stated in Chapter 5. Afterwards, Chapter 6 reports on the results and Chapter 7 will conclude on the purpose of this thesis. The last chapter gives recommendations.

1 Background information

The previous chapter introduced the purpose of this thesis. This chapter starts with a short introduction of the global warming issue. Afterwards the focus is on the Port of Rotterdam (PoR): how is the Port of Rotterdam Authority (PoRA) working towards a climate neutral operating inland fleet and what is the current situation concerning the CO_2 emission. The final section elaborates on CO_2 emission of inland ships and how it can be estimated.

1.1 Global warming

All types of motor vehicles, like inland ships, that do not solely use electricity of hydrogen produce several emissions. Examples are emission of particulate matter (PM), sulphur dioxide (SO_2), nitrogen oxide (NO_2) and carbon dioxide (CO_2). PM, SO_2 and NO_2 are examples of emissions that threaten our health when they are inhaled. CO_2 emission is a green house gas that is not considered a direct treat to our health, but causes the global mean surface temperature (GMST) to increase. Apart from burning fossil fuels, CO_2 is also released through natural processes such as respiration and volcano eruptions and through other human activities such as deforestation and land use changes. Human activities are the cause for an 47% increase of the CO_2 concentration in the atmosphere [7]. The goal of the Paris Agreement is to make sure this increase will not exceed 2% above pre-industrial levels and pursue efforts to lower the temperature increase to 1.5 °C [3]. The Intergovernmental Panel on Climate Change (IPCC) states on their website that due to human activities the GMST has currently increased about 1.0 °C above pre-industrial levels, with a likely deviation of 0.2 °C [11]. It also states that if the temperatures will keep increasing in the same pace it has been until now, it is likely that it will reach 1.5 °C between 2030 and 2052. Effects of global warming have already been observed: the number of hot days and temperature extremes on land have increased, the ocean's ecosystem has changed, ice in the south pole is melting and the geographic range of some insects, plants and vertebrates has been reduced [1]. Therefore, it is of high importance that actions are taken to prevent further global warming. Reduction of the CO_2 emission is one of those actions.

1.2 The Port of Rotterdam

The PoRA, the government and the port business community are working together to create a future-proof port. Remco Neumann, the Port of Rotterdam Authority's Corporate Social Responsibility program manager, states in [29] that the goal of a zero emission and noiseless industry and logistics in the harbor of Rotterdam by 2050 is a realistic goal. The mind shift towards sustainability as the new standard is currently happening which causes big companies to set higher requirements concerning CO_2 reduction for their suppliers and transport partners. The PoRA is taking actively part in this transition.

DCMR Milieudienst Rijnmond reported in [21] that the total CO_2 emission in Rotterdam was 29.9 megaton (Mton) in 2018, which was 19% of the total CO_2 emission in the Netherlands. The contribution of sea ships and inland vessels was an estimated amount of respectively 653 and 124 kiloton (kton). In the same year 29.475 sea ships and 107.000 inland vessels visited the PoR [28]. $323.240 * 10^3$ tonnes throughput was outgoing while $145.744 * 10^3$ tonnes came in [27]. Figure 1, that also originates from this report, shows the trend of the CO_2 emission from 2010 till 2018. The first goal is to reach a CO_2 reduction of 49% by 2030 with respect to 1990. Although 2019 showed a decrease from 107.000 to 100.000 inland ships that entered the PoR [30] this number dropped even more in 2020 due to the COVID-19 crisis. The first 6 months of 2020 showed a decline in the throughput of 9.1% compared with the same time period in 2019 [18]. One would think that less throughput indicates less visits of vessels and less CO_2 as a result. These numbers are not available thus far.

The PoRA strives to achieve a carbon-neutral and circular port with the following three steps: efficiency infrastructure (1), transitioning to a new energy system (2) and with a new raw materials and fuel system (3). The industry will take efficiency measures in the first step: residual heat will be used to heat homes and buildings and CO_2 will be captured and stored [25]. The project Porthos strives to store an annual amount of 2.5 million tonnes of CO_2 from the industry in empty gas fields beneath the North Sea as from 2024 [24]. All these developments require additional infrastructure. For step 2 the energy system needs to be changed: from oil and gas as energy resource to the use of electricity, hydrogen and green hydrogen. The third step involves the replacement of fossil fuels by biomass, recycled materials, green hydrogen and CO_2.

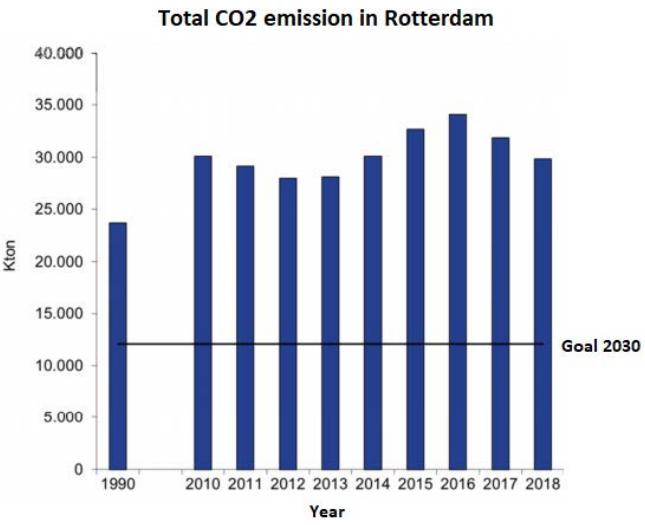

Figure 1: Trend of total CO_2 emission in Rotterdam from 2010 until 2018

In addition, the PoRA is developing a series of initiatives to improve the sustainability of the logistics chain in which Rotterdam participates to reduce CO_2 emission. They for example develop digital tools to help vessels sail more efficiently, introduce incentives for climate-friendly shipping by stimulating sea-going ships to use sustainable fuels, reward vessels that have a Green Award certificate with a discount on port dues and stimulate transition fuels like bio fuel and LNG. These transition fuels eventually will be replaced by electricity, hydrogen and synthetic fuels in the future [25]. The goal is that at least 150 inland vessels have a zero emission powertrain by 2030 [12]. The initiatives are not only focused on the logistics over water, but also on the road. These are outside the scope of this thesis, but can be found on website of the PoR.

1.3 CO_2 emission by inland shipping

To reduce the amount of emitted CO_2 by inland ships, one has to use more climate-friendly energy sources or investigate options to lower the energy consumption. Several scientific papers report on researches that focused on improving operation energy efficiency and how the energy consumption of an inland vessel can be reduced. For example, speed optimisation proved to cause significant reduction of the required energy [23]. Less energy consumption leads to less emissions. That is the reason why the PoRA, the Dutch government and the Municipality of Rotterdam introduced, amongst others, a temporary maximum speed limit in a certain area to ensure that, due to the construction of Maasvlakte 2 and the resulting increase in the freight flows, the air quality does not deteriorate. Another example is a generator ban at public berths introduced by the PoRA. If a vessel is moored it still needs energy to keep all the on-board activities going. With the shore-based power facilities the required energy comes from electricity instead of burning fuel [10]. This thesis does not focus on all the possibilities to reduce or optimise the energy consumption and will therefore not elaborate on these.

The amount of CO_2 an inland vessel emits depends on its energy source and the amount of fuel, or energy, it consumes. Other factors like the type of cargo, the cargo weight, the direction the vessel sails, its velocity, its motor capacity, but also the water levels, water depths and other types of water resistance it experiences during the journey determine the amount of required energy. This

thesis aims to estimate the CO_2 emitted by the inland vessels in the PoR by 2050. As mentioned in the introduction, it is difficult to assign measured CO_2 emission to a specific source. It is therefore estimated by the multiplication of the fuel or energy consumption with the emission factor of the energy source.

Estimation of the energy consumption
Since the impact of sea ships on emissions is much higher than the impact of inland ships less literature was available on the estimation of the CO_2 emission or energy consumption of inland vessels specifically. The developed methodology and models to calculate the emissions of the sea ships cannot be applied to the inland ships since some variables have important differences. For example, sea ships sail with higher speed, sail on water that is much deeper and wider and experience other weather conditions and water resistance. Trozzi proposed in [31] and [32] methodologies to calculate emissions and the fuel consumption. Both his models for the emission calculation require the fuel type (bunker fuel oil, marine diesel oil/marine gas oil (MDO/MGO), engine type (slow-, medium-, and high-speed diesel, gas turbine and steam turbine) and mode (cruise (opens sea), manoeuvring (approaching harbours) and hotelling (at the dock in port)). The PoRA does not have these data, which makes it difficult to use these models. His models to estimate the fuel consumption require specific information about the vessel or its engine which the PoRA also does not have.

De Vlaamse Milieu Maatschappij commissioned a research that was published in 2007 that aimed to calculate several emissions of sea ships, inland vessels and rail traffic for the years 1990 till 2030 with respect to 2005 [16]. Their methodology include the calculation of the energy consumption, fuel consumption, engine power together with the emission factor to obtain the emissions. [16] puts more focus on the approach that uses the fuel consumption. Perhaps this method was chosen because it suited the available data better. Another explanation could be that, besides CO_2 emission, other emissions like NO_2 and PM are included as well. The calculation of these emissions require more input than just the energy consumption. In this thesis the energy consumption is sufficient. [16] uses an existing model [5] to calculate the energy consumption per waterway. Because not all input data was available for all years the energy consumption is calculated only for 2001 and 2005. From these results the energy consumption per tonkilometer is obtained. Subsequently this information is used to calculate the energy consumption for the years for which only the number of tonkilometers is known. For the scenarios until 2030 a mean growth percentage for the number of tonkilometers is determined. These are also multiplied by the earlier calculated energy consumption per tonkilometer to obtain an estimate of the energy consumption for the years until 2030.

Emission factor
The majority of the inland vessels use a type of diesel oil, called EN590, as energy source. As mentioned in Chapter 1.2, the PoRA stimulates vessels to use transition fuels and eventually to have a zero emission powertrain. This sounds easier than it is. Most of the engines that are currently used are not suitable to use 100 % transition fuels, electricity or hydrogen. In order to use these a new engine needs to be purchased or the vessel can undergo certain conversions. These are both very expensive investments and do not happen often for that reason. In the meantime vessels can use an admixture of diesel oil with bio fuel. Although the use of an admixture is already obligated for vehicles on the road this is still not obligatory for inland vessels. Starting from January 2022 it is also obligated for inland ships to use an admixture. There will be 7% bio fuel mixed[1]. The National Climate Agreement states that the ambition is to add at least 30% bio fuel with the diesel oil by 2030 [15].

[1]According to N. Mak, sales engineer at Den Hartog b.v.

2 Data sets

The input data for this thesis are obtained from multiple data sets: AIS data, Verblijfsmonitor and two Excel sheets (called Excel sheet 1 and 2). The data set that will be used consist of the observations in the AIS data set that are found of the vessels that appear in Excel sheet 1. Excel sheet 2 supplies more specific information about each vessel. The Verblijfsmonitor data set makes it link an observation to a specific visit to the PoR. This chapter gives an overview of these data sets with a brief description.

2.1 Excel sheet 1

Aiara Lobo Gomes, data scientist at the PoRA, provided an Excel sheet that contains information of 4672 vessels about for example the mmsi number, the segment(s) it operates in, the total number of visits and the amount of tonnes goods transported, called sum throughput, to or from Rotterdam in 2018 (see Table 1). This data is received from terminals and barge operators. The purpose of this data set in this thesis is obtain a list of vessels for which AIS data will be imported and the segment they operate in. To accomplish this, only the columns *mmsi number* and *segment* are necessary. The segments in this data set are: 'tankers', 'pusher boats', 'containers', 'dry bulk', 'break bulk', 'passenger ships', 'bunker ships' and 'tow boats and other vessels'. Chapter 3.2 will show that the Monte Carlo simulations strongly depend on the throughput. Since the pusher boats, passenger ships, bunker ships and tow boats and other vessels do not contribute to the total throughput in the PoR, they are excluded in this thesis. The remaining segments are tankers, containers, dry bulk and break bulk. If the PoRA desires to do more statistical inference the advise is to start with extensive data cleaning, because several inexplicable anomalies were observed. The anomalies that do not involve the mentioned columns of interest are outside of the scope of this research.

Column name	**Data type**	**Description**
mmsi number	string	9-digits identification number
segment	string	segment in which the vessel operates
tonnage	float	maximum capacity of the vessel in tonnes
maximum draught	float	maximum draught of the vessel in meters

Table 1: Used columns of Excel sheet 1

This data set contains 741, 1208, 1528 and 888 vessels that operate in respectively the segments break bulk, containers, dry bulk and tankers. There are 692 vessels that operate in two segments and 425 in three segments. The numbers in Table 2 show that the majority of the vessels that operate in multiple segments operate in at least containers and dry bulk. Maaike Dalhuisen, Adviseur Business Strategy at the PoRA, mentioned that it is common for container ships to also transport a small(er) amount of dry bulk. This could explain the observed multiple segments per vessel. Multiple segments for a vessel leads to overestimation of the energy consumption calculation, because all the observations will be imported an equal amount of times as the number of segments for a vessel. Therefore it is important to ensure there exist only one segment per vessel. The initial approach was to produce results per segment in each group (see Chapter 3.1.2). Unfortunately, it turned out that there was insufficient data available to accomplish this. As a consequence the information about in which of the four segments a vessel operates is not necessary anymore. The data set will include the vessels that that operate in at least one of the mentioned segments. Nevertheless an estimate of the number of vessels per segment may be interesting, since this variable influences the movements of the vessel in the PoR. Therefore the segment for which the vessel showed the highest throughput is chosen as the segment. The result is that the data set contains 247, 314, 1381 and 881 vessels that operate in respectively the segments break bulk, containers, dry bulk and tankers. This are 2823 vessels in total. This is the list of vessels for which the AIS data will be attempted to retrieve as will be explained in Chapter 2.3.

Number of vessels	Segment 1	Segment 2	Segment 3
55	break bulk	containers	
175	break bulk	dry bulk	
452	containers	dry bulk	
2	containers	tankers	
8	dry bulk	tankers	
424	break bulk	containers	dry bulk
1	containers	dry bulk	tankers

Table 2: Combinations of segments in case a vessel operates in multiple

The model to calculate the energy consumption needs the *maximum draught* and *tonnage* as input. There exist cases where the vessel has two values for the variable *tonnage*. It is unknown why this occurs. For these vessels the maximum value is chosen under the assumption that this value corresponds to the main segment the vessel operates in. Furthermore there are 185 vessels that have more than 1 value for the variable *maximum draught*. One would think the maximum draught is vessel specific and should therefore only have one value. The reason why this is not always the case is unknown. It can not strictly be explained by the multiple segments a vessel operates in. There also exist cases where the value is very small (< 0.3 meters) or -1. The latter perhaps denotes that the value for this variable is unknown. Since the variable denotes the maximum draught, the maximum value is chosen in case a vessel has multiple values.

2.2 Excel sheet 2

A. Lobo Gomes provided a second Excel sheet that contains information about 26922 vessels. It includes information such as the mmsi number, barge category and owner of the vessel, but also characteristics like year of construction, length and beam. From this Excel sheet the columns *mmsi number*, *year of construction*, *length*, *beam*, *maximum draught* and *tonnage* are used (Table 3). Excel sheet 2 is merged with the obtained list of vessels in Excel sheet 1 in order to create one data set with all the information of the vessels. There are 145 vessels of Excel sheet 1 that do not have a record in Excel sheet 2. Therefore important information is missing of this group.

Column name	Data type	Description
mmsi number	string	9-digits identification number
year of construction	int	year the vessel was built
length	float	length of vessel in cm
beam	float	beam of vessel in cm
maximum draught	float	maximum draught of vessel in cm
tonnage	float	maximum capacity of the vessel in tonnes

Table 3: Used columns of Excel sheet 2

Figure 2 shows that, of the group of vessels from Excel sheet 1 that do have a record in Excel sheet 2, a very large part of the vessels have no value or value 0 for the *year of construction*. Chapter 2.5 will describe the method for imputation of the missing values in the final data set.

Maximum draught
Both Excel sheet 1 and 2 have a columns for the *maximum draught*. However, in Excel sheet 2, 45.304 observations have no value or value 0 for the *maximum draught* while there are none for the vessels in Excel sheet 1. Therefore the *maximum draught* of Excel sheet 1 will be used. There still exist cases where the value is -1, which also denotes the value is unknown. For these cases

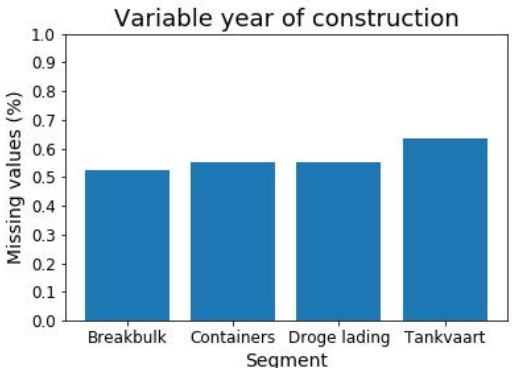

Figure 2: Missing values of variable year of construction

the *maximum draught* will be equal to the median value of the corresponding CEMT class [2]. To accomplish this an CEMT class has to be assigned to every vessel. A simple approach is used: the maximum capacity of the vessel determines its CEMT class. This means that the *maximum capacity* is compared with the tons deadweight for each CEMT class mentioned in [20].

Maximum capacity
Both Excel sheet 1 and 2 have a column *tonnage* that denotes the maximum capacity. Except for 1 missing value for the *tonnage* in Excel sheet 2, a value for all the vessels is present. For a little over 50% of the vessels these values are the same. For the remaining vessels the values in the two data sets are not the same. For the majority of this sub-group the values do not differ more than 10 tonnes. For the others differences of at most 500 tonnes are observed. It is unknown these differences exist. This could (partly) be the result of the assumptions that are made for the *tonnage* in Excel sheet 1. The decision is made to use the *tonnage* from Excel sheet 2.

2.3 AIS data

The most important data set for this research is the AIS data set. AIS stands for Automatic Identification System. It is an automatic tracking system that is installed on the vessel and frequently sends out static (i.e. vessel name, length, beam), dynamic (i.e. speed, position, course) and journey-specific (i.e. destination, estimated time of arrival) information of the vessel. A signal with static information is send out every 6 minutes, dynamic information every 3 minutes (in case it is moored) or 2-10 seconds (in case it is sailing) and journey-specific information every 6 minutes. Data of other vessels that have this system installed are also frequently received. It is a standardized procedure to automatically exchange nautical information between vessels and between vessels and shore. Inland AIS is a version of AIS, but is aimed at the specific needs of professional inland shipping [22]. The installation and use of an Inland AIS is obligatory for vessels that make use of waterways of CEMT class I or higher, which includes all the relevant classes for this thesis.

In case the inland vessel is also used as residence the owner has to give the PoRA consent to use the AIS data. This consent relates to identifying data that are incorporated with the AIS signal like vessel name, MMSI number, position, course, draught, destination and speed over ground. The PoRA uses this data in and for its products and services to promote trade and transport in the logistic chain, but also for internal testing, analysis and model training purposes [26]. Only the data of vessels for which consent was given is used in this research. Therefore, the results of this

[2]CEMT stands for Classification of European Inland Waterways. The classification of a waterway denotes to which vessels it is accessible. For each class the maximum values of vessel characteristics (i.e. length, beam, draught) is given. The most relevant classes are I till VII. For example, class I is accessible for vessels with a maximum weight of 400 tonnes and class VII for a maximum weight of 27000 tonnes.

research are based on a subset of all the inland ships that are active in the harbor.

Column name	Data type	Description
mmsi	string	9-digits identification number
recordTime	datetime	date and time of data point
draught	float	distance from the keel [3] to the waterline
location.coordinates	list[float, float]	longitude and latitude of location
speedOverGround	float	speed over ground in knots

Table 4: Used columns of the AIS data set

The AIS data set contains 11 columns in total from which 5 are used in this thesis (see Table 4). The columns *accurate*, *destination*, *eta*, *heading*, *location.type* and *courseOverGround* are excluded. Table 5 shows an anonymized sample of the columns that are used. The sailor has to manually enter the destination, eta and draught at the start of a new journey. This makes it very prone to errors and as a consequence the quality of the data decreases. Also, the column *destination* shows to have typographical errors, which causes operations like aggregation (in Python) to produce incorrect outputs. Furthermore, if the sailor forgets to enter the destination and eta at the start of a new journey the values from the previous journey will be retained. Moreover the option to enter information about these variables, along with the draught, was recently added by the PoRA on March 12th 2020. From this date every inland vessel owner needs to again give consent to the to save this data. Therefore the start date of the availability of these data varies per vessel. There are cases where the *draught* values have the same value for every journey or do not have a value at all. This makes that the quality of this information questionable. This is a significant challenge, since the energy consumption strongly depends on the draught information of the vessel. This will be discussed in Chapter 2.5.2.

mmsi	recordTime	draught	location.coordinates	speedOverGround
*********	2020-04-** 22:00:29	2.9	[4.01632, 51.97514]	4.424
*********	2020-04-** 22:00:54	2.9	[4.01802, 51.97476]	4.424
*********	2020-04-** 22:01:14	2.9	[4.01930, 51.97446]	4.476
*********	2020-04-** 22:01:35	2.9	[4.02062, 51.97416]	4.477
*********	2020-04-** 22:02:14	2.9	[4.02298, 51.97361]	4.527

Table 5: Anonymized sample from the AIS data set

AIS data from April 1 until September 30 2020 is retrieved of the vessels that were obtained in Chapter 2.1. Not all the vessels have observations in the AIS data set. This time frame is chosen, because of technical limitations and due to the data of the draught is only recently available. From the 2823 vessels in 2.1 observations of only 409 vessels can be retrieved (Table 6). These are the observations that are used to determine the *group* of each vessel (Chapter 3.1.2). For all the other calculations in this thesis only the observations that occurred in the PoR will be used.

Furthermore, several high values for the *speed over ground* are observed. There is a speed limit of 13 km/h, which is 3.61 m/s, in some areas of the PoR. Therefore the assumption is that values

[3] "The keel of a ship is similar to the spinal cord of humans. As the spine functions to keep our backbone upright by linking and supporting our body, the keel is the primary structural member and backbone of the vessel which runs along the centreline of the bottom plate around which the hull of the ship is built. It is the main longitudinal component of the ship to which every other main structural item is connected directly or indirectly."[4]

higher than 6.0 m/s are incorrect. These are excluded. After this data cleansing, a large part of both the observations and number of vessels is lost as a result. This will also show later in Table 8.

Segment	**Total number of vessels**	**Number of observations**
Break bulk	20	151.020
Containers	121	2.093.328
Dry bulk	112	1.178.711
Tankers	156	4.209.825
Total	**409**	**7.632.884**

Table 6: Characteristics AIS data April until September 2020 per segment

2.4 Verblijfsmonitor

The Verblijfsmonitor data set is a composed data set that contains information about all the ships that visited the PoR. This data is derived from data sources like AIS, radar and Portmaps. Besides the mmsi number, start and end date of the visit and ship specifications (e.g. length, beam) also the subclass, category, berth type and presumed cargo type can be found here. Every visit is given a visit ID, which is the column *vesselvisitid*. This data set is used to link an observation, or a set of observations, from the AIS data set to a visit ID. Therefore solely the columns *mmsinumber*, *vesselvisitid*, *start date*, *end date* and *port count* are used (see Table 7). The PoRA has divided the harbor into different subareas for operational purposes. This means that a visit to the harbor consists of multiple visits to subareas of the harbor. A visit ID has therefore multiple rows in this data set. The variable *port count* is a binary variable that indicates if the information in that row is about the total visit to the harbor (1) or about a visit to a subarea (0). Due to memory limitation solely the rows where *port count* is equal to 1 are used.

Column name	**Data type**	**Description**
mmsinumber	string	9-digits identification number
vesselvisitid	string	visit identifier
startdate	datetime	start date and time of visit
enddate	datetime	end date and time of visit
port_count	binary	1=harbor visit, 0=subarea visit

Table 7: Used columns of the Verblijfsmonitor data set

Table 8 shows the number of visit IDs per segment that are linked to the observations obtained from the AIS data set. The numbers in Table 6 refer to the observations that occurred inside and outside Rotterdam. Obviously, only the observations that occurred in the PoR have a visit ID.

Segment	**Number of vessels**	**Number of visits IDs**	**Number of observations**
Break bulk	3	18	10.004
Containers	17	511	240.933
Dry bulk	14	168	72.535
Tankers	89	1193	273.352
Total	**123**	**1890**	**596.824**

Table 8: Statistics of final data set per segment

2.5 Imputation of missing values

2.5.1 Year of construction

Chapter 2.2 described that more than half of the vessels in the data set of Excel sheet 2 have no value for the *year of construction*. From the remaining vessels in the final data set 2, 8, 5 and 57 vessels of the segment break bulk, containers, dry bulk and tankers the value of this variable misses. This is respectively 67%, 47%, 36% and 64% of the total number of vessels in the final data set. Due to insufficient data the distributions in Chapter 3.2.1 are not segment specific. Listwise deletion is not an option. Too much data would be lost. The consequence of imputing the missing values with one value, like the mean or median, is that the *year of construction* will be equal to this value for the majority of the vessels. Consequently, this value will have the highest probability to be chosen in the simulations, as will be seen in Chapter 3.2. This leads to invalid conclusions. Since it is unknown at this point if the year of construction is related to other characteristics of information a distribution is fitted to the total data set of Excel sheet 2. There are 5736 rows that have a value for the *year of construction*. The 44 rows that contain a value smaller than 1900 are assumed to be outliers and excluded for that reason. The remaining data points do not follow a specific distribution. Therefore the missing values are imputed with draws from the non-parametric distribution, a histogram, of the data. If the PoRA desires better estimations more effort should be invested in gathering more data about the year of construction of vessels or investigating other methods that could lead to better estimations. The consequence of keeping this variable in this thesis despite the large amount of missing values will be explained in Chapter 3.2.1.

2.5.2 Draught

The variable *draught* plays a significant role as it occurs twice in the model that will be used to estimate the energy consumption (Chapter 3.1). As mentioned in Chapter 2.3 the quality of this data is questionable. Unfortunately, 42% of all the observations do not have a value for the *draught* or have the value 0. This could denote that no consent to use this data was given (yet). There are also 10115 observations for which the *draught* is larger than the *maximum draught*. These values are assumed to be wrong and are therefore put to 0, which means they are treated as missing values. Van Dorsser et al. investigated in [6] what the minimum required operational draught is per type of ship (CEMT class). Every vessel was already assigned to a CEMT class (Chapter 2.2). All the missing values for the *draught* are imputed with the corresponding minimum required operational draught.

3 Method

This thesis consists of three parts: first the energy consumption of the inland ships in the harbor of Rotterdam will be estimated. Secondly, distributions are derived from the result of the first part. Lastly, simulations will be produced to create scenarios about the development of the CO emission of inland shipping from the year 2020 till 2050. The distributions are an important part of the simulations. Bolt developed a model in [5] to estimate the energy consumption for inland ships. This model is used in the first part. Empirical distributions are derived in the second part. The third part applies the Monte Carlo simulations technique. This chapter describes the methods and the application.

3.1 Estimation of the energy consumption of inland ships

The energy consumption is calculated with the model Bolt developed in [5]. This model explains how an estimate of the amount of energy an inland ship needs for its propulsion is constructed. This method is not suitable to make calculations for an individual ship, because in that case a method that includes more details has to be used [5]. Information about the type of fuel the vessel uses and the age of the engine are examples of factors that could make the estimation of the energy consumption more precise.

$$E = 2[53(\log(u_{rel}L) + 4)^{-2}(LB + 2LT_{gem}) \cdot u_{rel}^2 + C_p \frac{1}{2}\rho V^2 BT + C_z \rho g z BT]\frac{u_{rel}}{V}\Delta x \tag{1}$$

$$u_{rel} = max(V + u, V_{eff}) \tag{2a}$$

$$u = \frac{A_m/A_c}{1 - A_m/A_c - F_{nh}^2} \cdot V \tag{2b}$$

$$V_{eff} = \frac{1}{1 - e^{4(1-\frac{h}{T})}} \cdot V \tag{2c}$$

Variable	Description	Data set
L	length (m)	Excel sheet 2
B	beam (m)	Excel sheet 2
T_{gem}	Average draught (m)	not available
C_p	coefficient for residual resistance (in Dutch restweerstand)	0.15 [5]
ρ	mass density of water (km/m^3)	997 [5]
V	speed over water (m/s)	not available
T	draught measured at lowest point (m)	AIS data
C_z	coefficient for waterspiegeldalings-weerstand	0.20 [5]
g	gravitational acceleration (m/s^2)	9.812
z	sinkage (m)	not available
Δx	distance traveled over water (m)	not available
h	water way depth (m)	not available
A_m	midship section surface area (m^2)	not available
A_c	water way cross section (m^2)	not available

Table 9: Model variables for estimation of energy consumption

Bolt's model is constructed out of several components. The amount of energy that an inland ship needs depends on how much energy it costs to overcome the various types of resistance it suffers from. The resulting formula to calculate the energy consumption after taking all the different types

of resistance in consideration is shown in Equation 1. The formulas to calculate the variable u_{rel} are sub-equations 2a, 2b and 2c. This variable incorporates the effect of shallow water. Shallow water leads to higher water resistance, which has a negative effect on the speed over water of the vessel. If the water resistance increases, the vessel's energy consumption also increases if it wants to maintain its speed. u is larger than V_{eff} when the water way is shallow and respectively narrow, while the opposite is true when the water way is shallow but wide. Since the water ways in the PoR are deep and wide, one would expected that neither u nor V_{eff} play a significant role in this research. If so, this will likely make u_{rel} (almost) equal to V. Consequently, the expectation is that all variables (h, w and A_m) that are only found in the calculation of u_{rel} will not have a significant impact on the energy consumption (E in Equation 1). The results of the sensitivity analysis in Chapter 4 will show if this expectation is correct. For more detailed information about the model see [5]. Table 9 shows the description of all the variables and from which data set they initially originate.

Note that if a vessel does not move this model will output an energy consumption of 0 since the distance that is traveled, Δx, and the speed, V, are 0. This is not the case in reality. Aside from the energy required for propulsion, on board activities also need energy. Therefore, the diesel generator cannot be turned off when the vessel is moored. Nevertheless, as mentioned in Chapter 1.3, the PoR introduced a generator ban for inland ships at public berths in the harbor. This means that if the vessel is located in the harbor at public berths it must use the shore-based power facilities if this is possible. Shore-based power is not available throughout the PoR. More investigation is needed to develop a model that is suitable to calculate the energy consumption in these cases. This is out of this scope of this research. Therefore, these observations are excluded. One could therefore conclude that in this thesis the assumption is that if a vessel does not move, it does not use any energy and it emits no CO_2.

3.1.1 Missing input variables

Table 9 shows that the variables *average draught*, *speed over water*, *sinkage*, *distance traveled over water*, *water depth*, *midship section surface area* and *cross section of the water way* are not found in the data set. This section gives a description of each variable, the reason why it is missing and how its value will be chosen in this thesis. To verify that this chosen value is valid a sensitivity analysis will be conducted when needed, in Chapter 4. This means that the chosen values in this section are initial values that could change after review of the sensitivity analyses results.

Average draught
Sometimes, instead of the vessel laying horizontal in the water, the bow (i.e. front) or the stern (i.e. back) of the vessel are not equally deep in the water. This can have various reasons. Van Dorsser et al. state in [6] that if a vessel sails upstream its bow is loaded slightly deeper than the stern to avoid the risk of spinning around when getting grounded. Vice versa if the vessels sails downstream. That is the reason why it is common for loaded inland vessels to apply a trim [4]. Bram Visser and A. Lobo Gomes, both data scientists at the PoRA, assume that the *draught* in the data set corresponds to the draught measured at the lowest point. This means that the average draught, T_{gem}, is missing. This information could only be obtained if the sailor would submit this manually. The initial value for T_{gem} is chosen as T. Since in reality T_{gem} is smaller or equal to T but never larger, the assumption for T_{gem} is correct or overestimated. In case it is overestimated this will lead to a energy consumption that is higher than it should be. The impact of this overestimation is investigated in the sensitivity analysis in Chapter 4.1.

Speed over water
The speed over ground, V_a, is present in the data set. The speed that a vessel has over the water, V, can be different from V_a, because of the tidal streams. Therefore V is defined as V_a plus the

[4] "The trim of a vessel describes its floating position in length direction, namely if the bow (i.e. foremost part) or the aft (i.e. rear) of the vessel is deeper submerged into the water. The trim can have a significant impact on a vessel's energy demand for propulsion during sailing. The most efficient trim for a particular ship depends on its design, operational draft (i.e. draught) and speed." [14]

effect of the tidal stream on the vessel's speed. Since it was not possible at this time to include the tidal stream information for each observation, which requires the tidal stream on a exact date, time and location, V is initially chosen to be equal to V_a. A sensitivity analysis will be carried out in Chapter 4.2 to investigate how the tidal stream influences V and the output of the model, the energy consumption, as a consequence.

Sinkage
The sinkage, z, of the vessel is the difference between the actual waterline and the waterline when the vessel is empty (i.e. lightweight). When the ship is empty $z = 0$ and $T = T_{min}$, where T_{min} is the draught if the vessel is unloaded (i.e. empty draught). T is the *draught* found in the data set. When the vessel transports its maximum capacity with respect to the cargo weight (i.e. design capacity), both z and T are also maximized. This leads to $z = T - T_{min}$. The empty draught, T_{min}, is estimated with the model Van Dorsser et al. developed in [6]. This model uses the vessel's *beam*, *length* and *maximum draught* as input. With T from the data and the calculated T_{min} the *sinkage* can also be calculated.

Distance traveled over water
Since the data set only contains the speed over ground, V_a, the assumption is that the distance between the coordinates, calculated with the Haversine formula, corresponds to the distance over ground, Δx_a. Unfortunately, the model needs the distance traveled over water, Δx. The variable Δx can be calculated by the multiplication of V by Δt. As mentioned before, the variable V is the speed over water and is initially chosen as V_a. The time between observations, Δt, is obtained by subtracting the time of the current observation from the time of the previous observation. There exist cases where the time between observations is very large. It could be that the sailor turned off the AIS transponder. Unfortunately, this makes the value of Δx also very large and leads to a very large output of the model. Since it is unknown what happened between these observations and since they have a big the impact on the output, the observations where Δt is larger than a threshold are excluded. Chapter 2.3 stated that the AIS transponder should send out a signal every 6 minutes. This value is broadly taken and translated into a threshold of 10 minutes. Additionally, there also exist cases where Δt is 0, because there is no difference between the times of the observations but strangely enough the coordinates are not the same. The total CO_2 emission calculated with Δx was twice as large as the total CO_2 emission calculated with x_a, despite of the introduced threshold $\Delta t < 10$ minutes. It is unknown which information is incorrect. But since several errors in Δt were observed, the assumption is that the coordinates are of better quality than the time stamps. Therefore the decision is to use Δx_a instead of the calculated Δx in this thesis.

Unfortunately also errors were found in the coordinates. These errors were only noticed because the coordinate of an observation with a large value for Δx_a and the coordinate of its previous observation were manually looked up on Google Maps. An example was encountered where the coordinate of an observation was not near the coordinate of its predecessor and successor. It was also at a location that was on land instead of water. For each faulty observation of this type arise two faulty Δx_a values and two CO_2 emission values as a consequence. Some of these observations could be easily excluded, because their value for Δt was 0. Unfortunately, also observations where Δx_a is large and $\Delta t > 0$ were sometimes incorrect. Since it is not possible to automatically distinguish correct and incorrect observations for which Δx_a is large the following assumptions are introduced: observations where ($\Delta x_a > 100$ meters and $\Delta t = 0$ seconds) and observations where ($\Delta x_a > 3000$ meters) are excluded.

Water depth
The water depth, h, is of high importance for all the vessels. The PoRA guarantees a certain depth at every location in the harbor. Therefore dredging is an on-going activity. Figure 3, that was found in [8], gives a graphical view of the water depths. Unfortunately no legend of the colours is available. Shallow and deep waters are respectively yellow and blue. The areas that are colored deep blue are the routes the sea ships use. The water ways more to the East are used by inland ships and other relatively small vessels and are therefore less deep. Due to technical issues it was

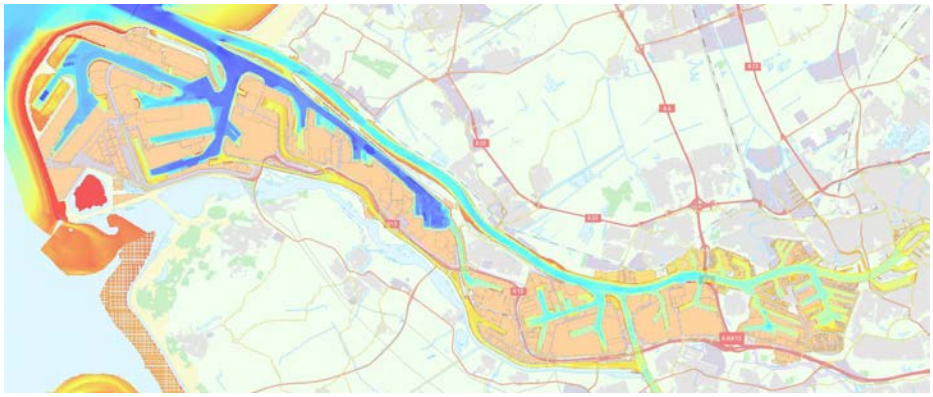

Figure 3: Water depth in the harbor of Rotterdam

not possible to import this data. Since it is not possible to come up with an assumption for the value of the *water depth* that is location specific a constant value of 15 meters is initially chosen for all the observations throughout the harbor. How changes in the value of this variable influence the output of the model will be investigated in Chapter 4.3.

Midship section surface area
The midship section surface area[5], A_m, is the surface area of the largest intersection of the vessel. This information does not appear in the data set. The shape of the midship section of cargo vessels is rectangular according to C. van Dorsser, knowledge developer at Koninklijke BLN-Schuttevaer. Information about the *beam*, which is the width at the widest part of the vessel, does appear in the data. The height of the vessel, is not available, but is obviously at least T_{max}. Therefore, the assumption is that the height is equal to T_{max}. The midship section surface area of a vessel is defined as the *beam* multiplied by T_{max}.

Water way cross section
There is no information was received about the cross section of the water way, A_c, in the harbor. Therefore it is calculated as the multiplication of the water depth, h, by the water way width, w. Unfortunately, it was not possible to receive the data of w as needed. Similar to h, one constant is chosen for all the observations throughout the harbor. This value for w is initially set to 100 meters. This variable is also included in the sensitivity analyses (see Chapter 4.4).

3.1.2 Groups

The PoRA asked to distinguish 4 types of groups within the total population of inland vessels that visit the PoR:

- Group 1: inland ships that operate only in the PoR or South-Holland
- Group 2: inland ships that operate on fixed, short routes
- Group 3: inland ships that operate on fixed, long routes
- Group 4: inland ships that do not operate on fixed routes

To accomplish this, polygons are used. A polygon is a set of coordinates that are located in a certain area. The polygons of the PoR and every province in the Netherlands, Belgium and Germany are included. For every observation will be determined in which polygon it was detected. If the observation was located outside the mentioned polygons, it is labeled as 'other'. Together with M. Dalhuisen, the following rules are defined: a route is fixed when the vessel visits the polygon more

[5]In Dutch grootspantoppervlakte

than 2 times per month, otherwise it is non fixed. A route is long when the distance from the PoR to the polygon is more than 260 kilometers, otherwise it is short. This number is an estimate of the distance between the PoR and the closest border of Germany.

If all the observations of a vessel were located in the PoR or South-Holland, then its *group* value is 1. More steps are required to determine if a vessel classifies as group 2,3 of 4. The polygon that the vessel visited the most is used to determine if the route was fixed or non fixed. If this number was 2 or higher, the route is seen as fixed, otherwise it is non fixed. If the distance from this polygon to the PoR is 260 kilometers or further away the route is labeled as long, otherwise short. These steps are performed for each month. As a consequence a vessel can have multiple group classes. The class that occurs the most often is chosen. There exist cases where multiple classes occur the same number of times. The assumption is that the order of importance with respect to the groups is 4, 3 and lastly 2. This means that, for example, if a vessel is classified 3 times for both group 4 and 3, group 4 is chosen.

Note that the reason why a vessel was located in a polygon is unknown. Possible reasons are that it delivered or picked up goods, the polygon was on the route from or to another polygon or it visited for other activities. Also, observations where the location is marked as 'other' could be anywhere outside of the Netherlands, Belgium of Germany. Therefore, the number of times a vessel visited polygon 'other' can quickly pass the threshold of 2. As a result the vessel is assigned to group 3, even though every occurrence in 'other' could haven been to different locations. The consequence is that there are probably more vessels classified as group 3 than there should be. This problem could be overcome by adding more data of other polygons. Due to memory limitations this was not done in this thesis. Table 10 shows that very few vessels are assigned to group 1. It is unknown if this is caused by the implementation of the rules. After all, if a vessel that operates only in the PoR or South-Holland but is located outside of the polygons just once, it is not assigned to group 1. These reasons cause the results in Table 10 to be unrealistic. Therefore the decision is to derive the distributions in Chapter 3.2.1 from the complete data set without any aggregation by group. Distribution A is an exception as will be explained in the section.

Segment	**Group**			
	1	**2**	**3**	**4**
Break bulk	0	1	1	1
Containers	0	6	10	1
Tankers	2	28	28	31
Dry Bulk	0	2	5	7

Table 10: Number of vessels per group per segment

3.2 Monte Carlo simulations

In the second part of this thesis, computer simulations are performed. One simulation produces one scenario. One scenario consists of the calculated CO_2 emission for every year from 2020 until 2050. The amount of CO_2 emission depends on the emission factor and the energy consumption. Chapter 3.2.2 describes how the emission factor is incorporated in the simulations. The model of Bolt in Equation 1 is again applied to calculate the energy consumption. But now, the values for the input variables will be random draws from the distributions that are derived from the results of the first part of this thesis. There are distributions for the vessel characteristics (*group*, *length*, *width*, *maximum capacity*, *maximum draught* and *year of construction*) and visit characteristic (*speed*, *distance traveled*, *capacity rate* during the visit and the total *number of visits* per year). Since there is randomness involved, simulations with these distributions are called stochastic computer

simulation or Monte Carlo simulations [6]. The idea is that if one would draw a large number of times from each distribution and calculate the CO_2 emission given these values this will give a distribution for the CO_2 emission. In other words, the simulations will indicate the probability that a range of outcomes will occur. The values for the CO_2 emission in 2050 that occur the most are assumed to have the highest probability to actually happen.

Algorithm of the simulations
One simulation comprises several steps in order to calculate the CO_2 emission for a year. The number of visits in a year is determined by the total throughput of that year. The PoRA made a throughput forecast for 2020, 2030 and 2040 for each segment. The throughput is the total amount of tonnes (or containers) that is transported in and from the PoR. The forecast showed an increase of 2.76% between 2020 and 2030 and an increase of 1.34% between 2030 and 2040. The assumption is that the increase in between these years is linear. Extrapolation leads to an estimate of the throughput in 2050. The characteristics of the vessel are determined by draws out of distributions A, D, E, G, H and I. The next section will elaborate on these distributions. The *maximum capacity* denotes the maximum tonnes a vessel can transport. Afterwards the *number of visits* to the PoR by the vessel is drawn from distribution B. For each visit of a vessel, the visit characteristics are drawn from distributions C, F and J. The *capacity rate* denotes the fraction of its maximum capacity the vessel transport on the visit. Therefore, the *capacity rate* times the *maximum capacity* gives the *capacity* (i.e. amount of tonnes) the vessel transport on the visit. This number is subtracted from the total throughput. These steps will continue until the throughput for the year is fully transported. In other words, in the simulation these steps will continue while the throughput is larger than 0.

The *draught* during a visit is seen as the *capacity rate* multiplied by its *maximum draught*. The *sinkage* is determined by the *draught* and the *empty draught*. The latter is calculated with the model of van Dorsser et al in [6] (see Chapter 3.1.1). Now all the input variables for the model of Bolt (Chapter 3.1) are present, the the energy consumption is calculated. The *group* of the vessel (distribution A) and its *engine age* determine if it uses diesel oil or electricity/hydrogen as *energy source*. With the latter no CO_2 is emitted. In case the vessel uses diesel oil, the CO_2 emission is calculated by multiplying the energy consumption by the emission factor (see Chapter 3.2.2). Therefore, the amount of CO_2 that is emitted each year depends on the vessels that use diesel oil and the emission factor.

Algorithm 1 shows the simulation in pseudo-code. Recall that in the simulations, the vessel characteristics are the group, length, width, maximum capacity, maximum draught and year of construction. The visit characteristics are the speed, distance traveled and capacity rate.

3.2.1 Distributions

Recall that the distributions are derived from the data of April till September 2020. This means that in the simulations the assumption is that these distributions will remain the same until 2050. The initial plan was to make the distributions segment and group specific. Unfortunately, as explained in Chapter 3.1.2, the distributions are based on the complete data set without any distinctions by segment or group.

There is not enough data to fit a parametric distribution to the data of every variable. The decision is to be consistent in the applied method to derive a distribution. Therefore, an empirical distribution, a histogram, is derived for each variable. This gives a probability per bin. The smallest bin size for which every bin has at least 1 observation is chosen. If there would exist bins without observations the probability that these bins are chosen is 0. Outlier removal is performed when needed to avoid the bin size to be too large and lose important information as a consequence. For

[6]"The term Monte Carlo was used by Neumann and Ulam during World War II as a code word for secret work at Los Alamos on problems related to the atomic bomb. That work involved simulation of random neutron diffusion in nuclear materials' [19].

Algorithm 1 Simulations

```
for iteration = 1, ..., 1000 do
    for year = 2020, ..., 2050 do
        Retrieve throughput for year
        while throughput > 0 do
            Pick vessel characteristics
            Pick the number of visits I
            Calculate the age of the engine of the vessel
            Determine the fuel type of the vessel
            for visit = 1, ..., I do
                Pick the visit characteristics
                Calculate the transported capacity, draught and sinkage
                Calculate the energy consumption
                Determine energy source
                Calculate the CO2 emission
                Subtract the transported capacity from throughput
            end for
        end while
    end for
end for
```

example, if 10 observations have a value between 1 and 2, 5 between 3 and 4 and one observation has the value 18, the bin size has to be at least 4.5 and more specific information about the distribution till 4 will be lost.

Two different models for the simulations are investigated: a model that includes dependence between some of the variables (model A) and a model that assumes all variables are independent (model B). The models are applied in thee 'pick vessel characteristics' and ' pick visit characteristics' step in Algorithm 1. Chapter 5 elaborates on these models. Both models make use of the following distributions:

- **Distribution A: group of the vessel**
  This distribution indicates the probability per group. There are 4 groups. Although it was decided, due to insufficient data, to not aggregate the results by *group*, this distribution is still needed for the application of the method described in Chapter 3.2.3. Table 11 shows the probability per group.

Group	Probability
1	0.016
2	0.301
3	0.366
4	0.317

Table 11: Distribution A

- **Distribution B: number of visit for the year**
  Distribution B indicates the probabilities of the *number of visits* to the PoR for a year by a vessel. The assumption is that one visit ID, and thus one visit, corresponds to a visit to one destination. In reality this is not always the case. For example, a vessel could transport cargo between terminals within the PoR, which should be seen as multiple visits and with that different *capacity rate*, *distance traveled* and *average speed* could apply. Unfortunately, this will not show in the data as used in this thesis. This can make the distribution look

(very) different. With the used data set this could not be investigated. Perhaps thorough investigation of a similar data set but with the variable port count set to 0 instead of 1 (see Chapter 2.4) it will show. This was not attempted, because it is unclear what the classification of the areas exactly indicate.

The largest value are seen as outliers and therefore an 5% trim is applied to the right side of the distribution. In other words, 5% of the vessels, which is equal to 6 vessels, with the largest values are excluded. Lastly, since this distribution is based on the number of visits in a year and the data set exist of 6 months, the number of visits are multiplied by 2 to get an estimate of the number of visits per year.

Table 12 and Figure 4 show respectively the statistics and histogram of the number of visits per vessel after excluding the outliers. The bin size is equal to 10 visits. The majority of the vessels, 75%, visited the harbor less that 38 times. One would expect that vessels that visited the PoR more than 38 times are classified as group 1 or 2, since this is not possible for the other groups due to long travel times.

Mean	**Std.**	**Min.**	**25%**	**50%**	**75%**	**Max.**
24.45	19.07	2.00	8.00	20.00	38.00	76.00

Table 12: Statistics of the number of visits per year (distribution B)

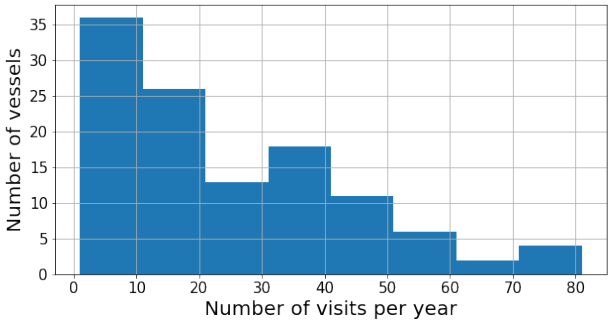

Figure 4: Histogram of the number of visits per year (distribution B)

- **Distribution C: capacity rate for the visit**
  This distribution indicates the *capacity rate* for the visit. Since the draught is maximized if the vessel transport its maximum capacity, the assumption is that the *capacity rate* is equal to the ratio of the current *draught* to the *maximum draught*. Therefore the *capacity rate* is calculated by the division of the *draught* by the *maximum draught*. Recall that, as mentioned in 2.3 and 2.5.2, the quality of the *draught* is probably low. There are cases where the calculated *capacity rate* is higher than 1. Since this is not possible and these values are therefore set to 1.

  Figure 5 shows that the majority of the vessels have a capacity rate between 0.40 and 0.63 (see Table 13). The minimum capacity rate found in the data is 0.24. The bin size is equal to 0.035 starting from 0.2. Therefore the assumption is that a vessel transports at least 20% of its maximum capacity.

Mean	Std.	Min.	25%	50%	75%	Max.
0.53	0.19	0.24	0.40	0.45	0.63	1.00

Table 13: Statistics of the capacity rate (distribution C)

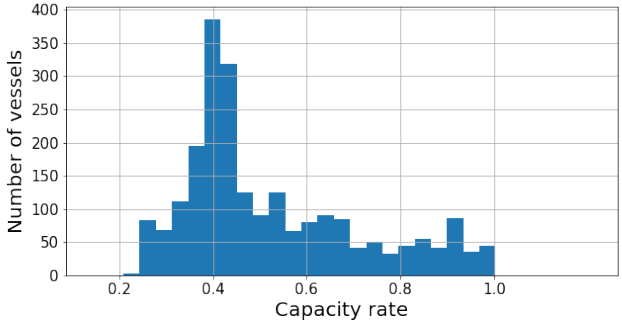

Figure 5: Histogram of the capacity rate (distribution C)

- **Distribution D: year of construction of the vessel**
  The *year of construction* is drawn from distribution D. Recall that the quality of this variable is questionable since the *year of construction* was missing for more than 50% of the original data set (see Chapter 2.2). Unfortunately, the method in Chapter 3.2.3 requires this information. Under the assumptions and missing values imputation (Chapter 2.5.1), the distribution and its statistics are shown in respectively Figure 6 and Table 14. The bin size is equal to 15 years.

Mean	Std.	Min.	25%	50%	75%	Max.
1985.68	19.22	1913.00	1973.00	1985.00	2003.00	2017.00

Table 14: Statistics of the year of construction (distribution D

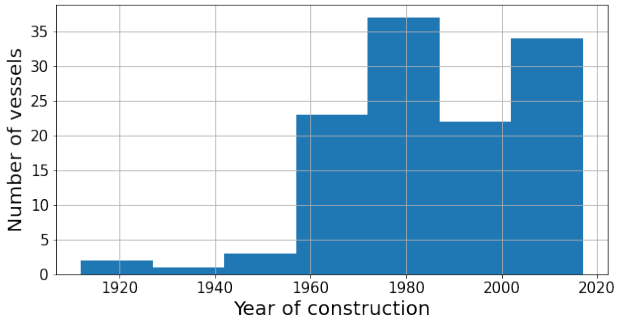

Figure 6: Histogram of the year of construction (distribution D)

- **Distribution E: maximum capacity of the vessel**
  Distribution E indicates the probabilities of the *maximum capacity* of the vessel. In the data set the column *tonnage* denotes the maximum capacity. The multiplication of the *maximum capacity* by the *capacity rate* (distribution C) determines the amount of tonnes cargo the vessel transports on a visit.

Table 15 and the histogram in Figure 7 show how the data for this variable is distributed. The majority of the vessels have a maximum capacity under 3183 tonnes. Recall that only the data of the segments tankers, dry bulk, break bulk and container are used. The bin size is equal to 565 tonnes.

Mean	**Std.**	**Min.**	**25%**	**50%**	**75%**	**Max.**
2606.03	1197.10	250	1687.50	2502.50	3182.25	5900.00

Table 15: Statistics of the maximum capacity (distribution E)

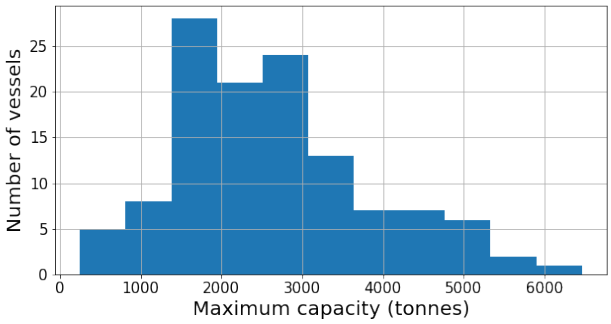

Figure 7: Histogram of the maximum capacity (distribution E)

- **Distribution F: distance traveled for the visit**
  This distribution indicates the distance that is traveled within the PoR on a visit. For every observation in the data set Δx is already calculated for the first part of this thesis. The values for this variable are summed up per visit. This gives the total distance that is traveled per visit. With this information a distribution for the *traveled distance* is derived. It is possible that the values for the total distance traveled per visit are higher than they should be. The same reasoning as mentioned in 'Distribution B' could apply: one visit can be in reality multiple visit. If this is the case, the values as observed in distribution F are equal or larger than the real values. This could also be the cause of the values being very wide spread. Logically, the largest values would be incorrect. The largest values are seen as outliers and therefore an 5% trim is applied to the right side of the distribution. This percentage includes 92 visits. In other words, the 92 visits that have the largest value for the *distance traveled* are excluded.

  Figure 8 and Table 16 show that during 50% of the visits less than approximately 2.1 kilometers is traveled. The maximum is a little over 10 kilometers while the minimum is only 0.95 meters. The latter raises the question if this value is realistic. The bin size is 437 meters.

Mean	**Std.**	**Min.**	**25%**	**50%**	**75%**	**Max.**
2749.44	2163.49	0.95	1121.12	2103.30	3834.81	10038.12

Table 16: Statistics of the traveled distance (m) (distribution F)

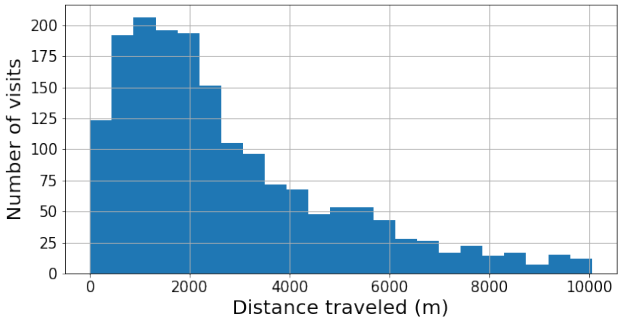

Figure 8: Histogram of the distance traveled (m) (distribution F)

- **Distribution G: width of the vessel**
  This distribution indicates the probabilities of the *width* of the vessel. This variable is found in the data set. The *width* is an input variable for the model of Bolt in Equation 1. Table 17 and Figure 9 show the distribution of the *width*. The bin size is equal to 1.4 meters. The majority of the vessels in this data set have a width between 6.40 meters and 11.45 meters.

Mean	**Std.**	**Min.**	**25%**	**50%**	**75%**	**Max.**
11.23	1.67	6.40	11.40	11.40	11.45	17.42

Table 17: Statistics of the width (m) (distribution G)

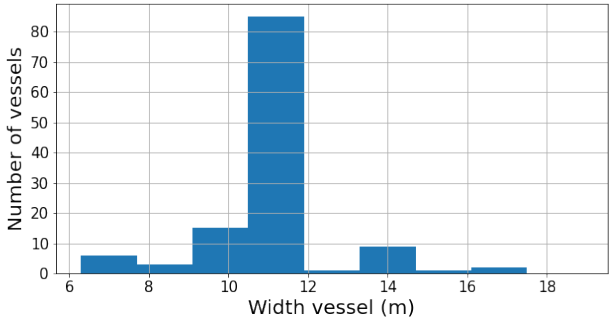

Figure 9: Histogram of the width (m) (distribution G)

- **Distribution H: length of the vessel**
  This distribution indicates the probabilities of the *length* of the vessel. This variable is also found in the data set. The variable *length* is an input variable for the model of Bolt in Equation 1. Table 18 and the histogram in Figure 10) show the distribution of the data for this variable. The bin size is equal to 12.75 meters.

Mean	**Std.**	**Min.**	**25%**	**50%**	**75%**	**Max.**
103.16	20.23	33.00	86.00	110.00	110.00	135.00

Table 18: Statistics of the length (m) (distribution H)

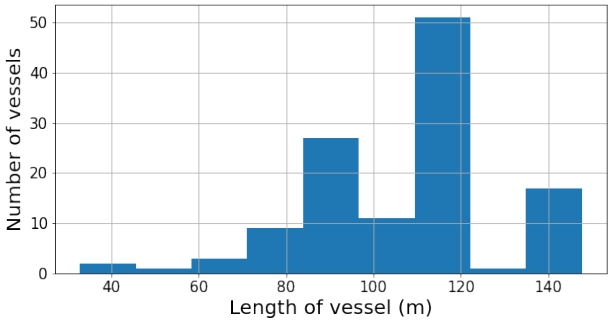

Figure 10: Histogram of the length (m) (distribution H)

- **Distribution I: maximum draught of the vessel**
  Distribution I indicates the probabilities of the *maximum draught* of the vessel. This variable can be found in the data set. It is necessary to pick the *maximum draught* of the vessel in order to thereafter multiply it by the *capacity rate* of the visit to determine the *draught* during a visit. The distribution, after the transformations mentioned in Chapter 2.2, is shown in Table 19 and Figure 10. The bin size is equal to 0.30 meters.

Mean	**Std.**	**Min.**	**25%**	**50%**	**75%**	**Max.**
3.31	0.49	2.50	2.92	3.34	3.61	4.65

Table 19: Statistics of the maximum draught (m) (distribution I)

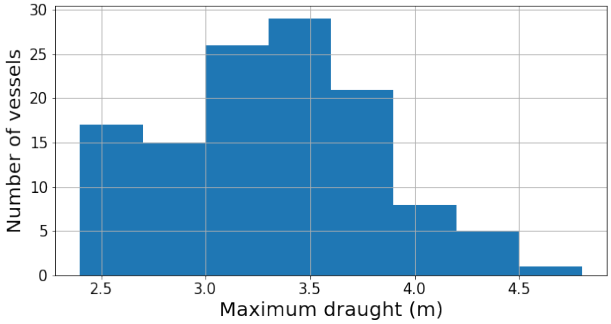

Figure 11: Histogram of the maximum draught (m) (distribution I)

- **Distribution J: average speed of the visit**
  In the simulations the value of the speed, V, is chosen as one value for the whole visit. This distribution indicates the probabilities of the average *speed* of the vessel during a visit. For every visit ID in the data set the average value of V is taken. Table 12 and Figure 20 show the distribution of the average speed for a visit. The bin size is equal to 0.20 m/s. The assumptions that were made for this variable are found in Chapter 2.3.

Mean	Std.	Min.	25%	50%	75%	Max.
2.76	0.94	0.12	2.15	2.70	3.38	5.54

Table 20: Statistics of the speed over water (m/s) (distribution J)

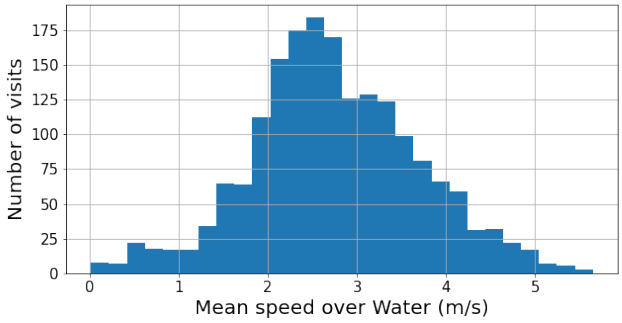

Figure 12: Histogram of the speed over water (m/s) (distribution J)

3.2.2 Emission factor

Recall that the amount of CO_2 an inland vessel emits depends on its energy source and the amount of fuel, or energy, it consumes. Through the simulations and the method in Chapter 3.2.3 the energy consumption for each visit and energy source for each vessel is obtained. To calculate the CO_2 emission the energy consumption is multiplied by the emission factor. The emission factor denotes the amount of emission per unit of energy source. The goal is that all the inland ships use electricity, hydrogen or other synthetic fuels, because their emission factor from tank to wheel [7] (TTW) is 0 (i.e. no emission). Although during the activities before these sources can be used as an energy resource (i.e. well to wheel [8] (WTT)) CO_2 is also emitted, this thesis will only focus on the TTW CO_2 emission. The current emission factor of diesel oil is 72 grams [9] of CO_2 per Mega-joule (MJ) (TTW). Recall from Chapter 1.3 that from January 2022 inland ships are obligated to use an admixture of diesel oil with at (at least) 7% bio fuel (e.g. bio diesel, bio ethanol, bio LNG) as energy source. This causes a decrease of the emission factor. It is unknown how much this decrease is. A. Lobo Gomes advised to put the emission factor to 0, since the TTW emission factor of bio fuels is very small (< 0.20 kilograms per unit of bio fuel [9]). As a result the emission factor of the admixture is calculated as the remaining diesel oil percentage multiplied by its emission factor. For example, if the fuel exists of 80% diesel oil and 20% bio fuel, the emission factor is $0.8 * 72 = 57.6$ grams of CO_2 per MJ.

The ambition mentioned in The National Climate Agreement is to let diesel oil used by inland ships consists of at least 30% bio fuel by 2030 [15]. No other literature was found on how this percentage will increase between 2022 and 2030 and between 2030 and 2050. The assumption in order to include the changing emission factor is that from 2020 until 2050 the increase of mixed bio fuel is linear with 2.875% per year. This leads to 87.5% bio fuel and 12.5% diesel oil in 2050. The percentages of bio fuel are found in Appendix A.1. Note that for both cases the assumption is that the emission factor of diesel oil itself remains the same until 2050. This is 72 grams CO_2 per Mega-joule (MJ).

[7] The emission factor from tank to wheel is the amount of emission that is omitted when an inland ship uses one unit of the energy source

[8] Well to wheel includes all the steps in the process from mining the energy source up to the inland ship

[9] This information originates from an Excel sheet provided by M. Dalhuisen

3.2.3 Energy source

The determination of the energy source of a vessel is solely based on assumptions, since no literature was found on this specific topic. Although one of the goals in the Green Deal [12] is that by 2030 at least 150 inland ships use an energy source that does not emit any CO_2, it is still up to the vessel owners to comply. Although the PoRA can, to some extend, encourage vessel owners to operate more efficient, improve sustainability and reduce the emitted CO_2, they do not have the power to impose such changes. In this thesis 2 classes of energy source are distinguished: 'diesel oil' and 'zero emission'. The class 'zero emission' includes all energy sources that have an emission factor of 0. M. Dalhuisen formulated the following assumption: from 2025, the vessels with *group* 1 or 2 that need a new engine will transfer from 'diesel oil' to 'zero emission'. The use of electricity or hydrogen is mainly suitable for short journeys. The energy source of vessels where the *group* equals 3 or 4, will always be 'diesel oil'. The assumption is also that a vessel needs a new engine every 20 years. Currently the number of vessels that have an engine that uses 100% bio fuel, electricity, hydrogen or other synthetic fuels is negligible. Therefore, all the inland ships, regardless of the *group*, are classified as 'diesel oil' until 2025.

First the number of years between the current year and the year of construction is calculated, called c_i. The remainder after dividing this number by 20 years is seen as the age of the engine (see Equation 3). For example, if the year of construction is 1996 and the current year is 2020, the age of the engine is 4 years, because 24 divided by 20 is 1 with a remainder of 4. This means that if the calculated age of the engine is 0 the assumption is that a new engine is purchased.

Since there are not a lot of vessels that are assigned to group 1, the probability for this group is very small (see Table 11). Luckily the probability for group 2 is larger. In case both probabilities were small the number of vessels that transfer to 'zero emission' would also be small, which would make the decrease in CO_2 emission over the years completely dependent on the changing emission factor.

$$\text{age of engine} = (c_i - \text{year of construction}) \ \% \ 20 \text{ with } i = 2020, ..., 2050 \tag{3}$$

Note that for every year all the vessels are randomly chosen, which means that the probability that the same vessels visit the PoR more than one year, is not taken into account. In reality a percentage of all the vessels is likely to also visit the PoR the following year(s). Since there are only 6 months of data available for this research, the recurrence rate per year for a vessel could not be calculated. The consequence is that in the simulations the number of vessels that transition to 'zero emission' is reset every year and thus is random. To include the possibility that the number of vessels that do not emit CO_2 can increase additional rules are added to the algorithm.

For y in 2025, ..., 2050 the following variables are introduced:

- N_z is the maximum number of vessels that are classified as 'zero emission' that were observed in a year. $N_z \geq 0$.
- $n_z(y)$ is the number of vessels that are classified as 'zero emission' in year y.
- $n_g(y)$ is the number of vessels where the *group* is 1 or 2 in year y.

Algorithm 2 shows the steps in pseudo-code. The following example clarifies the algorithm. At the start of 2025 $N_z = 0$.

- If $n_z(2025) = 5$ then $N_z < n_z(2025)$ and N_z is updated to $N_z = n_z(2025) = 5$ (see (1)).
- If thereafter $n_z(2026) = 2$ and $n_g(2026) = 4$, then $n_z(2026) = n_g(y)$ (see (2)). In other words the observed maximum number of vessels that were classified as 'zero emission' in a year up to 2026 was 7. In 2026, for 4 vessels was the *group* equal to 1 or 2. From this group, 2 vessels are classified as 'zero emission'. Because $n_g(2026) < N_z$, all the vessels where the *group* is 1 or 2 (i.e. 4 vessels) are classified as 'zero emission' regardless of the age of their engine. This means that $n_z(2026) = n_g(2026)$. N_z remains 7.

Algorithm 2 Determining the number of ' zero emission' vessels per year

$N_z = 0$
for $year = 2025, \ldots, 2050$ **do**
    if $N_z < n_z(y)$ **then**
        Update $N_z = n_z(y)$ (1)
    else $N_z > n_z(y)$
        if $n_g(y) < N_z$ **then**
            $n_z(y) = n_g(y)$ (2)
        else
            Random choose N_z vessels from the group $n_g(y)$ (3)
            Classify these vessels as 'zero emission' (4)
            Update $n_z(y)$ (5)
            if $n_z(y) > N_z$ **then**
                update $N_z = n_z(y)$ (6)
            end if
        end if
    end if
end for

- If in the following year $n_z(2027) = 4$ and $n_g(2027) = 10$, then the N_z (=7) vessels that are randomly chosen from the group $n_g(2027)$ (see (3)) are classified as 'zero emission' (see (4)). Due to the random choice, the possibility is that the vessels that were already classified as 'zero emission' before this step (as part of $n_z(2027)$) are not (all) part of the group of vessels that were picked at (3). Therefore in general holds that $N_z \leq n_z(y)(\text{new}) \leq n_z(y)(\text{old}) + N_z$. In this example: $7 \leq n_z(y)(\text{new}) \leq 10$, because the maximum value of $n_z(y)(\text{new})$ is restricted by $n_g(2027)$. If $n_z(y)(\text{new}) > N_z$, then N_z is updated to $n_z(y)(\text{new})$ (see (6)).

4 Sensitivity analysis

During the data collection and analysis it became clear that not all input variables for Bolt's model are of sufficient quality or are unusable. This is the case for the input variables *speed over water*, *water depth*, *water way width* and *average draught*. Prior to investing of additional time, collection of more data and effort to overcome technical issues, a sensitivity analysis (SA) could indicate if this time and effort would be well spend. SA assesses the sensitivity of the outcome of Bolt's model to the range of variation of each input. It can assess the impact of range of variation in the input values on the output [13]. The SA includes the calculation of the energy consumption for every value in the range of possible values of one input while the values of all other inputs remain unchanged. This is also referred to as nominal range SA. This chapter describes the nominal range SA application for each of the mentioned variables and the results.

Every SA uses the data of the observations from April until July 2020 of vessels for which the *segment* is equal to 'containers'. Recall the chosen initial values for the variables *average draught*, T_{gem}, *speed over water*, V, *waterway depth*, h, and the *water way width*, w:

- $T_{gem} = T$
- $V = V_a$
- $h = 15$ meters
- $w = 100$ meters

4.1 Average draught

Chapter 3.1.1 stated that it is common for inland vessels to have a different draught at the bow and stern. Van Dorsser et al. state in [6] that it is common for loaded inland ships to apply a trim (i.e. the bow or the stern is deeper submerged into the water) of about 2 to 5 centimeters over the length of the vessel. Also, the stern (i.e. back) is typically one meter lower than at the front in case the vessel is empty. Thus the vessel can lie horizontal in the water and if it is not, there the draught between the bow and the stern can differ with 1 meter at most. In this SA the *average draught*, T_{gem}, is calculated by taking the average of the draught at the bow and the stern of the vessel. Recall that T in the data set is the draught measured at the lowest point of the vessel, T_L. The draught at part of the vessel that is smaller than T_L is called T_H. Since T_H is equal or smaller than T_L, T_H can be defined as T_L minus a value l. Here l denotes the difference in draught between the bow and the stern of the vessel. The possible values of l are between 0 and 1.00 meter. If l is 0, then $T_L = T_H$, which indicates the vessel lies horizontal in the water. T_{gem} is defined as the average value of T_L and T_H (Equation 4). The nominal range SA calculates the energy consumption for every value of l. The expectation is that if l increase the output of the model increases, since the vessel will suffer from more water resistance as a consequence.

$$T_{gem} = \frac{T_L + T_H}{2} = \frac{T_L + (T_L - l)}{2} \text{ with } l = 0, ..., 1 \text{ meter} \tag{4}$$

l **(m)**	**Energy consumption (GJ)**	**Relative difference**
0.0	13117.61	1.000
0.2	13152.23	1.003
0.4	13186.85	1.005
0.6	13221.46	1.008
0.8	13256.08	1.011
1.0	13290.69	1.013

Table 21: SA of variable average draught

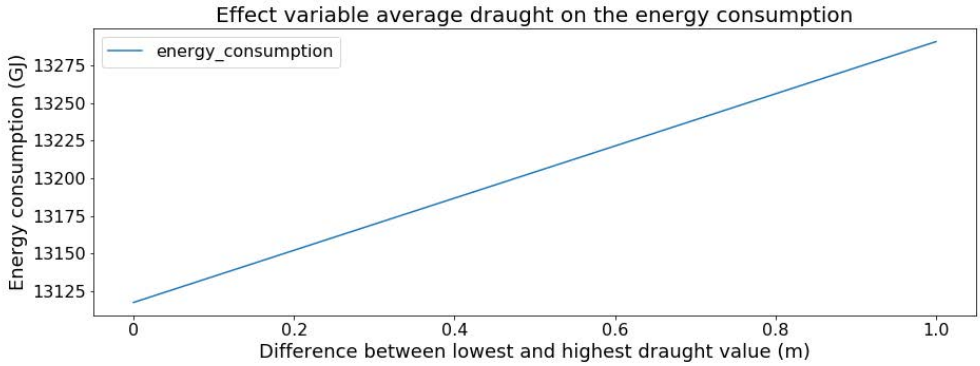

Figure 13: Graphical view of the SA results for the variable average draught

Figure 13 shows that the expectation was correct: if l increases, the calculated energy consumption, E, increases. Table 21 shows that if the difference in draught is 1.00 meter the energy consumption is 1.3% higher compared to no difference in draught. Note that the calculated values for E are in case all the vessels have that same value for l for every observation. For example, the calculated E with $l = 1.00$ meter is the energy consumption if all the vessels were empty at every observation. Obviously this is never the case in reality. Therefore the initial assumption that $T_{gem} = T$ will be remain unchanged.

4.2 Speed over water

Recall from Chapter 2.5 that the speed over water, V, is defined as the speed over ground, V_a, plus the effect of the tidal stream on the vessel's speed, e, as shown in Equation 5. The data for V_a are found in the data set. Therefore only e has to be calculated before the SA can be performed.

$$V = V_a + e \tag{5}$$

The PoRA keeps track of the tidal stream direction and rate for several measuring points throughout the PoR every 10 minutes. This generates a great amount of data every day. Due to the method that the PoRA chose to save these data, it is not possible to extract the value for a specific location on a specific date. Therefore, this information can only be used for analysis if all the data for a predefined time period is extracted. Due to computer memory limitations, data points of a maximum time period of two weeks can be extracted for the whole harbor. As a consequence, it is not possible to use these data in this thesis.

Since it is not possible to include information of the tidal stream for every obervation in the data set, the tidal stream data of a location near the Suurhoffbrug is used. According to Wendy Janssen, employee at the PoRA, this location is known for its strong tidal stream. She provided data about the tidal stream rate and direction at the Suurhoffbrug for every 10 minutes of 2018. As this location is known for its strong tidal stream, the assumption is that these data denote the maximum tidal stream values that can be found in the PoR. Figure 14 and Figure 15 show the data. For example, on March 1st at 14.30 hours, the tidal stream direction was 276 degrees and the corresponding rate was 1.28 m/s. The blank spaces show that there are missing values. Figure 15 shows that the tidal stream rate is periodic and the average rate decreases over time in this year. These data and the data set of this thesis are not from the same year. Unfortunately, this was the most recent data that could be delivered. In the future, when it is possible for the PoRA to import the tidal stream data of 2020, this can improve the results of this section. The times of the observations in the main data set are rounded to the nearest 10 minutes in order to merge it with the tidal stream data set.

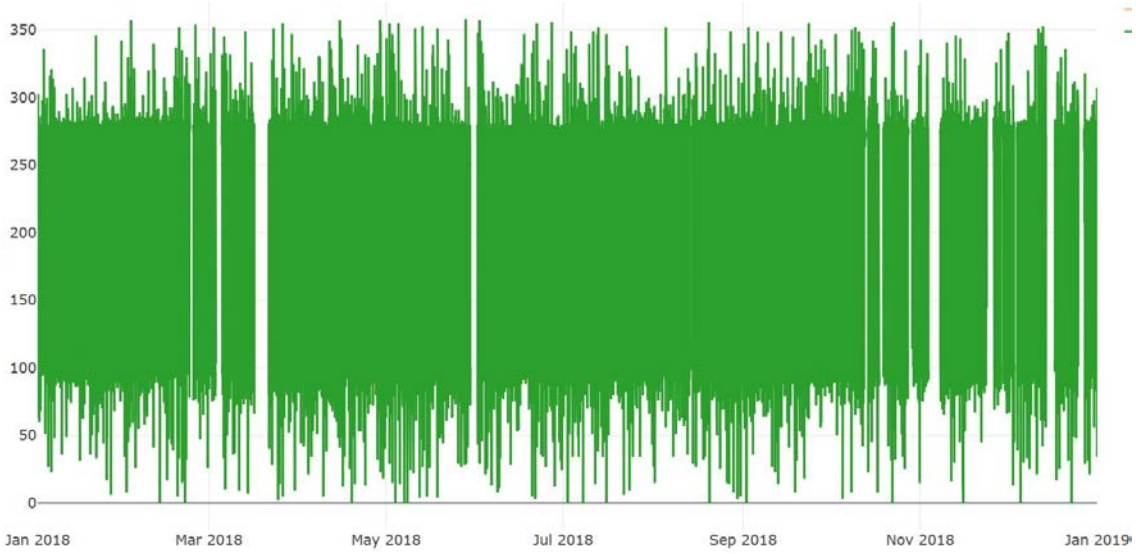

Figure 14: Tidal stream direction in degrees at Suurhoffbrug of 2018

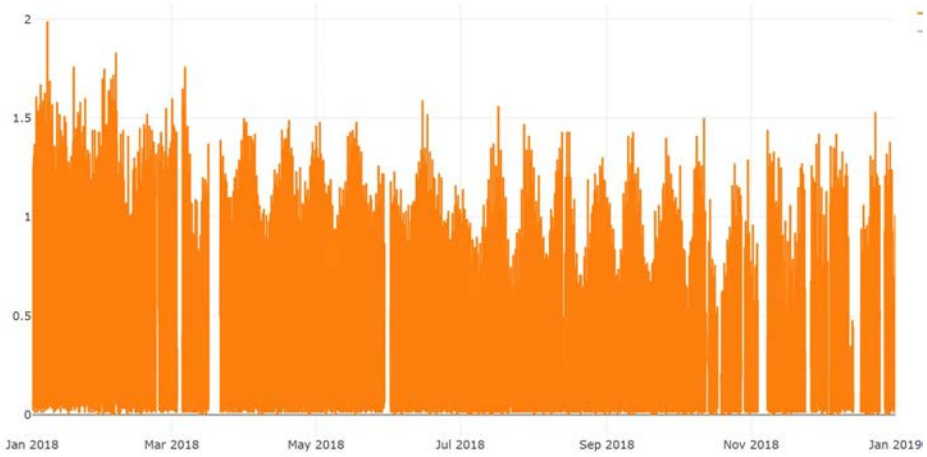

Figure 15: Tidal stream rate in m/s at Suurhoffbrug of 2018

No scientific literature was found on how to calculate e. In this thesis it was chosen to let e depend on the tidal stream rate, r, the tidal stream direction, d, and the heading of the vessel, h. The relationships between these variables are as follows:

- if the vessel sails in the same direction of the tidal stream, h is equal to d: the vessel sails downstream and fully benefits from the tidal stream rate. In this case $V = V_a + e = V_a + r$.
- if the difference between h and d is 180°: the vessel sails upstream and fully suffers from the tidal stream rate. Then $V = V_a + e = V_a - r$.
- if the difference between h and d is 90°: the effect of the tidal stream is canceled out, $e = 0$, and therefore $V = Va$.
- in all other cases $-r \leq e \leq r$, where the difference between h and d determines the value of e. This leads to Equation 6.

$$V = V_a + e = V_a + rcos(h - d) \tag{6}$$

The effect of the tidal stream on the vessel, e, is calculated by multiplying r by the cosines of the difference between h and d (Equation 6). For example, if $V_a = 3.0$ m/s, $r = 1.0$ m/s, $h = 105°$ and $d_1 = 45°$ (see Figure 16), then $V = 3.0 + 1 * cos(105 - 45) = 3.0 + 0.5 = 3.5$ m/s. This means that the vessel benefits from the tidal stream. On the other hand, if the direction is $d_2 = 315°$, then the vessel suffers from the tidal stream, since V is equal to 2.13 m/s. This is lower than V_a. Figure 16 support this as one can see that if $h = 105°$ and d_2 the vessel sails (not fully) upstream and therefore logically suffers (partially) from the tidal stream rate r.

The minimum values for the tidal stream are set to 0 m/s, which denote no tidal stream. In that case, $V = V_a$, which is called V_{min}. Recall that the assumption is that the provided tidal stream data denote the maximum values for the tidal stream that can be found within the PoR. Therefore, if these data is used V_{max} is calculated. Therefore, $V_{min} \leq V \leq V_{max}$. Since V is observation-specific, $V = V_{min} + Factor * (V_{max} - V_{min})$ is used, where $Factor$ is a value between 0 and 1. The nominal range SA calculates the energy consumption, E, while iterating over the possible values of $Factor$.

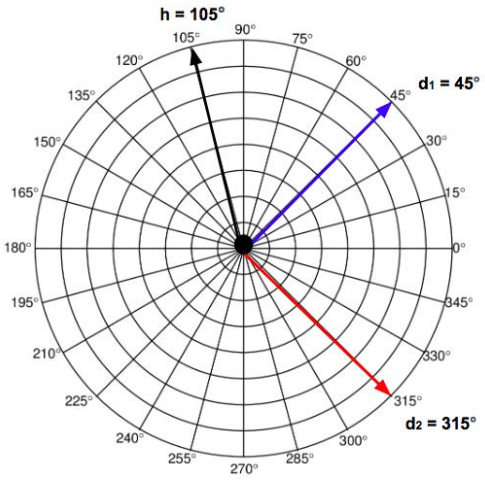

Figure 16: Example of the vessel's heading and the direction of the tidal stream

The expectation is that the effect of the tidal stream only becomes visible the result of an individual visit is observed. The vessel has to benefit or suffer from the tidal stream and as a consequence needs respectively less or more energy to maintain its speed. In this thesis the total energy consumption of a longer period is of interest. Since a vessel goes back and forth between locations it could be that the effect of the tidal stream cancels out over time. The average of V_a and V_{max} in the data are respectively 3.22 m/s and 3.20 m/s, which can support the expectation.

Factor	Energy consumption (GJ)	Relative difference
0.00	13112.17	1.000
0.20	13085.97	0.998
0.40	13065.45	0.996
0.60	13050.90	0.995
0.80	13041.45	0.995
1.00	13037.81	0.994

Table 22: Results of the SA of the variable speed over water

Figure 17 shows that if $Factor$ increases, E decreases. This means that for the observations in the data set, the energy consumption decreases if the tidal stream increases. However the change is

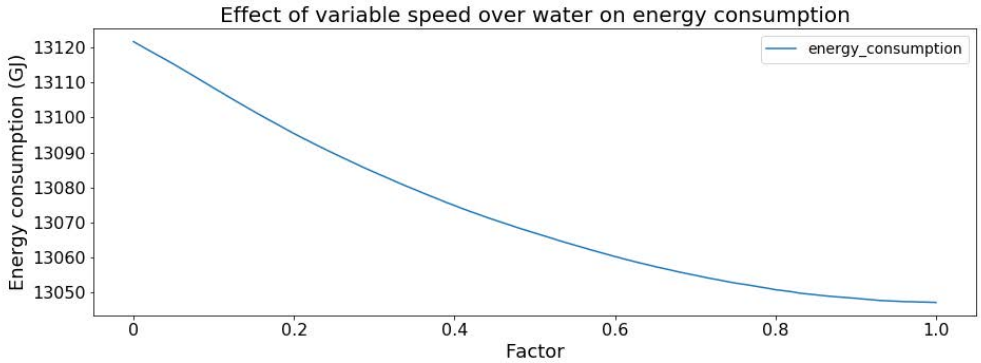

Figure 17: Graphical view of the SA results of the variable speed over water

very small. Table 22 shows that the total decrease of E, when the factor is varied from 0 till 1, is only 0.6% with respect to Factor = 0. This change can be ignored. From this result it can be concluded that, if the mentioned assumptions are correct, E is not sensitive to changes in V if the tidal stream is included. Therefore, the chosen initial value of V that equals V_a remains the same. This conclusion confirms the earlier mentioned expectation.

4.3 Water depth

The initial *water depth*, h, is chosen as 15 meters (see Chapter 3.1.1). The PoRA has the data of the *water depth* throughout the PoR, but these were again unusable due to computer memory limitations. The data from two locations, which are assumed to show extreme values, are used to conduct an SA for h. These data were also provided by W. Janssen. The data contain the water levels at a location near Hook van Holland (Figure 18), w_H, next to the North sea, and near Boerengat (Figure 19), w_B, in the centre of Rotterdam, for every 10 minutes in 2019. Both figures were also provided by W. Janssen. The year from which this data set originates is again not the same as the year of the data set used in this thesis. Unfortunately, this was the most recent data that could been delivered. The water near Hook van Holland is deeper than the water near Boerengat (Table 23), since the latter one is only used by inland vessels and not sea ships. To find the water depth, the water level is added to the depth that is guaranteed by the PoRA. This is shown in Equation 7). For the location near Hook van Holland, g_H, this is 16,20 meter below NAP. For the location near Boerengat, g_B, this is 5,45 meters below NAP. Therefore these are seen as respectively the upper bound, h_{max}, and lower bound, h_{min}.

Statistic	w_B (m)	w_H (m)
mean	5.72	16.34
std	0.62	0.67
min.	4.31	14.66
max.	7.83	18.58

Table 23: Statistics of water levels near Boerengat and Hook van Holland

$$h_{max} = g_H + w_H = 16.2 + w_H \tag{7a}$$

$$h_{min} = g_B + w_B = 5.45 + w_B \tag{7b}$$

The nominal range SA method for this variable is quite similar but less complicated compared to the SA of the previous variable. The values of w_H and w_B are date and time dependent. For each

observation, h is varied from h_{min} to h_{max} at the corresponding day and time. For example, the calculated h_{min} and h_{max} of April 5$^{\text{th}}$ 2018 are used for the observations on April 5$^{\text{th}}$ 2020. In the SA, the energy consumption, E, is calculated for every possible value. Again, a factor is used: $h = h_{min} + factor(h_{max} - h_{min})$, where the factor is varied from 0 to 1. Note that the same *water depth* is used for all observations. In other words, the same value for h is used for every vessel that was located anywhere in the PoR at the corresponding time and day. Therefore, the bare minimum and maximum value of E with respect to h is assumed to be calculated given the values for the other variables.

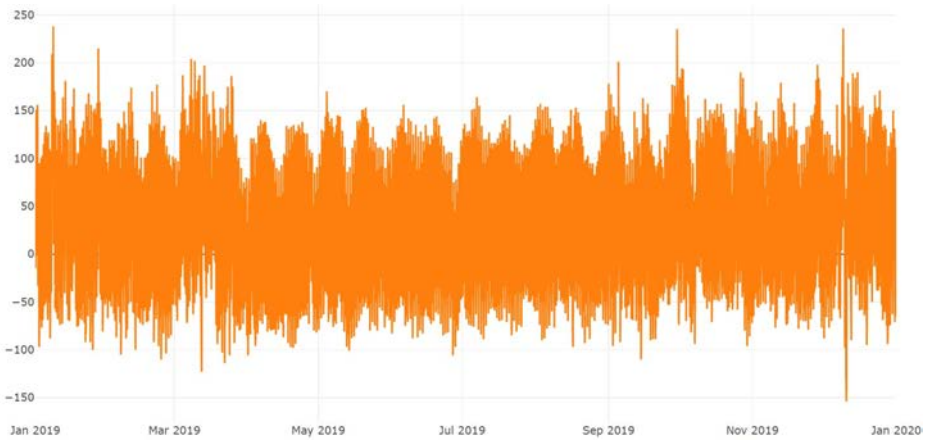

Figure 18: Water levels in cm in 2019 near Hoek van Holland

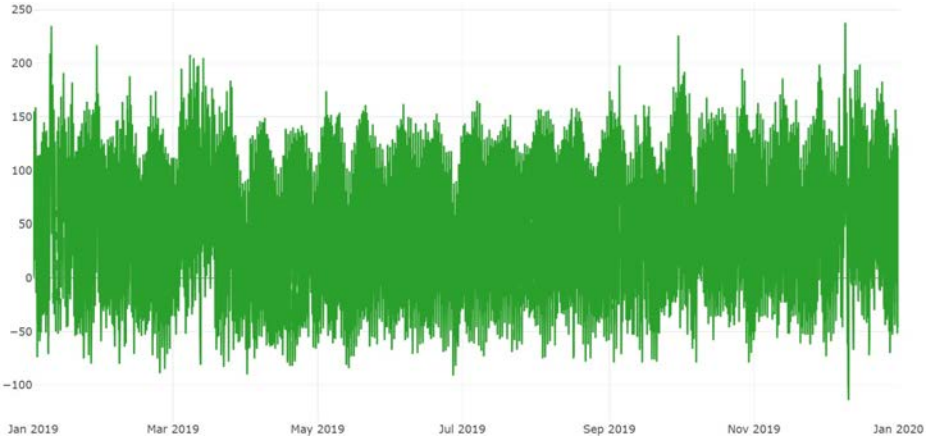

Figure 19: Water levels in cm in 2019 near Boerengat

Intuitively one would think that the water depth has significant impact on the energy consumption: a vessel experiences more water resistance in shallow water, which would make the energy consumption increase. However, the water ways in harbors are generally not shallow. Since the water ways in the PoR are also relatively wide, the water way cross section, A_c, will be large, which makes the numerator go to 0 and u very small (see Equation 2b). Since in the PoR $h >> T$, V_{eff} will approximately be equal to V (see Equation 2c). As a consequence u_{rel} (Equation 2) will approximately be equal to V either way. Therefore, variable h, but also A_c, are presumably not significant in this thesis.

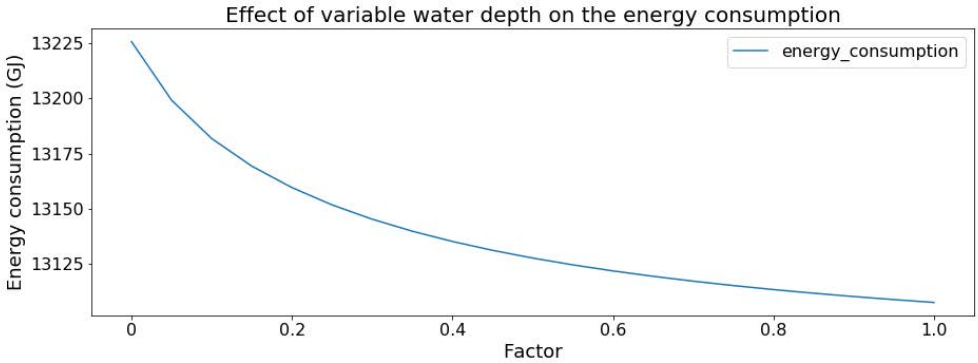

Figure 20: Graphical view of the SA results of the variable water depth

Factor	Energy consumption (GJ	Relative difference
0.00	13605.91	1.075
0.20	13076.21	1.033
0.40	12880.72	1.017
0.60	12773.88	1.009
0.80	12706.21	1.004
1.00	12659.42	1.000

Table 24: Results of the SA of variable water depth

Figure 20 shows that if h increases, E decreases. Table 24 shows that when factor=0, E is 7.5% higher with respect to Factor=1. In other words, if $h = h_{min}$ for all the observations in the data set, the total E would be 7.5% higher compared to if $h = h_{max}$. This means that h near Boerengat is low enough to have an impact on u_{rel}. Nevertheless, h_{min} and h_{min} represent the extreme possible values and in reality the inland vessels will not cross water ways in the PoR that are strictly equal to one of these extremes, but somewhere in between. Therefore h will be chosen as $h_{min} + 0.5(h_{max} - h_{min})$ in this thesis.

4.4 Water way width

The initial value for the water way *width* is 100 meters. There was no data available of w. Therefore, an SA can indicate how to choose its value. The variable w is solely needed for the calculation of Ac. The nominal range SA for the variable water way *width* is not difficult. With the use of [8] the possible values of w are 100 till 400 meters. Note that again the same water *width* is used for all the observations in data set of every vessel.

Width (m)	Energy consumption (GJ)	Relative difference
100	13432.53	1.000
150	13220.40	0.984
200	13117.08	0.977
250	13055.93	0.972
300	13015.52	0.969
350	12986.82	0.967
400	12965.39	0.965

Table 25: Results of the SA of the variable water width

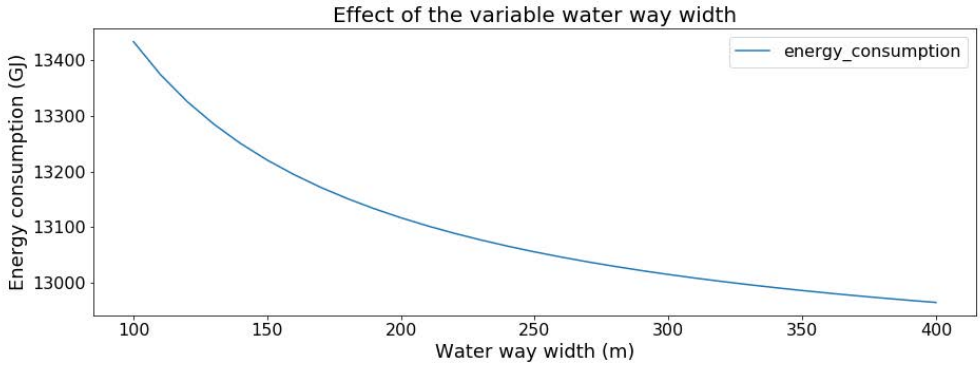

Figure 21: Graphical view of the SA results of the variable water way width

Figure 21 shows that if w increases, the energy consumption, E, decreases. If w increases, the impact of the water resistance on the vessel decrease, which causes E also to decrease. Nevertheless, the difference in E between the minimum and maximum value applied for w, is only 3.2% as shown in Table 25. This is the case if h of all observations would change from 100 to 400 meters. The conclusion from this result is that E is not sensitive to changes in w. The initial value of 100 meters appears to be a little low, since the majority of the water ways in PoR are wider than this. Therefore w will be chosen as 250 meters for all observations.

4.5 Summary

To summarize the results of all the sensitive analyses in this chapter:

- $T_{gem} = T$
- $V = V_a$
- $h = h_{min} + 0.5(h_{max} - h_{min})$, where h_{min} and h_{max} are respectively the water depth near Boerengat and Hook van Holland on the same date and time as the observation.
- $w = 250$ meters

With these assumptions values for all the input variables are present. Now, the method that includes the model of Bolt (Chapter 3.1) can be carried out to estimate the energy consumption.

5 Models

Intuitively, several variables that appear in the simulations correlate with each other. If there exist correlation this should be taken into account, because otherwise the results will not be realistic. Therefore a correlation test is performed. The results of this test are processed in model 1. In model 2 the assumption is that all the variables are independent. This chapter describes the correlation test results and explain the models that will be used in the algorithm of the simulations.

5.1 Model 1

Model 1 takes dependence between (some of) the variables into account. To investigate the correlations between the variables the Spearman Rank correlation coefficients are calculated. Strong correlations (< -0.5 or > 0.5) are observed between several variables. This model incorporates these correlations via multivariate distributions.

5.1.1 Correlations

The correlation matrix in Appendix A.2 shows the Spearman Rank correlation coefficients of all combinations. From this table it can be concluded that the variables *length*, *width*, *maximum capacity* and *maximum draught* have a high positive correlation. The variables *speed* and *distance traveled* are also positively correlated. The variables *draught* and *year of construction* do not show high correlation with one of the other variables. Intuitively, these correlation coefficients make sense. Logically, the width, length, maximum capacity and maximum draught of the vessel are dependent. Surely, a vessel with a width of 2 meters cannot have length of 100 meters. The width and the length determine the amount of tonnes a vessel can transport and when the maximum capacity is reached, the maximum draught is also at its maximum. One may expect that the year of construction can also correlate with these variables. Although the correlation matrix shows that it is not, this number could be the result of the large number of missing values for the variable *year of construction* that had to be imputed. If more data would be available, these correlations maybe be different. The same reasoning holds for the variable *draught*. Table 26 shows the highest correlation coefficients. The variables *speed* and *distance traveled* only show a high correlation with each other and not the other variables. Therefore the variables are split up in two groups: the vessel-specific variables (*length*, *width*, *maximum capacity*, *maximum draught*) and the visit-specific variables (*speed*, *distance traveled*).

	Length	**Width**	**Max. capacity**	**Max. draught**	**Speed**	**Distance trav.**
Length	1.00	0.82	0.91	0.69	0.14	0.11
Width	0.82	1.00	0.89	0.75	0.12	0.07
Max. capacity	0.91	0.89	1.00	0.83	0.19	0.14
Max. draught	0.69	0.75	0.84	1.00	0.19	0.15
Speed	0.14	0.12	0.19	0.19	1.00	0.72
Distance trav.	0.11	0.07	0.14	0.15	0.72	1.00

Table 26: Correlation matrix

This model takes these dependencies into account. The value for the remaining variables *group* (distribution A), *number or visits* (distribution B), *capacity rate* (distribution C) and the *year of construction* (distribution D) will be drawn out of the distributions as shown in Chapter 3.2.1. This means that in the simulations this will be draws from four different univariate distributions.

5.1.2 Multivariate distributions

The model integrates the dependencies of the previous section by combining the univariate distribution to a multivariate distribution. This means that with one draw out of a multivariate

distribution a value for all the incorporated variables is obtained. The requirement is that these values have the same covariance coefficients as found in the data set. An explicit formula was found for the multivariate normal distribution. Thus if the input variables follow a normal distribution it is easy to simulate multivariate data. Unfortunately, the values of not all the variables mentioned in the previous section are normally distributed, as was seen from the histograms in Chapter 3.2.1. Cario and Nelson described in [17] a method that is the solution to this problem. They present a model to simulate an k x 1 random vector $\mathrm{X} = (X_1, X_2, ..., X_k)'$ with arbitrary marginal distributions and covariance matrix. This means that vector X is the desired result that consists of a value for all the variables while the covariance coefficients are maintained. One starts with a standard multivariate normal random vector Z with a known covariance matrix and transforms it to achieve the desired marginal distributions for the components of the input variables. This is vector X. That is the reason why X is referred to as a NORTA (NORmal to Anything) distribution.

If this method is applied to the group of vessel-specific variables it is as follows: for one draw, it starts with Z. In this case $\mathrm{Z} = (Z_1, Z_2, Z_3, Z_4)'$, an 4 x 1 vector, since there are 4 variables: *length*, *width*, *maximum capacity* and *maximum draught*. The initial values of Z are random draws from a multivariate normal distribution given the covariance coefficients in Table 26 of these four variables. This means that the covariance coefficients of the components of Z are equal to those of the data (Table 26). X_{sim} will be a transformation of Z and thus denote the simulated values. Equation 8 shows vector X_{sim} where Φ is the univariate standard normal cumulative distribution function (cdf) and $F_X(u) \equiv \inf\{x : F_X(x) \geq u\}$ is the inverse cdf. Φ is applied to a realization of Z. Also, the cdf of the corresponding distributions in Chapter 3.2.1 are estimated. Since these are histograms the values of the bins are used. Thereafter, every value of Φ(Z) is compared with the bin edges of the cdf of the corresponding variable. The smallest bin that is still larger than Φ(Z) is chosen for the value of X_{sim}. Since a bin consists of a range of values the average value of the bin is chosen for the value of the variable. These steps are performed for every value, thus every variable, in Z.

$$\mathrm{X}_{\mathrm{sim}} = \begin{pmatrix} F_{X_1}^{-1}[\Phi(Z_1)] \\ F_{X_2}^{-1}[\Phi(Z_2)] \\ F_{X_3}^{-1}[\Phi(Z_3)] \\ F_{X_4}^{-1}[\Phi(Z_4)] \end{pmatrix} \tag{8}$$

For the visit-specific variables, *speed* and *distance traveled*, the same steps are taken to construct a multivariate distribution from which can be drawn. Recall that for the simulations the average speed and the total distance traveled per visit is used (see Chapter 3.2.1).

It came to notice that the algorithm is sensitive to large variations in scale of the input variables. Standardization of the variables works well to overcome this problem. The values of the *maximum capacity* are much larger than the values of the *length*, *width* and *maximum draught*. Similarly, the values of the *distance traveled* are much larger than the values of the *speed*. If the simple method in Equation 9 was applied to standardize the values of vector X, the covariance matrix of $\mathrm{X}_{\mathrm{sim}}$ was more similar to the covariance matrix of X than if no standardization was applied. Equation 10 was used to scale the resulting standardized value of the draw from the multivariate distribution ($x_{from.draw}$) back to obtain the value for a variable on its original scale (x_{new}).

$$x_{standardized} = \frac{x_{original} - \mu}{\sigma} \tag{9}$$

$$x_{new} = (x_{from.draw} * \sigma) + \mu \tag{10}$$

5.2 Model 2

The assumption in model 2 is that all variables are independent. This is the simplest model. The distributions can be used as described in Chapter 3.2.1. In the simulations will be drawn from these distributions. Each draw will give a bin as result. The variable value will then be obtained by

random choice between the minimum and maximum value of the chosen bin. In other words, the values within a bin are assumed to be uniformly distributed. For example, if the draw from the distribution gives bin 2 as the result and bin 2 denotes values between 3 and 6, the value for this variable is randomly chosen between 3 and 6. This method was chosen to avoid that a variable has a lot of the same values due to insufficient data. This is the method for all distribution except distribution A. For distribution A the draw is not only the bin, but also the *group*. Distribution A is still based on the numbers in Table 10 despite of low number of vessels within each group, because a *group* is required to determine the fuel type (Chapter 3.2.3)

Recall that the bin size for each distribution was chosen as the smallest bin size for which every bin has at least one observation. If more data would be available these bin sizes would probably be smaller, since there would be more observations in the range of values of a variable. In that case the average bin value could be chosen as the value for the variable. To investigate if the bin sizes affect the output the results of the simulations with different bin sizes are compared. Since the bin size is variable-specific, the bin size is increased by multiplying it with a fraction instead of adding an absolute value to it. The bin size will be increased by multiplicating it by a fraction 1.2, 1.4, 1.6, 1.8 and 2.0.

5.3 Model implementation

The difference in the algorithm of the simulations that uses model 1 or 2 is found in the 'pick vessel characteristics' and 'pick the visit characteristics' step (see Algorithm 1). For model 1 both of these steps consist of a draw out of one multivariate distributions while draws from 6 (univariate) distributions are necessary in case the algorithm uses model 2.

Recall that for every year all the vessels are randomly chosen. In reality a subset of all the vessels are likely to also visit the PoR the following year. Since there are only 6 months of data available for this research, the recurrence rate per year for a vessel could not be calculated.

6 Results

For all the observations in the data set the CO_2 emission is calculated with the method described in Chapter 3.1. The first section of this chapter reports on these results. Then-after in the second part, the results after application of the method in Chapter 3.2, the simulations, are described. The initial idea was to produce results per segment and group. Unfortunately there was not enough data to accomplish this. Therefore the results in this chapter are based on the complete data set without any type of distinction.

6.1 CO_2 emission

Recall that the energy consumption in megajoule (MJ) is calculated for every observation in the data set with Equation 1. The assumption is that all inland ships used 'diesel oil' without any bio fuel (see Chapter 3.2.3). The emission factor of diesel oil is 72 grams of CO_2 per Mega-joule (MJ) as mentioned in Chapter 3.2.2. Therefore the CO_2 emission is obtained by multiplying the energy consumption with the emission factor. All the values are aggregated per visit.

During the months April till September of 2020 122 vessels that operated in the segments containers, tankers, dry bulk and break bulk visited the harbor 1850 times and emitted $7.44 * 10^7$ kilograms, which is equal to 0.0744 megaton, of CO_2. Table 27 shows the total visits per month, the total CO_2 emission per month and the average and standard deviation of the CO_2 emission per visit for each month. There are 51 visit IDs that occurred in two months. The number of visits in April and May is much lower than in most of the others months. Although these months had the same number of visits, the CO_2 emission was higher in April. The lowest number of visits is observed in August. Perhaps this is a seasonal effect. This month is followed up by the month with the most number of visits, September. Despite of the month August having the lowest number of visits and total CO_2 emission, its average CO_2 emission per visit was not. Remarkable is that although the month September had the most visits and the highest average CO_2 emission per visit, the highest total CO_2 emission was observed in June and July. On the other hand, the month June showed a high number of visits and the highest total CO_2 emission although a higher average per visit was observed in the months July, August and September. These observations denote that the total CO_2 emission cannot be explained by solely the number of visits.

Month	**# visits**	**Total all visits (kg)**	**Avg. per visit (kg)**	**Std. per visit (kg)**
4	308	$1.31264 * 10^7$	$0.01893 * 10^7$	$0.05771 * 10^7$
5	309	$1.16866 * 10^7$	$0.01751 * 10^7$	$0.05222 * 10^7$
6	341	$1.36917 * 10^7$	$0.01953 * 10^7$	$0.05807 * 10^7$
7	332	$1.36345 * 10^7$	$0.02042 * 10^7$	$0.05763 * 10^7$
8	241	$9.43753 * 10^6$	$0.02026 * 10^7$	$0.05467 * 10^7$
9	370	$1.28653 * 10^7$	$0.02141 * 10^7$	$0.06324 * 10^7$

Table 27: Statistics of CO_2 emission per month

The boxplot in Figure 22 shows how the values for the CO_2 emission are distributed per month. Every month has outliers, which are shown as circles. These causes the average CO_2 emission to become larger. The values of September are spread the most. While its interquartile range is not the largest, the number of outliers and their values causes its average value to go up. That is the reason why this month showed to have the most visits and the largest average CO_2 emission per visit but not the highest total CO_2 emission. Unlike the average, the median is not prone to outliers. June has the largest median value (the green line) with few outliers, which causes the average CO_2 emission to be low compared to the other months but the total CO_2 emission high.

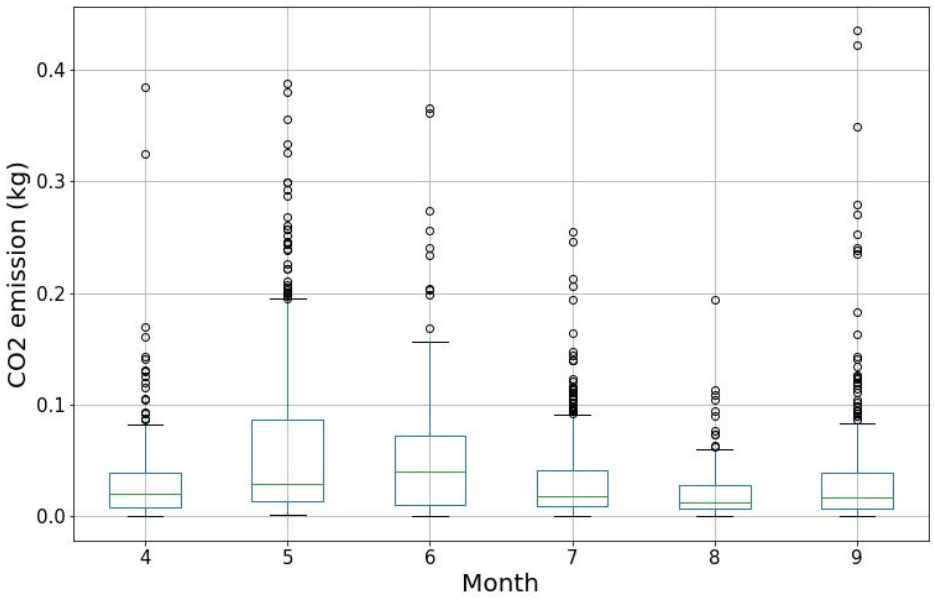

Figure 22: Boxplot of CO_2 emission * 10^6 per visit for each month. (Green line: median, blue lines: quartiles)

Distance traveled

The scatterplots in Figure 23 shows the distance traveled per visit against the CO_2 emission. Note that the x-axis and y-axis differ per month. The data points are the most spread for both axis in April. Although all months show several outliers between 15 and 50 kilometers, the outliers in April lie between 60 and 100 kilometers. There was far more distance traveled per visit in April than in the other months while the least distance was traveled per visit in May. This could explain why there was less CO_2 emission in April despite of the equal number of visits. It is remarkable that in the months April, May, June and September high CO_2 emission were observed for relatively a small number of kilometers. In these cases a wide spread of possible CO_2 emission is shown for a range of kilometers. For example, in the month September, the CO_2 emission of visits where approximately 10 kilometers was traveled differed between 40 * 10^3 and 320 * 10^3 kilograms. This means that the distance traveled on itself cannot explain the CO_2 a vessel emits. In all the plots run out, which could indicate that there are different slopes for different type of vessels.

Visit duration

Table 28 shows information about the visit duration for each month. The duration is the difference between the time of the first and last observation of a visit. The difference between the minimum and maximum visit duration of each month is very large. All the minimum values appear fairly small: approximately 6 to 19 minutes. It is unknown if these durations are possible in reality. The average values are bigger than the median values. This indicates again that there are outliers. The longest visit duration of nearly 154 hours is observed July while the lowest maximum value for the duration is approximately 102 hours in April. Recall that the month April showed visits where the most distance was traveled (Figure 23). This indicates that a lot of distance was traveled during relatively little time compared with the other months. Certainly, long visit durations could also be caused by other factors like waiting times. In this case, long duration would not necessarily indicate high CO_2 emission, because if the vessel is moored this does not emit much (or any) CO_2.

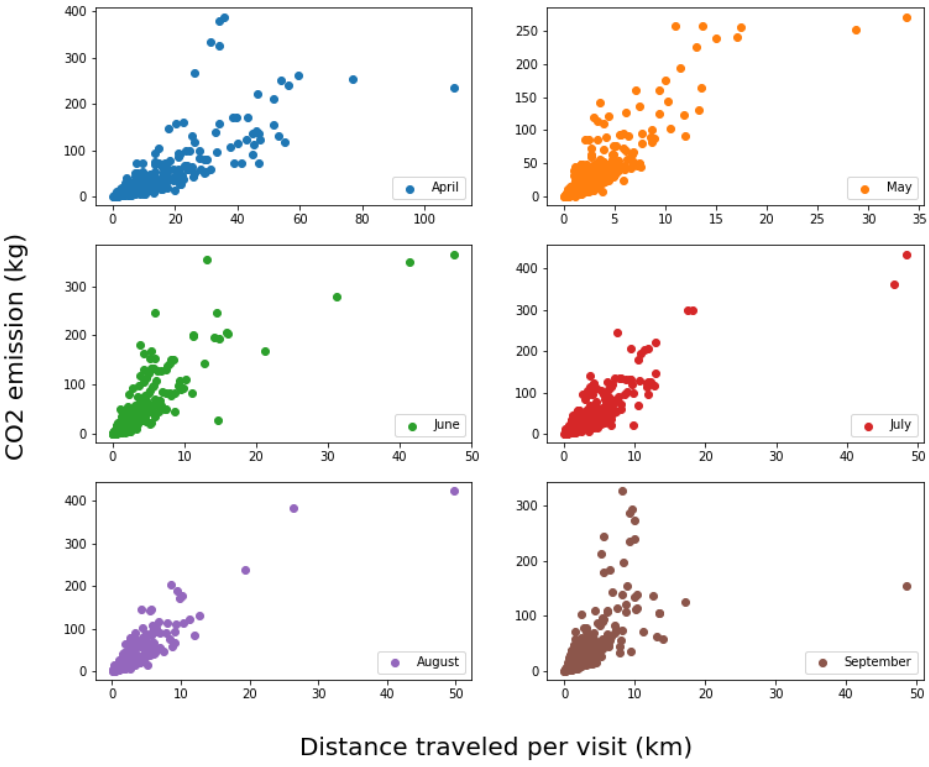

Figure 23: CO_2 emission x 10^3 kilograms per visit for each month

Month	Min. (hr)	Max. (hr)	Avg. (hr)	Median (hr)	Std. (hr)
4	0.10	102.11	25.56	21.87	19.59
5	0.31	119.16	25.97	21.01	21.30
6	0.14	138.54	25.38	19.35	22.62
7	0.13	153.84	27.24	20.50	22.28
8	0.12	113.38	23.61	18.21	19.58
9	0.19	135.05	24.98	19.97	21.80

Table 28: Statistics of visit duration in hours per month

6.2 Monte Carlo simulations

Recall from Chapter 5 that the simulations are run with two different models: model 1 includes dependency between several variables and the assumption in model 2 is that all the variables are independent. Furthermore, the effect of the decreasing emission factor (see Chapter 3.2.2), called model 1B and 2B, is investigated by comparing the results with the situation where the emission factor remains unchanged, called model 1A and 2A. Additionally, to investigate if the bin size of the distributions affects the outcomes, the results of the simulations with model B where the bin size is varied are also compared (see Chapter 5.2). The first and second section describe the results of the models separately, while the third section compares the results. Lastly, the fourth section

describes the results of the simulations when the bin size changes.

Unfortunately, due to memory limitations the real throughput forecast could not be used. An initial throughput of 500.000 tonnes is chosen for 2020. The increase in percentage that was found in the real forecast is applied. The exact numbers can be found in Appendix A.4. With these throughputs the run time was under 9 hours (maximum) and no memory problems were encountered. Due to this relatively low throughput it turned out that the implication of Algorithm 2 in Chapter 3.2.3 did not have an effect on the results. The number of vessels that was required to transport the given throughput was too low in order for this algorithm to have an impact on the results. Therefore, the results in this section do not include Algorithm 2. This means that the biggest impact on the results is caused by the changing emission factor and not the transition to 'zero emission'.

6.2.1 Model 1

- model 1A: constant emission factor
- model 1B: decreasing emission factor (Chapter 3.2.2)

Table 29, Figure 24 and Figure 25 show the statistics and the histograms of the simulation results of 2050 with model 1A and 1B. The algorithm is run twice to support the results of the first run. For both models the average value of both runs does not differ much. Table 29 shows that if the emission factor would remain the same the CO_2 emission in 2050 would be approximately 8 times higher than if the emission factor would decrease according to the assumptions in Chapter 3.2.2. Because the emission factor becomes lower over the years, the spread of the possible values in 2050 is also much more narrow for model 1B. This shows in the histograms of Figure 24 and Figure 25, but is also very clear in Figure 26 and 27 where the spread of all simulation values is shown. While the emission factor of model 1A remains 72 grams of CO_2 per MJ, the emission factor of 1B decreases to only $0.125 * 72 = 9.0$ grams of CO_2 per MJ. This is the reason why the results differ with a factor 8.

	Avg.	**Std.**	**Min.**	**Max**
Run 1				
model 1A	15343.79	1302.38	11222.60	20036.22
model 1B	1893.49	180.92	940.96	2504.53
Run 2				
model 1A	15379.55	1335.77	11220.54	19424.15
model 1B	1898.08	186.31	1162.66	2428.02

Table 29: Simulation results of CO_2 emission (kg) in 2050 with model 1A and 1B

A normality test, with the null hypothesis that the values of model 1A come from a normal distribution, gives a p-value of 0.1351. This means the null hypothesis cannot be rejected and the values could be normally distributed. The normality test for the results of the second run gives a p-value of 0.0652. Although this is much smaller than the p-value of the first run, the hypothesis still cannot be rejected and the same can be concluded. The same test is performed for both run with model 1B. The test gives a very small p-value of $1.3685 * 10^{79}$ and $3.9788 * 10^{79}$, which means that the null hypothesis can be rejected. The values are not normally distributed. The lack of normality may be due to the histogram being left-skewed.

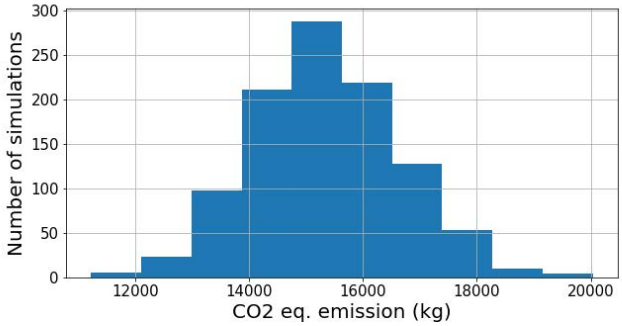

Figure 24: Histogram of CO_2 emission (kg) in 2050 with model 1A

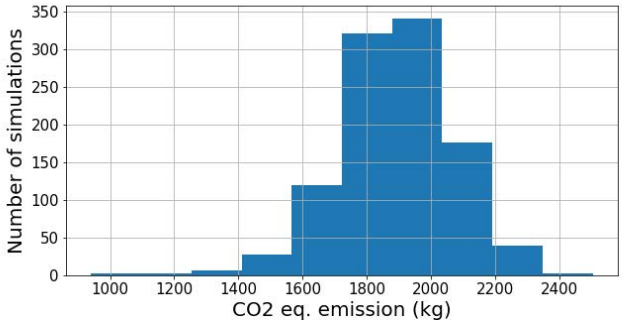

Figure 25: Histogram of CO_2 emission (kg) in 2050 with model 1B

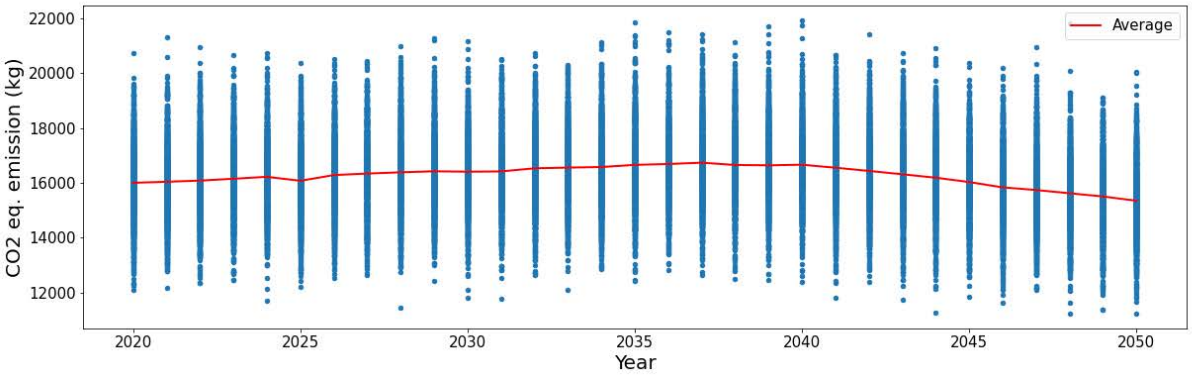

Figure 26: CO_2 emission (kg) of all simulations with model 1A

6.2.2 Model 2

- model 2A: constant emission factor
- model 2B: decreasing emission factor (Chapter 3.2.2)

Table 30, Figure 28 and Figure 29 show the statistics and the histograms of the simulation results of 2050 with model 2A and 2B. Again, for both models the average value of both runs does not

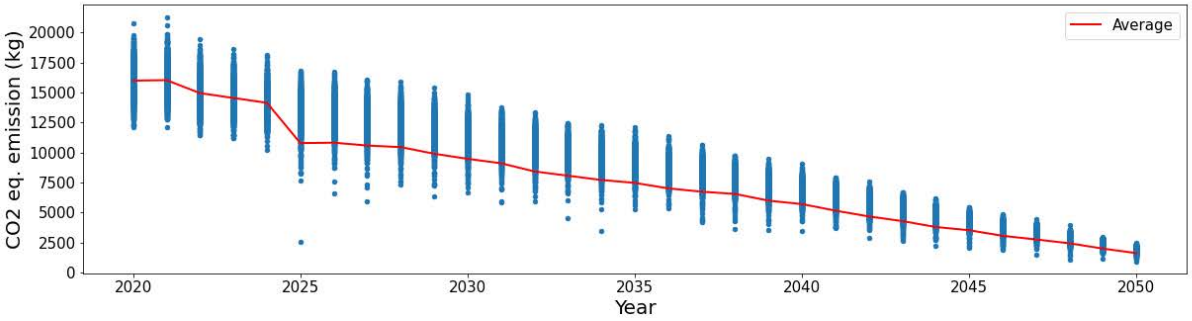

Figure 27: CO_2 emission (kg) of all simulations with model 1B

	Avg.	**Std.**	**Min.**	**Max**
Run 1				
model 2A	13684.04	2596.26	6854.59	25373.40
model 2B	1682.29	325.88	799.17	3171.68
Run 2				
model 2A	13589.92	2535.26	6341.54	23180.54
model 2B	1676.85	318.41	792.69	2744.22

Table 30: Simulation results of CO_2 emission (kg) in 2050 with model 2A and 2B

differ much. Since these models use the same emission factor as the previous models their averages differ again with approximately a factor 8. This time, both of the histograms in Figure 28 and 29 look like a normal distribution but have a large right tail. The resulting p-values 2.2335 * 10^{-5} and 0.0011 of the normality test for respectively the first and second run with model 2A, conclude that the values are not normally distributed (with significance level at 0.005). For model 2B the p-values are 1.4688 * $10^{-}34$ and 6.7150 * $10^{-}38$ from which also can be concluded that the values are not normally distributed. Figure 30 and Figure 31 show the same behaviour as the corresponding figures in the previous section: due to the decreasing emission factor the average CO_2 value and its standard deviation decreases over the years when a decreasing emission factor is used in the model.

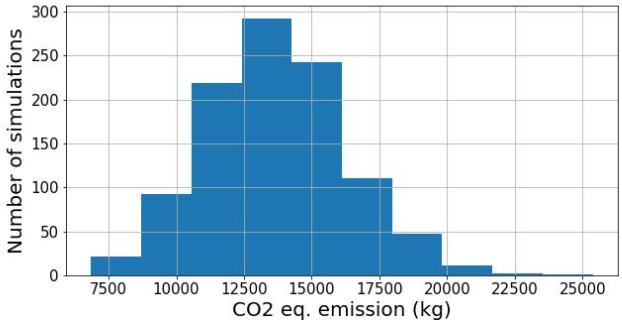

Figure 28: Histogram of CO_2 emission (kg) in 2050 with model 2A

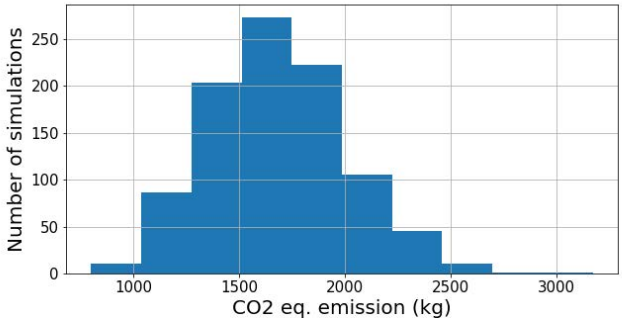

Figure 29: Histogram of CO_2 emission (kg) in 2050 with model 2B

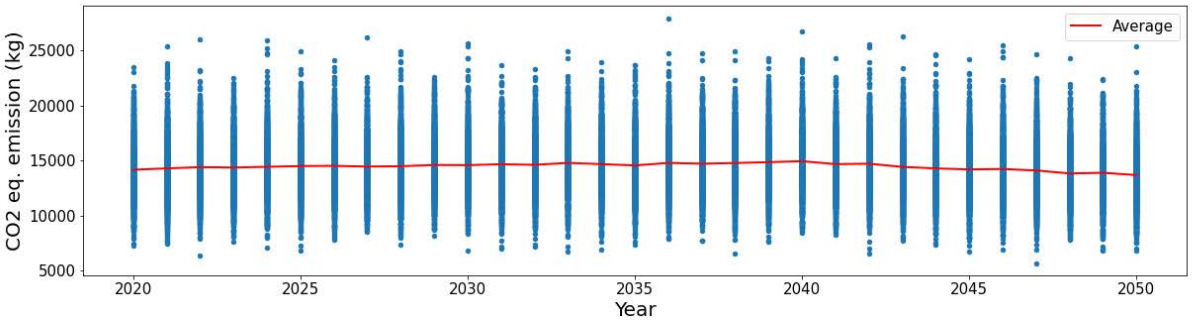

Figure 30: CO_2 emission (kg) of all simulations with model 2A

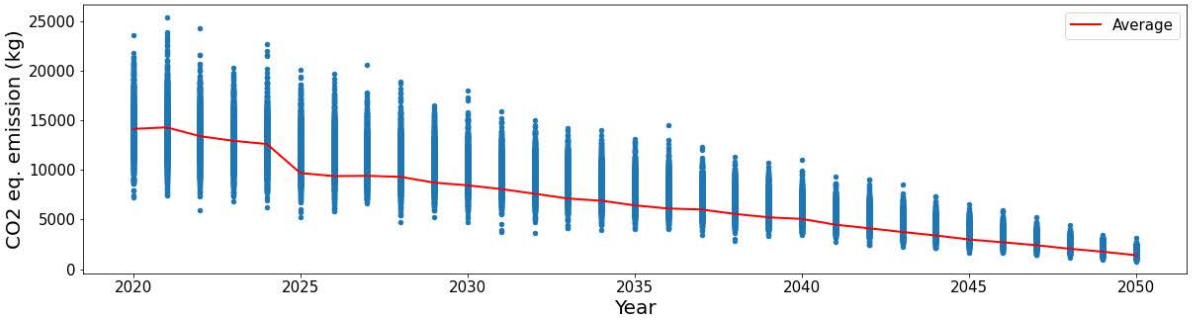

Figure 31: CO_2 emission (kg) of all simulations with model 2B

6.2.3 Comparison of model 1 and 2

The histograms of the simulation results for the year 2050 with model 1B (with dependency) and 2B (independency) are shown in Figure 32. Clearly, the distribution of model 1B is more narrow than the distribution of model 2B: the average value of model 1B is higher than the average value of model 2B, but its standard deviation is almost twice as small. Since in model 2B the assumption is that all variables are independent, a wider range of values for the variables length, width, maximum capacity, maximum draught, average speed and distance traveled is possible. This leads to more possible outcomes of the simulations. This is also visible in Figure 33: the standard deviation as well as the average value of the CO_2 emission is lower for model

1B. The emission factor changes from 2025. After 2025, the differences in average and standard deviation between the models decrease over the years and are almost equal by the year 2050. A Kolmogorov-Smirnov test is performed to test if the distributions of the results of the models are equal. The test returns a p-value of 7.97498 * $10^{-}90$, which means that the null hypothesis, that the results of model 1B and 2B are drawn from the same distribution, is rejected. The Wilcoxon signed-rank test, which is a non-parametric version of the paired T-test, also resulted in a very small p-value of 2.5851 * $10^{-}152$. Again, the null hypothesis that the results from model 1B and 2B follow the same distribution, is rejected. The distributions of the outcomes of the models are not the same. The Kruskal Wallis test tests if the averages of the results are equal. The result was a p-value of 6.6539 * $10^{-}76$, which denotes that the averages are not the same.

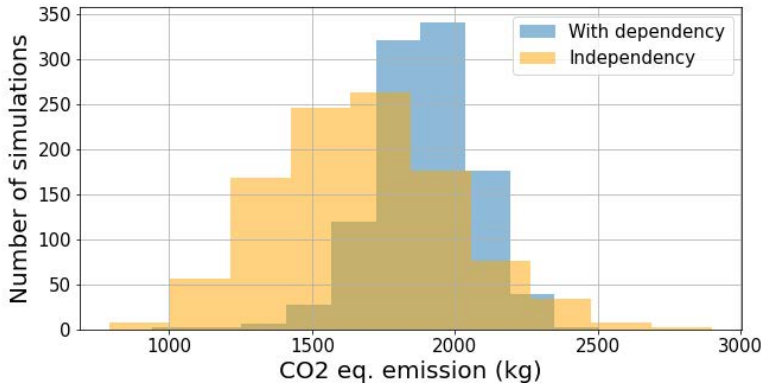

Figure 32: Comparison of the simulation output of the CO_2 emission in 2050

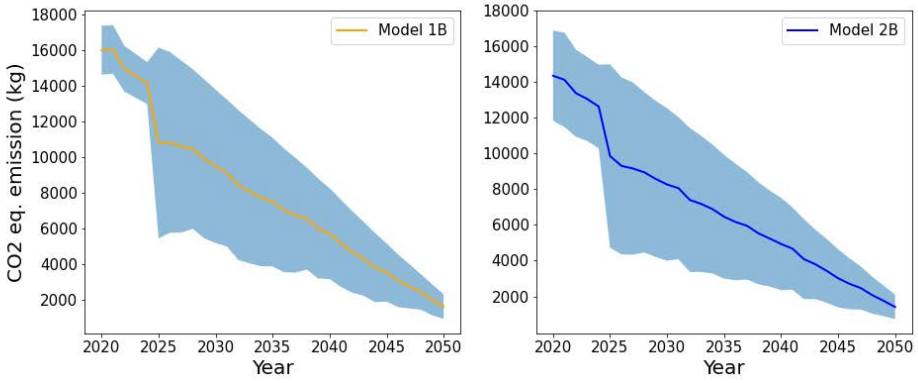

Figure 33: Average CO_2 emission and its standard deviation of model 1B and 2B

Figure 34 shows that the coefficient of variance, the ratio of the standard deviation to the average, of model 1B is smaller than that of model 2B for all years. The coefficient calculated for the years altogether is 0.0955 for model 1B and 0.1906 for model 2B. This is in line with the results that were seen in the previous figures.

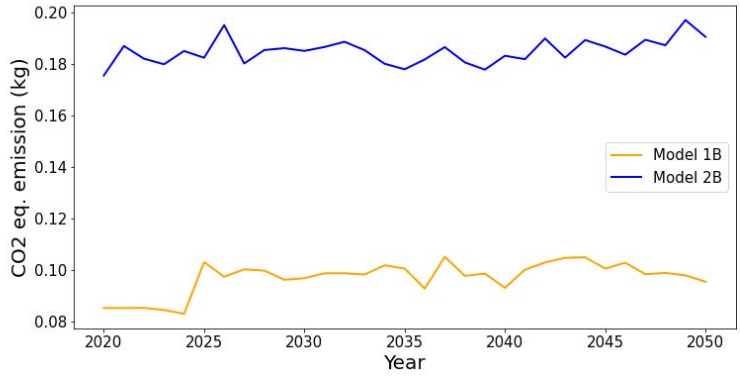

Figure 34: The coefficient of variance of model 1B and 2B

6.2.4 Different bin sizes

Figure 35 shows the average CO_2 emission for all years calculated with model 1 but with different bin sizes. Bin size 1.0 is equal to the bin size that was used in the results of model 2B that were previously seen. This bin size is changed by multiplicating this it by 1.2, 1.4, 1.6, 1.8 and 2.0. As the figure shows, the results are very similar. The bin size does not affect the results.

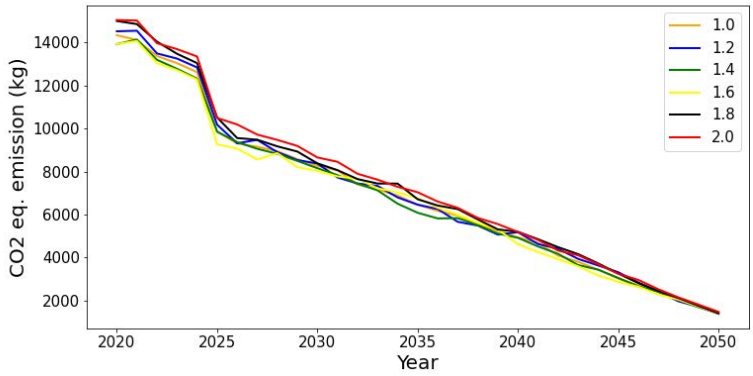

Figure 35: Average CO_2 emission of model 2B with different bin size

7 Conclusion

This thesis attempted to estimate the CO_2 that will be emitted by the inland vessels in the PoR by 2050. First the current CO_2 emission was quantified, secondly the distributions of several variables were derived and lastly the CO_2 emission was simulated until 2050. Under the assumptions and after several data transformations the calculated CO_2 emission of 122 inland vessels that operated in the segments 'containers', 'tankers', 'dry bulk' and 'break bulk' that visited the harbor 1850 times was $7.44 * 10^7$ kilograms of CO_2 emission. Unfortunately, it is difficult to validate this number. One can find a handful of sources that provide a number for the CO_2 emission of inland ships in the PoR, but these differ from each-other. Their method is also unspecified. The correctness of the results in this thesis are therefore invalidated.

Unfortunately, due to technical difficulties it was not possible to execute the simulations with the real throughput forecasts. Hence, the simulations were run with a initial fictional throughput of 500.000 tonnes in 2020. The growth factors of the real throughput forecasts were applied. The consequence is that in the simulations a very small number of vessels was needed to 'transport' the yearly throughput. The transition to zero emission energy sources depended on *group* classes 1 and 2 and the age on the engine, which were both of bad quality. Due to due to insufficient data many assumptions needed to be made. The probability that the *group* of a vessel was 1 or 2 and that it also needs a new engine, was low. Since just a few vessels were yearly needed in the simulations the number of 'zero emission' vessels was too low to have an impact on the results. Therefore, the decline of the CO_2 emission is determined by the decreasing emission factor. Specifically, the model in which the emission factor remained unchanged showed results that were (approximately) 8 times higher (in 2050) than those of the model where the emission changed according to the assumptions. This was because the value of the emission factor differed with a factor 8. Furthermore, if dependency between the some of the variables was taken into account, it led to simulation results with a higher average but where the values were less spread. Dependency between (some of) the variables should always be included, since this leads to realistic results. To conclude, although the main purpose of this thesis was partially fulfilled, the proposed method carried out with good quality data could produce the desired estimations.

8 Recommendations

Although the proposed method in this thesis appeared to be suitable to achieve the purpose, the expectation is that by decreasing the impact or avoiding the challenges that were faced with certain adjustments, the results can improve. This chapter proposes adjustments that perhaps can improve the method that was used in this thesis. These recommendations are aimed to obtain better results with the current state of the data sets and data infrastructure of the PoRA.

8.1 Estimation of the energy consumption of inland ships

The application of Bolt's model in Chapter 3.1 to estimate the energy consumption was challenging, because the data of several variables contained missing or incorrect values or were unusable. The model was applied to every observation of a subset of vessels in the AIS data set that were located in the PoR. Therefore the output was completely determined by the coordinates, the speed, the draught and the number of observations found in this data set. Errors and missing values affected the results tremendously. Hence the recommendation is to generalize the input values in order to produce results that are less dependent on each individual observation and therefore more stable. In this context, generalized input values are values that are not vessel and observation specific, but on average values of for example vessels classes. This section proposes a method to obtain generalized input values that presumably lead to more accurate results.

Choose sections
The expectation is that better results can be obtained by applying the model of Bolt per water way or section. If the behaviour of the vessels within a section is similar, good estimations of the variables can be made. If these estimations are used, every individual observations in the AIS data set are not necessary, which makes the results more stable. The impact of the errors in the data set can be therefore (partially) avoided. The next section will elaborate on the input variables.

The optimal number of sections has to be investigated. The sections can be chosen based on ,for example, traffic rate or water way characteristics. The latter is advised. The sections do not have to be the same size. More sections does not necessarily lead to better results. The most important requirement is that the behaviour of the vessels (and water way characteristics if desired) within a section should not differ much. Since it may be time consuming to execute this for the whole harbor, the most used waterways by inland vessels, the section(s) that presumably contribute(s) to the CO_2 emission the most or division of the harbor in a small number of sections can be used to start (or experiment). Mind that if the number of sections decreases it is likely that there will be more variability in value ranges of the input variables. If the energy consumption is then estimated while using the minimum and maximum values of the input values the results will be more spread and therefore less precise. On each input variable will be elaborated.

How to choose the input values
This section proposes a method to obtain more generalized values for the input variables. The assumption is that the maximum CO_2 emission is the most relevant, since this denotes the upper-bound of the emitted CO_2. This is calculated if the maximum possible values of all input variables are used to calculate the energy consumption. If the minimum values for all the variables are used this will lead to the to minimum CO_2 emission. The calculated minimum and maximum CO_2 emission will denote the upper and lower bound of the CO_2 emission in a section.

- **Draught**
  Chapter 2.3 mentioned that the quality of the data of the variable *draught* is questionable since this information has to be manually inserted by the sailor. The data also contained many missing values. There are several options to obtain an estimate of the value of this variable. Two examples follow. The first option is to use a model to estimate the maximum and minimum draught. In Chapter 3.1.1 the *minimum draught* was estimated via a model

found in [6]. This model needs the *length*, *width* and *segment* (container, dry bulk, etc.) of the vessel as input. The assumption is that these data are (mostly) accurate. In the same paper also a model to estimate the *maximum draught* can be found. The difficulty in the application of this model is that the intercept is not included as each ship type (i.e. containers, dry bulk, dumb barge, tanker) has its own intercept. The second option is to use an existing classification where each class is linked to vessel specifications. Examples are the CEMT or type of inland vessel (Spits, Kempenaar, Hagenaar, etc.). Both are linked to vessel specifications like the length, width, loaded draft and capacity. If the PoRA does not have such classification for each inland vessel a CEMT class could be assigned by comparing the data about the length, width (and/or capacity) with the class properties. The method that was used in this thesis can be found in 2.2. The loaded draft information of the CEMT class can be used as value for the draught. This is the maximum draught. If the data about the *draught* does not come from the AIS data set, more data could be used to estimate the the energy consumption, since this variable is only available starting from March 2020. This is a big advantage, because the small size of the data set influenced the results.

- **Average draught**
  From the results of the sensitivity analysis of this variable in Chapter 4.1 the conclusion was that the *average draught* can be chosen as the value of the draught.

- **Distance traveled**
  Since the sections are manually chosen the length, i.e. the maximum distance traveled, is known. This makes this variable a constant and therefore more accurate. As a consequence the AIS data set is not needed anymore for this variable.

- **Speed**
  The *speed* can be estimated by determining the minimum and maximum speed a vessel could (theoretically) have in the section. Note that this is an one-off activity and the values can be used for future calculations. If this can be achieved, the AIS data set is also for this variable not required anymore. If this is not possible the minimum and maximum observed speed of the sub-population can be used (after sanity checks). In this case the values are more dependent on the AIS data set then in the first option but are still section-specific and not observation-specific. Determining the speed per vessel class, if these are assigned for the variable *draught*, will make the results more precise.

According to the results of the sensitivity analysis in Chapter 4.2 the tidal streams have a small impact on the output. However the same tidal stream data of one location was used for every observation that occurred anywhere in the harbor, because it was not possible to import the complete data set for the whole harbor. In case the PoRA wants to include the tidal stream, the expectation is that for a specific section this information can be imported, since its size is much smaller. This data can be used to adjust the estimated minimum and maximum speed. The tidal stream data of a time period can be analyzed to obtain an estimate of the minimum, maximum and/or median value. Since the 'path' that vessels use within the section is probably clear, the vessel's direction with respect to the tidal stream is probably also clear. Otherwise it can be estimated.

- $\mathbf{u_{rel}}$
  Although the conclusion was that the variable u_{rel} was approximately equal to V, u_{rel} can, if desired, also be calculated with more precision. This variable needs, besides the *speed* and *draught*, the *midship cross section area*, the *water way cross section* and the *water depth* as input. The *midship cross section area* can be calculated by multiplicating the *width* by the *maximum draught*, as mentioned in Chapter 3.1.1. Chapter 3 mentioned that the data of the water way characteristics (depth, width and tidal streams) could not be used because of the data format currently. As mentioned at the previous variable, obtaining this data per section could perhaps avoid this problem. Furthermore, the expectation is that it is not necessary to import long time periods of data. The depth, width and consequently the water way cross

section will not differ much over time. Only for the water depth is a variation possible. The median water depth in a certain time period is presumably sufficient. This is also an one-off activity and the values can be used for future calculations. The value of u_{rel} will be much more precise than how it was calculated in this thesis.

In the best case the AIS data set is only used to determine when a vessel 'visited' the section. Every visit can be given a visit ID. The estimated energy consumption multiplied by the emission factor of diesel oil will give the CO_2 emission. Currently more or less all inland ships use diesel oil as energy source. Afterwards the results can be aggregated per vessel, visit ID or other variable. Analysis of the results can answer questions like where (i.e. what section) the most CO_2 is emitted and how the statistics of the variables differ between the sections. If, for example, the section with the most CO_2 emission shows higher values for the *speed*, perhaps a speed limit could lead to lower CO_2 emission. A dashboard could be made in which the values of (certain) variables can be manually changed, whereafter the energy consumption is recalculated and the results compared to see how they impact the outcome. If the variables *speed*, *draught* and, if present, u_{rel} are determined per vessel class the outcomes can be compared. For example, if vessels of a certain type of class were allowed or banned from the considered section. Overall the results can be useful in many ways.

Recall that only the data of the group of vessels that gave their consent may be used by the PoRA. To get an estimate of the total energy consumption it should be investigated what fraction the sub-population is with respect to the whole population. It should also be investigated if this group is a good representation of the whole population with respect to the type of vessels and number of times the section was visited. Secondly only the data of the vessels that operate in the segments that contribute to the throughput were used, because the simulations were based on the throughput. For a complete overview of the energy consumption the other segments could also be included. Lastly, the model of Bolt is not suitable for cases where the *speed* or the *distance traveled* is 0 (e.g. moored ships). Therefore the advise is to find a suitable model for the sections in which the vessel's movements show (mainly) this behaviour.

8.2 Groups

Although the expectation is that a good subdivision of the *group* solely based on the coordinates in the AIS data is difficult or perhaps not possible at this moment, this section proposes a different method that could perhaps lead to better results.

Chapter 3.1.2 explained that a group classification was attempted based on polygons. Unfortunately the proposed method gave bad results. Firstly, the polygons of all provinces in the Netherlands, Belgium and Germany were used. It turned out that this was unnecessary. The large number of polygons was the main reason the run-time was very large. To determine if a vessel sails long or short distances from the PoR was an time-consuming operation. This is obvious since every observation needs to be compared to all the considered polygons. Secondly, a part of the observations was wrongly classified as group 4. For example, if the polygons did not connect properly areas arised that technically did not belong to one of the considered polygons. Therefore, the observations that lie in these areas automatically end up in group 4.

The distributions in Chapter 3.2.1 and the results in Chapter 6.1 showed wide spread values and outliers, which were probably the effect of the absence of any aggregation by group. There are a couple of adjustments to the proposed method that simplify, still comply with the rules and could improve the results. Squares drawn around the PoR can be used instead of polygons to minimize the run-time. The advantage of squares compared to polygons or other shapes is that the coordinates only have to be compared to the coordinates at the corners of the squares. An example is shown in Figure 36. Corners A and C, but also corners B and D have the same latitude. Corners A and B, but also corners C and D have the same longitude. The coordinates of the observations have to be within these corners. The observations need to be compared with just 4 numbers instead of a (long) list of coordinates in a polygon. Logically, the smallest square should cover the area

in which vessels from group 1 operate. The PoRA can adjust each square to the size as well as the total number of the shapes that are considered to be the most useful for their decision-making process. The advise is to not use too many squares, because the run-time will increase. Also, more time should be invested in the classification rules. With the current rules, for example, a vessel that moves outside of the smallest square only once is not classified as group 1 although it should be. The classification rule to determine if a route is fixed or non fixed was based on the number of times a vessel visited the PoR. Perhaps this could be simplified by counting the number of unique vesselvisitid's per time period for each vessel.

Figure 36: The use of squares for group classification

8.3 Monte Carlo simulations

Although the proposed method for the Monte Carlo simulations led to the desired output, the results were not usable. The programming software Python in combination with the technical specifications of the server made it impossible to run the simulations with the real throughput forecast. Although a lot of time was invested into rewriting the code several times, it may be that it was still not fully optimized. Due to the small fictional throughput that was chosen, the results were completely determined by the emission factor. While additionally the emission factor over the years was not scientifically supported, since it is unknown how this will develop. Presumably experts on this subject could supply (somewhat) better predictions. The transfer from 'diesel oil' to any 'zero emission' did not play a significant role in the results while in reality it suppose to be the most important change towards the zero emission goal in 2050. It is interesting to know if the effect is more significant when the real or even a larger fictional throughput is used in the simulations. The transition to a zero emission energy source was determined by the *group* and *year of construction*, which were both of bad quality. On the *group* was already elaborated in the previous section. During the data exploration the conclusion already was that the quality of the data of the *year of construction* was not good enough to use. Perhaps the best option at the moment is to formulate new rules for the simulations concerning the energy transition.

If the information about the draught is not needed from the AIS data set, more historic data can be used to construct the distributions. This is a great advantage. The effect of the COVID-19 pandemic can be excluded (or integrated), the quality of the distributions will improve and the results as a consequence. The advise is then to make the distributions vessel classification-specific, since the behaviour can differ much. With more historic data a kind of recurrence rate for each vessel could also be determined in case the *year of construction* will still be used.

Lastly, instead of using model 1 to obtain a value for the *length*, *width*, *maximum capacity* and *maximum draught* in the simulations, a distribution of the vessel class can be used. This distribution will look similar to distribution A. A draw from this distribution will lead to realistic values for these variables, since a vessel class corresponds to specific vessel characteristics. This avoids unrealistic combinations of values of the variables in the simulations. Although model 1 took the correlation coefficients into account, (a small number of) unrealistic combinations of values of these variables could still have occurred.

References

[1] *A Degree of Concern: Why Global Temperatures Matter*. URL: https://climate.nasa.gov/news/2865/a-degree-of-concern-why-global-temperatures-matter/. (accessed: 16.10.2020).

[2] *About the Port Authority*. URL: https://www.portofrotterdam.com/en/port-authority/about-the-port-authority. (accessed: 18.01.2021).

[3] *ADOPTION OF THE PARIS AGREEMENT*. URL: https://unfccc.int/resource/docs/2015/cop21/eng/l09.pdf. (accessed: 1.9.2020).

[4] Mayur Agarwal. *Importance Of Ship's Keel and Types Of Keel*. URL: https://www.marineinsight.com/naval-architecture/importance-of-ships-keel-and-types-of-keel/. (accessed: 19.06.2020).

[5] Ernst Bolt. *Schatting energiegebruik binnenvaartschepen*. 2003.

[6] F. Vinke et al. C. van Dorsser. "The effect of low water on loading capacity of inland ships". In: *The European Journal of Transport and Infrastructure Research* 20(3) (2020), pp. 47–70.

[7] *Carbon Dioxide*. URL: https://climate.nasa.gov/vital-signs/carbon-dioxide/. (accessed: 16.10.2020).

[8] CHARMATZIS. *Port of Rotterdam map*. URL: https://portmaps.ad.portofrotterdam.com/PortMaps/Index.html?viewer=PortMaps.Portmaps_UH2&LayerTheme=9#. (accessed: 04.09.2020).

[9] *CO2-emissiefactoren*. URL: https://www.co2emissiefactoren.nl/. (accessed: 1.4.2020).

[10] *Facilities*. URL: https://www.portofrotterdam.com/en/shipping/inland-shipping/berths-for-inland-shipping/facilities. (accessed: 18.01.2021).

[11] *GLOBAL WARMING OF 1.5 ºC*. URL: https://www.ipcc.ch/sr15/chapter/spm/. (accessed: 16.10.2020).

[12] *Green Deal No. GD230 Zeevaart, Binnenvaart en Havens*. URL: https://www.greendeals.nl/green-deals/green-deal-zeevaart-binnenvaart-en-havens. (accessed: 14.10.2020).

[13] Sumeet R. Patil H. Christopher Frey. "Identification and Review of Sensitivity Analysis". In: *Risk Analysis* 22.3 (2002), pp. 553–578.

[14] Hapag-Lloyds. *How Vessel Trim Optimisation creates efficiencies*. URL: https://www.hapag-lloyd.com/en/news-insights/insights/2017/03/how-vessel-trim-optimisation-creates-efficiencies.html#:~:text=The%20trim%20of%20a%20ship,demand%20for%20propulsion%20during%20sailing.. (accessed: 06.01.2021).

[15] *Klimaatakkoord gepresenteerd op 29 juni 2019*. URL: https://www.klimaatakkoord.nl/documenten/publicaties/2019/06/28/klimaatakkoord. (accessed: 14.10.2020).

[16] De Vlaamse Milieu Maatschappij. *Emissiemodel voor spoorverkeer en scheepvaart in Vlaanderen*. 2007.

[17] Barry L. Nelson Marne C. Cario. "Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix". In: (1997).

[18] *Quickscan langere termijn impact coronacrisis op haven*. URL: https://www.portofrotterdam.com/nl/nieuws-en-persberichten/quickscan-langere-termijn-impact-coronacrisis-op-haven. (accessed: 14.10.2020).

[19] Dirk P. Kroese Reuven Y. Rubinstein. *Simulation and the Monte Carlo method*. Wiley series in probability and statistics. John Wiley Sons, Inc., 2017. ISBN: 9781118632208.

[20] Rijkswaterstaat. *Richtlijnen Vaarwegen 2020*. 2020.

[21] DCMR Milieudienst Rijnmond. *C02-uitstoot Rotterdam*. 2019.

[22] Centrale Commissie voor de Rijnvaart. *Informatieblad Inland AIS*. 2011.

[23] Centrale Commissie voor de Rijnvaart. *Mogelijkheden om het brandstofverbruik en de broeikasgasemissies in de binnenvaart te reduceren*. 2012.

[24] Port of Rotterdam. *Biggest Dutch project for CO2 reduction, Porthos, is on schedule*. URL: https://www.portofrotterdam.com/en/news-and-press-releases/biggest-dutch-project-for-co2-reduction-porthos-is-on-schedule. (accessed: 17.01.2021).

[25] Port of Rotterdam. *Carbon neutral in 3 steps*. URL: https://www.portofrotterdam.com/en/doing-business/port-of-the-future/energy-transition/carbon-neutral. (accessed: 17.01.2021).

[26] Port of Rotterdam. *Consent for use of AIS data*. URL: https://www.portofrotterdam.com/en/consent-for-use-of-ais-data. (accessed: 19.06.2020).

[27] Port of Rotterdam. *Throughput Port of Rotterdam*. 2019.

[28] Port of Rotterdam. *Vessels*. URL: https://www.portofrotterdam.com/en/our-port/facts-and-figures/facts-figures-about-the-port/vessels. (accessed: 17.01.2021).

[29] Port of Rotterdam. *Zero-emission port by 2050*. URL: https://www.portofrotterdam.com/en/news-and-press-releases/zero-emission-port-by-2050. (accessed: 19.06.2020).

[30] *Schepen*. URL: https://www.portofrotterdam.com/nl/onze-haven/feiten-en-cijfers/feiten-en-cijfers-over-de-haven/schepen. (accessed: 21.04.2020).

[31] C. Trozzi. *Emission estimate methodology for maritime navigation*. 2010.

[32] C. Trozzi. *Methodologies for estimating air pollutant emissions from ships: a 2006 update*. 2006.

[33] *Wat is het Klimaatakkoord?* URL: https://www.klimaatakkoord.nl/klimaatakkoord/vraag-en-antwoord/wat-is-het-klimaatakkoord. (accessed: 18.01.2021).

[34] *What is the Paris Agreement?* URL: https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement. (accessed: 18.01.2021).

A Appendices

A.1 Bio fuel percentages

Year	Scenario 1
2020	0
2021	0
2022	7
2023	9,875
2024	12,75
2025	15,625
2026	18,5
2027	21,375
2028	24,25
2029	27,125
2030	30
2031	32,875
2032	35,75
2033	38,625
2034	41,5
2035	44,375
2036	47,25
2037	50,125
2038	53
2039	55,875
2040	58,75
2041	61,625
2042	64,5
2043	67,375
2044	70,25
2045	73,125
2046	76
2047	78,875
2048	81,75
2049	84,625
2050	87,5

Table 31: Percentages bio fuel per year used in the simulations

A.2 Correlation matrix

	Length	**Width**	**Maximum capacity**	**Draught**	**Year of construction**
Length	1.00	0.82	0.91	0.23	0.38
Width	0.82	1.00	0.89	0.16	0.28
Maximum capacity	0.91	0.89	1.00	0.20	0.32
Draught	0.23	0.16	0.20	1.00	-0.08
Year of construction	0.38	0.28	0.32	-0.08	1.00
Maximum draught	0.69	0.75	0.84	0.18	0.33
Speed	0.14	0.12	0.19	0.05	0.16
Distance traveled	0.11	0.07	0.14	0.06	0.13

	Maximum draught	**Speed**	**Distance traveled**
Length	0.69	0.14	0.11
Width	0.75	0.12	0.07
Maximum capacity	0.83	0.19	0.14
Draught	0.18	0.05	0.06
Year of construction	0.33	0.16	0.13
Maximum draught	1.00	0.19	0.15
Speed	0.19	1.00	0.72
Distance traveled	0.15	0.72	1.00

A.3 Covariance matrices

This section shows the covariance matrices of the data, X, Z and X_{sim}. These are calculated after a 1000 draws. This is performed for two multivariate distributions: the vessel characteristics variables (length, width, maximum capacity and maximum draught) and the visit characteristics (speed and distance traveled).

Covariance matrices of the vessel-specific variables

$$Cov(\mathrm{X}) = \begin{pmatrix} 1.000 & 0.766 & 0.839 & 0.516 \\ 0.766 & 1.000 & 0.812 & 0.529 \\ 0.839 & 0.812 & 1.000 & 0.752 \\ 0.516 & 0.529 & 0.752 & 1.000 \end{pmatrix}$$

$$Cov(\mathrm{Z}) = \begin{pmatrix} 1.074 & 0.830 & 0.917 & 0.550 \\ 0.830 & 1.057 & 0.878 & 0.546 \\ 0.917 & 0.878 & 1.074 & 0.784 \\ 0.550 & 0.546 & 0.784 & 1.012 \end{pmatrix}$$

$$Cov(\mathrm{X}_{sim}) = \begin{pmatrix} 0.930 & 0.652 & 0.744 & 0.444 \\ 0.652 & 1.117 & 0.719 & 0.431 \\ 0.744 & 0.719 & 1.083 & 0.727 \\ 0.444 & 0.431 & 0.727 & 0.951 \end{pmatrix}$$

Covariance matrices of the visit-specific variables

$$Cov(\mathrm{X}) = \begin{pmatrix} 1.000 & 0.172 \\ 0.172 & 1.000 \end{pmatrix}$$

$$Cov(\mathrm{Z}) = \begin{pmatrix} 0.793 & 0.319 \\ 0.319 & 0.701 \end{pmatrix}$$

$$Cov(\mathrm{X}_{sim}) = \begin{pmatrix} 0.979 & 0.196 \\ 0.196 & 0.981 \end{pmatrix}$$

A.4 Throughputs

Year	Throughput (tonnes)
2020	500000.0
2021	501380.0
2022	502760.0
2023	504140.0
2024	505520.0
2025	506900.0
2026	508280.0
2027	509660.0
2028	511040.0
2029	512420.0
2030	513800.0
2031	514488.5
2032	515177.0
2033	515865.5
2034	516554.0
2035	517242.5
2036	517931.0
2037	518619.4
2038	519307.9
2039	519996.4
2040	520684.9
2041	516519.4
2042	512354.0
2043	508189.0
2044	504023.0
2045	499857.5
2046	495692.0
2047	491526.6
2048	487361.1
2049	483195.6
2050	479030.1

Table 32: Fictional throughput used in simulations