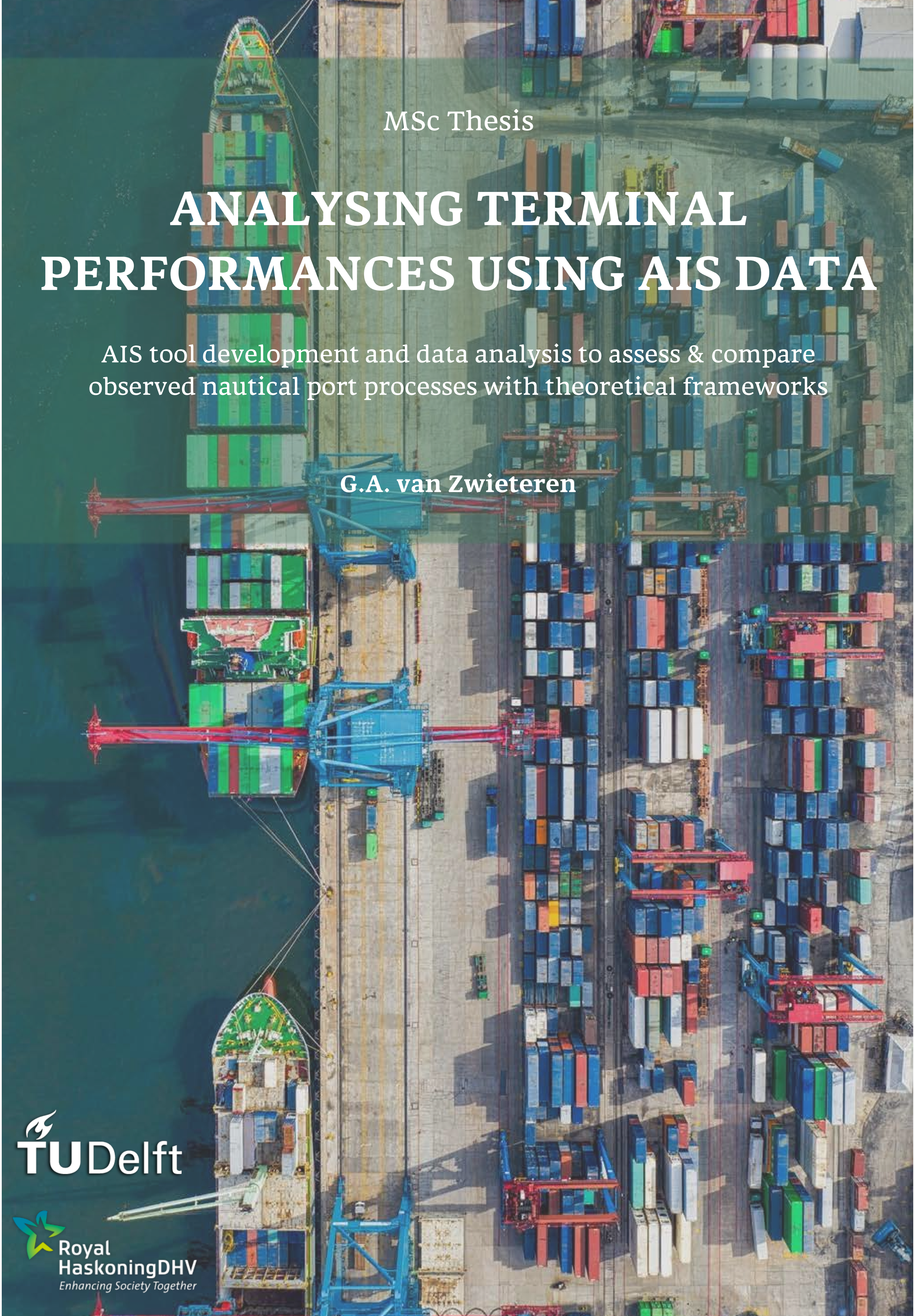
MSc Thesis

ANALYSING TERMINAL PERFORMANCES USING AIS DATA

AIS tool development and data analysis to assess & compare observed nautical port processes with theoretical frameworks

G.A. van Zwieteren



TU Delft

Royal HaskoningDHV
Enhancing Society Together

Analysing terminal performances using AIS data

AIS tool development and data analysis to assess & compare observed nautical port processes with theoretical frameworks

by

Gabriëlle van Zwieteren

In partial fulfilment of the requirements for the degree
Master of Science in Civil Engineering

To be defended publicly at Delft University of Technology,
on Thursday, October 15, 2020 at 12:00.

Student number	4316207	
Master track	Hydraulic Engineering	
Supervisors	(Chair) Prof.dr.ir. M. van Koningsveld	Delft University of Technology
	Ir. A.J. Lansen	Delft University of Technology
	Prof.dr.ir. P.H.A.J.M. van Gelder	Delft University of Technology
	Dr. F. Baart	Deltares
	Ir. S. Klaver	Royal HaskoningDHV

This report is available online at http://repository.tudelft.nl/

TU Delft

Royal HaskoningDHV
Enhancing Society Together

Preface

This research marks the final stage of my time as a student at Delft University of Technology. It therefore presents the final effort to obtain the Master of Science degree in Civil Engineering, with a master track in Hydraulic Engineering. The research was conducted in collaboration with the maritime department of Royal HaskoningDHV, a renowned engineering firm, by means of a graduate internship.

In my opinion, a thesis presents the final work a student performs, in which all previous years of studying and experiences can be bundled together to the highest reachable level. I was eager to start this final phase and hence, was very happy to find a combination between port planning and working with large data sets. My thesis started off by exploring all the possibilities and limitations AIS data has to offer. Throughout this research, I have obtained a love-hate relationship with AIS data, where all the endless possibilities and amounts of data keep amazing me and the rawness of the data could overwhelm me at times. Overall, I look back at the last few months with great pleasure and dedication, and am proud of the result obtained.

This thesis would not have been realised if it weren't for the help of others. I would like to take this opportunity to express my gratitude to the people that have guided me throughout this process.

First, I would like to thank the chair of my committee, Mark van Koningsveld, for sharing his inspiring views, knowledge and enthusiasm on the topic. I would like to thank the other committee members, Joost Lansen, who has introduced me to the topic, Pieter van Gelder and Fedor Baart. The positive, critical feedback I received from you all has been a continuous trigger for me to improve. The different backgrounds of the committee members has resulted in a great amount of knowledge input, leading to me receiving many fresh perspectives from every meeting. Last, I would like to thank my daily supervisor at Royal HaskoningDHV, Sebastiaan Klaver, for sharing your knowledge about port planning and always giving me constructive, detailed feedback. I would leave our weekly meetings with more energy and new motivation to tackle the Python errors and other complex issues that have come my way. Even though the Covid-19 virus has led me to perform this research from home, I never felt alone and always recognized and enjoyed your dedication towards this research.

Finally, I would like to express my gratitude towards my friends and family. I would like to thank my friends from Het Waterbouwdispuut, which have made the experiences during this Master unforgettable. My roommates in Rotterdam for your support during both joyful and more challenging times, the 'Covid time' would not be the same without you. Finally, I want to thank Roel and my family, for always believing in me and giving me unconditional support throughout my studies.

Gabriëlle van Zwieteren
October 2020

Executive summary

Ports are important intermodal hubs part of large global supply chains. Nowadays, ports are focusing more and more on digital transformations. Nonetheless, one aspect has been lacking behind when it comes to these digital transformations. *Permanent International Commission for Navigation Congresses* (*PIANC*) guidelines suggest that the most optimal design of a terminal is found by using queuing theory. Queuing theory defines the average waiting time in units of average service time, in order to find the number of berths based on an expected berth occupancy. Queuing theory mainly depends on how the inter arrival times and service times are distributed. In 1985 *United Nations Conference on Trade and Development* (*UNCTAD*) introduced certain recommendations for these distributions which are still being used today, but how accurate can these recommendations be in 2020? By using available *Automatic Identification System* (*AIS*) data the recommendations for these distributions can be verified. How relevant are these recommendations and are there more favorable methods for terminal design, possibly leading to higher efficiencies and improved use of port infrastructure?

To verify these recommended distributions, large amounts of AIS data are considered. Initially, AIS was implemented on all sea going vessels to improve safety and efficiency at sea. It is an automatic tracking system broadcasting static and dynamic information about the vessel. This research inspects the possibilities and usefulness of AIS data for research purposes. The service times, inter arrival times and berth occupancy are examined for different terminals, based on available AIS data, and compared to theoretical frameworks. Based on this knowledge, the main research question is introduced:

How are service times & inter arrival times distributed and can the berth occupancy be defined, at container-, dry bulk- and liquid bulk terminals, based on AIS data, and how do these compare with design guidelines?

AIS Tool development

In this research a tool is developed which transforms AIS data, into a data set containing information, for every vessel track, about the entry and exit times of the port, anchorage and terminal areas. The AIS Port processes tool is developed using Python programming language and is available on GitHub (Van Zwieteren, 2020). The tool uses a XGBoost classifier to predict whether or not a certain vessel berths at the terminal, predicting with an accuracy of 97%. Multiple research purposes can benefit from the usage of this tool as it is equipped to handle any type of AIS data set, for any terminal location.

AIS data analyses and comparisons to theoretical framework

The focus in this report lies on three different study parameters: the service times, the inter arrival times and the berth occupancy. For twelve different terminals the AIS Port processes tool is used to analyse the AIS data. To verify whether the expected distributions, based on PIANC guideline recommendations, correspond to the observed data, the *Kolmogorov-Smirnov* (*K-S*) goodness-of-fit test is performed.

First, the service time distributions are analysed. The service time is defined as the total time a vessel spends at the terminal. It is influenced by multiple factors, such as the vessel type, the quantity and type of cargo and the cargo handling rate. The smallest average service times are observed for container terminals, whilst the dry bulk vessels contain the highest average service times. All dry bulk terminals handle both import and export transfers, resulting in large differences within the service time distribution of these specific terminals. Whereas the container terminals, which only handle specifically containers and thus contain very similar handling equipment, lead to similar service times. PIANC expects the Erlang-k distribution to correctly represent the service time distributions. This contradicts with the observations from the data, where almost all distributions did not comply with any theoretical fit.

When investigating smaller sub sets of the data of container and dry bulk terminals (which are split based on different vessel classes) the Gamma distribution regularly appears as a good theoretical fit to the observed service time distributions. For liquid bulk terminals these smaller data sets contain no clear theoretical fit and their service times tend towards a Deterministic distribution. The Erlang-k distribution might have previously been a correct recommendation for the service times, but nowadays the difference between vessel types and sizes is larger, leading to more diverse and spread out service time distributions. Concluding, the container and dry bulk terminal service time distribution can be represented as a heterogeneous data set, built up of smaller homogeneous sub data sets, each best represented by a Gamma distribution.

Second, the inter arrival time distributions are investigated for the same terminals. The inter arrival time is defined as the time between two successive arrivals at a port. PIANC guidelines recommend using the Negative Exponential distribution to represent the inter arrival times. However, suggestions have been made that this representation nowadays might be too conservative based on improved terminal scheduling. The inter arrival time distributions of the dry- and liquid bulk terminals are best fitted by the Negative Exponential distribution. However, the container inter arrival times don't seem to correspond based on K-S goodness-of-fit tests, but visually do also follow the Negative Exponential distribution. When split into smaller data sets, again the Negative Exponential distribution is often the best representative for the inter arrival times. The data can thus be seen as *Independent and identically distributed* since the smaller data sets share the same probability distributions and are independent of each other. Overall, the conclusion is made that vessels still arrive in a stochastic manner at ports, despite the efforts of improved arrival scheduling.

Third, the berth occupancy is inspected. For the container and dry bulk terminals the berth occupancy can not be calculated due to the terminals having an inconsistent number of berths over time. These terminals are assessed by determining the adjusted length occupancy, defined as the total length occupied in comparison to the total length available. On average, a longer terminal is expected to be more flexible and is therefore assumed to contain a higher average occupancy. This relationship has been observed clearly for the container terminals, but was less obvious for the dry bulk terminals. For the liquid bulk terminals the berth occupancy was analyzed. Here, a higher number of berths is correlated to a higher average berth occupancy.

Queuing theory can be seen as a vital ingredient for terminal planning and design. The theory predicts the average waiting time in terms of average service time, from which conclusions about the number of berths can be made. The main inputs of queuing theory consist of the service time and inter arrival time distributions. In this research, twelve terminals are analysed and compared to distributions that PIANC guidelines recommend. Based on this research, it is recommended that for container and dry bulk terminals the service time distribution should be seen as a heterogeneous data set, built up of smaller homogeneous sub data sets (each representing a vessel class from the vessel mix). These smaller data sets can be best represented by the Gamma distribution, instead of the previously assumed Erlang-k distribution. The liquid bulk terminal service times are best represented by a Deterministic distribution, with an average service time of roughly 24 hours. For all terminal types, and corresponding to PIANC guidelines, the arrivals of vessels has a stochastic character, leading to the Negative Exponential distribution best representing the inter arrival times.

The work represented in this report improves the understanding of the essential distributions required when using queuing theory. The **AIS Port processes tool** has been developed and can be used to create various important insights into existing terminals, as well as lead to improvements for new terminal designs. This can have significant impacts on future terminal designs, providing economic advantages to vessel and terminal operators. Based on the developed tool and the research performed, multiple recommendations for further research are available.

Contents

List of Tables

List of Figures

Acronyms

AIS	Automatic Identification System
CDF	Cumulative distribution function
DWT	Dead Weight Tonnage
GNSS	Global Navigation Satellite System
IDD	Independent and identically distributed
IMO	International Maritime Organization
K-NN	K-Nearest Neighbors
K-S	Kolmogorov-Smirnov
kB	Kilobyte
LOA	Length Overall
LPP	Length between perpendiculars
MMSI	Maritime Mobile Service Identity
NED	Negative exponential distribution
PDF	Probability density function
PIANC	Permanent International Commission for Navigation Congresses
RBF	Radial basis function
RHDHV	Royal HaskoningDHV
SHAP	SHapley Additive exPlanations
SOG	Speed over ground
SVM	Support Vector Machine
TEU	Twenty Foot Equivalent Unit
UNCTAD	United Nations Conference on Trade and Development
VHF	Very High Frequency
VTS	Vessel Traffic Service
XAI	Explainable Artificial Intelligence

1 Introduction

The research objective and framework are introduced in this chapter. The problem description is followed by a summation of the current research gaps. Following the research gaps, the main research question is introduced followed by the sub research questions. Finally, the research scope and report outline are given.

1.1. Context introduction

In the 1960s ports all over the world experienced an exponential growth in container transport which advanced the ports into intermodal hubs that besides transportation services, also started to invest in distribution, logistics and other value adding services. Nowadays, ports are focusing on digital transformation and these transformations are all about controlling operations by continuously measuring and effectively using these growing data sources, to improve current logistics and infrastructure (Heilig et al., 2017). Despite these digital transformation efforts, the maritime industry is lacking behind in terms of digital innovation compared to other industries (Carlan et al., 2017). Collaboration between all actors of the port is necessary to fully benefit from all the growing possibilities these new technologies have to offer.

The arrival of vessels at ports has a stochastic character which leads to uncertainty for, among others, the terminal operator. Besides the random arrival patterns of vessels, the service times of a vessel depends on many variables resulting in a stochastic character as well. A fully occupied berth (in other words: 100% occupancy) can therefore not exist without a continuous queue of vessels waiting to enter the port. This relationship between service time and berth occupancy is based on Queuing Theory. This theory is used when designing a port, to estimate the expected average waiting times in terms of average service times, and depends on the terminal occupancy and the number of berths. *United Nations Conference on Trade and Development* (*UNCTAD*) and *Permanent International Commission for Navigation Congresses* (*PIANC*) guidelines demonstrate how different inter arrival and service times will lead to different expected waiting times, based on the combination of number of berths and berth occupancy (UNCTAD, 1985; PIANC WG158, 2014).

In 2000 the *International Maritime Organization* (*IMO*) decided that sea vessels should implement *Automatic Identification System* (*AIS*) to increase safety and efficiency at sea. By the end of 2004 all vessels larger than 300 gross tonnage were required to be equipped with AIS (de Boer, 2010). AIS is an automatic tracking system which broadcasts information over *Very High Frequency* (*VHF*) radio waves. Increasing amounts of AIS data are expected in the future, as regulation at sea keeps growing (Windward, 2014). With these increasing amounts, many valuable insights can be gained into the behavior of ships and operations at ports and waterways. AIS was not designed with the intention of being a research application and in the meantime the availability is not (yet) restricted by governments or the IMO (Robards et al., 2016).

1.2. Problem statement

Ports consist of multiple vital functions, such as facilitating the transfer of goods across the world. Ports not only help the exchange of cargo, but they also establish as energy hubs and significant employment possibilities (European Commission, 2015). It is important to realise that not a single port is the same and that the port and it's terminals all vary in size, function and layout. However, the underlying subsystems of the terminals are the same (Meijer, 2017).

When designing a port, the quay length is determined by an estimation of the berth occupancy and the number of berths. The number of berths can be chosen based on the outcomes of queuing theory. Queuing theory is based on the inter arrival- and service time distribution and returns the expected average waiting times in terms of service times. Based on a chosen maximum allowable waiting time (in terms of service time), the number of berths can be selected based on the expected occupancy factor (Ligteringen, 2017). To apply the theory the port must be simplified and no complex variables can be included (such as weather influences) (Bellsolà Olba et al., 2014).

Queuing theory is thus used to determine the average waiting time, resulting in an estimation of the number of berths, using different distributions for the service and inter arrival times. This average waiting time is greatly dependent on the choice of the inter arrival- and service time distributions. In port transportation systems the uncertainty of arrival times is one of the main problems (Meijer, 2017; Parolas, 2016).

UNCTAD published a handbook for planners in developing countries for port development in 1985 (UNCTAD, 1985) in which this relationship between berth occupancy and service times was introduced. For the inter arrival times and the service times different distributions were selected. For example, for break-bulk cargo terminals the Erlang-1 arrival distribution is assumed and for dry bulk cargo terminals (where there is some tendency towards scheduling) the smoother Erlang-2 distribution is assumed. Statistically, the negative exponential distribution (NED) proves to represent the random arrival process at ports quite well since it is dependent on many external factors and different port operators (Bellsolà Olba et al., 2018). Following the publishing of the UNCTAD Handbook (UNCTAD, 1985) no research has been done to define the inter arrival times and service times based on AIS data. Over the following years, after the UNCTAD publication, *PIANC* reports represent the same outputs and guidelines as were defined in the initial UNCTAD Handbook.

Since 2004 all vessels larger than 300 gross tonnage are required to use *Automatic Identification System* (*AIS*), which sends out the vessels position and various other properties. The AIS data represents the reality, thus including different external factors (such as weather influences). Since 2004 numerous studies have been performed using AIS data.

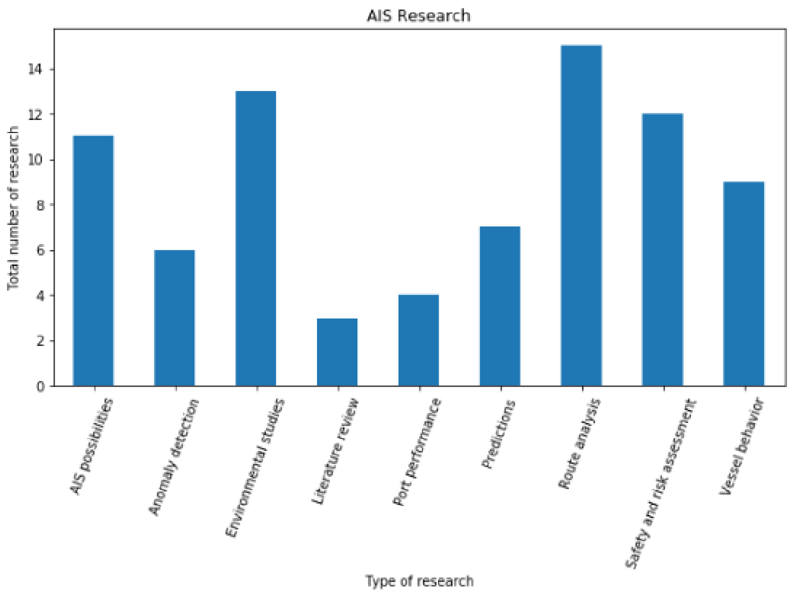


Figure 1.1: AIS research type

A total of 80 different research studies have been analysed for this research. Figure 1.1 represents these researches and their different types, performed between 2004 and 2019 (classified in appendix A.2.1). A total overview of all research found is given in appendix A.2.3. Thus based on the performed literature review the conclusion can be drawn that there has been no research so far which uses AIS data to determine if these statistical distributions, given by PIANC, are similar to the reality.

Furthermore, in order to use AIS data the raw data should be cleaned, processed and transformed into an output where, per customer, different statistics are returned, for example the service time and waiting time. No method exists yet which transforms raw data into cleaned processed data. This data should represent the different processes a vessel follows in a port.

To summarise, this has led to the following problems:

- Service- and inter arrival times are assumed to follow certain distributions, based on theoretical assumptions. These distributions have not been compared to reality based on AIS data and therefore remain unvalidated.
- The theoretical guidelines used in port planning do not take into account actual port operations with all possible complex variables and external influences. The guidelines and chosen distributions are based on assessing ports as simple transport systems. Thus the guidelines following these assumptions, might not be accurate representations of reality.
- AIS data collected is untreated and might contain errors such as duplicate messages. No method exists which cleans and processes raw AIS data and transforms it into a clear, useful data set which highlights the various port processes a vessel follows in a port.

1.3. Research objectives

The problem statement has led to the main research question:

How are service times & inter arrival times distributed and can the berth occupancy be defined, at container-, dry bulk- and liquid bulk terminals, based on AIS data, and how do these compare with design guidelines?

In order to answer the research question, the following sub questions and objectives are formulated, split into four sections:

1. AIS data: processing and possibilities
   (a) *How can AIS data be used to define different processes a vessel follows in a port?*
   Raw AIS data should be thoroughly cleaned and processed before it can be analysed. After the data is processed, vessel tracks can be defined for vessels entering the port and selected terminal. Based on these timestamps different port processes can be defined for every vessel track.

   (b) *What is the most optimal procedure of extracting vessel tracks that berth at a terminal, using AIS data?*
   The data contains all AIS messages sent by all types of vessels. For the research the interest lies specifically in vessels that berth at a certain terminal. Therefore, from a data set containing all data, only the data should be extracted regarding vessels that actually berth. AIS messages do not contain information about vessels berthing or not, so multiple methods will be tested in which an attempt is made to extract these certain vessels.

   (c) *What are performance analyses which can be generated using AIS data?*
   Using AIS data varying types of port performance analyses can be performed. All possible analyses are demonstrated to get an overview of the many different performance indicators that can be generated using AIS data.

2. Service time distribution
   (a) *How are service times distributed along container-, dry bulk- and liquid bulk terminals, based on AIS*

data, and how do they compare to PIANC guidelines?
A comparison is made between different types of terminals, to see which distribution best fits each different set of service times, according to AIS data. The different commodity types that were studied are containers, dry bulk and liquid bulk. The fitted distributions will be compared to current design guidelines.

   (b) *How are the service times distributed per vessel class along container-, dry bulk- and liquid bulk terminals, based on AIS data?*
   After analysing the service times of an entire vessel mix, the service times based on certain vessel classes will be examined. The difference between service times within one single terminal will be inspected.

   (c) *How do the three terminal types (container, dry bulk, liquid bulk) compare based on service times?*
   The three terminal types are compared based on the service times. A distinction is made between the different vessel classes in between the service times.

3. Inter arrival time distribution

   (a) *How are inter arrival times distributed along container-, dry bulk- and liquid bulk terminals, based on AIS data, and how do they compare to PIANC guidelines?*
   A comparison is made between different types of terminals, to see which distribution best fits each different set of inter arrival times of vessels, according to AIS data. The different commodity types accessed are containers, dry bulk and liquid bulk. The fitted distributions will be compared to current design guidelines.

   (b) *How are the inter arrival times distributed per vessel class, along container-, dry bulk- and liquid bulk terminals, based on AIS data?*
   Next to the inter arrival time distribution being measured for the entire mix of vessel classes arriving, the inter arrival times based on certain vessel classes will be examined.

4. Berth occupancy

   (a) *Can the berth occupancy be defined for container-, dry bulk- and liquid bulk terminals, based on AIS data, and is there a correlation between the occupancy and the terminal size?*
   The number of berths present at a terminal should be constant over time before the berth occupancy can be defined. The possibilities of the berth occupancy are discussed and other methods are examined to determine a type of terminal occupancy, such as the length occupancy. The correlation is examined between the size of a terminal (number of berths or quay length) and the occupancy.

1.4. Scope of the research

To further define the research area a concise scope is specified. More in-depth information about the scope is defined in part I of this thesis, based on the performed literature study. In this research three different terminals are analysed: container terminals, dry bulk terminals and liquid bulk terminals.

Considering that terrestrial AIS data has a maximum coverage of 40 km (Zhao et al., 2014), the port boundary for this research is set at the sea side of the designated anchorage area(s). Analyzed data sets include AIS messages from the moment that vessels enter or pass the anchorage area until the moment they leave or pass this area.

To deliver a tool which is as generic as possible only four parameters of the AIS data are selected: *Maritime Mobile Service Identity* (*MMSI*), longitude, latitude, timestamp. This also limits the chance of parameters containing wrongful input or missing data. From the unique identification number, the *MMSI*, multiple vessel characteristics are coupled to the data. The following parameters are merged with the AIS data, based on a different data source (*Known at RHDHV*): vessel type, *Length Overall*, *Dead Weight Tonnage*, *Twenty Foot Equivalent Unit* capacity.

1.5. Research approach

In order to answer the research question and sub questions the gathered AIS data will be cleaned, enriched and transformed. With this processed data, a new pandas Data Frame will be created with entry and exit times for the port area, the anchorage area and the terminal area. A pandas Data Frame is a two-dimensional, size-mutable tabular data set (Pandas, 2020). Using this new data frame, tools can be built to extract the service time distribution and the inter arrival time distribution. With the service time the berth occupancy can be determined. The approach for this research is visualised in figure 1.2. An iterative approach is used by re-adjusting the AIS tool based on (intermediate) results.

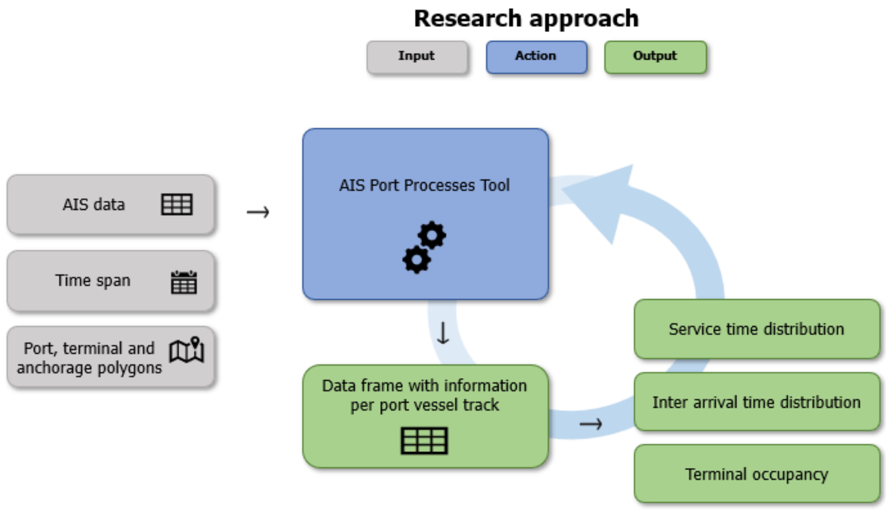


Figure 1.2: Research approach

1.6. Report outline

The report outline is as follows:

- Part I: Literature, materials and method
  - In chapter 2 relevant background information is given regarding current port processes and design guidelines, and the study parameters are defined.
  - In chapter 3 all available data sets, specifically AIS data, are introduced. The methodology of the research is addressed.
- Part II: Results
  - In chapter 4 the results of the AIS port processes tool are presented and discussed.
  - Chapter 5 represent the results based on the service time distribution.
  - The inter arrival time distribution results are discussed in chapter 6.
  - Finally, the berth occupancy results are discussed in chapter 7.
- Part III: Discussion, conclusions and recommendations
  - In chapter 8 the results are discussed.
  - Chapter 9 summarises the final conclusions for this research.
  - Recommendations for further research are given in chapter 10.

I

Literature, materials and method

- Chapter 2: Background literature and definitions of the study parameters
- Chapter 3: Available data and methodology

Part I elaborates on valuable literature research, the materials used and the methodology of the research. First, in chapter 2, the port processes are introduced in order to clearly define the study parameters. Current port design guidelines are discussed and queuing theory is described. In chapter 3 all data sources are introduced, including a thorough description of AIS data and its capacities and limitations. Next, the methodology used in this research is clarified. The methodology is split up into the AIS tool set up, the required statistical analysis and the distribution fitting approach.

2

Background literature and definitions of the study parameters

In this chapter the theoretical framework of the research is presented in order to define the different study parameters. First, an introduction is given to different port processes based on relevant literature. Next, the different port and terminal design guidelines are discussed. Queuing Theory is thoroughly explained and the different distributions which can be implemented in this theory are introduced. To summarise, chapter 2.2 presents the definitions of the three study parameters for this research: the service time, the inter arrival time and the berth occupancy. Based on these definitions of the study parameters a well founded approach can be designed and implemented for the AIS tool, following in chapter 3.

2.1. Theoretical framework

In order to define the required study parameters it is important to outline and elaborate on the current port processes a vessel follows when approaching and leaving the terminal. Furthermore, literature is discussed which focuses on queuing theory using Kendall's notation, as this is an important input for terminal- planning and design.

2.1.1. Port processes

The port processes are briefly discussed in which the key study parameters, inter arrival-, service times and occupancy, are introduced.

Inter arrival times

The process in the port starts for a vessel when it requires to access a port. Ports have various regulations when it comes to entering a port at a specific time. Entrance windows vary between ports depending on the (current) level of congesting, the tidal variations of the port (leading to draft restrictions) and the operating hours of the port (Kim et al., 2016). When a vessel approaches a port it might slow down its speed if that is favorable for the estimated arrival time. The *Vessel Traffic Service* (*VTS*) of the port is responsible for providing information regarding the berth availability and weather conditions (Bellsolà Olba et al., 2014). The inter arrival time is defined as the time between two successive arrivals at the port (PIANC WG121, 2014). The processes following the arrival will be specified using figure 2.1 in which the different stages are visualised of a vessel entering and leaving the port.

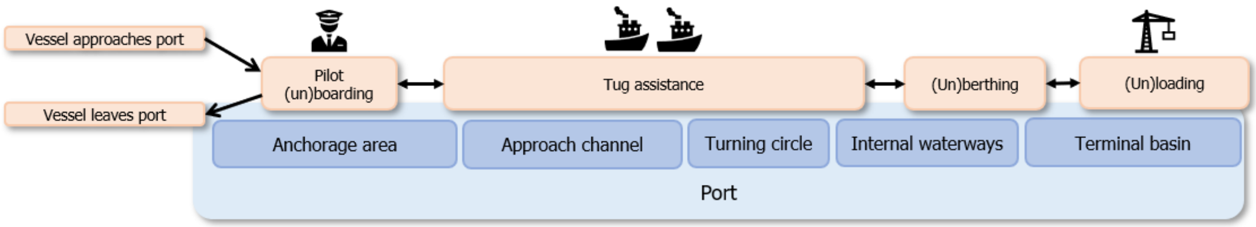


Figure 2.1: Port processes

A vessel will drop anchor in the designated anchorage area, when the berth is not available. The waiting time is defined as the time the vessel is present in the anchorage area. Among others, a few factors can influence the waiting time in the anchorage area: berth availability, tugboats not available, pilots not available, bad weather conditions (high waves or tide limited time windows). Furthermore, the berth availability also depends on the expected service time, the availability of the equipment onshore and requirements for other vessels that will arrive shortly after, or are already at the quay (Meijer, 2017; Günther & Kim, 2006).

Depending on the type and the size of the vessel, the weather conditions, the permit of the captain and specific regulations of the port, a pilot is required or not (Mašović, 2019). If a pilot is necessary, it will sail towards the ship, and in extreme weather conditions board the ship by helicopter. The pilot will assist the vessels captain in entering (and leaving) the port to ensure the vessel safely and efficiently reaches the terminal. In many territorial waters a pilot is required, however forms of exemptions are available for instance by acquiring Pilotage Exemption Certificates (PEC) (International Harbour Masters Association, n.d.).

Tug boats are boats that attach to the vessel in order to reduce the risk of contact with port infrastructure or grounding in restricted areas by attaching and assisting the large vessels. The requirements for tug assistance are similar to those of the pilot requirements. Small vessels have relatively good maneuvering capabilities and when sailing in the port they will most likely stop under their own power with a minimal stopping length. Nonetheless, for larger vessels, starting from vessels of approximately 50,000 tonnage or larger, tug assistance is essential (PIANC WG121, 2014). The maneuvering abilities are affected by the vessels hull shape, the mass, the propulsion system, the power, and the rudder system and dimensions. Lack of course control during stopping requires tugs to attach and assist the vessel inside the port (Ligteringen, 2017). Regulations are not only given by the port master but can also be in the form of governmental regulations, such as the Oil Pollution Act of 1990 (OPA 90) which requires tug assistance for all vessels containing hazardous cargo when maneuvering in ports (Carral Couce et al., 2015).

Tugs will try and make fast as quickly as possible, after the vessel passes the port entrance and enters protected waters. When no effective tug boat control is established yet, the vessels should maintain a minimum speed of around 4 knots (relative to the water) in order to have enough rudder control (Ligteringen, 2017). On average tug fastening takes about 5 to 20 minutes, but now and then circumstances are disadvantageous, when the waves are too high or the vessel moves too fast, and the tugs will not be able to attach maintaining safe operations. The limit speed of vessels is often around 5 to 6 knots and the maximum significant wave height is around a height of 1.5 to 3 meters, dependent on the capabilities of the tugs and its crew (PIANC WG121, 2014).

When the pilot and tugs are attached the vessel sails in a certain path towards the quay determined by the size and complexity of the port, divided into the turning circles, inner channels and crossings. Each of these sections has their own specific requirements for vessel sailing and maneuvering (Bellsolà Olba et al., 2014). Once the vessel has reached the quay or jetty, the mooring can be done to safely attach the vessel to the berth where pilot, tugboats and lineman work closely together (Mašović, 2019).

Service times

The time the vessel is located at the berth, thus the time taken to serve the vessel (unloading and possible loading process) is defined as the service time (Groenveld, 2001). The (un)loading of course depends on the commodity type the vessel is carrying. For container terminals for example, the lashings are immediately taken off once the vessel is safely moored at the berth. Prior to arrival the containers to be unloaded are identified and the containers to be transferred are stacked and arranged at the terminal in the right order. For container unloading the ship-to-shore gantry cranes are used, which can be as high as a cathedral. These cranes play a vital role in the commercial success of the container terminal (Ligteringen, 2017).

For vessel operators the productivity is measured based on service times at terminals. However, at a national level the productivity is much more focused on the amount of cargo throughput over the terminal for a certain time period. The service time is variable and dependent on the amount of cargo (un)loaded and the capacity of the terminals (on)shore facilities (El-Naggar, 2010).

Moreover, Ducruet, Itoh, and Merk found in their research that for container terminals, larger ports are on average more time efficient, based on multiple regression analyses. Larger ports were defined as ports with a higher number of total DWT of all vessel calls at the port. This higher time efficiency is based on the abilities of offering more modern terminal handling equipment and the higher sole traffic size (Ducruet et al., 2014). With regards to the the differences in the service time between container and dry bulk terminals, the dry bulk terminals are expected to require (on average) longer service times. The (un)loading processes of dry bulk vessels are quite complicated as these are specialized vessels that are often designed for specific types of cargo. Therefore, a number of crucial measures and considerations are necessary to ensure suitable and safe (un)loading of the vessel at the terminal (Abdul Rahman, Othman, Sanusi, MD Arof, & Ismail, 2019).

The service time at LNG Terminals is dependent on the type of process required by the vessel, and the possibilities of the terminal. For example, LNG terminals in North West Europe contain a large variety of options at terminals: unloading, storing an re-gassing an LNG vessel, loading the vessel, transshipping of storing the vessels LNG (*LNG Terminalling*, 2019). Often terminals maintain a certain allowable lay time, which describes the allowable time for (un)loading operations. This time depends on, among others, the vessel size, the equipment and technologies at the terminal and the nature of the operations. Regularly this time is limited between 12 and 60 hours (Council of European Energy Regulators, 2017). Thus, service times for these terminals are expected to be in this range.

Occupancy

Berth occupancy is the time that the berth is physically occupied by a vessel, relative to the total number of operating hours of the terminal. Operating hours vary between ports but it is common for ports to operate 365 days per year and 24 hours per day (Zamanirad et al., 2017). The occupancy can therefore be determined by using the service times. However, when designing a terminal an estimation for the occupancy is necessary in order to use queuing theory to determine the quay length. Based on the stochastic character of the arrivals at terminals a fully occupied berth can never exist without there being a continuous queue of vessels. In other words, 100% occupancy is not a realistic design guideline. A lower occupancy results in more flexibility for the terminal, and less waiting times for the vessels (PIANC WG 135, 2014). The occupancy thus increases for higher number of berths.

For container terminals the berth occupancy has been seen to lie in between 35 and 70% based on Thoresen's Port Designer's Handbook (Thoresen, 2003). Moreover, it is often recommended to apply occupancy factors of around 35% for container terminals due to the stringent conditions set by the shipping lines, to ensure the minimum waiting time. Whilst for general cargo terminals a much higher occupancy is suggested of 70%, due to these vessels more easily accepting longer waiting times (Ligteringen, 2017). Finally, liquid bulk terminals are often designed using an optimal berth occupancy of roughly 50 to 65% (Kox, 2016).

Permission is required for a vessel to depart. After the permission is granted the pilot comes onboard and the tugs are attached to the vessel. Once the pilot does not need the tug assistance anymore the tugs will detach and after passing the port entrance the pilot will leave the vessel as well (Bellsolà Olba et al., 2014). It must be taken into account that vessels occasionally visit multiple terminals in the same port, transport between the terminals will most likely require tug and pilot assistance as well.

It is important to note that these processes are sensitive to various external influences, which can vary the durations and possible sequence of these events. As mentioned in subchapter 3.1.1 physical environments can influence the processes at a port, such as varying climate conditions for different port locations. Moreover, the economic situations of a country and for example various types port regulations, can influence these port processes explained.

2.1.2. Port and terminal design

PIANC Guidelines

When designing a port, engineers and port planners often use guidelines to ensure safe and efficient design. The *Permanent International Commission for Navigation Congresses* (*PIANC*), founded in 1885, is an organization which has been investing in the expert guidance and technical advice for waterborne transport infrastructure. Working groups and commissions write technical reports which can be used for harbour engineering (PIANC, n.d.; Jianghao & Degong, 2018). The following three PIANC reports have been a substantial resource for this thesis research:

- WG158 Masterplans for the development of existing ports (2014)
- WG135 Design Principles for Small and Medium Marine Container Terminals (2014)
- WG121 Harbour Approach Channels – Design Guidelines (Updated Version 2014)

Capacity and simulation models

The different capacities of the separate parts of the port are discussed, starting with the capacity of the approach channel. Over time a port can become congested and it will reach a point where the waiting times are too long and the decision for a two-way channel has to be taken into account. There are no precise guidelines for the determination of the maximum capacity, however there are multiple logistical simulations possible to determine the capacity. In the last 50 years a growing trend can be seen in use of these traffic simulation models for the design and planning of ports (Dragović et al., 2017). PORTSIM, FLEXSIM, SimPort, ARENA and HarbourSim are a few examples of simulation models used for port planning. The capacity of the waterway is the maximum traffic volume which can be handled by the approach channel, while still meeting the service and safety requirements.

The estimation of the capacity is not an easy matter, acceptable waiting times are minimal when set by vessels and will vary between different vessel types. The waiting times can be estimated by queuing theory when the port is defined as a simple service system. The part of the approach system with the lowest capacity will define the capacity of the entire system. It is common for the approach channel to end in a turning circle, where often only one vessel is allowed at a time. The capacity in the turning circle is therefore lowest leading to the limiting factor of the system (PIANC WG121, 2014).

In 2016 Branislav Dragović did a literature review on all the possible simulation modelling in ports. The most popular approach for container terminals is the discrete-event simulation even though more and more new techniques arise, such as network based modelling, simulation-based education, agent based modelling and so on (Dragović et al., 2017). In these simulation models a vital piece of input is the choice for the vessel arrival distribution. A theoretical distribution can be chosen, for example one of the distributions as mentioned in Sub-Chapter 2.1.3. Nevertheless, if historical data is available, it will be the best way to determine the stochasticity of the arrival pattern (Bellsolà Olba et al., 2018; Zhou et al., 2017). Xavier Bellsolà Olba found by using the large AIS data availability that historical data analysis leads to the best estimation of the vessel demand. For new ports which do not yet have the AIS data, similar ports can be extrapolated to make the estimations. Particular peculiarities should not be forgotten, for example influences by seasonality (Bellsolà Olba et al., 2018).

Design quay length

The focus in this research is on the length of the quay wall, neglecting other design parameters such as the strength and stability of the quay wall. In designing the quay length it is important to obtain an assumption for the *LOA*. Once the LOA is chosen, the simplistic method to calculate the quay length for more than 1 berth is given by (Ligteringen, 2017):

$$L_{n>1} = 1.1 * n * (L_{OA,Average} + 15) + 15 \tag{2.1}$$

Where n represents the number of berths.This calculation includes a 15 meter berthing gap between the vessels moored at the quay, plus an additional 15 meter length at the outer side of the berth. *UNCTAD* performed a study which researched the probability of excess waiting time as a result of simultaneous mooring of multiple above-average vessels, based on the above definition of design quay length. Once the berth length is equal to 110% of the average berth length plus gap, there is no additional waiting time. Hence this has resulted in a design factor of 1.1 (Ligteringen, 2017).

2.1.3. Queuing theory

When designing the terminal the quay length is an important variable. Equation 2.1 is an example of how the terminal quay length can be calculated. An important factor for determining the quay length is the number of berths. This number of berths is subsequently dependent on the maximum allowable waiting times which can be determined using queuing theory.

In figure 2.2 a visualisation of the process of using queuing theory in order to determine the number of berths is given. Queuing theory uses the three input parameters to anticipate the waiting time in terms of service time, for different combinations of number of berths and occupancy. Using a chosen limit for the maximum allowable waiting time (in units of service time) the number of berths based on the expected occupancy can be selected (Ligteringen, 2017). Information regarding the input parameters will follow an overall introduction to queuing theory.

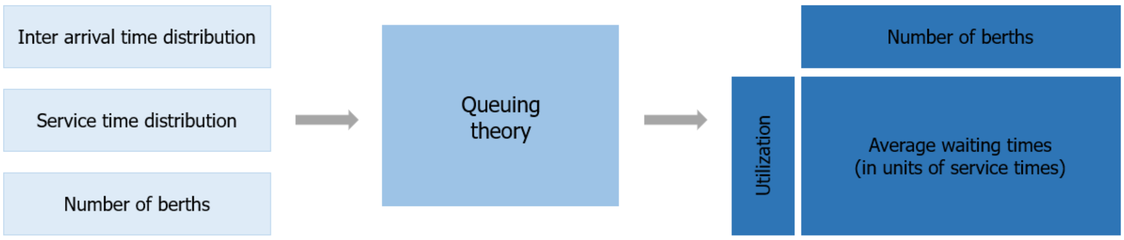


Figure 2.2: Queuing theory process

Queuing theory represents a model to assess different behaviors of waiting lines. The paper published in 1909 *The Theory of Probabilities and Telephone Conversations* by A. K. Erlang described the number of telephone circuits needed for a phone service, without clients waiting too long for an available phone line (Erlang, 1909). The model returned an optimum number of circuits and operators to handle the expected number of phone calls. The theory Erlang developed is applicable to much more than just phone lines (Berry, 2006). From a business perspective, queuing theory advises the development of efficient and cost-optimal workflow systems. If a system is not able to deal with over-capacity negative outcomes will arise (*Queuing theory: Definition, history & real-life applications*, 2020).

Queuing theory inspects the total system of waiting in line. The arrival rate, the number of arrivals, the number of servers, the queuing discipline, the average service time and the capacity of the waiting area are elements included in the queuing theory approach (*Queuing theory: Definition, history & real-life applications*, 2020). A standard notation to describe queuing theory was first proposed by David George Kendall in 1953 (Kendall, 1953). His notation covers a large range of various queuing situations and can be represented as A/S/c/K/N/D where (*Queuing theory: Definition, history & real-life applications*, 2020):

- A = inter arrival time distribution
- S = service time distribution
- c = number of servers
- K = capacity of the queue (neglected if unlimited)
- N = amount of possible customers (neglected if unlimited)
- D = queue discipline (default: first-in-first-out)

For this research the inter arrival time distribution is defined as the distribution of the time intervals between successive arrivals of vessels. The service time distribution represents the time vessels stay at the terminal and the number of servers is the number of berths at the terminal (Groenveld, 2001). The assumption is made that there are no specific priority rules and vessels are served in order of arrival, thus the vessels will attend the terminal one by one and the basic model with 3 input characters is sufficient (A/S/c). Furthermore, the capacity of the queue is unlimited which means that when vessels arrive at a location where a long queue is already formed, the vessel doesn't leave but joins the waiting line (El-Naggar, 2010).

Often the queuing theory is used in terminal design when selecting an optimal amount of berths. It must be noted that in order to use queuing theory the port systems should be schematised with only simple facilities and no variables such as weather influences can be included (Bellsolà Olba et al., 2014). The result of queuing theory must only be used as a guidance for terminal planning, because the theory does not include the impact of the variety of vessels approaching the service systems, or any of the dynamics between these vessels (Ho & Bateman, 2013).

For different combinations of inter arrival (A) and service times (S) different outcomes (visualised in tables) have been generated, as shown in the lecture notes of R. Groenveld *Service Systems in Ports and Inland Waterways*, (Groenveld, 2001). An example of a table with a M/E2/n distribution combination is given in table 2.1, in which the inter arrival time is represented by a Markov process (exponential distribution) and the service time is represented by an Erlang-2 distribution. The utilization (u) is the occupancy divided by the the number of berths. These tables can also be generated using the OpenQTSim python package, which is available on GitHub (Van Koningsveld & Uijl, 2020).

Utilisation (u)	Number of berths (n)								
u = p / n	1	2	3	4	5	6	7	8	...
0.10	0.08	0.01	0.00	0.00	0.00	0.00	0.00	0.00	...
0.20	0.19	0.03	0.01	0.00	0.00	0.00	0.00	0.00	...
0.30	0.32	0.08	0.03	0.01	0.00	0.00	0.00	0.00	...
0.40	0.50	0.15	0.06	0.03	0.02	0.01	0.01	0.00	...
0.50	0.75	0.26	0.12	0.07	0.04	0.03	0.02	0.01	...
0.60	1.13	0.43	0.23	0.14	0.09	0.06	0.05	0.03	...
0.70	1.75	0.73	0.42	0.27	0.19	0.14	0.11	0.09	...
0.80	3.00	1.34	0.82	0.57	0.42	0.33	0.27	0.22	...
0.90	6.75	3.14	2.01	1.45	1.12	0.91	0.76	0.65	...

Table 2.1: Average waiting time of ships in queue M/E2/n (in units of average service time) (Table IV, (Groenveld, 2001))

Distributions

In the selected Kendall notation (A,S,c) two inputs represent a theoretical distribution. Besides being an important input for the theoretical queuing theory models, the inter arrival and service distributions are used in simulations of port capacity analyses (Kuo et al., 2006). In many traffic simulation models for ports the impact of the arrival patterns of vessels tend to be underestimated (Van Asperen et al., 2003).

One of the most common distributions is the *Negative exponential distribution* (*NED*). The probability density function of this NED is given by:

$$f(t) = \lambda * e^{-\lambda t} \tag{2.2}$$

Where λ is the *mean* positive value, often called the 'rate parameter' since it describes the parameter of the distribution. t is a random parameter and $f(t)$ is the *Probability density function* (*PDF*), the PDF of the NED follows the line as shown in Figure 2.4 by line *k = 1*. The letter **M** (from Markovian, representing complete randomness) represents this completely random arrival distribution (Adan & Resing, 2015). The exponential distribution is distributed as the time between events in a Poisson process. The Poisson distribution is discrete process, representing the number of occurrences per time interval (Cooper, 2005).

The General distribution, represented by **G**, can be used when no assumption is made and the distribution can thus take on any form. Another distribution is the deterministic distribution, represented by the letter **D** which can be used to define the service distribution for a fixed amount of service time. This distribution is visualised by using the cumulative density function (CDF), as shown in figure 2.3.

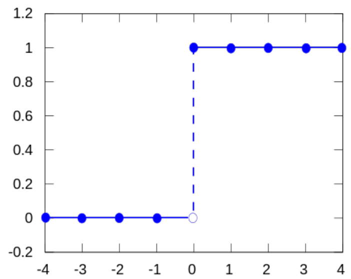


Figure 2.3: Cdf of deterministic distribution (source: https://en.wikipedia.org/wiki/Degenerate_distribution, accessed 7-3-2020)

Another distribution is the Normal distribution is often used for defining measurements using the following PDF (*Pre 1.5 Normal distribution*, n.d.):

$$f_y(y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp(-\frac{(y-\overline{y})^2}{2\sigma_y^2}) \tag{2.3}$$

Where $\overline{y}$ represent the mean and σ_y^2 the variance of the variable. The Normal distribution is always symmetric around it's mean. The Gamma distribution is given by (*1.3.6.6.11 Gamma Distribution*, n.d.):

$$f(x) = \frac{(\frac{x-\mu}{\beta})^{\gamma-1} exp(-\frac{x-\mu}{\beta})}{\beta\Gamma(\gamma)} \quad x \geq \mu; \beta > 0 \tag{2.4}$$

Where γ is the shape parameter, β is the scale parameter, μ is the location parameter, and Γ is the Gamma function defined as:

$$\Gamma(a) = \int_0^{\inf} t^{a-1} e^{-t} dt \tag{2.5}$$

Finally, the Beta distribution is the normalized constant distribution of Gamma distribution given by (*Chapter 8 Beta and Gamma*, n.d.):

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)+\Gamma(b)} x^{a-1}(1-x)^{b-1} \tag{2.6}$$

The Beta distribution can be seen as a continuous probability based on two parameters. The Gamma distribution is expected to better represent the distributions for this research, due to the Beta distribution focusing on modelling the uncertainty of a certain probability of success for an experiment (Taboga, 2017).

A more general distribution is the Erlang-k distribution, represented by the letter **E-k**. The Erlang-k distribution is made up of k negative exponential distributions and requires two input parameters μ and k. The Erlang distributions are entirely theoretical curves (El-Naggar, 2010). The Erlang distribution is a specific case of the Gamma distribution, when the shape parameter in the Gamma distribution is an integer. The PDF of the Erlang distribution is formulated by:

$$f(t) = \frac{(k*\mu)^k * t_{k-1} * e^{-k*\mu*t}}{(k-1)!} \tag{2.7}$$

Figure 2.4 shows the Erlang-k distributions for a mean of 1 and different numbers of k which represent a positive integer (the shape factor) and μ which represents the scale factor equal to $1/\lambda$, where λ is the rate factor: a positive real number. When the number (k) increases the distribution becomes more constant, as illustrated in the Figure 2.4. Thus, based on these Erlang distributions a process can be described as an entirely random process (k = 1) or a purely constant process (k = ∞) (El-Naggar, 2010).

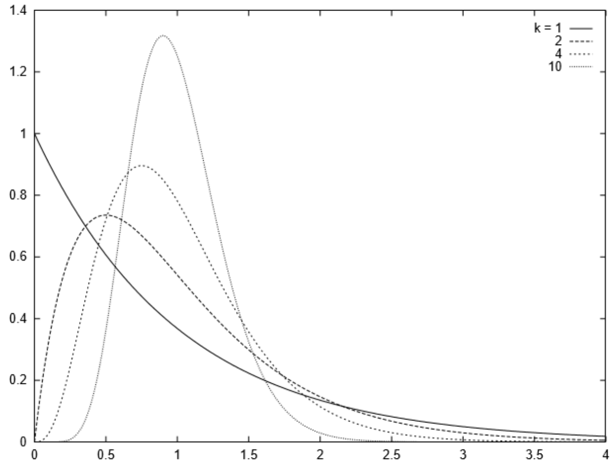


Figure 2.4: Erlang distribution (source: Queuing systems *Ivo Adan* figure 2.2)

When the shape parameter is equal to one, the Erlang distribution reduces to an exponential distribution (k = 1 in figure 2.4). In other words, the Gamma distribution with the shape parameter as one, is an exponential distribution (*scipy.stats.gamma*, 2019). Furthermore, the Beta distribution is actually an uniform distribution when both shape parameters are one (Taboga, 2017).

Inter arrival time distribution

Literature suggests that very often the *NED* represents the inter arrival times when the arrivals are completely stochastic and independent of each other (Groenveld, 2001; PIANC WG121, 2014; UNCTAD, 1985). However, for a terminal that has a shipping line with a regular service the inter arrival time often can be distributed by the Erlang-k distribution (Bellsolà Olba et al., 2018). The distribution coefficient k tends to decrease as the scale of the system grows (Kuo et al., 2006). Finally, *UNCTAD* states that for specialised terminals the distribution will often follow an Erlang-2 distribution (UNCTAD, 1985).

Furthermore, *UNCTAD* assumes that for break-bulk cargo terminals the Erlang-1 inter arrival distribution (thus NED) is assumed and for dry bulk cargo terminals (where there is some tendency towards scheduling) the smoother Erlang-2 distribution is assumed. These above mentioned distributions might be too conservative for the container vessels due to the improved scheduling (PIANC WG158, 2014). Thus for a terminal where shipping lines visit with scheduled arrivals the NED or Erlang distribution might not be right. Nonetheless, even when shipping lines visit in a scheduled manner, the arrival times are dependent on external factors such as weather influences or engine failures, which still return that a stochastic character of arrival would suit the arrival pattern best (Bellsolà Olba et al., 2018). Besides these external factors, Pachakis and Kiremidjian declared that the superposition of multiple container lines, which each could have a uniform arrival rate, still returns roughly a Poisson arrival distribution process.

For container terminals an extensive amount of (theoretical) research has been performed regarding the arrival processes of vessel at ports. Kuo et al. performed a literature review in 2006 into the vessel arrival distribution literature, from which he concluded that when the terminal and port are noted as one system, the inter arrival time follows an Erlang-1 distribution, thus NED.

Furthermore, van Vianen concluded in his literature review for his PhD thesis that most literature recommends using the NED for vessel inter arrival time distributions. Figure 2.5 gives an overview of all proposed inter arrival time distributions (van Vianen, 2015).

IATDist	Reference	n_s [-]	Cargo[1]	IAT Dist	Reference	n_s [-]	Cargo[1]
Weibull	Tengku-Adnan et al. (2009)	408	DB	NED	Kia et al. (2002)	372	C
	Tahar and Hussain (2000)	-	C		Demirci (2003)	297	C
Erlang-2	UNCTAD (1985)	-	DB		Pachakis and Kiremidjian (2003)	142	C
Erlang-k	Kuo et al. (2006)[2]	7,729	C				
NED	UNCTAD (1985)	-	DB		Van Asperen et al. (2003)	-	LB
	Radmilovich (1992)	-	-		Dragovic et al. (2006)	711	C
	Kozan (1997)	679	C		Bugaric and Petrovic (2007)	-	DB
	Shabayek and Yeung (2002)	12,610	C		Legato and Mazza (2013)	1030	C

[1] Where C stands for containers, DB for dry bulk and LB for liquid bulk.
[2] Kuo et al. (2006) discovered for the arriving of container vessels in the port of Kaohsiung that the ship interarrival time distribution followed an Erlang-k distribution. The distribution coefficient (k) tends to decrease as the system's scale grows. The interarrival time at the public container terminal appears to be more scattered than at dedicated container terminals.

Figure 2.5: An overview of inter arrival time distributions proposed by literature in *Van Vianen* PhD's thesis (table 3.4)

Where n_s represents the number of vessels in the data set used. Remarkable is the proposed Weibull distribution. The Weibull distribution, is mostly used for reliability applications of by testing material strengths. However, it has been used for a large amount of other applications such as the wind-speed analysis, but also survival data analysis (Lai et al., 2006). Different forms of the Weibull distribution exist but the most common is(Lai et al., 2006):

$$F(t) = 1 - exp\Big[-(\frac{t-\tau}{\alpha})^{\beta}\Big],\ t \geq \tau \tag{2.8}$$

Where β is the shape parameter, τ is the location parameter and α represents the scale parameter.

Service time distribution

The service times often obtains a much more regular distribution compared to the inter arrival times. However, the time taken to unload and load vessels varies considerably on the type of vessel, the quantity and type of cargo and the rate at which the cargo is handled. Therefore, often these distributions follow the Erlang-k distribution (UNCTAD, 1985). The service time, as earlier defined (subchapter 2.1.1), is dependent on a number of different stages such as the (un)mooring and the (un)loading. Therefore, PIANC expects the Erlang-k distribution to be a natural choice since it contains the characteristics of the multiple phases (built up of k phases) (PIANC WG121, 2014).

For terminals with a more constant service time a low Erlang distribution (k= 1-4) will lead to higher estimates for the queuing time than expected. However, this seems to rarely be the case based on data available to UNCTAD in 1985 (UNCTAD, 1985). Only for certain bulk terminals, with vertical operations, high Erlang numbers have been used, such as k = 8.

Furthermore, van Vianen concluded in his literature review for his PhD thesis, besides assessing the inter arrival distributions, that most literature recommends an Erlang-k distribution for the service time. Figure 2.6 gives an overview of all proposed service time distributions, where W_s represents the *ship service time*.

W_sDist	Reference	n_s [-]	Cargo[1]	W_sDist	Reference	n_s [-]	Cargo[1]
Normal	Tahar and Hussain (2000)	150	C	Erlang-k	Shabayek and Yeung (2002) [k:117]	12,610	C
	Bugaric and Petrovic (2007)	-	DB		Kozan (1997) [k:4]	679	C
NED	Radmilovich (1992)	-	-		Kia et al. (2002) [k:4]	372	C
	Demirci (2003)	297	C		Altiok (2000) [k:4]	248	DB
Beta	Legato and Mazza (2013)	1,030	C		Dragovic et al. (2006) [k: 3,7,12]	711	C
Gamma	Jagerman and Altiok (2003)	304	DB		UNCTAD (1985) [k:2]	-	DB

[1] Where C stands for containers and DB for dry bulk.

Figure 2.6: An overview of service time distributions proposed by literature in *Van Vianen* PhD's thesis (table 3.6)

Occupancy

In a single service system (where the third part of the Kendall notation is equal to one) the occupation can be easily defined. The terminal can tend to one vessel at a time and the occupation(ρ) is defined by multiplying the arrival rate (λ) by the mean service time. The mean service time is equal to 1 divided by the average service rate(μ), returning an occupancy formula:

$$\rho = \frac{\lambda}{\mu} \tag{2.9}$$

The more berths the terminal will have, the lower the waiting time will be, as the chance of all berths being occupied decreases (PIANC WG 135, 2014). The only way to decrease the waiting time and increase the berth occupancy, without increasing the number of berths, is by persuading vessel line operators to schedule more regularly (UNCTAD, 1985).

The occupancy of the berth is often referred to the berth utilization, for single service systems these two are the same. In order to control the waiting times a multi service system can be designed where more than 1 berth will be able to handle vessels. The utilization(Ψ) then becomes equal to the occupancy divided by the number of berths (n):

$$\Psi = \frac{\rho}{n} \tag{2.10}$$

The theoretical ultimate capacity is the maximum that could be handled by the terminal, as if the berth is 100% of the time occupied by vessels. When designing the terminals, typical occupancy values are based on fulfilling average ratios between the waiting times and service times, depending on the type of terminal and number of berths. For bulk terminals this average ratio 0.3, for general cargo it is 0.2 and for container terminals this is 0.1, as the container vessels are much less patient (PIANC WG 184, 2019; PIANC WG158, 2014).

Waiting times

The three main influences on the waiting time for vessels at the port are: the arrival rate (λ), the service rate (μ) and the number of berths available (n). As mentioned, terminal design is influenced by the service times in terms of the maximum allowable waiting times. In order to get an idea of these acceptable waiting times *PIANC WG 121* gives some rough estimates (where the values represent the average waiting time in percent of the service time) (PIANC WG121, 2014):

- Container vessels: 5-10%
- Gas carriers: 10%
- General cargo: 30%

- Liquid bulk carriers: 30%
- Ore carriers: > 40%

2.1.4. Distribution fitting

In order to compare the reality with certain theoretical distributions the AIS data can be fitted to various distributions. Distribution fitting is defined as the process where the best statistical distribution is found to fit a certain data set. Multiple techniques are available which estimate the parameters of certain distributions which define the distribution. The possible parameters are the scale, shape, location and threshold factors, however not all distributions consist of all of these parameters (McNeese, 2016).

Comparing the *Probability density function* of a certain distribution with a histogram of the data is the best way to find the most optimal parameters for that distribution. A histogram is made up of intervals, known as bins, which contain the sum of data in that certain interval. It must be noted that the number of bins should be large enough to visualise all important peaks and the size of every bin should also be large enough to contain multiple data points. Small data sets are therefore not very suitable to be displaced in a histogram (Gast, 2013).

In order to test whether a fitted statistical PDF is a good match for a data set a goodness-of-fit test can be performed. A goodness-of-fit test compares the PDF or *Cumulative distribution function* to the data using a fit statistic or discrepancy measure (Royle & Kéry, 2016). According to the literature review, the most appropriate approaches are the *Kolmogorov-Smirnov* (*K-S*) test and the Chi-Square test (Kuo et al., 2006). Both goodness-of-fit tests will be further explained.

Kolmogorov-Smirnov Test

The *K-S* test compares how well the theoretical fitted CDF compares to the actual CDF of the data set. The test calculates how well the two distributions compare by using the p-value which is the significant limit which represents the probability of the Null Hypothesis being accepted. In this case the null hypothesis is a hypothesis that says there is no statistical significance between the two variables (Rakshit, 2018). Once the p-value is below a certain limit, often 0.05, the null hypothesis can be rejected. In other words, when the p value is below the limit, there is no fit with the distribution. When the p value is above the limit, the fitted distribution can be assumed to not diver greatly from the observed data distribution. It is however often possible for large data sets to not have any theoretical fit (Wicklin, 2019).

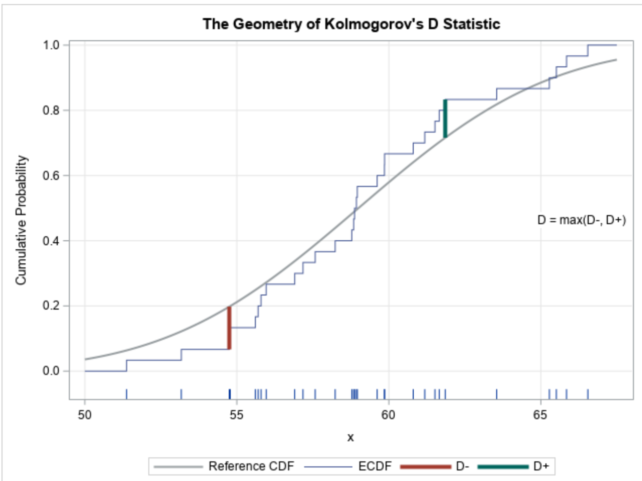


Figure 2.7: D statistic based on Kolmogorov-Smirnov Test (Wicklin, 2019)

The test can be applied in a renowned Python-package *scipy-stats* (*scipy.stats.kstest*, 2019). The resulting p-statistic from the K-S test should be greater than 0.05 in order to suggest that the theoretical fitted distribution is not that different compared to the data distribution (Allen, 2018). The second value returned represents the D statistic. The D statistic, as shown in figure 2.7, is the maximum absolute difference between the two CDFs. It is calculated as the maximum of either D- or D+ (Wicklin, 2019).

Chi-square test

The Chi-Square test is the sum of the squared error for every bin of the CDF, as shown in equation 2.11 (Bolboacă et al., 2011), where O represents the observed data, E the predicted data and f the mutually exclusive classes. It measures the statistical difference between the expected and observed frequency. The p value of the Chi-square test is based on the whether or not two distributions (CDFs) are similar and how independent they are of each other (Okada, 2020). Once the p value exceeds a certain limit (5%), the Null Hypothesis of the two distributions being the same is rejected, thus there is a similarity between the two distributions (Hamel, 2019).

$$\chi^2 = \sum_{i=1}^{f} \frac{(O_i - E_i)^2}{E_i} \tag{2.11}$$

Again, it must be noted that when large data sets are available the possibilities of no theoretical distributions fitting the data, are higher. Using the Python-package scipy-stats the Chi-Square test can be performed (default the degrees of freedom as f - 1).

2.2. Definition of study parameters

To start off, a distinction is made between a single service and multi service system. The most simple type is a terminal with one berth: a single server system. For multi service system terminals there are multiple berths at which vessels can berth. Furthermore, a distinction is made between two types of multi service system terminals: a system where the number of berths is constant versus a system where the number of berths is varying over time.

2.2.1. Service times

The service time is defined as the time the vessel spends at the berth, including berthing, mooring, unloading, loading and unmooring the vessel at the terminal (PIANC WG 135, 2014), as explained in subchapter 2.1.1. Loading related affairs such as documentation checks and inspections are also included and during unloading and loading possible downtime can arise due to extreme weather conditions or maintenance of terminal equipment (Zamanirad et al., 2017).

Using AIS data with a predefined terminal location the service time is determined as the time a vessel is physically located (almost) directly next to the quay wall. A polygon will be drawn over the terminal location. It is difficult to determine exactly when the berthing starts, as it will vary between terminals, and it is uncertain where the antenna is located on the vessel. In this research the service time can be defined as the time a vessel is present in the polygon, thus the difference in time between entering and exiting the polygon. The terminal polygon should be drawn over the port layout, roughly three times as wide as the vessel width to include the (un)berthing times.

It must be noted that some vessels present in the polygon will not berth. These vessels should not be included in the service time calculations. Lastly, the service time approach will not differ between the different types of terminals. For example, for multi service terminals the service time will still regard each single vessel and thus is not expected to depend on the number of berths.

2.2.2. Inter arrival times

The inter arrival time is the time between two successive arrivals of vessels at a port (PIANC WG121, 2014), as mentioned in subchapter 2.1.1. Once the port area is defined, the inter arrival time can be calculated using the AIS data as the time between the moment a vessel enters the port area and the moment that a successive vessel enters this same area. The port area is defined as a polygon including all terminals and port infrastructure, plus the possible anchorage areas. For multi service terminals, a larger number of berths is expected to return a more frequent inter arrival time at the port, as it is able to handle more vessels (due to the larger number of berths).

2.2.3. Berth occupancy

The **berth occupancy** is defined as the time the berth is physically occupied by a vessel (a.k.a. the service time) relative to the total time that is available, as described in subchapter 2.1.1. This must not be confused with **berth commitment** time which is an expansion of this concept, where the time is included where the berth remains unavailable without there actually being a vessel physically present at the berth. This additional time could be due to transit time of vessels sailing towards and away from the berth, or downtime by maintenance or reparations of equipment and infrastructure at the terminal. When designing a terminal it is important to take note of these two different occupancy definitions. For example, for ports with a very long uni-directional access channel an average berth occupancy would lead to overestimating capacities, where the berth commitment would not neglect the downtime due to the long approach channel (PIANC WG 184, 2019).

Once the service time is defined, the berth occupancy can be defined as the service time divided by the total time that is available to serve the vessels, also known as the operating time. As mentioned in chapter 2.1.1 it is common for ports to operate 365 days per year 24 hours per day. The default is therefore set to 365 days - 24 hrs/day, but this can be manually adjusted. Possible downtime of a terminal due to infrastructure failure or weather conditions will not be filtered out and thus the service times and thus the berth occupancy will include these external factors.

For terminals with multiple berths it is important to make a distinction between the berths in order to determine the total berth occupancy at the terminal. When the number of berths is constant, the berth occupancy can be defined. However, often the number of berths varies over time and this method is no longer valid. A way to define the terminal performance in terms of occupancy is to use the vessel lengths with regards to the total length of terminal available. This comes with a lot of uncertainties and substantiated conclusions can not be drawn from this length occupancy.

Container terminals often consist of a long quay wall where a varying number of vessels can berth at the same time, since the shore-side equipment can manoeuvre alongside the quay. On the contrary, liquid bulk terminals consist of a fixed number of berths where the (un)loading takes place at a central manifold location. For dry bulk terminals a distinction can not as easily be made. Often for export dry bulk terminals the loading of vessels takes place using conveyor belts which results in these terminals being more similar to jetty lay-outs. However, for import terminals the onshore equipment consists of cranes unloading the vessels, thus these terminal designs are more similar to container terminals (Ligteringen, 2017).

Every port is different and thus it is difficult to determine the berth commitment from AIS if no other port layout information is known. Thus the berth commitment time falls outside of the scope of this research. The berth occupancy for different types of ports can not be compared without taken port layout factors into account. Similar types of terminals can be compared to each other. For example, it is expected that complex ports and ports with long approach channels will therefore have a smaller berth occupancy.

3
Available data and methodology

For this research the materials consist of all available data sets. In this chapter these different data sources are introduced. Specifically, the AIS data set is thoroughly discussed. Furthermore, the AIS tool set up is presented and discussed.

3.1. Available data sources

The research objective focuses solely on the possibilities of AIS data. However, for some analyses and for certain design steps of the tool different data sources are consulted. First, the AIS data will be thoroughly introduced, where-after the other useful data sources are inspected.

3.1.1. AIS data

In this subchapter AIS is introduced, its capabilities and limitations are considered and a literature review on AIS research is performed.

What is AIS?

The International Maritime Organization (IMO) is a specialized agency by the United Nations that is responsible for the safety and security of maritime traffic and prevention of pollution (marine and atmospheric) by vessels (International Maritime Organization, 2020). In December 2000 the IMO decided that sea vessels were required to implement Automatic Identification Systems (AIS) to increase efficiency and safety at sea. Since 2004 all sea and passenger vessels larger than 300 gross tonnage were required to use AIS. Besides enhancing safety and efficiency at sea, AIS also has been seen to improve the waterway management and vessel traffic surveillance (VTS) (de Boer, 2010). Regulations have been sharpened ever since and starting from December 2014, besides that every sea vessel (even below the gross tonnage limit) uses AIS, now for all inland vessels longer than 20 meters the use of AIS is required as well (Zhou et al., 2017).

Vessels using with AIS transmitters send out static and dynamic data about the vessel at different intervals. The transmitters broadcast this data over Very High Frequency (VHF) radio waves, where base stations along coastlines and inland waterways can receive these radio waves and therefore track the vessels (Zhang et al., 2016; Rawson et al., 2014). Besides AIS stations the AIS messages can also be tracked by a few satellites in the earth's orbit. These can be very helpful for vessels at sea which fall out of the reach of base stations at coastlines, which with satellite AIS can be tracked all day and night. The maximum range of these terrestrial AIS base stations is roughly 40 km ($\approx$ 20 nautical miles), however weather conditions, surrounding topography and transceiver location, height and type can influence this maximum range negatively (Tu et al., 2016; Zhao et al., 2014; Robards et al., 2016).

The information from the AIS data can be subdivided in static (broadcasted every six minutes) and dynamic information (broadcasted at varying intervals) (Meijer, 2017; D. Chen et al., 2016; Rawson et al., 2014):

- Static information: *IMO* number, *MMSI* number, Call sign and number, Type of vessel, length, beam, location of antenna on vessel (using GPS)
- Dynamic information: vessel's position with accuracy indication, time (UTC), course over ground (COG), heading over ground (SOG), heading, navigational status, destination, type of cargo, vessel's draught

Optional: rate of turn, angle of heel, pitch and roll, route plan-waypoints, number of persons on board, short text messages with important navigational information

All vessels containing an AIS unit will have a *Global Navigation Satellite System* (*GNSS*), based on the GPS location. The accuracy of the send data will depend on the sensors on board. For the latitude and longitude the accuracy can be up to 1/10,000 minute, equal to 0.18 meter. Due to the GPS having intrinsic behavior, the *IMO* assumes an accuracy of 10 meters for latitude and longitude is representative (Renso et al., 2013; Perez et al., n.d.).

The IMO initially required AIS to be implemented in order to enhance safety and efficiency and to increase situational awareness. It is therefore originally not developed for the purpose of research or intended as an archive or public medium. Several entities, such as INTERTANKO and INTERCARGO, actually desired to limit the public access to the AIS data. In 2004, at the 79th meeting of IMO's safety committee (MSC) concerns were discussed but no specific regulations or rules were made by the IMO and governments have since not restricted the assess to AIS data (Bellsolà Olba et al., 2014; Robards et al., 2016).

Considering that AIS is a worldwide standard and the AIS messages are sent on very regularly bases, it provides a very valuable source of information. For example, it can locate illegal actions, such as pollution, fishing in protected areas and smuggling. Besides these outputs, the AIS data has also seen to improve insights into vessel behavior (Fiorini et al., 2016; de Boer, 2010).

Capacity and limitations of AIS

As the regulations will most likely increase in the future, accelerating amounts of data will become available. Nevertheless, vessels are becoming more and more aware that they are being traced and they might turn of the AIS broadcasting to avoid detection. These new developments, along with the fact that the AIS messages are not developed for research intentions, will limit the possibilities by increasing uncertainties in the data. When using AIS data for research, the data should therefore always be examined very critically before any conclusions can be made (de Boer, 2010; Windward, 2014).

The quality of AIS data is discussed in multiple studies, where the different types of AIS data are checked. For example Sotirov and Alexandrov analysed AIS data for a three year period and found that more than 20% of the data was incorrect. The incorrect data was mostly from the following data parameters: call sign, antenna position, hazardous cargo type, vessels draught and destination (Sotirov & Alexandrov, 2017). Incorrect AIS data can occur based on three different causes. First, the data can contain non-deliberate errors due to incomplete or non-secure transmissions or lack of knowledge onboard. Second, the data can be deliberately falsified, and thirdly, the data can be spoofed, which means data is broadcasted from an outside source as if it is coming from the vessel (Ray et al., 2015).

Many AIS data providers exist online, with either free or paid services. Often providers allow current AIS data (and with for example 2 days back) to be viewed for free, however when more data is required, over longer time periods, this will not be for free. A few web-based data sources of popular AIS providers are given in appendix A.1.

AIS Literature Review

Since 2004 numerous studies have benefited from AIS data. It has been adopted in various types of studies and in this literature research an examination is done on what types of AIS research have been performed so far. 80 research studies have been analysed. The list of studies will not fully cover all the research done, but it will cover most and therefore generate a good picture of what has been done. The research is spread out between 2005 and 2019 as shown in figure A.1 in appendix A.2.2. As mentioned, the number of studies, defined per research type (classified in appendix A.2.1) is given in figure 1.1. A total overview of all research found is given in appendix A.2.3.

Besides the research done using AIS data, there have also been simulation models that have been calibrated using AIS, that give very reliable results. For example, models that predict vessel collisions use extensive AIS data analysis to determine the vessels paths. The simulation results from these models prove the importance of an AIS data analysis and how these analysis inputs can improve simulation models. (Bellsolà Olba et al., 2018).

R. Meijer, I. Parolas and A. Dobrkovic all focused on the prediction of estimated arrival times (ETA) for vessels using AIS data. They found that using only AIS data is sufficient for predicting ETA and that the ETA can be improved using AIS data analysis (Meijer, 2017; Parolas, 2016; Dobrkovic et al., 2015). H. Ni Ni studied the distribution pattern of vessel arrivals for the port in Yangshan, where he used AIS data to analyse the frequency of ship arrivals and see if they fit a Poisson distribution. He concluded that the arrival distribution depends on the size of the group interval and that for smaller samples the Poisson distribution fits the distribution well. However, for larger samples the Poisson distribution should be modified (Ni Ni et al., 2011).

Available AIS data for this research
The AIS data analysed in this research comes from a private data source, maintained by *Royal HaskoningDHV*.

More information about the specific data source of RHDHV is not publicly available.

From the AIS data the following parameters will be used:

- *Maritime Mobile Service Identity* (*MMSI*)
- Longitude
- Latitude
- Timestamp

The *MMSI* is a unique nine digit identification number (Valsamis et al., 2017). Often the AIS data comes from the antenna which is located at the center of the bridge of each vessel. However, it is not certain where the bridge is located with regards to the vessel geometry. This will influence the location of the vessel portrayed in the waterways and at the quay.

External influences that might affect the data are:

- Physical environment: consisting of wind, waves, current, ice and visibility influences.
- Vessel characteristics: vessel capacity, onboard equipment, propulsion and steering capabilities.
- Human factor: behavior of port masters, officers, pilots, tug masters and crew (on board and on land). Port masters have control of when vessels are allowed to enter the port and captains can decide to slow down when they know there is no vacant berth at the port.
- Port operations: general port operations, as well as terminal operations, will have a large impact on the data since it will affect the time at berth (the service time).
- Economic developments: import and export demands and the prosperity of the economy.

Prior research has demonstrated that the parameters of the physical environment: wind, current and visibility and from the vessel characteristics: the vessel size, all have a significant influence on the vessel speed and path. However, for these external influences the impact did not vary between different class sizes (de Boer, 2010). In the analysis of the data, these external influences are taken into account, but will not be extracted from the data, since the data represents the reality in the most optimal way possible.

3.1.2. Other data sources
More information about the back end of the RHDHV data base is not publicly available.

The interest in this research lies in the following parameters: the vessel type, the *LOA*, the *DWT* and the *TEU* capacity.

TEU capacity and *DWT* are parameters often used to describe the load capacity of a certain vessel. *TEU* is used for container vessels to express the number of containers carried by the vessel. The following dimensions are represent 1 *TEU*: 20 feet long, 8 feet high and 8 feet wide (Ligteringen, 2017). The *DWT* is used for weight bulk carriers, and is defined as the maximum load of the vessel, including fuel, oil, crew and supplies expressed in metric tons (PIANC WG121, 2014). The *LOA* is the length overall, measured between the vessel's bow and stern, which must not be confused with the *Length between perpendiculars* (*LPP*), as visualised in Figure 3.1.

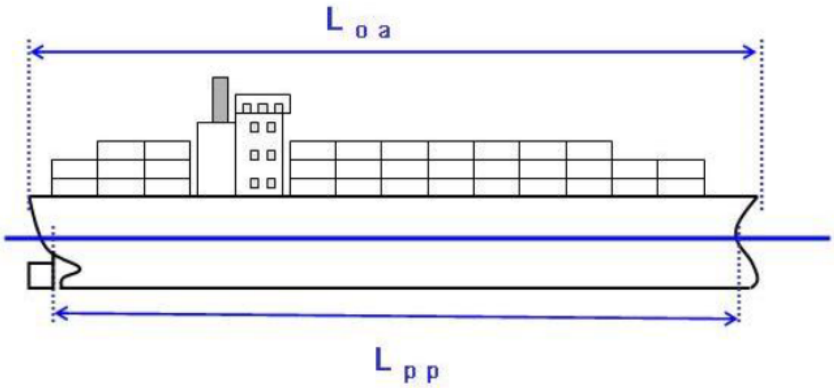


Figure 3.1: Typical ship dimension (source: PIANC WG121, 2014, Figure 1.2)

The vessel type is based on the type of vessel registered with the MMSI number. For the three different terminal types the possible vessel types are selected:

Terminal type	Vessel type
Container	General Cargo, Other Dry Cargo, Passenger/General Cargo, Refrigerated Cargo, Inland Waterways Dry Cargo / Passenger, Inland Waterways Others Non Seagoing, Container, Other Activities, Other Activities cont, None
Dry bulk	Bulk Dry, Bulk Dry/Liquid, Other Bulk Dry, Self Discharging Bulk Dry, General Cargo, Other Dry Cargo, Passenger/General Cargo, Refrigerated Cargo, Inland Waterways Dry Cargo / Passenger, Inland Waterways Others Non Seagoing, Inland Waterways Tanker, Other Activities, Other Activities cont, None
Liquid bulk	Chemical, Gas tankers, Oil, Other liquids, Inland Waterways Tanker, Bulk Dry/Liquid, Other Activities, Other Activities cont, General Cargo, Inland Waterways Others Non Seagoing, None

Table 3.1: Vessel type classification

If the type of the vessel is undefined, it is returned as type 'None'.

Furthermore, in some specific cases the Sea-web data base is consulted as a verification data source. When outliers arise and are investigated, the Sea-web data can function as a confirmation of the AIS data and visualised results.

3.2. Methodology

The methodology of this research follows the research approach as introduced in figure 1.2. The methodology in this research is split up into three different steps. First, the AIS tool set up is discussed. Next, the different required statistical parameters necessary for the study parameters are introduced. Finally, the method of performing goodness-of-fit tests on the observed distributions and theoretical distributions are elaborated.

3.2.1. AIS tool set up

Before the three study parameters can be generated, the AIS data should undergo steps as shown in figure 3.2. Based on several inputs and the raw AIS data, a data frame can be created which includes the entry and exit timestamps for every vessel track in the port, the anchorage area and the terminal. The steps represent how the AIS Port Processes Tool will be build.

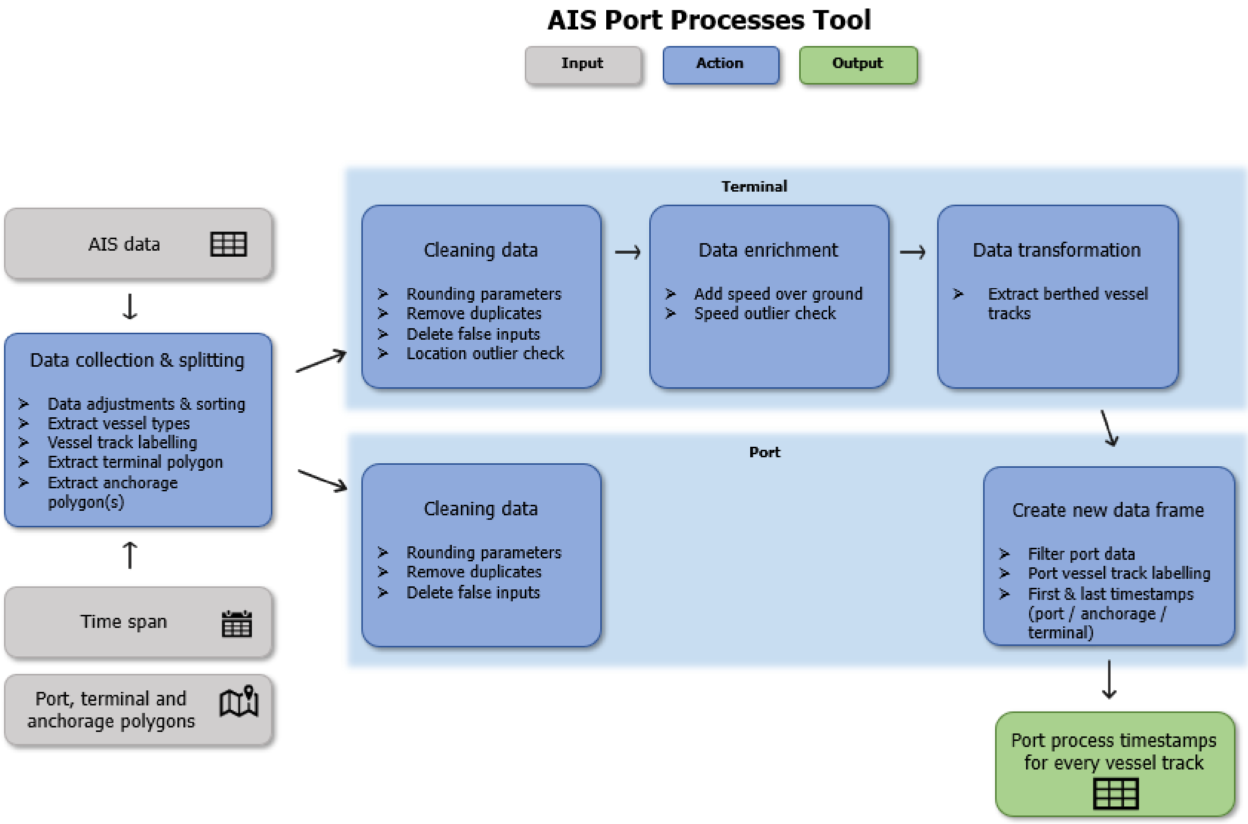


Figure 3.2: Tool flow chart

All data will be treated using Python programming language. The following steps can be used as a guideline to fulfill all research objectives:

- Firstly, AIS data is extracted. The data contains eight different parameters: MMSI, latitude, longitude, timestamp, vessel type, *LOA*, *DWT* and *TEU* capacity. Rows are sorted by MMSI number and timestamp, in order to later on characterise different vessel paths. A first filter step is performed based on the terminal type, using the vessel type. With the input of the port, anchorage and terminal locations the data can be split into smaller data frames for each of these locations.

- Next, data cleaning steps will be performed for both the terminal and port data. The data parameters will be rounded accordingly to minimize unnecessary computing time. Duplicate rows are removed, unrealistic inputs are deleted and a location outlier check is performed. Thereafter, the data of the terminal is enriched by adding *Speed over ground* (*SOG*). Vessel tracks in the terminal, which do not berth, should be removed from the data. A method will be defined to filter these vessel tracks.

- A new data frame will be created where per vessel track in the port, the port entry and exit, the anchorage entry and exit and the terminal entry and exit time are generated. The data frame will only represent vessel tracks that eventually berth at the specified terminal location.

Finally, from this new data frame, the service-, inter arrival- time and berth occupancy can be calculated. For the berth occupancy the operating hours should be known. With these distributions various distributions can be fitted to the data to determine the best fit.

3.2.2. Statistical analysis for general study parameters

As mentioned and extensively elaborated in chapter 2.2, the study parameters are the service times, the inter arrival times and the berth occupancy. First, the service times are defined based on the data frame created by the AIS tool. The service times are calculated for every single row, meaning for every single vessel track, and returned in hour units.

- Service times [hours] = terminal exit time - terminal entry time

The inter arrival times can be determined in two ways, based on the port arrival time or the terminal arrival time. They are defined in chapter 2.2.2 as the time between two successive arrivals at the **port**. Thus for the determination of inter arrival times the following steps are taken:

- Sort the data frame based on port terminal entry column
- Inter arrival times [hours] = port entry time (current vessel track) - port entry time (previous vessel track)

The port, anchorage and terminal timestamps are normalized by subtracting the first timestamp of the data frame (smallest port entry timestamp) from all the other timestamps. This way all data will be based on the first moment of the data set.

The berth occupancy is determined as the time the berth is physically occupied by a vessel relative to the total operating hours. As mentioned in chapter 2.2.3 the berth occupancy can only be calculated when the number of berths is constant over time, which means the berths and their locations are fixed. The required input for this step are the number of berths and the total operating hours per year (default: 365 x 24 hours). The berth occupancy is then calculated as follows:

- Generate a new data frame which splits the total time span into hours
- Calculate for every hour how many vessels are present in the terminal polygon
- Occupancy [per hour] = number of vessels present divided by the total number of berths
- Average occupancy [%] = average occupancy times *total hours in a year* divided by *operating hours per year*

If the number of berths is not fixed a different approach can be taken. The length occupancy over time can be calculated if the total length of the terminal is known. The adjusted length occupancy is determined as follows:

- Generate a new data frame which splits the total time span into hours
- Add, for every hour, the sum of vessel lengths present in the terminal polygon
- Length occupancy [per hour] = total length present divided by the total available quay length
- Average occupancy [%] = average length times *total hours in a year* divided by *operating hours per year*
- Adjusted length occupancy [per hour] = add 15 meters to every vessel length in order to take the design length range between vessels into account (subchapter 2.1). Then add all adjusted vessel lengths present per hour, divided by the total number of length available.

The length occupancy can be useful for container and dry bulk terminals. For liquid bulk terminals the berths are made of single jetties, where the vessels are unloaded at a central location thus shore-side facilities are fixed on a limited area (Ligteringen, 2017). Therefore, the length occupancy is irrelevant for liquid bulk terminals.

3.2.3. Distribution fitting on processed AIS data

Once the service time and inter arrival time distributions are plotted, multiple distributions can be fit to the data. Based on literature research the following distributions will be tested (for more information see subchapter 2.1.3):

- For inter arrival times

  - Exponential distribution
  - Gamma distribution
  - Erlang-2 distribution
  - Weibull distribution
- For service times
  - Exponential distribution
  - Gamma distribution
  - Erlang-2 distribution
  - Erlang-3 distribution
  - Erlang-4 distribution
  - Erlang-5 distribution
  - Normal distribution
  - Beta distribution

Two very common goodness-of-fit tests were introduced in subchapter 2.1.4: the K-S test and the Chi-square test. The choice is made to use the K-S goodness-of-fit test as well as visual and logical interpretations to decide which distributions fit the observed data. It must be taken into account that for the K-S test the sensitivity of the test is highest at the middle of the distribution and lower at the tails. In other words, the test is more sensitive to deviations at the middle of the distribution, compared to at the tails (*Kolmogorov-Smirnov test One- & two-sample, and related tests*, n.d.). An advantage of the K-S test statistic is that it does not depend on the underlying CDF. Also, the K-S is an exact test which does not depend on an adequate sample size, in order to be valid. Whilst, a minimum sample size is necessary for the Chi-square goodness-of-fit test (*1.3.5.16. Kolmogorov-Smirnov Goodness-of-Fit Test*, n.d.). This advantage is useful and important further on in the research when smaller sub sets of data are tested.

Visual interpretations focus on what the visual results of the observed and theoretical CDFs represent. Logical interpretations focus on what distributions would be expected for these certain distributions. For example, as mentioned, the Weibull distribution is usually used to represent wind speed statistics or reliability analysis of materials. Furthermore, the Beta distribution is a normalized constant of the Gamma distribution and mostly used for defining probabilities of occurrences. Combining this information, together with visual interpretations and the K-S goodness-of-fit test, will lead to robust conclusions of which theoretical distribution(s) best fit a certain observed data set.

II

Results

AIS Tool development

- Chapter 4: Results: AIS Port processes tool

AIS data analyses and comparisons to theoretical framework

- Chapter 5: Results: Service time distributions
- Chapter 6: Results: Inter arrival time distributions
- Chapter 7: Results: Terminal occupancy

Part II considers all results found during the research, necessary to answer all research questions. First, the results based on building the IS Port processes tool including all possibilities and limitations are discussed. Second, the results for the service time distributions are presented, based on the three different terminal types. Third, the inter arrival time distribution results are discussed, again based on the different terminal types. Finally, the results for the terminal occupancy of the different terminals are presented.

4

Results: AIS Port processes tool

The first section of the research focuses on the AIS Port processes tool. A method is generated in which different port processes are defined using raw AIS data. Specifically the extracting of vessels that berth is highlighted and different methods are compared. Finally, the possibilities as well as limitations of the developed AIS tool are discussed.

4.1. Using AIS data to define port processes

In chapter 3.2.1 the subsequent steps were briefly introduced on how to turn raw AIS data into a data frame containing information about the different port processes. A tool is created which follows the steps, as shown in figure 4.1. The necessary inputs are an AIS data set, a time span and locations of the port, terminal and anchorage area(s). The output is a data frame containing information about the entry and exit times for the port, the anchorage and the terminal areas. The tool is developed using Python programming language and is available on GitHub (Van Zwieteren, 2020). It can be used to assess and analyse any terminal from any port.

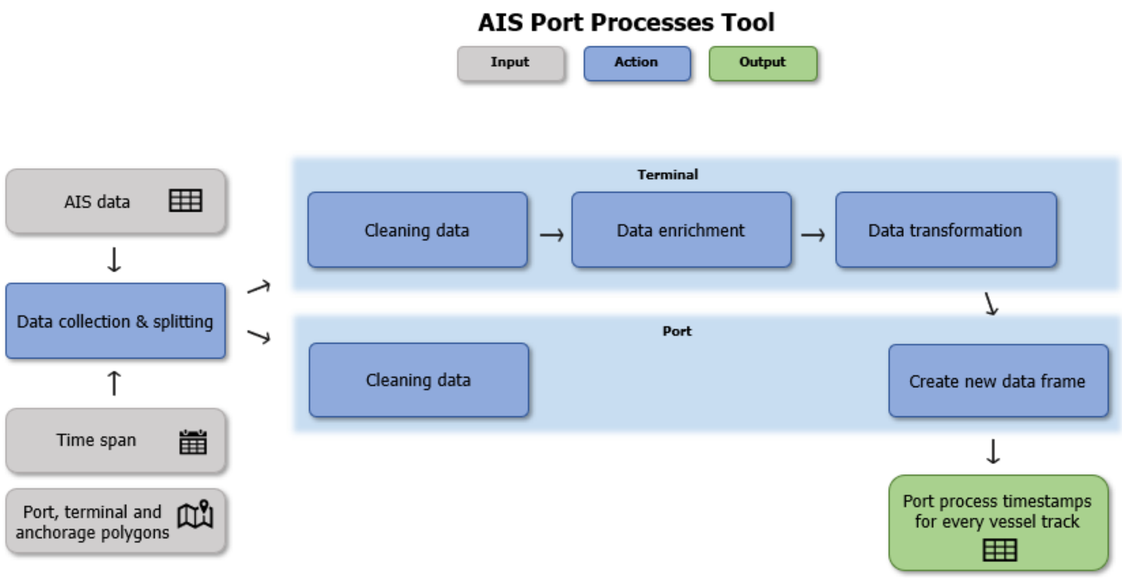


Figure 4.1: Flowchart AIS port processes tool

The *action* steps, as shown above in blue, will be highlighted in the following subchapters.

4.1.1. Data collection and splitting

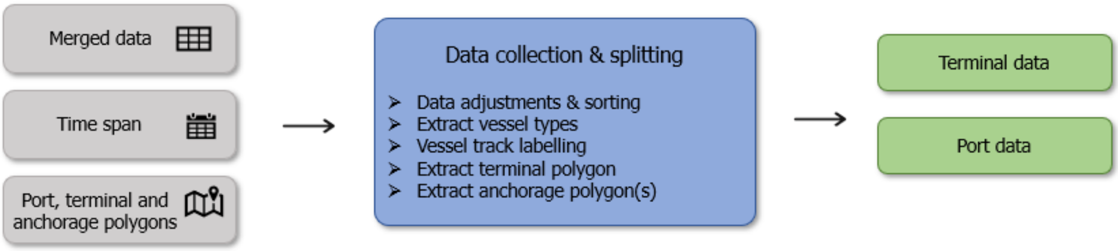


Figure 4.2: AIS port processes tool: Data collection and splitting

Data adjustments and sorting

The tool will be as generic as possible, thus it should be able to accept different types of data. A crucial step will become the reformatting of the data to the required format. The column names can be renamed to: 'mmsi', 'lat', 'lon', 'timestamp', 'type', 'loa', 'DWT' and 'teu_capacity'. Where the 'lat' and 'lon' represent the latitude and longitude of the position of the vessel, respectively. *LOA* is the length overall in meters and *DWT* represents the dead weight tonnage. Furthermore, the data should be sorted by MMSI number and by timestamp, in order to obtain the vessel paths chronologically.

More information about the specific data extraction steps are not publicly available.

Extract vessel types

Before cleaning the data, the data should be filtered based on the vessel type, as shown in table 3.1. In this table a distinction is made between the three terminal types: container, dry bulk and liquid bulk. During this filter step it is important to also keep all data from vessels with unspecified ('None') vessel types.

For this research twelve different port locations were investigated, consisting of 4 container terminals, 4 dry bulk terminals and 4 liquid bulk terminals. The chosen terminals with their specific port, anchorage and terminal polygons are visualised in appendix C. The time span is set between 01-05-2019 and 01-07-2020. The twelve terminals summed together lead to a total raw data set containing 449,343,847 rows (AIS messages). The average terminal consists of 37,445,320 messages and on average the filtering based on vessel types removes 75.3 % of the messages. More specific information per terminal can be found in appendix D.

Vessel track labelling

First, information is added to the data frame about whether or not the message is coming from the terminal or anchorage area. This leads to two new columns: 'in terminal' and 'in anchorage'. In these columns a 1 represents the message coming from the specific terminal and a 0 represents the opposite (message does not come from this location). For the twelve terminals on average 2.7% of the messages come from the terminal (appendix D).

Next, the vessel tracks are labelled in order to distinguish multiple visits from the same vessel as separate tracks. For the terminal, every vessel track is labelled. A vessel track is classified as a number assigned to a certain vessel, for all the messages sent during one trip to the terminal and back (in the terminal polygon). The track number is equal to zero when the vessel is not in the terminal polygon. An AIS message is classified as a **new** vessel track when:

- The previous message (row) in the data contains a different MMSI number.

The data is sorted based on MMSI and timestamp, thus when two subsequent rows differ based on MMSI numbers, this automatically results in a new vessel track.

- The previous message (row) did not come from the terminal location **and** the previous message (row) is from more than 2 hours earlier.

The assumption was made that once the vessel leaves the terminal polygon that the vessel has left the terminal completely. However, during this research some complications arose. Due to possible (coordinate) errors, or a terminal polygon being drawn too small, one actual vessel track might be registered as two separate tracks. When for the same vessel (same *MMSI*) the time difference between the latest timestamp of the previous vessel track and current timestamp, is smaller than 2 hours, the vessel track will be classified as the same as the one before. Once this time difference is larger than 2 hours, a new vessel track will be defined. In table 4.1 an example is given, purely fictive, where both situations (new- and same vessel track) are illustrated. This limit is set to 2 hours, after analyzing the complications based on the twelve data sets.

MMSI	Timestamp	In terminal	Vessel track number
123456789	'2020-01-01 01:00:00'	Yes	4
123456789	'2020-01-01 **01:40:00**'	Yes	**4**
123456789	'2020-01-01 02:00:00'	No	0
123456789	'2020-01-01 02:20:00'	No	0
123456789	'2020-01-01 **04:10:00**'	Yes	**5**
123456789	'2020-01-01 04:15:00'	No	0
123456789	'2020-01-01 **04:40:00**'	Yes	**5**
123456789	'2020-01-01 05:00:00'	Yes	5
123456789	'2020-01-01 07:00:00'	Yes	5

Table 4.1: Vessel track labelling example

Polygon extractions

The terminal data set is created by filtering the entire port data set based on the column 'in terminal'. Only AIS messages located in the terminal (in terminal = 1) are kept. The same can be done for the anchorage data set, though using the 'in anchorage' column.

4.1.2. Data cleaning

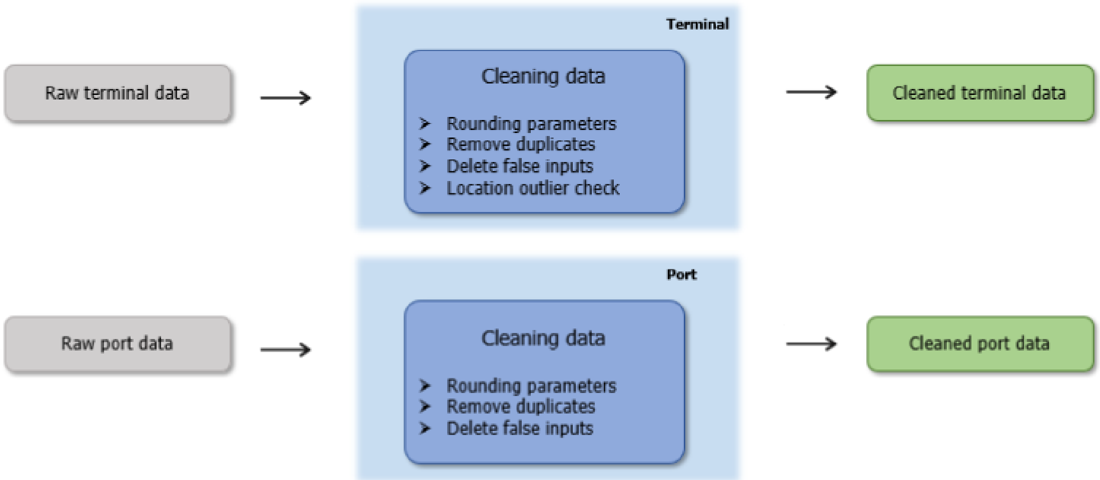


Figure 4.3: AIS port processes tool: Data cleaning for terminal and port

The port data frame is roughly 69 times as large as the terminal data based on the number of messages (appendix D), thus with regards to computation time, first only the data in the terminal is thoroughly cleaned.

The last cleaning step, the location outlier check, accounts for the most computational time. Therefore, the port data frame is cleaned using only data rounding, cleaning duplicates and deleting false inputs. If necessary the port data could be thoroughly cleaned as well, however in this research the most interest lies in the terminal data.

Data rounding

First, to further decrease the data size the parameters are rounded. It goes without saying that the MMSI will be rounded to zero decimals, which in the *RHDHV* data it already is, since it is a unique nine digit identification number (Valsamis et al., 2017). The latitude and longitude are rounded to 6 decimals, corresponding to an resolution of roughly 10 cm. Finally, the timestamp is rounded per second. An average reduction for the terminal data frame of 19.2 % in data size is found based on file size in *kB*.

Removing duplicate rows

The data base might contain duplicate rows. A code is written in which a loop goes through every row in the data and checks if the previous row's latitude and longitude are the same as the current row's latitude and longitude. If so, the current row will be dropped. The same is done for when the previous row's timestamp is exactly the same as the current's row timestamp. It is sufficient to check only the previous row, instead of all rows in the database, because the data is sorted by MMSI and timestamp. For all twelve original terminal data sets, containing a total of 3,227,276 messages (rows), 325,736 messages were removed. The average percentage removal per terminal is 9.09%. For the port data, containing an original total (for all twelve terminals together) of 126,070,391 messages, 12,027,049 messages are duplicates. Per terminal location the average removal of duplicates in port data is 7.83%.

Deleting false inputs

As mentioned before, MMSI numbers should have a length of 9 digits exactly. Physically it is not possible for latitudes to be outside the range of -90 to 90 degrees, and for longitudes to be outside the range -180 to 180 degrees. A code is used to delete all rows where the values contain one of these 3 impossible inputs. In the data set used there are zero rows which contain these false inputs, and thus 0% is removed.

Location outlier check

The location outlier check is done by considering three successive points in time, all for the same MMSI number. Vessel paths are considered where the time between two successive messages is not larger than 2 hours. The distance is calculated, as *dist* between point 1 and point 3. The middle point is defined as the exact center of these points 1 and 3. Next, the actual point 2 is checked by measuring the distance between this point 2 and the middle point (MP), this distance is called *x*. As visualised in figure 4.4 once x is smaller than *dist* the location is not registered as an outlier and thus will not be removed. On the other hand, when x is larger than *dist* the data point (row) will be dropped.

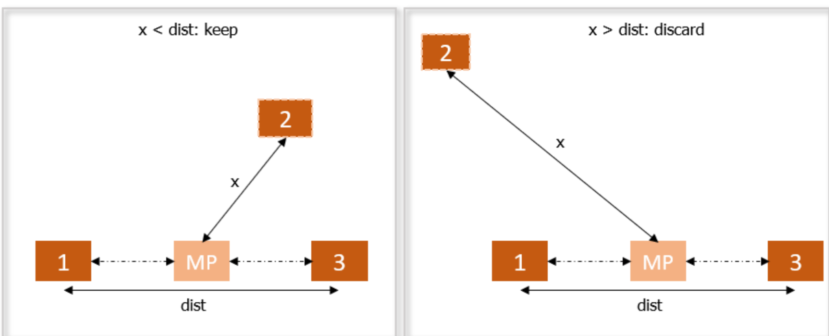


Figure 4.4: Visualisation of location outlier check

An exception is made that the outlier check is only performed when the *dist* is more than 30 meters, to prevent data with information about sharp turns and minimal movements, being deleted. The distance minimum is chosen as 30 meters in order to only take into account possible speed outliers for vessels which are sailing. At the berth, the vessel will make relatively sharp turns because the distance is small between each message. Thus this minimum distance limit will prevent these messages to be (possibly) seen as outliers. For the data set containing twelve locations, the location outlier check has removed 2,940 rows, with an average of 2940 messages per terminal. The average number of location outliers across the twelve terminals is low (0.10%).

4.1.3. Data enrichment

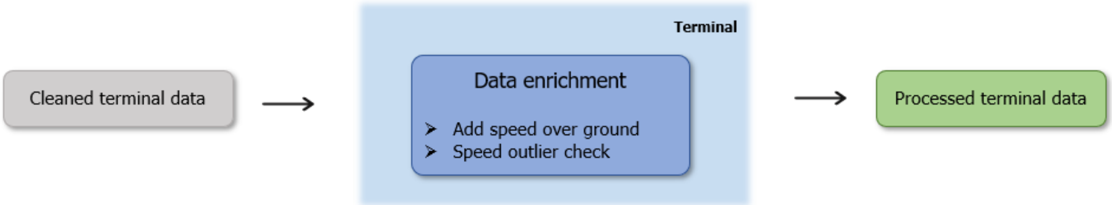


Figure 4.5: AIS port processes tool: Data enrichment for terminal

Adding speed over ground (sog)

From the parameters latitude and longitude, each vessel track shows per timestamp a coordinate. With these timestamps and coordinates the vessel speed can be derived. The best way to define the vessel speed is by using three successive messages for the same MMSI number. The data is split once there is more than 2 hours between two messages of the same MMSI number, that way it is seen as a 'new' track. It is noted that vessel might turn of AIS signals when berthed, however when considering the vessel speed, it is not seen as a problem for the system to consider those points as 'new track', since the speed will be very low at that point. A distinction is made in determining the speed:

- For all **'middle' messages** of a vessel track: speed is equal to the average of the speed the vessel had between the previous and current location, and the speed the vessel has for the current and next row. This average will return the exact speed at that timestamp moment. x represents the coordinate, t represents the timestamp, i represent the current row, $i-1$ the previous row and $i+1$ the next row.

$$sog = \frac{\frac{x_i - x_{i+1}}{t_i - t_{i+1}} + \frac{x_i - x_{i-1}}{t_i - t_{i-1}}}{2} \tag{4.1}$$

- For every **first message** of a vessel track: speed is equal to the distance between the current and next row, divided by the time between the current and next row.

$$sog = \frac{x_i - x_{i+1}}{t_i - t_{i+1}} \tag{4.2}$$

- For every **last message** of a vessel track: speed is equal to the distance between the current and the previous row, divided by the time between the current and previous row.

$$sog = \frac{x_i - x_{i-1}}{t_i - t_{i-1}} \tag{4.3}$$

- For all **single messages** (all vessel tracks that only contain 1 message) the speed is set to zero.

Speed outlier check

To remove unreasonable speeds the speed outlier check is performed. It removes all speed values above 25 m/s and below 0 m/s, which is a very high upper limit considering for example, an inland maximum speed limit of 3.6 m/s in the Port of Rotterdam (Port of Rotterdam, n.d.). The lower limit is set to 0, since the interest lies in vessels berthing, thus very low speeds are expected. When necessary for different research applications, the lower speed limit can be increased when for example only sailing vessels are requested.

Based on the cleaned total terminal data from all twelve terminals, the speed outlier check removed 578 messages (rows). The average removal of speed outliers is 0.07% (appendix D). To conclude, only data points remain between vessel speeds of 0 and 25 m/s.

Robustness of data cleaning and enrichment

Above steps all have contributed to the robustness of the data. The average terminal contained 9.09% duplicate rows. In total when regarding the sum of all twelve terminals, the number of duplicates consist of 11.8%, as shown in the pie chart in figure 4.6.

More information about the reason for this large number of duplicates is not publicly available.

All other cleaning and enrichment steps, the faulty inputs, location outlier and speed outlier checks, only minimally contribute to cleaning the data.

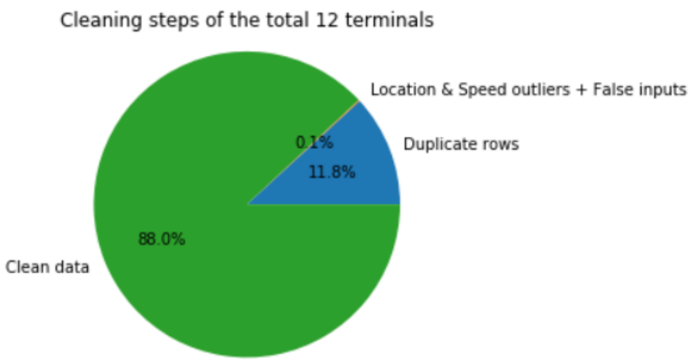


Figure 4.6: Visualisation cleaning and enrichment steps

4.1.4. Data transformation: extracting berthed vessel tracks

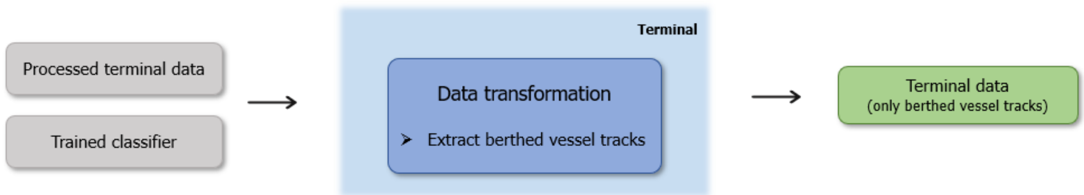


Figure 4.7: AIS port processes tool: Data enrichment for terminal

When extracting the data for the specified location and time span, all AIS messages are returned. This means that not only berthed vessels, but all other passing vessels will be recorded. Before analysing the inter arrival and service times of terminals, the vessels that don't berth should be removed, and only the vessels that berth should be considered.

The problem of extracting all of the berthed vessel tracks can be seen as a classification problem. A classification problem is a problem where a certain category is undefined for a new data set. In this situation the undefined parameter is whether or not the vessel has berthed. Once a method is defined which predicts this undefined parameter, a labelled data set (the validation set) can be used to test the quality of the prediction (GeeksforGeeks, n.d.). A classification problem can be solved in two ways: either by defining the rules manually, or by using the data to learn what the rules are.

Four different approaches, based on two different methods, are considered in order to define which vessels have berthed and which have not. The methods used are:

1. Identifying vessels lying still based on manually chosen limits
    - Approach using raw, basic data
    - Approach using pre-filtered data based on vessel type and with more information per vessel
2. Use a supervised machine learning algorithm, trained on a labelled data set
    - Approach using raw, basic data
    - Approach using pre-filtered data based on vessel type and with more information per vessel

An underlying goal is to keep the tool as generic as possible, thus for the first approach to both methods data is used containing only: *MMSI*, timestamp, latitude and longitude. However, after extensive research and multiple attempts, no sufficient solution of correctly defining berthed vessel tracks was found. Therefore, a different approach based on the same two methods as before, were performed on filtered data (based on vessel type filtering) and included: *LOA*, *TEU* capacity and *DWT*.

In order to validate the results obtained by these methods, the Sea-web database is used. Sea-web Ships is an online database and service which provides its users to extract data from different databases (Sea-web Ships, 2020). From Sea-web for certain terminals, data can be extracted for a given time span, which include the MMSI number, the time the vessel arrived at the port and the sailed time (the timestamp the vessel left).

Method 1: Identifying vessels lying still based on manually chosen limits

As mentioned by PIANC WG 135 (PIANC WG 135, 2014), the service time is defined as the entire process of the vessel at the berth including (un)berthing, (un)mooring and (un)loading. It is expected that vessels berthing at the terminal will spend a relatively long amount of time at the quay wall, compared to vessels not berthing at the terminal. Tug boats assist the vessels in their maneuvering towards the berth and are therefore assumed to move around a lot, not lying still for a longer amount of time (Ligteringen, 2017). Other vessels, such as cargo vessels that pass by and do not berth, are also expected to not lie still in the terminal area for a long time. Based on these assumptions the berthing of a vessel is defined as follows:

- Berthing = A vessel should 'lie still' for at least *Duration limit* hours
- Lie still = A vessel is classified as 'lying still' when:
    - The distance between successive data messages is less than the *Distance limit*
    - The speed between successive data messages is less than the *Speed limit*

This approach returned three different limits to be determined: Duration limit, Distance limit and Speed limit. Multiple combinations of limits were tested to find which combination best suits the determination of 'berthing'.

The results and more information about both approaches to this method, using concise and pre-filtered data, are given in appendix E.1 and E.2.

Method 2: Use a supervised machine learning algorithm, trained on a labelled data set

Machine learning uses programmed algorithms which learn from existing data to make acceptable predictions. A distinction is made between classification and regression models. Classification models attempt to predict the categorical class (discrete) for an output whilst regression models predict a numerical value (continuous) (Siguenza-guzman et al., n.d.). Evidently this research is dealing with a classification problem. Classification problems can be solved using many different machine learning algorithms, such as Support Vector Machine, Naive Bays, Decision Trees and Random Forests. Before these algorithms can be used the data quality should be sufficient. The 'Garbage in - garbage out' principle is important to take into account, meaning that poor data quality used will lead to data output being unreliable (Kilkenny & Robinson, 2018).

Processes should be implemented to ensure good quality data. Besides the cleaning and enrichment steps that are performed, feature selection is implemented. Feature selection is the process of selecting variables from the data that are expected to have the largest influence on the prediction ability of the model (Omara et al., 2018). An attempt is made to select features which represent different parts of the data, such as features focusing on the total vessel track (total time, total messages) in comparison to features focusing on the difference between two successive messages (average timestamp interval).

For the first approach using raw, basic data nine features have been chosen to represent the data, as shown in table E.10.

Code name	Feature
$timeinpolygon$	The total time the vessel was present [s]
$meansogpertrack$	The average speed of all data messages [m/s]
$avgtimestampinterval$	The average time between two successive messages [s]
$messagestot$	The total number of messages sent [-]
$avgdistance$	The average distance between two successive messages [m]
$averagespeedsmallest75speed$	The average speed for 75% of the slowest speed messages [m/s]
$stdspeed$	The standard deviation of speed [m/s]
$stddistance$	The standard deviation of distance between two successive messages [m]
$messagefrequency$	The frequency at which messages are sent [/hr]

Table 4.2: Features

For the second approach using the pre-filtered data four extra features are added.

Code name	Feature
$teucapacity$	The TEU capacity of the vessel [TEU]
loa	The length overall of the vessel [m]
DWT	Dead weight tonnage of the vessel [DWT]
$stdlocation$	The standard deviation of the location [m]

Table 4.3: Features added in second approach

Once the data is merged with the Sea-web data, a labelled data frame with features is available for the testing of different machine learning algorithms. The choice of an algorithm is important and different considerations are made to choose which best fits best to the classification problem (Çigşar & Ünal, 2019). Multiple classification algorithms will be tested returning accuracy scores.

Twelve different terminals will be used to generate a sufficiently large data set. These terminals differ from the terminals used in this research, in order to obtain a diverse data set on which the algorithms can be trained. From this large data set the data is split into a training set (80%) and a test set (20%). The data is split as a means to estimate the predicting accuracy, and for the training of the algorithms only the training data set will be used (Dobbin & Simon, 2011). After the model is trained on this labelled data, it can predict outcomes for an unlabelled data set (Uddin et al., 2019).

More information about the train data set, the terminals, the different classifiers and their accuracy scores can be found in appendices E.3 and E.4.

Result: Keep only berthed vessel tracks in terminal data

The best method, as described in the appendix, is the fourth approach using the pre-filtered data and supervised machine learning approach. The accuracy scores for the second approach are summarised in the table below.

Approach	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
1. Basic data	92.60	66.15	13.08
2. Pre-filtered data	97.06	99.05	7.13

Table 4.4: Result: basic versus pre-filtered data.

With the classifier trained on twelve different locations, any terminal location can be run on the model. The interest lies only in vessel tracks that berth thus only the true positives (TP) vessel tracks are returned by the model. If required, the data is returned not only in a data frame containing per row every vessel track, but also the original data with all data messages for the berthed vessel tracks can be returned. In this research the concise data frame containing one entire vessel track per row is sufficient. However, it might be of interest for other research objectives to examine the entire vessel track, thus keeping all data messages of the berthed tracks.

Of all the twelve terminals researched (note that these are not the exact same twelve terminals as used to train the classifier), an average of 24.54% of the tracks is classified as an obvious not-berthed vessel tracks. On average compared to the total number of vessel tracks, 56.50 % of the tracks is classified as berthed (appendix D).

4.1.5. New data frame: entry and exit times for port, anchorage and terminal

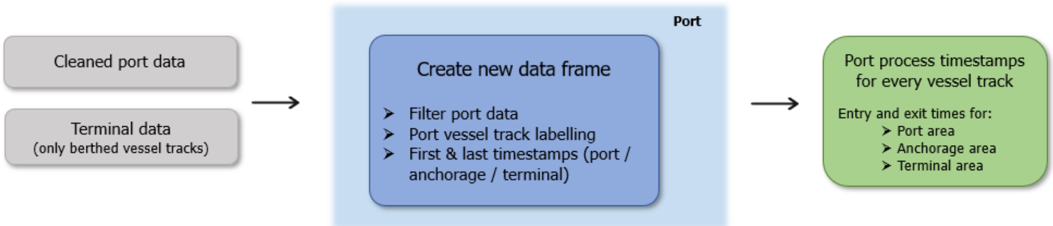


Figure 4.8: AIS port processes tool: New data frame

Filter port data

Once the XGBoost classifier has returned a concise terminal data frame containing all *MMSI* numbers that berth, the port data can be filtered. Initially the port data is only filtered based on MMSI numbers that berth. This filter step is not sufficient enough since the research objective lies in finding entry and exit times for only vessel tracks that berth at the terminal in the same port track. The data set after filtering might still contain port vessel tracks that do not berth at the terminal, for that certain port track. These vessels will visit the terminal at a different moment in time. In later steps these non-visiting terminal port tracks will be removed.

Port vessel track labelling

In subchapter 4.1.1 the port data was extended with two columns representing if the certain data message is coming from the terminal area or anchorage area and the port and anchorage data frames were labelled for each vessel track. Nonetheless the entire port data frame is still unlabelled, thus vessel tracks for the entire port are generated. A new port vessel track is created when:

- A row has a different MMSI number than the row before
- A row has the same MMSI number as the row before, **but** the row before is more than twelve hours earlier

First and last timestamps for port, anchorage and terminal

Lastly, the new data frame can be created. First, all port vessel tracks are removed which do not enter the terminal. Afterward, the port entry and exit times are defined as:

- Port entry time = the first timestamp of the entire port vessel track
- Port exit time = the last timestamp of the entire port vessel track

Next, the terminal entry and exit times can be defined as:

- Terminal entry time = the first timestamp of the **current** terminal vessel track
- Terminal exit time = the last timestamp of the **current** terminal vessel track

Situations can occur where one port vessel track visits a terminal more than once in the same port track. In this situation every terminal visit will create their own row in the data frame. Thus, in one data frame there can be port vessel tracks having multiple entries. For the analysed 12 terminals in this research on average 3.03% of the terminal vessel tracks were tracks that re-entered the terminal polygon in the same port track.

Once these terminal vessel tracks are known, it is important to filter the new data from which only **berthed** vessel tracks are kept. This is done by merging the data frame created in subchapter 4.1.4 with the newly created data frame based on an inner-join, keeping only the berthed vessel tracks. Finally, the anchorage entry and exit times can be defined in two different ways:

If the port vessel track has only one terminal vessel track:

- Anchorage entry time = the first timestamp in the anchorage area **if** this timestamp is before the terminal entry time
- Anchorage exit time = the last timestamp in the anchorage area **before** before the terminal entry time

If the timestamp in the anchorage area is after the first timestamp in the terminal, it means that the vessel did not enter the anchorage area before arriving at the terminal, meaning no waiting time has occurred and the vessel most likely passed the anchorage area while leaving the port. If there is no anchorage timestamp at all it is due to the vessel never entering the anchorage area, not before or after the berthing.

When there are no waiting times for a vessel track the tool returns the timestamp ’1970–00–00 00:00:00’. This does not affect later analysis as the anchorage entry and exit times are only assessed in comparison to each other. In other words, the waiting time is defined as the exit anchorage time minus the entry anchorage time, returning 0 when no data messages have been retrieved from the anchorage area.

If the port vessel track has multiple terminal vessel tracks:

- First terminal vessel track of port vessel track:
  - Anchorage entry time = the first timestamp in the anchorage area **if** this timestamp is before the terminal entry time
  - Anchorage exit time = the last timestamp in the anchorage area **before** before the terminal entry time
- Last terminal vessel track of port vessel track:
  - Anchorage entry time =

    - the first timestamp in the anchorage area **after** the previous terminal vessel track terminal exit time
    - **if** there is no timestamp after the previous terminal vessel track terminal exit time, then use anchorage entry time of previous terminal vessel track
  - Anchorage exit time = the last timestamp in the anchorage area **before** before the terminal entry time
- 'Middle' terminal vessel tracks of port vessel track:
  - Anchorage entry time =
    - the first timestamp in the anchorage area **after** the previous terminal vessel track terminal exit time
    - **if** there is no timestamp after the previous terminal vessel track terminal exit time, then use anchorage entry time of previous terminal vessel track
  - Anchorage exit time = the last timestamp in the anchorage area **before** before the terminal entry time

4.1.6. Conclusion: using AIS data to define port processes

During this research a tool is created which uses raw AIS data based on a defined port polygon and turns this into a data frame containing entry and exit times for the port, anchorage and terminal area. The tool is available on GitHub (Van Zwieteren, 2020). A few inputs are necessary in order to obtain this final table:

- Port polygon: coordinates (latitude and longitude) of port polygon corners
- Time span: two dates between which data is selected, if left blank all data available is selected.
- Terminal type: choose container / dry bulk / liquid bulk
- Number of anchorage areas: choose one / two
- Anchorage polygon: coordinates (latitude and longitude) of anchorage polygon corners, two different anchorages areas can be defined. If only one anchorage area is specified, the second anchorage area coordinates have no influence.
- Terminal polygon: coordinates (latitude and longitude) of terminal polygon corners
- Visualise: choose yes / no, visualises the different specified polygons with a selection of the data set in google maps

A table will be generated with this tool, an example of what the table looks likes is:

	term_track_number	port_track_number	mmsi	loa	DWT	teu capacity	type
0	1	1	123456789	140	3300	0	General Cargo
1	2	2	123456789	140	3300	0	General Cargo
2	3	2	123456789	140	3300	0	General Cargo

	port_entry_time	port_exit_time	terminal_entry_time	terminal_exit_time	
	2020-01-01 01:00:00	2020-01-01 05:00:00	2020-01-01 03:00:00	2020-01-01 04:30:00	
	2020-01-01 02:00:00	2020-01-01 18:00:00	2020-01-01 03:30:00	2020-01-01 14:50:00	
...	2020-01-01 02:00:00	2020-01-01 18:00:00	2020-01-01 12:30:00	2020-01-01 16:30:00	...

	anchorage_entry_time	anchorage_exit_time
	2020-01-01 01:20:00	2020-01-01 02:40:00
	2020-01-01 01:40:00	2020-01-01 02:30:00
...	2020-01-01 01:40:00	2020-01-01 02:30:00

Table 4.5: Example output of AIS tool

This chapter focuses on the first research objective *AIS data: processing and possibilities*. With this developed tool the first sub research question has been completed:

- *How can AIS data be used to define different processes a vessel follows in a port?*

A tool has been built which transforms raw AIS data to a data frame summarising the entry and exit times for certain port locations. In this research the focus initially was on three parts: the inter arrival time distribution, the service time distribution and the waiting time distribution. For the inter arrival time distribution, data is required which covers the entire port. For the service time data at and around the terminal is sufficient, whilst for the waiting times the anchorage area data was required.

The current Python Tool indicates the entry and exit times for these 3 areas: the port, the anchorage and the terminal. The data is structured in a way that for every vessel track (customer to the port) these 6 timestamps are listed as one row of data, as shown in table 4.5.

Using this tool a considerable amount of statistical analyses can be performed for different (performance) factors of the port and terminal. The tool is available on GitHub (Van Zwieteren, 2020). These possibilities will be further discussed in subchapter 4.2. However, in using this tool one should be aware of its limitations and assumptions, as will be discussed in subchapter 4.1.7.

The second sub research question has also been answered:

- *What is the most optimal procedure of extracting vessel tracks that berth at a terminal, using AIS data?*

As was extensively discussed in subchapter 4.1.4 and appendix E, four different approaches have been investigated, each trying to extract the berthed vessel tracks from a AIS data set. The most optimal approach was found by using supervised machine learning techniques. Using information from Sea-web as validation data, for twelve different terminal locations, a labelled data set was created by merging the AIS data with the Sea-web data. On 80% of the labelled data various algorithms were trained. Eventually, the algorithms were tested on the remaining 20% of the data to see how well they could predict a vessel track berthing or not.

The XGBoost classifier turns out to best predict the vessel berthing or not, with an overall accuracy of 97.06%. It uses 13 features, all representing information about the specific vessel track, to predict whether the vessel has berthed or not. The classifier can be used on any container, dry bulk or liquid bulk terminal. Limitations are discussed in 4.1.7.

4.1.7. Discussion: AIS tool limitations

During the process of building this tool multiple assumptions were made. Hard inputs as summarised in subchapter 4.1.6 are necessary for the working of the tool. One should take into account the following assumptions made in this tool, before using it to analyse the data.

Assumptions

- AIS Data: the AIS data is assumed to represent reality. *Limitations based on the RHDHV AIS data source are not publicly available.*

  Furthermore, the assumption is made that all vessels do not turn of their AIS signals before berthing at the terminal.

- Type filtering: the AIS data is filtered based on the vessel type. A list of types per certain terminal (container, dry bulk or liquid bulk) is made using classifications from RHDHV, as shown in table 3.1. The assumption is made that no other vessel types (which are not on the specified list) berth at the terminal.
- The container, dry bulk and liquid bulk vessels are split into different vessel classes, as shown in tables B.1, B.3, B.4.
- A new terminal vessel track is defined when a new vessel enters the terminal polygon, **or** when the same vessels re-enters the terminal polygon and the last AIS message was more than 2 hours before, as discussed in subchapter 4.1.1.
- The latitude and longitude are rounded to 6 decimals, thus have a resolution of 10 cm.
- All MMSI numbers contain 9 digits, all other MMSI numbers are deleted from the data.
- A location outlier module was created, as shown in subchapter 4.1.1. This module assumes no location outlier can occur when the distance between two subsequent messages is smaller than 30 meters.
- The speed over ground (sog) is defined as the average speed of the speed just before and just after the timestamp.
- The speed outlier check assumes a minimum limit of 0 m/s and a maximum limit of 25 m/s.
- Extracting berthed vessel tracks: the assumption is made that the XGBoost Classifier predicts correctly whether or not the vessel track has berthed at the terminal (subchapter 4.1.4).
- Sea-web data: the validation data set for the training of the XGBoost classifier is based on Sea-web data. Therefore, the assumption is made that Sea-web contains the correct information regarding vessel arrivals at terminals.
- Port vessel track labelling: a new port vessel track is defined as the moment a new MMSI enters the port polygon **or** if the same vessel (MMSI) re-enters the port, more than twelve hours after the last AIS message was sent. More information regarding this assumption in subchapter 4.1.5.
- Anchorage entry and exit times: the anchorage entry and exit times are assumed to occur **before** the vessel track enters the terminal. All messages after the vessels leave the terminal, and possibly re-enter the anchorage area, are left neglected (subchapter 4.1.5).
- When a vessel track re-enters the terminal, in the same port track, the anchorage entry and exit times are defined as the last before entry/exit of the anchorage area. Thus, it might occur that a vessel enters the terminal twice, for both vessel tracks the same entry and exit times for the anchorage area will be used (subchapter 4.1.5).

A lot of these assumptions contain inputs which can be adjusted in further research, if necessary for different research objectives.

To summarise, AIS data is extracted from the RHDHV data base.

More information regarding the RHDHV data source is not publicly available.

However, it is very important to realize that even when these data bases are correctly registering the AIS messages, there can still be errors in the data due to vessels wrongfully sending AIS messages. For example, vessels tend to stop sending AIS messages when berthed at a terminal. Therefore, the chance exists that vessels stop sending AIS messages too soon (when arriving), or start sending again too late (when leaving).

Furthermore, limitations of the tool are present, based on these assumptions but also based on the data and tool capabilities.

Limitations

- Port polygon: the port locations of the polygon should be manually entered into the tool. The chosen port boundaries are therefore affected by personal interpretations.
- Anchorage polygon(s): a maximum of two anchorage areas can be defined. If more specific anchorage areas are present, this can not be included in the current tool. Again, the chosen anchorage boundaries are manually input and can therefore be affected by personal interpretations.
- Terminal polygon: only one terminal polygon can be input. When a terminal contains a unique or abnormal shape this might be hard to implement as 1 terminal polygon. Again, the chosen terminal boundaries are manual inputs, thus open to personal interpretations.
- The current tool focuses on the above 3 mentioned locations. It therefore presents the port processes: port entrance/exit, anchorage entrance/exit and terminal entrance/exit. The tool does not focus specifically on the processes that occur between these locations, such as the pilot boarding, the tug assistance or the sailing times between these processes. This leads to recommendations for expansion of the tool.

4.2. AIS tool applications

The developed AIS tool produces a considerable amount of statistical analyses possibilities. Performance indicators can be defined using the tool to test and compare different ports and terminals. With this, the last sub research objective of 'AIS data: processing and possibilities' will be answered:

- *What are performance analyses which can be generated using AIS data?*

The following three analyses will be thoroughly investigated in this research.

- Service times

Using the entry and exit times of the terminal, the service time can be computed of vessels. With these service times a distribution can be formed and the theory about service time distributions can be challenged. Besides that, the service time distributions can be visualised for smaller sub populations, such as for a specific vessel type.

- Inter arrival times

Using the entry times of the port the inter arrival times can be calculated. With these times a distribution can be formed which can be compared with current theory about inter arrival times, based on different terminal types.

- Occupancy

The utilization of a terminal can be calculated using the information of how many vessels are present in the terminal polygon at a certain time. This can be translated into an occupancy based on number of berths or total terminal length. These will be necessary manual inputs (terminal length / number of berths).

Besides the three main research topics, a lot of other analyses are possible using the tool. A few possible applications are shortly introduced. These form potential extra research questions, interesting for further research.

- Waiting times

The waiting times can be calculated using the entry and exit times for the anchorage area(s). In addition to the waiting times, the ratio waiting in terms of service times can be found. This is interesting when comparing terminals based on performance indicators. Again, the distinctions can be made between terminal types, as well as more specifically vessel types. Two examples for the waiting times and the waiting times in terms of service times are given for the Rotterdam APM-2 Container Terminal, in figures 4.9 and 4.10 below.

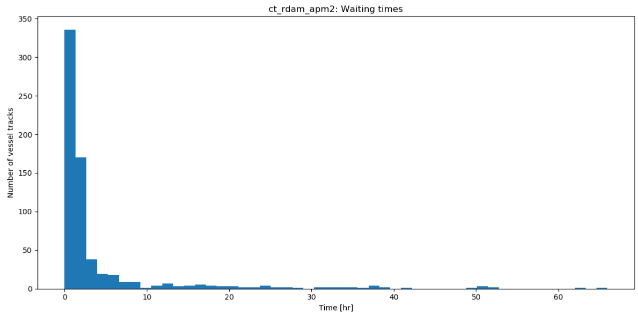


Figure 4.9: Waiting time distribution (Rotterdam APM-2 Terminal)

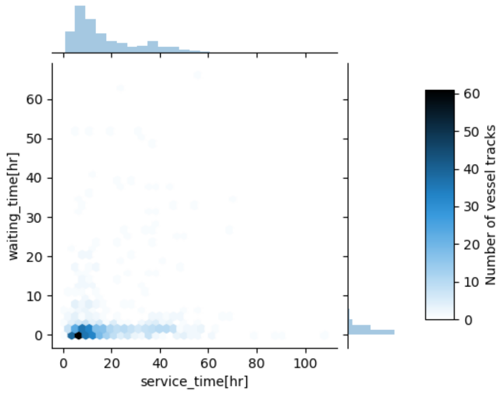


Figure 4.10: Waiting time versus service time (Rotterdam APM-2 Terminal)

- Arrivals over time

Based on the AIS data obtained from the tool, the number of arrivals at a certain terminal or port can be calculated and visualised. Correlations and trends between terminals over different time frames can be investigated. As shown in figure 4.11 the number of arrivals can be categorized per vessel class. Or as shown in figure 4.12 the number of arrivals can be categorized per month, for different vessel classes. Both figures are examples of the Rotterdam APM-2 Container Terminal.

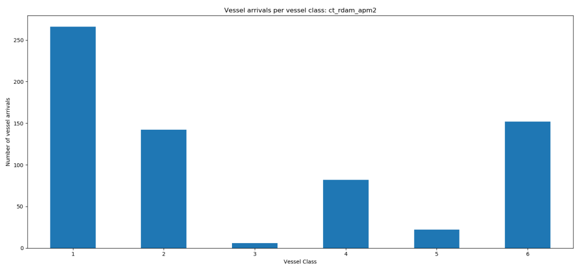


Figure 4.11: Arrivals per class (Rotterdam APM-2 Terminal)

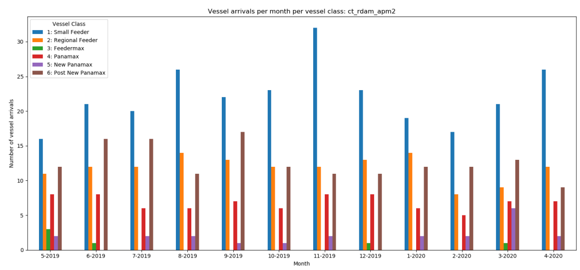


Figure 4.12: Arrivals per month, per vessel class (Rotterdam APM-2 Terminal)

- Multiple parameters, visualised over time

Besides plotting the arrivals over time, multiple other parameters can be plotted over time. Such as the average service time over time, for example as in figure 4.13, which can also be split into different vessel classes (figure 4.14), both from the Rotterdam APM-2 Container Terminal. Visualising over time will give insights about possible external influences on a port. For example, visualising the number of arrivals over time could show possible effects of Covid-19.

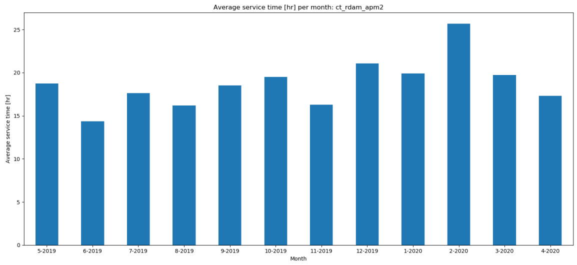


Figure 4.13: Average service time per month (Rotterdam APM-2 Terminal)

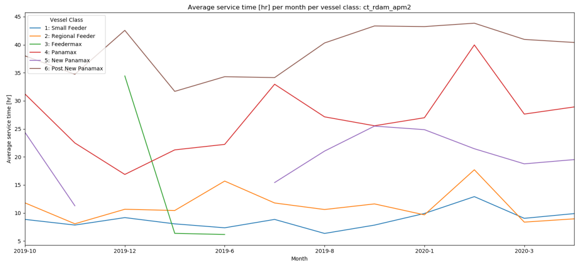


Figure 4.14: Average service time per month, split into different vessel classes (Rotterdam APM-2 Terminal)

5

Results: Service time distribution

In this chapter the results for the service time distributions are discussed. First, the service time distribution will be analysed for container terminals, then for dry bulk terminals and finally for liquid bulk terminals. For every terminal a goodness-of-fit test is performed based on multiple fitted theoretical distributions. Every terminal data set is split into specific vessel class data sets. Again, the goodness-of-fit test is used to find the most optimal theoretical fitted distribution. Finally, comparisons are made between the service time distributions of the three terminal types: container terminals, dry bulk terminals and liquid bulk terminals.

5.1. Container terminals: service time distributions

5.1.1. Service time distributions using all vessels

The service times of all the berthed vessel tracks are plotted in a histogram using 100 bins. Multiple distributions, as stated in subchapter 3.2.3, are fit to the service time distribution. The *CDF* of the original data is plotted, together with the CDF of the theoretical fitted distributions. For four container terminals the service times are plotted and goodness-of-fit test is performed. The first terminal, Rotterdam APM-2 Terminal will be thoroughly demonstrated next. The other three terminals are extensively described and analysed in appendix F.1.

First, the container terminal Rotterdam APM-2 is analysed, the service time histogram is shown in figure 5.1. All distributions are fitted and visualised, as shown in figure 5.2. In total, 1816 vessels berth in the selected time span. Based on the *K-S* test none of the distributions are possible fits with the data, based on the limit of 5% of the Null hypothesis (results in table 5.1). This result corresponds to what is visually observed. None of the CDF seem to correctly follow the service time distribution.

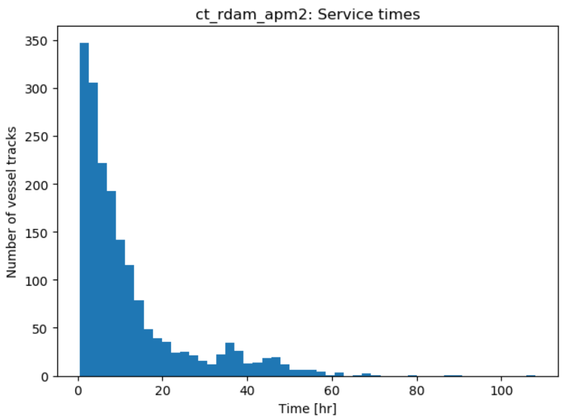


Figure 5.1: Rotterdam APM-2 Terminal service times (histogram)

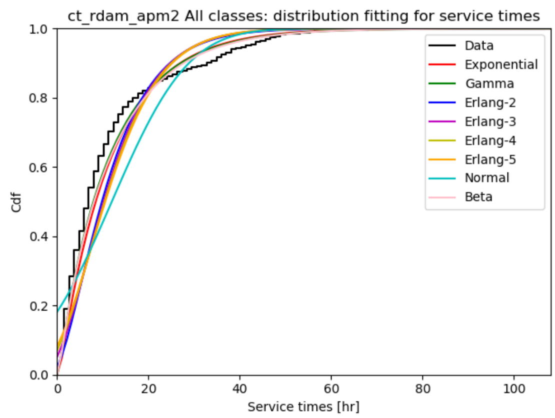


Figure 5.2: Rotterdam APM-2 Terminal with fitted distributions

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p [%]	Lim
Exponential	0.50	11.50	1.00		0.07	0.00	No
Gamma	0.50	12.52	0.90		0.05	0.00	No
Erlang-2	-1.45	6.72	2.00		0.13	0.00	No
Erlang-3	-4.54	5.51	3.00		0.15	0.00	No
Erlang-4	-7.41	4.85	4.00		0.15	0.00	No
Erlang-5	-10.06	4.41	5.00		0.15	0.00	No
Normal	12.00	13.16			0.19	0.00	No
Beta	0.50	809.77	0.88	61.17	0.06	0.00	No

Table 5.1: Service time distribution fitting for container terminal: Rotterdam APM-2

The same methodology is applied to the Rotterdam APM, Rotterdam Euromax and Le Havre Atlantic container terminals, as demonstrated in appendix F.1.1. Based on these first four terminals, no fit can be found for any of the fitted theoretical distributions, as summarised below in table 5.2.

Terminal	Rotterdam APM2	Rotterdam APM	Rotterdam Euromax	Le Havre Atlantic
Theoretical fit	None	None	None	None

Table 5.2: Service time distribution fitting for container terminals: Best theoretical fits

UNCTAD specifies in their handbook that the service time varies considerably between different terminals based on the type of vessel, the quantity and type of cargo, and the rate at which the cargo is handled (UNCTAD, 1985). It therefore assumes the service time distribution to follow an Erlang-k distribution. Along with UNCTAD, most literature available suggest the Erlang-k distribution for container as well as dry bulk terminals (see table 2.6).

Altogether this implies that the expected Erlang-k distributions do not correspond to the four service time distributions found. The service time distribution, such as an expected Erlang-k distribution, is used in queuing theory models in order to predict the approximate waiting times, which can lead to the number of berths. To recap, the service time is expected to depend on (among others):

- The vessel type
- The quantity of the cargo
- The type of the cargo
- The handling rate at the terminal

The quantity and handling rate of the cargo are unknown and unavailable when using AIS data. However, the vessel type (and thus the type of cargo) are known parameters, since these parameters were merged together with the data for the benefit of an earlier pre-processing step (subchapter 4.1.4). To repeat, the type of the vessel, the vessel length and the vessel DWT are known parameters in this research. With these parameters the vessel can be classified into a certain vessel class category. In subchapter 5.1.2 a distinction is made between different vessel classes in order to eliminate the possible influence such a variable might have on the service time distribution. The hypothesis is generated that the theoretical distributions will better fit to these smaller sub data sets, where the great diversity of vessel types is reduced.

Interpretations of, and comparisons between, the container terminals

Comparisons between the container terminals are visualised in figure 5.3 and table 5.3. The Rotterdam APM terminal service time distribution results in a very different visual interpretation due to an extreme outlier. The outlier is the container vessel LEXA MAERSK which stays at the terminal for 539 hours, which is verified using Sea-web data base and visual inspections of the data. More information about the investigation of this outlier is given in appendix F.1.1. The conclusion is made that this outlier represents the reality and this vessel did not leave the terminal polygon during the specific time period. To better inspect the differences between the terminals the y axis are adjusted, as shown in figure 5.4.

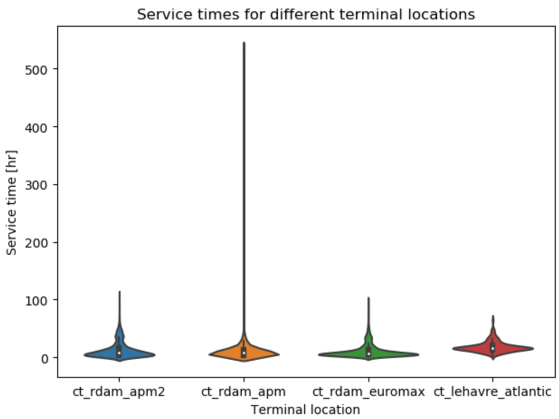


Figure 5.3: Service times for container terminals

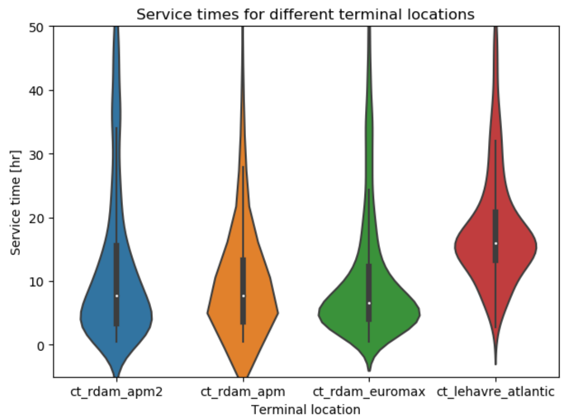


Figure 5.4: Service times for CT (zoom)

For the Le Havre Atlantic Terminal the service time distribution visually also takes on a different shape. When plotted as a histogram (figure F.7 in appendix F.1.1), the terminal seems to obtain a more spread and divers service time (two visual peaks). This terminal contains a much shorter quay wall (800 meters) and receives less vessel arrivals over the same time period (383 arrivals). The average service time of this terminal is relatively higher and the service time distribution is more spread out (relative differences between 25% and 75% quartiles, table 5.3). As mentioned, the service time is dependent on many different factors. The higher service time for this terminal is expected to be due to lower cargo handling rates, since the type of cargo is the same between the terminals. The higher average service time can also be caused by a larger quantity of cargo (un)loaded at the terminal. The quantity of cargo transfer is an unknown parameter in this research, thus no concise conclusions can be drawn. However, the type (and thus size) of the vessel might correlate to the average service time. This correlation will be investigated later on in this research (subchapter 5.1.2). Finally, the lower number of vessel arrivals could be the cause of the more diverse service time (25% and 75% quartile differences), as there are simply less number of data points used in this distribution.

	Terminal information		Service times [hr]			
Terminal	No. of arrivals [vessels]	Terminal length [m]	Mean	25% Quartile	Median (50%)	75% Quartile
Rotterdam APM2	1816	1500	12.0	3.30	7.30	14.40
Rotterdam APM	2108	1500	9.90	3.5	7.30	12.80
Rotterdam Euromax	3001	1900	10.36	3.86	6.38	11.69
Le Havre Atlantic	383	800	17.01	10.65	15.32	20.03

Table 5.3: Service time distribution for all container terminals

Besides the differences found for the Le Havre Atlantic terminal the shape of the violinplots are similar and the three terminals in Rotterdam all return very comparable average and median service times. The conclusion made by Ducruet et al. about larger ports obtaining higher time efficiencies, seems to correspond to the average service times between the Rotterdam (larger) port and Le Havre (smaller) port (Ducruet et al., 2014). The

Port of Rotterdam contains lower average service times for it's terminals, compared to the Le Havre terminal. However, the conclusion is somewhat unstable since the amount of cargo transferred between vessel and terminal is unknown. If the amount of transferred cargo at the Rotterdam terminals and Le Havre terminals is very different, this conclusion is baseless. More information about their research is addressed in subchapter 2.1.1.

5.1.2. Service time distributions using only specific vessel classes

The container vessels will be split into different vessel classes according to table B.1 for the container vessels, and table B.2 for general cargo vessels. For all terminals the number of vessel tracks per vessel class are visualised in appendix F.1.2 (figures F.9, F.10, F.11 and F.12). The terminals are merged together and per vessel class the number of arrivals is shown in figure 5.5. Especially the APM-2 and APM terminal in Rotterdam contain a large amount of inland waterway vessels, classified as container class 1. This is not unexpected as it is known that for example the Rotterdam APM-2 terminal contains a separate specific part of the quay, fully focused on the inland vessels. In the previous subchapter the observation was made that for the Le Havre Atlantic terminal the average service time was higher compared to the service times of the other terminals. This specific terminal receives mostly container vessels of container class 4: the New Panamax. The expectation is made that larger vessels which can carry more TEU, also will (un)load more TEU at a terminal (on average). Therefore, the higher average service time at the Le Havre Atlantic terminal could be caused by a larger quantity TEU (un)loading (due to the large amount of class 4 vessels).

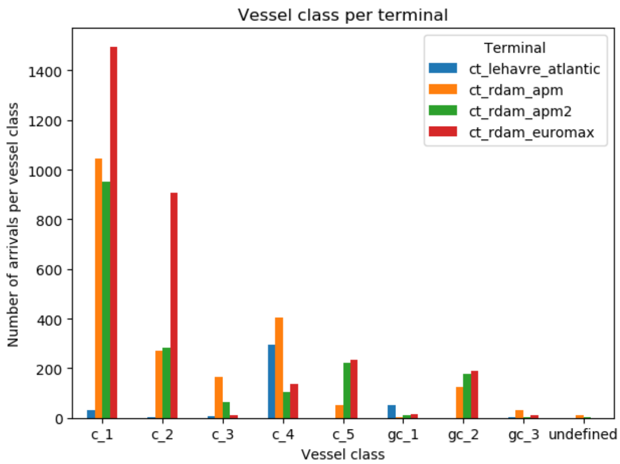


Figure 5.5: Container terminals: Arrivals per vessel class (all terminals)

In next steps, only the five different container vessel classes (c1, c2, c3, c4, c5) are investigated. For these five classes, and for every separate container terminal, again the same methodology is applied. The best theoretical fit is chosen for the service time distribution based on K-S goodness-of-fit tests, combined with visual interpretations of the CDF. Extensive data analyses of these separate classes is demonstrated in appendix F.1.2. A summary of the possible theoretical distributions per vessel class and per terminal, is shown in table 5.4.

Terminal	Rotterdam APM2	Rotterdam APM	Rotterdam Euromax	Le Havre Atlantic
Container class 1	B	B	E-3, E-4, B	Unreliable
Container class 2	E-5	G, E-4, E-5	E-2, B	Unreliable
Container class 3	E-3, E-4, G	G, E-4, B	Unreliable	Unreliable
Container class 4	G, B, E-5	None	E-2, B, G	G
Container class 5	N, G, B	N, G, B	G, B	Unreliable

Table 5.4: Service time distribution fitting for container terminals: Best theoretical fits, per vessel class (*B = Beta, E- = Erlang-, G = Gamma, N = Normal*)

The main findings and insights of the different container classes are:

- Container Class 1: The APM-2 and APM terminal both return the only fit as the Beta distribution. The Euromax Terminal returns the Erlang-3, Erlang-4 and Beta distributions as possible fits. A clear common distribution is the Beta distribution.
- Container Class 2: The second class does not have one distribution that fits for all three terminals, however the Erlang-5 and Beta distribution both fit on two out of three terminals.
- Container Class 3: This class is more difficult due to there only being two reliable terminals. Erlang-4, Gamma and Beta are all three common distributions.
- Container Class 4: The Rotterdam APM Terminal is the first to not fit any distributions. However, the other three terminals do have possible theoretical fits. The Gamma distribution is the only common fit among the three.
- Container Class 5: For this class clearly the Gamma and Beta distributions can represent the service time distributions well, based on three different terminals.

It is clear that splitting the total data set (including all vessel class arrivals, into smaller specific vessel class data sets) has been beneficial in terms of service time distribution fitting. Not all terminals return reliable results for every vessel class. The minimum amount of vessel arrivals for container terminals in order to be classified as reliable is set at 30 arrivals. Eventually, all data sets evaluated contained at least 50 vessel arrivals. A common factor between the terminals is for most classes the Gamma or Beta distribution. Since the Beta distribution is a normalized constant of the Gamma distribution and the Gamma distribution is an expected distribution for these kinds of processes, the Gamma distribution is selected as the best representative of the service time distributions for separate container vessel classes.

Besides summarising the best fits per vessel class, the service time distribution for every distribution is plotted (figures 5.6, 5.7, 5.8 and 5.9). In some cases certain vessel classes do not contain enough vessel tracks in order to make robust conclusions. However, from these figures an overall conclusion can be made. As repeatedly mentioned, the service times are dependent on different factors. The type of the vessel and the quantity of the cargo are two of these factors. Larger container vessel classes are able to carry larger amounts of cargo (TEU containers) and are thus, on average, expected to (un)load more cargo. This assumption is verified using these figures. The overall conclusion is drawn that for a higher vessel class, the service time distribution shifts towards the right of the plot, resulting in higher service times for these distributions.

A remark should be made about the APM Terminal, which contains the previously discussed outlier (extreme service time) in class 4. This leads to the less smooth curves for the total distribution and fourth class (figure 5.7).

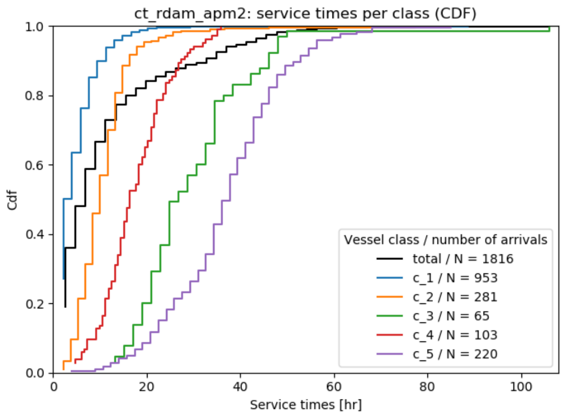


Figure 5.6: Service time distribution per class Rotterdam APM-2

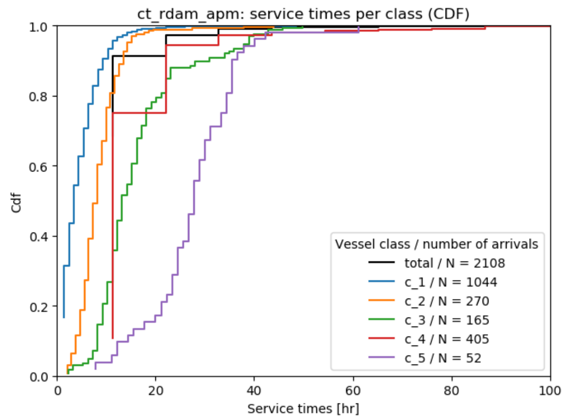


Figure 5.7: Service time distribution per class Rotterdam APM

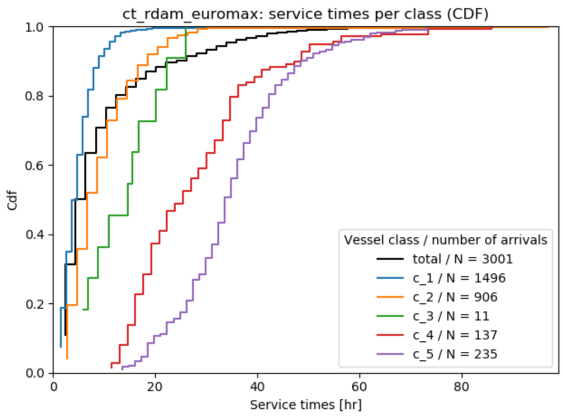


Figure 5.8: Service time distribution per class Rotterdam Euromax

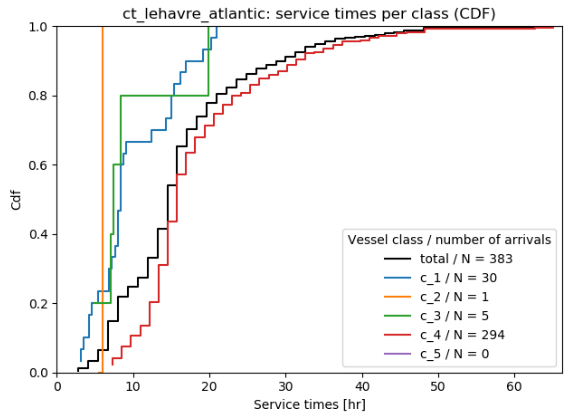


Figure 5.9: Service time distribution per class Le Havre Atlantic

5.2. Dry bulk terminals: service time distributions

5.2.1. Service time distributions using all vessels

The same steps will be taken in analysing the dry bulk terminals as were taken for the container terminals. A histogram with 100 bins is made for the service times and multiple distributions are fitted to the CDF. Four different dry bulk terminals are analysed: Rotterdam EMO, Vlissingen OVET, Rotterdam EECV and Dunkirk Western Bulk Terminal. All results can be found in appendix F.2.1. A summary of the possible theoretical fits found is given in table 5.5.

Terminal	Rotterdam EMO	Vlissingen OVET	Rotterdam EECV	Dunkirk Western Bulk
Theoretical fit	None	None	None	E-2

Table 5.5: Service time distribution fitting for dry bulk terminals: Best theoretical fits (*E-* = *Erlang-*)

For three out of four terminals no theoretical distribution is found that can describe the data. For the Dunkirk terminal the Erlang-2 is chosen as the best distribution to represent the data. UNCTAD assumes the service time distribution of dry bulk terminals to follow an Erlang-k distribution. This assumption corresponds to the Dunkirk Western Bulk service time distributions. However, for the other three terminals this does not. With regards to the choice made to split the data for the container terminals, the same deliberation is made and the vessel classes will be split to further investigate. Again the hypothesis will be tested of the theoretical distributions obtaining better fits to these smaller sub data sets, due to the reduction of diversity in the vessel mix.

Interpretations of, and comparisons between, the dry bulk terminals

Comparisons between the dry bulk terminals are visualised in figure 5.10 and table 5.6. For the Rotterdam EMO, the Vlissingen OVET and the Rotterdam EECV terminals a large number of the service times are relatively low and a peak forms around roughly 5 hours (figures F.29, F.33, F.35 in appendix F.2.1). For the Rotterdam EMO and Rotterdam EECV terminal some vessel tracks with extremely small service times are further investigated using Sea-web and visualisations of the tracks. The vessel tracks either do not correctly represent a berthed vessel track, due to errors in the AIS tool, or the small service times are correctly represented and vessels (mostly Inland Waterway Tankers) tend to contain very short service times. These situations are more elaborately discussed in appendix F.2.1. These large numbers of small service times are clearly visible in figure 5.10 for the Rotterdam EMO and Rotterdam EECV terminal. As was observed for the container terminals, again the smallest terminal (based on quay length) corresponds with the smallest number of vessel arrivals, found for the Dunkirk Western Bulk.

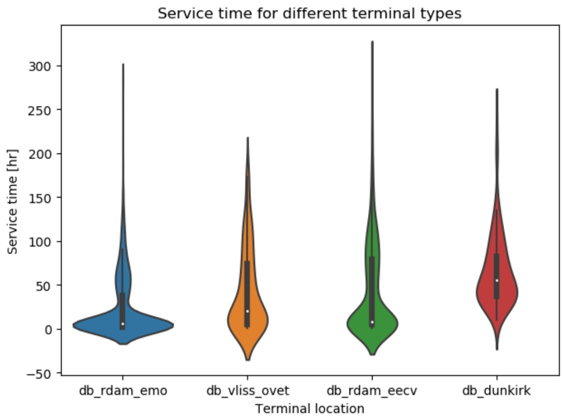


Figure 5.10: Service times for dry bulk terminals

	Terminal information		Service times [hr]			
Terminal	No. of arrivals [vessels]	Terminal length [m]	Mean	25% Quartile	Median (50%)	75% Quartile
Rotterdam EMO	907	2700	23.21	2.30	5.75	38.09
Vlissingen OVET	127	950	43.53	5.22	20.29	73.72
Rotterdam EECV	514	1090	42.68	4.46	8.54	79.27
Dunkirk Western Bulk	94	675	65.48	36.59	54.96	82.27

Table 5.6: Service time distribution for all dry bulk terminals

Furthermore, the distributions are not very similar and vary largely among themselves. The dry bulk terminals will handle different selections of cargo among the different terminals, leading to the large differences between the different terminals. All terminals show a large difference between the 25 and 75% quartile values (table 5.6), representing a very diverse spread out service time distribution. For the container terminals a smaller number of arrivals corresponded with the most diverse service time distribution (subchapter 5.1.1). This correlation is not found for the dry bulk terminals. Loading and unloading operations are very different for dry bulk terminals, leading to terminals often being designed for one-way traffic only. For example, unloading systems and equipment consist of grabs, vertical conveyors, bucket elevators and pneumatic systems. Whilst the loading systems are virtually always continuous processes, where there are multiple vessel loaders fed by a belt conveyor system (Ligteringen, 2017). It is important to note that all four terminals handle both importing and exporting cargo, resulting in the large differences within the service time distribution of every terminal. Comparisons between the terminal types (container, dry bulk, and liquid bulk) will be discussed later on in subchapter 5.4.

5.2.2. Service time distributions using only specific vessel classes

The dry bulk vessels will be split based on the classification made, as shown in table B.3. In the appendix F four figures represent the number of vessel tracks per vessel class (figures F.39, F.40, F.41 and F.42). A summary of these four is given in figure 5.11. The choice is made to focus solely on the dry bulk vessels (db1, db2, db3, db4, db5). Thus, the large amount of undefined vessels (most of which represented by *Inland Waterways Tanker*) are neglected in this research and will not be analysed and assessed. On these five dry bulk vessel classes thorough data analyses has been performed, as shown in appendix F.2.2.

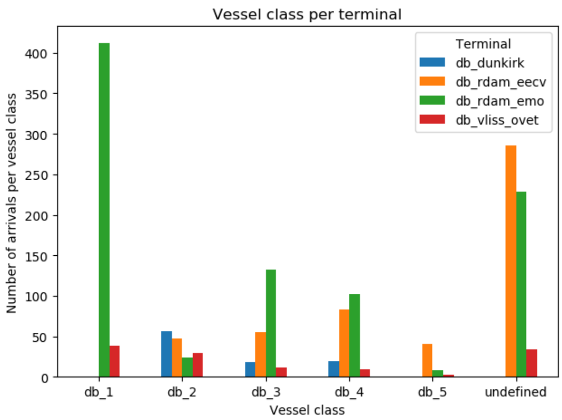


Figure 5.11: Dry bulk terminals:Arrivals per vessel class (all terminals)

A similar conclusion can be drawn as for the container terminals: splitting the total data set into smaller data sets (based on vessel class) has been beneficial in terms of fitting theoretical distributions to the service time distributions. As mentioned, all results are elaborately discussed in appendix F.2.2. An overview of the best fits found per vessel class is shown below in table 5.7.

Terminal	Rotterdam EMO	Vlissingen OVET	Rotterdam EECV	Dunkirk Western Bulk
Dry bulk class 1	Exp, B	G	Unreliable	Unreliable
Dry bulk class 2	Unreliable	G, E-2	E-k	G, B
Dry bulk class 3	E-5	Unreliable	B, G, E-2	Unreliable
Dry bulk class 4	E-4, B	Unreliable	B, G	Unreliable
Dry bulk class 5	Unreliable	Unreliable	G	Unreliable

Table 5.7: Service time distribution fitting for dry bulk terminals: Best theoretical fits, per vessel class (*B = Beta, E- = Erlang-, G = Gamma, N = Normal, Exp = Exponential*)

Where:

- Dry bulk Class 1: Two terminals contain enough data, the fits are different between the terminals. For the Rotterdam EMO terminal is best represented by the Exponential or Beta distribution and the Vlissingen OVET terminal is best represented by the Gamma distribution.

- Dry bulk Class 2: Three terminals can be used for the second vessel class. Two terminals are represented by the Gamma distribution. The Rotterdam EECV terminal is best fit by the Erlang-k distributions.
- Dry bulk Class 3: The third class is analysed for two terminals. The Rotterdam EMO terminal is best fitted by an Erlang-5 distribution whilst the Rotterdam EECV has the Beta, Gamma or Erlang-2 distributions as best fits.
- Dry bulk Class 4: Again only two terminals can be used. For both terminals the Beta distribution is one of the possible fits.
- Dry bulk Class 5: This class is only represented by the Rotterdam EECV terminal which, for which the service time distribution is most similar to a Gamma distribution.

A similar conclusion is made for the dry bulk terminals (compared to the container terminals). All of the smaller data sets correspond to at least one of the tested theoretical distributions. However, it must be noted that quite a few of the vessel class data sets did not contain enough data points (> 30) in order to return reliable results. The Gamma distribution seems to represent a lot of the different distributions, for all the dry bulk vessel classes. Nonetheless, the Erlang-k distribution also fits for roughly half of the fitted data sets.

The service time distributions are plotted for all the separate vessel classes. For the first three terminals (figures 5.12, 5.13 and 5.14) the conclusion can be drawn that a higher vessel class leads to an (on average) higher service time, since the service time distributions shift towards the right of the plot. To recap, the service time is dependent on characteristics such as the type of vessel and quantity of cargo. The assumption made for the container terminals again holds: larger vessels are able to carry larger amounts of cargo, therefore on average (un)loading more cargo at terminals. Thus, the higher average service time for higher vessel classes is again not unexpected.

For the Dunkirk Terminal the third class seems to have higher service times than the fourth service class, in which the previous drawn conclusion does not hold anymore. However, this terminal (specifically these two classes) contain a minimum amount of vessel arrivals, which makes these remarks less reliable.

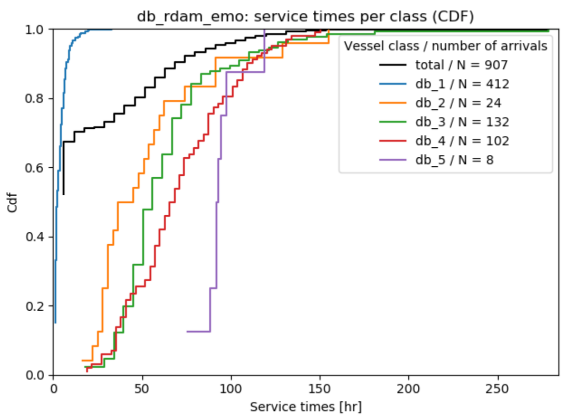


Figure 5.12: Service time distribution per class Rotterdam EMO

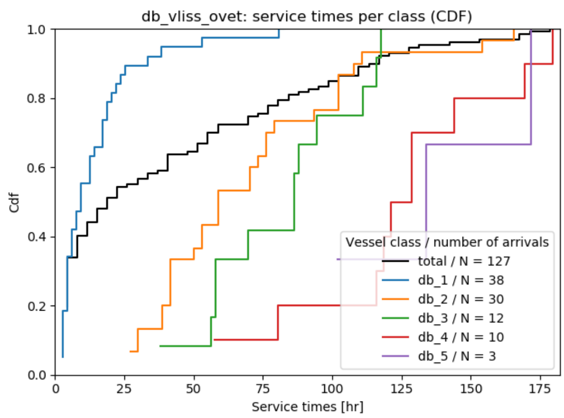


Figure 5.13: Service time distribution per class Vlissingen OVET

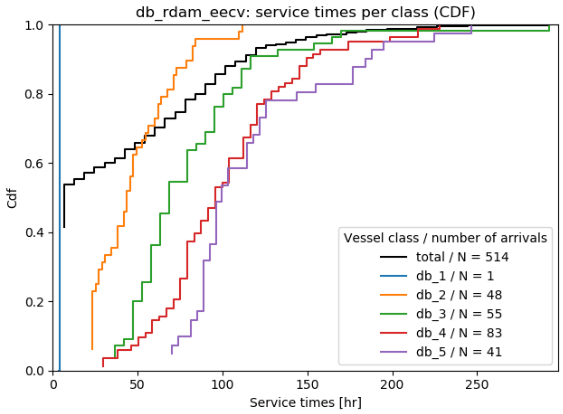


Figure 5.14: Service time distribution per class Rotterdam EECV

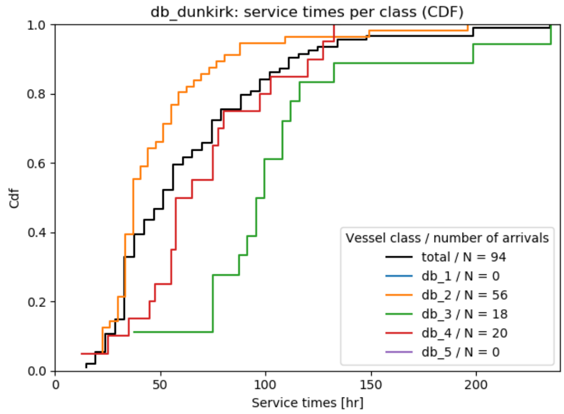


Figure 5.15: Service time distribution per class Dunkirk Western Bulk

5.3. Liquid bulk terminals: service time distributions

5.3.1. Service time distributions using all vessels

Again, a histogram is plotted to visualise the different service time for every terminal. In total four different LNG terminals are considered: Rotterdam GATE, Zeebrugge LNG, Dunkirk LNG and France Montoir LNG terminals. The distributions, as fitted for the container and dry bulk terminals, are fitted to the CDF of the service time. All results regarding the liquid bulk terminals can be found in appendix F.3.1. An overview of the possible theoretical fits for these terminals service time distributions are given in table 5.8.

Terminal	Rotterdam Gate	Zeebrugge	Dunkirk	France - montoir
Theoretical fit	None	None	None	None

Table 5.8: Service time distribution fitting for liquid bulk terminals: Best theoretical fits

PIANC and UNCTAD do not specifically mention LNG Terminals and their expected service time distributions. Accordingly, the Erlang-k distributions are assumed to best represent the service times. The service time distribution for all four LNG terminals does not correspond to any of the fitted theoretical distributions, as summarised in the table above and thoroughly assessed in appendix F.3.1. Therefore, the total data set will be split into smaller sub sets to investigate whether this will have a positive effect with regards to possible theoretical fits.

Interpretations of, and comparisons between, the liquid bulk terminals

When analysing the LNG terminals one matter immediately stands out. The service times seem to lie around a peak between 22-25 hours, which can also be seen in the visuals in figure 5.16 and average & median values in table 5.9. These average are not unexpected as the lay time often is set between 12 and 60 hours (more information in subchapter 2.1.1). The majority of the service times (for all terminals between 25% and 75% quartiles) all lie between these lay time limits (12-60 hours). Thus, the LNG terminals apparently do require a service time between 12 and 60 hours, dependent on the type of operations required and efficiency of the terminal.

Furthermore, the terminals service time distributions are very similar and the number of berths (1 or 2) does not seem to influence the distribution (figure 5.16). The service time for LNG terminals is very dependent on the type of operations required. Therefore, the assumption is made that the terminals, especially the Rotterdam GATE, Zeebrugge and Dunkirk terminals, all perform relatively similar operations, leading to the similar service times. Finally, the type of vessels is expected to be very similar among the four different terminals, leading to the similar service times. This last hypothesis will be validated by splitting the data sets dependent on the vessel class.

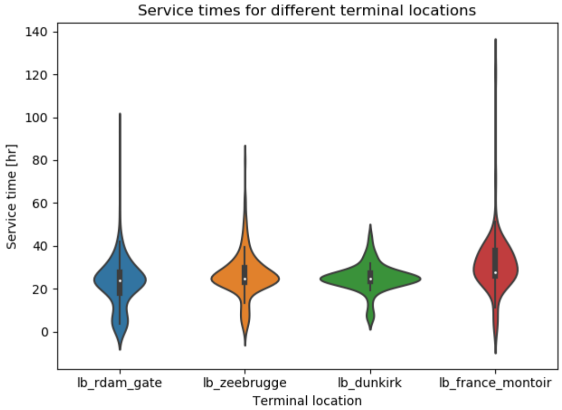


Figure 5.16: Service times for liquid bulk terminals

	Terminal information		Service times [hr]			
Terminal	No. of arrivals [vessels]	No. of berths	Mean	25% Quartile	Median (50%)	75% Quartile
Rotterdam GATE	173	2	22.98	17.91	23.85	27.60
Zeebrugge	185	2	26.47	22.91	24.90	29.64
Dunkirk	68	1	25.20	23.39	24.85	27.19
France-Montoir	121	2	32.11	26.04	27.47	37.68

Table 5.9: Service time distribution for all dry bulk terminals

5.3.2. Service time distributions using only specific vessel classes

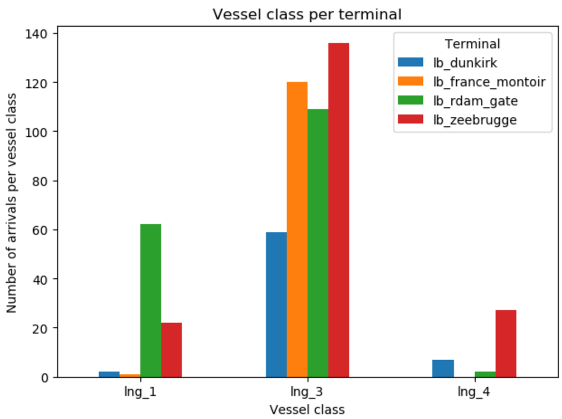


Figure 5.17: Liquid bulk terminals: Arrivals per vessel class (all terminals)

The LNG vessels will be split according to the classification made, as shown in table B.4. The number of arrivals per vessel class are plotted in figure 5.17. For every separate terminal figures F.61, F.62, F.63 and F.64 in appendix F represent the number of arrivals. It is clear that the bulk of the vessels arriving are vessels from the LNG3 class. This is not completely unexpected as LNG terminals are built to facilitate a specific type and class of vessel. No terminal receives vessel from class 2. Class 1 is mostly noticeable in the Rotterdam GATE terminal, and can only be tested on this terminal and the Zeebrugge Terminal (with some precaution: only 22 vessel arrivals). Class 4 can only be analysed using the Zeebrugge Terminal, however, also using this terminal some hindsight should be taken into account due to the number of arrivals (27 vessels).

The effect of splitting the data into smaller sub data sets based on vessel classifications, in order to increase the number of theoretical fits to the service time distribution, is not as useful as it was for the container and dry bulk terminals. The new (smaller) data sets often contain no possible theoretical fit, something which was different for the other two terminal types. All classes are elaborated extensively in appendix F.3.2. Summarising per vessel class the following remarks are made.

- LNG Class 1: Two terminals were analysed. The Rotterdam GATE terminal service time distribution was best fitted by the Erlang-k distributions. The Zeebrugge LNG terminal is best presented by the Exponential distribution.
- LNG Class 2: No terminals analysed that received this vessel class.
- LNG Class 3: All four terminals receive a considerable amount of vessels from the third vessel class. However, none of the terminals correspond to any of the fitted theoretical distributions.
- LNG Class 4: Only the Zeebrugge LNG terminal receives vessels from this vessel class. The Normal distribution was selected as best fit on the distribution.

An overview of the possible fits per vessel class is given in table 5.10 below.

Terminal	Rotterdam Gate	Zeebrugge	Dunkirk	France - montoir
Liquid bulk class 1	E-k	Exp	Unreliable	Unreliable
Liquid bulk class 2	-	-	-	-
Liquid bulk class 3	None	None	None	None
Liquid bulk class 4	Unreliable	N	Unreliable	-

Table 5.10: Service time distribution fitting for liquid bulk terminals: Best theoretical fits, per vessel class (*E- = Erlang-, N = Normal, Exp = Exponential*)

No clear theoretical fit can be found for the different vessel classes. The majority of the vessels arriving are from class 3. Since the entire data set also returned no possible fits, the chance of theoretical fits on this class (very similar to total data) was low. Visually, the distributions tend to shift towards deterministic distributions, with an average service time of roughly 24 hours.

For all the liquid bulk terminals the service time distributions are plotted for the different LNG vessel classes. Due to the minimal amount of arrivals for certain classes, it is difficult to make concise conclusions. The LNG class 3 (green line) is often very similar to the total data set (black line), due it having the most amount of arrivals, compared to the other classes. Furthermore, the LNG Class 1 (blue line) always contains smaller service times compared to the other vessel classes.

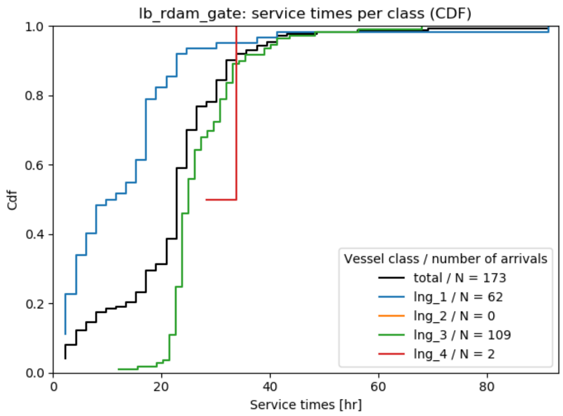


Figure 5.18: Service time distribution per class Rotterdam GATE

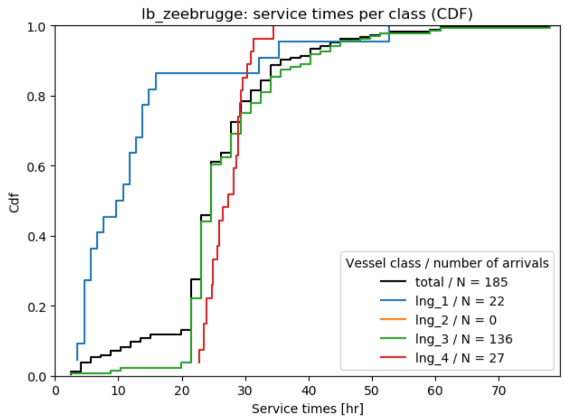


Figure 5.19: Service time distribution per class Zeebrugge LNG

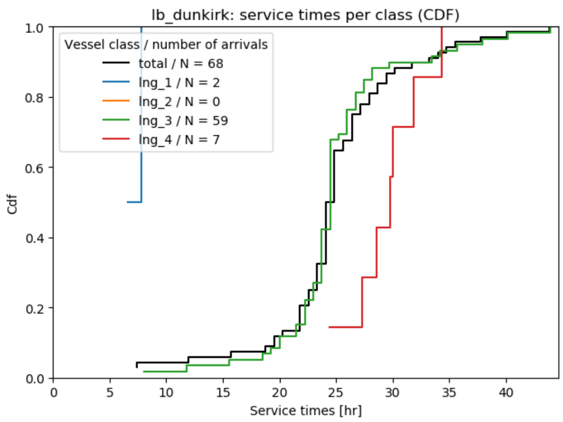


Figure 5.20: Service time distribution per class Dunkirk LNG

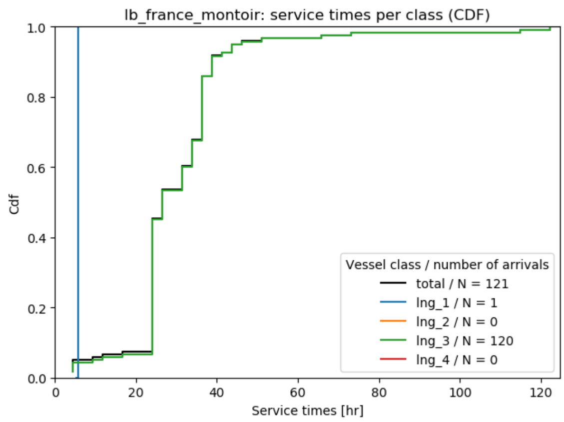


Figure 5.21: Service time distribution per class France Montoir

5.4. Comparisons and correlations between the terminal types

In the previous sections the service time distributions have been compared in between the terminal types and the terminals themselves (based on different vessel classes). The next three figures represent the service time distribution for different vessel classes, for each separate terminal. It must be noted that every figure represents a different categorization of vessel classes.

The container and dry bulk vessels clearly both tend to increase the service times for higher vessel classes (figures 5.22 and 5.23). As mentioned, the service time is dependent on multiple factors, for example the type of vessel class and the quantity of the cargo. As earlier discussed, it makes sense that a higher vessel class (indicating larger vessels), will most likely transport more cargo, and therefore require a longer service time. The liquid bulk terminals are more challenging to assess based on the smaller number of vessel arrivals for certain classes. The median service times for the liquid bulk seem to be very similar between the LNG3 and LNG4 classes.

All twelve terminals are combined into one violinplot where the total data set is taken into account. To better interpret this figure the y axis are adjusted, demonstrated in figure 5.25. Again, the most variance between the terminal types is found within the dry bulk terminals. At a dry bulk terminal a more variable set of vessels might arrive, compared to the other two terminal types. Furthermore, these terminals handle different cargo types, whereas the container terminals for example all contain the same container handling techniques, leading to less diversity between the container terminals. Additionally, the service times of dry bulk terminals have been mentioned to be very complicated and therefore variable. Both the container, as well as the liquid bulk terminals show little variance between the four terminal types. Especially the liquid bulk terminals are very similar, which is probably caused by the vessel mix arriving at these terminals consisting of mostly of LNG3 type vessels. The equipment for LNG (un)loading is very dependent on the vessel class, similar between the analysed terminals, and will deliver a very consistent (un)loading process. Therefore, the very similar service times are not unexpected.

Thus, the service time of the dry bulk terminals are most variable and diverse (highest spread between the 25% and 75% quartiles). The average service time for the container terminals is the lowest (between 10 and 17 hours). After the container terminals, the lowest average service time is of the liquid bulk terminals (between 23 and 32 hours). The dry bulk terminals have the highest average service times (between 23 and 65 hours). This is not unexpected as it corresponds to the statement of Abdul Rahman et al. that dry bulk terminals require very complicated procedures and are dependent on many variables (subchapter 2.1.1). For the container and liquid bulk terminals the median and average values are similar, but for dry bulk vessels the median values represent much smaller service times. This is due to a large number of vessels staying relatively short, as discussed in subchapter 5.2.1. On average though, the service times are much larger for these terminal types.

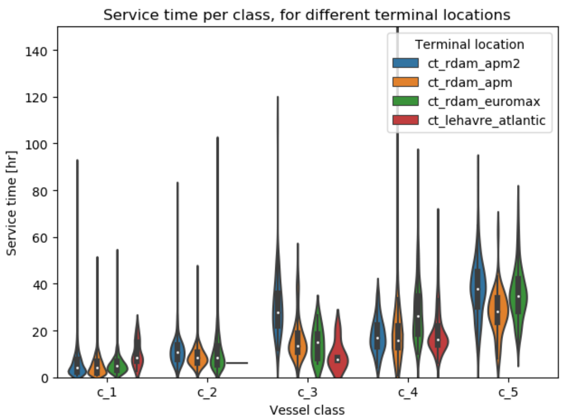


Figure 5.22: Service time distribution per class, Containers

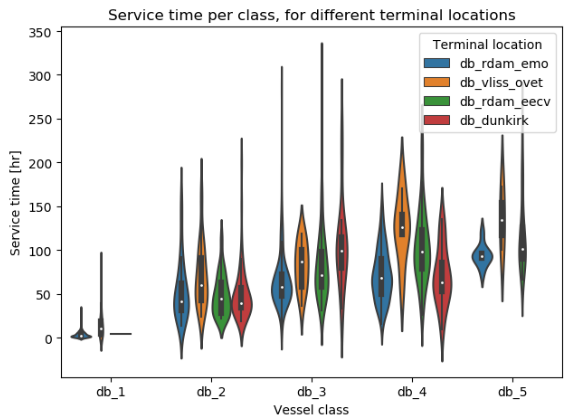


Figure 5.23: Service time distribution per class, Dry Bulk

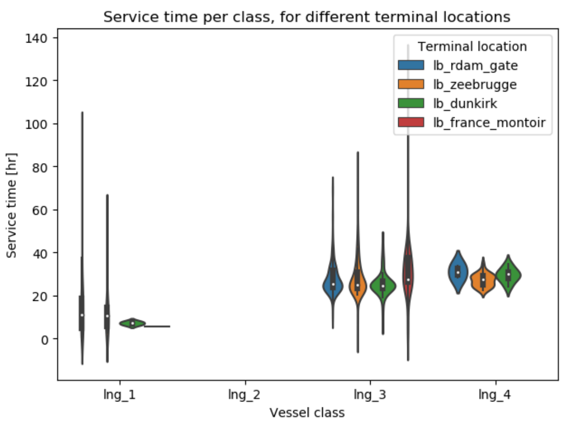


Figure 5.24: Service time distribution per class, Liquid bulk

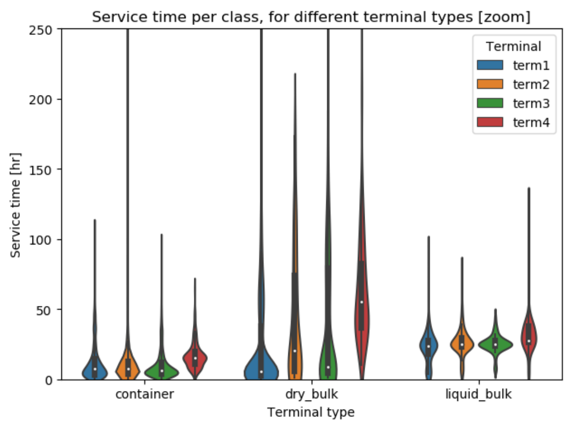


Figure 5.25: Service time distribution per terminal type

5.5. Discussion and conclusions: service time distributions

All the results are summarised and discussed in this next subchapter. First, the results and disclaimers are discussed. Next, the research objectives are repeated, together with the answers found.

5.5.1. Discussion and limitations of the service time distribution results

The results in this section are all based on the definition of the service times:

- Service time = The time present at the berth, including (un)berthing, (un)mooring and (un)loading (PIANC WG 135, 2014).

In this research the service time results depend mainly on the following:

- The terminal polygon: the terminal is manually defined by the user of the AIS tool. In this research the terminal location boundaries were set to include on one side just a little more than the quay wall (to include possible GPS outliers in the location of the messages). Opposite of the quay wall the terminal is defined as roughly 3 times the average vessel width, into the channel. The outer edges of the terminal were chosen at the locations the quay wall begins and ends. Examples are given in appendix C.
- The AIS data: the AIS data is assumed to represent the reality based on all the vessels present in the terminal polygon. Errors such as vessels which turn off their AIS signals too soon, are not taken into account and thus will be present in the current data sets.
- The AIS tool: the AIS tool is developed to (among others) return the entry and exit timestamps of a vessel track in the terminal polygon. Therefore, the assumptions and limitations of the tool (as discussed in subchapter 4.1.7) also apply to these service time results.
- External influences: certain weather circumstances in the port might influence the service times of the vessel. The loading and unloading process is highly dependent on the onboard and onshore facilities and (available) equipment. These influences are not registered by the AIS data.

Furthermore, in this research only three terminal types have been assessed. Conclusions have been drawn based on specific vessel classes, each defined separately per terminal type. It should be noted that only four terminals of every terminal type have been analysed. Possibly the drawn conclusions and remarks might be adjusted when additional terminals are investigated. Besides the relative small amount of terminals analysed, the size of the sub data sets can become too low. A limit for container and dry bulk vessels is set at 30 messages, and for liquid bulk terminals at 22 messages. When the sub data sets contain less than 22 or 30 messages, the results are immediately registered as unreliable. Furthermore, for data sets with a small number of messages extra precaution is taken when analysing the results. Recommendations to decrease the effects of these uncertainties are given in chapter 10.

As mentioned, the AIS tool splits the data into smaller sub data sets based on certain vessel classifications (based on vessel type). For the container terminals only specifically the container vessel classes are analysed and for the dry bulk vessels only the specifically the dry bulk vessels are investigated. Thus, a part of the arrivals are not analysed in the second part of the investigations (results per vessel class). For the container vessels the number of vessels left out of these results is not relatively large and is thus not expected to influence the conclusions much (visualised in appendix F.1.2, figures F.9, F.10, F.11 and F.12). However, for the dry bulk terminals quite a considerable part of the arrivals is left out of these second inspections (visualised in figures F.39, F.40, F.41 and F.42). This can influence the final results and conclusions made. Finally, the liquid bulk vessels are all divided into one of possible LNG classes analysed. Thus, for these terminals no vessel arrivals are excluded from this second research objective.

Finally, the distributions chosen are based on the previous literature suggestions about service time distributions for terminals. Possible other theoretical distributions might fit the data as well, or even better, but these have not been taken into account.

5.5.2. Conclusions and answers of research questions

In this chapter the answers to all sub research objectives have been given to the *Service time distribution* research objectives. The first sub research question is:

- *How are service times distributed along container-, dry bulk- and liquid bulk terminals, based on AIS data, and how do they compare to PIANC guidelines?*

First, four container terminals have been analysed based on the service time distribution (subchapter 5.1): the Rotterdam APM-2 terminal, the Rotterdam APM terminal, the Rotterdam Euromax terminal and the Le Havre Atlantic terminal. Current PIANC and UNCTAD guidelines assume the container terminals to follow an Erlang-k distribution as the service time distribution (UNCTAD, 1985). These four terminals do not show any fit between any of the Erlang-k distributions measured (Erlang-2/ Erlang-3/ Erlang-4/ Erlang-5) and the service time distributions. Comparisons between the four terminals show that the Le Havre terminal contains a more diverse and higher average service time. The more diversity is expected to be dependent on the lower number of total vessel arrivals. The higher average service time can be either due to the lower cargo handling rates of the terminal or larger amounts of cargo transferred between the terminal and vessel, both effects can not be verified because the cargo handling rates and quantity of transferred cargo are both unknown parameters in this research.

Next, four dry bulk terminals were assessed (subchapter 5.2): the Rotterdam EMO terminal, the Vlissingen OVET terminal, the Rotterdam EECV terminal and the Dunkirk Western Bulk terminal. Again UNCTAD expects an Erlang-k distribution for the terminals based on the service times (UNCTAD, 1985). For three of the four terminals no fit was found based for any of the chosen theoretical distributions. The Dunkirk terminal service time distribution was best fit by an Erlang-2 distribution. For all terminals, a large bulk of vessels contains a very short service time (<5 hours), whilst the average service times is relatively high compared to the other two terminals. This diversity, as well as the large difference between the different analysed dry bulk terminals, is due to the terminals handling different types of cargos, each using diverse cargo handling techniques.

Lastly, four liquid bulk terminals were analysed (subchapter 5.3). The scope for this research was set to inspect only LNG terminals: the Rotterdam GATE terminal, the Zeebrugge LNG terminal, the Dunkirk LNG terminal and the France Montoir terminal. PIANC and UNCTAD reports do not specifically specify what type of service time distributions should be used for LNG or liquid bulk terminals (PIANC WG121, 2014; PIANC WG158, 2014; UNCTAD, 1985). Thus again the Erlang-k distributions should fit the service times distributions. None of the four terminals fit the Erlang-k distribution, or any other fitted theoretical distribution. Based on visual results these distributions are similar to the Deterministic distribution with an average service time of roughly 24 hours. All four LNG terminals contain very similar distributions, which is expected to be based on the very specific cargo handling operations of these terminals, as well as due to the very specific vessels arriving at these terminals.

Almost none of the terminals (all types) fit to any of the theoretical distributions, which might be due to the service time distributions being dependent on too many factors (as discussed in the conclusions of subchapter 5.1.1). Therefore, the hypothesis is made that these theoretical distributions might fit better on smaller subsets where the split has been made based on the vessel class. Additionally, the hypothesis is created of the vessel class being correlated to the average service times, thereby influencing the distributions. The second sub research objective focuses on sub data sets of the terminals, and was formulated as follows:

- *How are the service times distributed per vessel class along container-, dry bulk- and liquid bulk terminals, based on AIS data?*

First, the container terminals are split into five classes. For all the classes theoretical fits have been found which represent the service time distributions. Class 1 is often fit by the Beta distribution. Class 2 by the Erlang-5 or Beta distribution. Class 3 by the Erlang-4, Gamma or Beta distribution. Class 4 by the Gamma and Class 5 by the Gamma or Beta distributions. A common factor between the terminals is the Gamma or Beta distribution. Since the Beta distribution is a normalized constant of the Gamma distribution and the Gamma distribution is an expected distribution for these kinds of processes, the Gamma distribution is selected as the best representative of the service time distributions for separate container vessel classes. Furthermore, the conclusion is made that the service times increase for higher vessel classes, which confirms this hypothesis for the container vessels (correlation between vessel class and average service time).

For the dry bulk terminals a similar conclusion is made. A lot of the sub data sets based on a specific vessel class, are fitted by theoretical distributions. However, not all the classes can be represented by all four terminals due to a lack of vessels arriving in these specific classes. For the first class the fit is either an Exponential or Beta distribution, or a Gamma distribution. The second class is mostly represented by the Gamma distribution. The third class is either fitted by an Erlang-5 distribution, or by the Beta, Gamma or Erlang-2 distribution. Class 4 is represented best by a Beta distribution and class 5 by a Gamma distribution. Again, the Gamma distribution seems to represent a lot of vessel classes for these four terminals. Additionally, as expected a higher vessel class leads to higher service times (on average).

Finally, the liquid bulk terminals are split into four classes. The first class is fit best by Erlang-k or Exponential distributions. The second class can not be analysed because no terminals receive these vessels. The third class is represented by all four terminals, of which none of them fit to any distributions. The normal distribution fits the fourth class. Visually the service time distributions of the LNG terminals lie much more around the median of the distributions, thus expected would be more of a Deterministic distribution. The majority of all of the terminals is represented by vessels from Class 3, which do not fit any distributions. Due to this majority of vessel coming from Class 3, no clear correlation can be found between the average service times between vessel classes.

The hypothesis was introduced where smaller sub data sets will obtain (better) theoretical fits, due to a reduction in the diversity of the vessel types. This hypothesis seems valid when assessing the container and dry bulk terminals. Originally, UNCTAD introduced the recommended Erlang-k distribution in 1985 for the service time distribution. It is expected that the diversity of the vessel mix arriving in 1985 was much less, compared to all the vessel types arriving at terminals nowadays. Therefore, the Erlang-k distribution might have previously been a correct recommendation, when the vessel mix was less diverse. However, nowadays the difference between the vessel types is big, which leads to a very spread out service time distribution. The service time distribution can thus better be represented as a heterogeneous data set consisting of multiple homogeneous data sets. These homogeneous data sets will represent the different vessel classes, each with their own service time distribution.

The final sub research question is based on how the three terminal types compare among each other:

- *How do the three terminal types (container, dry bulk, liquid bulk) compare based on service times?*

The most variance within the terminal types is found for the dry bulk terminals. These four terminals differ the most based on the service time distributions. The dry bulk terminals can handle different types of cargo and vessel types, leading to diverse cargo handling rates and efficiencies between the terminals. It is important to note that all four dry bulk terminals handle both importing and exporting cargo, resulting in the large differences within the service time distribution of every specific terminal. The container terminals handle only specifically containers and therefore have very similar container handling techniques and thus service time distributions. The service time distributions are similar between the LNG terminals as well, based on these terminals requiring very specific handling methods.

Furthermore, the conclusion is made that for the container and dry bulk terminals a higher vessel class leads to (on average) higher service times. The hypothesis of there being a correlation between the service times and the vessel classes is thus validated. On average, a higher vessel class (indicating a larger vessel based on vessel size) will contain more cargo which leads to more cargo transfer between the terminal and the vessel, thus requiring longer service times.

6
Results: Inter arrival time distribution

In this chapter the results are discussed for the inter arrival time distributions. The same twelve terminals will be investigated as were analysed for the service time distributions. First, the inter arrival time distributions are inspected for all twelve terminals using the entire data sets. Multiple theoretical distributions are fitted to the data and goodness-of-fit tests are performed. Next, the data is split into sub data sets based on vessel classifications, in order to test if the arrivals are independent and identically distributed. This test is necessary since the arrivals are expected to follow a Poisson process.

6.1. Container terminals: inter arrival time distribution

6.1.1. Inter arrival time distributions using all vessels

The inter arrival time has been classified as the time between two subsequent arrivals at the port. Possible theoretical fits to this distribution have been discussed in subchapter 3.2.3. First, a histogram with 100 bins is plotted, where-after, the possible theoretical distributions are fitted on the inter arrival time distributions. The same four container terminals are analysed as in chapter 5.1. Again, the Rotterdam APM-2 terminal will be discussed as an example. The other terminal analyses are demonstrated in appendix G.1.1.

The inter arrival time distribution for the Rotterdam APM-2 Terminal is plotted in figure 6.1. At first glance the inter arrival time distribution seems to follow an Exponential distribution. Based on *K-S* tests however none of the fitted distributions pass the limit (table 6.1). Visually the Exponential distribution seems to fit the distribution the best (figure 6.2).

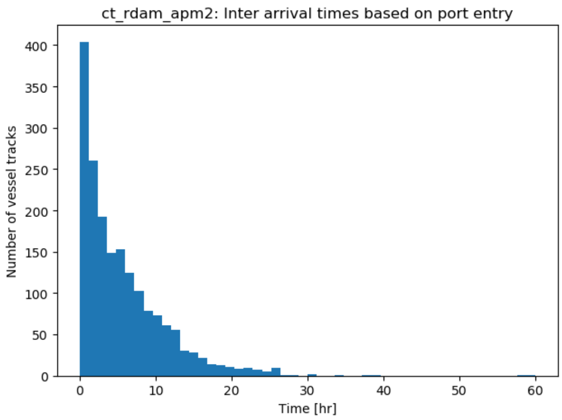


Figure 6.1: Rotterdam APM-2 Terminal inter arrival times

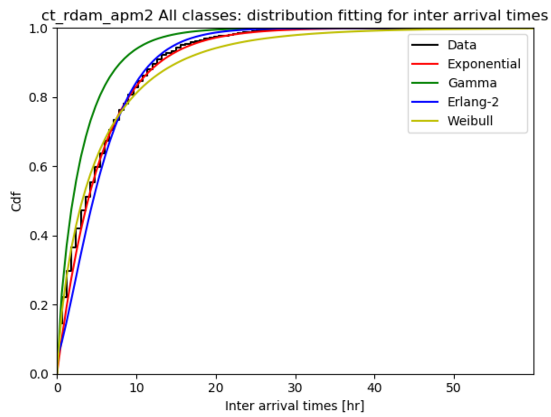


Figure 6.2: Rotterdam APM-2 Terminal with fitted distributions

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	5.61	1.00	0.07	0.00	No
Gamma	0.00	4.29	0.76	0.22	0.00	No
Erlang-2	-1.08	3.34	2.00	0.09	0.00	No
Weibull	0.00	5.12	0.77	0.07	0.00	No

Table 6.1: Inter arrival time distribution fitting for container terminal: Rotterdam APM-2

For all four terminals no theoretical distribution fits are found for the inter arrival time distributions, based on the K-S tests, as summarised in table 6.2. This contradicts the visual interpretations of all the CDFs which do show a possible fit with most often the Exponential distribution. Design guidelines would expect the inter arrival time distributions to follow Exponential distributions.

Location	Rotterdam APM2	Rotterdam APM	Rotterdam Euromax	Le Havre Atlantic
Theoretical fit	None	None	None	None

Table 6.2: Inter arrival time distribution fitting for container terminals: Best theoretical fits

The guidelines expect the Exponential distribution to correctly represent the inter arrival time distribution, thus the time between the arrivals. As introduced in subchapter 2.1.3, the arrivals of the vessels will occur in a Poisson distribution. The arrivals of vessels can be modelled as the Poisson distribution if the arrivals are *Independent and identically distributed* (*IDD*). To test if the data is IID the data should be split into smaller data sets. The sub data sets are IID when they share the same PDF and can be classified as independent events (Stephanie, 2016).

Interpretations of, and comparisons, between the container terminals

In order to compare all four terminals based on the inter arrival time distribution, for every terminal a violinplot is generated, as shown in figure 6.3. The terminal length and number of arrivals per terminal is also presented, in table 6.3. A clear correlation is found between the number of arrivals and the terminal length. Based on these four terminals a conclusion can be made that a longer terminal length will receive more vessels over the same time period. This is expected due to a longer terminal being able to service more vessels at once. Nonetheless, the number of arrivals will also depend on the terminal's geographical location. However, more arrivals will lead to shorter inter arrival times. Furthermore, the violinplot demonstrates that the shorter Le Havre Terminal has a much more variable inter arrival time distribution (difference between 25% and 75% quartiles), corresponding to the smaller number of arrivals and higher average inter arrival time.

	Terminal information		Inter arrival times [hr]			
Terminal	No. of arrivals [vessels]	Terminal length [m]	Mean	25% Quartile	Median (50%)	75% Quartile
Rotterdam APM2	1816	1500	5.61	1.42	4.06	8.04
Rotterdam APM	2108	1500	4.83	1.28	3.48	6.80
Rotterdam Euromax	3001	1900	3.39	0.88	2.33	4.74
Le Havre Atlantic	383	800	26.44	8.50	20.28	35.99

Table 6.3: Inter arrival time distribution for all container terminals

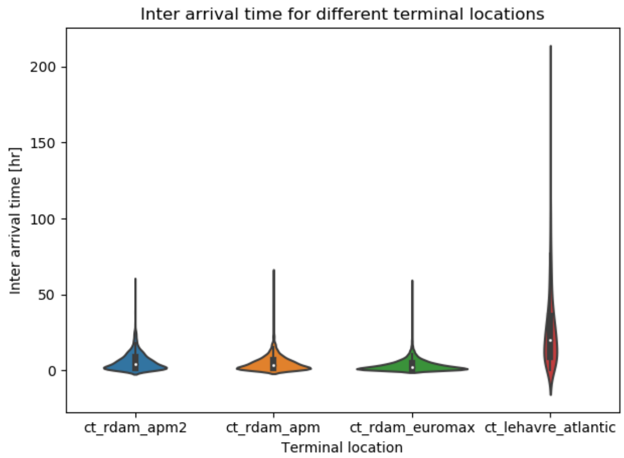


Figure 6.3: Inter arrival time, per terminal location (container terminals)

6.1.2. Inter arrival time distributions using only specific vessel classes

In order to test whether splitting the data into smaller sub sets will positively influence the distribution fitting, the data is split based on the different container vessel categories. The same classification is done as was performed for the service time distributions in subchapter 5.1.2.

Overall, splitting the data into smaller data sets has had a beneficial effect with regards to the distribution fitting to the inter arrival times. Per container class the best possible fits are as follows.

- Container Class 1: For three out of four terminals the only possible fit is the Exponential distribution. For the Le Havre Atlantic Terminal both the Exponential as well as Erlang-2 distributions are good fits to the inter arrival time data.
- Container Class 2: Two of the three terminals do not fit any distributions based on the K-S test limit. Both terminals visually do seem to fit the Exponential distribution really well. The third terminal does obtain a fit based on the test: the Exponential distribution.
- Container Class 3: For the two analysed terminals both can fit multiple distributions. The Exponential distribution is chosen as the most suitable fit in both situations.
- Container Class 4: The best fits are either the Exponential or Gamma distributions, or the Exponential or Weibull distributions. The Exponential distribution is the common distribution between all four terminals.
- Container Class 5: Two terminals return the best fit of either a Weibull or Exponential distribution, the third terminal's best fit is the Exponential or Gamma distribution.

Table 6.4 summarises the theoretical fits for every vessel class and every terminal. Not for every terminal/vessel class combination a fit was found based on the K-S tests. However, visually and for the terminals that did obtain a fit, a common distribution often found is the Exponential distribution. This corresponds with literature where both PIANC as well as UNCTAD guidelines suggest a (negative) Exponential distribution for containers (break bulk cargo) (PIANC WG158, 2014; UNCTAD, 1985). The split in the data set has confirmed that the arrivals can be classified as *IDD* since the smaller sub data sets follow the Exponential distributions. This confirmation is based on the assumption that the arrivals of vessels are independent of each other. Whilst it can occur that vessels slow down, therefore adjusting their arrival time, affecting the distribution. This possible influence is neglected in the statement that the inter arrival times are IID.

The Weibull distribution also regularly returns as a possible theoretical fit on the distribution. However, the Exponential distribution is not only more often a possible fit, it also is a better fit based on the physical expectations. As mentioned, the Weibull distribution is a distribution created to represent reliability analyses of materials or wind speeds statistics. It is therefore not a good representative of the inter arrival time distributions in this situation.

Location	Rotterdam APM2	Rotterdam APM	Rotterdam Euromax	Le Havre Atlantic
Container class 1	Exp	Exp	Exp	Exp, E-2
Container class 2	None	Exp, G	None	Unreliable
Container class 3	Exp	G, Exp	Unreliable	Unreliable
Container class 4	Exp, G	Exp, W	Exp, W	Exp, G
Container class 5	W, Exp	Exp, G	Exp, W	Unreliable

Table 6.4: Inter arrival time distribution fitting for container terminals: Best theoretical fits, per vessel class (*W = Weibull, E- = Erlang-, G = Gamma, Exp = Exponential*)

Finally, the difference in between the classes of every terminal for the inter arrival times is analysed based on the following four figures. Where, for the service times a very obvious increase was registered between the classes, here the difference between the classes is nil. Except for the Le Havre Atlantic Terminal where some variance is observed between the classes. However, these results are unreliable due to the small amount of vessel arrivals these classes have (figure F.12). This nil difference between the vessel classes corresponds to literature suggesting the arrivals of vessels being completely independent and stochastic of each other (more information in subchapter 2.1.3). For container terminals specifically PIANC suggests the random arrival process might be too conservative based on improved scheduling at these terminals (PIANC WG158, 2014). However, for the terminals analysed in this research this possible more improved scheduling is not visible in the visualisations of the inter arrival time distributions. It must be noted that besides scheduling, the arrival time is dependent on other (external) factors, such as weather influences or engine failures (Bellsolà Olba et al., 2018).

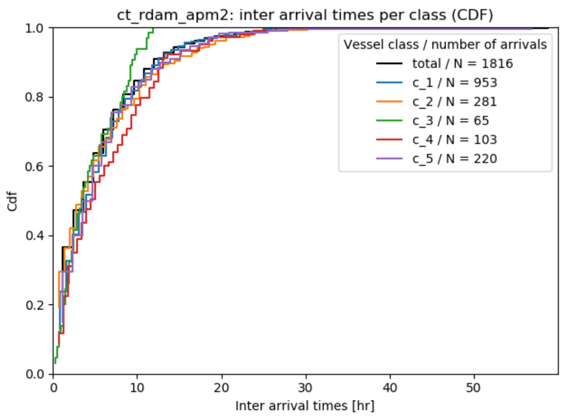


Figure 6.4: Inter arrival time distribution per class Rotterdam APM-2

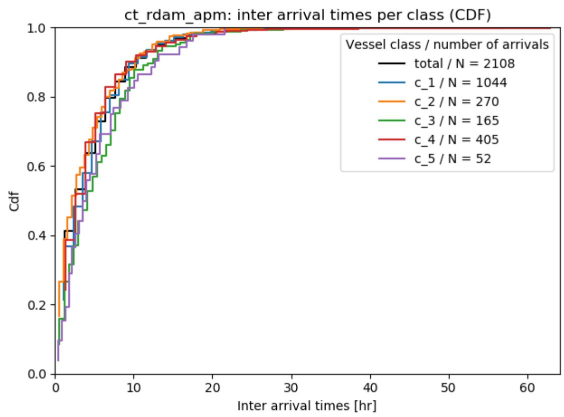


Figure 6.5: Inter arrival time distribution per class Rotterdam APM

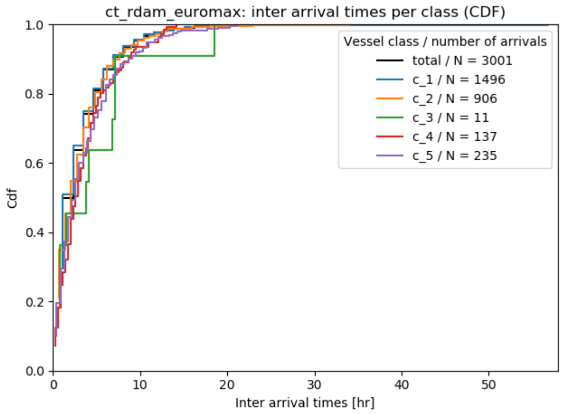


Figure 6.6: Inter arrival time distribution per class Euromax

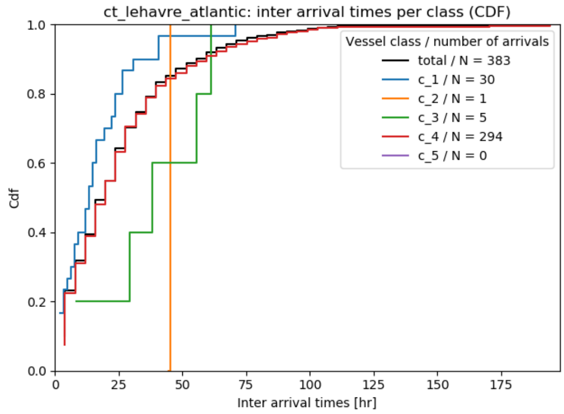


Figure 6.7: Inter arrival time distribution per class Le Havre Atlantic

6.2. Dry bulk terminals: inter arrival time distribution

6.2.1. Inter arrival time distributions using all vessels

For the four dry bulk terminals the same steps will be performed to find the best fits for the inter arrival time distributions. The four terminals examined are the same as previously analysed terminals in subchapter 5.2.1. The results of analysing the inter arrival time distributions for the dry bulk terminals are thoroughly demonstrated in appendix G.2.1. A summary of the possible theoretical distributions is given in the table below.

Location	Rotterdam EMO	Vlissingen OVET	Rotterdam EECV	Dunkirk Western Bulk
Theoretical fit	W, Exp	G	Exp	Exp, G

Table 6.5: Inter arrival time distribution fitting for dry bulk terminals: Best theoretical fits (*G = Gamma, W = Weibull, Exp = Exponential*)

Unlike the container terminals having no fits on their total data sets, the dry bulk terminals have theoretical distributions that correctly fit the data. The Rotterdam EMO terminal inter arrival time is best represented by the Weibull or Exponential distribution, whilst the Vlissingen OVET terminal is best fitted by the Gamma distribution. For the Rotterdam EECV terminal the Exponential distribution and for the Dunkirk Western Bulk Terminal either the Exponential or Gamma distributions fit.

The Exponential distribution is expected to represent the inter arrival times for the dry bulk terminals. For three of the four terminals analysed this expectation corresponds to the observed distributions. However, also the Gamma distribution has been found to represent two of the terminals. Again, the data set will be split to further inspect the distributions of the inter arrival time and validate the assumption of the data being *IDD*.

Interpretations of, and comparisons between, the dry bulk terminals

The inter arrival time distribution is compared between all four terminals. Figure 6.8 shows that the Vlissingen OVET and the Dunkirk Western Bulk both contain much more variable inter arrival times. This corresponds with the smaller number of vessel arrivals (table 6.6). For the container terminals a correlation was found between the terminal lengths and the number of vessel arrivals at the terminal. This correlation is less obvious for the dry bulk terminals. The number of vessel arrivals is expected to be dependent on much more than just the vessel length (such as location of the port and weather influences). However, the more spread out inter arrival time distributions do correspond to the smaller number of total vessel arrivals. The two terminals with the smaller number of vessel arrivals (Vlissingen OVET and Dunkirk Western Bulk) also obtain a much higher average service time. This makes sense as less vessel arrivals immediately leads to higher average inter arrival times.

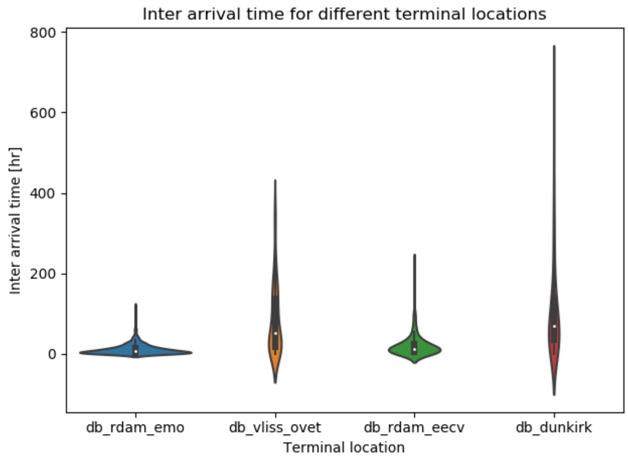


Figure 6.8: Inter arrival time, per terminal location (dry bulk terminals)

	Terminal information		Inter arrival times [hr]			
Terminal	No. of arrivals [vessels]	Terminal length [m]	Mean	25% Quartile	Median (50%)	75% Quartile
Rotterdam EMO	907	2700	11.23	2.61	7.54	15.41
Vlissingen OVET	127	950	79.83	12.86	48.15	127.75
Rotterdam EECV	514	1090	19.80	4.61	12.55	24.23
Dunkirk Western Bulk	94	675	107.38	34.51	69.86	136.48

Table 6.6: Inter arrival time distribution for all dry bulk terminals

6.2.2. Inter arrival time distributions using only specific vessel classes

The same split into smaller data sets is made for the dry bulk terminals (as was introduced in subchapter 5.2.2). Again, only specifically the dry bulk vessel classes are analysed. Results for all the terminals are demonstrated in appendix G.2.2. The following summary can be made for the inspected vessel classes:

- Dry bulk Class 1: Two terminals are analysed. The Rotterdam EMO terminal returns a nicely fitted Exponential or Weibull distribution. The Vlissingen OVET terminal is fitted by the Gamma distribution, but visually less suitable.
- Dry bulk Class 2: All the terminals have one or two possible distributions that fit their data sets. The recurring distribution fitted on all the data is the Exponential distribution.
- Dry bulk Class 3: The two terminals analysed are fitted by the Exponential distribution. One of the terminals could also be represented by the Weibull distribution.
- Dry bulk Class 4: For one of the two terminals analysed the data is almost perfectly represented by an Exponential or Erlang-2 distribution. The other terminal has a less robust fit visually, however a Gamma distribution is a possible fit.
- Dry bulk Class 5: The best fit on the (only analysed) EECV Terminal is the Exponential or Gamma distribution.

The results for all the specific terminals are given in table 6.7. A common factor between all of the classes is that often the Exponential distribution is a fit. However, the results are quite diverse and many data sets are unreliable based on the small amount of data points. Therefore, the conclusions about the Exponential distribution are not fully robust and must be taken into account with caution.

Location	Rotterdam EMO	Vlissingen OVET	Rotterdam EECV	Dunkirk Western Bulk
Dry bulk class 1	Exp, W	G	Unreliable	Unreliable
Dry bulk class 2	Unreliable	Exp	Exp, G	Exp, E-2
Dry bulk class 3	Exp	Unreliable	Exp, W	Unreliable
Dry bulk class 4	Exp, E-2	Unreliable	G	Unreliable
Dry bulk class 5	Unreliable	Unreliable	Exp, G	Unreliable

Table 6.7: Inter arrival time distribution fitting for dry bulk terminals: Best theoretical fits, per vessel class (*W = Weibull, E- = Erlang-, G = Gamma, Exp = Exponential*)

The inter arrival time distribution per vessel class, for every terminal, is plotted below in the next four figures. Again, no noteworthy differences between the different dry bulk vessel classes are present. Distributions that stand out are the vessel classes with very little vessel arrivals and can therefore be neglected. UNCTAD presents the possibility of the dry bulk terminals having a more smooth Erlang-2 inter arrival time distribution (UNCTAD, 1985). However, the tendency towards some amount of scheduling is counteracted by the influence of external factors leading to the fully stochastic arrival processes as observed at these terminals.

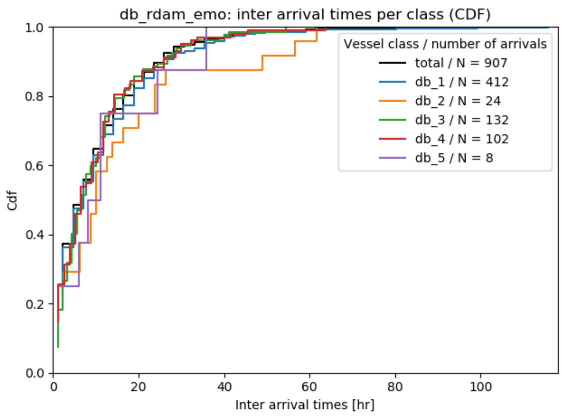


Figure 6.9: Service time distribution per class Rotterdam EMO

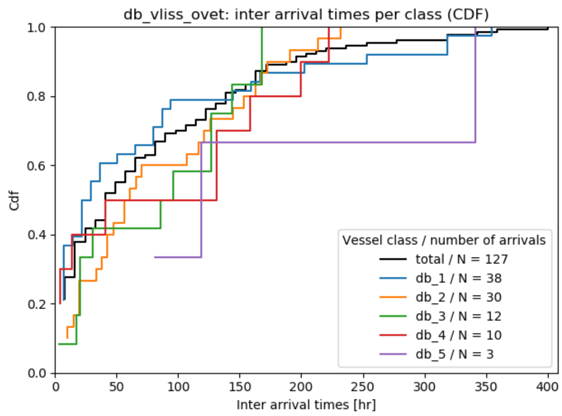


Figure 6.10: Service time distribution per class Vlissingen OVET

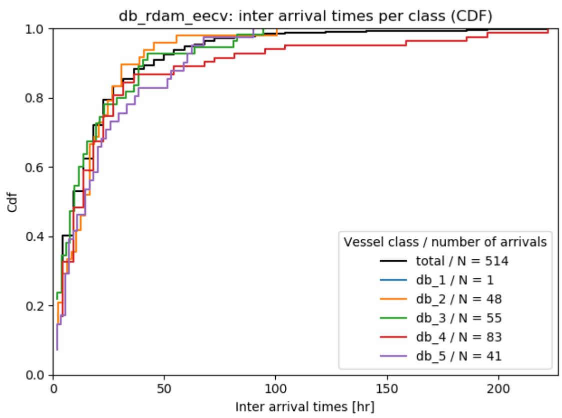


Figure 6.11: Service time distribution per class Rotterdam EECV

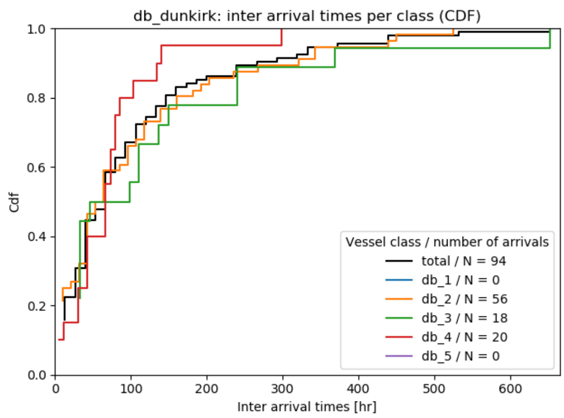


Figure 6.12: Service time distribution per class Dunkirk Western Bulk

6.3. Liquid bulk terminals: inter arrival time distribution

Four liquid bulk terminals have been assessed based on their service time distribution in subchapter 5.3. The same four terminals will be assessed based on their inter arrival time distribution to investigate how these are distributed across four similar LNG terminals.

6.3.1. Inter arrival time distributions using all vessels

The four terminals are thoroughly analysed and discussed in appendix G.3.1. For all four terminals the inter arrival time distribution was fitted by one or more theoretical distributions, based on visual interpretations and K-S tests. Three terminals were best fit by the Exponential distribution and the Dunkirk LNG terminal was best fit by the Erlang-2 distribution, as summarised in the table below.

Location	Rotterdam Gate	Zeebrugge	Dunkirk	France - montoir
Theoretical fit	Exp	Exp	E-2	Exp

Table 6.8: Inter arrival time distribution fitting for liquid bulk terminals: Best theoretical fits (*E- = Erlang-, Exp = Exponential*)

UNCTAD expects specialized terminals (and thus these LNG terminals) to follow an Erlang-2 distribution for the inter arrival times (UNCTAD, 1985). Whilst, other literature suggest the distribution is best represented as a Poisson process: an exponential distribution (Van Asperen et al., 2003). Both expectations are found in the observed data sets. Three terminals follow an Exponential distribution and one terminal an Erlang-2 distribution. At the Zeebrugge LNG terminal the Erlang-2 distribution represents a smoother inter arrival time distribution, possibly due to more terminal scheduling and less influences from external factors.

Interpretations of, and comparisons between, the liquid bulk terminals

From table 6.9 a smaller number of berths seems to be correlated to less vessels arriving, leading to higher average inter arrival times. Figure 6.13 visualises the inter arrival time distributions for all the analysed LNG terminals. The Dunkirk terminal stands out by having a more diverse range of inter arrival times. This is not unexpected as the terminal has the least number of vessel arrivals, leading to higher average inter arrival times.

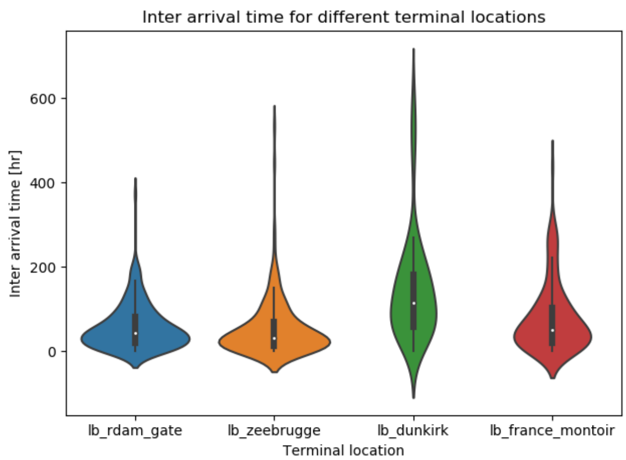


Figure 6.13: Inter arrival time, per terminal location (liquid bulk terminals)

	Terminal information		Inter arrival times [hr]			
Terminal	No. of arrivals [vessels]	No. of berths	Mean	25% Quartile	Median (50%)	75% Quartile
Rotterdam GATE	173	2	57.94	18.55	44.00	81.52
Zeebrugge	185	2	54.49	12.22	31.42	69.06
Dunkirk	68	1	147.54	56.97	114.16	181.37
France-Montoir	121	2	79.95	20.06	50.19	103.17

Table 6.9: Liquid bulk terminals: number of berths and number of arrivals

6.3.2. Inter arrival time distributions using only specific vessel classes

As performed for previous terminal types, the data is again split into smaller sub sets. The K-S test is performed and the best theoretical fits will be chosen. As discussed in subchapter 5.3.2 three LNG vessel classes are analysed: LNG1, LNG3 and LNG4. Results for these goodness-of-fit tests and analyses are discussed in appendix G.3.2. Per vessel class the following remarks are made:

- LNG Class 1: The first class can be represented by an Exponential distribution in both cases (two terminals were analysed).
- LNG Class 2: No terminals analysed that received this vessel class.
- LNG Class 3: A common distribution between all four terminals assessed is the Exponential distribution. For one terminal also the Weibull distribution was a good fit, and for another terminal the Erlang-2 was also a very suitable distribution.
- LNG Class 4: The Exponential distribution fits best for the Zeebrugge LNG terminal (the only terminal analysed).

The results are summarised in table 6.10. A very clear relationship is found between all the LNG vessel classes. The Exponential distribution fits for almost all analysed terminals. In some cases either the Weibull or the Erlang-2 distributions also present a suitable fit. As mentioned before, the Weibull distribution is unexpected based on the physical properties behind this distribution. Furthermore, the Exponential and Erlang-2 distributions are not unexpected as these are the expected distributions based on current design guidelines (as discussed in subchapter 2.1.3).

Location	Rotterdam Gate	Zeebrugge	Dunkirk	France - montoir
Liquid bulk class 1	Exp	Exp, G	Unreliable	Unreliable
Liquid bulk class 2	-	-	-	-
Liquid bulk class 3	Exp, E-2	Exp, W	E-2, Exp	Exp
Liquid bulk class 4	Unreliable	Exp	Unreliable	-

Table 6.10: Inter arrival time distribution fitting for liquid bulk terminals: Best theoretical fits, per vessel class (*E- = Erlang-, G = Gamma, Exp = Exponential*)

The different inter arrival time distributions per vessel class are plotted in the following figures. Even though a lot of the vessel classes in terminals do not contain many vessel tracks, the inter arrival time distribution is very similar between the classes of every terminal. Outliers include vessel classes with very little amount of arrivals, which therefore can be neglected. The same conclusions can be drawn as were made for the inter arrival time distributions for the container and dry bulk terminals. There is no significant difference between the inter arrival times of the different vessel classes. Thus the inter arrival time is not dependent on the type of the vessel. Furthermore, the distributions all represent the same PDF (Exponential) thus are expected to be *IDD*.

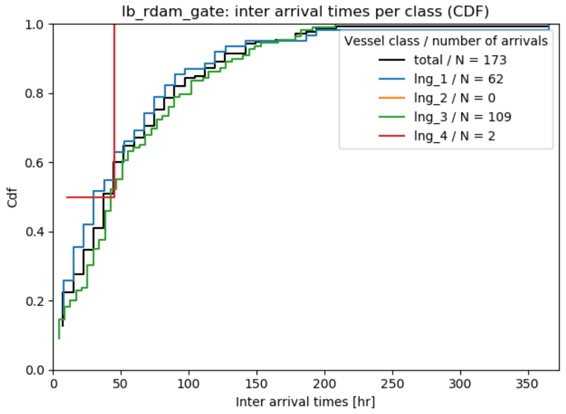


Figure 6.14: Inter arrival time distribution per class Rotterdam Gate

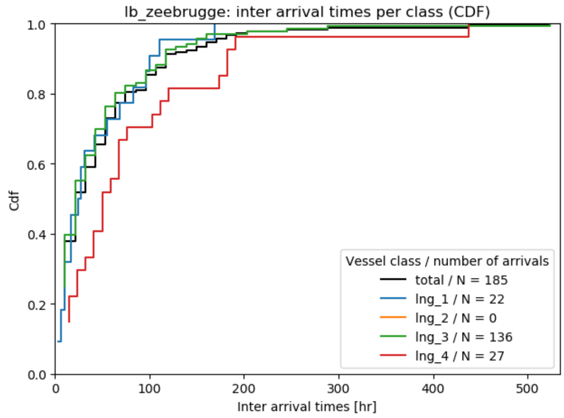


Figure 6.15: Inter arrival time distribution per class Zeebrugge LNG

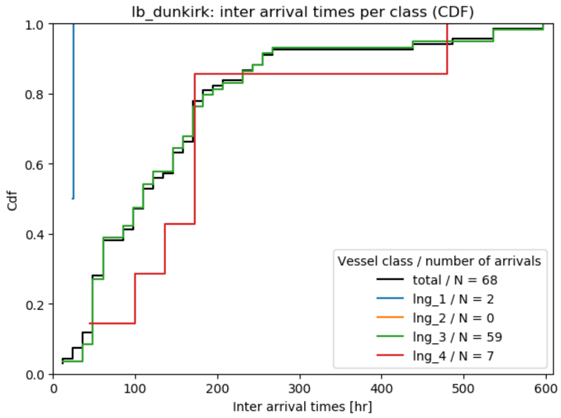


Figure 6.16: Inter arrival time distribution per class Dunkirk LNG

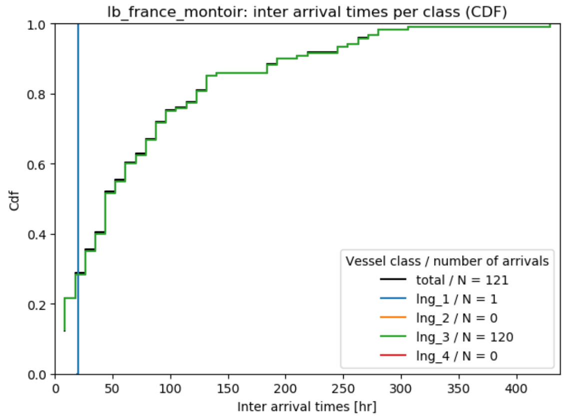


Figure 6.17: Inter arrival time distribution per class France Montoir

6.4. Discussion and conclusions: inter arrival time distributions

First, all results regarding the inter arrival time distributions are discussed. Second, the sub research questions are answered with regards to these inter arrival time distributions.

6.4.1. Discussion and limitations of the inter arrival time distribution results

The results discussed in this previous chapter all depend on the definition of the inter arrival time, which has been defined as:

- Inter arrival time = The time between two successive arrivals of vessel at a port (PIANC WG121, 2014).

For these results, the inter arrival time is largely dependent on the following:

- The port polygon: the port polygon is defined as the total port area, at which the outer boundary (at sea side) is located just past the anchorage area(s). This port polygon is manually defined by the user of the AIS tool.
- The AIS data: the AIS data is assumed to represent the reality. For the service time an important limitation was that errors such as vessels which turn of their AIS signals too soon, are not taken into account and thus will be present in the current data sets. For the inter arrival times this limitation is expected to influence the results less due to the inter arrival time being dependent on the port entry and exit times,

not the terminal entry and exit times.

- The AIS tool: as discussed in subchapter 5.5.1 the AIS tool is developed to (among others) return the entry and exit timestamps of a vessel track in the port polygon. Therefore, the assumptions and limitations of the tool also apply to these service time results.
- External influences: certain weather circumstances at sea might influence the arrival processes at ports. Also, it is common for vessels to reduce speed when they are aware of a terminal being fully occupied. These effects are not registered in the AIS data.

Additionally, the focus of this research lies on three different terminal types: containers, dry bulk and liquid bulk. Different conclusions could be possible when different terminal types are assessed. Furthermore, only four terminals of every terminal type are analysed, leading to always some extent of uncertainty when conclusions are drawn.

As discussed in subchapter 5.5.1 the AIS tool splits the data into smaller sub data sets based on certain vessel classifications (based on vessel type). Following, a part of the arrivals are not analysed in the second part of the investigations (results per vessel class) due to these arrivals not belonging to the specific analysed vessel classes. The same considerations made for the service times are valid here. For the container vessels the number of vessels left out of these results is not relatively large and is thus not expected to influence the conclusions much. However, for the dry bulk terminals quite a considerable part of the arrivals is left out of these second inspections, this can influence the final results and conclusions made. Finally, for the liquid bulk terminals no vessel arrivals are excluded from this second research objective, thus no influences will be present for this matter on the results and conclusions.

Finally, the goodness-of-fit test was performed on four different theoretical distributions. These distributions were chosen based on current design guidelines of terminals and previous research & literature available. Moreover, the Weibull distribution was occasionally found to be the best fit to the inter arrival time distribution. However, as discussed in subchapter 2.1.3, the Weibull distribution physically is based on (among others) the reliability applications of testing material strengths. Whilst the Erlang-k and Exponential distributions have been previously linked to queuing theory by literature. Therefore, the assumption is made that the Weibull distribution might be a good fit on some specific vessel class/terminal combinations, however it is not as suitable to represent the inter arrival time distribution.

6.4.2. Conclusions and answers of research questions

In this previous chapter all results regarding the research objective *Inter arrival time distribution* have been presented. Two sub research questions were defined, which based on these results, can now be answered. The first research question is:

- *How are inter arrival times distributed along container-, dry bulk- and liquid bulk terminals, based on AIS data, and how do they compare to PIANC guidelines?*

The three different terminal types were separately analysed. First, four container terminals have been investigated. Design guidelines expect the inter arrival time distributions of break bulk cargo to follow an Negative Exponential Distribution (NED), however some notes have been made about improved port terminal planning leading to more constant arrivals (more towards Erlang-2 distributions) (PIANC WG158, 2014; UNCTAD, 1985). When fitting the possible theoretical distributions on the observed data none of the four terminals have any good fits based on the K-S tests. This contradicts the visual interpretations of all the CDFs which do show a possible fit with most often the Exponential distribution. Thus the Exponential distribution is the best representative distribution for the inter arrival times. The suggestions about more improved scheduling leading to less random patterns of arrivals is not visible at these terminals, as they are best represented by Exponential distributions. It must be noted that besides scheduling, the arrival time is dependent on other (external) factors, such as weather influences or engine failures.

The dry bulk terminals are thus also expected to follow a NED shape. Again four terminals are analysed.

All four terminals have possible distributions that fit on the inter arrival time data. The Exponential distribution is a possible fit on three out of four terminals. The liquid bulk terminals can be classified as specialised terminals, as all four terminals analysed are specific LNG terminals. UNCTAD states that the inter arrival time distribution for specialised terminals often follow the Erlang-2 distribution (UNCTAD, 1985). Whilst Van Asperen again expects the liquid bulk terminals to follow the (Negative) Exponential distribution (Van Asperen et al., 2003). For the four terminals analysed all of the inter arrival time distributions had possible theoretical fits. Three terminals are best represented by a Exponential distribution, whilst one terminal has a best fit on the Erlang-2 distribution. The terminal with the Erlang-2 distribution might contain this less stochastic arrival pattern due to more terminal scheduling or less influences from external factors.

For all three terminal types a strong correlation was observed between the number of arrivals and the average inter arrival times. It makes sense that a higher number of arrivals will lead to (on average) less time between two successive arrivals, thus returning a lower average inter arrival time. Furthermore, for the container terminals a longer terminal corresponded to more arrivals and for liquid bulk terminals more berths corresponded to more arrivals as well. However, the number of arrivals is not only expected to dependent on the terminal size (length/ number of berths), but also on external influences such as the geographical location. Finally, no remarkable difference is found between the terminal types as they mostly all contain the Exponential inter arrival time distribution leading to very similar plots and figures. To repeat, the number of arrivals seems to have the largest impact on the differences between the distributions.

Similar to the split made for the service time distributions, the data is split again into smaller sub data sets. The same vessel classifications are used. Since the Exponential distribution is expected as the inter arrival time distribution, a split is expected to result in again Exponential distributions, but for smaller data sets. The arrivals can be modelled as a Poisson process if the arrivals are *Independent and identically distributed* (*IDD*). To test if the data is IID the data should be split into smaller data sets. When they share the same PDF and are independent events the distribution can be classified as IID. Therefore, the second sub research question is:

- *How are the inter arrival times distributed per vessel class, along container-, dry bulk- and liquid bulk terminals, based on AIS data?*

First, the container terminals are assessed and every terminal data set is split into five different data sets, representing each one specific container vessel class. The most terminals and classes obtained a theoretical fit based on the Exponential distribution. Not specifically for every terminal and every class a theoretical fit was found. However, for the data sets where no fit was found based on the K-S test, often based on visual results the Exponential distribution would fit. For all the dry bulk terminals the results were more diverse. A common factor between all the separate classes was that the Exponential distribution was often a very good fit. Finally, the liquid bulk terminals again showed the most classes and terminals fitting best by an Exponential distribution. However, in some cases the Weibull or Erlang-2 distribution would also be a very good fit. Nonetheless, the Weibull distribution is a distribution originally created based on material strengths test and often used in wind speed calculations. As discussed in subchapter 2.1.3, the Weibull distribution is actually an unexpected distribution to represent the inter arrival times.

To conclude, the inter arrival time distribution is very often fitted by the Exponential distribution, as proposed in PIANC and UNCTAD guidelines. However, the guidelines suggest the distribution to fit to the entire data, and do not specify distributions for smaller data sets. Since the inter arrival times for all the sub data sets follow the Poisson process, leading to an Exponential inter arrival time distribution, the data can be classified as IID. Nonetheless, this confirmation is based on the assumption that the arrivals of vessels are independent of each other. Whilst it can occur that vessels slow down, therefore adjusting their arrival time, affecting the distribution. This possible influence is neglected in the statement that the inter arrival times are IID. All in all, the inter arrival time distribution of the assessed terminals are best represented by an Exponential distribution.

7
Results: Berth occupancy

This chapter focuses on the occupancy of different terminals. Independently, four container, four dry bulk and four liquid bulk terminals will be analysed based on their occupancy. Consequently, conclusions are drawn about the different terminal types and their occupancies.

7.1. Occupancy for container terminals

7.1.1. Berth occupancy for container terminals

Container terminals are first inspected based on their berth occupancy. The berth occupancy is defined as follows:

- Berth occupancy = the time that a berth is physically occupied by a vessel, relative to the total number of operating hours of the terminal.

Two different steps are necessary in order to find the berth occupancy of terminals. First, the number of berths should be known. Second, the location of these berths should be defined. For container terminals both of these steps raise complications. Often a container terminal is defined by a long quay wall where a variable amount of vessels can berth at the same time. Onshore equipment, such as gantry cranes, are in such a way flexible that often they can move alongside the quay to assist the (un)loading process at any location.

Multiple methods have been investigated to see if it is possible to find precise number and locations of berths. More results on these methods are given in appendix H, where the Vlissingen OVET Terminal is taken as an example. Below in figure 7.1 an example is given of a container terminal with the central location of every vessel track, visualised using a heat map. Thus, the conclusion is made that for container terminals the specific **berth** occupancy is not possible to find.

Figure 7.1: Rotterdam APM Terminal heat map containing every vessel track

7.1.2. Length occupancy for container terminals

A different option to gain insight into the occupancy of the terminal is to use the vessel lengths, compared to the entire length available at the terminal. Therefore, the *LOA* is necessary and the entire length of the terminal quay wall. As mentioned in subchapter 3.2.2 the total occupied length is defined as the length of all ships present at a certain time, including the 15 meter range between every vessel present. Figure 7.2 visualises how a berth and adjusted length occupancy are calculated. Following, the occupancy is averaged over the selected time span to generate the average occupancy.

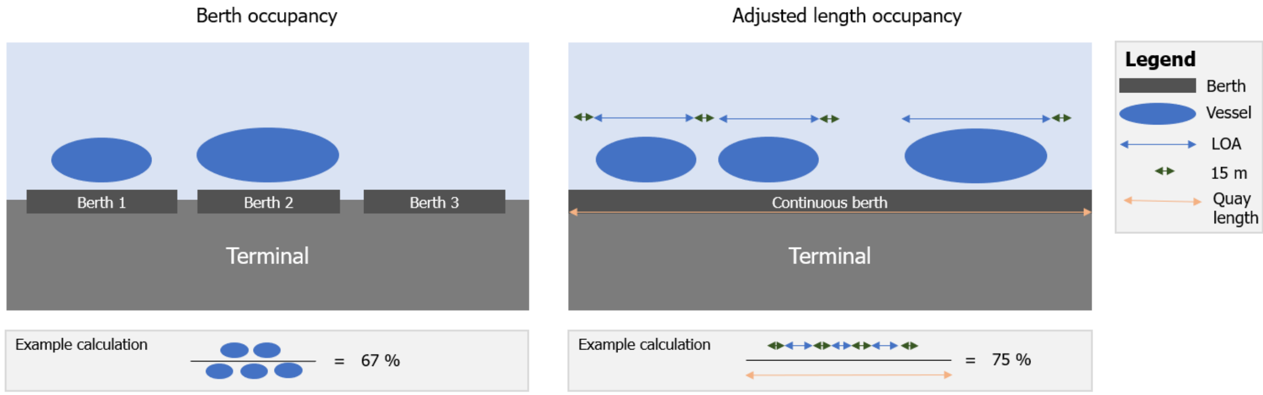


Figure 7.2: Berth- and adjusted length occupancy

The four terminals analysed are the same as analysed for the service time and inter arrival time distributions. The Rotterdam APM-2 terminal occupancy is visualised in figure 7.3, the other three terminals are visualised in appendix I.1 (figures I.2, I.3, I.4). All visualisations of these four terminals are based on the adjusted length occupancy. This means the 15 meter range between every vessel is included. None of the terminal occupancies based on length pass the 100% limit. The largest average occupancy is found for the Rotterdam Euromax terminal, with an adjusted occupancy of 40.01%. The lowest occupancy is for the Le Havre Atlantic Terminal, with an adjusted occupancy of 23.05%. The relatively low average occupancy is expected to be caused by the fact that container terminals often tolerate only a little amount of waiting time, leading to longer quay lengths in the design process, based on the smaller design occupancy numbers.

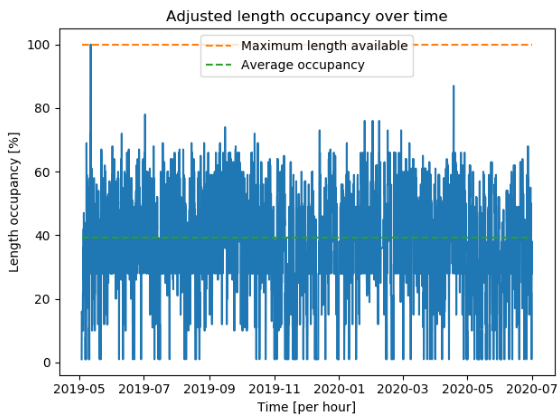


Figure 7.3: Rotterdam APM-2: Adjusted length occupancy

As discussed in subchapter 2.1.1 and 2.1.3 container vessels set stringent conditions about the maximum waiting times. The typical average waiting time in terms of service time is set at 10% (PIANC WG 184, 2019). The typical berth occupancy of 35% is often applied to these terminals (Ligteringen, 2017). However, the berth occupancy for the analysed terminals can not be determined based on the inconsistent number of berths over time. For these container terminals the adjusted length occupancy has been defined, and thus not the berth occupancy. Therefore, comparisons between the theoretical recommendations for the berth occupancy and the actual observed length occupancy are not 100% reliable. Nonetheless, the average length occupancy is observed to be between 23 and 40%, which is similar to the recommended 35%.

7.1.3. Occupancy comparisons between container terminals

The four length occupancies are compared, as shown in figure 7.4. The first three analysed terminals are very similar as regards to their mean and median values. The fourth terminal, the Le Havre Atlantic terminal has an obvious different shape. This terminal also has a relatively smaller quay length, which might influence this occupancy. Figure 7.5 shows the relation between the terminal length and the average adjusted terminal occupancy. For these four terminals the conclusion can be drawn that a higher terminal length will lead to a higher average occupancy. This is expected, as in design steps the occupancy can be increased when there are more servers present, as the terminal is then expected to be more flexible. In other words, with a higher number of berths, a terminal can obtain a higher occupancy rate with the same amount of maximum waiting time in terms of service time.

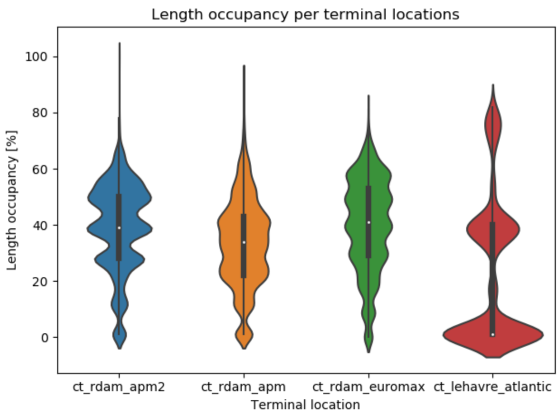


Figure 7.4: Container terminals: occupancy

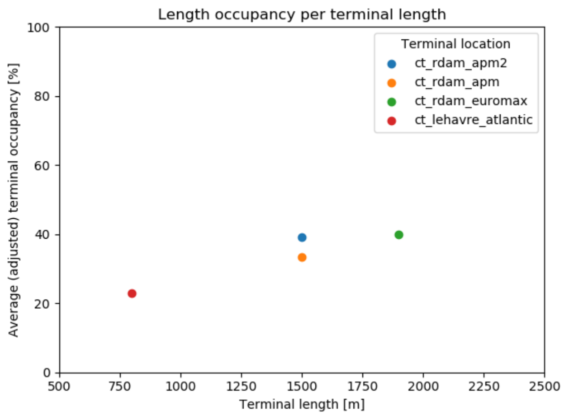


Figure 7.5: Container terminals: terminal length versus occupancy

7.2. Occupancy for dry bulk terminals

7.2.1. Berth occupancy for dry bulk terminals

As discussed in the previous subchapter (7.1.1), the berth occupancy is difficult to determine for container and dry bulk terminals. The example of the Vlissingen OVET terminal is extensively discussed in appendix H. Therefore, the same approach will be taken for the dry bulk terminals, as was taken for the container terminals.

7.2.2. Length occupancy for dry bulk terminals

For all four dry bulk terminals, as previously analysed, the adjusted length occupancy is defined per hour. Figure 7.6 represents the Rotterdam EECV adjusted length occupancy over time. In appendix I.2 figures I.5, I.6 and I.8 represent the Rotterdam EMO, Vlissingen OVET and Dunkirk Western Bulk terminals. The largest average occupancy is found for the Rotterdam EECV terminal, with an average adjusted occupancy of 53.86%. The lowest occupancy is found for the Vlissingen OVET terminal, with an average adjusted occupancy of 14.16%. More comparisons between the dry bulk terminals are made in the next subchapter.

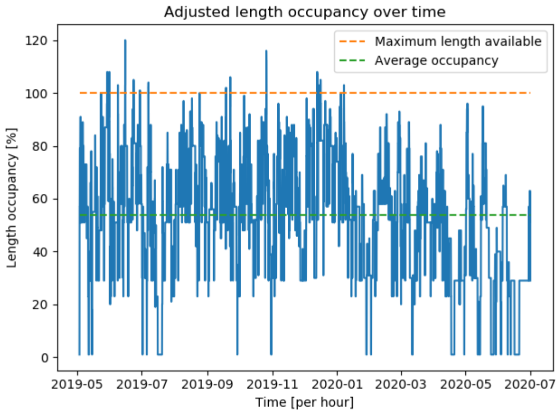


Figure 7.6: Rotterdam EECV: terminal occupancy

As previously mentioned, the dry bulk vessels are known for accepting longer waiting times in terms of service times, compared to the container vessels (Ligteringen, 2017). The average waiting time in terms of average service time is often set at 30% in terminal designs (PIANC WG 184, 2019). However, again the comparison between the berth occupancy and the adjusted length occupancy is not completely correct. Both occupancy measures focus on different aspects of the terminal. However, the occupancy of the dry bulk terminals are expected to be higher than the container terminals (between 40 and 80% dependent on the number of berths), based on the assumption that dry bulk vessels tolerate longer waiting times (Kox, 2016). This does not necessarily seem to be the case for these four dry bulk terminals, compared to the four container terminals.

Inspection of occupancy outliers

The Rotterdam EECV terminal's occupancy passes the 100% a few times. This could occur based on various reasons, therefore the terminal is further analysed. The maximum moments of an occupancy above 100% are inspected. On the 15-06-2019 between 12:59:12 and 13:59:12 the highest occupancy was measured: 120%. The second and third highest occupancies (116% and 115%) both were also only present for one hour. These outliers are expected to be due to the definition of the service time. The service time, as defined for in this research, is the time at the berth including the (un)berthing of the vessel. Therefore, a vessel might have actually already left the quay, but still be present in the terminal. Other reasons could be based on possible errors in the AIS tool (most likely in the extracting of berthed vessel tracks), or based on errors in AIS data.

A longer period of occupancy > 100% is examined. Again, the 15-06-2019 is examined where between 11 AM and 2 PM an occupancy of 108% was measured. Between this time 5 vessels were berthed at the terminal: 3 dry bulk vessels and 2 inland waterway tankers. Comparing these 5 vessel tracks with Sea-web data confirms that Sea-web also had 5 vessels registered at roughly the same time spans (figure 7.7).

mmsi	loa	DWT	type	breadth	teu_capacity	port_entry_time	port_exit_time	terminal_entry_time	terminal_exit_time
477067300	292.00	180958.0	Bulk Dry	45.00	0.0	2019-05-23 23:29:57	2019-05-30 19:28:34	2019-05-24 05:02:15	2019-05-30 15:00:37
566662000	300.00	208025.0	Bulk Dry	50.00	0.0	2019-05-26 13:52:10	2019-06-01 09:03:09	2019-05-26 18:54:01	2019-06-01 03:54:21
241237000	299.92	205888.0	Bulk Dry	50.05	0.0	2019-05-27 01:09:34	2019-06-01 12:02:25	2019-05-27 06:04:35	2019-06-01 06:47:24
205542590	81.30	4207.0	Inland Waterways Tanker	9.46	0.0	2019-05-27 16:55:14	2019-06-02 17:54:01	2019-05-29 03:55:47	2019-05-29 14:42:21
244660629	124.95	3464.0	Inland Waterways Tanker	11.40	0.0	2019-05-27 22:47:24	2019-05-30 05:49:03	2019-05-29 10:10:52	2019-05-29 16:50:46

Figure 7.7: Five vessel tracks present at the moment of maximum occupancy (108%)

Summarising, the surpassing of the occupancy can be caused by several factors:

- Errors due to the time span of every data point. The length of the vessels present is calculated by determining the total length of vessels present per hour. In other words, an overlap is possible of two vessels, both present in the same 'hour interval', but leaving/entering after each other, thus not present at the terminal at the same time.
- Errors in the AIS data
- Errors in the AIS tool: wrongfully extracting vessels that berth.
- Vessels berthing parallel to other vessels, for example for bunkering activities. These vessels parallel can also be wrongfully classified as berthed by the terminal.
- An incorrectly defined terminal length. The terminal length might actually be longer, due to incorrect measurements or inputs.
- The 15 meter margin is not correct. The assumed 15 meter margin between vessels might actually be smaller.

For the example given in figure 7.7 the five vessels present were also registered by Sea-web, thus the first and second would not hold.

7.2.3. Occupancy comparisons between dry bulk terminals

Comparing the occupancies between the terminals leads to the following figure (figure 7.8). The four terminals all have different occupancy distributions, with very different median, as well as very different maximum, occupancies. The length is plotted against the average adjusted length occupancy in order to find any correlations (figure 7.9). The conclusion that for container terminals a longer terminal leads to a higher average occupancy, does not hold for the dry bulk terminals. The occupancy is dependent on much more than just the terminal length, such as the port location, and thus no clear (linear) relationship is found between these two based on the dry bulk terminals.

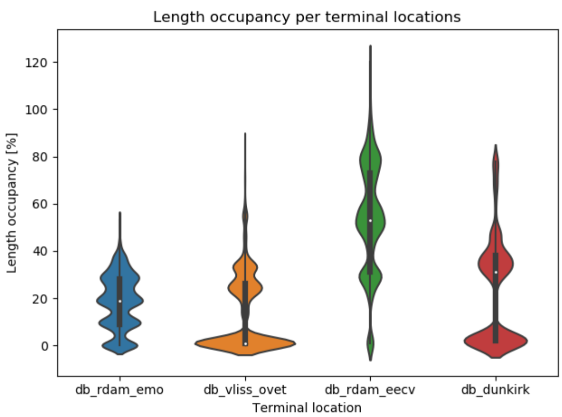


Figure 7.8: Dry bulk terminals: occupancy

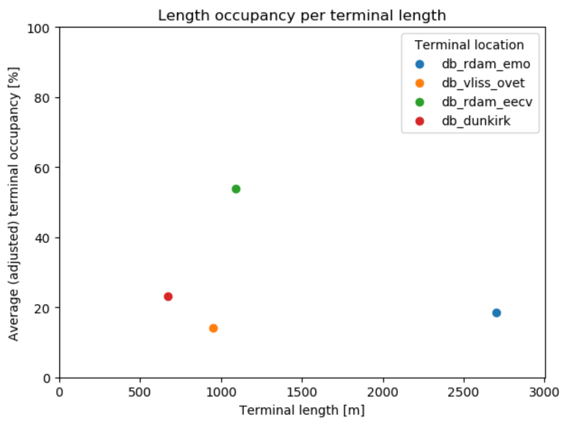


Figure 7.9: Dry bulk terminals: terminal length versus occupancy

7.3. Occupancy for liquid bulk terminals

7.3.1. Berth occupancy for liquid bulk terminals

For the chosen liquid bulk terminals the number and location of berths is very clear. The Rotterdam GATE, Zeebrugge LNG and France Montoir LNG terminals all contain 2 berths, whilst the Dunkirk LNG terminal only contains 1 berth. The Rotterdam GATE terminal is visualised in figure 7.10 and the other three terminals are plotted in appendix I.3, figures I.10, I.11 and I.12. The average occupancy for these four terminals lies between

17 and 24%. The optimal value for liquid bulk terminals is often set at an average berth occupancy of 50 and 65% (Kox, 2016). The observed average berth occupancies are much lower compared to these design guidelines.

Three out of four terminals at some moment in the time span actually pass the 100% occupancy limit. These possible outliers will be separately examined per terminal.

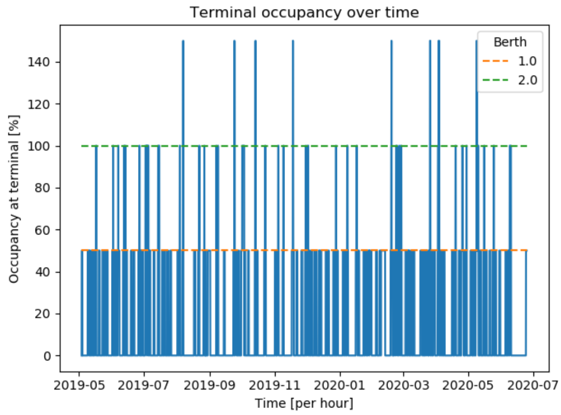


Figure 7.10: Rotterdam GATE: berth occupancy

Inspections of occupancy outliers

The outliers for the three terminals are investigated. Two (typical) examples will be given here for the Rotterdam GATE terminal. For the Rotterdam GATE terminal a few moments include an occupancy of 150%, as if three vessels were actually present at the terminal (which include only 2 LNG jetties). A period is chosen where for a consecutive period an occupancy of 150% is found, between 23-09-2019 11 PM and 24-09-2019 3 AM (four hours in total). Three vessels are indeed present at this time, as shown in figure 7.11. Based on a list of berthed vessels from Sea-web only the two Inland Waterway Tankers berthed during this time period. Thus, apparently the BORIS DAVYDOV (MMSI 209356000) did not berth at the terminal.

mmsi	loa	DWT	type	breadth	teu_capacity	port_entry_time	port_exit_time	terminal_entry_time	terminal_exit_time
209356000	299.0	96765.0	Gas tankers	50.00	0.0	2019-09-22 20:28:53	2019-09-24 09:44:50	2019-09-23 00:59:21	2019-09-24 07:40:16
244670246	110.0	2950.0	Inland Waterways Tanker	11.40	0.0	2019-09-23 08:10:10	2019-09-26 00:10:17	2019-09-23 22:16:20	2019-09-24 04:05:26
244670103	86.0	1644.0	Inland Waterways Tanker	9.56	0.0	2019-09-23 18:02:04	2019-09-26 10:43:36	2019-09-23 22:15:09	2019-09-24 03:57:43

Figure 7.11: Three vessel tracks present at the moment of maximum occupancy (150%) [example 1]

The vessel path of this specific vessel track is further investigated. The vessels sent AIS messages between 2019-09-23 00:59:21 and 2019-09-24 07:40:16 are shown in figure 7.12. The vessel does show characteristics of a vessel that would actually berth at the terminal.

Figure 7.12: Vessel path of BORIS DAVYDOV

The second example contains the moment between 2019-11-18 00:01:23 and 2019-11-18 06:01:23, where for 6 hours the occupancy again was 150%. The following 3 vessels were present at the time, as shown in figure 7.13. These three vessel tracks are almost the same as the vessel tracks which according to Sea-web berthed at the terminal, during this time frame.

mmsi	loa	DWT	type	breadth	teu_capacity	port_entry_time	port_exit_time	terminal_entry_time	terminal_exit_time
244790048	85.73	1650.0	Inland Waterways Tanker	9.56	0.0	2019-11-15 17:59:42	2019-11-18 14:14:45	2019-11-17 23:51:37	2019-11-18 06:56:43
311000632	299.00	96840.0	Gas tankers	50.00	0.0	2019-11-16 08:10:00	2019-11-19 00:24:21	2019-11-17 22:10:18	2019-11-18 21:38:54
538002921	315.17	106896.0	Gas tankers	50.04	0.0	2019-11-16 17:39:58	2019-11-18 09:36:52	2019-11-16 21:39:24	2019-11-18 07:36:38

Figure 7.13: Three vessel tracks present at the moment of maximum occupancy (150%) [example 2]

To summarise, multiple factors can influence the occupancy passing the 100%. Some were already mentioned in subchapter 7.2.2. However, the two last factors (regarding the the terminal length and 15 meter margin) are not applicable for liquid bulk terminals. These outliers are either due to errors in the AIS data or in the created AIS tool. The AIS tool might be classifying certain vessel tracks as berthed, whilst they did not actually berth at the terminal.

7.3.2. Length occupancy for liquid bulk terminals

LNG terminals consists of one or more jetties which function as the terminal berths. A jetty often consists of an approach bridge connected to a jetty head. The jetty head is a platform containing loading arms, breasting and mooring dolphins and multiple service areas (Ligteringen, 2017). An example of a L jetty arrangement is given in figure 7.14 below. The adjusted length occupancy can be defined, but the occupancy of the terminal is not expected to depend on the length, but on whether or not a jetty is occupied or not. Thus, the adjusted length occupancy will not be defined for the LNG terminals as the results will not yield any useful information.

Figure 7.14: Example of LNG Jetty (source: Costa Norte LNG Terminal, Colon (https://www.hydrocarbons-technology.com/projects/costa-norte-lng-terminal-colon/)

7.3.3. Occupancy comparisons between liquid bulk terminals

The berth occupancy distribution of the four terminals are plotted side by side in figure 7.15. The Rotterdam GATE terminal and Zeebrugge LNG terminals obtain a very similar shape. The same similar shape is found for the France Montoir LNG terminal, with the exception of this terminal not having any outliers above 100%. The different shape observed for the Dunkirk LNG terminal is expected as it only has one berth, compared to all the other terminals having two berths.

Furthermore, the number of berths is plotted against the average terminal occupancy to see if any relationships between these two are noticeable. It is not feasible to make conclusions about this relationship based only on these four terminals, since three out of four terminals have two berths. The average occupancies of the liquid bulk terminals are very similar to one another, where the average occupancies of the two other terminal types are more diverse among themselves. This is assumed to be caused by the LNG terminals handling very specific cargo, and thus very specific vessel types.

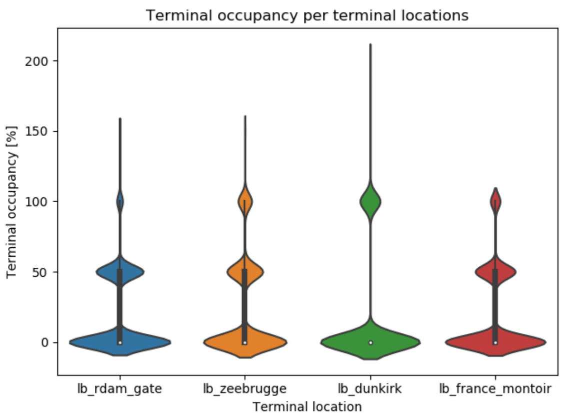


Figure 7.15: Liquid bulk terminals: occupancy

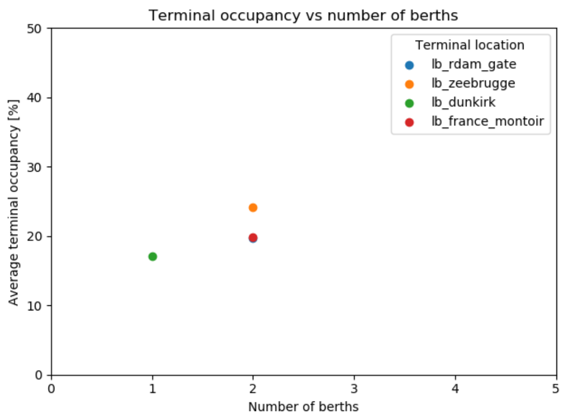


Figure 7.16: Liquid bulk terminals: no. of berths vs occupancy

7.4. Discussion and conclusions: berth occupancy

The results regarding the occupancy are discussed in the next subchapter. The answers to the sub research questions regarding the berth occupancy are given.

7.4.1. Discussion and disclaimers of the berth occupancy

The berth occupancy is defined as:

- Berth occupancy = The number of vessels present at the terminal, in comparison with the total number of available berths, over time.

For the container and dry bulk terminals the berth occupancy could not be defined due to an inconsistent number of berths at the terminal. Therefore, for these terminals a new occupancy measure was defined:

- Adjusted length occupancy = The length of all vessels present including a 15 meter margin between every two vessels, in comparison with the total available terminal length, over time.

The vessels present at the terminal are defined and mainly dependent on the service times. Possible limitations and error sensitivity for these service times have been discussed in subchapter 5.5.1. To summarise:

- The terminal polygon: this polygon is manually defined by the user of the AIS tool leading to errors based on personal interpretations. These errors are reduced by the fact that the polygons are drawn as consistently as possible between the twelve different terminals.
- The AIS data: the AIS data is assumed to represent reality. Possible errors in the AIS data sources or in the AIS signals are not deleted from the data set.
- The AIS tool: the tool is developed using multiple assumptions and chosen limits, which must be remembered when assessing these results. These limitations are thoroughly discussed in subchapter 4.1.7.

It is very important to realise that AIS data represents the reality. That means that when the AIS signals are correctly send by all vessels and the data is correctly registered by the AIS data bases, the data will present actual operations at terminals. Comparisons have been made to design occupancy values, which are created for the design capacity. Terminals might run consistently below their originally designed capacity and occupancy values, leading to difficulty in comparisons due to the complexity of these values.

Furthermore, for every terminal type only four terminals have been analysed, leading to uncertainties when making conclusions. This small number of observed terminals can significantly impact the results, when there is a large spread between the average occupancies this can cloud any potential correlations.

7.4.2. Conclusions and answers of research questions

The research objective focused on the *Berth occupancy* was defined as follows:

- *Can the berth occupancy be defined for container-, dry bulk- and liquid bulk terminals, based on AIS data, and is there a correlation between the occupancy and the terminal size?*

In this past chapter the occupancy was defined for the three different terminal types, using AIS data. For the container and dry bulk terminals the berth occupancy can not be defined based on the number of berths. These terminals contain a variable amount of berths, dependent on the size of the vessels present. Therefore, for these terminals a different method was defined in order to inspect the occupancy: the adjusted length occupancy. The adjusted length occupancy is the length of all vessels present, including a 15 meter margin, compared to the total available terminal length.

For the container and dry bulk terminals the adjusted length occupancy has been plotted over time and possible outliers (of more than 100% occupancy) have been investigated. Possible outliers represent moments in the selected time span where the occupancy was greater than 100%. These outliers were inspected and multiple reasons are possible which can lead to these outliers. First of all, the occupancy is calculated with an one hour interval, thus the occupancy data might present two vessels present at the same time (in that same hour), whilst they actually do not berth at the same exact time. Vessels can send incorrect AIS signals or the developed AIS tool can't register the signals correctly, both leading to incorrect representations of the occupancy. Furthermore, vessel can berth parallel to each other and the terminal length might be incorrectly defined. Finally, the 15 meter margin between vessels based on design guidelines might not correspond to reality.

For the liquid bulk terminals the number of berths is available, based on the number of LNG jetties the terminal has. For all the liquid bulk terminals the berth occupancy has been plotted over time and again possible outliers (of > 100% occupancy) have been investigated. The same impacts are considered as possible reasons for these outliers, with the exception of course of the 15 meter length margin or incorrect definition of terminal length.

With regards to possible relationships between the size of a terminal and the terminal's occupancy the length or number of berths have been plotted against the occupancy. For all terminals, a higher occupancy is expected when the terminal is larger (longer quay wall or larger number of berths). In other words, with a higher number of berths, a terminal can obtain a higher occupancy rate with the same amount of maximum waiting time in terms of service time. For the container terminals a conclusion was drawn that longer terminals have a higher adjusted length occupancy. For dry bulk terminals this was not the case, and no clear relationship could be determined. Finally, for the liquid bulk terminals the data was verified where three of the four LNG terminals have two berths, the other terminal has one berth. Therefore, it has been decided that possible conclusions would not be reliable, based on the non-diversity and small number of the terminals analysed.

The difference between the average adjusted length occupancy of the dry bulk terminals is the largest, varying between 14% and 54%. Whilst, the container terminals vary between 23% and 40%. In terminal design it is expected that container vessels are less patient and therefore contain a lower maximum waiting time, leading to average design occupancy values of 35%. Dry bulk vessels are assumed to tolerate longer waiting times, leading to a design occupancy of roughly 60-70%. Comparisons between the theoretical recommendations for the berth occupancy and the actual observed length occupancy are not 100% reliable because the occupancies are defined in two very different ways. Nonetheless, the occupancy of the container terminals corresponds to the design guidelines, whilst the occupancy of dry bulk terminals is actually much lower compared to the guidelines. However, this could be based on the difference in how the occupancy is defined. The dry bulk terminal might be fully utilized (based on equipment or berth locations), whilst the quay is not 100% occupied physically (based on vessel length). The dry bulk terminal occupancy is larger compared to the container terminal occupancy, corresponding to the expectation that these vessels will allow higher waiting times, leading to higher occupancy values.

The liquid bulk terminals all have an average berth occupancy between 17% and 24%. The liquid bulk terminals are all LNG terminals, handling a very specific cargo type, and thus very specific vessels. LNG terminals are often designed to handle berth occupancy values of 50-65%. Thus, the observed average berth occupancies are much lower compared to these design guidelines. It must be remembered that comparisons have been made to design occupancy values, which are created for the design capacity of the terminal. Terminals might run consistently below their originally designed capacity and occupancy values due to various reasons and external influences.

III

Discussion, conclusions and recommendations

- Chapter 8: Discussion
- Chapter 9: Conclusions
- Chapter 10: Recommendations

Part III reflects on the research objectives and summarises all results. Recommendations are given for the current design guidelines, the use of the AIS tool and possible further research.

8
Discussion

This chapter covers the relationships and principles of the results found in chapters 4, 5, 6 and 7. Similarities, as well as contradictions, with previously published guidelines are discussed. First, the AIS tool development results are deliberated. Second, the AIS data analyses for the different study parameters are discussed. Afterwards, the reliability and the generalizability of the obtained results are addressed.

8.1. AIS Tool development

To start of, the first section of the results are discussed: the AIS Tool development. In this research a tool is created which transforms raw AIS data into a data frame containing information of entry and exit times of different port processes. The tool requires several inputs, as elaborated in subchapter 4.1.6, which can be summarised into:

- Geographical information: port polygon, anchorage polygon(s), terminal polygon
- Port information: terminal type, number of anchorage areas
- Time span

These inputs are required manually and are therefore sensitive to personal interpretations. In this research the inputs are selected as consistently as possible between all the terminals. However, it is impossible to perform these steps without any consequences due to the manual inputs. Furthermore, multiple assumptions were used in the creation of the tool, which are discussed thoroughly in subchapter 4.1.7. The assumptions with the greatest influences on this research, which can not be neglected when analysing the results, are:

- Reliability of the AIS data
- Type filtering and vessel class categorizations
- Extracting berthed vessel tracks based on XGBoost classifier, *including Sea-web data as validation data set*

The results obtained for the service times, inter arrival times and berth occupancy all rely on the AIS data. The data used is extracted from the RHDHV data base.

More information about the specific data source of RHDHV is not publicly available.

For the reliability of the results the AIS data is thus very important. The quality of the AIS data has been discussed and examined in multiple researches from which varying results were found. Specifically a few parameters, which are not used in this research, are often incorrect. Nonetheless, the AIS data might be correctly registered and received, but the vessels should always send AIS data at the correct times. For example, when berthing at terminals AIS signals might be switched off. Therefore, the chance exists that vessels stop sending AIS messages too soon (when arriving), or start sending again too late (when leaving). Thus, even when the data bases correctly register all AIS messages, the AIS messages might not represent the exact reality. To summarise, a certain precaution should be taken when believing the AIS data to represent the reality 100% accurately.

Second, the type filtering and vessel class categorizations influence the results. The type filtering is one of the first filter steps which removes a large part of the AIS data, based on the vessel type. *More information regarding this filter step is not publicly available.* The vessel class categorizations splits the data into smaller data sets based on certain vessel characteristics. These characteristics (dependent on either the LOA or DWT) are therefore assumed to be true, which might not be the case 100% of the time, as errors in these characteristics can occur.

Finally, the tool uses a trained XGBoost classifier in order to predict whether vessels berth at the terminal or not. The training of the XGBoost algorithm is based on twelve different terminals, using Sea-web data as a validation data source. Therefore, the assumption has been made that Sea-web correctly registers all berthed vessel tracks. Furthermore, the algorithm predicts with an accuracy of 97.06%. Thus, when analysing the results and interpreting the conclusions, it must be noted that some of the vessel tracks might actually not have berthed at the terminal.

8.2. AIS data analyses and comparisons to theoretical framework

The AIS data analysis focuses on the service times, the inter arrival times and the berth occupancy of twelve different terminals. First of all, the generated AIS tool is used to transform the raw AIS data into information about the different entry and exit times for certain locations of the port. The limitations and assumptions applied during the creation of the AIS tool are very important. Especially the manual inputs are very easily affected by personal interpretations. The assumptions and inputs should be kept in mind when analysing all results.

The service times and inter arrival times are compared to what current design guidelines (based on PIANC and UNCTAD reports) suggest. These guidelines suggest possible distributions for situation where external influences do not play a role. In reality there are many different external influences which can affect these service and inter arrival time distributions. For example, the service time can be affected by weather influences slowing down the (un)loading processes. The service time is also influenced by possible downtime and the availability of onshore and/or onboard equipment. In addition, the inter arrival times are likewise affected by weather influences. Certain wave or tidal conditions can limit the entrance of a port. Moreover, the inter arrival time is also affected by vessels slowing down their speed when approaching a port, if known that the desired terminal is occupied. To summarise, it is important to realize that the results contain all external influences present. As mentioned, queuing theory is based on a simplified port system where external influences are not included. The observed distributions are not adjusted in such a way that these influences are excluded.

Twelve different terminals are analysed from which conclusions can be drawn about the service time, inter arrival time and occupancy. In this research multiple influences on the service times are discussed. The type of vessel, the type and quantity of the cargo and the cargo handling rate are seen as the four main factors influencing the service time of every vessel. The cargo handling rate is subsequently dependent on the equipment used to (un)load the vessel. Container and liquid bulk terminals are most often designed for both unloading and loading, thus for both importing and exporting cargo. These cargo handling rates are not expected to vary much between unloading and loading operations. However, for dry bulk terminals, it is important to differentiate between export and import terminals, as these terminals are mostly designed for one-way traffic only (Ligteringen, 2017). In this research all four dry bulk terminals handle both import and export cargo, which makes it valid to compare these four terminals. Nonetheless, the service times are expected to differ greatly if specifically import and export terminals are compared.

When splitting the terminals into groups based on the terminal type, only four terminals per terminal type remain. In other words, all conclusions for certain terminal types are based on only four terminals. A large spread in the results can cloud any potential correlations and relationships. This small amount of terminals limits the reliability of the conclusions and must be taken into consideration when analysing the results. When splitting the data of certain terminals into smaller data sets for specific vessel classes the data sets become much smaller (less AIS messages). For some vessel classes of certain terminals these data sets become so small that they are classified as unreliable. Limits based on a minimal number of arrivals are implemented, below these limits all data sets are registered as unreliable. Besides these limits, extra precaution in analysing results is taken when data sets are relatively small. This subsequently influences the reliability of the results of every specific vessel class. Again, this must be remembered when analysing results and making conclusions.

Furthermore, after the split only certain vessel classes are analysed. Thereby, a part of the arrivals are not analysed in this second step (analysis based on vessel class), due to these arrivals not belonging to the specific analysed vessel classes. For the container vessels the number of vessels left out of these results is not relatively large and is thus not expected to influence the conclusions much. However, for the dry bulk terminals quite a considerable part of the arrivals is left out of these second inspections, this can influence the final results and conclusions made. Finally, for the liquid bulk terminals no vessel arrivals are excluded from this second research objective, thus no influences will be present for this matter on the results and conclusions.

Finally, the berth occupancy can only be defined when the number of berths is constant over time. This is not the case for the eight analysed container and dry bulk terminals. For these terminals a new method is defined to represent a terminals occupancy: the adjusted length occupancy. However, the reliability of this new occupancy measure is questionable. The length occupancy suggests that the terminal occupancy is fully dependent on the length of the terminal. Another very important factor influencing the terminal occupancy is the availability of onshore equipment. For example, at container terminals when all cranes are occupied the terminal can not serve an extra vessel, while there might physically be place for one (dependent on the available terminal length).

8.3. Reliability of the results

In this chapter multiple limitations have been discussed that affect the results and their reliability. It is important to quantify the impacts these influences have on the reliability of the results, where this is possible. First of all, the developed AIS tool contains some necessary inputs which are added manually. Deviations in the way that these inputs are defined, for example the polygon locations, affect all three study parameters. For one single terminal these effects are negligible, as these manual choices are consistent across all the results of that terminal. For example, when the port polygon would be chosen slightly different, it will influence the arrival times of the vessels. The assumption is made that all vessels will have roughly the same deviation to the arrival time, thus leading to the same type of total inter arrival time distribution. Regarding the effects during comparisons of all terminals, the manual inputs will lead to slightly different parameters of the distribution. Nonetheless, the assumption is made that the distributions will still contain the same shapes, despite for example slightly affected location parameters.

The reliability of the AIS data is a more difficult influence to assess. As mentioned in subchapter 8.1, the accuracy of AIS data has been the topic of multiple previous researches. Based on previous literature, the conclusion is made that the reliability of the AIS data is strongly increased due to the limited number of parameters used in this research. Only four parameters of AIS data are used (MMSI, timestamp, latitude, longitude), thus the AIS data used is expected to lead to reliable results. Furthermore, the XGBoost classifier predicts if a vessel is berthed or not with an 97% accuracy. Specifically, 99% of the berthed vessels tracks are correctly predicted as berthed. However, 7% of the total number of berthed vessel tracks, are vessels that did not actually berth, thus are wrongfully predicted (table E.19). When analysing the results it must therefore be kept in mind that roughly 7% of the data contains vessel tracks that did not actually berth, and that 1% of the vessel tracks that berth is missing from the data (due to a non-berthed prediction). Besides these wrongful predictions present in the data, a certain precaution is necessary when assessing the AIS data as 100% correct. As mentioned in subchapter 8.1 there are multiple possible influences that can affect the quality of the AIS data.

8.4. Generalizability of the results

The generalizability of the results are inspected in two sections. First, the possible generalizations of the AIS Tool are discussed. Second, the generalizability of the data analyses results and comparisons to guidelines are discussed. As previously mentioned, the developed AIS port processes tool can be used for any terminal location. For any combination of manual inputs, the AIS data will be transformed to a data set containing information about the entry and exit times of the port, anchorage and terminal areas. Additionally, other AIS data sources can be assessed with the developed tool due to the possibility of adjusting the column names.

Correlations and comparisons in this research are based on twelve different analysed terminals. The terminals represent three different terminal types: container, dry bulk and liquid bulk. This does not necessarily mean that other terminal types can't be compared with these results. The possibility exists where other terminal types are assessed with the obtained results from this research, however some caution is recommended. With regards to the location of the terminals, different types of locations have been selected. In total, six different port locations have been used, as shown in figure 8.1. The analysed terminals all are located in the Netherlands, Belgium or France. However, the results found can be applied in other similar terminals as well. The terminal sizes range between 675 and 2700 meters, which suggests that other similar terminals with a terminal length between this range, will most likely comply with the found results. However, no extra research has been performed to test this assumption. With regards to the liquid bulk terminals, only specifically LNG terminals have been analysed. Since the results are quite explicit, they are not assumed to be very useful for other types of liquid bulk terminals.

Figure 8.1: Port locations

In this thesis multiple influences on the service times and inter arrival times have been discussed. Based on all this influences the service time and inter arrival time will never be the same for different terminals. It has also been observed that for a specific terminal the service and inter arrival time can be very variable. Therefore, for new terminals these times are expected to differ as well. However, the typical distributions for the inter arrival and service times, are predicted to be similar as to the observed terminals, but with different parameters.

9 Conclusions

Chapter 9 discusses and reflects on the research as a whole. This chapter is compartmentalized into different subchapters, each representing the one of the different research objectives. First, conclusions based on the developed AIS tool with its limitations and capabilities are discussed. Second, the results based on the service time distributions are elaborated. Third, the inter arrival time distributions results are summarised. Finally, the berth occupancy results are discussed.

The main research question is addressed in this chapter:

How are service times & inter arrival times distributed and can the berth occupancy be defined, at container-, dry bulk- and liquid bulk terminals, based on AIS data, and how do these compare with design guidelines?

The research question was approached by being split into four different research objectives (each with one or more sub research questions): AIS data processing and possibilities, service time distribution, inter arrival time distribution and berth occupancy. Next, all four research objectives are addressed.

AIS data: processing and possibilities

During this research a tool is created which uses raw AIS data together with manually specified port-, anchorage- and terminal locations, in order to obtain information about different port processes a vessel follows in a port. The tool returns a data frame in which for every single vessel track the entry and exit times of these three locations are summarised: port, anchorage and terminal. The three sub research questions for this section are:

(a) *How can AIS data be used to define different processes a vessel follows in a port?*

(b) *What is the most optimal procedure of extracting vessel tracks that berth at a terminal, using AIS data?*

(c) *What are performance analyses which can be generated using AIS data?*

A tool has been developed using Python programming language which uses (raw) AIS data and transforms it into a data frame as specified above. For every individual vessel track three different locations are specified by their entry and exit times. The tool therefore serves as a base method to define the different port processes a vessel can follow in a port. The tool is currently in its first edition and therefore certain limitations and assumptions apply which should always be taken into account when using it.

Multiple approaches have been used to find the best way possible of extracting berthed vessel tracks from a data set containing all possible vessel tracks (both vessels that do or do not berth). The best approach is found by training a machine learning algorithm XGBoost on a mixed labelled data set consisting of twelve different terminals. The predictions are made with an overall accuracy of 97%.

The developed AIS port processes tool can be used to define the service time-, inter arrival time distributions and berth or length occupancy. However, much more (statistical) analyses are possible with this tool. For example, the waiting time of a specific vessel track can easily be defined. Furthermore, the number of arrivals for a specific terminal can be generated, from which subsequently more (visual) analyses can be performed.

This can lead to insights into modern issues at ports and terminals, such as the possible influences of Covid-19 on the number of vessel arrivals.

Service time distribution

Using the tool, twelve different terminal locations are inspected and the service time of every single vessel track is calculated. For the service time distributions three different sub research questions were formulated:

(a) *How are service times distributed along container-, dry bulk- and liquid bulk terminals, based on AIS data, and how do they compare to PIANC guidelines?*

(b) *How are the service times distributed per vessel class along container-, dry bulk- and liquid bulk terminals, based on AIS data?*

(c) *How do the three terminal types (container, dry bulk, liquid bulk) compare based on service times?*

For the four analysed container terminals no fit was found for any of the service time distributions. Current PIANC and UNCTAD guidelines expect these distributions to follow an Erlang-k distribution but for these terminals this was not the case. The majority of the dry bulk terminals similarly had no fit with any of the theoretical distributions for the service times. For the liquid bulk terminals specifically four LNG terminals were analysed. Again, the expected distributions for the service time are the Erlang-k distributions. None of the terminals showed any fit between the service time distribution and any of the tested theoretical distributions.

The service time is expected to depend on, among other things, the vessel type, the quantity and type of cargo and the handling rate. To more thoroughly inspect the service time distributions, all data sets were split into different data sets based on vessel classes. A common distribution which is often fitted for the service time of container terminals is the Gamma distribution. For the dry bulk terminals similarly the Gamma distribution was most often found as a good fit of the service time distributions. Finally, the mass of all liquid bulk vessels belong to the third LNG vessel class, to which none of the distributions could fit. Visually these distributions are similar to the Deterministic distribution with an average service time of roughly 24 hours.

For the container as well as dry bulk vessels, a larger vessel class resulted in a higher average service time. For both terminals, the different classes reciprocally did show similar distribution shapes, but each with different location parameters. The hypothesis of there being a correlation between the service times and the vessel classes is thus validated for these two terminal types. On average, a higher vessel class (indicating a larger vessel based on vessel size) will contain more cargo, which leads to more cargo (un)loading at the terminal, thus requiring longer service times. For the liquid bulk vessels this was more difficult to inspect due to the majority of vessels being LNG Class 3 vessels. No clear correlation can be found between the average service times between vessel classes for these terminals.

The four dry bulk terminals showed the most variance in between their service time distributions, not unexpected due the complicated (un)loading procedures of these terminals. The dry bulk terminals handle different types of cargo, each requiring different cargo handling techniques, in comparison to container terminals where the terminals contain very similar handling techniques and only transfer specifically containers. It is important to note that all four dry bulk terminals handle both import and export cargo, resulting in the large differences within the service time distribution of every specific terminal. The LNG terminals also contained very similar average and median service times, due to the very specific terminal commodity and unique (un)loading requirements.

The hypothesis was introduced where smaller sub data sets will obtain (better) theoretical fits, due to a reduction in the diversity of the vessel types. This hypothesis is validated when assessing the container and dry bulk terminals. Originally, UNCTAD introduced the recommended Erlang-k distribution in 1985 for the service time distribution. It is expected that the diversity of the vessel mix arriving in 1985 was much less, compared to all the vessel types arriving at terminals nowadays. Therefore, the Erlang-k distribution might have previously been a correct recommendation, when the vessel mix was less diverse. However, nowadays the difference between the vessel types is big, which leads to a very spread out service time distribution. The service time distribution can thus better be represented as a heterogeneous data set consisting of multiple homogeneous data sets. These homogeneous data sets will represent the different vessel classes, each with their own service time distribution.

Inter arrival time distribution

For the same twelve terminals, as analysed for the service time, the inter arrival time distributions are inspected. The inter arrival time distribution results are based on the following two sub research questions:

(a) *How are inter arrival times distributed along container-, dry bulk- and liquid bulk terminals, based on AIS data, and how do they compare to PIANC guidelines?*

(b) *How are the inter arrival times distributed per vessel class, along container-, dry bulk- and liquid bulk terminals, based on AIS data?*

All terminals are expected to follow an Exponential distribution based on PIANC guidelines. However, suggestions have been made that this distribution might be too conservative due to improved scheduling at terminals (PIANC WG158, 2014). The container inter arrival times don't seem to correspond based on the K-S goodness-of-fit tests, but visually do follow the Negative Exponential distribution. Three of the four analysed dry bulk terminals confirm the PIANC recommendations, where the Exponential distribution is a fit. For the liquid bulk terminals, the majority of the terminals (three out of four) follow the Exponential distribution whilst one terminal was best represented by an Erlang-2 distribution. The suggestion is made that terminals that follow the Erlang-2 distribution possibly contain more terminal scheduling or are less impacted by external factors. However, for all the other terminals represented by Exponential inter arrival times, suggestions about more improved scheduling leading to less random patterns of arrivals is not visible.

The arrivals can be modelled as a Poisson process if the arrivals are *Independent and identically distributed* (*IDD*). To test if the data is IID the data should be split into smaller data sets. When these smaller data sets share the same PDF and are independent events the distribution can be classified as IID. Thus, again the data sets per terminal are split based on vessel classifications. Based on the results of the smaller data sets, most terminals and their different vessel classes follow the Exponential distribution. The data is IID based on the assumption that the arrivals of vessels are independent of each other. As expected, the different classes contain very similar inter arrival times among the classes, for every specific terminal type. The conclusion is made that vessels still arrive in a stochastic manner at ports, despite the efforts of improved arrival scheduling. Apparently the effects of external influences can not be ignored with regards to the improved terminal scheduling. For all terminals, less arrivals correspond to higher average inter arrival times, which makes sense as there is simply more time between the arrivals.

Berth occupancy

The final research objective focuses on the possibilities of defining the berth occupancy, based on AIS data, with it's only sub research question being:

(a) *Can the berth occupancy be defined for container-, dry bulk- and liquid bulk terminals, based on AIS data, and is there a correlation between the occupancy and the terminal size?*

For the twelve terminals, the eight container and dry bulk terminal's occupancy were defined by using the length occupancy, since the berth occupancy was not possible based on the inconsistent number of berths over time. The adjusted length occupancy is defined as the total length occupied in comparison to the total length available. The average occupancy for container terminals was between 23 and 40%, and for the dry bulk terminals between 14 and 54%. Comparisons to design guidelines are not completely reliable as guidelines contain recommendations for the berth occupancy, where as here the length occupancy has been defined. In terminal design it is expected that container vessels are less patient compared to dry bulk vessels and therefore contain a lower maximum waiting time, leading to lower average design occupancy values (35% for containers compared to 60-70% for dry bulk vessels). This correlation is similar to what has been observed between the container and dry bulk length occupancy values, where the average occupancy is higher for dry bulk vessels.

Furthermore, a longer terminal length leads to a higher average length occupancy for the container terminals, but for the dry bulk terminals this correlation is less visible. In practice, a longer terminal is expected to be more flexible and is therefore assumed to contain a higher average occupancy. However, much more factors influence the occupancy of the berth, such as economical situations and geographical locations of the terminal. Additionally, the reliability of the length occupancy is questioned, as to if this is a correct and useful method of defining the terminal's occupancy.

For the four LNG terminals the number of berths were constant over time and thus the berth occupancy could be generated. The average berth occupancy is between 17 and 24%, which is much lower compared to recommended design occupancy values of 50-60%. It must be noted that comparisons have been made to design occupancy values created for the design capacity of the terminal. Thus, terminals might run consistently below their originally designed occupancy values due to various external influences. The three liquid bulk terminals, containing two berths, have very similar occupancy distributions. The terminal containing one berth has a slightly lower average berth occupancy. In practice, the number of berths is often related to the maximum allowable occupancy, where a higher number of berths is correlated with a higher average occupancy. This corresponds with the results found in this research. However, some caution is necessary when using this conclusion, as it is less reliable due to the small amount of terminals analysed.

10 Recommendations

In this chapter the recommendations are discussed based on the performed research. First, the recommendations for design guidelines are discussed. Second, recommendations for using the developed AIS tool are summarised. Finally, multiple recommendations for further research are introduced.

10.1. Recommendations for design guidelines

Queuing theory is used to inspect and simulate systems where customers wait in line to be served. The theory is often used for designing and simulating waterways and terminals. It requires a schematised port system where only the simplest facilities can be included. Generally the three first parameters are used: the inter arrival time distribution, the service time distribution and the number of servers. Based on these three parameters the waiting time in units of service time can be calculated based on different combinations of number of berths and occupancy factors. With these results, the number of berths can be selected from which the quay length can be calculated.

Current PIANC design guidelines assume the inter arrival time distribution to follow a completely random arrival process, leading to a Negative Exponential distribution. For the service time distribution the assumption is made that the service time has more consistency, leading to the smoother Erlang-k distribution. Based on the research performed the recommendation for further port planning is to adjust a few parameters.

First of all, the inter arrival time is indeed most often presentable by the Exponential distribution. Suggestions have been made that terminals nowadays contain more planning resulting in the Exponential distribution being too conservative. The conclusion is made that vessels still arrive in a stochastic manner at ports, despite the efforts of improved arrival scheduling. Apparently the effects of external influences can not be ignored with regards to the improved terminal scheduling. Furthermore, no theoretical distributions fit the service time distribution for all terminal types. The service time distributions are expected to depend on many factors, with large influences being the vessel sizes and the amount of cargo being (un)loaded. The external influences are reduced by splitting the data set into smaller sub data sets, for specific vessel sizes. It turns out that the service time distributions of these smaller data sets, do have good fits for the service time distribution, of which most are Gamma distributions.

Based on the results, the recommendation is given to define the service time distribution as a heterogeneous data set consisting of multiple homogeneous data sets. These homogeneous data sets will represent the different vessel classes, each with their own service time distribution. With an estimation of the expected vessel mix, a new service time distribution for the entire terminal can be created. The vessel mix represents the number of arrivals for each vessel class, which can be found from the developed AIS tool.

Concluding, it is recommended to use queuing theory as it is a vital ingredient for terminal planning and design. The service time distribution of the entire data should be seen as a heterogeneous data set, built up of a number smaller homogeneous sub data sets (each representing a vessel class). These smaller data sets can be best represented by the Gamma distribution, instead of the previously assumed Erlang-k distribution. The liquid bulk terminal service times are best represented by a Deterministic distribution, with an average service time of roughly 24 hours, dependent on the delivered services. For all terminal types, and corresponding to

PIANC guidelines, the arrivals of vessels are stochastic, leading to the Negative Exponential distribution best representing the inter arrival times.

10.2. Recommendations for the use of the AIS Port processes tool

10.2.1. Possible improvements for the AIS tool

For further research the AIS tool can be optimised as there is always room for improvement. At this moment, large data sets require considerable amounts of computation time. For example, 65 million AIS messages take up to 4 hours of computation time. The assumption is made that running the developed tool on the back end of a server data base will decrease the computation time significantly. Besides optimizing the tool with regards to more effective & smart coding, multiple recommendations are given to increase the possibilities and capabilities of the tool. First of all, the number of processes visualised can be (relatively easily) increased. For example, in this version of the tool only four coordinates for every polygon can be entered. This number of coordinates can be increased to assist more complex polygon shapes. Besides the shape of the polygon, the number of the polygons can also be expanded. Currently, one port, one terminal and one or two anchorage areas can be entered. Naturally, more of these locations exist in ports, and thus by adding these the AIS tool will be equipped to handle more complex port systems.

At this moment the tool can present the processes for entering and leaving the port, the anchorage area(s) and the terminal. As visualised in figure 2.1, multiple processes can be defined and the entry and exit times for these specific processes can generated. These processes can be expanded by, for example, introducing an approach channel polygon. With such an extra polygon the traffic intensity of the approach channel could be presented.

10.2.2. Recommendations for the usage of the tool

The developed AIS tool has been used to define the inter arrival time, the service time and the berth occupancy. However, it can be used for much more (different) research purposes. It is recommended that when using the tool for any kind of objective, the limitations and assumptions made are read, as extensively described in subchapter 4.1.7.

Port planners and engineers can gain insights into the performances of an existing terminal by using the developed AIS tool (available on GitHub (Van Zwieteren, 2020)). The different types of vessels can be visualised to obtain information regarding the vessel mix arriving at the terminal. The developed tool provides the basis for a potential feedback report with multiple aspects and information about the terminal. For example, a report for an existing terminal could be generated with information about the service times, the inter arrival times, the waiting times and the (berth or length) occupancy. For each of these parameters information can be provided about the average value, the distribution type, possible fluctuations over time and possible differences per vessel type. Hence, this report will give insight into different terminal performances. Subsequently with such a report, existing terminals can easily be compared to each other based on these performance indicators. Additionally, with the use of the developed tool, insights can be gained about the different processes the vessel follows in the port, and how long each different process takes. For example, the average duration between arrival, time in anchorage, sailing towards terminal, time at terminal and the sailing out can be inspected. With these insights possible bottlenecks will be uncovered. All things considered, the AIS tool can be used to create various important insights into existing terminals, as well as lead to improvements for new terminal designs. This can have significant impacts on future terminal designs, providing economic advantages to vessel and port operators.

10.3. Recommendations for further research

Since the amount and availability of AIS data is expected to increase in the future, it is crucial for more research to be performed based on and by using the AIS data. In this research AIS data has been introduced as a valuable data set, where from minimal amounts of information (only using a few data columns) large amounts of (statistical) analyses can be performed. Every data set representing reality will contain some errors and limitations, and thus caution when using the data is definitely necessary. Some recommendations for further research are given next.

As discussed in subchapter 8.3 the reliability of the results are questioned based on the size of the analysed data sets. First of all, only twelve terminals have been assessed in this research. The results and conclusions are based on only four terminals per terminal type. The first recommendation is therefore to **increase the number of analysed terminals**. With an increase of the size of the analysed data, the results become more reliable and the conclusions more robust. This increase in analysed terminals might uncover new relationships and correlations which now are possibly clouded due to diverse results. Besides expanding the number of analysed terminals, the **time span should be increased**. For example, it can be very interesting to assess possible seasonality influences. In this research AIS data based on 14 months is used, between 01-05-2019 and 01-07-2020. The collected AIS data is continuously growing, as a result more and more data is available. Especially the smaller sub data sets (based on vessel class splits) are often too small to be able to return reliable results. When the time span is increased, more data is available and these smaller sub data sets are expected to grow. Following, more reliable results and more robust conclusions can be generated for the specific vessel classes.

The number of arrivals per terminal or port can be a very valuable parameter in port planning and design steps, as well as when analysing current processes in ports. By using the current AIS tool the different **vessel mixes (number and type of vessels) arriving at terminals** can be visualised. By comparing multiple terminals, including different types and sizes, characteristic vessel mixes can be defined. These can subsequently be used in port capacity planning. Besides that, the (vessel mix) arrivals can also be used for inspecting external influences on the port and/ or terminal. For example, the 2020 evolving COVID-19 situation and related significant impact to global trade and impact on supply and demand, can be inspected by comparing multiple years of arrival data.

By expanding the number and type of polygons processed by the AIS tool, **more port processes can be visualised and inspected**. For example, the channel occupancy can be defined by selecting a polygon to represent a certain channel. Furthermore, the tug fastening or pilot on-boarding processes could be specifically analysed. Besides expending the number of processes that can be researched, **the different terminal types can be increased**. This means more terminal types are possible to analyse, instead of only the researched container, dry bulk and liquid bulk terminals.

A more complex next research step would be to **include information** to the analyses of the port processes **about the physical environment**. The inter arrival time, service time and (thus) the berth occupancy all depend on multiple influences from the physical environment of the analysed port. By including data sets which contain information about wind, waves, current or visibility, more profound analyses can be performed. Furthermore, it is interesting to inspect the influence of certain days of the week, or months of the year, on the difference in service times. The analysed terminals are operating 365 days per year, 24 hours per day. Valuable insights can be gained by inspecting the correlations between the time of the day (night versus day) and the efficiency of the terminal.

Besides including information of external influences, **information about the equipment onboard and onshore** can lead to very valuable, more thorough insights and analyses. For example, when the number of operable cranes is known, more insights about the capacity and occupancy of terminals with varying number of berths can be gained. **Including information about the call size** (quantity of cargo transfer) of every berthing vessel will also lead to more extensive insights about the service time. In this research differences in service times are often assumed to be due to (among others) the difference in call size. The amount of cargo transferred largely impacts the total service time required for (un)loading the vessel. The service time greatly depends on these factors, thus gaining more knowledge about the equipment and it's effects on different service times would be of tremendous relevance.

Besides recommendations consisting of expansions of the current AIS tool, a separate research recommendation is addressed. The inter arrival time is influenced by, among others, the external impacts of the physical environment. The communication between the arriving vessels and the terminal operator, prior to the vessels arrival, is expected to be another great influence on the arrival times. Vessels are foreseen to slow done and reduce their sailing speed once the terminal or berth is occupied at the estimated time of their arrival. Further study into the **impact of vessels slowing down** when approaching the port due to occupied terminals should be investigated, as this may have a significant impact to certain findings, and in parallel provide economic benefits to both port and vessel operators.

References

1.3.5.16. Kolmogorov-Smirnov Goodness-of-Fit Test. (n.d.). Retrieved from https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm#:~:text=TheKolmogorov-Smirnovtest(Chakravart,populationwithaspecificdistribution.&text=Anotheradvantageisthatit,theapproximationstobevalid).

1.3.6.6.11 Gamma Distribution. (n.d.). Retrieved from https://www.itl.nist.gov/div898/handbook/eda/section3/eda366b.htm

Aarsæther, K. G., & Moan, T. (2009). Estimating Navigation Patterns from AIS. *The Journal of Navigation*, *62*(4), 587–607.

Abdul Rahman, N. S. F., Othman, M. K., Sanusi, I. A., MD Arof, A., & Ismail, A. (2019). Evaluation of Delay Factors on Dry Bulk Cargo Operation in Malaysia: A Case Study of Kemaman Port. *Asian Journal of Shipping and Logistics*, *35*(3), 127–137. Retrieved from http://dx.doi.org/10.1016/j.ajsl.2019.09.001 doi: 10.1016/j.ajsl.2019.09.001

Adan, I., & Resing, J. (2015). *Queueing systems* (Tech. Rep.). Eindhoven: Deparment of Mathematics and Computing Science Eindhoven University of Technology. doi: 10.1016/0005-1098(66)90015-X

AIS Partner Stations. (n.d.). Retrieved from https://stations.vesselfinder.com/

Aiyappa, S. N., & Ramamurthy, B. (2018). An efficient approach towards clustering using K-means algorithm. *International Journal of Civil Engineering and Technology*, *9*(2), 705–714.

Allen, M. (2018). *81. Distribution fitting to data.* Retrieved from https://pythonhealthcare.org/2018/05/03/81-distribution-fitting-to-data/

API Data Packages. (n.d.). Retrieved from https://developer.fleetmon.com/api-data-packages/

Baldauf, M., Wismar, H., Benedict, K., & Motz², F. (2008). Aspects of Technical Reliability of Navigation Systems and Human Element in Case of Collision Avoidance. *NAV08 /ILA37*. Retrieved from https://www.researchgate.net/publication/242306672

Bellsolà Olba, X., Daamen, W., Vellinga, T., & Hoogendoorn, S. (2014). *An approach to port network capacity* (Tech. Rep.). Wuhan: Delft University of Technology. Retrieved from https://www.researchgate.net/publication/283713926

Bellsolà Olba, X., Daamen, W., Vellinga, T., & Hoogendoorn, S. P. (2018, 10). State-of-the-art of port simulation models for risk and capacity assessment based on the vessel navigational behaviour through the nautical infrastructure. *Journal of Traffic and Transportation Engineering (English Edition)*, *5*(5), 335–347. doi: 10.1016/j.jtte.2018.03.003

Berry, R. (2006). *Queuing Theory.*

Bolboacă, S. D., Jäntschi, L., Sestraş, A. F., Sestraş, R. E., & Pamfil, D. C. (2011). Pearson-Fisher Chi-Square Statistic Revisited. *Information*, *2*(3), 528–545. doi: 10.3390/info2030528

Bomberger, N. A., Rhodes, B. J., Seibert, M., & Waxman, A. M. (2005). *Associative learning of vessel motion patterns for maritime situation awareness* (Tech. Rep.). doi: 10.1109/ICIF.2006.301661

Braca, P., Katsilieris, F., & Coraluppi, S. (2013). *Detection of Malicious AIS Position Spoofing by Exploiting Radar Information* (Tech. Rep.). Retrieved from https://www.researchgate.net/publication/252932808

Carlan, V., Sys, C., Vanelslander, T., & Roumboutsos, A. (2017, 1). Digital innovation in the port sector: Barriers and facilitators. *Competition and Regulation in Network Industries*, *18*(1-2), 71–93. doi: 10.1177/1783591717734793

Carral Couce, L., Carral Couce, J. C., & Fraguela Formoso, J. A. (2015, 1). Operation and handling in escort tugboat manoeuvres with the aid of automatic towing winch systems. *Journal of Navigation*, *68*(1), 71–88. doi: 10.1017/S0373463314000435

Chapter 8 Beta and Gamma. (n.d.). Retrieved from https://bookdown.org/probability/beta/beta-and-gamma.html

Chen, D., Zhao, Y., Nelson, P., Li, Y., Wang, X., Zhou, Y., ... Guo, X. (2016, 11). Estimating ship emissions based on AIS data for port of Tianjin, China. *Atmospheric Environment*, *145*, 10–18. doi: 10.1016/j.atmosenv.2016.08.086

Chen, L., Zhang, D., Ma, X., Wang, L., Li, S., Wu, Z., & Pan, G. (2015, 5). Container Port Performance Measurement and Comparison Leveraging Ship GPS Traces and Maritime Open Data. *IEEE Transactions on Intelligent Transportation Systems*, *17*(5), 1227–1242. doi: 10.1109/TITS.2015.2498409

Çigşar, B., & Ünal, D. (2019). Comparison of Data Mining Classification Algorithms Determining the Default Risk. *Scientific Programming*, *2019*. doi: 10.1155/2019/8706505

Cooper, J. C. (2005). The Poisson and Exponential Distributions. *Applied Probability Trust*. Retrieved from http://ms.appliedprobability.org/data/files/Abstracts37/37-3-6.pdf

Council of European Energy Regulators. (2017). *LNG TF: Removing LNG barriers on gas markets* (Tech. Rep. No. December).

de Boer, T. (2010). *An analysis of vessel behaviour based on AIS data Application of AIS data in a nautical traffic model* (Unpublished doctoral dissertation). Technical University of Delft.

De Vreede, I. (2016). *Managing Historic Automatic Identification System data by using a proper Database Management System structure* (Unpublished doctoral dissertation). Delft University of Technology.

Dobbin, K. K., & Simon, R. M. (2011). Optimally splitting cases for training and testing high dimensional classifiers. *BMC Medical Genomics*, *4*(1), 31. Retrieved from http://www.biomedcentral.com/1755-8794/4/31 doi: 10.1186/1755-8794-4-31

Dobrkovic, A., Iacob, M. E., Hillegersberg, J. v., Mes, M. R., & Glandrup, M. (2015). Towards an Approach for Long Term AIS-Based Prediction of Vessel Arrival Times. *Logistics and Supply Chain Innovation: Bridging the Gap between Theory and Practice*, 281–294.

Dr. Dataman. (2019). *Explain Any Models with the SHAP Values — Use the KernelExplainer.* Retrieved from ExplainAnyModelswiththeSHAPValues—UsetheKernelExplainer

Dragović, B., Tzannatos, E., & Park, N. K. (2017, 3). Simulation modelling in ports and container terminals: literature overview and analysis by research field, application area and tool. *Flexible Services and Manufacturing Journal*, *29*(1), 4–34. doi: 10.1007/s10696-016-9239-5

Ducruet, C., Itoh, H., & Merk, O. (2014). Time Efficiency at World Container Ports. *International Transport Forum*(2014-08), 30. Retrieved from www.internationaltransportforum.org/jtrc/DiscussionPapers/jtrcpapers.html%5Cnwww.internationaltransportforum.org/jtrc/DiscussionPapers/jtrcpapers.html%5Cnwww.internationaltransportforum.org/jtrc/DiscussionPapers/jtrcpapers.html

El-Naggar, M. E. (2010). Application of queuing theory to the container terminal at Alexandria seaport. *Journal of Soil Science and Environmental Management*, *1*(4), 77–85.

Erlang, A. K. (1909). The theory of probabilities and telephone conversations. *NYT Tidsskrift for Matematik*, *20*, 33.

European Commission. (2015). *Exhange of views between port CEOs and Transport Commissioner Bulc.* Retrieved from https://ec.europa.eu/transport/modes/maritime/ports/ports_en

ExactAIS. (n.d.). Retrieved from https://www.exactearth.com/products/exactais

Felski, A., & Jaskólski, K. (2012, 10). Information unfitness as a factor constraining Automatic Identification System (AIS) application to anti-collision manoeuvring. *Polish Maritime Research*, *19*(3), 60–64. doi: 10.2478/v10012-012-0032-4

Ferraro, G., Baschek, B., de Montpellier, G., Njoten, O., Perkovic, M., & Vespe, M. (2010, 1). On the SAR derived alert in the detection of oil spills according to the analysis of the EGEMP. *Marine Pollution Bulletin*, *60*(1), 91–102. doi: 10.1016/j.marpolbul.2009.08.025

Ferraro, G., Bernardini, A., David, M., Meyer-Roux, S., Muellenhoff, O., Perkovic, M., … Topouzelis, K. (2007, 4). Towards an operational use of space imagery for oil pollution monitoring in the Mediterranean basin: A demonstration in the Adriatic Sea. *Marine Pollution Bulletin*, *54*(4), 403–422. doi: 10.1016/j.marpolbul.2006.11.022

Fiorini, M., Capata, A., & Bloisi, D. D. (2016, 12). AIS Data Visualization for Maritime Spatial Planning (MSP). *International Journal of e-Navigation and Maritime Economy*, *5*, 45–60. doi: 10.1016/j.enavi.2016.12.004

Gast, T. D. (2013). *Probability Distribution Functions for Geomechanical Properties from Well Log Data* (Unpublished doctoral dissertation). Technical University of Delft.

GeeksforGeeks. (n.d.). *Regression and Classification.* Retrieved from https://www.geeksforgeeks.org/regression-classification-supervised-machine-learning/

Gini index vs Entropy. (n.d.). Retrieved from https://thatascience.com/learn-machine-learning/gini-entropy/

Goerlandt, F., & Kujala, P. (2011, 1). Traffic simulation based ship collision probability modeling. *Reliability Engineering and System Safety*, *96*(1), 91–107. doi: 10.1016/j.ress.2010.09.003

Groenveld, R. (2001). *Service Systems in Ports and Inland Waterways*. TU Delft: VSSD.

Grossmann, M. (2019). *Collision risk assessment in coastal waters* (Unpublished doctoral dissertation). Delft University of Technology.

Guerriero, M., Coraluppi, S., & Carthel, C. (2010). *Analysis of AIS intermittency and vessel characterization using a Hidden Markov Model* (Tech. Rep.). NATO.

Günther, H.-O., & Kim, K.-H. (2006). Container terminals and terminal operations. In *Container terminals and terminal operations* (28(4) ed., pp. 437–445). OR Spectrum.

Gursky, J. (2020). *Boosting Showdown: Scikit-Learn vs XGBoost vs LightGM vs CatBoost in Sentiment Classification.* Retrieved from https://towardsdatascience.com/boosting-showdown-scikit-learn-vs-xgboost-vs-lightgbm-vs-catboost-in-sentiment-classification-f7c7f46fd956

Gustriansyah, R., Suhandi, N., & Antony, F. (2019). Clustering optimization in RFM analysis based on k-means. *Indonesian Journal of Electrical Engineering and Computer Science*, *18*(1), 470–477. doi: 10.11591/ijeecs.v18.i1.pp470-477

Guzman, H. M., Gomez, C. G., Guevara, C. A., & Kleivane, L. (2013). Potential vessel collisions with Southern Hemisphere humpback whales wintering off Pacific Panama. *Marine Mammal Science*, *29*(4), 629–642.

Hamel, G. (2019). *Python for Data 25: Chi-Squared Tests.* Retrieved from https://www.kaggle.com/hamelg/python-for-data-25-chi-squared-tests

Hansen, M. G., Jensen, T. K., Lehn-Schioler, T., Melchild, K., Rasmussen, F. M., & Ennemark, F. (2013, 11). Empirical ship domain based on AIS data. *Journal of Navigation*, *66*(6), 931–940. doi: 10.1017/S0373463313000489

Harati-Mokhtari, A., Wall, A., Brooks, P., & Wang, J. (2007). Automatic Identification System (AIS): Data Reliability and Human Error Implications. *The Journal of Navigation*, *60*(3), 373–389.

Harati-Mokhtari, A., Wall, A., Harati-Mokhtari, A., Brooks, P., & Wang, J. (2008). *Automatic Identification System (AIS): A Human Factors Approach* (Tech. Rep.). Retrieved from https://www.researchgate.net/publication/254062770

Heilig, L., Lalla-Ruiz, E., & Voß, S. (2017, 12). Digital transformation in maritime ports: analysis and a game theoretic framework. *NETNOMICS: Economic Research and Electronic Networking*, *18*(2-3), 227–254. doi: 10.1007/s11066-017-9122-x

Heymann, F., Noack, T., & Banyś, P. (2013). *Plausibility analysis of navigation related AIS parameter based on time series* (Tech. Rep.).

Ho, J. H., & Bateman, S. (2013). *Maritime Challenges And Priorities in Asia: Implications for Regional Security*. Oxon: Routledge.

Hornauer, S., & Hahn, A. (2013). Towards marine collision avoidance based on automatic route exchange. In *Ifac proceedings volumes (ifac-papersonline)* (Vol. 9, pp. 103–107). IFAC Secretariat. doi: 10.3182/20130918-4-JP-3022.00049

Høye, G. K., Eriksen, T., Meland, B. J., & Narheim, B. T. (2007, 1). Space-based AIS for global maritime traffic monitoring. *Acta Astronautica*, *62*(2-3), 240–245. doi: 10.1016/j.actaastro.2007.07.001

Hu, L. Y., Huang, M. W., Ke, S. W., & Tsai, C. F. (2016). The distance function effect on k-nearest neighbor classification for medical datasets. *SpringerPlus*, *5*(1). doi: 10.1186/s40064-016-2941-7

Huang, Z., Shao, Z., Pan, J., & Ji, X. (2015). Berthing speed control law for large vessels based on AIS Data. In *Fifth international conference on transportation engineering.*

International Harbour Masters Association. (n.d.). *Harbour Master: Safety.* Retrieved from https://www.harbourmaster.org/harbour-master/safety

International Maritime Organization. (2020). *Introduction to IMO.* Retrieved from http://www.imo.org/en/About/Pages/Default.aspx

Itani, S., Rossignol, M., Lecron, F., & Fortemps, P. (2019). Towards interpretable machine learning models for diagnosis aid: A case study on attention deficit/hyperactivity disorder. *PLOS ONE*, *14*(4), 1–20. doi: 10.1371/journal.pone.0215720

Iyer, A., S, J., & Sumbaly, R. (2015). Diagnosis of Diabetes Using Classification Mining Techniques. *International Journal of Data Mining & Knowledge Management Process*, *5*(1), 01–14. doi: 10.5121/ijdkp.2015.5101

Jain, A. (2016). *Complete Machine Learning Guide to Parameter Tuning in Gradient Boosting (GBM) in Python.* Retrieved from https://www.analyticsvidhya.com/blog/2016/02/complete-guide-parameter-tuning-gradient-boosting-gbm-python/

Jianghao, C., & Degong, D. (2018). *Analysis of PIANC Guideline and ROM standard in design of appraoch channel and harbor bason* (Tech. Rep.). Panama: PIANC-World Congress Panama City.

Kaljouw, S. (2019). *Tugboat resting location optimization using AIS analysis* (Unpublished doctoral dissertation). Erasmus University Rotterdam.

Kendall, D. G. (1953). Stochastic Processes Occurring in the Theory of Queues and their Analysis by the Method of the Imbedded Markov Chain. *The Annals of Mathematical Statistics*, *24*(3), 338–354. doi: 10.1214/aoms/1177728975

Kilkenny, M. F., & Robinson, K. M. (2018). Data quality: “Garbage in – garbage out”. *Health Information Management Journal*, *47*(3), 103–105. doi: 10.1177/1833358318774357

Kim, J. G., Kim, H. J., Jun, H. B., & Kim, C. M. (2016). Optimizing ship speed to minimize total fuel consumption with multiple time windows. *Mathematical Problems in Engineering*. doi: 10.1155/2016/3130291

Kolmogorov-Smirnov test One- & two-sample, and related tests. (n.d.). Retrieved from https://influentialpoints.com/Training/kolmogorov-smirnov_test-principles-properties-assumptions.htm

Kox, S. (2016). *A Tool for Determining Marine Terminal Dimensions and Costs in a Project's Feasibility Phase.* (Doctoral dissertation, Technical University of Delft). doi: cab027b6-fe58-4be5-a97d-21d88e7fb0ec

Kujala, P., Hänninen, M., Arola, T., & Ylitalo, J. (2009, 8). Analysis of the marine traffic safety in the Gulf of Finland. *Reliability Engineering and System Safety*, *94*(8), 1349–1357. doi: 10.1016/j.ress.2009.02.028

Kuo, T. C., Huang, W. C., Wu, S. C., & Cheng, P. L. (2006). A case study of inter-arrival time distributions of container ships. *Journal of Marine Science and Technology*, *14*(3), 155–164.

K.W., S., Loh, N. C., Lin, C., Booth, V., Chan, J. W., Yip, A. C., … Lau, A. K. (2013). Policy change driven by an AIS-assisted marine emission inventory in Hong Kong and the Pearl River Delta. *Atmospheric Environment*, *76*, 101–112.

Lagueux, K. M., Zani, M. A., Knowlton, A. R., & Kraus, S. D. (2011). Response by vessel operators to protection measures for right whales Eubalaena glacialis in the southeast US calving ground. *Endangered Species Research*, *14*(1), 69–77. doi: 10.3354/esr00335

Lai, C.-d., Murthy, D., & Xie, M. (2006). Springer Handbook of Engineering Statistics. In *Springer handbook of engineering statistics* (pp. 63–78). doi: 10.1007/978-1-84628-288-1

Lampe, O. D., Kehrer, J., & Hauser, H. (2010). Visual analysis of multivariate movement data using interactive difference views. In *Vmv 2010 - vision, modeling and visualization* (pp. 1–8). doi: 10.2312/PE/VMV/VMV10/315-322

Lane, R. O., Nevell, D. A., Hayward, S. D., & Beaney, T. W. (2010). Maritime anomaly detection and threat assessment. In *13th conference on information fusion, fusion 2010.* doi: 10.1109/icif.2010.5711998

Ligteringen, H. (2017). *Ports and Terminals* (Second ed.). Delft: Delft Academic Press.

LNG Terminalling. (2019). Retrieved from https://www.fluxys.com/en/products-services/activities/lng

Löptien, U., & Axell, L. (2014, 7). Ice and AIS: ship speed data and sea ice forecasts in the Baltic Sea. *The Cryosphere Discussions*, *8*(4), 3811–3828. doi: 10.5194/tcd-8-3811-2014

Made Smart Group AIS Data Store. (n.d.). Retrieved from https://www.madesmart.nl/worlds-largest-ais-data-store/

Maheswari, K. (2019). Finding Best Possible Number of Clusters using K-Means Algorithm. *International Journal of Engineering and Advanced Technology*, *9*(1S3), 533–538. doi: 10.35940/ijeat.a1119.1291s419

Mao, S., Tu, E., Zhang, G., Rachmawati, L., Rajabally, E., & Huang, G.-B. (2018). *An Automatic Identification System (AIS) Database for Maritime Trajectory Prediction and Data Mining* (Tech. Rep.). Singapore. Retrieved from http://www.marinecadastre.gov/ais/.

Mašović, M. (2019). *Port Call Efficiency Optimization, Using Data Analysis, Process Mining and Discrete Event Simulation* (Unpublished doctoral dissertation). Technical University of Delft.

McNeese, D. B. (2016). *Deciding Which Distribution Fits Your Data Best.* Retrieved from https://www.spcforexcel.com/knowledge/basic-statistics/deciding-which-distribution-fits-your-data-best#:~:text=Distributionfittingistheprocess,andSmallestExtremeValuedistributions.&text=Statisticaltechniquesareusedtoestimatetheparametersof

Meijer, R. (2017). *ETA prediction Predicting the ETA of a container vessel based on route identification using AIS data* (Doctoral dissertation, Technical University of Delft). Retrieved from http://repository.tudelft.nl/.

Mjelde, A., Martinsen, K., Eide, M., & Endresen, O. (2014, 10). Environmental accounting for Arctic shipping - A framework building on ship tracking data from satellites. *Marine Pollution Bulletin*, *87*(1), 22–28. doi: 10.1016/j.marpolbul.2014.07.013

Mou, J. M., Tak, C. v. d., & Ligteringen, H. (2010, 5). Study on collision avoidance in busy waterways by using AIS data. *Ocean Engineering*, *37*(5-6), 483–490. doi: 10.1016/j.oceaneng.2010.01.012

Naus, K., Makar, A., & Apanowicz, J. (2007). Usage AIS Data for Analyzing Ship's Motion Intensity. *International Journal on Marine Navigation and Safety of Sea Transportation*, *1*(3).

Ni Ni, H., Hu, Q., & Shi, C. (2011). Studying Probability of ship arrival of Yangshan port with AIS. *International Journal on Marine Navigation and Safety of Sea Transportation*, *5*(3).

Norris, A. (2007). AIS Implementation: Succes or failure? *Journal of Navigation*, *60*(1), 1–10.

Okada, S. (2020). *Gentle Introduction to Chi-Square Test for Independence.* Retrieved from https://towardsdatascience.com/gentle-introduction-to-chi-square-test-for-independence-7182a7414a95

Omara, H., Lazaar, M., & Tabii, Y. (2018). Effect of Feature Selection on Gene Expression Datasets Classification Accuracy. *International Journal of Electrical and Computer Engineering (IJECE)*, *8*(5), 3194–3203. doi: 10.11591/ijece.v8i5.pp3194-3203

Pachakis, D., & Kiremidjian, A. S. (2003). Ship Traffic Modeling Methodology for Ports. *Journal of Waterway, Port, Coastal and Ocean Engineering*, *129*(5). doi: https://doi.org/10.1061/(ASCE)0733-950X(2003)129:5(193)

Pan, J., Jiang, Q., Hu, J., & Shao, Z. (2012). An AIS data visualization model for assessing maritime traffic situation and its applications. In *Procedia engineering* (Vol. 29, pp. 365–369). doi: 10.1016/j.proeng.2011.12.724

Pandas. (2020). *pandas.Dataframe.* Retrieved from https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html

Pandis, N. (2017). Logistic regression: Part 1. *American Journal of Orthodontics and Dentofacial Orthopedics*, *151*(4), 824–825. Retrieved from http://dx.doi.org/10.1016/j.ajodo.2017.01.017 doi: 10.1016/j.ajodo.2017.01.017

Parolas, I. (2016). *ETA prediction for containerships at the Port of Rotterdam using Machine Learning Techniques* (Unpublished doctoral dissertation). Technical University of Delft.

Patrous, Z. S. (2018). *Evaluating XGBoost for User Classification by using Behavioral Features Extracted from Smartphone Sensors* (Unpublished doctoral dissertation). KTH Royal Institute of Technology.

Perez, H. M., Chang, R., Billings, R., & Kosub, T. L. (n.d.). *Automatic Identification Systems (AIS) Data Use in Marine Vessel Emission Estimation* (Tech. Rep.). Retrieved from http://www.epa.gov/ttnchie1/conference/ei18/session6/perez.pdf

PIANC. (n.d.). *The World Association for Waterborne Transport Infrastructure.* Retrieved from https://www.pianc.org/about

PIANC WG 135. (2014). *Design principles for small and medium marine container terminals.* Retrieved from http://www.pianc.org

PIANC WG 184. (2019). *Design principles for dry bulk marine terminals.* Brussels. Retrieved from http://www.pianc.org

PIANC WG121. (2014). *Harbour Approach Channels Design Guidelines* (Tech. Rep.). Brussels: Maritime Navigation Commission. Retrieved from http://www.pianc.org

PIANC WG158. (2014). *Masterplans for the development of existing ports* (Tech. Rep.). Maritime Navigation Commission. Retrieved from http://www.pianc.org

Port of Rotterdam. (n.d.). *Sailing speed for inland shipping.* Retrieved from https://www.portofrotterdam.com/en/shipping/inland-shipping/other/sailing-speed-for-inland-shipping

Pre 1.5 Normal distribution. (n.d.). Retrieved from https://ocw.tudelft.nl/course-readings/pre-1-5-normal-distribution/

Qu, X., Meng, Q., & Suyi, L. (2011, 11). Ship collision risk assessment for the Singapore Strait. *Accident Analysis and Prevention*, *43*(6), 2030–2036. doi: 10.1016/j.aap.2011.05.022

Queuing theory: Definition, history & real-life applications. (2020). Retrieved from https://queue-it.com/blog/queuing-theory/#:~:text=Queuingtheory(orqueueingtheory,customer%2Cjob%2Correquest.

Rajabi, A., Khodadad Saryazdi, A., Belfkih, A., & Duvallet, C. (2018, 1). Towards Smart Port: An Application of AIS Data. In *Proceedings - 20th international conference on high performance computing and communications, 16th international conference on smart city and 4th international conference on data science and systems, hpcc/smartcity/dss 2018* (pp. 1414–1421). Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/HPCC/SmartCity/DSS.2018.00234

Rakshit, S. (2018). *Statistics with R Programming*. McGraw-Hill Education.

Rane, S. (2018). *The balance: Accuracy vs. Interpretability.* Retrieved from https://towardsdatascience.com/the-balance-accuracy-vs-interpretability-1b3861408062

Rawson, A., Rogers, E., Foster, D., & Phillips, D. (2014, 3). Practical application of domain analysis: Port of london case study. *Journal of Navigation*, *67*(2), 193–209. doi: 10.1017/S0373463313000684

Ray, C., Gallen, R., Iphar, C., Napoli, A., & Bouju, A. (2015). DeAIS project: Detection of AIS spoofing and resulting risks. *MTS/IEEE OCEANS 2015 - Genova: Discovering Sustainable Ocean Energy for a New World*. doi: 10.1109/OCEANS-Genova.2015.7271729

Redoutey, M., Scotti, E., Jensen, C., Ray, C., & Claramunt, C. (2008). Efficient vessel tracking with accuracy guarantees. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (Vol. 5373 LNCS, pp. 140–151). doi: 10.1007/978-3-540-89903-7{_}13

Renso, C., Spaccapietra, S., & Zimányi, E. (2013). *Mobility Data: Modeling, Management, and Understanding*.

Ristic, B., La Scala, B., Morelande, M., & Gordon, N. (2008). Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction. In *Proceedings of the 11th international conference on information fusion, fusion 2008.* doi: 10.1109/ICIF.2008.4632190

Robards, M. D., Silber, G. K., Adams, J. D., Arroyo, J., Lorenzini, D., Schwehr, K., & Amos, J. (2016, 1). Conservation science and policy applications of the marine vessel Automatic Identification System (AIS)-A review. *Bulletin of Marine Science*, *92*(1), 75–103. doi: 10.5343/bms.2015.1034

Rong, H., Teixeira, A., & Soares, C. G. (2015). Simulation and analysis of maritime traffic in the Tagus River Estuary using AIS data. In *Maritime technology and engineering* (pp. 185–194). London, UK: Taylor and Francis Group.

Royle, J. A., & Kéry, M. (2016). Modeling Static Occurrence and Species Distributions Using Site-occupancy Models. In *Applied hierarchical modeling in ecology* (pp. 551–629). Academic Press. doi: https://doi.org/10.1016/B978-0-12-801378-6.00010-2

Sampath, P., & Parry, D. (2013). *Trajectory Analysis using Automatic Identification Systems in New Zealand Waters* (Tech. Rep.). Retrieved from http://gpsd.berlios.de/AIVDM.html

Saxena, R. (2017). *How the naive bayes classifier works in machine learning.* Retrieved from https://dataaspirant.com/2017/02/06/naive-bayes-classifier-machine-learning/%0D

Schmelzer, R. (2019). *Understanding Explainable AI.* Retrieved from https://www.forbes.com/sites/cognitiveworld/2019/07/23/understanding-explainable-ai/#4d39a5967c9e

Schøyen, H., Hjelmervik, K., Wang, H., & Osen, O. (2017, 7). Measuring Container Port Complementarity and Substitutability with Automatic Identification System (AIS) Data – Studying the Inter-port Relationships in the Oslo Fjord Multi-port Gateway Region. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, *11*(2), 79–84. doi: 10.12716/1001.11.02.08

Schwehr, K. D., & McGillivary, P. A. (2007). Marine ship automatic identification system (AIS) for enhanced coastal security capabilities: An oil spill tracking application. In *Oceans conference record (ieee).* doi: 10.1109/OCEANS.2007.4449285

scipy.stats.gamma. (2019). Retrieved from https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.gamma.html#scipy.stats.gamma

scipy.stats.kstest. (2019). Retrieved from https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kstest.html#:~:text=PerformtheKolmogorov-Smirnovtest,givendistributionG(x).

Sea-web Ships. (2020). *The Ultimate Maritime Reference Tool.* Retrieved from https://maritime.ihs.com/entitlementportal/home/information/seaweb_ships#verticalTab2

Service Plans Big Ocean Data. (n.d.). Retrieved from https://www.bigoceandata.com/try-us/service-plans/

Services MarineTraffic. (n.d.). Retrieved from https://www.marinetraffic.com/en/online-services/plans/comparison-list

SHAP- Explainers. (2018). Retrieved from https://shap.readthedocs.io/en/latest/

Sharma, N. (2020). *Understanding the Mathematics Behind Naive Bayes.* Retrieved from https://heartbeat.fritz.ai/understanding-the-mathematics-behind-naive-bayes-ab6ee85f50d0

Shelmerdine, R. L. (2015, 4). Teasing out the detail: How our understanding of marine AIS data can better inform industries, developments, and planning. *Marine Policy*, *54*, 17–25. doi: 10.1016/j.marpol.2014.12.010

Shen, W. (2012). *A statistical analysis of average vessel behaviour by AIS data* (Unpublished doctoral dissertation). UNESCO-IHE Institute for Water Education.

Shipfinder Prices. (n.d.). Retrieved from http://www.shipfinder.com/Home/Price

Shu, Y., Daamen, W., Ligteringen, H., & Hoogendoorn, S. (2013, 1). Vessel speed, course, and path analysis in the Botlek area of the port of Rotterdam, Netherlands. *Transportation Research Record*(2330), 63–72. doi: 10.3141/2330-09

Shu, Y., Daamen, W., Ligteringen, H., & Hoogendoorn, S. P. (2017). Influence of external conditions and vessel encounters on vessel behavior in ports and waterways using Automatic Identification System data. *Ocean Engineering*, *131*, 1–14. doi: 10.1016/j.oceaneng.2016.12.027

Shucksmith, R. J., & Shelmerdine, R. L. (2015, 9). A risk based approach to non-native species management and biosecurity planning. *Marine Policy*, *59*, 32–43. doi: 10.1016/j.marpol.2015.05.001

Siguenza-guzman, L., Saquicela, V., & Vandewalle, J. (n.d.). *Literature review of data mining applications in academic libraries* (No. 0). Leuven, Belgium.

Silber, G. K., Adams, J. D., & Fonnesbeck, C. J. (2014). Compliance with vessel speed restrictions to protect North Atlantic right whales. *PeerJ*, *2014*(1). doi: 10.7717/peerj.399

Sisodia, D., Shrivastava, S., & Jain, R. (2010). ISVM for face recognition. In *2010 international conference on computational intelligence and communication network.* doi: 10.1109/CICN.2010.109

Sisodia, D., & Sisodia, D. S. (2018). Prediction of Diabetes using Classification Algorithms. *Procedia Computer Science*, *132*(Iccids), 1578–1585. Retrieved from https://doi.org/10.1016/j.procs.2018.05.122 doi: 10.1016/j.procs.2018.05.122

Slundberg. (n.d.). *GitHub slundberg/shap.* Retrieved from https://github.com/slundberg/shap

Sotirov, S., & Alexandrov, C. (2017). Improving AIS data reliability. In *18th annual general assembly of the international association of maritime universities - global perspectives in met: Towards sustainable, green and integrated maritime transport, iamu 2017* (Vol. 1, pp. 237–244).

Stephanie. (2016). *IID Statistics: Independent and Identically Distributed Definition and Examples.* Retrieved from https://www.statisticshowto.com/iid-statistics/

Taboga, M. (2017). *Beta distribution.* Retrieved from https://www.statlect.com/probability-distributions/beta-distribution#:~:text=TheBetadistributionisa,ofsuccessofanexperiment.

Talavera, A., Aguasca, R., Galván, B., & Cacereño, A. (2013). Application of Dempster-Shafer theory for the quantification and propagation of the uncertainty caused by the use of AIS data. *Reliability Engineering and System Safety*, *111*, 95–105. doi: 10.1016/j.ress.2012.10.007

Thoresen, C. A. (2003). *Port Designer's Handbook: Recommendations and Guidelines.*

Tsou, M.-C. (2010). Discovering Knowledge from AIS Database for Application in VTS. *The Journal of Navigation*, *63*(3), 449–469.

Tu, E., Zhang, G., Rachmawati, L., Rajabally, E., & Huang, G.-B. (2016, 6). Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey. Retrieved from http://arxiv.org/abs/1606.00981

Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making*, *19*(1), 1–16. doi: 10.1186/s12911-019-1004-8

UNCTAD. (1985). *Port development: a handbook for planners in developing countries*. United Nations.

Valsamis, A., Tserpes, K., Zissis, D., Anagnostopoulos, D., & Varvarigou, T. (2017, 5). Employing traditional machine learning algorithms for big data streams analysis: The case of object trajectory prediction. *Journal of Systems and Software*, *127*, 249–257. doi: 10.1016/j.jss.2016.06.016

Van Asperen, E., Dekker, R., Polman, M., & Swaan Arons, H. d. (2003). Modeling ship arrivals in ports. In *Winter simulation conference proceedings* (Vol. 2, pp. 1737–1744). Rotterdam. doi: 10.1109/wsc.2003.1261627

Van Koningsveld, M., & Uijl, J. d. (2020). *OpenQTSim.* Retrieved from https://zenodo.org/record/3706930 doi: 10.5281/ZENODO.3706930

van Vianen, T. (2015). *Simulation-integrated Design of Dry Bulk Terminals* (Unpublished doctoral dissertation). Delf University of Technology.

Van Zwieteren, G. (2020). *TUDelft-CITG/AIS-port-processes-tool.* Delft. Retrieved from https://zenodo.org/record/4055787 doi: 10.5281/ZENODO.4055787

Wicklin, R. (2019). *What is Kolmogorov's D statistic?* Retrieved from https://blogs.sas.com/content/iml/2019/05/15/kolmogorov-d-statistic.html

Wijaya, W. M., & Nakamura, Y. (2013). Predicting ship behavior navigating through heavily trafficked fairways by analyzing AIS data on apache HBase. In *Proceedings - 2013 1st international symposium on computing and networking, candar 2013* (pp. 220–226). doi: 10.1109/CANDAR.2013.39

Wiley, D., Hatch, L., Thompson, M., Schwehr, K., & Macdonald, C. D. (2013). Marine Sanctuaries and Marine Planning: Protecting endangered marine life. *National Ocean Policy*. Retrieved from https://www.researchgate.net/publication/258554984

Willems, N., Van De Wetering, H., & Van Wijk, J. J. (2009). Visualization of vessel movements. *Computer Graphics forum*, *28*(3), 959–966.

Windward. (2014). *AIS Data on the High Seas: An Analysis of the Magnitude and Implications of Growing Data Manipulation at Sea* (Tech. Rep.). Windward EU.

Winther, M., Christensen, J. H., Plejdrup, M. S., Ravn, E. S., Eriksson, O. F., & Kristensen, H. O. (2014). Emission inventories for ships in the arctic based on satellite sampled AIS data. *Atmospheric Environment*, *91*, 1–14. doi: 10.1016/j.atmosenv.2014.03.006

Xiao, F. (2014). *Ships in an Artificial Force Field: A Multi-agent System for Nautical Traffic and Safety* (Doctoral dissertation, Delft University of Technology). Retrieved from www.wps.nl

Xiao, F., Ligteringen, H., Van Gulijk, C., & Ale, B. (2013). *Comparison study on AIS data of ship traffic behavior* (Tech. Rep.). Delft University of Technology.

Young, B. L., & Huddleston, S. H. (2017). *Predicting vessel trajectories from AIS data using R* (Unpublished doctoral dissertation). Naval PostGradaute School California.

Zamanirad, S., Mazaheri, S., & Ghafourian, M. (2017). Introducing a Method for More Precise Prediction of Berth Occupancy Ratio in Bulk Liquid Terminals. *International Journal of coastal and offshore engineering*, *1*(2), 21–26. doi: 10.18869/acadpub.ijcoe.1.2.21

Zhang, W., Goerlandt, F., Kujala, P., & Wang, Y. (2016, 9). An advanced method for detecting possible near miss ship collisions from AIS data. *Ocean Engineering*, *124*, 141–156. doi: 10.1016/j.oceaneng.2016.07.059

Zhang, W., Goerlandt, F., Montewka, J., & Kujala, P. (2015, 12). A method for detecting possible near miss ship collisions from AIS data. *Ocean Engineering*, *107*, 60–69. doi: 10.1016/j.oceaneng.2015.07.046

Zhao, Z., Ji, K., Xing, X., Zou, H., & Zhou, S. (2014, 1). Ship surveillance by integration of space-borne SAR and AIS - Review of current research. *Journal of Navigation*, *67*(1), 177–189. doi: 10.1017/S0373463313000659

Zhou, Y., Daamen, W., Vellinga, T., & Hoogendoorn, S. (2015). Vessel classification method based on vessel behavior in the port of Rotterdam. *Scientific Journals of the Maritime University of Szczecin*, *42*(114), 86–92.

Zhou, Y., Daamen, W., Vellinga, T., & Hoogendoorn, S. (2017). *AIS data analysis for the impacts of wind and current on ship behavior in straight waterways* (Tech. Rep.). Delft: Department of Hydraulic Engineering, Delft University of Technology. Retrieved from https://www.researchgate.net/publication/324007907

A

Appendix A: AIS sources & literature

A.1. AIS Data Sources

A few web-based data sources of popular AIS providers (take note that much more providers exist):

- Shipfinder: Free access for AIS data with a 12h delay. Vessel tracking, arrival alerts and overviews of vessels in port can be accessed at a monthly price depending on the requested features. No data extractions are possible (*Shipfinder Prices*, n.d.).
- MarineTraffic: Free access for the last and next 3 days of vessels. Various packages available for viewing vessels with a maximum of 365 days back. The monthly data export allows 3000 rows (*Services MarineTraffic*, n.d.).
- Made Smart Group: AIS data upon request, they claim to be the World's Largest AIS Data Store. Services include vessel incident analysis, vessel traffic movement and density analysis, vessel maneuverability analysis, historical weather and sea state (*Made Smart Group AIS Data Store*, n.d.).
- Big Ocean Data: Offers a free trial and multiple service plans. The most expensive service plans allows data extraction (*Service Plans Big Ocean Data*, n.d.).
- FleetMon: Real time vessel tracking, for free for 15 minutes per day for a max of 10 vessels. Different packages for different services. API Data packages are available on request for certain prices, it might allow some small data packages for students for free (*API Data Packages*, n.d.).
- ExactEarth: Offers various packages, real time, density mapping and archive possibilities. Upon request and for varying prices (*ExactAIS*, n.d.).
- VesselFinder: By connecting an individual AIS-receiving station VesselFinder allows you to access all the AIS data that is delivered by other partners. In total there are more than 200 stations (*AIS Partner Stations*, n.d.).

A.2. Previous AIS research

A.2.1. Categorizations of AIS research

A classification is made between the different research categories:

- AIS possibilities: AIS processing, AIS problem analysis, AIS data analysis, performance analysis (of space-based AIS)
- Anomaly detection
- Environmental studies: ship emissions, vessel effects on whales, bio-fuel, oil spills detection, ice forecasting
- Literature review
- Port performance: port competition, port call efficiency, ship arrival distribution
- Predictions: predicting ETA, behavioral predictions, vessel trajectories predictions
- Route analysis: spatial planning, vessel trajectories, vessel patters, traffic visualization, traffic simulation
- Safety and risk assessment: speed control, near miss detection / collision prediction, (collision) risk assessment, performance analysis in case of collision avoidance
- Vessel behavior: influence external factors, movement intensity, ship domain analysis

A.2.2. AIS research over the years

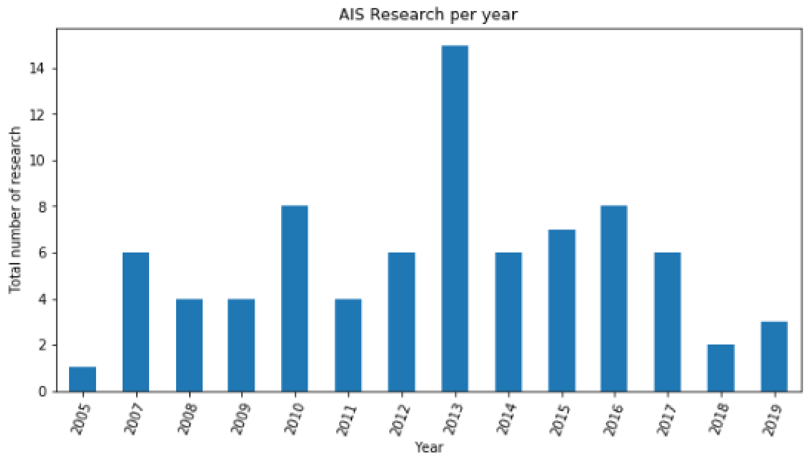


Figure A.1: AIS research per year

A.2.3. References from AIS literature review

The previous research so far that benefited from AIS data:

- AIS possibilities: (Rajabi et al., 2018), (Shelmerdine, 2015), (Harati-Mokhtari et al., 2007), (Redoutey et al., 2008), (Harati-Mokhtari et al., 2008), (Felski & Jaskólski, 2012), (Høye et al., 2007), (Guerriero et al., 2010), (Tsou, 2010), (Braca et al., 2013), (De Vreede, 2016)
- Environmental studies: (D. Chen et al., 2016), (Guzman et al., 2013), (Shucksmith & Shelmerdine, 2015), (K.W. et al., 2013), (Winther et al., 2014), (Mjelde et al., 2014), (Lagueux et al., 2011), (Silber et al., 2014), (Wiley et al., 2013), (Ferraro et al., 2007), (Ferraro et al., 2010), (Schwehr & McGillivary, 2007), (Löptien & Axell, 2014)
- Literature review: (Robards et al., 2016), (Zhao et al., 2014), (Norris, 2007)
- Port performance: (Schøyen et al., 2017), (Mašović, 2019), (Ni Ni et al., 2011), (L. Chen et al., 2015)
- Predictions: (Meijer, 2017), (Parolas, 2016), (Wijaya & Nakamura, 2013), (Valsamis et al., 2017), (Young & Huddleston, 2017), (Mao et al., 2018), (Dobrkovic et al., 2015)
- Route analysis: (Fiorini et al., 2016), (Kaljouw, 2019), (Xiao et al., 2013), (Pan et al., 2012), (Talavera et al., 2013), (Sampath & Parry, 2013), (Heymann et al., 2013), (Xiao, 2014), (Rong et al., 2015), (Bomberger et al., 2005), (Ristic et al., 2008), (Aarsæther & Moan, 2009), (Lampe et al., 2010), (Lane et al., 2010).
- Safety and risk assessment: (Zhang et al., 2016), (Huang et al., 2015), (Grossmann, 2019), (Tu et al., 2016), (Zhang et al., 2015), (Qu et al., 2011), (Kujala et al., 2009), (Rawson et al., 2014), (Mou et al., 2010), (Goerlandt & Kujala, 2011), (Baldauf et al., 2008), (Hornauer & Hahn, 2013).
- Vessel behavior: (de Boer, 2010), (Zhou et al., 2017), (Naus et al., 2007), (Hansen et al., 2013), (Willems et al., 2009), (Shu et al., 2017), (Shen, 2012), (Shu et al., 2013), (Zhou et al., 2015).

B

Appendix B: Vessel classification

In order to analyse smaller subsets of AIS data for terminals, the total data set will be split into smaller sub data sets based on the vessel classifications. The classification is determined by the port authority or an equivalent party, or can follow the classification defined by classification societies (Zhou et al., 2015). For this research the following classifications will be used, based on current vessel class classifications made by *RHDHV*.

Class No.	Vessel Class	LOA [m]
1	Small Feeder	< 145
2	Regional Feeder	145 - 185
3	Feedermax + Panamax	185 - 223
4	New Panamax	223 – 366
5	Post New Panamax	> 366

Table B.1: Container vessel classification, RHDHV

Class No.	DWT
1	< 5,000
2	5,000 - 10,000
3	10,000 - 15,000
4	15,000 - 20,000
5	20,000 - 30,000
6	> 30,000

Table B.2: General cargo vessel classification, RHDHV

Class No.	Vessel Class	DWT
1	Small handy	< 10,000
2	Handy + Handymax + Supramax	10,000 - 65,000
3	Panamax	65,000 - 85,000
4	Mini Capesize + Capesize	85,000 - 200,000
5	Very Large Bulk Carrier (VLBC) / Very Large Ore Carrier (VLOC)	> 200,000

Table B.3: Dry bulk vessel classification, RHDHV

Class No.	Vessel Class	LOA [m]
1	Small Spherical / Membrane LNG Carriers	< 250
2	Medium Spherical / Membrane LNG Carriers	250 - 275
3	Large Spherical / Membrane Carriers	275 - 300
4	Very large Spherical / Membrane LNG Carriers	> 300

Table B.4: LNG vessel classification, RHDHV

C
Appendix C: Terminal locations

Visualised in the figures below are the 12 chosen terminals for this research. Polygons are drawn over the entire terminal and the terminal area.

C.1. Container terminals

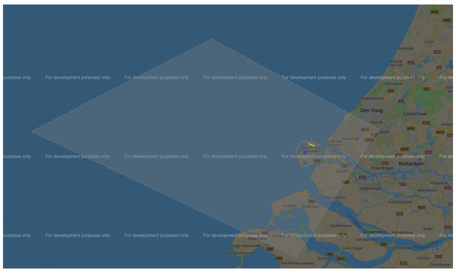


Figure C.1: CT: Rotterdam APM-2 Terminal (blue terminal)

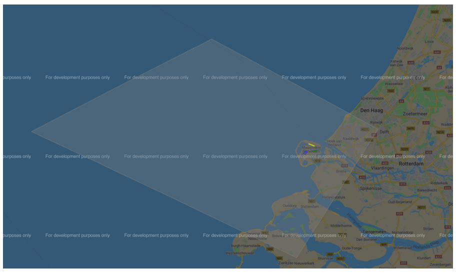


Figure C.2: CT: Rotterdam APM Main Quay (green terminal)

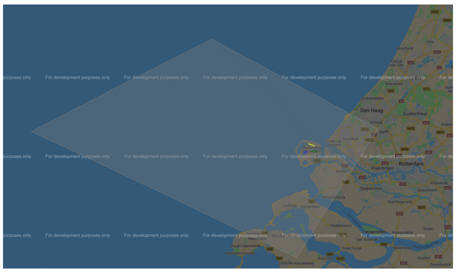


Figure C.3: CT: Rotterdam EuroMax Terminal (yellow terminal)

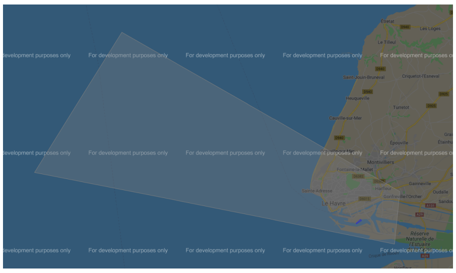


Figure C.4: CT: Le Havre Atlantic Container Terminal

C.2. Dry bulk terminals

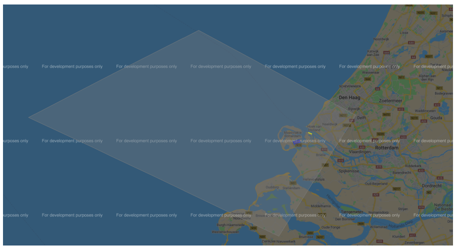


Figure C.5: DBT: Rotterdam EMO Terminal (blue terminal)

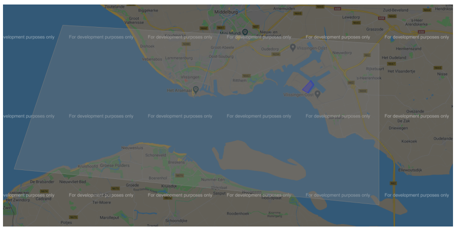


Figure C.6: DBT: Vlissingen OVET Terminal

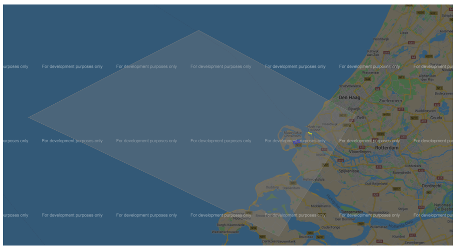


Figure C.7: DBT: Rotterdam EECV Terminal (yellow terminal)

Figure C.8: DBT: Dunkirk Western Bulk Terminal

C.3. Liquid bulk terminals

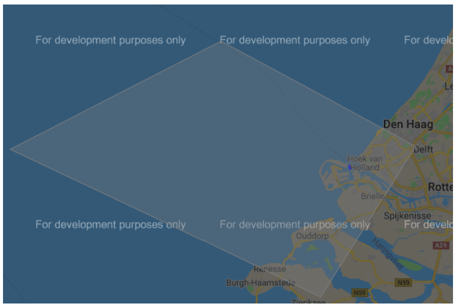


Figure C.9: LBT: Rotterdam GATE Terminal

Figure C.10: LBT: Zeebrugge LNG Terminal

Figure C.11: LBT: Dunkirk LNG Terminal

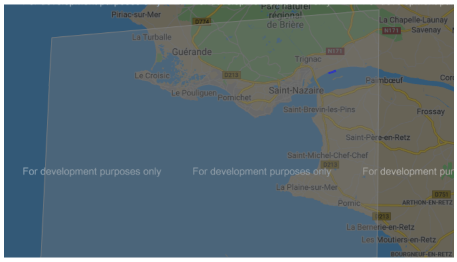


Figure C.12: LBT: France Montoir LNG Terminal

D

Appendix D: AIS tool terminal information

D.1. Results processing and cleaning terminals using AIS tool

Terminal locations	Rotterdam APM2	Rotterdam APM	Rotterdam Euromax	Le Havre Atlantic
Initial rows	65214798	65214798	65214798	15334176
Category filtering removed	46422855	46422855	46422855	12239297
After category filtering	18791943	18791943	18791943	3094879
Rows in terminal polygon	564679	557181	757182	139617
Number of duplicates (terminal)	143788	100837	7162	15527
Number of faulty inputs (terminal)	0	0	0	0
Number of location outliers (terminal)	1061	451	166	104
Left after cleaning terminal	41983	455893	683902	123986
Number of duplicates (port)	1811575	1811575	1811575	358158
Number of faulty inputs (port)	0	0	0	0
Left after cleaning port	16980368	16980368	16980368	2736721
Number of speed outliers (terminal)	312	43	93	15
Left after processing terminal	419518	45585	683809	123971
Number of tracks terminal	3534	3388	3997	496
Obvious non-berthing tracks	577	409	322	93
Number of berthed vessel tracks	1979	2132	3029	384
Rows with all data from berthed tracks	368816	40382	643503	123092
Time 5,run_all_steps [min]	243.17	306.53	1067.75	28.2

Table D.1: Processing and cleaning information: container terminals

Terminal locations	Rotterdam EMO	Vlissingen OVET	Rotterdam EECV (Northern/ Outer Quay)	Dunkirk Western Bulk
Initial rows	65214798	16465317	65214798	5578930
Category filtering removed	48871959	12308326	48871959	4656812
After category filtering	16342839	4156991	16342839	922118
Rows in terminal polygon	465651	81479	473752	35277
Number of duplicates (terminal)	21735	5889	16777	1849
Number of faulty inputs (terminal)	0	0	0	0
Number of location outliers (terminal)	696	88	163	15
Left after cleaning terminal	443220	75502	14613917	33413
Number of duplicates (port)	1728922	443579	1728922	38510
Number of faulty inputs (port)	0	0	0	0
Left after cleaning port	14613917	3713412	6746050	883608
Number of speed outliers (terminal)	76	10	6	4
Left after processing terminal	443144	75492	456806	33409
Number of tracks terminal	2460	756	2518	109
Obvious non-berthing tracks	541	294	1721	5
Number of berthed vessel tracks	1342	144	543	94
Rows with all data from berthed tracks	410744	72082	430554	32988
Time 5,run_all_steps [min]	105,85	14,766	96,38	4,77

Table D.2: Processing and cleaning information: dry bulk terminals

Terminal locations	Rotterdam Gate	Zeebrugge	Dunkirk	France Montoir
Initial rows	65214798	11754861	5578930	3342845
Category filtering removed	40390110	9400689	4744746	2520993
After category filtering	24824688	2354172	834184	821852
Rows in terminal polygon	62107	58624	9669	22058
Number of duplicates (terminal)	3669	5649	748	2106
Number of faulty inputs (terminal)	0	0	0	0
Number of location outliers (terminal)	37	152	3	4
Left after cleaning terminal	58401	52823	8918	19948
Number of duplicates (port)	2172890	75552	29127	16664
Number of faulty inputs (port)	0	0	0	0
Left after cleaning port	22651798	2278620	805057	805188
Number of speed outliers (terminal)	8	9	2	0
Left after processing terminal	58393	52814	8916	19948
Number of tracks terminal	240	392	91	400
Obvious non-berthing tracks	40	53	8	266
Number of berthed vessel tracks	173	185	68	122
Rows with all data from berthed tracks	55627	49820	8400	18842
Time 5,run_all_steps [min]	98,48	12,57	2,71	3,38

Table D.3: Processing and cleaning information: liquid bulk terminals

12 terminals together	Average
Category filtering removed [%]	75.32
Rows in terminal vs total rows [%]	2.72
Number of times port rows compared to terminal rows	68.68
Average size reduction (based on kB) after parameter rounding [%]	19.2
Duplicates removed from terminal data [%]	9.09
Duplicates removed from port data [%]	7.83
False inputs from terminal data [%]	0.00
False inputs from port data [%]	0.00
Location outliers from terminal data [%]	0.10
Speed outliers from terminal data [%]	0.07
Cleaned and enriched terminal data vs raw terminal data [%]	83.94
Obvious vessel tracks removed vs total vessel tracks [%]	24.54
Berthed vessel tracks vs total vessel tracks [%]	56.50

Table D.4: Overview of average values for all 12 terminals together (based on tables D.1, D.2 and D.3)

E

Appendix E: Extracting berthed vessel tracks

In order to remove vessel tracks from the terminal area that do not berth, four different methods are investigated to see which method best extracts the berthed vessel tracks. The problem of extracting all of the berthed vessel tracks can be seen as a classification problem. A classification problem is a problem where a certain category is undefined for a new data set. In this case the undefined parameter is whether or not the vessel has berthed or not. Once a method is defined which predicts this undefined parameter, a labelled data set (the validation set) can be used to test the quality of the prediction (GeeksforGeeks, n.d.). A classification problem can be solved in two ways: either by defining the rules manually, or by using the data to learn what the rules are. The method of defining limits to classify vessels lying still, is first attempted due to it being most interpretative.

E.1. Approach I: Limits to define vessel lying still - basic data

In order to extract the berthed vessels an approach was generated to identify the vessels that berth. In this first approach the original data containing *MMSI*, latitude, longitude and timestamp is used, where there has been no filtering based on categories.

As mentioned by PIANC WG 135 (PIANC WG 135, 2014), the service time is defined as the entire process of the vessel at the berth including (un)berthing, (un)mooring and (un)loading. It is expected that vessels berthing at the terminal will spend a relatively long amount of time at the quay wall, compared to vessels not berthing at the terminal. Tug boats assist the vessels in their maneuvering towards the berth and are therefore assumed to move around a lot, not lying still for a longer amount of time (Ligteringen, 2017). Other vessels, such as cargo vessels that pass by and do not berth, are also expected to not lie still in the terminal area for a long time. Based on these assumptions the berthing of a vessel is defined as follows:

- Berthing = A vessel should ’lie still’ for at least *Duration limit* hours
- Lie still = A vessel is classified as ’lying still’ when:
  - The distance between successive data messages is less than the *Distance limit*
  - The speed between successive data messages is less than the *Speed limit*

This approach returned three different limits to be determined: Duration limit, Distance limit and Speed limit. Multiple combinations of limits were tested to find which combination best suits the determination of ’berthing’.

In order to validate the results the Sea-web database is used. Sea-web Ships is an online database and service which provides its users to extract data from different databases (Sea-web Ships, 2020). From Sea-web for certain terminals, data can be extracted for a given time span, which include the MMSI number, the time the vessel arrived at the port and the sailed time (the timestamp the vessel left).

The following steps are taken:

- First, the vessel tracks are labelled based on the first moment a vessel enters the terminal polygon until the moment the vessel leaves the polygon.
- Next, all non-valid rows are equal to 0, these will be removed. These are AIS messages (rows) that do not enter the smaller polygon. The total duration that a vessel has in the polygon is calculated, and added as a new column. When a vessel track only sends one message during their stay in the smaller polygon, this total duration will be equal to zero, all these vessel tracks are removed.
- The distance between two successive rows is determined. Afterwards, per row of each vessel track the limits based on the vessel ’lying still’ are checked. If the row satisfies both limits it is categorised as ’lying still’.
- The total time of continuous lying still is calculated for each vessel track. If for any part (any subset of rows) the duration of lying still exceeds the duration limit, the entire vessel track will be classified as ’berthed’.

- As mentioned, the tool can be validated with the Sea-web data base. The two data frames are merged together.
- Once the data frame is merged, a confusion matrix, a performance measure for machine learning models, is plotted to visualise how well the tool predicts vessels that actually berth (based on Sea-web port data).

When running multiple limit combinations, the problem arises that when the limits are becoming too 'strict' a lot of berthed vessels are not represented by the tool. On the other hand when the limits are more 'loose' the tool interprets tug boats as berthed vessels as well. As an example, a container terminal from the South-Hampton port is used. A selection of run parameters is visualised in table E.1. The accuracy represents the correct predictions (for both not-berthed as berthed vessels), divided by the total amount of vessel tracks. Corrected predicted berths represents the number of correctly predicted berths divided by the total number of actual berthed vessels. False predicted berths represents the number of vessels that where predicted to be berthed by the tool, divided by the total number of actual berthed vessels.

Distance limit [m]	Speed limit [m/s]	Duration limit [s]	Accuracy [%]	Corrected predicted berths [%]	False predicted berths [%]
4	0.2	3600	97.226	72.619	17.85
5	0.2	3600	97.445	77.381	19.04
10	0.2	3600	97.664	84.524	22.62
15	0.2	3600	97.664	85.714	23.81
10	0.1	3600	97.664	83.333	21.429
10	0.3	3600	97.737	84.524	21.429
10	0.5	3600	97.737	84.524	21.429
10	1.0	3600	97.737	84.524	21.429
10	0.5	1800	97.372	90.476	33.333

Table E.1: South-Hampton: predicting qualities of first approach with different limits

From this table it is clear that by using these three limits the overall accuracy is high (> 97%). It must be noted that this accuracy also represents the number of correctly predicted not-berthed vessels. From the total number of vessel tracks (1370), a large section (93.87%) represents vessel that will not berth (1286 vessel tracks). The research objective focuses solely on the vessels that berth at the terminals, thus it is important to zoom into those predictive qualities. From the confusion matrix in Figure E.1, the false negatives (23) represent the vessel tracks which are predicted to have not berthed (*predicted label = 0*) but actually did berth (*true label = 1*) and the false positives (15) represent the opposite, vessels that did not berth, but were predicted as if they did berth. As mentioned earlier, the last two columns from table E.1 present the predictive qualities for these. When trying different sets of limits, when the corrected number of berths improves, the number of wrongfully predicted number of berths degrades (increases in percentage, leading to more wrongfully predicted berths). When regarding only one limit, an optimum could be found graphically, however the situation includes three limits which makes finding the optimum more complex.

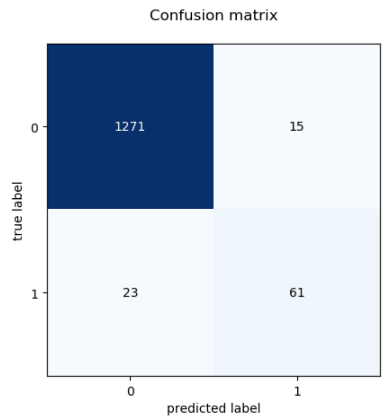


Figure E.1: Confusion matrix representing inputs from first row of table E.1 (South-Hampton)

Besides the previous mentioned issues, this example only focuses on one terminal. Eventually, a tool is required which is as generic as possible and ideally the limits will not have to be adjusted for every terminal. In table E.2 three sets of data are compared. Firstly, the above South-Hampton container terminal. Secondly, two container terminals together: South-Hampton and Singapore Brani Terminal. And third, three different terminals together: South-Hampton container terminal, Lisbon Dry bulk terminal and Rotterdam Liquid Bulk Terminal. All are tested for a distance limit of 10 meters, a speed limit of 0.5 m/s and a duration limit of 3600 seconds.

Location(s)	Accuracy [%]	Correctly predicted berths [%]	Falsely predicted berths [%]
South-Hampton	97.737	84.524	21.429
South-Hampton and Singapore	89.332	59.556	25.6
South-Hampton, Lisbon and Rotterdam	95.113	86.55	54.971

Table E.2: South-Hampton: predicting qualities of first approach with different limits

From this table it can be seen whilst the overall accuracy is still sufficiently high, the percentage of correctly predicted berths varies extremely for the second situation (only 59.56%), and for the third situation the number of falsely predicted berthed vessels is more than the half the total number of vessel berths (54.97%)! The extraction of berthed vessel tracks is very important for reaching the research objective and thus a different approach will be tested in order to try and improve the predictions of vessels berthing or not, to eventually extract all non-berthed vessels.

E.2. Approach II: Limits to define vessel lying still - filtered data

The second approach uses the same defined limits, but with different input data. Data filtered on vessel type will be used, which means all tug and pilot boats, among others, are filtered out. The same steps are performed in the same order. Multiple locations have been selected for different variations of the three limits. A concise overview of these results follows.

E.2.1. Container Terminals

Distance limit [m]	Speed limit [m/s]	Duration limit [s]	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
5	0.5	3600	77.367	25.793	11.185
10	0.5	3600	79.049	41.486	20.534
10	1.0	3600	79.027	41.486	20.618
10	1.0	1800	78.805	50.250	30.217

Table E.3: Container Terminal: Rotterdam Euromax

Distance limit [m]	Speed limit [m/s]	Duration limit [s]	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
5	0.5	3600	55.991	32.940	13.911
10	0.5	3600	62.340	58.005	27.297
10	1.0	3600	62.197	57.743	27.297
10	1.0	1800	61.912	62.598	32.677

Table E.4: Container Terminal: Barcelona BEST + Lisbon Container Terminals

E.2.2. Dry Bulk Terminals

Distance limit [m]	Speed limit [m/s]	Duration limit [s]	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
5	0.5	3600	77.983	62.222	413.333
10	0.5	3600	75.380	66.667	471.111
2	0.5	3600	88.395	22.222	160.000
5	1.0	3600	77.983	62.222	413.333
5	1.0	1800	72.234	71.111	540.000

Table E.5: Dry bulk terminal: Vlissingen OVET

Distance limit [m]	Speed limit [m/s]	Duration limit [s]	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
5	0.5	3600	82.609	34.472	62.422
10	0.5	3600	81.385	45.652	82.609
5	1.0	3600	82.651	34.783	62.422
2	1.0	3600	84.466	13.354	27.640
5	1.0	1800	81.385	46.273	83.230

Table E.6: Dry bulk terminal: Vlissingen OVET + New Holland Terminal

E.2.3. Liquid Bulk Terminals

Distance limit [m]	Speed limit [m/s]	Duration limit [s]	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
5	0.5	3600	92.453	66.667	11.111
10	0.5	3600	94.340	83.333	16.667
10	1.0	3600	94.340	83.333	16.667
10	0.2	3600	94.340	88.889	22.222
10	0.5	1800	87.736	83.333	55.556
10	0.5	7200	90.566	61.111	16.667

Table E.7: Liquid bulk terminal: Lisbon REPSOL

Distance limit [m]	Speed limit [m/s]	Duration limit [s]	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
5	0.5	3600	87.258	60.563	25.352
10	0.5	3600	86.981	67.606	33.803
3	0.5	3600	87.258	46.479	11.268
5	1.0	3600	87.535	60.563	23.944
5	0.2	3600	87.535	61.972	25.352
5	0.5	1800	83.102	74.648	60.563
5	0.5	7200	87.258	52.113	16.901

Table E.8: Liquid bulk terminal: Vlissingen TOTAL + Belfast

E.2.4. Terminals combined

Distance limit [m]	Speed limit [m/s]	Duration limit [s]	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
5	0.5	3600	71.285	29.678	13.280
10	0.5	3600	73.566	50.604	27.565
10	1.0	3600	73.531	50.503	27.565
10	1.0	1800	72.357	54.024	34.507

Table E.9: Combined terminals: Barcelona CT + New Holland DBT + Belfast LBT

Regarding the corrected predicted berths percentage, the model is not performing well. For locations a corrected predicted berths could be reached of about 89 % however in those situations the false number of predicted berths would be too large, and vice versa. Again, the same problem arises as was present in the first approach and thus the conclusion is made that this method is again not sufficient enough to robustly extract not berthed vessel tracks.

E.3. Approach III: Machine learning algorithm - basic data

A machine learning approach will be used, because the problem of extracting berthed vessels can not be solved based on manual inputs. The data used will first only be the generic data containing **no** vessel category filtering, and using the *MMSI*, latitude, longitude and timestamp. Machine learning uses programmed algorithms which learn from existing data to make acceptable predictions. A distinction can be made between three types of machine learning:

- Unsupervised
- Semi-supervised
- Supervised

Unsupervised machine learning is where the machine will have to learn based on unlabelled data, whereas supervised machine learning is based on training the model based on labelled data. In this research, data through Sea-web is imported, which means the existing data can be labelled (column 'berthed': yes/ no). After the model is trained on this labelled data, it can predict outcomes for an unlabelled data set (Uddin et al., 2019).

Supervised machine learning can focus on two types of problems: classification and regression problems. Classification models attempt to predict the categorical class (discrete) for an output whilst regression models predict a numerical value (continuous) (Siguenza-guzman et al., n.d.). Evidently this research is dealing with a classification problem. Classification problems can be solved using many different machine learning algorithms, such as Support Vector Machine, Naive Bays, Decision Trees and Random Forests. Before these algorithms can be used the data quality should be sufficient. The 'Garbage in - garbage out' principle is important to take into account, meaning that poor data quality used will lead to data output being unreliable (Kilkenny & Robinson, 2018).

Processes should be implemented to ensure good quality data. Besides the cleaning and enrichment steps that are performed, feature selection is implemented. Feature selection is the process of selecting variables from the data that are expected to have the largest influence on the prediction ability of the model (Omara et al., 2018). Nine features have been chosen to represent the data, as shown in table E.10. An attempt is made to select features which represent different parts of the data, such as features focusing on the total vessel track (total time, total messages) in comparison to features focusing on the difference between two successive messages (average timestamp interval).

Code name	Feature
$timeinpolygon$	The total time the vessel was present [s]
$meansogpertrack$	The average speed of all data messages [m/s]
$avgtimestampinterval$	The average time between two successive messages [s]
$messagestot$	The total number of messages sent [-]
$avgdistance$	The average distance between two successive messages [m]
$averagespeedsmallest75speed$	The average speed for 75% of the slowest speed messages [m/s]
$stdspeed$	The standard deviation of speed [m/s]
$stddistance$	The standard deviation of distance between two successive messages [m]
$messagefrequency$	The frequency at which messages are sent [/hr]

Table E.10: Features

The data frame from the data base should be merged together with the data from Sea-web. Before the merging the data with all features is reduced to only one single row per vessel track. The data will be merged together based on the MMSI numbers. The merging can take place based on different types of join methods, as visualised in Figure E.2. A natural join will return a new data set only representing the values which both data frames contain. An outer join will keep all values from both data frames. A left join, will keep all values from the 'left' data frame, and add the overlapping values from the 'right' data frame, the opposite occurs for the right join. For this situation it is important to check every vessel track if it is berthed or not, therefore the data is merged as full outer join. If the Sailed Time (from Sea-web) is within 6 hours of the last timestamp from the vessel track, the row is assumed to have berthed.

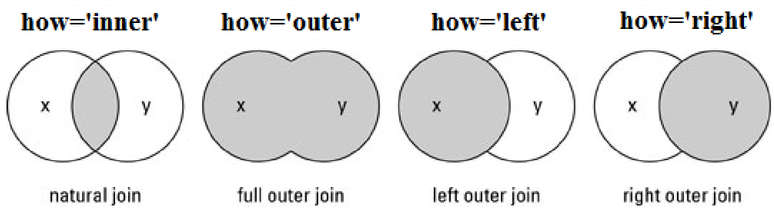


Figure E.2: Methods to join two data sets http://www.datasciencemadesimple.com/join-merge-data-frames-pandas-python/, accessed on 28-4-2020)

Once the data is merged a labelled data frame with features is available for the testing of different machine learning algorithms. The choice of an algorithm is important and different considerations are made to choose which best fits best to the classification problem (Çigşar & Ünal, 2019). Multiple classification algorithms will be tested returning accuracy scores. 16 different terminals will be used to generate a sufficiently large data set. From this large data set the data is split into a training set (80%) and a test set (20%). The data is split as a means to estimate the predicting accuracy, and for the training of the algorithms only the training data set will be used (Dobbin & Simon, 2011). The data is split randomly into the test and training set. In order to correctly compare the algorithms this random split is fixed in order to return the same set of train and test data for each run. Furthermore, a distinction is made between the X and the y of the data set. X represents all the feature columns and y represent the target variable, in this case: berthed or not berthed.

E.3.1. Logistic Regression

Logistic regression is a machine learning algorithm used for classification problems as it predicts the probability of a category for the target variable. The target variable contains data coded as 1 (yes) or 0 (no) as it predicts the probability P(y=1) as a function of all variables X. Since it it takes into account all features (variables) separately it is important to only use correct features because too many (unimportant) features might dilute the outcome (Uddin et al., 2019; Pandis, 2017).

E.3.2. K-Nearest Neighbors (K-NN)

The *K-Nearest Neighbors* (*K-NN*) model works in a way that it finds the k nearest neighbors of a new data point, according to for example the Euclidean distance, the different categories these neighbours are in are counted and the new data point is accordingly placed in the category of which the most neighbors are present. k is an input parameters, default set to 5. In this situation the default parameters, $k = 5$ and the Euclidean distance are used (Hu et al., 2016).

E.3.3. Support Vector Machine (SVM)

The *Support Vector Machine* (*SVM*) aims to find an hyperplane in the data which separates the two classes the best way possible. The distance between the support vectors (points closed to the hyperplane) and the hyperplane should be maximized (Sisodia & Sisodia, 2018; Sisodia et al., 2010). The SVM algorithm uses a set of mathematical functions representing the kernel. The most used type of kernel is the Gaussian *Radial basis function* (*RBF*), for a non-linear problem. In this situation both the linear kernel (similar to logistic regression) and RBF kernel are used.

E.3.4. Naïve Bays

The Naïve Bayes classifier uses the Bayes theorem, by adopting previous knowledge with current knowledge. The Bayes theorem calculates the probabilities given that something else has already occurs, as follows:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \tag{E.1}$$

The classifier predicts the probability that the target feature belongs to a particular category. All attributes contribute equally to the final prediction of the target feature. The classifier assumes all the features are independent, which is obviously not the case in this situation (Saxena, 2017; Sharma, 2020; Sisodia & Sisodia, 2018).

E.3.5. Decision Tree

Decision tree classifier are relatively easy to interpret, it is formed as a tree which follows a flowchart. The classifier uses the nodes and internodes to predict and classify target variables. The root node is the node representing the different features whilst leaf nodes are the target classifications. For every node, the highest information gain is chosen between all the attributes (Iyer, S, & Sumbaly, 2015). Several input parameters are needed in order to train the decision tree classifier. The maximum depth of the tree, set to 3, is based on interpretation reasons. A higher maximum depth will lead to more accurate predictions, therefore the classifier is tested for a depth of 6 as well. The tree does not further split values after this maximum depth has been reached. This maximum is set to prevent over-fitting. Finally, the criterion is based on how the split is made at a node, and how impure these nodes become. Gini is chosen as it is known to give better predictions. The gini value in the decision tree represents how incorrectly labelled the node is, the closer to 0 the more pure the leaf is (Sisodia & Sisodia, 2018; *Gini index vs Entropy*, n.d.; Uddin et al., 2019).

To further optimise the decision tree the most ideal number of layers is found by running the model for multiple max depth values. From Figure E.3 it is noted that a good max depth would be 4 layers, since the increase between 3 and 4 is still quite large, whilst between 4 and 5 layers this increase is relatively less, the curve has flattened. The decision tree classifier returns an accuracy of 93.6% for a max depth of 4.

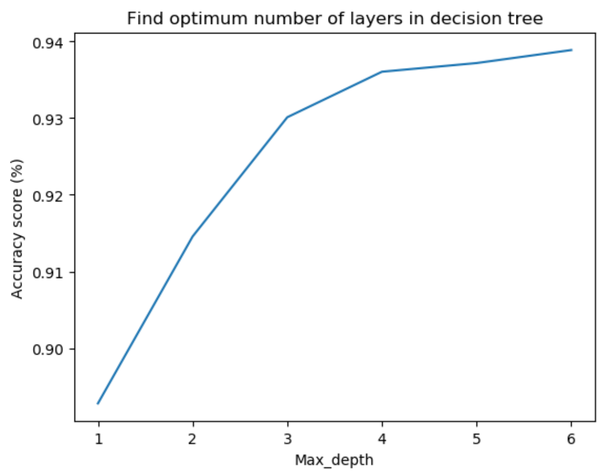


Figure E.3: Optimal max depth for Decision Tree (criterion = gini)

E.3.6. Random Forest

A random forest classifier can be seen as an n number of decision trees together. The classifier trains the various trees on different parts of the training data. The classifier works by picking K (random) number of data points from data set and building one decision tree from that data set. Following, a N number of trees is selected to build and repeat this first step. For an unidentified data point, every N tree will predict the outcome, and the feature is categorized into the class which has the majority of the outputs of the trees (Uddin et al., 2019).

Again the maximum depth has to be specified for every tree. Besides this depth, the number of trees n $estimators$ has to be defined. In this situation the maximum depth is run for 3 and 6, and the number of estimators is set to 100. The criterion for node splitting is again set to Gini.

E.3.7. XGBoost

Gradient Boosting is very similar to a random forest, as it grows multiple decision trees. The only difference is that that Gradient Boosting actively 'boosts' trees, based on outputs of already grown trees. XGBoost is one of many algorithms which uses this gradient boosting (Gursky, 2020). Again the number of estimators and max depth are required inputs, both set to similar inputs as for the Random Forest Classifier.

E.3.8. Overview of classification algorithms accuracy

For all classifiers the accuracy is determined, by determining how well the predicted output for the target variable matches the real output. In the data a large part will contain vessel tracks which are obvious not to have berthed. Considering current port processes, as mentioned in Subchapter 2.1.1, it is expected that berthing vessels need at least 30 minutes as service time, and will send at least 5 messages during their stay close to & attached to the terminal. In order to optimise the models the following will be removed from the data:

- If the vessel track has less than 5 messages in total
- If the vessel track total duration is less than 30 minutes

After merging of the raw data with Sea-web data and removing all rows with a track label of zero the data contains 55,752 rows. After this check has been done 20,349 rows have been removed (36.45%), leaving a data of 35,403 rows. These adjusted accuracies are shown in the second column in table E.11.

However, as mentioned in the first approach, it is important to note that the data set contains 93.87% vessel tracks that will not berth. As an example, the confusion matrix is given in Figure E.4 which represents the output for the Decision Tree classifier, for a max-depth of 3 with the adjusted (smaller) data set containing 35,403 rows. The confusion matrix represents how well the test section of the data set (20%) performs when predicted using the classifier. In total the test set represents 7081 vessel tracks. The classifier correctly predicts 326 vessels (true positives). 83 vessel tracks (1.17%) were 'false positives', which means that these tracks were classified as berthed, whilst they actually did not berth. 412 vessel tracks (5.8%) were 'false negatives', these vessels actually did berth, but they were interpreted as 'not berthing' by the classifier. Nonetheless, the percentage of correctly predicted berths (326) compared to the total number of actual berths (412+326) is only 44.173%. The percentage of wrong berth predictions (83) compared to the total number of actual berths (412+326), is 11.247%.

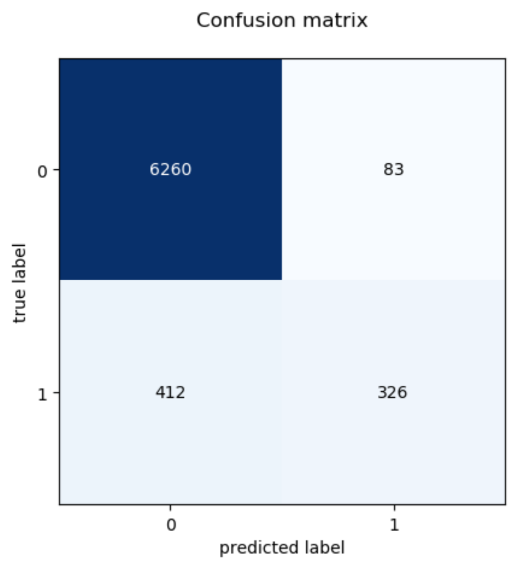


Figure E.4: Confusion matrix for Decision Tree (max depth = 3, criterion = gini)

For all classifiers the percentages of correct and the percentage of wrong berth predictions are calculated, returning the following table:

Classifier	Adjusted accuracy [%]	Corrected predicted berths [%]	False predicted berths [%]
Logistic Regression	91.244	26.558	10.569
K-NN	93.560	60.163	21.951
SVM (kernel = Linear | Gaussian RBF)	90.665 | 92.430	18.564 | 40.379	8.130 | 13.008
Naïve Bayes	59.992	94.038	377.913
Decision Tree (max depth = 3 | 6)	93.009 | 93.885	44.173 | 53.252	11.247 | 11.924
Random Forest (max depth = 3 | 6)	92.953 | 94.139	37.398 | 55.014	5.014 | 11.247
XGBoost (max depth = 3 | 6)	92.953 | 94.139	70.596 | 74.119	11.382 |10.434

Table E.11: Predicting qualities of second approach for all classifiers (for adjusted data set)

The Naive Bayes Classifier performs very well based on the correctly predicted berthed vessels. However, the number of false predicted values is enormous, more than 370% more than the total amount of berthed vessels. Furthermore, the XGBoost Classifier performs the best, based on the number of correctly predicted berthed vessels, together with a low amount of falsely predicted berthed vessels. However, 74% correctly predicted berthed vessels is not sufficient since this is a vital step for reaching the research objectives. For this approach, as well as for the first approach, none of the classifiers tested give an accurate prediction.

E.4. Approach IV: Machine learning algorithm - filtered data

The second method is approached again, now using more information about the vessel tracks. As mentioned, the data is filtered based on category type and four extra parameters are appended to the data set: *TEU* capacity, *LOA*, *DWT* and the standard deviation of the location. These first three parameters are vessel characteristics and found in the private AIS data base from RHDHV. The last new parameter is defined as the standard deviation of every AIS message compared to the same vessel tracks center location.

E.4.1. Classifier choice - Explainable AI

In the first approach to this method multiple classifiers were tested from which can be concluded that the XGBoost classifier will predict the vessel berthing or not, with the best accuracy (74.12% when regarding correctly predicted berths). However, a consideration must be made by balancing the interpretability versus the accuracy. A lot of deep learning models, such as XGBoost, perform best when considering the accuracy, but are complex to interpret. This trade-off is discussed thoroughly on the internet under the name *Explainable Artificial Intelligence* (*XAI*) (Itani et al., 2019; Rane, 2018; Schmelzer, 2019). This emerging field aims to improve the ability to understand the models and the choices the models make. A more visible, simpler model, such as a decision tree, is transparent and can easily be understood, but therefore also under-performs slightly in terms of accuracy. Therefore, in this second approach both the XGBoost and Decision Tree models are trained and tested. The classifiers are trained based on the same locations as were trained in the last approach.

With the three new parameters and the data being filtered based on category, the XGBoost and Decision Tree results are summarised in table E.12. It is clear that again the XGBoost performs the best compared to the Decision Tree. The same conclusion is made when training the model on different terminal types separately, as results show in tables E.13, E.14, E.15.

Classifier	Max depth	n estimators	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
XGBoost	4	100	94.264	89.941	10.848
XGBoost	6	100	93.994	88.955	10.848
XGBoost	4	250	94.048	88.560	10.256
XGBoost	4	50	93.994	88.955	10.848
Decision Tree	4	-	92.857	87.968	14.004
Decision Tree	6	-	93.074	85.207	10.454

Table E.12: Method 2: Second approach, XGBoost versus Decision Tree, combined terminals

Classifier	Max depth	n estimators	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
XGBoost	4	100	94.057	92.761	11.260
XGBoost	6	100	93.885	92.493	11.528
Decision Tree	4	-	92.679	90.885	13.673
Decision Tree	6	-	93.454	92.761	13.137

Table E.13: All container terminals

Classifier	Max depth	n estimators	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
XGBoost	4	100	95.911	83.784	16.216
XGBoost	6	100	96.082	86.486	17.568
Decision Tree	4	-	94.208	77.027	22.973
Decision Tree	6	-	93.867	75.676	24.324

Table E.14: All dry bulk terminals

Classifier	Max depth	n estimators	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
XGBoost	4	100	92.079	87.097	12.903
XGBoost	6	100	91.089	80.645	9.677
Decision Tree	4	-	67.742	67.742	16.129
Decision Tree	6	-	88.119	70.968	9.677

Table E.15: All liquid bulk terminals

E.4.2. Feature optimization

Now that the classifier is selected an important step in training the model to perform better is feature optimization, also known as parameter tuning (Patrous, 2018). In order to improve the model two different parameter optimization approaches are performed: boosting parameters and tree-specific parameters.

Boosting parameters improve the boosting operation of the classifier, and is improved by adjusting the learning rate and number of estimators. The learning rate is defined as the relative impact every tree has on the final prediction/result. As mentioned before, XGBoost updates after every tree is generated based on an initial estimate and thus this learning parameter affects the degree of change occurring between the tree generating.

Often a lower learning rate is preferred since this improves the model robustness, however a lower rate needs a larger number of trees to find all relations (Jain, 2016). The classifier is trained based on different learning rates from which results can be found in table E.17. From these results an optimal learning rate of 0.2 is chosen, however the change between the different rates was barely significant.

As mentioned, the number of estimators is defined as the number of trees generated by the model. It must be taken into account that increasing the number of estimators should be carefully done since over-fitting could occur (Jain, 2016). From multiple runs, results in table E.16, 100 estimators have been selected as most optimal.

n estimators	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
100	94.318	89.349	10.059
50	93.885	88.166	10.454
75	94.156	89.349	10.651
125	94.210	89.349	10.454

Table E.16: Number of estimators optimization

Learning rate	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
Default (0.3)	94.264	89.941	10.848
0.1	93.723	88.757	11.637
0.2	94.318	89.349	10.059

Table E.17: Learning rate optimization

Next, the tree-specific parameters are optimized by tuning the maximum depth. The max depth is defined as the maximum depth of a tree, earlier an optimal max depth of 4 was chosen (results in table E.12).

E.4.3. Feature importance

Four extra features have been added to the model, compared to the first approach of this method. The leading method of interpreting tree-based classifiers is claimed to be *SHapley Additive exPlanations* (*SHAP*) (Slundberg, n.d.). SHAP is a method which explains the classifiers output based on a game theoretic approach (*SHAP- Explainers*, 2018). A SHAP value is defined as the (relative) amount each feature contributes to the final target variable. Once plotted it can be seen as an extensive version of a feature importance plot since the SHAP plot also visualises the positive or negative influence on the feature outcome (Dr. Dataman, 2019). The SHAP plot for the XGBoost trained classifier is visualised in figure E.5.

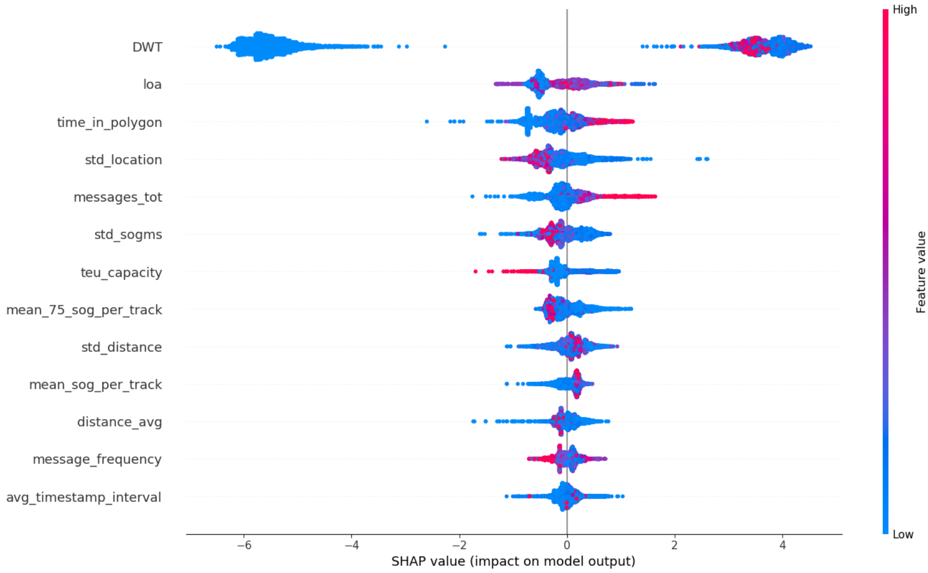


Figure E.5: SHAP importance plot

The features are ranked by their importance, ascending from highest importance (*DWT*) to lowest importance (average timestamp interval) on the left vertical axis. The SHAP value on the horizontal axis represents the impact every feature has on the outcome, a negative value shows that a certain feature leads to a lower prediction (Dr. Dataman, 2019). The target variable is classified as berthed vessel track = 1 and not berthed vessel tracks = 0. Therefore a lower prediction (negative SHAP value) is equal to a higher possibility of the feature leading to a not berthed vessel track. On the other hand, a higher/positive SHAP value will lead to a higher possibility of the vessel track berthing at the terminal. Finally, the color on the right vertical axis represents if that certain feature has a high or low value. A few feature importances are clarified.

The first thing that tracks attention is the *DWT*. A low value of DWT (blue color) leads to a high possibility of the target variable being 0, thus displaying a not berthed vessel track. This could be explained due to smaller vessel sizes not berthing at the terminal, however a large number of DWT values are left empty thus a confident conclusion about the DWT influence can not be made. Furthermore, a longer duration in the terminal polygon (higher value of time in polygon) leads to a larger chance of vessel tracks berthing, which makes sense as a berthed vessel track will stay relatively longer in the terminal than a vessel just sailing by. Once the standard deviation with regards to the location center is low (blue) there will be higher probability of vessel tracks berthing (positive SHAP value), which likewise makes sense since the berthed vessel tracks will stay at one location for a longer time. Finally, a higher total amount of messages (pink) will return a larger chance of vessels berthing (positive SHAP value), which can be based on the same explanation as for the time in polygon.

E.4.4. Wrongfully predicted vessel tracks

To further improve the model the wrongfully predicted values are investigated. During this investigation a problem surfaced in which one actual vessel track in the terminal polygon is defined by the vessel track labelling as two separate vessel tracks. In sub-chapter 4.1.1 this complication was explained and the solution demonstrated. Applying this new method of vessel track labelling leads to a large increase of model accuracy, as shown in table E.18.

Vessel track labelling	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
Old	94.318	89.349	10.059
New	98.202	98.081	5.970

Table E.18: New method of defining terminal vessel tracks

This new method reduces the total number of vessel tracks, for the chosen 12 terminals, to 10,290 tracks and thus the test data size (20%) equals 2058 tracks. The new confusion matrix is shown in figure E.6, from which it can be seen that 28 vessel tracks were wrongfully predicted as berthed and 9 vessel tracks were wrongfully predicted as not berthed.

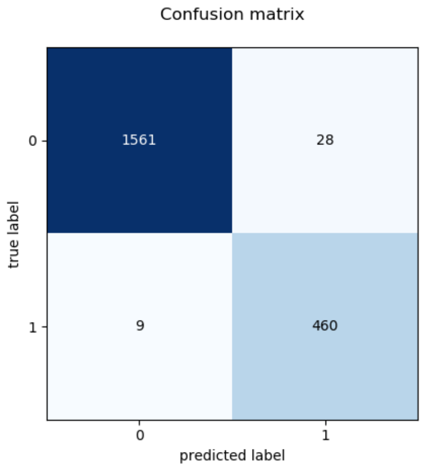


Figure E.6: Confusion matrix for new method of vessel track labelling

In order to further verify the model the wrongfully predicted vessel tracks are analyzed. For all separate 12 locations these vessel tracks are visualised, as shown in the following subchapter.

Visualisation wrongfully predicted vessel tracks

For all 12 locations the wrongfully predicted vessel tracks are visualised, for both the false negatives (FN) and false positives (FP). False negatives represent the situation where the vessel track was classified as not berthed, whilst it actually did berth. On the other hand the false positives represent vessel tracks which were classified as berthed while they actually did not berth. The number of false positives or false negatives per terminal location are given in each figure title as the number in [].

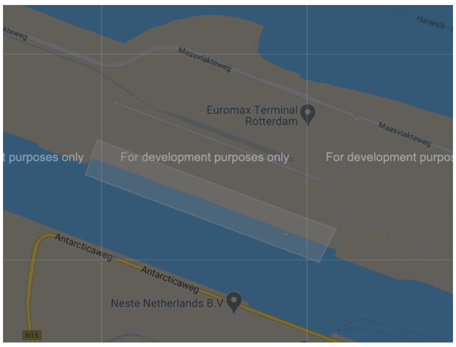


Figure E.7: FN: CT Rotterdam EuroMax [1]

Figure E.8: FP: CT Rotterdam EuroMax [15]

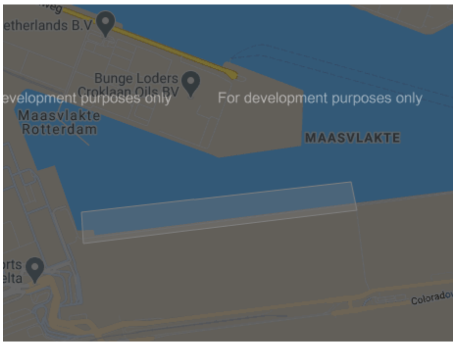


Figure E.9: FN: CT Rotterdam APM Terminals - Main Quay [0]

Figure E.10: FP: CT Rotterdam APM Terminals - Main Quay [12]

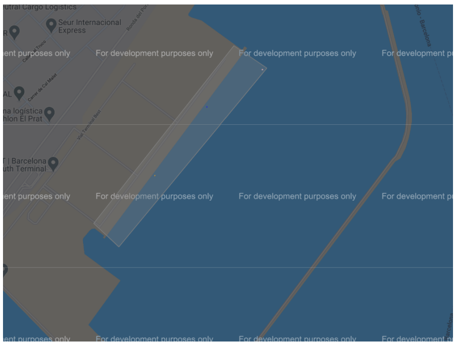


Figure E.11: FN: CT Barcelona Europe South Terminal [3]

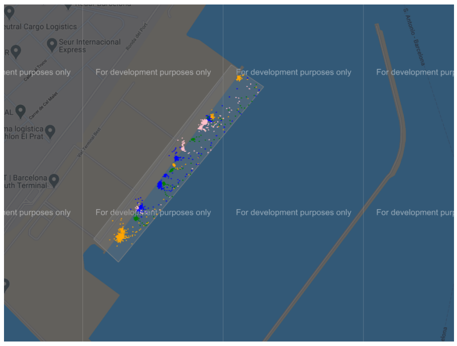


Figure E.12: FP: CT Barcelona Europe South Terminal [21]

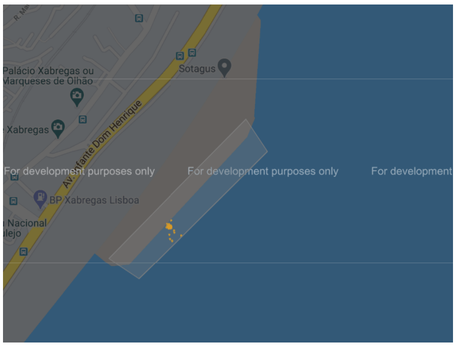


Figure E.13: FN: CT Lisbon Santa Apolonia [1]

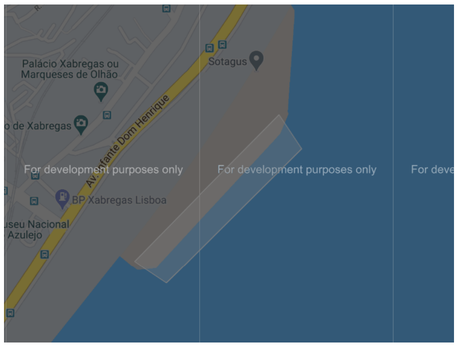


Figure E.14: FP: CT Lisbon Santa Apolonia [0]

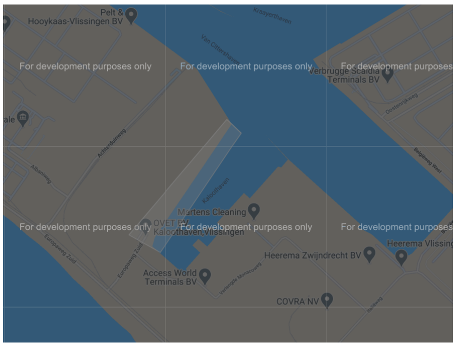


Figure E.15: FN: DBT Vlissingen OVET Terminal [0]

Figure E.16: FP: DBT Vlissingen OVET Terminal [1]

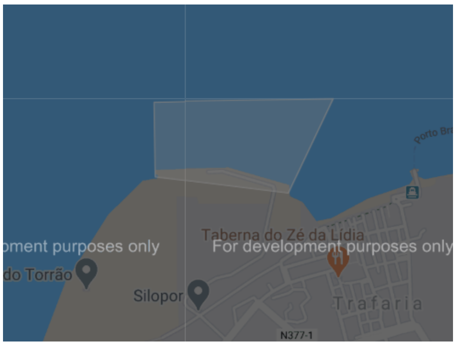


Figure E.17: FN: DBT Lisbon [0]

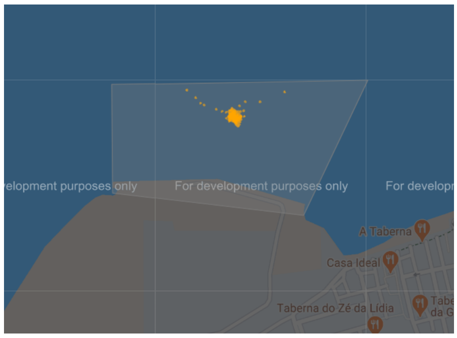


Figure E.18: FP: DBT Lisbon [1]

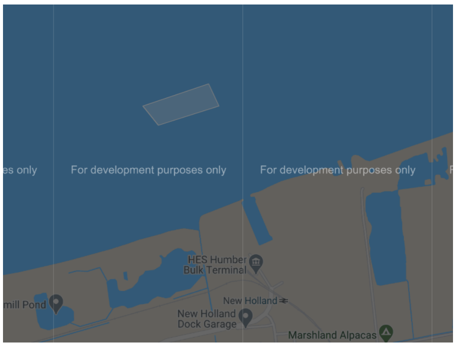


Figure E.19: FN: DBT New Holland [0]

Figure E.20: FP: DBT New Holland [4]

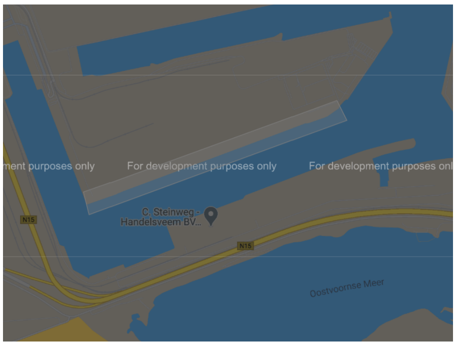


Figure E.21: FN: DBT Rotterdam EMO Terminal [0]

Figure E.22: FP: DBT Rotterdam EMO Terminal [5]

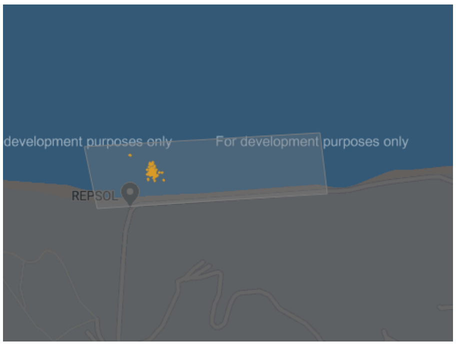


Figure E.23: FN: LBT Lisbon REPSOL [1]

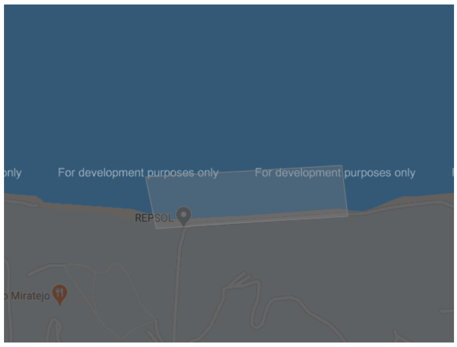


Figure E.24: FP: LBT Lisbon REPSOL [0]

Figure E.25: FN: LBT Vlissingen [4]

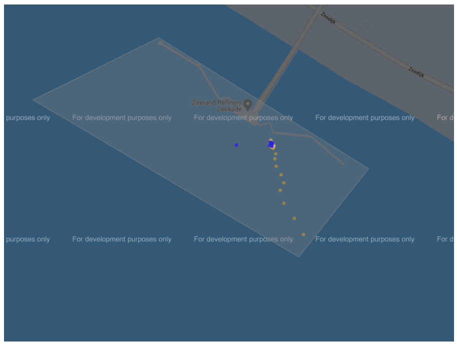


Figure E.26: FP: LBT Vlissingen [3]

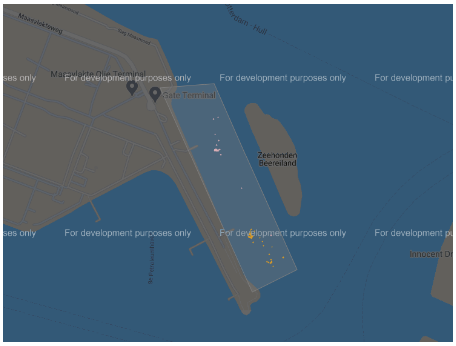


Figure E.27: FN: LBT Rotterdam GATE [2]

Figure E.28: FP: LBT Rotterdam GATE [0]

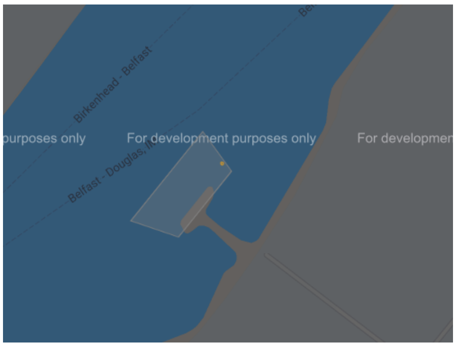


Figure E.29: FN: LBT Belfast Puma Energy East Wharf [1]

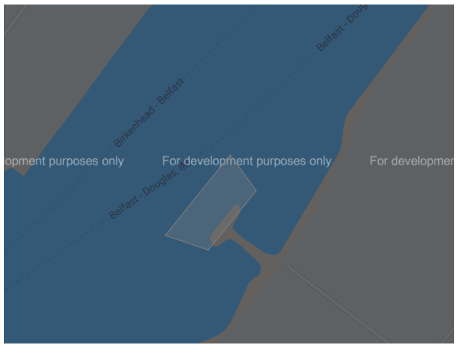


Figure E.30: FP: LBT Belfast Puma Energy East Wharf [0]

From these visualisations it is clear that the false negatives, vessels wrongfully predicted as not berthed, are often one of the following:

- Vessel tracks containing only very little AIS messages, only a few random messages/points on the figures
- Relatively short vessel tracks with a small amount of messages
- Tracks with no clear 'center' of the vessel track, as if it visually did not lay still at one point in time

It is logical that these false negative vessel tracks were classified as not berthed, as they mostly show a behavioural pattern expected from vessels not berthing at the terminal. For the false positives, the vessel tracks wrongfully predicted as berthed, the following types were often found:

- Vessel tracks with a lot of messages at one location
- Vessel track with a small amount of messages in total

The first type of vessel tracks is obvious to be classified as berthed, the second a little less. Besides visualising these vessel tracks it must be noted that the vessel tracks are compared to what Sea-web defines as berthed vessels, thus Sea-web is seen as 'true' in this research. Wrongfully predicted vessel tracks could be due to noise in AIS data or possible errors in Sea-web data. No unexpected results were found thus the classifier is now working sufficiently for the purpose of this research.

E.4.5. Pre-classifier step: filtering obvious vessels

Now that the classifier is performing sufficiently one last step is added to improve the model. In the first approach of this method this step was already performed by removing vessel tracks with less than 5 messages and vessel tracks that stay less than 30 minutes in the terminal polygon. In this step the focus lies on the time in polygon only, since these two filters overlap almost fully. Applying a filter step to the data, for every vessel track less than 30 minutes, before training the model returns the following new model accuracy:

Filter	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
No	98.202	98.081	5.970
Yes	97.056	99.050	7.126

Table E.19: Filter step added

With this filter step roughly 30% of the vessel tracks is removed. Only 0.0418% of the vessel tracks is removed wrongfully, being a vessel track which actually berthed but was removed in this filter step. This will most likely be due to errors in AIS data or Sea-web validation data, since it is highly unexpected that a vessel with berth at a terminal in less than 30 minutes (including (un)mooring and (un)loading). For both situations, with and without the filter step, there are advantages and disadvantages. The choice is made to apply the filter step since this improves the number of correctly predicted berths. However this filter step also slightly increases the number of falsely predicted berths. This does not weigh out the fact that the filtering step removes a large amount of the input data (30%) while still performing very well.

E.4.6. Comparison between first and second approach of this method

Since a new method was defined and implemented for determining the vessel track, the first approach to this method (using basic data) should also be retested. Furthermore, the obvious vessel removal should be the same as for the first approach. Thus, the same 12 locations are rerun and tested, using only the basic data no pre-filtering of the vessels based on certain vessel types. The results are presented in table E.20, from which the obvious conclusion can be made that the pre-filtering of the data, including extra parameters (the LOA, TEU capacity and DWT), returns a more accurate prediction model.

Approach	Accuracy [%]	Corrected berths predicted [%]	False predicted berths [%]
1. Basic data	92.593	66.154	13.076
2. Pre-filtered data	97.056	99.050	7.126

Table E.20: Results: basic versus pre-filtered data.

F

Appendix F: Results for goodness-of-fit tests for the service time distribution

F.1. Container terminals

F.1.1. Container terminals: total vessel mix

Rotterdam APM-2 Terminal

Results are visualised and demonstrated in chapter 5.1.1.

Rotterdam APM Terminal

For the container terminal Rotterdam APM all distributions are fitted and visualised, as shown in figure F.2. In total, 2108 vessels berth at the APM terminal.

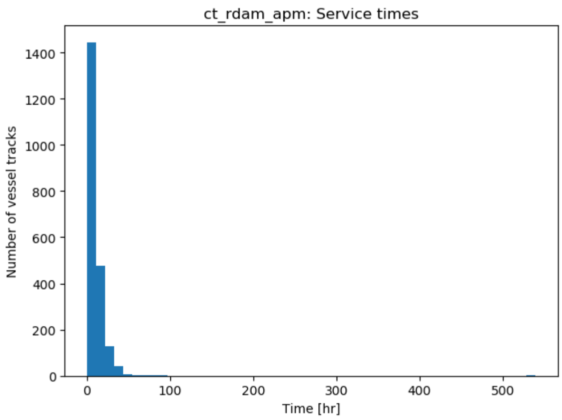


Figure F.1: Rotterdam APM Terminal service times (histogram)

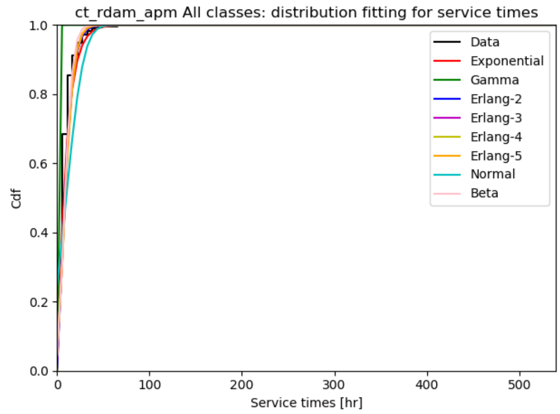


Figure F.2: Service time distribution: Rotterdam APM Terminal

Straight away it is clear that the service time distribution includes outliers. Investigation into the larger service times (> 100 hours) leads to the conclusion that there is only one vessel track with an extreme service time. The container vessel LEXA MAERSK stays at the APM terminal between May 18th until June 9th 2020, staying a total of 539 hours. Two actions are taken in order to inspect this outlier. First, the Sea-web data base is used to see if a different data source also registered this unusually long stay. The Sea-web data base also revealed this vessel staying for such a long time at the terminal. A final check consists of visualising all AIS messages sent between the entry and exit time of the terminal. This supports the fact that the vessel did not leave the terminal polygon between these two dates (figure F.3).

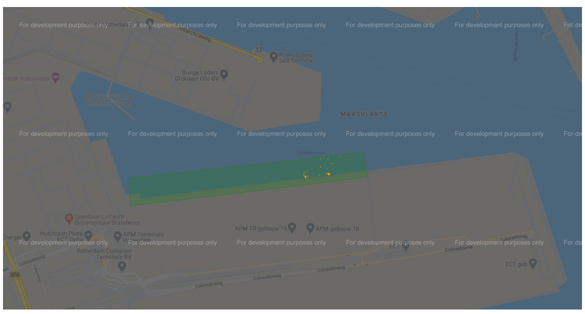


Figure F.3: Service time outlier: visualisation AIS messages

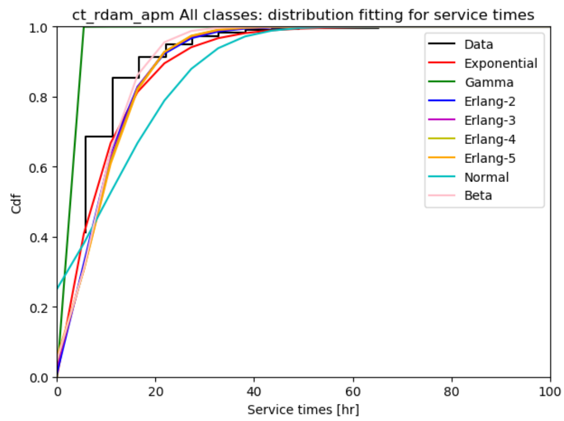


Figure F.4: Service time distribution: Rotterdam APM Terminal (zoom in)

In order to visually inspect the fitted distributions a new plot is given in figure F.4, where the x axis has been manually adjusted. Based on the K-S test hypothesis, none of the distributions fit the data (table F.1 in appendix F). Visually, the best fit would most likely be the Erlang-5 distribution. However, none of them fully represent the data.

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p [%]	Lim
Exponential	0.50	9.44	1.00		0.04	0.01	No
Gamma	0.50	3.44	0.00		0.98	0.00	No
Erlang-2	-0.76	5.35	2.00		0.07	0.00	No
Erlang-3	-2.87	4.27	3.00		0.08	0.00	No
Erlang-4	-4.89	3.71	4.00		0.09	0.00	No
Erlang-5	-6.77	3.34	5.00		0.09	0.00	No
Normal	9.95	14.77			0.26	0.00	No
Beta	-7.45	9.76E+13	6.53	3.81E+13	0.08	0.00	No

Table F.1: Service time distribution fitting for container terminal: Rotterdam APM

Rotterdam Euromax Terminal

The third terminal accessed is the longer Euromax terminal in Rotterdam. The histogram of the service time distribution, as well as the CDF with fitted distributions, are shown in figures F.5 and F.6. The Euromax terminal receives a total of 3001 vessels. Again, the K-S test is performed on all fits, however none fit (results in table F.2 in appendix F). This corresponds with the visual interpretation of the CDF's.

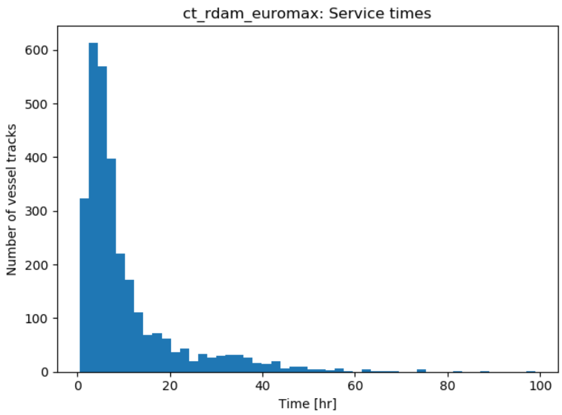


Figure F.5: Rotterdam Euromax Terminal service times (histogram)

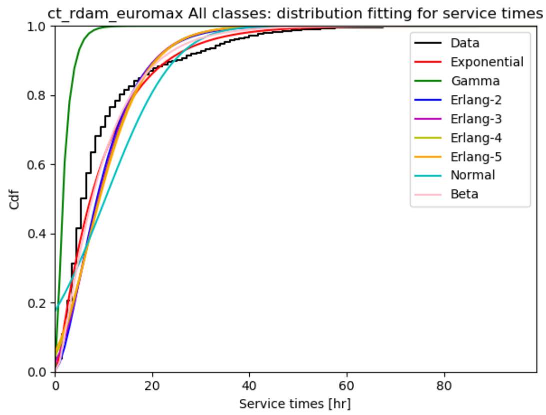


Figure F.6: Rotterdam Euromax Terminal with fitted distributions

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p [%]	Lim
Gamma	0.50	1.77	0.92		0.63	0.00	No
Erlang-2	-0.35	5.35	2.00		0.15	0.00	No
Erlang-3	-2.72	4.3					
Erlang-4	-5.03	3.85	4.00		0.18	0.00	No
Erlang-5	-7.17	3.51	5.00		0.18	0.00	No
Normal	10.36	11.10			0.22	0.00	No
Beta	0.50	4.00E+09	1.23	5.16E+08	0.10	0.00	No

Table F.2: Service time distribution fitting for container terminal: Rotterdam Euromax

France Le Havre Terminal

Finally, a much shorter terminal (800 meters long) is analysed. The service time distribution of the Le Havre Atlantic Terminal (figure F.7) clearly is less skewed towards the left in comparison to the other three analysed terminals. Whilst the other terminals contained visually one peak, this terminal might have a more diverse service time distribution, due to its two peaks. However, it should be noted that this terminal also facilitates much less arrivals (383 arrivals) in the same time span.

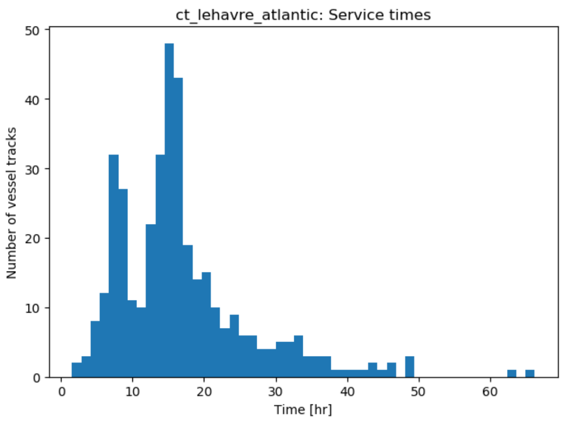


Figure F.7: Le Havre Atlantic service times (histogram)

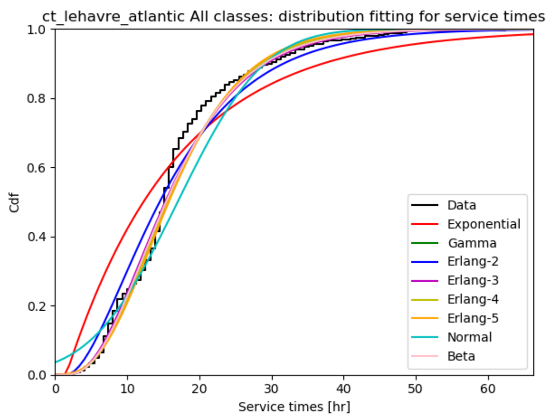


Figure F.8: Le Havre Atlantic with fitted distributions

Despite the first remarks, again no correct fit has been found visually, as well as using the K-S goodness-of-fit test (results in table F.3 in appendix F).

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p [%]	Lim
Exponential	1.55	15.45	1.00		0.22	0.00	No
Gamma	0.85	4.92	3.28		0.08	0.01	No
Erlang-2	1.48	7.76	2.00		0.12	0.00	No
Erlang-3	1.12	5.29	3.00		0.08	0.02	No
Erlang-4	-0.15	4.29	4.00		0.09	0.01	No
Erlang-5	-1.77	3.75	5.00		0.09	0.00	No
Normal	17.01	9.43			0.16	0.00	No
Beta	0.86	1.71E+13	3.27	3.46E+12	0.08	0.01	No

Table F.3: Service time distribution fitting for container terminal: Le Havre Atlantic

F.1.2. Container terminals: specific vessel classes

Arrivals per vessel class

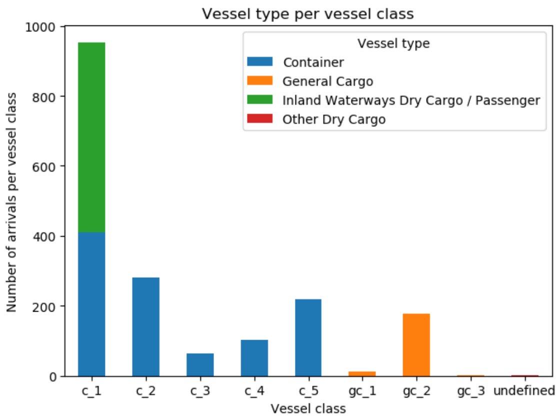


Figure F.9: Arrivals per vessel class: Rotterdam APM-2 Terminal

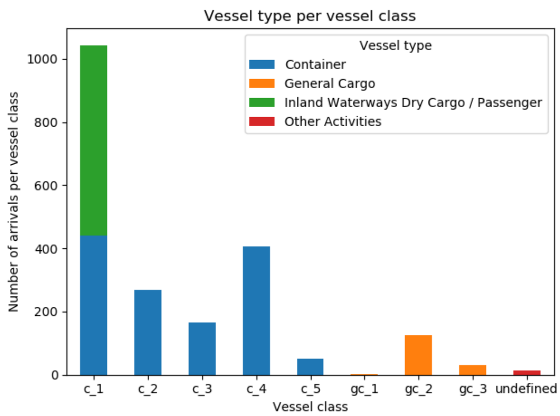


Figure F.10: Arrivals per vessel class: Rotterdam APM Terminal

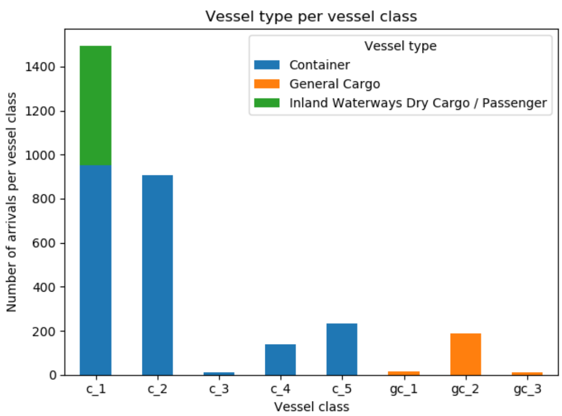


Figure F.11: Arrivals per vessel class: Rotterdam Euromax Terminal

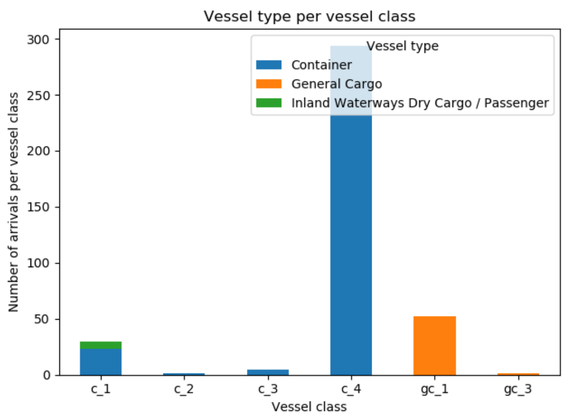


Figure F.12: Arrivals per vessel class: Le Havre Atlantic Terminal

Container Class 1: Small feeders

The numerical results of the distribution fitting can be found in appendix F in tables F.4, F.5, F.6 and F.7 for the Rotterdam APM-2, Rotterdam APM, Rotterdam Euromax and Le Havre Atlantic Terminals respectively. The visual results are presented in figures F.13, F.13, F.15 and F.16.

Figure F.13 represents the Rotterdam APM-2 terminal. From the goodness-of-fit test the **Beta** distribution is found as the only possible fit for the distribution with a relatively p value. This corresponds with the visual interpretation, however visually other distributions also seem to fit. Visually interpreting this plot is difficult and therefore the chosen distribution is based mainly on the K-S test (table F.4). Figure F.13 represents the APM terminal. Based on the K-S test the **Beta** distribution is seen as the only possible fit. Visually again a lot of distributions might fit. The Beta is chosen based on the p limit of the K-S test. However, when analysing the D value the Exponential distribution might also be a possible fit (table F.5).

Figure F.15 represents the Euromax terminal. Visually a lot of distributions fit. Based on the K-S test the **Erlang-3, Erlang-4 and Beta** distribution all fit on the data (table F.6). Finally, figure F.16 visualises the Le Havre Atlantic Terminal, visually can be fit by the Gamma, the Erlang-2, Erlang-3, Erlang-4 and Erlang-5 all could fit the data. Numerical results of the K-S test (table F.7) suggest all distributions could fit except the beta distribution. However, clearly less data is available based on the smoothness of the data visualisation compared to the other three terminals. Only 30 vessel tracks arrive, thus this result is unreliable.

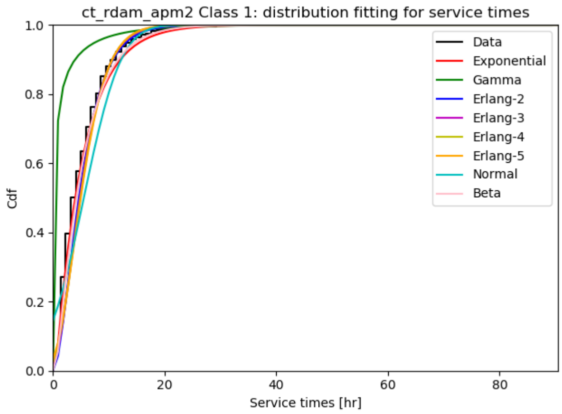


Figure F.13: Service time CDF Class 1 Rotterdam APM-2 Terminal

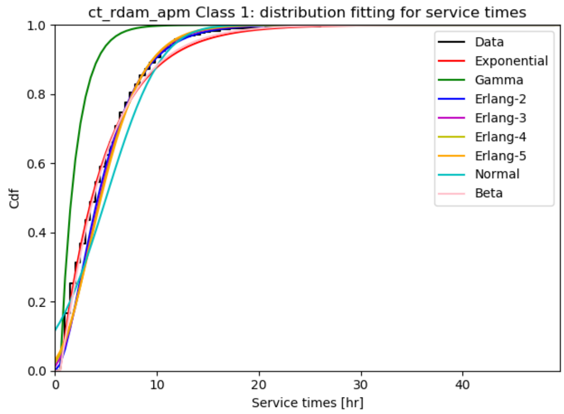


Figure F.14: Service time CDF Class 1 Rotterdam APM Terminal

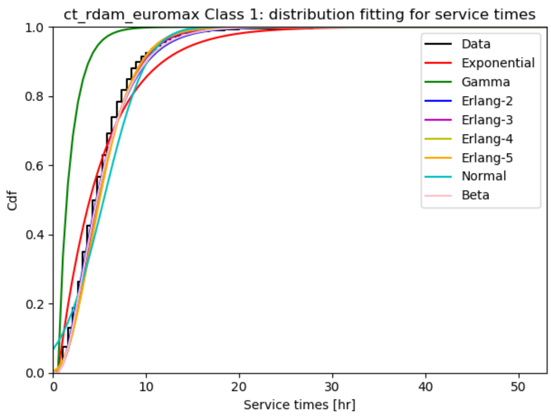


Figure F.15: Service time CDF Class 1 Rotterdam Euromax Terminal

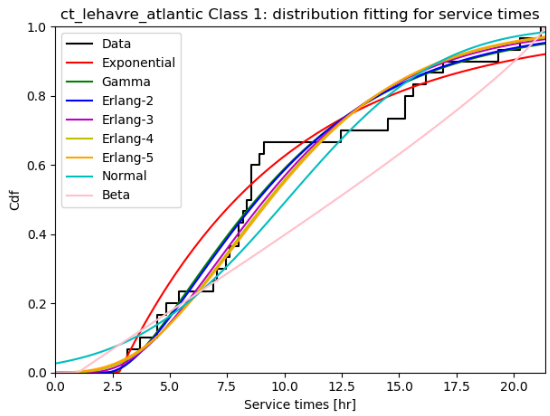


Figure F.16: Service time CDF Class 1 Le Havre Atlantic Terminal

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	0.50	5.04	1.00		0.05	0.02	No
Gamma	0.50	10.72	0.12		0.69	0.00	No
Erlang-2	0.01	2.77	2.00		0.09	0.00	No
Erlang-3	-1.03	2.19	3.00		0.10	0.00	No
Erlang-4	-2.03	1.89	4.00		0.11	0.00	No
Erlang-5	-2.95	1.70	5.00		0.11	0.00	No
Normal	5.54	5.22			0.17	0.00	No
Beta	0.50	2.71E+09	1.18	6.31E+08	0.02	0.72	Yes

Table F.4: Service time distribution fitting for Class 1 container terminal: Rotterdam APM-2

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	0.50	4.52	1.00		0.05	0.01	No
Gamma	0.50	1.60	1.00		0.42	0.00	No
Erlang-2	-0.02	2.52	2.00		0.07	0.00	No
Erlang-3	-0.92	1.98	3.00		0.08	0.00	No
Erlang-4	-1.79	1.70	4.00		0.08	0.00	No
Erlang-5	-2.59	1.52	5.00		0.08	0.00	No
Normal	5.02	4.21			0.14	0.00	No
Beta	0.50	259.72	1.09	61.39	0.03	0.17	Yes

Table F.5: Service time distribution fitting for Class 1 container terminal: Rotterdam APM

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	0.55	4.90	1.00		0.16	0.00	No
Gamma	0.55	1.57	0.89		0.59	0.00	No
Erlang-2	0.48	2.49	2.00		0.04	0.02	No
Erlang-3	-0.07	1.84	3.00		0.02	0.52	Yes
Erlang-4	-0.74	1.55	4.00		0.03	0.13	Yes
Erlang-5	-1.38	1.37	5.00		0.04	0.04	No
Normal	5.45	3.65			0.10	0.00	No
Beta	0.40	8.28E+11	2.19	3.58E+11	0.03	0.19	Yes

Table F.6: Service time distribution fitting for Class 1 container terminal: Rotterdam Euromax

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	2.78	7.34	1.00		0.20	0.15	Yes
Gamma	2.31	4.22	1.85		0.14	0.60	Yes
Erlang-2	2.17	3.97	2.00		0.14	0.54	Yes
Erlang-3	1.09	3.01	3.00		0.16	0.35	Yes
Erlang-4	-0.03	2.54	4.00		0.18	0.27	Yes
Erlang-5	-1.09	2.24	5.00		0.19	0.23	Yes
Normal	10.12	5.22			0.24	0.05	Yes
Beta	1.00	20.38	0.97	0.84	0.31	0.01	No

Table F.7: Service time distribution fitting for Class 1 container terminal: Le Havre Atlantic

Container Class 2: Regional feeders

Next, the regional feeders are investigated. Again the Le Havre Terminal has too little data (only 1 vessel track) in order to make robust conclusions. First, the APM-2 terminal is analysed. Based on the numerical results of the goodness-of-fit test only the Erlang-5 distribution fits the data (table F.8). However, when focusing on the D values also the Erlang-3, Erlang-4 and Beta might fit. Visually the fit isn't perfect and thus taking all the above into consideration the **Erlang-5** distribution is chosen as the best fit for this data set. The APM terminal for the second class contains a lot of possible fits: the Gamma, Erlang-3, Erlang-4, Erlang-5 and Beta distribution(table F.9). Based on the D statistic of the K-S test the Gamma, Erlang-4, Erlang-5 and Beta all fit the best. Taken the visualisation into account the **Gamma, Erlang-4 of Erlang-5** seem to fit best.

The Rotterdam Euromax terminal performs best for the **Erlang-2 or Beta** distribution (table F.10). These two distributions perform best when taking the K-S test limit, statistic and visualisation into account.

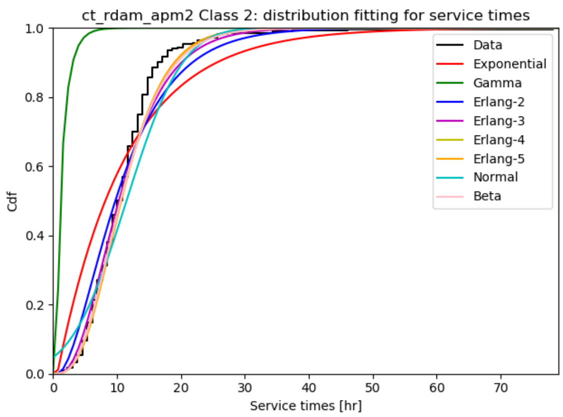


Figure F.17: Service time CDF Class 2 Rotterdam APM-2 Terminal

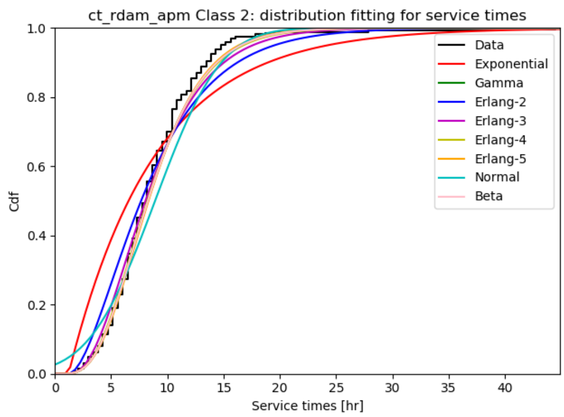


Figure F.18: Service time CDF Class 2 Rotterdam APM Terminal

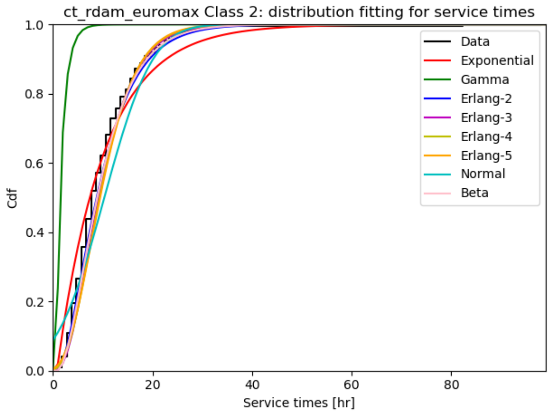


Figure F.19: Service time CDF Class 2 Rotterdam Euromax Terminal

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	0.67	10.83	1.00		0.27	0.00	No
Gamma	0.67	1.55	0.60		0.92	0.00	No
Erlang-2	0.61	5.45	2.00		0.13	0.00	No
Erlang-3	0.41	3.70	3.00		0.09	0.01	No
Erlang-4	-0.13	2.91	4.00		0.08	0.04	No
Erlang-5	-0.96	2.49	5.00		0.08	0.07	Yes
Normal	11.50	6.86			0.14	0.00	No
Beta	-0.17	6.88E+09	4.02	2.36E+09	0.08	0.04	No

Table F.8: Service time distribution fitting for Class 2 container terminal: Rotterdam APM-2

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	1.21	7.69	1.00		0.26	0.00	No
Gamma	0.32	1.95	4.40		0.05	0.52	Yes
Erlang-2	1.16	3.87	2.00		0.14	0.00	No
Erlang-3	1.00	2.64	3.00		0.08	0.07	Yes
Erlang-4	0.56	2.09	4.00		0.05	0.50	Yes
Erlang-5	-0.07	1.79	5.00		0.05	0.53	Yes
Normal	8.90	4.60			0.11	0.00	No
Beta	0.33	4.16E+12	4.38	2.12E+12	0.05	0.52	Yes

Table F.9: Service time distribution fitting for Class 2 container terminal: Rotterdam APM

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	0.81	9.52	1.00		0.17	0.00	No
Gamma	0.81	1.50	0.71		0.82	0.00	No
Erlang-2	0.78	4.78	2.00		0.03	0.25	Yes
Erlang-3	-0.04	3.46	3.00		0.06	0.00	No
Erlang-4	-1.41	2.94	4.00		0.08	0.00	No
Erlang-5	-2.75	2.62	5.00		0.08	0.00	No
Normal	10.34	7.63			0.13	0.00	No
Beta	0.74	6.49E+12	2.21	1.50E+12	0.04	0.11	Yes

Table F.10: Service time distribution fitting for Class 2 container terminal: Rotterdam Euromax

Container Class 3: Feedermax & Panamax

For the third vessel class only the Rotterdam APM-2 Terminal and Rotterdam APM terminals are reliable data sources. The Rotterdam Euromax (11 vessel tracks) and the Le Havre Terminal (5 vessel tracks) both have too little data to return reliable conclusions.

The results for Rotterdam APM-2 terminal are presented in table F.11 and report that all distributions are possible, except the Exponential distribution. From figure F.20 the best fit is virtually chosen to be either the **Beta, Erlang-3, Erlang-4 or Gamma** distribution. For the Rotterdam APM terminal the K-S test (table F.12) returns the **Gamma, Erlang-4 or Beta** distributions as possible fits. This corresponds with the visual in figure F.21 from which it is clear that the data is not represented that well by the distributions.

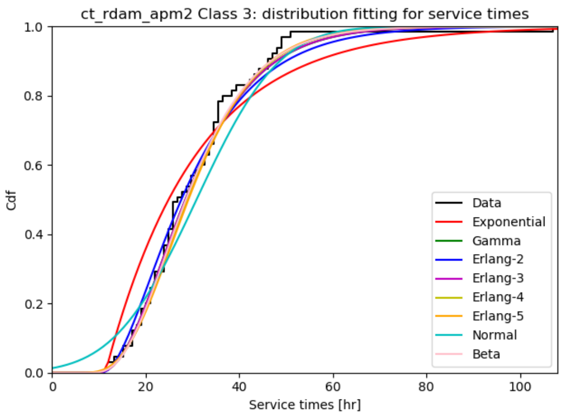


Figure F.20: Service time CDF Class 3 Rotterdam APM-2 Terminal

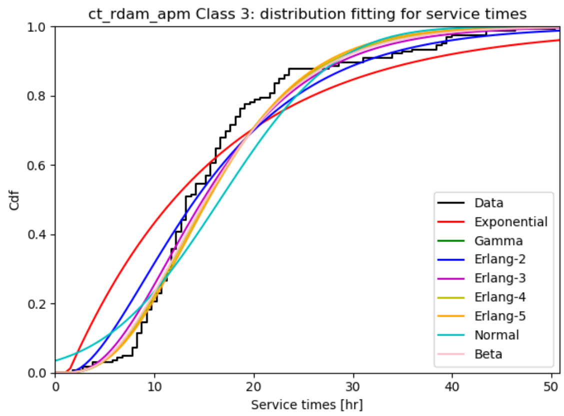


Figure F.21: Service time CDF Class 3 Rotterdam APM Terminal

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	11.40	19.48	1.00		0.21	0.01	No
Gamma	9.01	7.38	2.96		0.07	0.92	Yes
Erlang-2	10.69	10.10	2.00		0.09	0.64	Yes
Erlang-3	8.93	7.32	3.00		0.07	0.92	Yes
Erlang-4	6.70	6.05	4.00		0.07	0.92	Yes
Erlang-5	4.40	5.30	5.00		0.08	0.84	Yes
Normal	30.88	13.89			0.14	0.16	Yes
Beta	7.64	5.55E+12	3.67	8.85E+11	0.06	0.97	Yes

Table F.11: Service time distribution fitting for Class 3 container terminal: Rotterdam APM-2

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	1.35	15.28	1.00		0.30	0.00	No
Gamma	0.06	4.28	3.87		0.09	0.10	Yes
Erlang-2	1.20	7.71	2.00		0.17	0.00	No
Erlang-3	0.81	5.27	3.00		0.11	0.04	No
Erlang-4	-0.09	4.18	4.00		0.10	0.09	Yes
Erlang-5	-1.39	3.60	5.00		0.11	0.04	No
Normal	16.63	9.13			0.16	0.00	No
Beta	0.23	3.160.63	3.63	693.93	0.10	0.09	Yes

Table F.12: Service time distribution fitting for Class 3 container terminal: Rotterdam APM

Container Class 4: New Panamax

For the fourth container vessel class all 4 terminals are assumed to contain enough vessel tracks. First, the Rotterdam APM-2 terminal is assessed. All distributions could fit based on the K-S test (table F.13) and based on the visual results (figure F.22 the **Gammma, Beta and Erlang-5** distributions are chosen as best fits. The Rotterdam APM terminal is difficult to visually interpret, based on the extreme service time outlier (as mentioned in subchapter 5.1.1). However, when zooming in and based on the K-S tests no distributions are found that fit the data (table F.14).

For the Rotterdam Euromax terminal all distributions are found to fit the data, except the exponential distribution (table F.15). The best fits, together with the D statistic values and visual interpretations are the **Erlang-2, Beta and Gamma** distributions (figure F.24). Finally, the Le Havre Atlantic terminal has a slightly less fit based on the initial visual result (figure F.25). The **Gamma** distribution is chosen as best fit based on the K-S tests (table F.16).

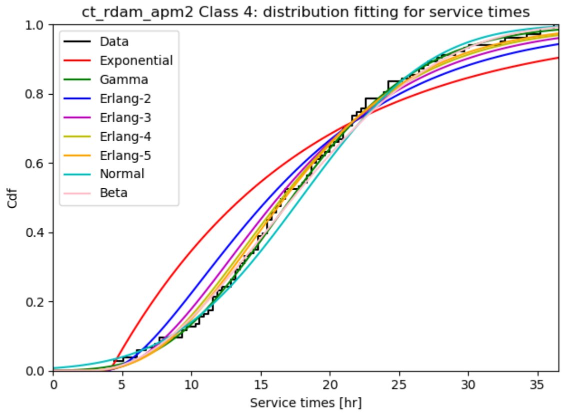


Figure F.22: Service time CDF Class 4 Rotterdam APM-2 Terminal

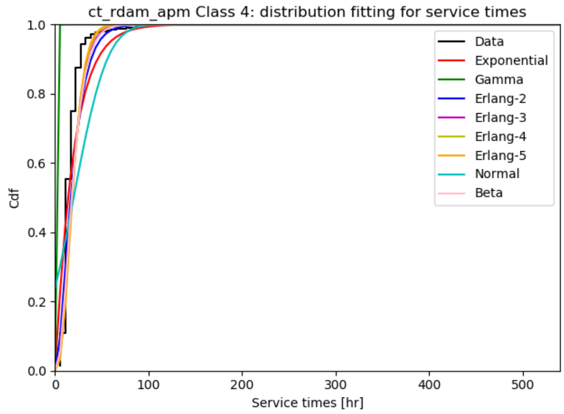


Figure F.23: Service time CDF Class 4 Rotterdam APM Terminal

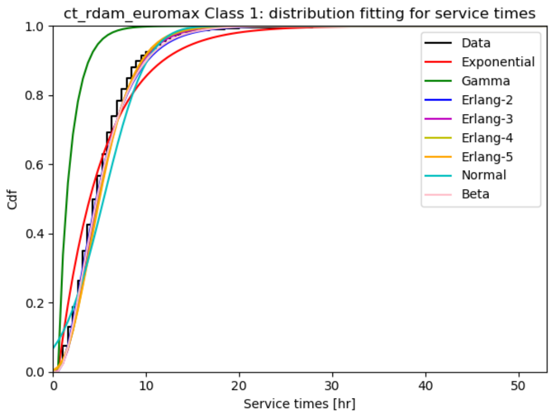


Figure F.24: Service time CDF Class 1 Rotterdam Euromax Terminal

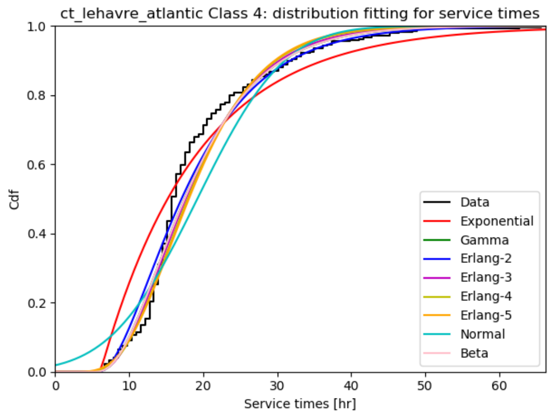


Figure F.25: Service time CDF Class 4 Le Havre Atlantic Terminal

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	4.12	13.82	1.00		0.25	0.00	No
Gamma	-8.74	2.06	12.94		0.04	0.99	Yes
Erlang-2	3.61	7.16	2.00		0.14	0.04	No
Erlang-3	2.57	5.12	3.00		0.09	0.34	Yes
Erlang-4	1.33	4.15	4.00		0.07	0.72	Yes
Erlang-5	0.05	3.58	5.00		0.06	0.91	Yes
Normal	17.94	7.37			0.08	0.55	Yes
Beta	2.06	43.33	2.53	4.38	0.05	0.95	Yes

Table F.13: Service time distribution fitting for Class 4 container terminal: Rotterdam APM-2

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	0.50	19.56	1.00		0.33	0.00	No
Gamma	0.50	3.64	0.00		0.99	0.00	No
Erlang-2	0.41	9.82	2.00		0.21	0.00	No
Erlang-3	0.11	6.65	3.00		0.14	0.00	No
Erlang-4	-0.78	5.21	4.00		0.15	0.00	No
Erlang-5	-2.49	4.51	5.00		0.16	0.00	No
Normal	20.06	27.98			0.32	0.00	No
Beta	0.17	1.40E+14	2.82	1.97E+13	0.14	0.00	No

Table F.14: Service time distribution fitting for Class 4 container terminal: Rotterdam APM

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	9.95	18.81	1.00		0.15	0.00	No
Gamma	9.43	9.30	2.08		0.06	0.78	Yes
Erlang-2	9.52	9.62	2.00		0.05	0.87	Yes
Erlang-3	7.40	7.12	3.00		0.09	0.23	Yes
Erlang-4	4.65	6.03	4.00		0.10	0.14	Yes
Erlang-5	1.98	5.36	5.00		0.10	0.11	Yes
Normal	28.76	13.64			0.11	0.06	Yes
Beta	9.43	5.01E+09	2.07	5.37E+08	0.06	0.78	Yes

Table F.15: Service time distribution fitting for Class 4 container terminal: Rotterdam Euromax

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	6.14	13.08	1.00		0.25	0.00	No
Gamma	5.39	5.30	2.61		0.11	0.00	No
Erlang-2	5.93	6.64	2.00		0.12	0.00	No
Erlang-3	4.88	4.78	3.00		0.12	0.00	No
Erlang-4	3.36	3.97	4.00		0.13	0.00	No
Erlang-5	1.76	3.49	5.00		0.14	0.00	No
Normal	19.22	9.24			0.19	0.00	No
Beta	5.40	4.07E+08	2.60	7.65E+07	0.11	0.00	No

Table F.16: Service time distribution fitting for Class 4 container terminal: Le Havre Atlantic

Container Class 5: Post New Panamax

The last container class contains the largest vessels based on length. The container terminal Le Havre Atlantic does not receive any vessels from this vessel class. The Rotterdam APM-2 Terminal visual is abnormal and only the Normal, Gamma and Beta distribution seem to fit the data (figure F.26). The Erlang-k distributions seem to contain a similar shape, but start much more at lower service times. Using the K-S tests these three distributions (**Normal, Gamma and Beta**) are possible fits to the data (table F.17).

For the Rotterdam APM terminal the same three distributions (**Normal, Beta and Gamma**) are selected as the best possible fits based on the K-S test and visual results (table F.18 and figure F.27). Finally for the Rotterdam Euromax Terminal the service times visually take on a different shape (figure F.28), compared to the other 2 terminals. Four distributions fit the data based on the K-S test (table F.19), the best fits are the **Gamma and Beta** distributions.

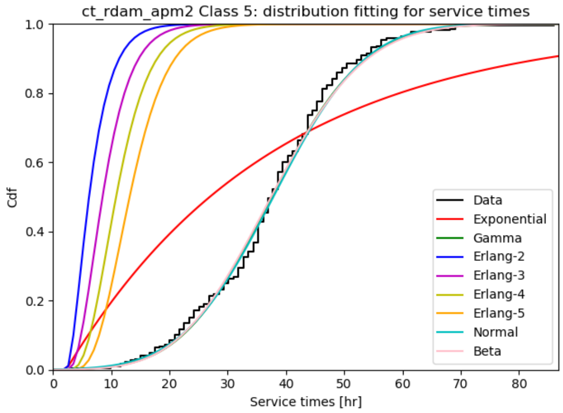


Figure F.26: Service time CDF Class 5 Rotterdam APM-2 Terminal

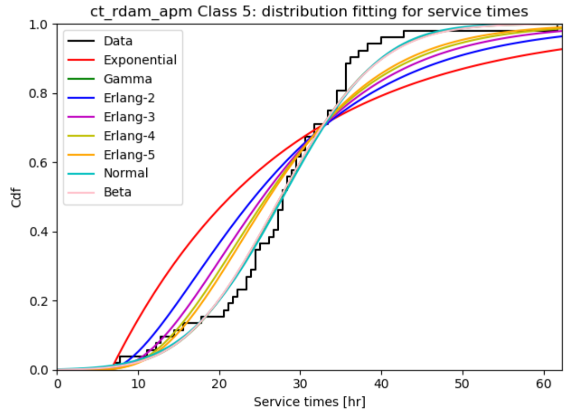


Figure F.27: Service time CDF Class 5 Rotterdam APM Terminal

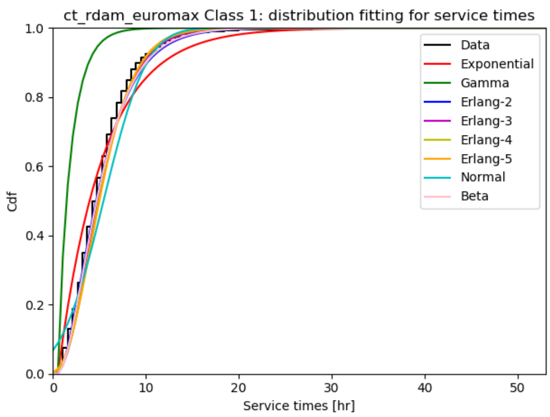


Figure F.28: Service time CDF Class 5 Rotterdam Euromax Terminal

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	2.27	35.56	1.00		0.33	0.00	No
Gamma	-142.81	0.84	214.59		0.07	0.24	Yes
Erlang-2	2.26	2.34	2.00		0.95	0.00	No
Erlang-3	2.25	2.34	3.00		0.91	0.00	No
Erlang-4	2.23	2.34	4.00		0.88	0.00	No
Erlang-5	2.22	2.34	5.00		0.82	0.00	No
Normal	37.83	12.34			0.06	0.38	Yes
Beta	-3.38	224.65	22.36	47.76	0.07	0.18	Yes

Table F.17: Service time distribution fitting for Class 5 container terminal: Rotterdam APM-2

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	6.67	21.23	1.00		0.34	0.00	No
Gamma	-77.51	0.83	126.42		0.09	0.75	Yes
Erlang-2	6.03	10.93	2.00		0.24	0.00	No
Erlang-3	4.93	7.66	3.00		0.20	0.02	No
Erlang-4	3.58	6.08	4.00		0.18	0.06	Yes
Erlang-5	2.14	5.15	5.00		0.17	0.10	Yes
Normal	27.90	9.40			0.09	0.80	Yes
Beta	-62.30	5.236.31	90.18	5.145.04	0.10	0.72	Yes

Table F.18: Service time distribution fitting for Class 5 container terminal: Rotterdam APM

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	12.42	23.21	1.00		0.28	0.00	No
Gamma	-2.20	33.15	11.41		0.05	0.55	Yes
Erlang-2	12.24	11.69	2.00		0.18	0.00	No
Erlang-3	11.63	80.00	3.00		0.12	0.00	No
Erlang-4	10.26	6.34	4.00		0.10	0.03	No
Erlang-5	8.53	5.42	5.00		0.08	0.08	Yes
Normal	35.63	11.24			0.07	0.23	Yes
Beta	-2.19	5.81E+05	11.41	1.75E+05	0.05	0.55	Yes

Table F.19: Service time distribution fitting for Class 5 container terminal: Rotterdam Euromax

F.2. Dry bulk terminals

F.2.1. Dry bulk terminals: total vessel mix

Rotterdam EMO Terminal

First, the EMO Dry bulk Terminal in Rotterdam is investigated. A lot of vessels seem to have a relatively short service time. By selecting a few vessel tracks an indication is formed about the small service times. Most of the vessels do berth at the terminal, based on Sea-web data and vessel track visualisations. However, a few vessel tracks seem to represent vessels that do not actually berth at the terminal. Two examples of these vessel tracks are demonstrated. The first vessel is the NEDSHIP Inland Waterway vessel with a vessel path as shown in figure F.31. The second is the AMOUREUS, an inland cargo ship, with its path visualised in figure F.32. These both fall under the wrongfully predicted vessel tracks, and should actually have been filtered out. In total 907 vessel tracks arrive during the selected time span.

None of the distributions pass the limit of the Null hypothesis based on the K-S test (table F.20). This corresponds with the visual represented in figure F.29, which shows none of the distribution representing the data well.

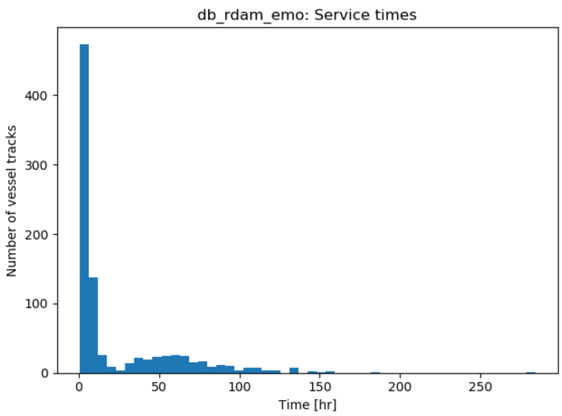


Figure F.29: Rotterdam EMO Terminal (histogram)

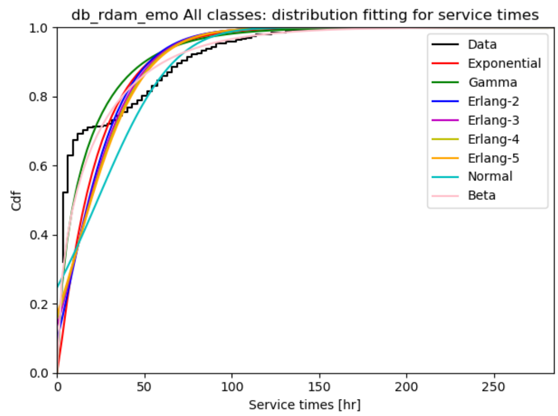


Figure F.30: Rotterdam EMO Terminal with fitted distributions

Figure F.31: Rotterdam EMO wrongfully predicted vessel (1)

Figure F.32: Rotterdam EMO wrongfully predicted vessel (2)

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	0.51	22.70	1.00		0.33	0.00	No
Gamma	0.51	33.95	0.55		0.16	0.00	No
Erlang-2	-8.13	15.67	2.00		0.34	0.00	No
Erlang-3	-17.18	13.46	3.00		0.33	0.00	No
Erlang-4	-25.13	12.08	4.00		0.33	0.00	No
Erlang-5	-32.28	11.10	5.00		0.33	0.00	No
Normal	23.21	33.96			0.31	0.00	No
Beta	0.51	629.43	0.46	12.70	0.16	0.00	No

Table F.20: Service time distribution fitting for dry bulk terminal: Rotterdam EMO

Vlissingen OVET

For the Vlissingen OVET dry bulk terminal the most service times again lie around smaller service times. In comparison with the Rotterdam EMO terminal the amount of smaller service times is less. Visually this also leads to more distributions that could match with the data set, as shown in figure F.34. However, none of the distributions seem to accurately match the data. Based on the K-S test only the Beta distribution matches with the data (table F.21). Nonetheless, based on the visual interpretation and the relatively high D statistic, the choice is made that no distributions match the data sufficiently.

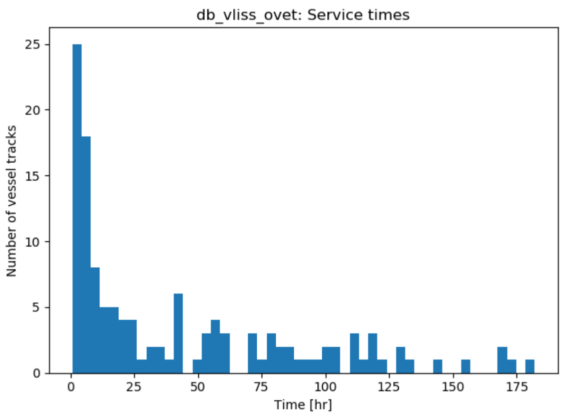


Figure F.33: Vlissingen OVET Terminal (histogram)

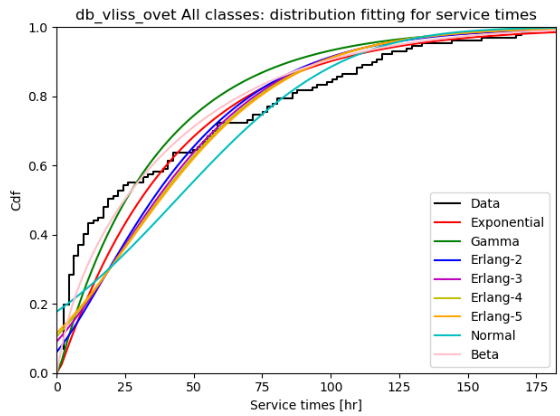


Figure F.34: Vlissingen OVET Terminal with fitted distributions

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	0.74	42.80	1.00		0.20	0.00	No
Gamma	0.74	38.32	0.94		0.15	0.00	No
Erlang-2	-10.80	27.16	2.00		0.21	0.00	No
Erlang-3	-23.72	22.42	3.00		0.21	0.00	No
Erlang-4	-35.09	19.66	4.00		0.20	0.00	No
Erlang-5	-45.33	17.77	5.00		0.20	0.00	No
Normal	43.53	47.05			0.19	0.00	No
Beta	0.74	330.64	0.65	4.98	0.10	0.14	Yes

Table F.21: Service time distribution fitting for dry bulk terminal: Vlissingen OVET

Rotterdam EECV

For the Rotterdam EECV terminal again the predominant of the service times is low (figure F.35). Multiple vessel tracks with (extremely) small service times are investigated. In total of all the 514 vessel tracks, 5 vessels stay less than 1 hour at the terminal (service time < 1 hour). Two examples are featured, first the VICTORIA Inland tanker stays at the terminal for just a little less than 1 hour. Sea-web also suggests the same vessel arrival. The same vessel arrives again, later in the time span, and stays around 50 minutes. Again, Sea-web confirms this short stay. The conclusion is made that these short arrivals are correctly represented in the data and that they do occur at dry bulk terminals.

Based on K-S tests no distributions fit the data (table F.22), which corresponds with the visual interpretation of figure F.36.

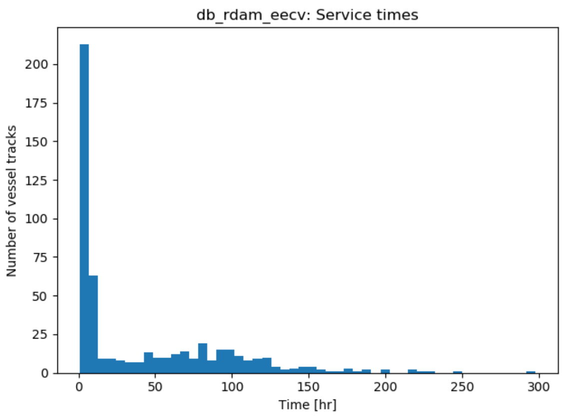


Figure F.35: Rotterdam EECV Terminal (histogram)

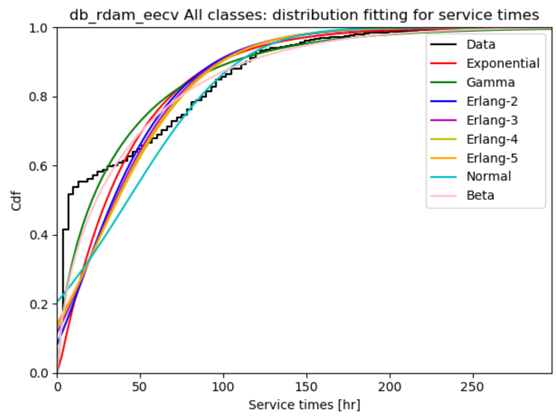


Figure F.36: Rotterdam EECV Terminal with fitted distributions

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	0.81	41.87	1.00		0.33	0.00	No
Gamma	0.81	68.89	0.58		0.20	0.00	No
Erlang-2	-13.27	27.98	2.00		0.32	0.00	No
Erlang-3	-27.73	23.47	3.00		0.30	0.00	No
Erlang-4	-40.28	20.74	4.00		0.30	0.00	No
Erlang-5	-51.52	18.84	5.00		0.29	0.00	No
Normal	42.68	51.89			0.26	0.00	No
Beta	0.81	509.40	0.54	5.94	0.21	0.00	No

Table F.22: Service time distribution fitting for dry bulk terminal: Rotterdam EECV

Dunkirk Western Bulk

The Dunkirk Western Bulk receives the least amount of arrivals (94) compared to the other three terminals. The visual fit is clearly possible for more distributions, compared to the other visual fits (figure F.38). This corresponds with the K-S test which returns that the Erlang-2, Erlang-3, Erlang-4, Erlang-5 or Beta distribution can all represent the data. The best fit, based on these and visual results, is the Erlang-2 distribution.

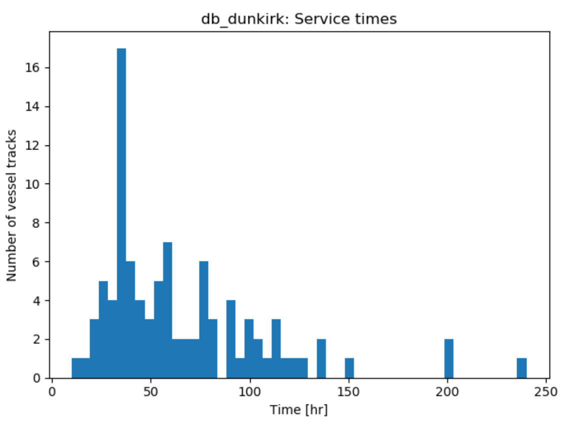


Figure F.37: Dunkirk Western Bulk Terminal (histogram)

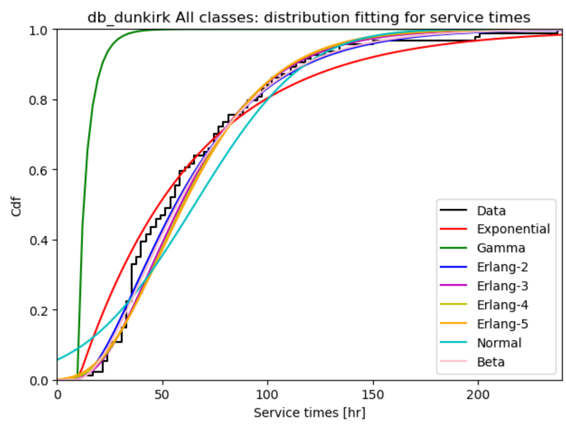


Figure F.38: Dunkirk Western Bulk Terminal with fitted distributions

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	10.16	55.32	1.00		0.21	0.00	No
Gamma	10.16	6.68	0.67		0.89	0.00	No
Erlang-2	8.93	28.27	2.00		0.08	0.52	Yes
Erlang-3	4.18	20.43	3.00		0.12	0.10	Yes
Erlang-4	-3.50	17.24	4.00		0.13	0.06	Yes
Erlang-5	-11.32	15.36	5.00		0.14	0.05	Yes
Normal	65.48	41.22			0.16	0.02	No
Beta	8.36	1.30E+13	2.20	4.98E+11	0.10	0.34	Yes

Table F.23: Service time distribution fitting for dry bulk terminal: Dunkirk Western Bulk

F.2.2. Dry bulk terminals: specific vessel classes

Arrivals per vessel class

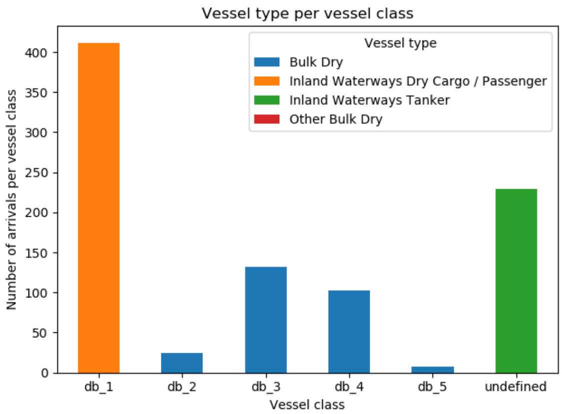


Figure F.39: Arrivals per vessel class: Rotterdam EMO Terminal

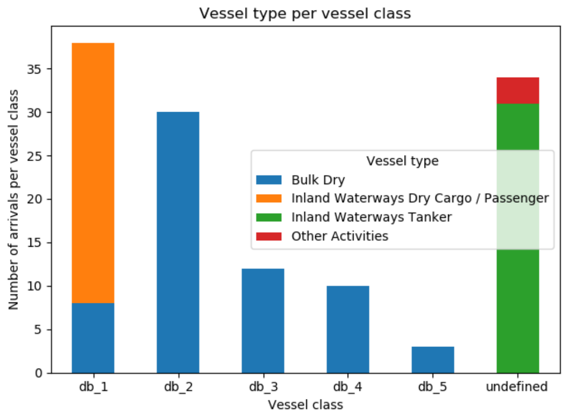


Figure F.40: Arrivals per vessel class: Vlissingen OVET Terminal

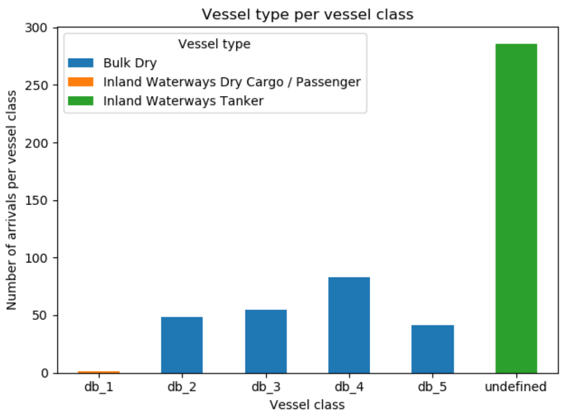


Figure F.41: Arrivals per vessel class: Rotterdam EECV Terminal

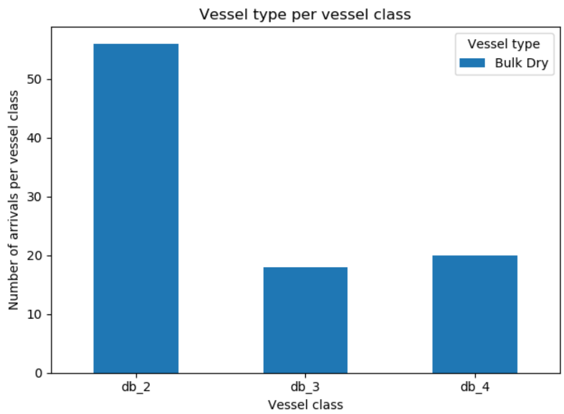


Figure F.42: Arrivals per vessel class: Dunkirk Western Bulk Terminal

Dry bulk Class 1: Small handy

For the first vessel class only the first two terminals can be used for reliable conclusions based on the small number of vessel arrivals in the Rotterdam EECV and Dunkirk Western Dry Bulk terminals. The Rotterdam EMO terminal K-S test presents three possible distributions: the Exponential, Gamma and Beta distributions. Visually and based on the test results (table F.24) the **Exponential and Beta** distributions represent the data the best. The second terminal, the Vlissingen OVET terminal has a lot of distributions that pass the K-S limit test (table F.25). All distributions except the Beta distribution are possible fits. Visually the best distributions is chosen to be the **Gamma** distribution (figure F.44). Between these two figures (F.43 and F.44) the influence of the number of vessel tracks (data messages) is clearly visible. The Rotterdam EMO terminal receives 412 vessels, whilst the Vlissingen OVET terminal receives a total of 38 arrivals of this specific vessel class.

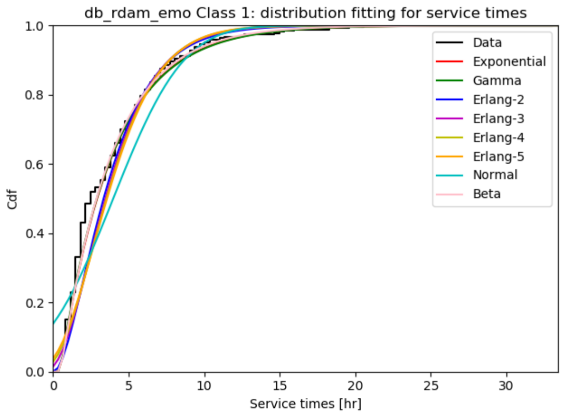


Figure F.43: Service time CDF Class 1 Rotterdam EMO Terminal

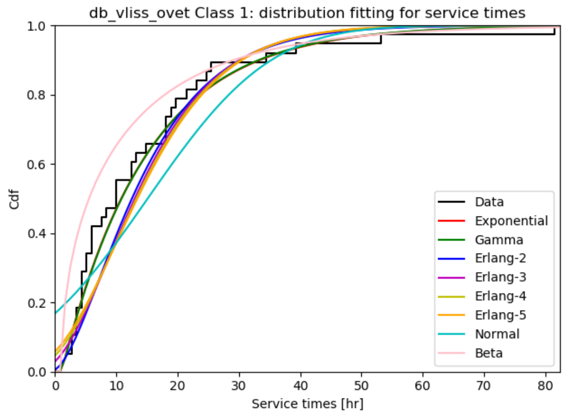


Figure F.44: Service time CDF Class 1 Vlissingen OVET Terminal

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	0.51	3.48	1.00		0.06	0.07	Yes
Gamma	0.51	3.56	0.98		0.06	0.11	Yes
Erlang-2	0.03	1.98	2.00		0.15	0.00	No
Erlang-3	-0.79	1.59	3.00		0.16	0.00	No
Erlang-4	-1.56	1.39	4.00		0.16	0.00	No
Erlang-5	-2.27	1.25	5.00		0.16	0.00	No
Normal	3.99	3.66			0.17	0.00	No
Beta	0.51	2.60E+13	1.03	7.87E+12	0.06	0.10	Yes

Table F.24: Service time distribution fitting for Class 1 dry bulk terminal: Rotterdam EMO

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	1.09	14.03	1.00		0.10	0.87	Yes
Gamma	1.08	13.78	1.02		0.10	0.85	Yes
Erlang-2	-0.80	7.96	2.00		0.18	0.14	Yes
Erlang-3	-4.24	6.45	3.00		0.18	0.14	Yes
Erlang-4	-7.51	5.66	4.00		0.18	0.15	Yes
Erlang-5	-10.54	5.13	5.00		0.18	0.17	Yes
Normal	15.12	15.73			0.19	0.13	Yes
Beta	1.09	1319.43	0.47	60.60	0.30	0.00	No

Table F.25: Service time distribution fitting for Class 1 dry bulk terminal: Vlissingen OVET

Dry bulk Class 2: Handy + Handymax + Supramax

For this vessel class all the terminals receive similar numbers of arrivals. The lowest number of arrivals is for the Rotterdam EMO terminal with only 24 vessels arriving. The highest is for the Dunkirk Western Bulk Terminal with 56 arrivals. The minimum amount of vessel arrivals in order to be classified is again set at 30 vessels. Thus, three terminals (all except Rotterdam EMO) are analysed, however, the small number of data points is taken into consideration.

First, the Vlissingen OVET Terminal is assessed. Again, all distributions pass the limit of the K-S test (table F.26). Based on the visual interpretation of figure F.45 the **Gamma and Erlang-2** distributions are chosen as the best theoretical distributions that fit.

The Rotterdam EECV terminal also obtains a lot of possible fits: all distributions except the Beta distribution (table F.27). The conclusion is made, using the visual results as in figure F.46, that all the **Erlang-k** distributions fit the data. Finally, the Dunkirk Western Bulk service time distribution can be represented by the Gamma, Erlang-2, Erlang-3 or Beta distributions, based on the K-S test (table F.28). Taking figure F.47 into account returns that the **Gamma or Beta** distribution fit the data best.

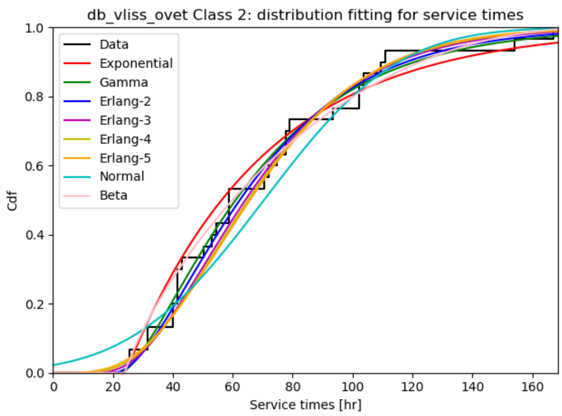


Figure F.45: Service time CDF Class 2 Vlissingen OVET Terminal

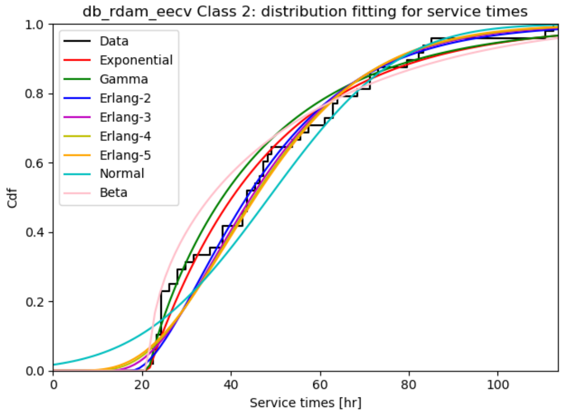


Figure F.46: Service time CDF Class 2 Rotterdam EECV Terminal

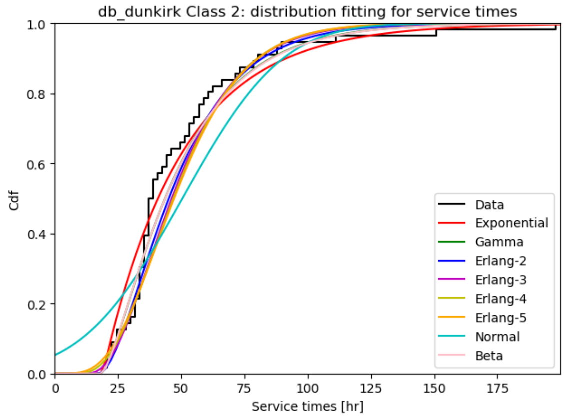


Figure F.47: Service time CDF Class 2 Dunkirk Western Bulk Terminal

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	5.64	80.54	1.00	0.13	0.71	Yes	Yes
Gamma	5.64	90.75	0.75	0.18	0.26	Yes	Yes
Erlang-2	-7.68	46.93	2.00	0.12	0.75	Yes	Yes
Erlang-3	5.64	3.40	0.23	0.61	0.00	No	Yes
Erlang-4	-1.52	13.91	4.00		0.19	0.31	Yes
Erlang-5	-8.05	12.44	5.00		0.19	0.29	Yes
Normal	54.13	34.00			0.19	0.33	Yes
Beta	12.06	7.50E+10	1.65	2.91E+09	0.13	0.81	Yes

Table F.26: Service time distribution fitting for Class 2 dry bulk terminal: Vlissingen OVET

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	21.55	27.06	1.00		0.13	0.37	Yes
Gamma	21.55	29.52	0.86		0.17	0.13	Yes
Erlang-2	17.44	15.58	2.00		0.14	0.28	Yes
Erlang-3	11.67	12.31	3.00		0.13	0.37	Yes
Erlang-4	6.41	10.55	4.00		0.13	0.40	Yes
Erlang-5	1.60	9.40	5.00		0.12	0.41	Yes
Normal	48.61	22.92			0.13	0.34	Yes
Beta	21.55	169.08	0.50	2.88	0.20	0.04	No

Table F.27: Service time distribution fitting for Class 2 dry bulk terminal: Rotterdam EECV

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	18.99	31.49	1.00		0.18	0.04	No
Gamma	18.54	21.69	1.47		0.12	0.37	Yes
Erlang-2	17.11	16.68	2.00		0.15	0.13	Yes
Erlang-3	11.89	12.86	3.00		0.18	0.05	Yes
Erlang-4	6.09	11.10	4.00		0.19	0.04	No
Erlang-5	0.49	10.00	5.00		0.19	0.03	No
Normal	50.48	31.07			0.19	0.03	No
Beta	18.56	1.14E+13	1.46	5.21E+11	0.12	0.38	Yes

Table F.28: Service time distribution fitting for Class 2 dry bulk terminal: Dunkirk Western Bulk

Dry bulk Class 3: Panamax

For the third vessel class only two terminals contain enough vessel arrivals in order to make reliable conclusions: the Rotterdam EMO and Rotterdam EECV terminals. The Rotterdam EMO terminal service time distribution can be fitted by the Erlang-3, Erlang-4, Erlang-5 or Beta distribution based on the K-S test (table F.29). The best fit is chosen to be the **Erlang-5** distribution. Based on the visualisation of the Rotterdam EECV terminal, as shown in figure F.49, more distributions are expected to fit the data. This corresponds with the results from the K-S test in table F.30: all distributions pass the K-S limit test. The best fit is chosen to be the **Beta, Gamma or Erlang-2** distribution, based on the D statistic and visual interpretations.

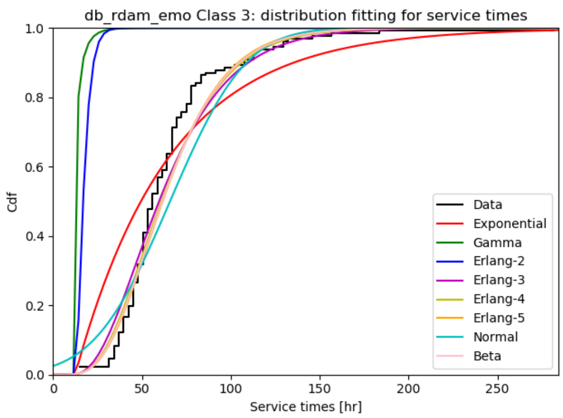


Figure F.48: Service time CDF Class 3 Rotterdam EMO Terminal

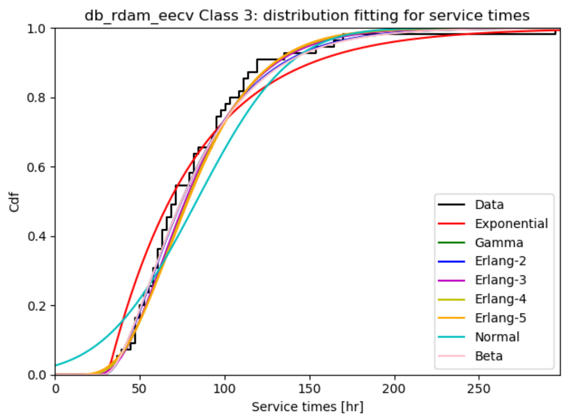


Figure F.49: Service time CDF Class 3 Rotterdam EECV Terminal

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	12.50	53.34	1.00		0.30	0.00	No
Gamma	12.50	6.62	0.21		0.97	0.00	No
Erlang-2	12.49	2.66	2.00		0.97	0.00	No
Erlang-3	9.08	18.92	3.00		0.12	0.05	Yes
Erlang-4	5.88	14.99	4.00		0.11	0.10	Yes
Erlang-5	1.91	12.78	5.00		0.11	0.08	Yes
Normal	65.84	33.55			0.18	0.00	No
Beta	3.65	3.67E+12	4.58	2.70E+11	0.11	0.09	Yes

Table F.29: Service time distribution fitting for Class 3 dry bulk terminal: Rotterdam EMO

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	31.16	52.29	1.00		0.18	0.06	Yes
Gamma	28.88	27.73	1.97		0.06	0.99	Yes
Erlang-2	28.73	27.36	2.00		0.06	0.98	Yes
Erlang-3	22.32	20.38	3.00		0.09	0.75	Yes
Erlang-4	14.60	17.21	4.00		0.10	0.56	Yes
Erlang-5	6.98	15.29	5.00		0.11	0.47	Yes
Normal	83.45	42.99			0.14	0.19	Yes
Beta	29.05	2.38E+10	1.92	8.38E+08	0.06	0.98	Yes

Table F.30: Service time distribution fitting for Class 3 dry bulk terminal: Rotterdam EECV

Dry bulk Class 4: Mini Capesize + Capesize

Similar as for Class 3, this fourth class also only can be represented by the Rotterdam EMO and EECV terminals. The Rotterdam EMO terminal visually presents very good fits (figure F.50 for multiple distributions. The K-S test limit is passed by all the distributions except the Exponential distribution (table F.31), the best fit is found in the **Erlang-5 or Beta** distributions. The Rotterdam EECV also returns multiple possible distributions: the Gamma, Erlang-3, Erlang-4, Erlang-5, Normal or Beta distribution (table F.32). The best distributions based on the K-S statistic are the **Beta and Gamma** distributions, which corresponds to the best visual fits as well (figure F.51).

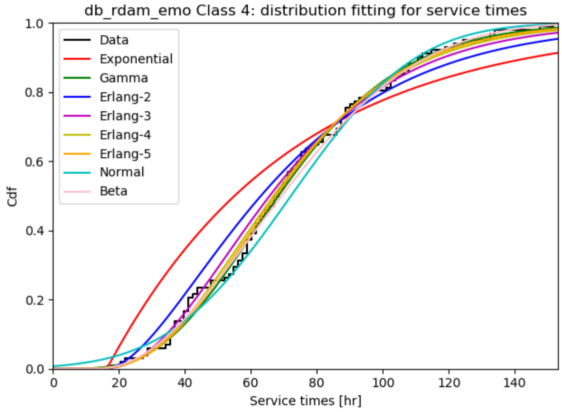


Figure F.50: Service time CDF Class 4 Rotterdam EMO Terminal

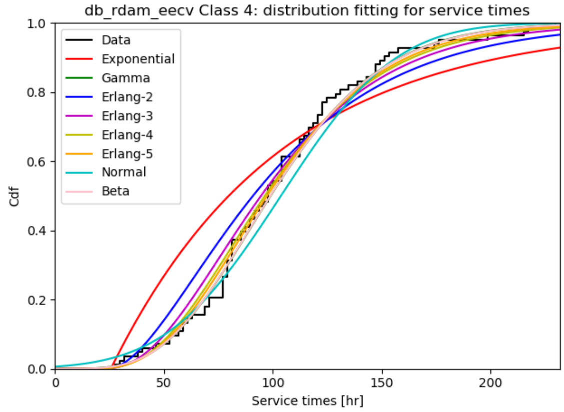


Figure F.51: Service time CDF Class 4 Rotterdam EECV Terminal

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	16.35	55.84	1.00		0.23	0.00	No
Gamma	-6.68	11.54	6.84		0.06	0.84	Yes
Erlang-2	15.36	28.41	2.00		0.13	0.05	Yes
Erlang-3	12.55	19.88	3.00		0.09	0.43	Yes
Erlang-4	7.88	16.08	4.00		0.06	0.80	Yes
Erlang-5	2.68	13.90	5.00		0.05	0.93	Yes
Normal	72.19	29.66			0.08	0.58	Yes
Beta	11.94	180.99	2.43	4.86	0.05	0.97	Yes

Table F.31: Service time distribution fitting for Class 4 dry bulk terminal: Rotterdam EMO

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	25.66	78.08	1.00		0.28	0.00	No
Gamma	-13.69	14.45	8.13		0.08	0.73	Yes
Erlang-2	23.89	39.93	2.00		0.18	0.01	No
Erlang-3	19.58	28.06	3.00		0.13	0.10	Yes
Erlang-4	13.45	22.57	4.00		0.11	0.26	Yes
Erlang-5	6.76	19.40	5.00		0.09	0.42	Yes
Normal	103.74	41.49			0.09	0.45	Yes
Beta	-11.89	3.65E+07	7.83	2.47E+06	0.08	0.71	Yes

Table F.32: Service time distribution fitting for Class 4 dry bulk terminal: Rotterdam EECV

Dry bulk Class 5: Very Large Bulk Carrier + Very large Ore Carrier

The fifth dry bulk class contains the largest possible dry bulk vessels. The only terminal that receives enough vessels of this class size is the Rotterdam EECV terminal. Based on the K-S test all distributions except the Beta distribution are fits to the service time distribution (table F.33). Visually the distributions don't perfectly fit the data, the best fit would be the **Gamma** distribution (figure F.52).

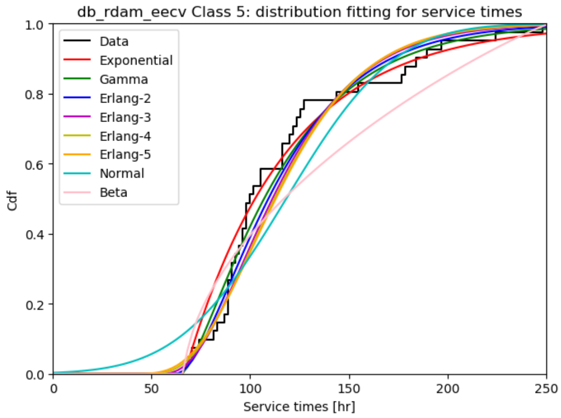


Figure F.52: Service time CDF Class 5 Rotterdam EECV Terminal

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	66.74	51.56	1.00		0.19	0.09	Yes
Gamma	65.00	35.42	1.50		0.12	0.56	Yes
Erlang-2	62.04	28.13	2.00		0.11	0.68	Yes
Erlang-3	53.93	21.45	3.00		0.14	0.40	Yes
Erlang-4	45.23	18.27	4.00		0.15	0.30	Yes
Erlang-5	36.82	16.30	5.00		0.15	0.26	Yes
Normal	118.30	43.09			0.19	0.09	Yes
Beta	66.74	185.19	0.57	1.10	0.22	0.04	No

Table F.33: Service time distribution fitting for Class 5 dry bulk terminal: Rotterdam EECV

F.3. Liquid bulk terminals

F.3.1. Liquid bulk terminals: total vessel mix

Rotterdam GATE Terminal

First, the Rotterdam GATE terminal is analysed. The terminal contains 2 berths and a total of 173 vessels arrive over the selected time span. Figure F.53 visualises the service time distribution, in which a peak is observed around 22-25 hours. This corresponds with the average and median values of the service time being 22.98 and 23.85 hours respectively. Figure G.42 visualises the CDF of the service time including all possible distributions. From this figure, as well as from numerical data based on the K-S test (table F.34) it is clear that no distributions fit on the data.

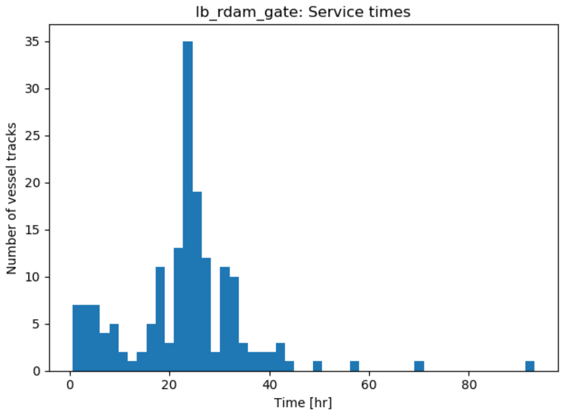


Figure F.53: Rotterdam GATE Terminal (histogram)

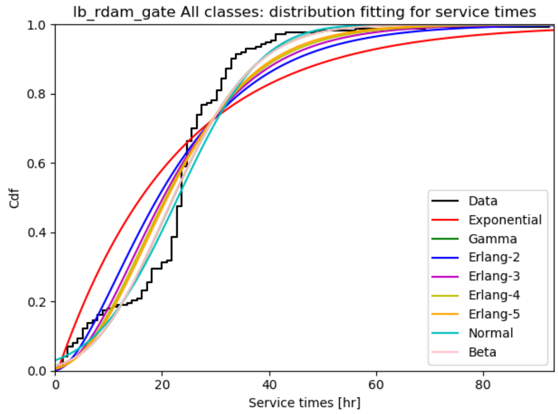


Figure F.54: Rotterdam GATE Terminal with fitted distributions

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	0.51	22.47	1.00		0.30	0.00	No
Gamma	-32.27	2.59	21.32		0.18	0.00	No
Erlang-2	-0.25	11.61	2.00		0.25	0.00	No
Erlang-3	-2.40	8.46	3.00		0.23	0.00	No
Erlang-4	-4.66	6.91	4.00		0.22	0.00	No
Erlang-5	-6.84	5.96	5.00		0.21	0.00	No
Normal	22.98	12.19			0.15	0.00	No
Beta	-30.73	2.04E+07	20.05	7.61E+06	0.18	0.00	No

Table F.34: Service time distribution fitting for liquid bulk terminal: Rotterdam GATE

Zeebrugge LNG Terminal

Next, the Zeebrugge LNG terminal is examined. Again this terminal contains two berths and a similar amount of vessels arrive (185 vessels) as did in the Rotterdam GATE terminal (173 vessels). Another similarity between the two terminals is the location of the peak (figure F.55) which corresponds for this terminal with a mean of 26.47 hours and a median of 24.90 hours. For the Zeebrugge Terminal again no distributions fit the service time distribution, based on the numerical results (table F.35) and the visualisation (figure G.44).

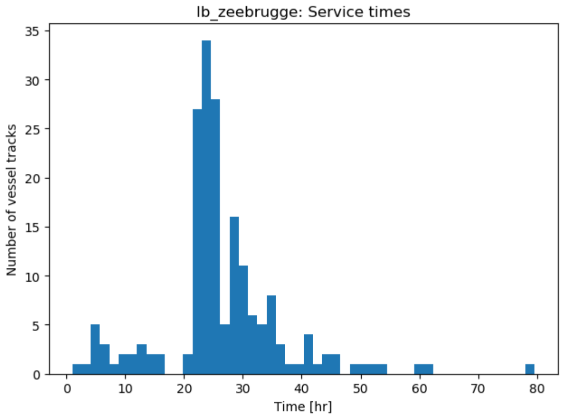


Figure F.55: Zeebrugge LNG Terminal (histogram)

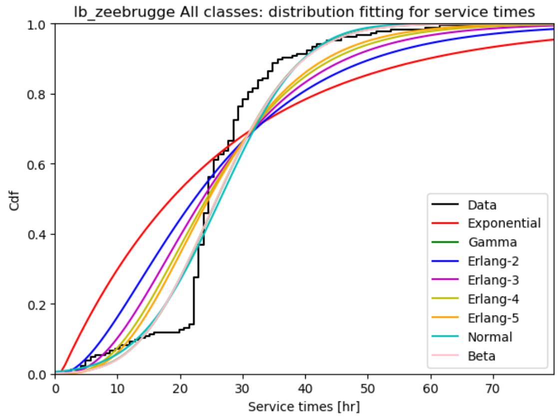


Figure F.56: Zeebrugge LNG Terminal with fitted distributions

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	0.96	25.51	1.00		0.43	0.00	No
Gamma	-35.12	1.69	36.43		0.22	0.00	No
Erlang-2	0.76	12.86	2.00		0.36	0.00	No
Erlang-3	0.30	8.72	3.00		0.32	0.00	No
Erlang-4	-0.58	6.76	4.00		0.30	0.00	No
Erlang-5	-1.78	5.65	5.00		0.28	0.00	No
Normal	26.47	10.38			0.20	0.00	No
Beta	-30.78	7202.90	31.05	3873.78	0.22	0.00	No

Table F.35: Service time distribution fitting for liquid bulk terminal: Zeebrugge

Dunkirk LNG Terminal

The Dunkirk LNG terminal contains only 1 berth. In comparison to the two previous terminals the number of arrivals is therefore also a lot lower (68 arrivals). The peak of the service time distributions, as shown in figure F.57, lies around the same values, corresponding to a mean of 25.20 hours and a median of 24.85 hours. All fitted distributions do not pass the K-S limit test and do not fit visually on the service time distribution (table F.36 and figure G.46).

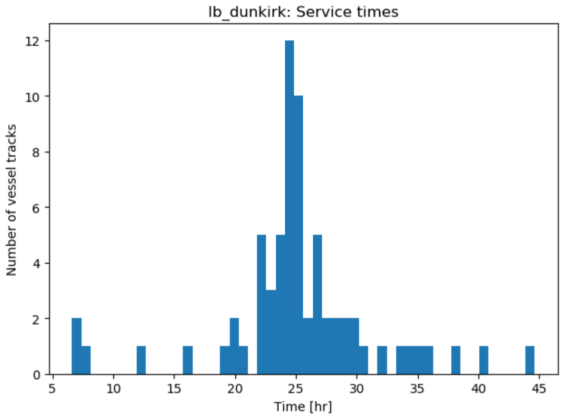


Figure F.57: Dunkirk LNG Terminal (histogram)

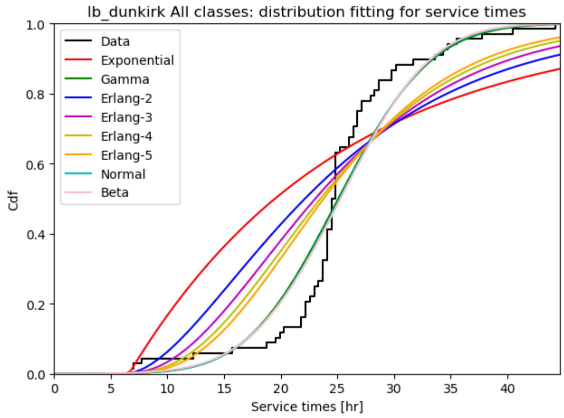


Figure F.58: Dunkirk LNG Terminal with fitted distributions

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	6.61	18.59	1.00		0.43	0.00	No
Gamma	-113.59	0.30	465.83		0.18	0.02	No
Erlang-2	6.15	9.52	2.00		0.36	0.00	No
Erlang-3	5.43	6.59	3.00		0.33	0.00	No
Erlang-4	4.62	5.14	4.00		0.31	0.00	No
Erlang-5	3.79	4.28	5.00		0.29	0.00	No
Normal	25.20	6.34			0.18	0.03	No
Beta	-4021.95	4205.09	15328.55	598.19	0.17	0.03	No

Table F.36: Service time distribution fitting for liquid bulk terminal: Dunkirk

France Montoir LNG Terminal

The last LNG terminal analysed contains two berths. Again, the peak of the service time distribution seems to lie around the same values as the earlier examined terminals (figure F.59). The mean of the distribution is 32.11 hours and the median 27.47 hours. From figure G.48 it is clear that no distributions fit the data set, which corresponds with the outcomes of the K-S tests (table F.37).

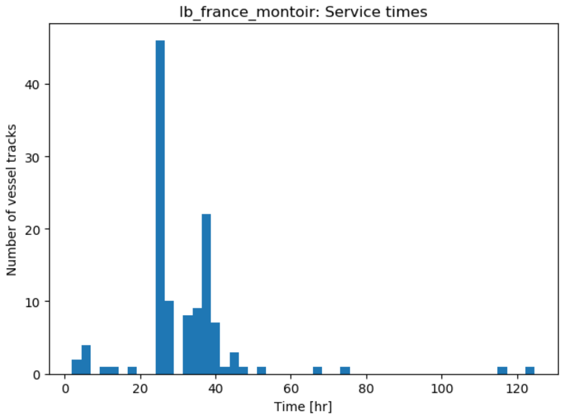


Figure F.59: France Montoir LNG Terminal (histogram)

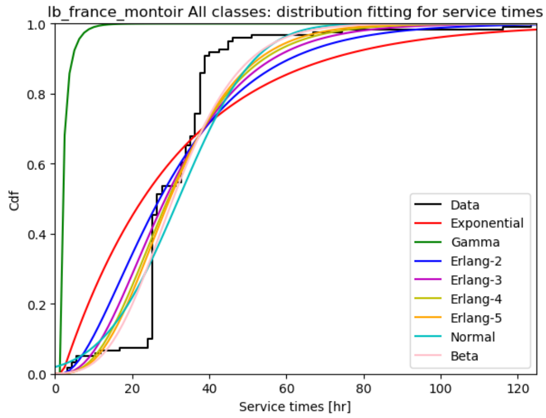


Figure F.60: France Montoir LNG Terminal with fitted distributions

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	1.96	30.15	1.00		0.46	0.00	No
Gamma	1.96	3.29	0.26		0.94	0.00	No
Erlang-2	1.51	15.30	2.00		0.38	0.00	No
Erlang-3	0.49	10.54	3.00		0.33	0.00	No
Erlang-4	-0.99	8.28	4.00		0.31	0.00	No
Erlang-5	-2.69	6.96	5.00		0.29	0.00	No
Normal	32.11	15.48			0.24	0.00	No
Beta	-10.34	2.99E+07	9.13	6.42E+06	0.25	0.00	No

Table F.37: Service time distribution fitting for liquid bulk terminal: France Montoir

F.3.2. Liquid bulk terminals: specific vessel classes

Arrivals per vessel class

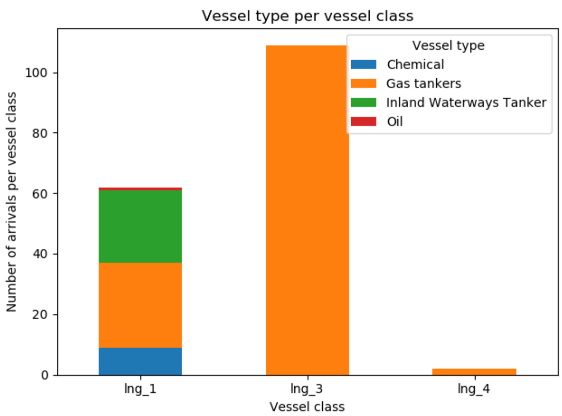


Figure F.61: Arrivals per vessel class: Rotterdam GATE Terminal

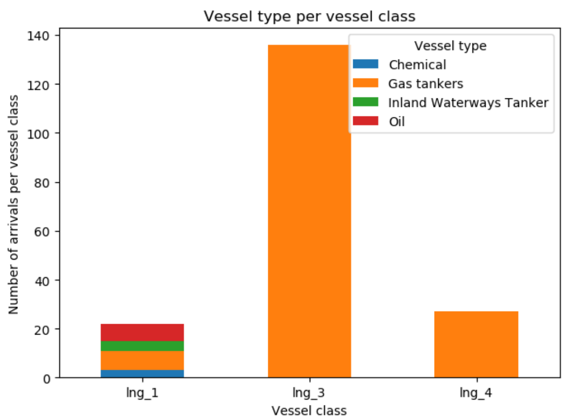


Figure F.62: Arrivals per vessel class: Zeebrugge LNG Terminal

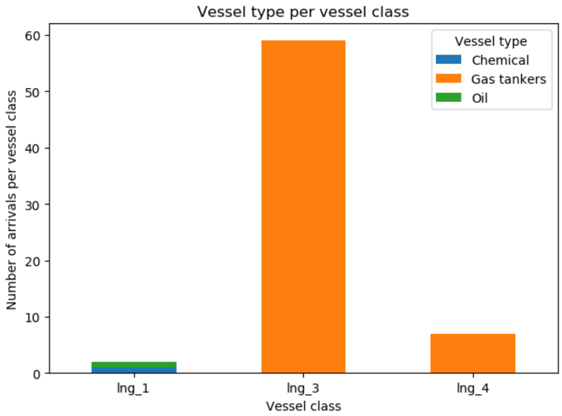


Figure F.63: Arrivals per vessel class: Dunkirk LNG Terminal

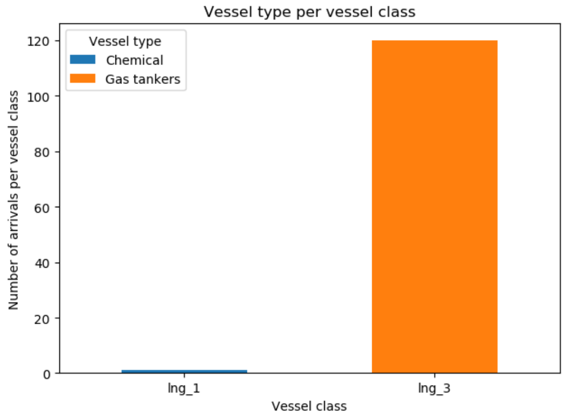


Figure F.64: Arrivals per vessel class: France Montoir LNG Terminal

LNG Class 1: Small Spherical / Membrane LNG Carriers

As mentioned, for this first class only the Rotterdam GATE terminal and the Zeebrugge LNG terminal are taken into account. The K-S test based on the Rotterdam GATE's service times results in almost all distributions fitting the data, except the Gamma distribution (table F.38). Visually the fits are not perfect and the best fits are the **Erlang-k** distributions (figure F.65).

The Zeebrugge terminal has multiple distributions that fit based on the K-S test (table F.39): all distributions fit except the Beta distribution. Based on the D statistic and visual results (figure F.66) the best distribution to fit the service times is the **Exponential** distribution (not perfect but best fit).

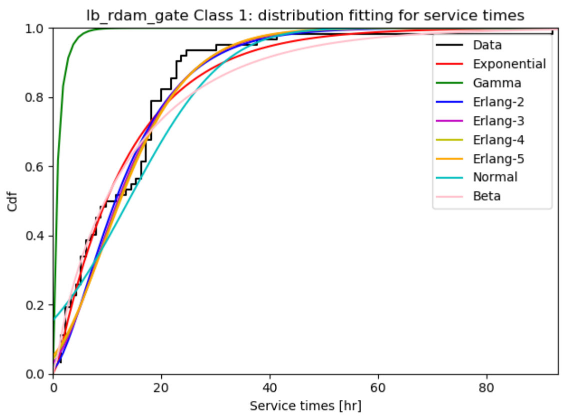


Figure F.65: Service time CDF Class 1 Rotterdam GATE Terminal

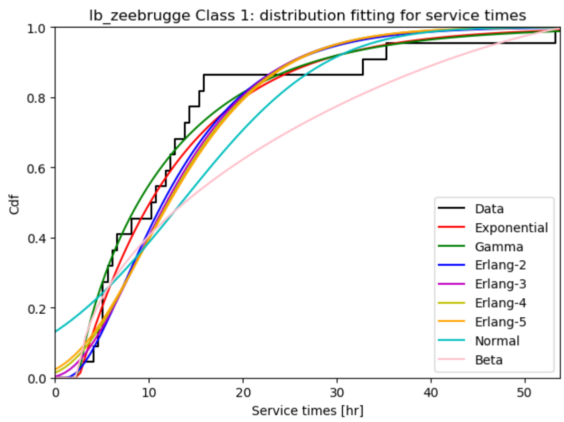


Figure F.66: Service time CDF Class 1 Zeebrugge LNG Terminal

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	0.51	13.43	1.00		0.14	0.19	Yes
Gamma	0.51	2.11	0.34		0.77	0.00	No
Erlang-2	-1.10	7.52	2.00		0.12	0.33	Yes
Erlang-3	-3.96	5.97	3.00		0.12	0.31	Yes
Erlang-4	-6.65	5.15	4.00		0.12	0.27	Yes
Erlang-5	-9.14	4.62	5.00		0.13	0.26	Yes
Normal	13.94	13.71			0.16	0.06	Yes
Beta	0.51	4531.59	0.77	234.39	0.13	0.22	Yes

Table F.38: Service time distribution fitting for dry bulk terminal: Rotterdam GATE

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	2.55	10.93	1.00		0.15	0.65	Yes
Gamma	2.55	12.98	0.78		0.18	0.45	Yes
Erlang-2	1.15	6.17	2.00		0.18	0.45	Yes
Erlang-3	-1.43	4.97	3.00		0.18	0.43	Yes
Erlang-4	-3.95	4.36	4.00		0.19	0.37	Yes
Erlang-5	-6.31	3.96	5.00		0.20	0.32	Yes
Normal	13.48	12.00			0.28	0.06	Yes
Beta	2.55	52.15	0.53	1.24	0.31	0.02	No

Table F.39: Service time distribution fitting for dry bulk terminal: Zeebrugge LNG

LNG Class 2: Medium Spherical / Membrane LNG Carriers

None of the four analysed terminals receive vessels from this second vessel class.

LNG Class 3: Large Spherical / Membrane LNG Carriers

All four terminals contain enough vessel arrivals of the third LNG vessel class. First, the Rotterdam GATE terminal is investigated. No fit is found based on all the distributions and testing using the K-S goodness-of-fit test (table F.40). This corresponds with what is concluded from the visualisations (figure F.67). For the Zeebrugge LNG terminal visually no fits seem possible as well (figure F.68) and no distributions pass the K-S test limit.

For the Dunkirk LNG terminal and the France Montoir LNG terminal similar conclusions can be drawn. For both terminals none of the distributions fit the data well enough, both based on the visual interpretations (figures F.69 and F.70) and the K-S tests (tables F.42 and F.43).

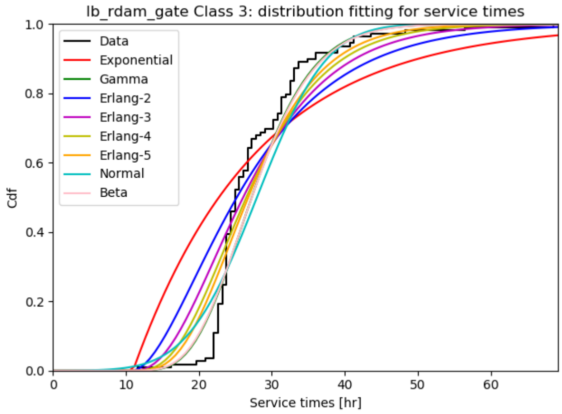


Figure F.67: Service time CDF Class 3 Rotterdam GATE Terminal

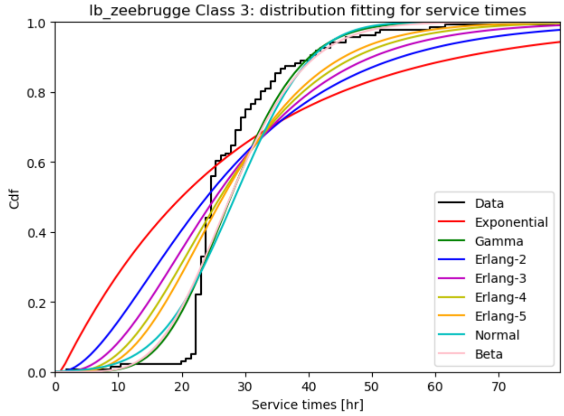


Figure F.68: Service time CDF Class 3 Zeebrugge LNG Terminal

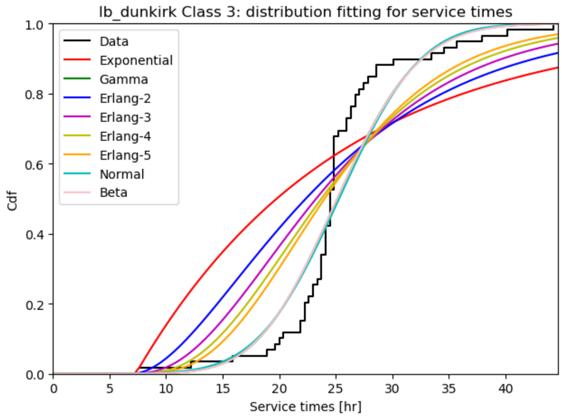


Figure F.69: Service time CDF Class 3 Dunkirk LNG

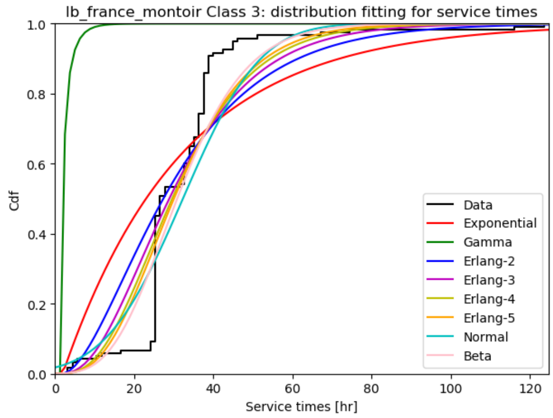


Figure F.70: Service time CDF Class 3 France Montoir LNG

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	10.97	17.00	1.00		0.45	0.00	No
Gamma	6.99	2.16	9.73		0.16	0.01	No
Erlang-2	10.79	8.59	2.00		0.34	0.00	No
Erlang-3	10.57	5.80	3.00		0.29	0.00	No
Erlang-4	10.28	4.42	4.00		0.25	0.00	No
Erlang-5	9.91	3.61	5.00		0.22	0.00	No
Normal	27.97	7.51			0.20	0.00	No
Beta	6.93	1.793.95	9.39	790.53	0.16	0.01	No

Table F.40: Service time distribution fitting for dry bulk terminal: Rotterdam GATE

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	0.96	27.45	1.00		0.50	0.00	No
Gamma	-10.74	2.02	19.40		0.21	0.00	No
Erlang-2	0.75	13.84	2.00		0.42	0.00	No
Erlang-3	0.49	9.31	3.00		0.37	0.00	No
Erlang-4	0.18	7.06	4.00		0.34	0.00	No
Erlang-5	-0.18	5.72	5.00		0.31	0.00	No
Normal	28.42	9.50			0.21	0.00	No
Beta	-6.35	1.03E+05	13.64	4.01E+04	0.22	0.00	No

Table F.41: Service time distribution fitting for dry bulk terminal: Zeebrugge LNG

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	7.32	17.98	1.00		0.44	0.00	No
Gamma	-43.90	0.46	150.57		0.18	0.04	No
Erlang-2	6.99	9.15	2.00		0.37	0.00	No
Erlang-3	6.62	6.22	3.00		0.33	0.00	No
Erlang-4	6.21	4.77	4.00		0.30	0.00	No
Erlang-5	5.76	3.91	5.00		0.28	0.00	No
Normal	25.29	5.66			0.19	0.03	No
Beta	-34.97	1.70E+06	113.50	3.20E+06	0.18	0.05	No

Table F.42: Service time distribution fitting for dry bulk terminal: Dunkirk LNG

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	1.96	30.37	1.00		0.46	0.00	No
Gamma	1.96	3.26	0.25		0.95	0.00	No
Erlang-2	1.53	15.40	2.00		0.38	0.00	No
Erlang-3	0.63	10.57	3.00		0.33	0.00	No
Erlang-4	-0.70	8.26	4.00		0.31	0.00	No
Erlang-5	-2.25	6.91	5.00		0.29	0.00	No
Normal	32.33	15.36			0.25	0.00	No
Beta	-9.04	2.28E+07	8.82	4.85E+06	0.25	0.00	No

Table F.43: Service time distribution fitting for dry bulk terminal: France Montoir LNG

LNG Class 4: Very large Spherical / Membrane LNG Carriers

As mentioned earlier, only the Zeebrugge LNG terminal receives these largest LNG Class 4 vessels. Based on K-S tests all the distributions fit the service times for this terminal (table F.44). The **Normal** distribution visually fits the best to the service time distribution (figure F.71).

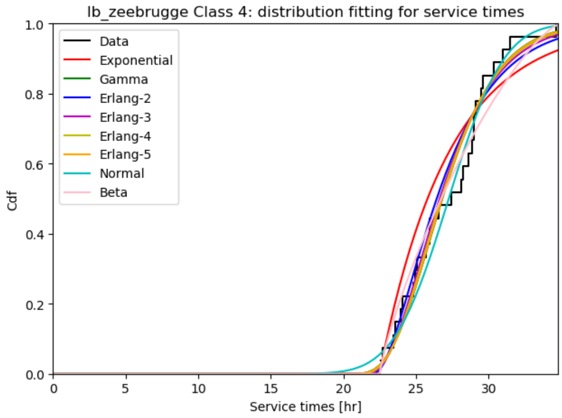


Figure F.71: Service time CDF Class 4 Zeebrugge LNG Terminal

	Distribution parameters				K-S test		
Distribution	Location	Scale	Shape	Shape (b)	D	p	Lim
Exponential	22.49	4.77	1.00		0.18	0.30	Yes
Gamma	20.07	1.35	5.34		0.15	0.53	Yes
Erlang-2	22.10	2.58	2.00		0.17	0.40	Yes
Erlang-3	21.49	1.92	3.00		0.16	0.45	Yes
Erlang-4	20.86	1.60	4.00		0.16	0.49	Yes
Erlang-5	20.26	1.40	5.00		0.15	0.52	Yes
Normal	27.26	2.98			0.11	0.91	Yes
Beta	22.49	12.45	0.83	1.34	0.12	0.84	Yes

Table F.44: Service time distribution fitting for dry bulk terminal: Zeebrugge LNG

G

Appendix G: Results for goodness-of-fit tests for the inter arrival time distribution

G.1. Container terminals

G.1.1. Container terminals: total vessel mix

Rotterdam APM-2 Terminal

Results are visualised and demonstrated in chapter 6.1.1.

Rotterdam APM Terminal

For the Rotterdam APM Terminal similar conclusions can be drawn. Visually the **Exponential or Gamma** distribution is the best fit on the data (figure 6.2). However, based on K-S test results none of the distributions pass the Null Hypothesis limit (results in table G.1).

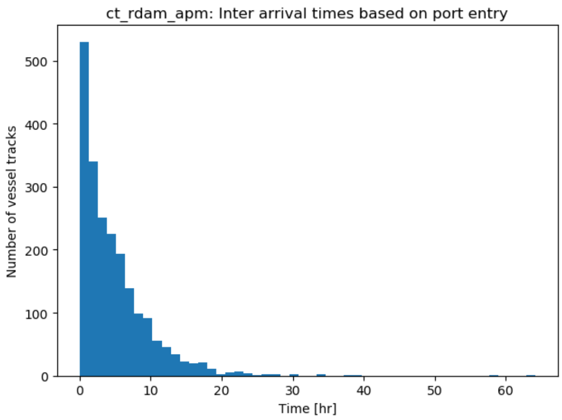


Figure G.1: Rotterdam APM Terminal inter arrival times (histogram)

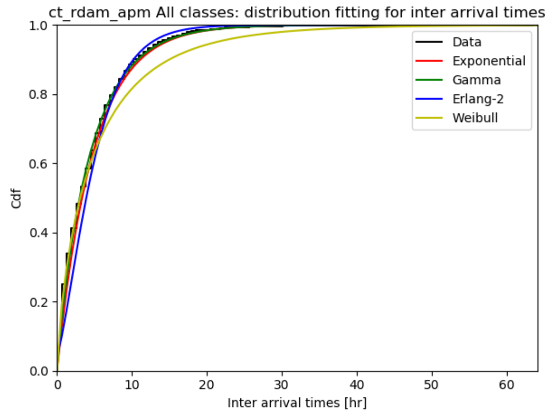


Figure G.2: inter arrival time distribution: Rotterdam APM Terminal

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	4.83	1.00	0.04	0.00	No
Gamma	0.00	5.06	0.90	0.05	0.00	No
Erlang-2	-0.86	2.85	2.00	0.09	0.00	No
Weibull	0.00	5.03	0.76	0.07	0.00	No

Table G.1: Inter arrival time distribution fitting for container terminal: Rotterdam APM

Rotterdam Euromax Terminal

The Rotterdam Euromax terminal again shows a very similar distribution shape (figure G.3) as to the two previous terminals. Visually almost all of the distributions seem to fit the data (figure G.4), however none of them pass the K-S test limit (table G.2).

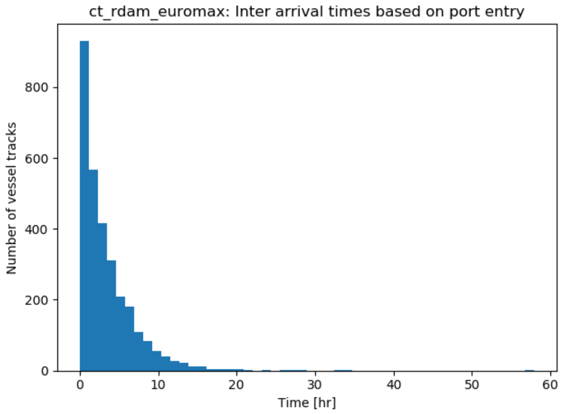


Figure G.3: Rotterdam Euromax inter arrival times (histogram)

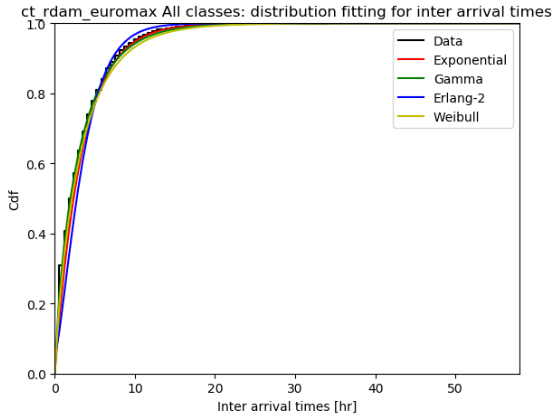


Figure G.4: Rotterdam Euromax Terminal with fitted distributions

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	3.39	1.00	0.05	0.00	No
Gamma	0.00	4.21	0.76	0.06	0.00	No
Erlang-2	-0.62	2.01	2.00	0.09	0.00	No
Weibull	0.00	3.25	0.88	0.04	0.00	No

Table G.2: Inter arrival time distribution fitting for container terminal: Rotterdam Euromax

France Le Havre Atlantic Terminal

The inter arrival time distribution of the Le Havre Atlantic Terminal in France is slightly different. The shape of the histogram is similar to the other terminals, but slightly less smooth (figure G.5). This could be due to the relatively smaller amount of vessel arrivals (383 arrivals) compared to the other terminals (1816, 2108 and 3001 arrivals). The *CDF* of the distribution does seem to nicely fit the theoretical **Exponential or Gamma** distributions. Again, no fit is found based on the K-S test for any of the theoretical distributions (table G.3).

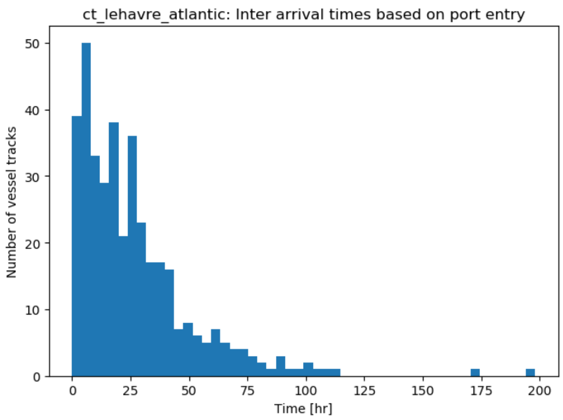


Figure G.5: Le Havre Atlantic inter arrival times (histogram)

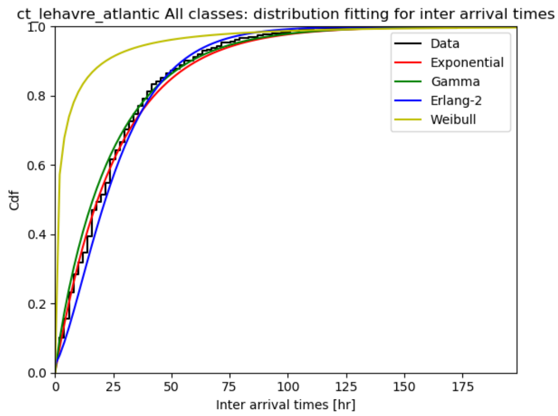


Figure G.6: Le Havre Atlantic with fitted distributions

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	26.44	1.00	0.07	0.08	Yes
Gamma	0.00	27.58	0.89	0.11	0.00	No
Erlang-2	-3.34	14.89	2.00	0.07	0.06	Yes
Weibull	0.00	2.96	0.42	0.59	0.00	No

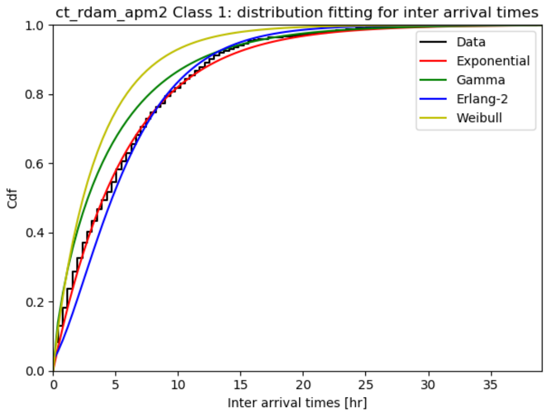
Table G.3: Inter arrival time distribution fitting for container terminal: Le Havre Atlantic

G.1.2. Container terminals: specific vessel classes

Container Class 1: Small feeders

For the Rotterdam APM-2 the only theoretical fit based on the K-S test is the **Exponential** distribution (table G.4). This corresponds with the visualisation of the CDF in figure G.7. For the Rotterdam APM terminal the **Exponential** distribution is chosen as best fit, based on the K-S test results (table G.5) and visual interpretation (figure G.8).

The Rotterdam Euromax Terminal also only fits the **Exponential** distribution for the inter arrival time, based on the K-S test limits (table G.6). This corresponds with the visual results, where the Exponential distribution is also the best fit to the data (figure G.9). Finally, the Le Havre Atlantic terminal has three possible theoretical distributions that fit the data. All the chosen distributions except the Weibull distribution fit the data. The best fit based on the K-S D statistic and the visual results both **Exponential and Erlang-2** distributions fit the data well (table G.7 and figure G.10).



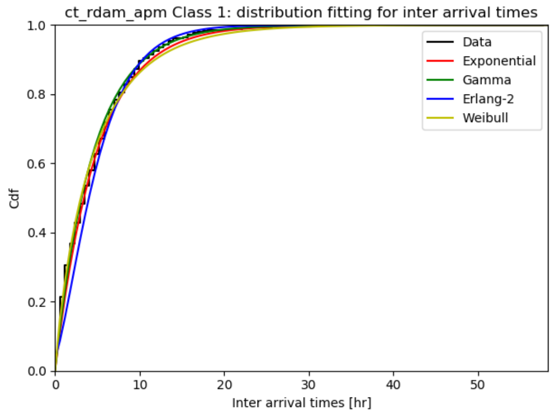


Figure G.7: Inter arrival time CDF Class 1 Rotterdam APM-2 TerminalFigure G.8: Inter arrival time CDF Class 1 Rotterdam APM Terminal

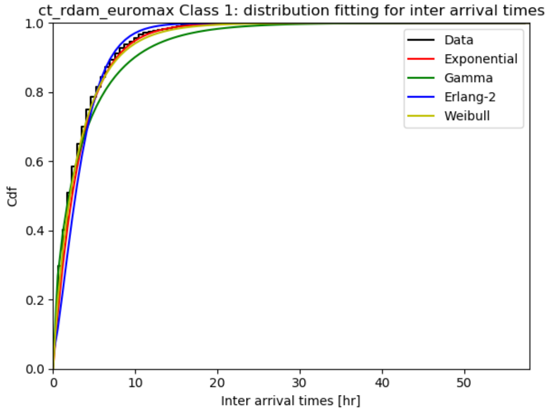


Figure G.9: Inter arrival time CDF Class 1 Rotterdam Euromax

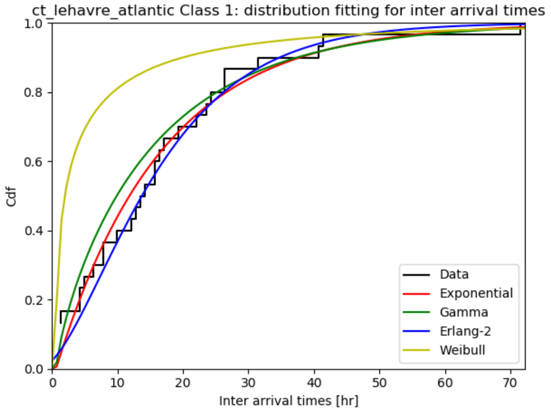


Figure G.10: Inter arrival time CDF Class 1 Le Havre Atlantic Terminal

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	5.81	1.00	0.04	0.07	
Gamma	0.00	6.44	0.73	0.14	0.00	
Erlang-2	-0.87	3.34	2.00	0.08	0.00	
Weibull	0.00	3.55	0.94	0.21	0.00	

Table G.4: Inter arrival time distribution fitting for Class 1 container terminal: Rotterdam APM-2

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	4.94	1.00	0.04	0.08	Yes
Gamma	0.00	4.65	0.98	0.07	0.00	No
Erlang-2	-0.82	2.88	2.00	0.08	0.00	No
Weibull	0.00	4.61	0.88	0.06	0.00	No

Table G.5: Inter arrival time distribution fitting for Class 1 container terminal: Rotterdam APM

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	3.36	1.00	0.03	0.15	Yes
Gamma	0.00	6.54	0.58	0.10	0.00	No
Erlang-2	-0.55	1.95	2.00	0.09	0.00	No
Weibull	0.00	3.14	0.90	0.04	0.01	No

Table G.6: Inter arrival time distribution fitting for Class 1 container terminal: Rotterdam Euromax

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.63	16.16	1.00	0.11	0.85	Yes
Gamma	0.63	19.33	0.78	0.17	0.34	Yes
Erlang-2	-2.21	9.49	2.00	0.11	0.89	Yes
Weibull	0.63	3.01	0.45	0.50	0.00	No

Table G.7: Inter arrival time distribution fitting for Class 1 container terminal: Le Havre Atlantic

Container Class 2: Regional feeders

The second container class is represented by the first three terminals. The Rotterdam APM-2 Terminal has no theoretical fits based on the K-S test (table G.8). However, visually the exponential distribution seems to fit the data very well (figure G.11). For the Rotterdam APM terminal both the **Gamma and Exponential** distributions pass the K-S test limit (table G.9). Based on the visualisation of these tests and the CDF of the data, both distributions seem to fit the data very well (figure G.12).

The Rotterdam Euromax terminal has no theoretical fits based on the K-S test (table G.10), but visually does seem to fit well by the Exponential distribution (figure G.13).

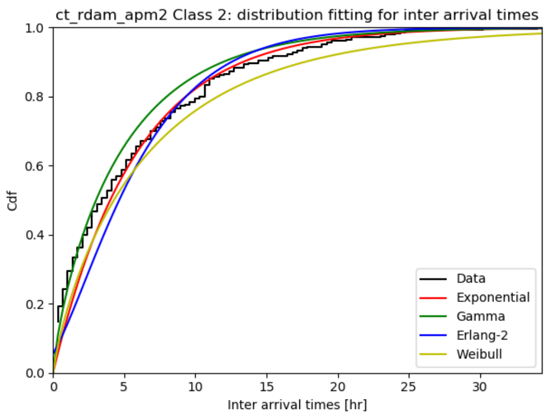


Figure G.11: Inter arrival time CDF Class 2 Rotterdam APM-2

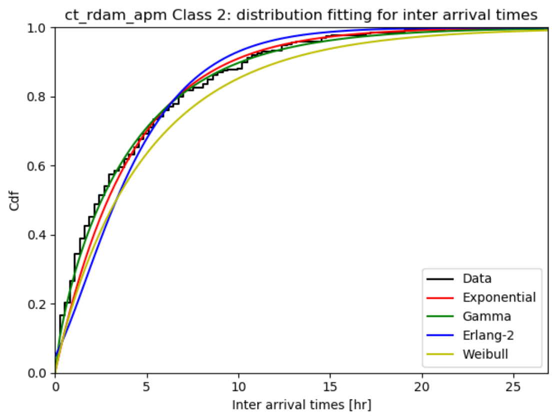


Figure G.12: Inter arrival time CDF Class 2 Rotterdam APM Terminal

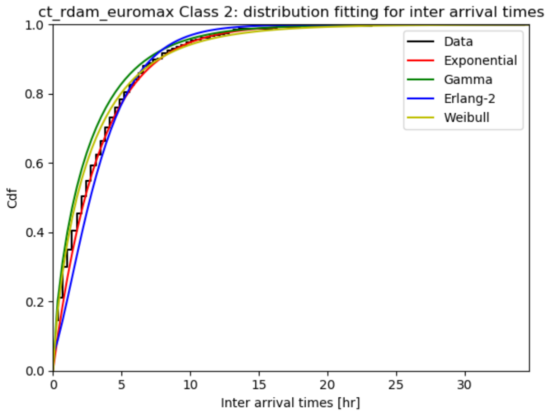


Figure G.13: Inter arrival time CDF Class 2 Rotterdam Euromax Terminal

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	5.80	1.00	0.09	0.01	No
Gamma	0.00	6.26	0.78	0.08	0.04	No
Erlang-2	-1.34	3.57	2.00	0.13	0.00	No
Weibull	0.00	6.60	0.85	0.09	0.03	No

Table G.8: Inter arrival time distribution fitting for Class 2 container terminal: Rotterdam APM-2

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	4.15	1.00	0.08	0.09	Yes
Gamma	0.00	5.34	0.77	0.06	0.24	Yes
Erlang-2	-0.86	2.51	2.00	0.14	0.00	No
Weibull	0.00	4.96	0.93	0.10	0.01	No

Table G.9: Inter arrival time distribution fitting for Class 2 container terminal: Rotterdam APM

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	3.38	1.00	0.06	0.00	No
Gamma	0.00	3.96	0.69	0.13	0.00	No
Erlang-2	-0.64	2.01	2.00	0.10	0.00	No
Weibull	0.00	2.74	0.81	0.10	0.00	No

Table G.10: Inter arrival time distribution fitting for Class 2 container terminal: Rotterdam Euromax

Container Class 3: Feedermax & Panamax

The third container class is only analysed for the APM-2 and APM terminal. The Rotterdam APM-2 terminal fits all four possible distributions based on the K-S test. Visually the fit is less suitable. The best fit is chosen to be the **Exponential** distribution based on results from table G.11 and figure G.15. The Rotterdam APM terminal fits three of four distributions based on the K-S test: the Exponential, Gamma and Erlang-2 distributions. Visually the best fits are chosen to be the **Gamma and Exponential** distributions, which corresponds to the D-statistic results (table G.12, figure G.14).

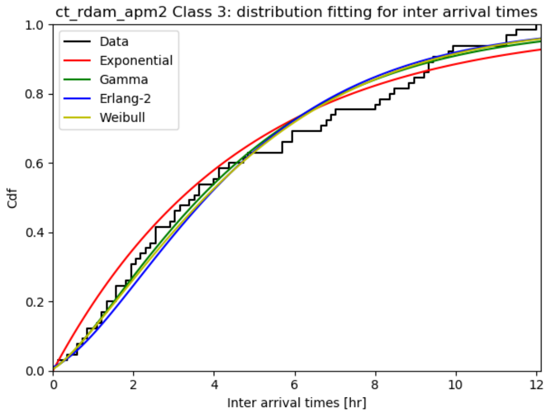


Figure G.14: Inter arrival time CDF Class 3 Rotterdam APM-2

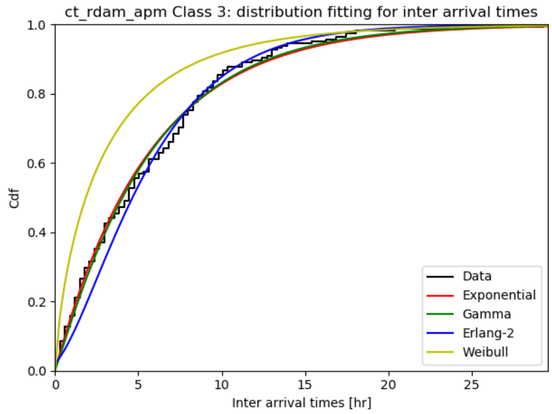


Figure G.15: Inter arrival time CDF Class 3 Rotterdam APM Terminal

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	4.61	1.00	0.10	0.49	Yes
Gamma	-0.15	3.01	1.59	0.09	0.66	Yes
Erlang-2	-0.37	2.49	2.00	0.10	0.55	Yes
Weibull	-0.09	5.10	1.33	0.09	0.64	Yes

Table G.11: Inter arrival time distribution fitting for Class 3 container terminal: Rotterdam APM-2

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	5.67	1.00	0.06	0.54	Yes
Gamma	0.00	5.25	1.08	0.06	0.59	Yes
Erlang-2	-0.68	3.17	2.00	0.09	0.12	Yes
Weibull	0.00	2.99	0.76	0.27	0.00	No

Table G.12: Inter arrival time distribution fitting for Class 3 container terminal: Rotterdam APM

Container Class 4: New Panamax

For the fourth container class all four terminals can be analysed. Based on the results from table G.13 and figure G.16 the best fit on the Rotterdam APM-2 Terminal is the **Exponential or Gamma** distribution. For the Rotterdam APM terminal the best fit is either the **Exponential or Weibull** distribution, based on the K-S tests and visuals (table G.14 and figure G.17).

The Rotterdam Euromax Terminal is best fit by the **Exponential or Weibull** distribution as well, based on numerical and visual results (table G.15 and figure G.18). Finally, the Le Havre Atlantic terminal is best represented by the **Exponential or Gamma** distribution (table G.16 and figure G.19).

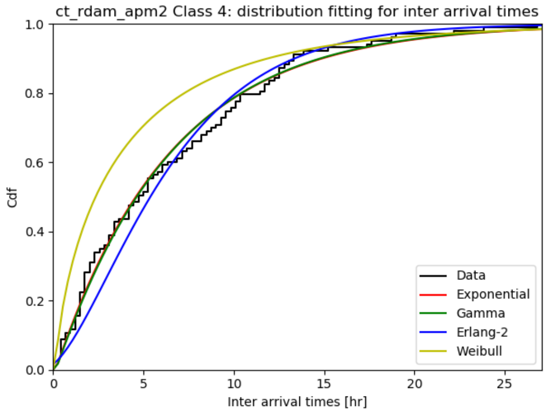


Figure G.16: Inter arrival time CDF Class 4 Rotterdam APM-2

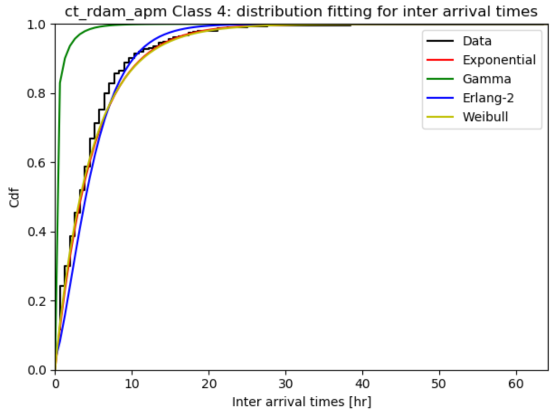


Figure G.17: Inter arrival time CDF Class 4 Rotterdam APM Terminal

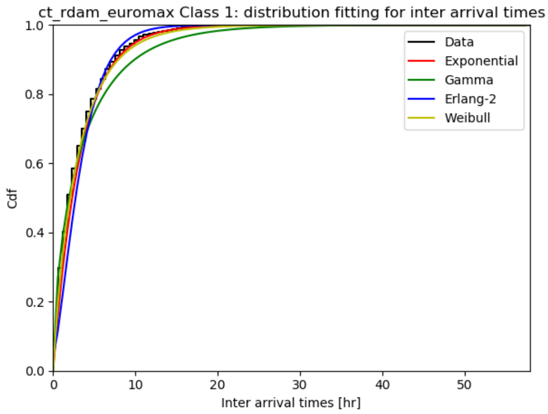


Figure G.18: Inter arrival time CDF Class 4 Rotterdam Euromax

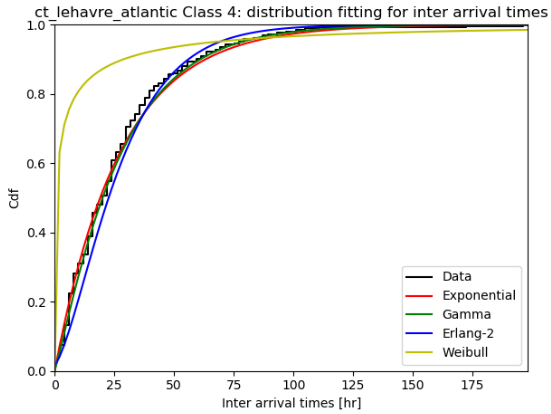


Figure G.19: Inter arrival time CDF Class 4 Le Havre Atlantic Terminal

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.15	6.40	1.00	0.06	0.78	Yes
Gamma	0.15	6.23	1.03	0.06	0.87	Yes
Erlang-2	-0.67	3.61	2.00	0.12	0.08	Yes
Weibull	0.15	3.67	0.72	0.26	0.00	No

Table G.13: Inter arrival time distribution fitting for Class 4 container terminal: Rotterdam APM-2

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.01	4.90	1.00	0.04	0.42	Yes
Gamma	0.01	2.69	0.16	0.72	0.00	No
Erlang-2	-0.68	2.79	2.00	0.09	0.00	No
Weibull	0.01	4.80	0.96	0.06	0.15	Yes

Table G.14: Inter arrival time distribution fitting for Class 4 container terminal: Rotterdam APM

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	3.74	1.00	0.07	0.60	Yes
Gamma	0.00	3.57	1.05	0.06	0.78	Yes
Erlang-2	-0.45	2.10	2.00	0.08	0.30	Yes
Weibull	0.00	3.35	0.97	0.10	0.10	Yes

Table G.15: Inter arrival time distribution fitting for Class 4 container terminal: Rotterdam Euromax

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	7.49	31.50	1.00	0.30	0.67	Yes
Gamma	-278.76	1.23	257.24	0.21	0.98	Yes
Erlang-2	1.62	18.68	2.00	0.24	0.94	Yes
Weibull	-1.77E+08	1.77E+08	1.13E+07	0.21	0.98	Yes

Table G.16: Inter arrival time distribution fitting for Class 4 container terminal: Le Havre Atlantic

Container Class 5: Post New Panamax

The largest vessel class for the containers is analysed for the Rotterdam APM-2, APM and Euromax terminals. The Rotterdam APM-2 terminal is best fit on the **Weibull or Exponential** distributions, however the Erlang-2 distribution is also a fit based on the K-S test (table G.17). For the Rotterdam APM terminal also has three possible fits based on the K-S test: the Exponential, Gamma or Erlang-2 distributions (table G.18). Although the best fits visually are the **Exponential or Gamma** distributions (figure G.21). The Rotterdam Euromax terminal again has three possible fits based on the K-S test, again a different combination: the Exponential, Gamma or Weibull distributon (table G.19). The chosen best fits are the **Exponential or Weibull** distributions (figure G.22).

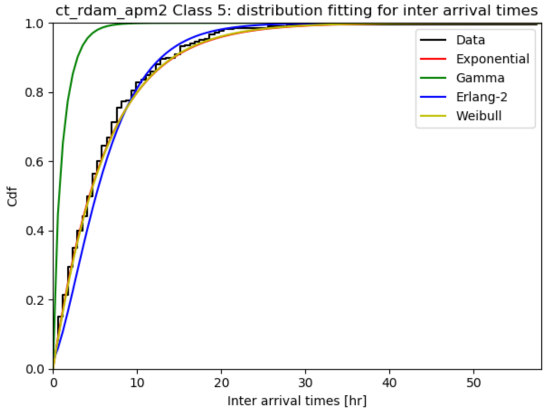


Figure G.20: Inter arrival time CDF Class 5 Rotterdam APM-2

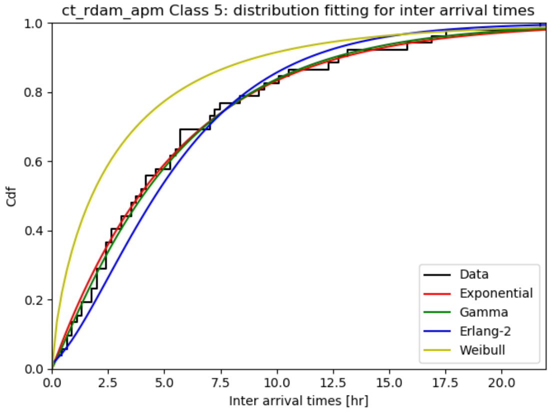


Figure G.21: Inter arrival time CDF Class 5 Rotterdam APM Terminal

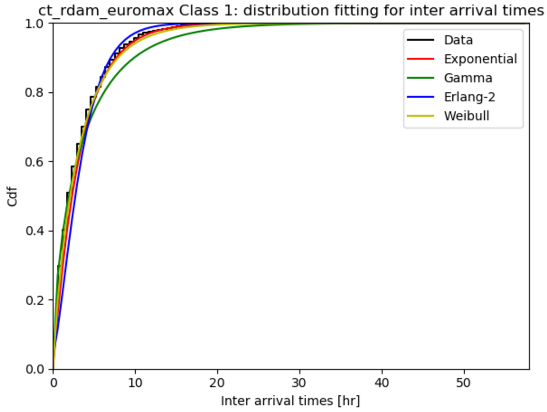


Figure G.22: Inter arrival time CDF Class 5 Rotterdam Euromax Terminal

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.01	6.22	1.00	0.04	0.83	Yes
Gamma	0.01	1.59	0.73	0.57	0.00	No
Erlang-2	-0.78	3.51	2.00	0.08	0.12	Yes
Weibull	0.01	6.26	1.02	0.04	0.91	Yes

Table G.17: Inter arrival time distribution fitting for Class 5 container terminal: Rotterdam APM-2

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.01	5.58	1.00	0.08	0.85	Yes
Gamma	-0.01	4.99	1.12	0.06	0.99	Yes
Erlang-2	-0.58	3.09	2.00	0.10	0.64	Yes
Weibull	0.01	2.91	0.73	0.32	0.00	No

Table G.18: Inter arrival time distribution fitting for Class 5 container terminal: Rotterdam APM

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	3.86	1.00	0.03	1.00	Yes
Gamma	0.00	4.13	0.95	0.03	1.00	Yes
Erlang-2	-0.56	2.21	2.00	0.11	0.01	No
Weibull	0.00	3.86	0.97	0.02	1.00	Yes

Table G.19: Inter arrival time distribution fitting for Class 5 container terminal: Rotterdam Euromax

G.2. Dry bulk terminals

G.2.1. Dry bulk terminals: total vessel mix

Rotterdam EMO Terminal

The Rotterdam EMO terminal inter arrival time distribution has one theoretical distribution that fits: the Weibull distribution (table G.24). However, the D statistic and p value for the K-S test are very similar for the Exponential distribution. Visually both the **Weibull and Exponential** distributions are good fits to the data (figure G.24).

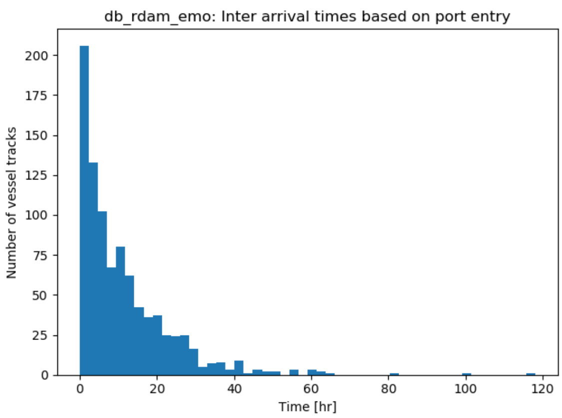


Figure G.23: Rotterdam EMO Terminal (histogram)

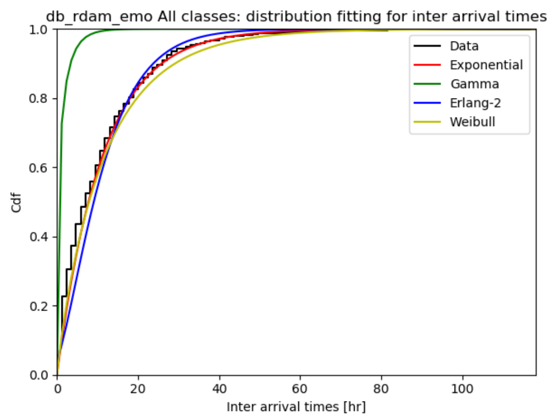


Figure G.24: Rotterdam EMO Terminal with fitted distributions

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	11.23	1.00	0.05	0.03	No
Gamma	0.00	3.51	0.33	0.62	0.00	No
Erlang-2	-2.05	6.64	2.00	0.11	0.00	No
Weibull	0.00	11.79	0.92	0.04	0.18	Yes

Table G.20: Inter arrival time distribution fitting for dry bulk terminal: Rotterdam EMO

Vlissingen OVET

The Vlissingen OVET terminal has a less smooth inter arrival time distribution (figure G.25) compared to the Rotterdam EMO inter arrival time distribution (figure G.23). This is most likely due to the smaller amount of total vessel arrivals of the Vlissingen terminal (127 versus 907 arrivals). The CDF of the Vlissingen inter arrival times is fit by the theoretical **Gamma** distribution (table G.21), which visually best represents the data as well (figure G.26).

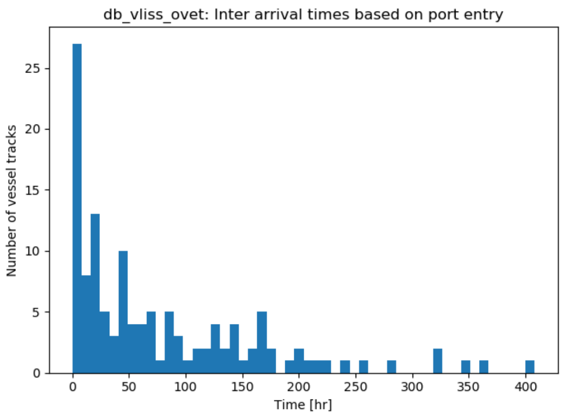


Figure G.25: Vlissingen OVET Terminal (histogram)

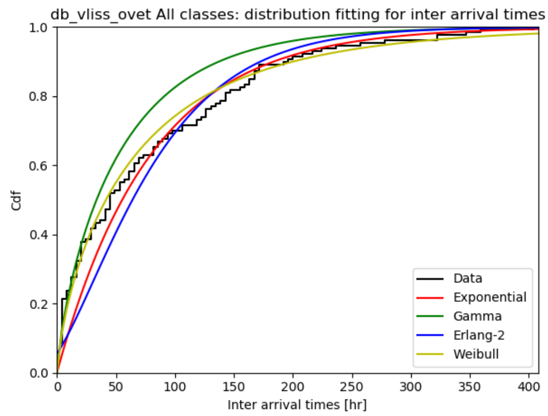


Figure G.26: Vlissingen OVET Terminal with fitted distributions

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	79.83	1.00	0.12	0.05	No
Gamma	0.00	74.41	0.75	0.15	0.00	No
Erlang-2	-18.38	49.10	2.00	0.16	0.00	No
Weibull	0.00	67.11	0.76	0.07	0.49	Yes

Table G.21: Inter arrival time distribution fitting for dry bulk terminal: Vlissingen OVET

Rotterdam EECV

The Rotterdam EECV has a much smoother inter arrival time distribution curve, compared to the Vlissingen terminal. Again, this is assumed to be due to the higher number of vessel arrivals (514 arrivals). The **Exponential** distribution is the only theoretical distribution that passes the K-S test (table G.22). The Gamma distribution visually is very similar (figure G.28).

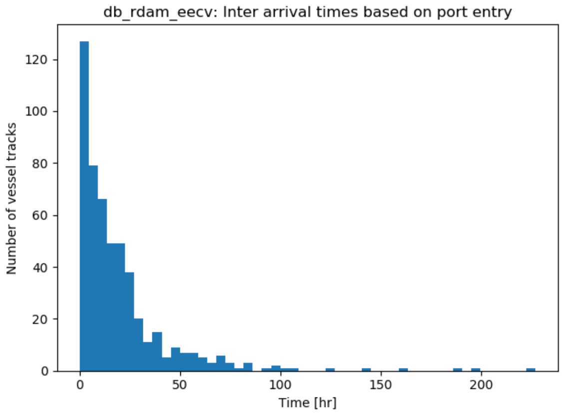


Figure G.27: Rotterdam EECV Terminal (histogram)

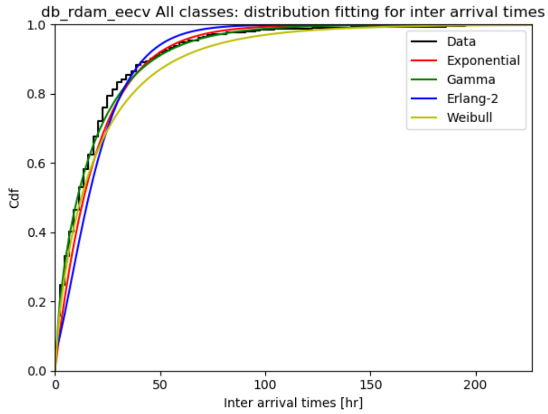


Figure G.28: Rotterdam EECV Terminal with fitted distributions

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	19.80	1.00	0.05	0.10	Yes
Gamma	0.00	27.42	0.68	0.07	0.01	No
Erlang-2	-3.93	11.86	2.00	0.11	0.00	No
Weibull	0.00	20.06	0.78	0.08	0.01	No

Table G.22: Inter arrival time distribution fitting for dry bulk terminal: Rotterdam EECV

Dunkirk Western Bulk

The fourth terminal analysed is the Dunkirk Western Bulk terminal. This terminal is expected to have a less smooth curve for the inter arrival time distribution based on the smaller amount of vessel arrivals (94 arrivals). This hypothesis is confirmed by figure G.29. From table G.23 and figure G.30 the best fits to the data are selected to be the **Exponential or Gamma** distributions. However, these are not perfect fits.

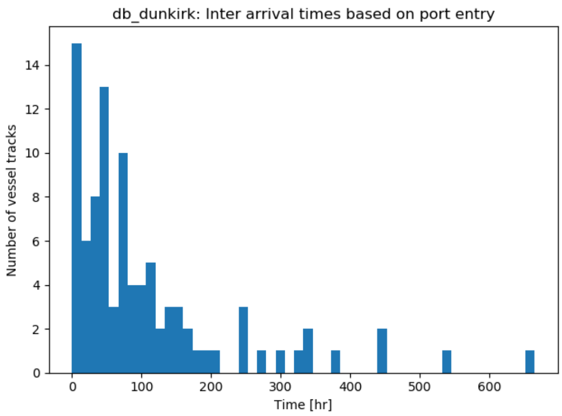


Figure G.29: Dunkirk Western Bulk Terminal (histogram)

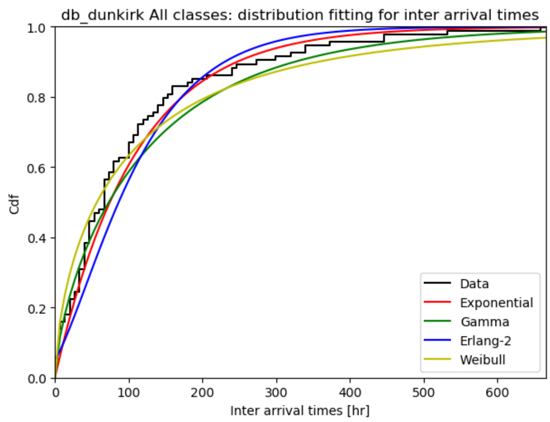


Figure G.30: Dunkirk Western Bulk Terminal with fitted distributions

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	107.38	1.00	0.12	0.13	Yes
Gamma	0.00	190.53	0.67	0.10	0.30	Yes
Erlang-2	-22.74	65.06	2.00	0.14	0.04	No
Weibull	0.00	99.72	0.65	0.16	0.02	No

Table G.23: Inter arrival time distribution fitting for dry bulk terminal: Dunkirk Western Bulk

G.2.2. Dry bulk terminals: specific vessel classes

Dry bulk Class 1: Small handy

The first dry bulk class is analysed for the Rotterdam EMO and Vlissingen OVET terminals. The Rotterdam EMO terminal has two possible fits based on the K-S test: the Exponential or the Weibull fit (table G.24). Both distributions are very similar and therefore the **Exponential and Weibull** distributions are both chosen as best fit (figure G.31). The Vlissingen OVET terminal only has one theoretical distribution that passes the K-S test: the **Gamma** distribution (table G.25). However, the number of arrivals for this class should be taken into account (38 vessels), leading to not a visually good fit (figure G.32).

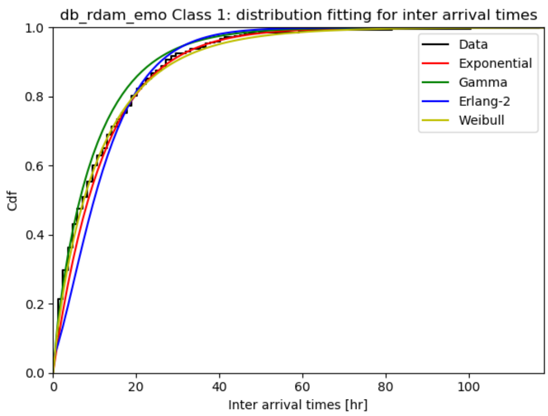


Figure G.31: Inter arrival time CDF Class 1 Rotterdam EMO Terminal

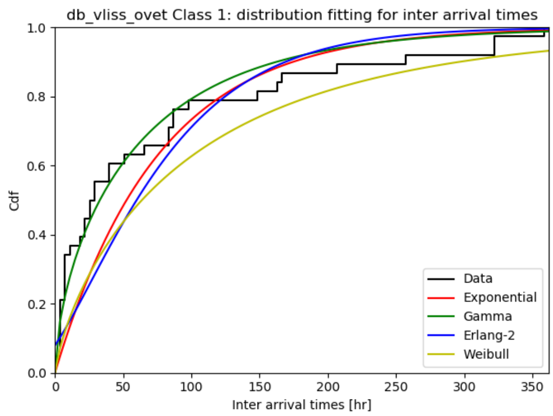


Figure G.32: Inter arrival time CDF Class 1 Vlissingen OVET Terminal

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	12.18	1.00	0.06	0.09	Yes
Gamma	0.00	11.98	0.84	0.08	0.01	No
Erlang-2	-2.20	7.19	2.00	0.13	0.00	No
Weibull	0.00	11.23	0.87	0.03	0.81	Yes

Table G.24: Inter arrival time distribution fitting for Class 1 dry bulk terminal: Rotterdam EMO

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	75.70	1.00	0.24	0.02	No
Gamma	0.00	103.72	0.60	0.10	0.83	Yes
Erlang-2	-22.85	49.27	2.00	0.25	0.01	No
Weibull	0.00	101.83	0.78	0.23	0.03	No

Table G.25: Inter arrival time distribution fitting for Class 1 dry bulk terminal: Vlissingen OVET

Dry bulk Class 2: Handy + Handymax + Supramax

For the second class three terminals are analysed. Nonetheless, the Vlissingen OVET terminals has a relatively low amount of data available (30 arrivals). For the Vlissingen OVET terminal the best fit is the **Exponential** distribution (table G.26 and figure G.33). The Rotterdam EECV terminal is visually and numerically best fit by the **Exponential or Gamma** distribution (table G.27 and figure G.34). Finally, the Dunkirk Western Bulk terminal is best fitted by either the **Erlang-2 or Exponential** distributions (table G.28 and figure G.35).

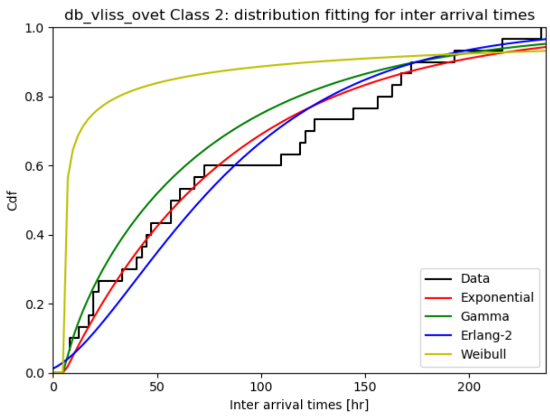


Figure G.33: Inter arrival time CDF Class 2 Vlissingen OVET Terminal

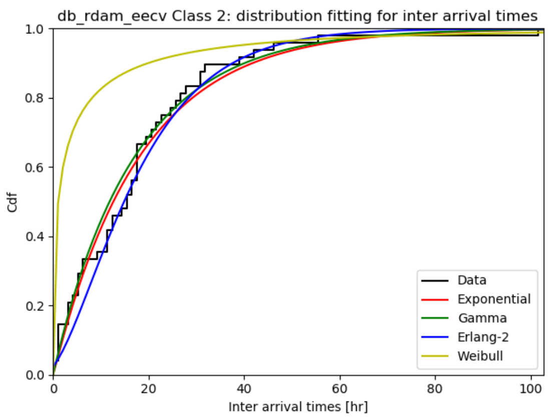


Figure G.34: Inter arrival time CDF Class 2 Rotterdam EECV Terminal

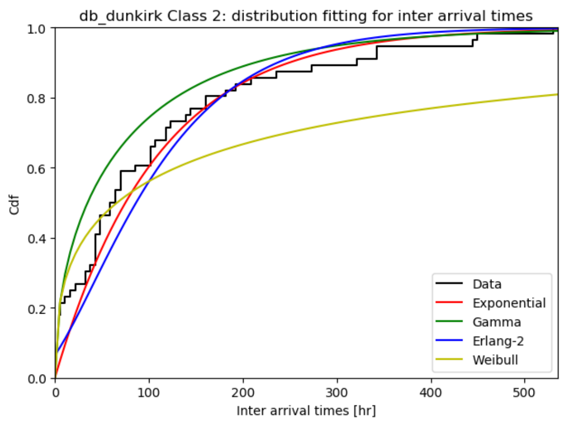


Figure G.35: Inter arrival time CDF Class 2 Dunkirk Western Bulk Terminal

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	5.64	80.54	1.00	0.13	0.71	Yes
Gamma	5.64	90.75	0.75	0.18	0.26	Yes
Erlang-2	-7.68	46.93	2.00	0.12	0.75	Yes
Weibull	5.64	3.40	0.23	0.61	0.00	No

Table G.26: Inter arrival time distribution fitting for Class 2 dry bulk terminal: Vlissingen OVET

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.07	18.11	1.00	0.11	0.53	Yes
Gamma	0.07	17.87	0.96	0.14	0.31	Yes
Erlang-2	-2.37	10.28	2.00	0.12	0.49	Yes
Weibull	0.07	2.50	0.40	0.53	0.00	No

Table G.27: Inter arrival time distribution fitting for Class 2 dry bulk terminal: Rotterdam EECV

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	108.27	1.00	0.17	0.06	Yes
Gamma	0.00	160.90	0.49	0.24	0.00	No
Erlang-2	-27.92	68.10	2.00	0.16	0.09	Yes
Weibull	0.00	159.40	0.42	0.20	0.02	No

Table G.28: Inter arrival time distribution fitting for Class 2 dry bulk terminal: Dunkirk Western Bulk

Dry bulk Class 3: Panamax

For the third vessel class category only the Rotterdam EMO and Rotterdam EECV terminals are analysed. The Rotterdam EMO terminal clearly has multiple fits (figure G.36). The best fit is the **Exponential** distribution (table G.29). For the Rotterdam EECV either the **Weibull or Exponential** distribution fit the data well (table G.30, figure G.37).

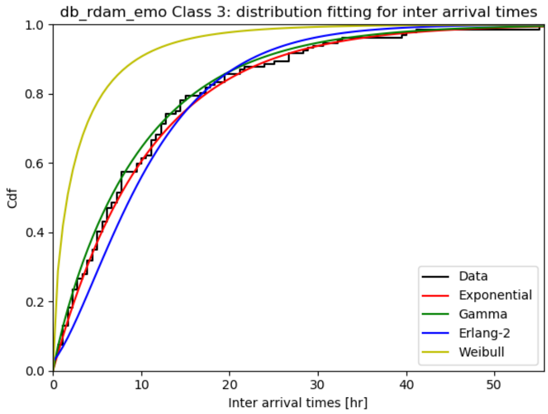


Figure G.36: Inter arrival time CDF Class 3 Rotterdam EMO Terminal

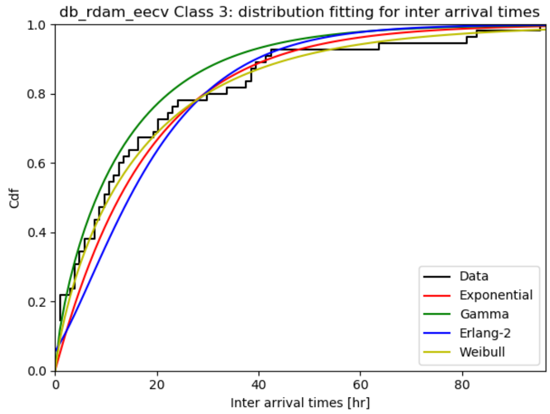


Figure G.37: Inter arrival time CDF Class 3 Rotterdam EECV Terminal

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	10.78	1.00	0.04	0.99	Yes
Gamma	0.00	10.90	0.90	0.08	0.39	Yes
Erlang-2	-1.51	6.15	2.00	0.10	0.12	Yes
Weibull	0.00	2.78	0.68	0.44	0.00	No

Table G.29: Inter arrival time distribution fitting for Class 3 dry bulk terminal: Rotterdam EMO

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.02	18.23	1.00	0.12	0.39	Yes
Gamma	0.02	18.38	0.74	0.11	0.49	Yes
Erlang-2	-4.25	11.25	2.00	0.15	0.14	Yes
Weibull	0.02	16.67	0.82	0.07	0.96	Yes

Table G.30: Inter arrival time distribution fitting for Class 3 dry bulk terminal: Rotterdam EECV

Dry bulk Class 4: Mini Capesize + Capesize

The same two terminals are analysed for the fourth class. The Rotterdam EMO terminal seems to be representable by all four theoretical distributions (table1 G.31). The best fits are either the **Exponential or Erlang-2** distributions. For the Rotterdam EECV terminal the best fit numerically is the **Gamma** distribution (table G.32). However, visually none of the distributions perfectly fit the data (figure G.39).

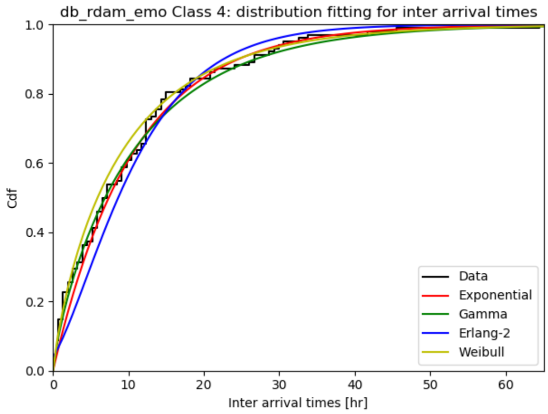


Figure G.38: Inter arrival time CDF Class 4 Rotterdam EMO Terminal

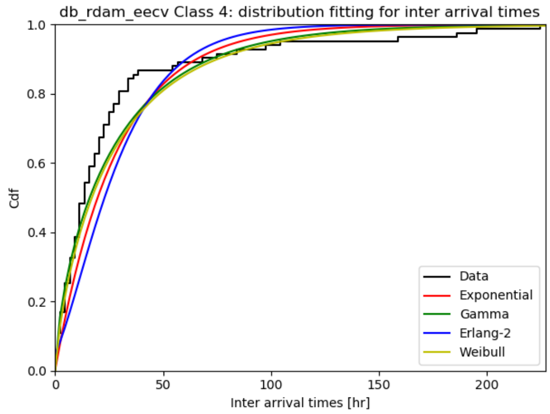


Figure G.39: Inter arrival time CDF Class 4 Rotterdam EECV Terminal

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	10.67	1.00	0.06	0.79	Yes
Gamma	0.00	13.83	0.80	0.05	0.92	Yes
Erlang-2	-1.94	6.31	2.00	0.11	0.17	Yes
Weibull	0.00	9.11	0.85	0.10	0.26	Yes

Table G.31: Inter arrival time distribution fitting for Class 4 dry bulk terminal: Rotterdam EMO

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	28.46	1.00	0.14	0.06	Yes
Gamma	0.00	42.67	0.67	0.12	0.17	Yes
Erlang-2	-5.86	17.16	2.00	0.21	0.00	No
Weibull	0.00	26.83	0.81	0.13	0.11	Yes

Table G.32: Inter arrival time distribution fitting for Class 4 dry bulk terminal: Rotterdam EECV

Dry bulk Class 5: Very Large Bulk Carrier + Very large Ore Carrier

As mentioned earlier in the service time distribution analysis, only one terminal receives enough vessels from this class: the Rotterdam EECV terminal. The terminal can fit three possible distributions based on the K-S test (table G.33). The best fit is either the **Exponential or the Gamma** distribution (figure G.40).

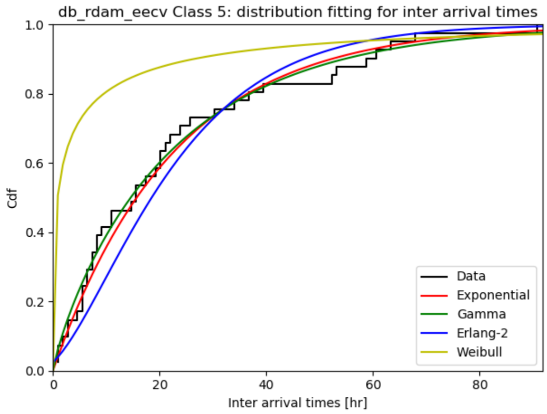


Figure G.40: Inter arrival time CDF Class 5 Rotterdam EECV Terminal

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	22.65	1.00	0.07	0.98	Yes
Gamma	0.00	26.98	0.84	0.10	0.81	Yes
Erlang-2	-3.07	12.86	2.00	0.16	0.22	Yes
Weibull	0.00	2.47	0.35	0.58	0.00	No

Table G.33: Inter arrival time distribution fitting for Class 5 dry bulk terminal: Rotterdam EECV

G.3. Liquid bulk terminals

G.3.1. Liquid bulk terminals: total vessel mix

Rotterdam GATE Terminal

The Rotterdam GATE terminal is analysed first. In total 173 vessels arrive at the terminal in the chosen time span. The inter arrival time distribution is visualised in figure G.41. By plotting the CDF multiple distributions can be tested based on their goodness-of-fit towards the distribution (figure G.42). Based on the results for the K-S test three distributions fit the data: the Exponential, Gamma and Erlang-2 distributions. The theoretical distribution that best fits the data is the **Exponential** distribution, based on the most optimal D-statistic and visual interpretations (table G.34).

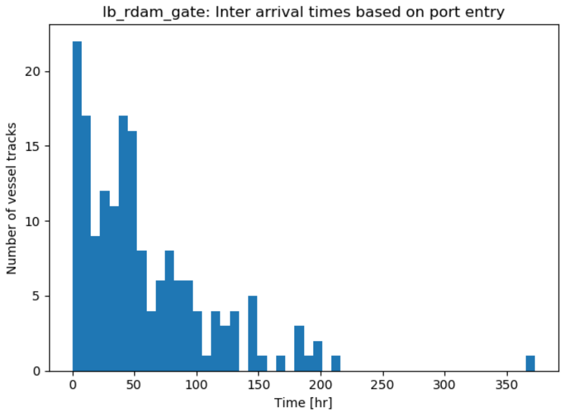


Figure G.41: Rotterdam GATE Terminal (histogram)

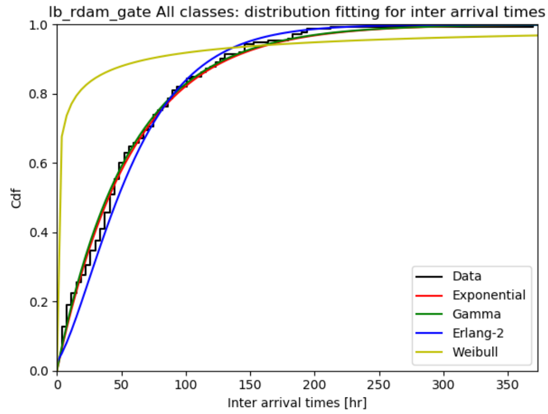


Figure G.42: Rotterdam GATE Terminal with fitted distributions

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	57.94	1.00	0.07	0.40	Yes
Gamma	0.00	58.01	0.97	0.08	0.21	Yes
Erlang-2	-7.91	32.92	2.00	0.08	0.19	Yes
Weibull	0.00	2.30	0.24	0.63	0.00	No

Table G.34: Inter arrival time distribution fitting for liquid bulk terminal: Rotterdam GATE

Zeebrugge LNG Terminal

Next, the Zeebrugge LNG terminal is analysed. The inter arrival time distribution in histogram form is plotted in figure G.43. The CDF form with possible fitted distributions are plotted in figure G.44. Based on the K-S goodness-of-fit test (table G.35) three distributions are possible (pass the limit test): the Exponential, Gamma and Weibull distribution. Similar to the previous analysed terminal, the **Exponential** distribution fits the inter arrival time distribution the best based on the K-S test results and visual interpretations.

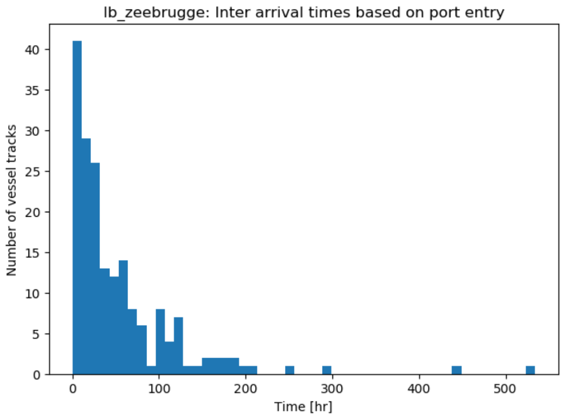


Figure G.43: Zeebrugge LNG Terminal (histogram)

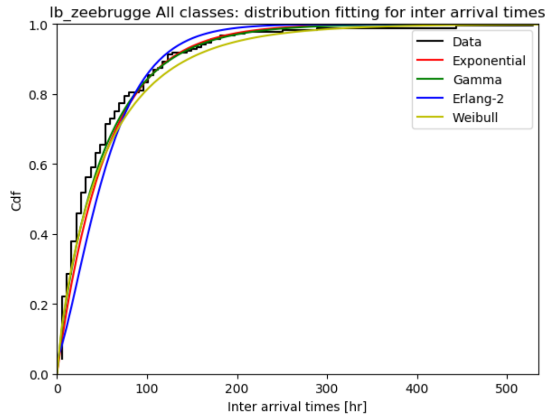


Figure G.44: Zeebrugge LNG Terminal with fitted distributions

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	54.49	1.00	0.08	0.16	Yes
Gamma	0.00	61.94	0.86	0.08	0.15	Yes
Erlang-2	-9.51	32.00	2.00	0.16	0.00	No
Weibull	0.00	54.29	0.84	0.09	0.09	Yes

Table G.35: Inter arrival time distribution fitting for liquid bulk terminal: Zeebrugge

Dunkirk LNG Terminal

The Dunkirk LNG terminal is the only terminal that only contains 1 berth. This is expected to lead to a higher average inter arrival time, due to simply less vessels arriving at the terminal. Results confirm that the highest inter arrival times are found for the Dunkirk terminal, as shown in table G.36.

Terminal	No. Berths [-]	Median [hr]	Mean [hr]
Rotterdam GATE	2	44.00	57.94
Zeebrugge LNG	2	31.42	54.49
Dunkirk LNG	1	114.16	147.54
France Montoir LNG	2	50.20	79.95

Table G.36: LNG terminals Inter arrival time distribution information

The same theoretical distributions are fit to the data, as shown in figure G.46. The only distribution that passes the K-S limit test is the Erlang-2 distribution (table G.37). Visually the **Erlang-2** distribution fits the data the best, however the fit is not perfect.

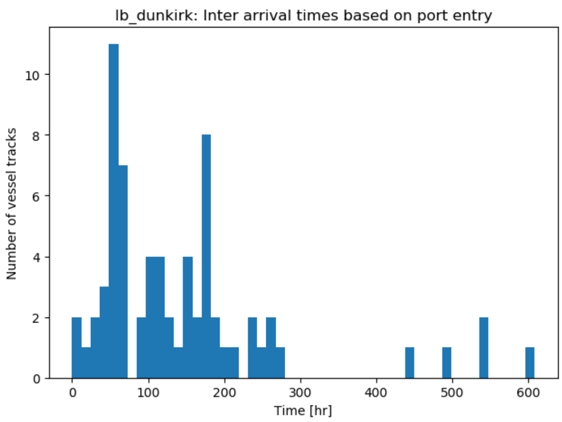


Figure G.45: Dunkirk LNG Terminal (histogram)

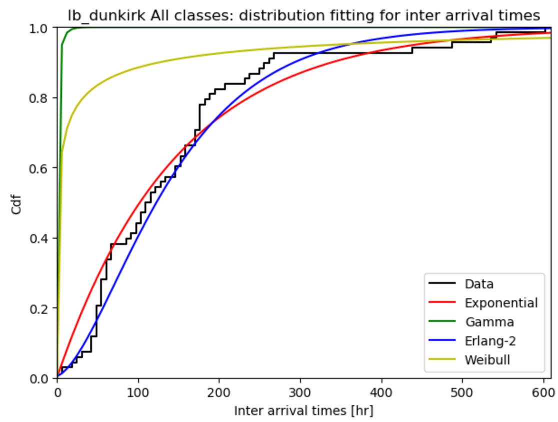


Figure G.46: Dunkirk LNG Terminal with fitted distributions

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	147.54	1.00	0.19	0.01	No
Gamma	0.00	8.09	0.14	0.97	0.00	No
Erlang-2	-6.64	77.09	2.00	0.11	0.41	Yes
Weibull	0.00	5.47	0.27	0.75	0.00	No

Table G.37: Inter arrival time distribution fitting for liquid bulk terminal: Dunkirk

France Montoir LNG Terminal

Finally, the LNG terminal in France Montoir is analysed. All four terminals fit based on the K-S test (table G.38). The best fit is found to be the **Exponential** distribution based on K-S test results and the visualisation as in figure G.48.

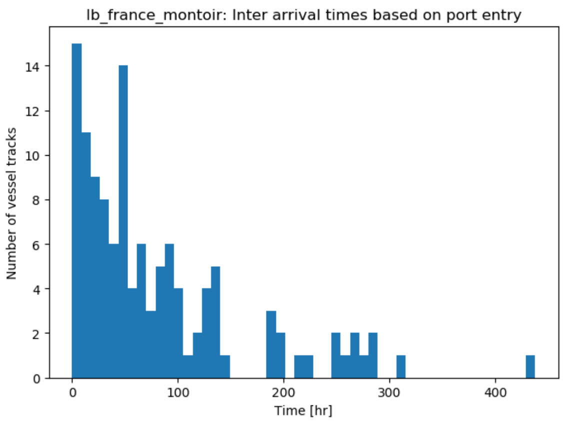


Figure G.47: France Montoir LNG Terminal (histogram)

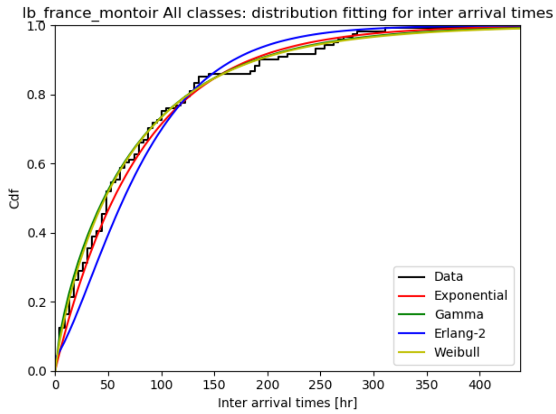


Figure G.48: France Montoir LNG Terminal with fitted distributions

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	79.95	1.00	0.05	0.95	Yes
Gamma	0.00	102.52	0.75	0.09	0.31	Yes
Erlang-2	-13.47	46.71	2.00	0.12	0.07	Yes
Weibull	0.00	72.29	0.86	0.08	0.45	Yes

Table G.38: Inter arrival time distribution fitting for liquid bulk terminal: France Montoir

G.3.2. Liquid bulk terminals: specific vessel classes

LNG Class 1: Small Spherical / Membrane LNG Carriers

The first LNG class is the class represented by small spherical or membrane LNG carriers. Two terminals research a sufficient amount of vessels from this class in order to be analysed. First, the Rotterdam GATE terminal inter arrival time distribution is fitted by either the Exponential, Erlang-2 or Weibull distributions (based on the K-S tests, see table G.39). The best fit is chosen as the **Exponential** distribution (figure G.49). For the second terminal, the Zeebrugge LNG terminal the best fit is either the **Exponential or Gamma** distribution, based on K-S tests and visual interpretations (table G.40) and figure G.50).

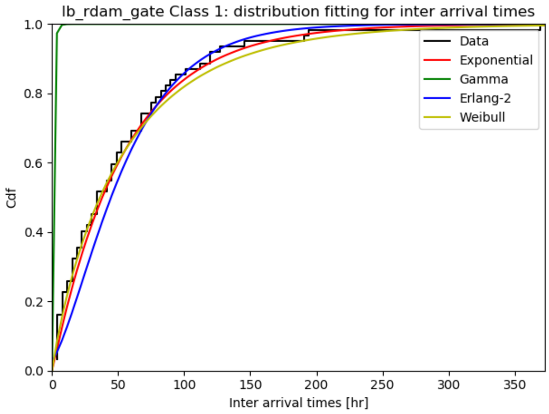


Figure G.49: Inter arrival time CDF Class 1 Rotterdam GATE Terminal

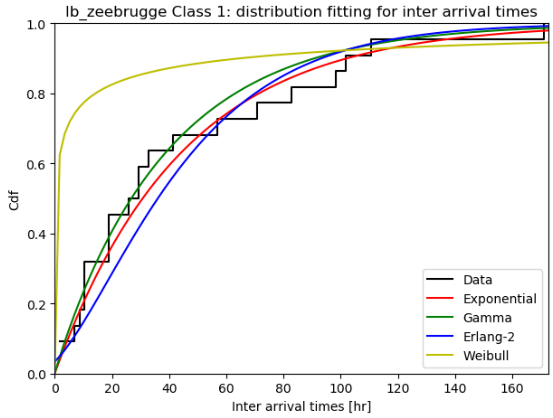


Figure G.50: Inter arrival time CDF Class 1 Zeebrugge LNG Terminal

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.32	54.43	1.00	0.05	0.99	Yes
Gamma	0.32	1.96	0.27	0.95	0.00	No
Erlang-2	-7.79	31.27	2.00	0.12	0.34	Yes
Weibull	0.32	54.35	0.89	0.07	0.95	Yes

Table G.39: Inter arrival time distribution fitting for dry bulk terminal: Rotterdam GATE

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	44.36	1.00	0.11	0.95	Yes
Gamma	0.00	40.33	0.97	0.11	0.95	Yes
Erlang-2	-7.35	25.85	2.00	0.17	0.48	Yes
Weibull	0.00	1.91	0.24	0.65	0.00	No

Table G.40: Inter arrival time distribution fitting for dry bulk terminal: Zeebrugge LNG

LNG Class 2: Medium Spherical / Membrane LNG Carriers

None of the four analysed terminals receive vessels from this second LNG vessel class.

LNG Class 3: Large Spherical / Membrane LNG Carriers

The third LNG class is represented by all four terminals. First, the Rotterdam GATE terminal is analysed. The best fitted theoretical distributions are the **Erlang-2 or Exponential** distributions (results in table G.41 and figure G.51). Next, the Zeebrugge LNG terminal is analysed. This terminals inter arrival time distribution is best fitted by the **Exponential or Weibull** distribution (results in table G.42 and figure G.52)

The Dunkirk LNG terminal obtains a less tight fit based on visual results of the inter arrival time CDF (figure G.53). Based on K-S test results the Erlang-2 distribution is a fit on the data set (table G.43). Based on the visual interpretation of the distributions the **Erlang-2 or Exponential** distribution both are selected as possible fits. Finally, the France Montoir LNG terminal is assessed. The best fit is found to be the **Exponential** distribution (results in table G.43 and figure G.54).

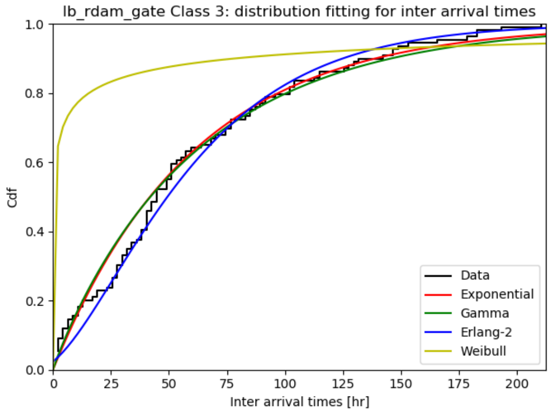


Figure G.51: Inter arrival time CDF Class 3 Rotterdam GATE Terminal

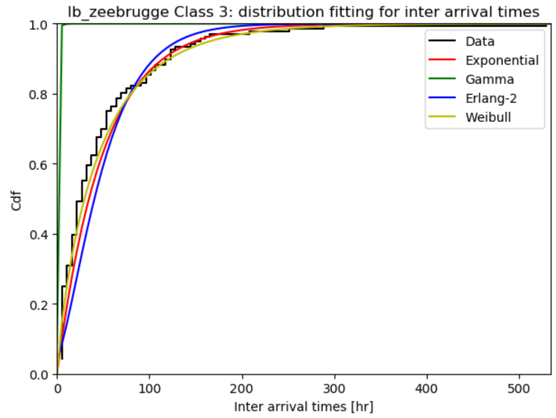


Figure G.52: Inter arrival time CDF Class 3 Zeebrugge LNG Terminal

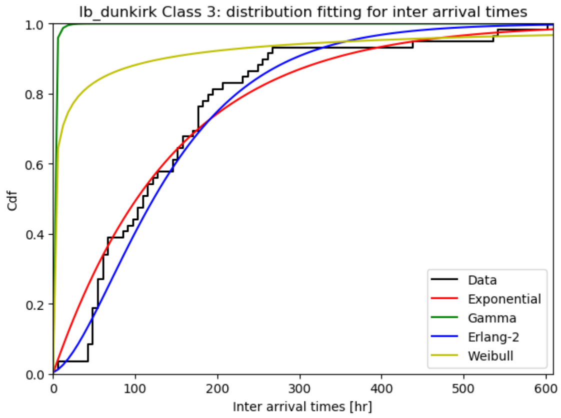


Figure G.53: Inter arrival time CDF Class 3 Dunkirk LNG

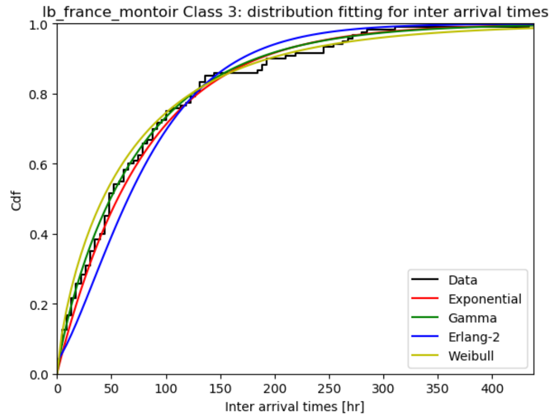


Figure G.54: Inter arrival time CDF Class 3 France Montoir LNG

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	60.30	1.00	0.11	0.13	Yes
Gamma	0.00	66.45	0.94	0.11	0.11	Yes
Erlang-2	-8.00	34.15	2.00	0.07	0.71	Yes
Weibull	0.00	1.80	0.22	0.63	0.00	No

Table G.41: Inter arrival time distribution fitting for dry bulk terminal: Rotterdam GATE

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	50.00	1.00	0.11	0.07	Yes
Gamma	0.00	3.78	0.04	0.96	0.00	No
Erlang-2	-8.77	29.38	2.00	0.16	0.00	No
Weibull	0.00	45.84	0.85	0.11	0.09	Yes

Table G.42: Inter arrival time distribution fitting for dry bulk terminal: Zeebrugge LNG

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	147.31	1.00	0.23	0.00	No
Gamma	0.00	8.18	0.11	0.97	0.00	No
Erlang-2	-6.43	76.87	2.00	0.11	0.41	Yes
Weibull	0.00	5.48	0.26	0.79	0.00	No

Table G.43: Inter arrival time distribution fitting for dry bulk terminal: Dunkirk LNG

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	0.00	80.45	1.00	0.05	0.96	Yes
Gamma	0.00	90.13	0.85	0.07	0.61	Yes
Erlang-2	-13.54	47.00	2.00	0.12	0.08	Yes
Weibull	0.00	67.22	0.79	0.12	0.07	Yes

Table G.44: Inter arrival time distribution fitting for dry bulk terminal: France Montoir LNG

LNG Class 4: Very large Spherical / Membrane LNG Carriers

For the fourth class only the Zeebrugge LNG terminal is inspected. Three out of four distributions fit the inter arrival time distribution: the Exponential, the Erlang-2 and the Weibull distribution (table G.45). Visually the fit is not perfect due to the less number of vessel arrivals (figure G.55). The best fit is chosen as the **Exponential** distribution.

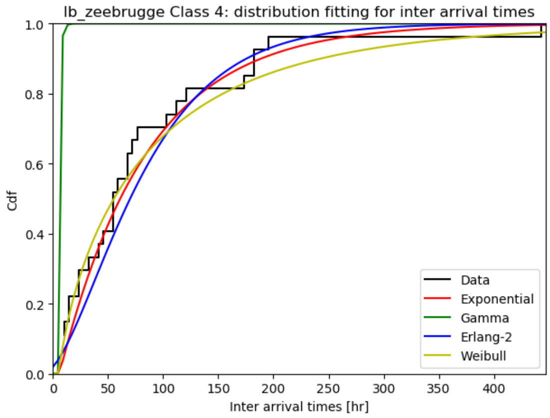


Figure G.55: Inter arrival time CDF Class 4 Zeebrugge LNG Terminal

	Distribution parameters			K-S test		
Distribution	Location	Scale	Shape	D	p	Lim
Exponential	6.02	79.37	1.00	0.10	0.95	Yes
Gamma	6.02	2.40	0.19	0.88	0.00	No
Erlang-2	-10.02	47.70	2.00	0.14	0.59	Yes
Weibull	6.02	79.67	0.77	0.10	0.93	Yes

Table G.45: Inter arrival time distribution fitting for dry bulk terminal: Zeebrugge LNG

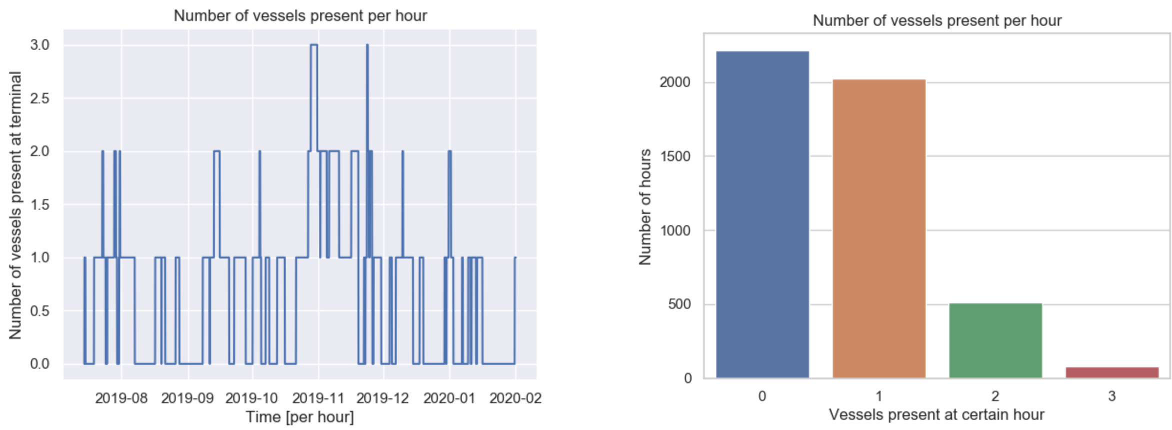
H

Appendix H: Finding the number and locations of berths

In order to determine the berth occupancy, the number of berths and their locations should be defined. Preferably these numbers and locations should be automatically generated using the available AIS data. In this Appendix multiple methods are discussed which try to find the number and location of the berth. Examples for the Vlissingen OVET dry bulk terminal are given.

H.0.1. Number of berths at terminal

In order to determine the number of berths a code is written where per hour, the number of vessels present is counted. First, the timestamps are split into an array starting at the first timestamp of the entire data set, continuing until the very last timestamp of the entire data set, with a frequency of 1 hour. Next, for every hour present in the array, the number of vessels present is counted. Finally, a new data frame is returned which includes every timestamp with the number of vessels present. This new data frame is plotted over time, as total vessels per hour and the percentage of total time is calculated. An example of the output for the Vlissingen Dry Bulk Terminal is shown in figures H.1 and H.2.



(a) Number of vessels present per hour

(b) Total number of hours per number of vessels present

Figure H.1: Vlissingen OVET Dry Bulk Terminal (number of vessels/ hour)

	timestamp	percentage_of_total_time
0	2216.00000	45.85000
1	2021.00000	41.82000
2	513.00000	10.61000
3	83.00000	1.72000

Figure H.2: Total percentage of time number of vessels is present

In this particular example the most logical assumption would be to assume this terminal has 2 berths. The 1.7% of the total number of hours where 3 vessels are present could be due to an error in the previous cleaning or pre-processing steps, or it might represent the moment that one vessels is leaving and one is arriving. Another reason could be that the terminal can handle two medium to large vessels at the same time, or three smaller vessels. From the data itself it is not possible to find the actual cause, it could be any one of these reasons mentioned.

H.0.2. Location of berths at terminal

To determine the location of the berths several approaches have been performed. First, all coordinates can been visualised using GoogleMapPlotter with the scatter function, also a heat map is used for the visualisation of all coordinates. For the example of the Dry Bulk Terminal in Vlissingen, the scatter map and the heat map are shown in Figure H.3. It is clear that not 2 or 3 berths can easily be defined on this data using this approach.

(a) Scatter plot (b) Heatmap

Figure H.3: Vlissingen OVET Dry Bulk Terminal (01-07-2019 – 01-02-2020)

In order to find the locations of the berths a clustering technique is used, which represents a form of unsupervised machine learning. Unsupervised machine learning focuses on unlabelled data, such as here where the location of the berths are unknown. The algorithm will try to find group similarity among a data set. Clustering uses distances and thus the shape of the cluster is only affected by the distance only. The K-Means algorithm tries to optimise the minimum distance between the data points and its cluster centroid. The number of K clusters must be selected before the model can be run (Maheswari, 2019).

Besides trying to locate the the centroid of the clusters of the total data, the K-Means algorithm can be implemented for every vessel track individually. For each track, the clustered centroid location is determined, using the K-means algorithm with a predefined number of clusters being equal to 1. Using only the coordinates of these centroids of each vessel will eliminate a lot of data points, possibly improving the scatter map and heat map. The scatter map and heat map using only the centroid locations per vessel track are visualised in Figure H.4. Again, no clear distinction can be made for the berth locations.

(a) Scatter plot

(b) Heatmap

Figure H.4: Vlissingen OVET Dry Bulk Terminal: Only centroid locations (01-07-2019 – 01-02-2020)

If the assumption is made that the number of berths is known, the K-Means clustering technique can be applied to find the centers of the clusters. For both the entire data set with coordinates, as well as for only the centers per vessel track, the centers are determined. In Figure H.5 the yellow circles represent the large data set, and the blue circles represent the center found for only the vessel centers.

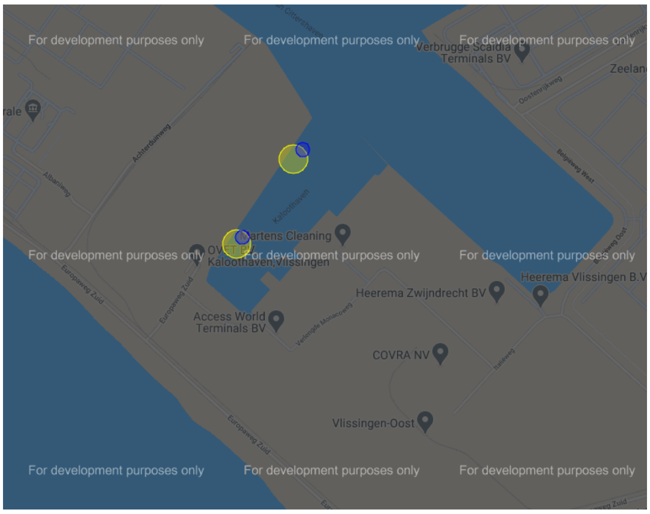
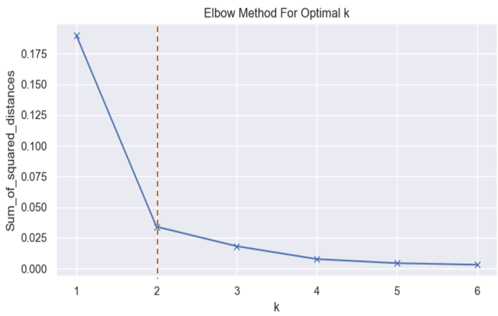


Figure H.5: K-Means cluster centers (for K = 2)

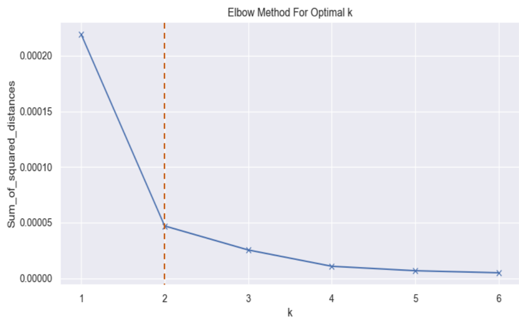
Another approach to determine the number of clusters at the terminal, can be done using validity models such as the elbow and silhouette statistic methods. The Elbow method returns a curve where the within sum of squares (a measure of the variability of points in the cluster) is plotted in contrast to the total number of clusters. If an extra cluster is chosen, the total within sum of squares will not increase. In the curve when an angle is formed between two number of clusters, the 'elbow' shape represents the optimal number of clusters. In Figure H.6 the Elbow Method is demonstrated for the example in Vlissingen, where an optimal number of clusters is found at 2 clusters. The within sum of squares can be calculated as follows:

$$WSS = \sum_{j=1}^{k} \sum_{i=1}^{k} ||x_i^{(j)} - c_j||^2 \tag{H.1}$$

Where k represents the number of clusters, n the number of objects, x_i is the i th element in the cluster and c_j is the center of the jth cluster (Gustriansyah et al., 2019; Aiyappa & Ramamurthy, 2018; Maheswari, 2019).



(a) All data points



(b) Only 1 coordinate per vessel track

Figure H.6: Elbow Method - Vlissingen

When the steps are repeated for Container Terminals, Liquid Bulk Terminals and other Dry Bulk Terminals the conclusion is made based on the heat maps that for Container Terminals and Dry Bulk Terminals it is impossible to distinguish, with confidence, the exact location of the berth, let alone define which vessels berthed at which berth. This is most likely due to the vessels not berthing at the same locations, the antenna on the vessel is not at the same location (relative to the vessel) and the vessels are in different sizes.

For the number of berths however, the three methods combined: Elbow method, number of vessels per hour plot and the total percentage values for every number of vessels, together give an idea for the maximum number of berths. Nonetheless, for container terminals this is not always the case, and it becomes tricky when the maximum number of berths varies over time.

I

Appendix I: Results for occupancies

I.1. Container terminals

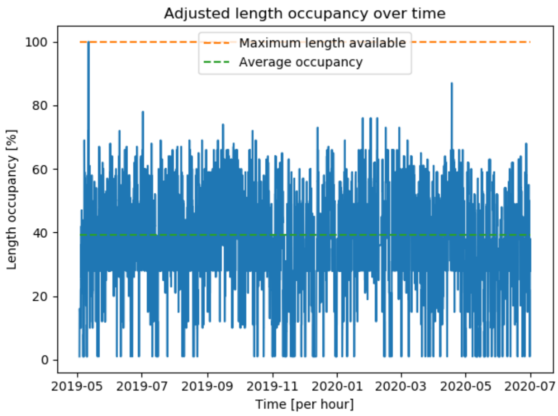


Figure I.1: Rotterdam APM-2: Adjusted length occupancy

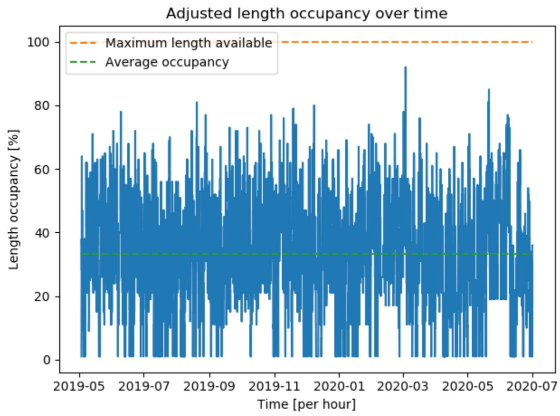


Figure I.2: Rotterdam APM: Adjusted length occupancy

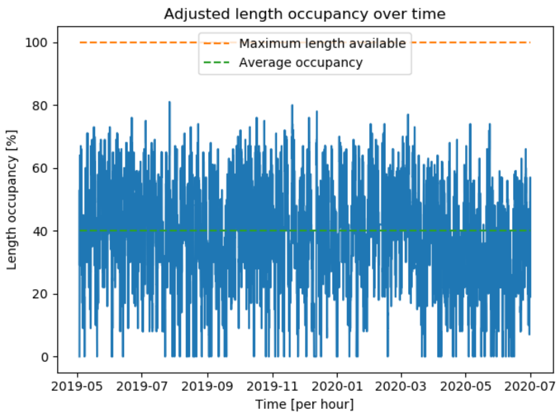


Figure I.3: Rotterdam Euromax: Adjusted length occupancy

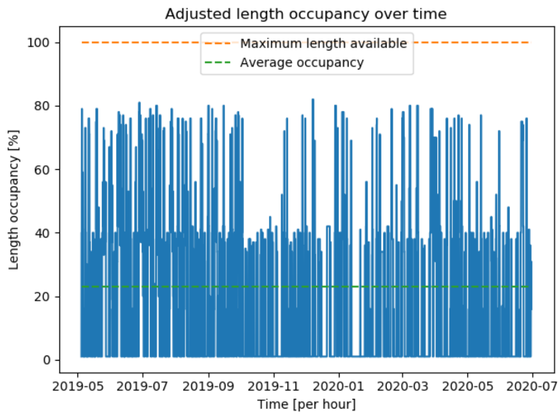


Figure I.4: Le Havre Atlantic: Adjusted length occupancy

I.2. Dry bulk terminals

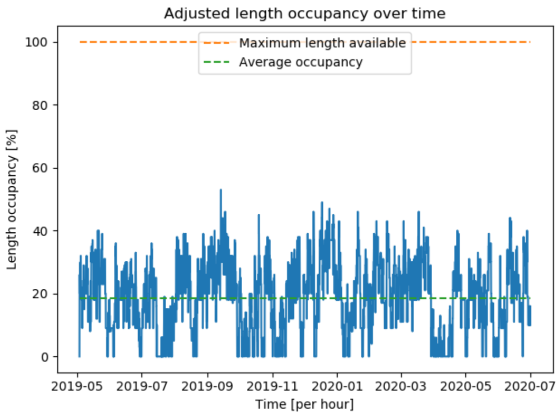


Figure I.5: Rotterdam EMO: Adjusted length occupancy

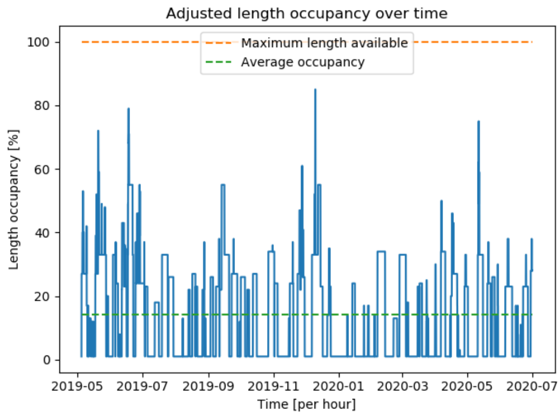


Figure I.6: Vlissingen OVET: Adjusted length occupancy

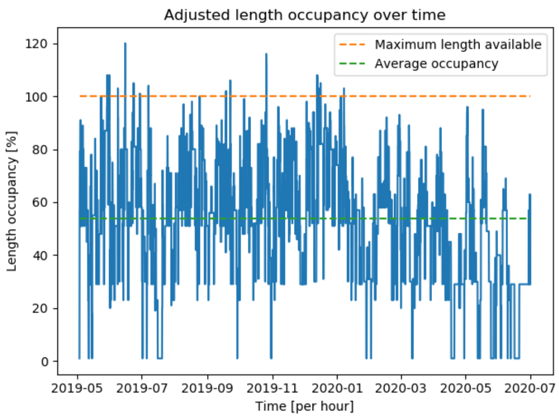


Figure I.7: Rotterdam EECV: Adjusted length occupancy

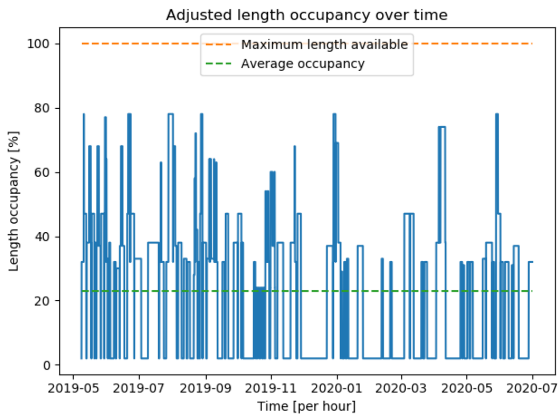


Figure I.8: Dunkirk Western Bulk: Adjusted length occupancy

I.3. Liquid bulk terminals

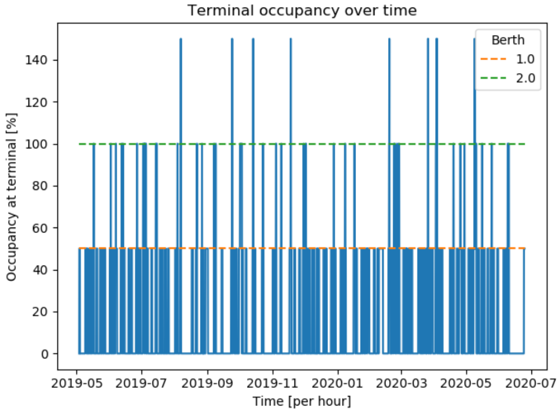


Figure I.9: Rotterdam GATE: berth occupancy

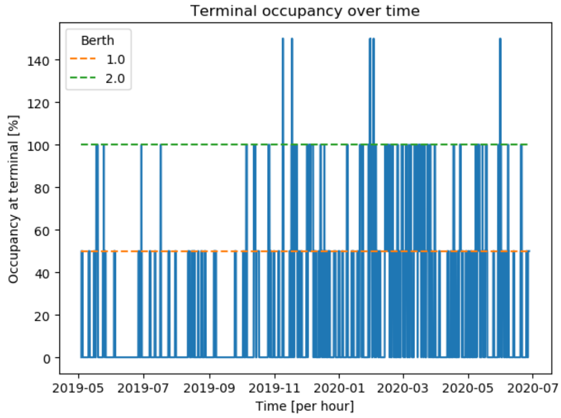


Figure I.10: Zeebrugge LNG: berth occupancy

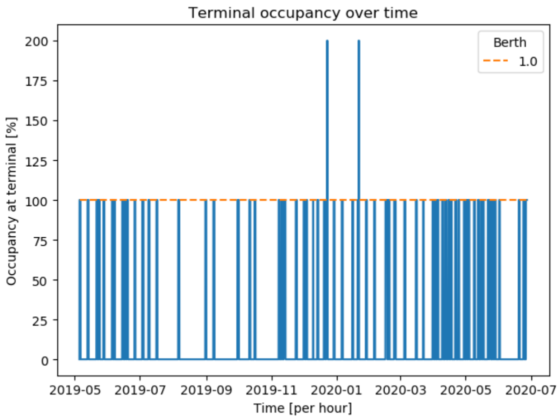


Figure I.11: Dunkirk LNG: berth occupancy

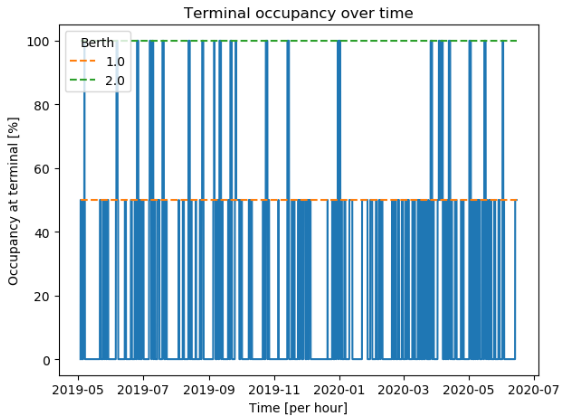


Figure I.12: France Montoir LNG: berth occupancy