Erasmus University Rotterdam

Erasmus School of Economics

Master's Thesis - Operations Research and Quantitative Logistics

Tugboat resting location optimization using AIS data analysis

A case study in the Port of Rotterdam

Author:

Sylvia Kaljouw

415643

Supervisor:

Dr. P.C. Bouman

Second Assessor:

Dr. S. Sharif Azadeh

April 5, 2019

Abstract

In this thesis we investigate how well Automatic Identification System data can describe the nautical chain of the Port of Rotterdam and we use this data analysis to optimize the positioning of tugboat resting locations in the port. We find that the quality of AIS data is not flawless, however, it can accurately describe certain parts of the nautical chain, such as locations where pilot tenders and tugboats meet deep-sea vessels to start or end service and service durations. Because of missing data, we find that the AIS data cannot be used to describe interarrival times of nautical service requests accurately. Using the findings of this analysis, we then model the problem of positioning tugboat resting locations in the Port of Rotterdam as a two-stage stochastic problem with uncertainty in the demand and return parameters. We are able to find upper and lower bounds on the optimal objective value of the stochastic model and we find the optimal solution within reasonable computation time. We test the effect of changing to the optimal locations with a discrete-event simulation model, where we find that a yearly saving of almost €65,000 or 3.8% of total costs is achieved.

Keywords: Stochastic programming, facility location problem, Automatic Identification System, Port of Rotterdam, nautical chain.

Contents

1 Introduction

Every year roughly 30,000 sea-going vessels visit the Port of Rotterdam, giving it a throughput of 467.4 million tons per year[1]. This makes Rotterdam the largest port in Europe and the tenth largest port in the world. However, as the ports of Hamburg and Antwerp are close to the Port of Rotterdam and have generally overlapping hinterland regions to supply, there is tough competition between these ports in the Hamburg-Le Havre range for attracting shipping lines. A shipping line's choice for a port generally depends on capacity considerations, cost of handling, quality of hinterland transport to end destination and quality of nautical services. Tavasszy et al. (2011) emphasize the influence of generalized port costs on the port transshipment volumes. Intuitively, higher generalized port costs make a port less attractive for the shipping company and therefore cause lower port transshipment volumes. One important factor included in the generalized port costs are the nautical service costs. Therefore, the costs of the nautical services can be seen as an influencing factor for the shipping line's port decision. Because of this direct relation, the nautical services are the focus of this thesis.

The nautical service providers can be defined as all agents who provide some service to a vessel during a visit to the Port of Rotterdam. This includes a piloting organization, tugboat services, boatmen services, the terminal and the port authority. However, in this thesis, only the piloting organization, tugboat services, and the terminal are assumed to have an influence on the quality of the nautical service chain (Verduijn, 2017). As all of these services serve the same vessel, disturbances at one service create delays and planning issues throughout the nautical chain. The costs of such a congestion are considerably large as a one hour delay is already estimated to cause additional operating costs for the shipping company in the range of 300-800 dollars depending on the size of the ship (Murray, 2015), which is excluding all costs made by the other stakeholders as a consequence of this delay. Moreover, if the service times for the various services are unpredictable, the shipping company will have to plan buffer time for the port visit, again resulting in considerable costs. Naturally, these costs affect the shipping line's port choice when ports are relatively close together as is the case in

[1]From: https://www.portofrotterdam.com/en/our-port/facts-figures-about-the-port. Accessed: 2 October 2018

the Hamburg-Le Havre range.

These are important reasons for the Port of Rotterdam to monitor the performance of its nautical chain. One of the projects which was set up by the Port of Rotterdam to improve the understanding of the nautical chain is the Swarmport project which is a collaboration between NWO, TNO, TU Delft, University of Maastricht, Rotterdam Port Authority, Intertransis, ECT and Smartport. The ultimate goal of the Swarmport project is "to design self-organisational strategies for port processes that increase resilience, flexibility and reliability". The self-organization in this definition is on the level of service organizations and not on the level of individual tugboats or pilots. As part of the Swarmport project, this master's thesis contributes by investigating strategies which can be used to operate tugboats with high efficiency and reliability. Subsequently, the main research question for this thesis on a strategic level is: "What is the most efficient positioning of tugboat resting locations in the Port of Rotterdam when considering real ship location data?" The main data set available to answer this question comes from a data source called Automatic Identification System (AIS). This data has never been used to describe the nautical chain of a port yet, therefore, a second underlying research question in this thesis is: "Can AIS data accurately describe the Port of Rotterdam nautical chain?" This question is relevant for the tugboat resting location optimization and is also an important contribution to future research in the Swarmport project.

To answer the main research question of positioning tugboat resting locations, we investigate which locations in the port would be most efficient for tugboats to switch crews and rest between towing jobs. This is a dynamic problem in the sense that a ship can leave for a service request from one location and return back to a different location, depending on which location is the most efficient in that moment. Additionally, from inspection of historical data we know that the service demand and tugboat return rates for the different areas in the port are stochastic, which is a complicating factor in solving this problem. We model the problem as a two-stage stochastic programming problem based on the well-known facility location problem, where the objective is to minimize the expected distance traveled between resting locations and demand locations. Before solving the model we first calculate upper and lower bounds on the objective value to find the value of solving the stochastic problem. We then solve the stochastic problem using a rewriting method which transforms the problem into a single stage problem. To validate the quality of the optimal solution, the kilometers traveled and jobs delayed when applying these resting locations are compared to those when applying the current resting locations using a discrete-event simulation model and a greedy dispatching approach.

As indicated, the main data source that we work with to solve the resting location optimization problem is Automatic Identification System (AIS) data, which contains among others the location data of the vessels visiting the port as well as location data of the nautical service vessels such as pilot tenders and tugboats. As we cannot use this raw data directly for the optimization problem described above, an extensive data analysis is included in this thesis research to answer the second research question. The first step of this analysis involves creating event detection rules, where the question to be answered is what pattern in the raw data defines a specific event, for example a tugboat service completion event. These rules are created via a clustering approach based on four months of AIS data of deep-sea vessels, tugboats and pilot tenders. Once we found event detection rules, statistical distributions on for example service durations are defined.

In this thesis, finding statistical distributions from the event data in the Port of Rotterdam is an essential task towards identifying strategies for efficient tugboat operations. However, the identification of these distributions in itself is also an important practical and scientific contribution. A considerable amount of research has already been done in the field of predicting the arrival time of deep-sea vessels at the anchoring area outside of the port, as summarized in Dobrkovic et al. (2016). This extensive amount of research has led to considerable accuracy in predictions for vessel arrival times at the outside of the port. However, the predictions of arrivals at specific locations within the harbour and service times in the nautical chain are lacking such precision. This in itself complicates the planning of activities of the nautical chain and as a consequence it affects the overall reliability of the nautical chain services. Therefore, accurate estimates of arrival and service patterns as analyzed in this thesis can have a significant practical impact.

The remainder of this thesis is organized as follows. First, Chapter 2 gives a detailed description of the nautical chain processes, followed by an overview of previous related work, both as part of project Swarmport and outside of this project in Chapter 3. Chapter 4 gives a problem description of the tugboat resting location problem. Next, the solution method for the tugboat resting location problem is given in Chapter 5. In Chapter 6 the AIS data and the port call data which are used in this study are discussed in more detail, after which in Chapter 7 the process of data analysis is described. Afterwards, Chapter 8 describes the results of the resting location problem, Chapter 9 discusses some of the limitations of the research and future research directions. Finally, Chapter 10 briefly summarizes the methods and the findings.

2 The nautical chain

In this chapter we explain the nautical service process in more detail. That is, we explain the processes of arrival, departure and shift movements. Additionally, we explain why the terminal operations are treated as a black box in this research.

The nautical chain consists of all services that are used by a sea-going vessel when visiting the port in order to make its movements in the port area smooth and safe. In general, the nautical chain consists of the Harbour Master, piloting organization, tugboat companies, boatmen companies and a terminal. However, since the Harbour Master and the boatmen companies are not crucial for the efficiency of the nautical chain (Verduijn, 2017), they are out of scope for this research. Three movements can be identified in a vessel's visit to the port: arrival (berthing), departure (unberthing) and shifting-berth. The shifting-berth movement is optional, but can also occur more than once during a visit. The below descriptions of the arrival and departure movements have been paraphrased from Verduijn (2017, p. 5-7). We first explain the arrival movement.

An incoming ship informs the Harbour Master of its arrival 24 hours before the actual arrival in the harbour. Six hours before the ship arrives at the location where the pilot should board, the piloting organization is informed of the ship's arrival. Next, 2.5 hours before arrival at the pilot boarding location, the pilot is planned and ordered. After this, a check is done on whether the berth is available at the terminal, if not, the ship needs to wait in the Anchor Area outside of the port and the pilot service needs to be canceled. Pilot services can be cancelled for free at the latest one hour before arrival at the pilot boarding location. Only once the terminal has allocated an available berth to the ship, the ship can proceed to the terminal. If necessary, a pilot is ordered again and boards the ship at the dedicated location. Once the pilot has boarded the ship, it is decided whether tugboats are needed and if so, how many should be ordered. These tugboats are then ordered and a meeting place and time are decided. If tugboats are not immediately available, the ship waits and slows down. When both tugboats and pilot are in place, the mooring can be done and the terminal can initiate operation. This is the end of the berthing movement. (Verduijn, 2017)

Although terminal operations are the main reason for a visit to the Port of Rotterdam, these operations are treated as a black box in this research, that is, the terminal operating time is treated as a given. The reason for this assumption is that the terminal operations are privatised and are optimized by specialized algorithms of the terminals which we do not have access to. The terminal operations are thus represented in this research by the time the ship is lying at the berth. This time also includes, for example, bunkering of the ship.

The second movement investigated in this research is the departure of a vessel, this process has some different elements. During the terminal activities, the ship's estimated departure time is updated on a regular basis. If the estimated departure time is less than two hours away, this time is firmed as the actual departure time, upon which the Ship Agent orders the departure process, including a pilot and tugboat order. Pilot and tugboat now both need to be available from the start, so the ship has to wait at the berth place if either service is not available. If both services are available and the terminal operations and bunkering are finished, the ship departs and leaves the port. (Verduijn, 2017)

The last movement considered is the shifting-berth process. In this process, again the ship leaves the terminal at some estimated departure time. Similar to the unberthing movement, the ship has to wait at the berth and can only leave once both the tugboat and the piloting services are available. Once these are available, instead of leaving the port, the vessel is tugged to a different berth, where the pilot and tugboats leave the vessel and terminal operations commence once more. Other than the arrival and departure movements, this movement can be repeated multiple times during one visit or it can be skipped during a visit.

3 Overview of previous research

This chapter focuses on discussing previous research related to the data and the research questions of this thesis. The chapter is split into two parts, the first part focuses on previously executed research under project Swarmport and the second part focuses on other related literature.

3.1 Previous research in Swarmport

Over the past years, the Swarmport project team has investigated the nautical chain by identifying agents and their characteristics in the nautical chain and investigating relationships between these agents. For each of the agents and relationships their importance for the nautical chain process was analyzed. Additionally, an agent based simulation model (ABSM) was created. This section captures some highlights of the Swarmport research which are relevant for this thesis.

3.1.1 Agents and relations

The research on agents and relations was done by Nikghadam et al. (2019) and Verduijn (2017). From this research, many interesting conclusions were found. However, here we only discuss a few which are relevant in answering the research questions addressed in this thesis.

An important insight from Verduijn (2017) is the fact that boatmen and harbour master, although critical parts of the nautical chain, almost never cause a delay in the arrival or departure of a ship. Therefore, these are not considered critical in the research for efficiency in the nautical chain. For this reason, as stated before, they are not considered in the remainder of this thesis.

Another interesting conclusion from Verduijn (2017) is that there is no central command in the nautical chain. This means that none of the actors have the ability to control activities other than their own in the nautical chain. This raises the question of whether this is in fact the most efficient way of working. Centralizing the command in the nautical chain might have significant impact on the efficiency of the nautical chain services. Given this information, it should be noted that the measures such as optimizing the tugboat resting locations investigated in this thesis could not have the desired efficiency effects on the complete chain if it turns out that the inefficiencies are to a large extent created by the choice of having a decentralized way of working. Therefore, it could be that to optimize the overall nautical chain processes, the optimization of, for example, the tugboat process should be less tight to give other parts of the process more room for improvement.

Nikghadam et al. (2019) also gave some relevant insights into the agents and relations in the nautical chain. One of those is that the pilot organization is the only actor in the nautical chain which has relationships with all other actors in the chain and is therefore an important factor for efficiency in the chain. Another interesting observation is that the terminal has the weakest relationships with the other actors in the chain. The terminal also does not seem to be interested in improving these relationships and most likely will not cooperate in any knowledge sharing activities. From this observation, we know that it is most realistic to do any optimizations of the tugboat activities without detailed information on terminal activities.

3.1.2 Simulation model

A simulation model for the nautical chain of the Port of Rotterdam was developed in Python by TNO. However, this model does not accurately represent the actual situation yet. The data analysis in this research is therefore particularly useful for making this Swarmport simulation model more realistic. For this reason, in the data analysis we not only investigate the tugboat process, but also the pilot process. The current state of the model is as follows.

Ships arrive in the port area at an arbitrarily chosen rate, with an arbitrarily chosen speed over ground. The simulation checks whether any pilots are available for the ship, if so, the pilot is assigned to the incoming ship using a first-come, first-served distance-based approach after which the ship can continue. If not, the ship has to wait in the anchor area. Similarly, the simulation checks whether a berth is available, if so, the ship proceeds to the port, if not the ship waits at the anchor area. Once these two services are arranged, the ship continues to the entrance of the port, now a check is done on the necessary available tugboats. Again, if these are not available, the ship waits at the entrance of the port until the service is available. Once the service is available, the ship can proceed to the berth, where terminal activities start.

At departure, again, it is checked whether the tugboat and piloting services are available. If they are not available, the ship has to wait at the berth. Once all services are available and at the ship's location, the ship can leave the port again. Then the tugboats are disconnected and the pilot leaves the ship. This marks the end of the port visit.

There are several other aspects of the process which have not been implemented in the simulation or for which the implementation would benefit from improvements. Firstly, in the current state of the simulation model the shifting-berth movement is not considered which means that all ships only visit a single terminal during a port visit. As this situation can lead to distorted simulation results it is a future research area for which the current data analysis is again useful. Secondly, as mentioned the planning decisions for piloting and tugboat services are currently made in a first-come, first-served distance-based manner. As this approach is quite simple and might be unrealistic, this is also an area which needs investigation and possibly improvement. However, as the actual planning algorithms used by pilot and tugboat organizations are sensitive information and therefore are unavailable it is hard to indicate whether or not such a simple dispatching algorithm is realistic.

3.2 Literature review

This section describes some relevant literature related to the research questions investigated in this thesis. First, papers related to the tugboat resting location problem and possible solution methods are discussed, then the relevant literature regarding AIS data analysis is described.

3.2.1 Optimizing tugboat resting locations

As will be shown in Chapter 4, the tugboat resting location problem can be written as a stochastic case of the facility location problem with reverse flows. Extensive research has been done on the topic of stochastic and deterministic facility location problems. A relatively recent overview can be found in Melo et al. (2009).

One variant of the facility location problem which has particularly similar characteristics to the current problem is given in Amin and Zhang (2013). In Amin and Zhang (2013) a closed-loop supply chain network is considered, implying that the problem to be solved includes both demand flows and return flows. The problem is modeled as a mixed-integer programming problem where for each demand location a demand rate and a return rate are identified. This idea of defining a set of demand locations instead of defining individual customers is useful for the current problem, because it limits the number of nodes in the problem and it does not cause a large deviation from the real problem since often the locations of service requests from different customers are very similar. For example, all customers with a specific section of the port as their destination request service from the same starting point. In line with the current problem, Amin and Zhang (2013) also consider the extension of uncertainty in demand and return rates by adjusting the model to be a stochastic programming model, however the number of scenarios of the uncertain parameters is much smaller than that in this thesis. Exact methods are used to solve the models.

Santoso et al. (2005) propose a stochastic programming approach to solve the problem of supply chain network design. In this problem, not only the number and location of the facilities needs to be decided upon, but also which machines are to be procured for these facilities. The problem is modeled as a two-stage stochastic programming model where the uncertainty comes from processing and transportation costs, demands, supplies and capacity of the facilities. In the first-stage problem it is to be decided which facilities should be opened and which machines should be bought, whereas in the second-stage problem the distribution of products from suppliers through processing units to customers given the facility decisions from the first stage needs to be optimized. The authors use a sample average approximation (SAA) scheme integrated with accelerated Benders decomposition to solve the problem, where the Benders decomposition is used to solve the sample average approximation problems. The advantage of using SAA is that a stochastic problem with many (possibly infinite) scenarios is brought down to a more manageable problem with only a sample of the scenarios. This approach can be useful for problems similar to the current situation where the number of scenarios in the empirical demand and return distributions turns out to be too large for efficient problem solving. However, in the current problem the number of scenarios is manageable and therefore such a scenario reduction approach is not necessary.

A different approach to solving a two-stage stochastic problem which is suitable for the current problem is proposed by Carøe and Schultz (1999). The method proposed here is an algorithm based on dual decomposition and Langrangian relaxation. One of the important features of this method is that it can be used to solve problems with integer recourse, in contrast with many other methods such as the one proposed in Santoso et al. (2005). The authors also extend the method to be able to handle multi-stage stochastic problems. The proposed method is tested with two numerical examples, for which the performance is significantly better than that of solving the problem with CPLEX.

3.2.2 AIS data analysis

As is explained in more detail in Chapter 6, one of the main data sets for this thesis consists of AIS (Automatic Identification System) data, which contains information on characteristics and locations of vessels. Over the past two decades, a considerable amount of research has been done on vessel patterns and collision avoidance based on AIS data. A recent review of this literature is given by Robards et al. (2016).

Pallotta et al. (2013) propose to use the TREAD (Traffic Route Extraction and Anomaly Detection) method to analyze vessel patterns in AIS data. The method is also used for vessel route prediction. The problem description in the paper is somewhat similar to the data analysis in the current research in the sense that in both cases patterns in the vessel behaviour are investigated. However, there are also some notable differences in problem definition. In

Pallotta et al. (2013) the patterns that are identified are route patterns, that is they are patterns in the locations of the ships regardless of the amount of time spent in a position. The current research is focused on finding time-wise patterns in the AIS data, since this is where we can discover delays and places where ships spend more or less time. Another significant difference is that in Pallotta et al. (2013) an unsupervised method is used to discover vessel route patterns, whereas in the current studies, a supervised approach is taken. Nevertheless, the authors make some interesting discoveries on the AIS data which are useful for the data analysis in the current research. Most importantly, they discover that the speed registered in their AIS data was not accurate and should therefore be calculated from time and location data instead.

Another paper investigating the potential of AIS data for various different purposes such as spatial planning is Shelmerdine (2015). The authors apply, among others, density maps and interpolations to get more insight into shipping activity around the island Shetland. Additionally, they give some insight into the quality of the AIS data, which is of particular interest for this thesis. In Shelmerdine (2015), MMSI, latitude, longitude, timestamp and, most interestingly, speed are assumed to be of excellent quality, which contradicts the statement on speed information in Pallotta et al. (2013). This difference might be caused by the difference in location of the case study. The quality of the ship type information is assumed to be moderate, therefore the authors suggest to use more globally defined ship type groups, where they base the definition more on the industry, rather than on activity.

Mou et al. (2010) investigate the behaviour of collision involved ships based on AIS data from the busy waterways Maas West inner and Maas West outer just outside the Port of Rotterdam. In this study, the AIS data is used to create linear regression models and the SAMSON model for risk assessment is used. Both the ship type with the most risk and the area in the waterways with the most risk are identified based on these models. Interestingly, this research uses AIS data over a period of only 62 days, which seems short, as most of the previous researches used roughly a year's worth of data. However, the data set still consists of over 600,000 records, which seems plenty to draw a significant conclusion. The authors do not do any data quality analysis. A Normal distribution is fitted on the speed of the ships in the waterways. The quality of this fit is checked using a Kolmogorov-Smirnov test.

Next, De Boer (2010) investigates the possibility to use AIS data to describe patterns in vessel behaviour in the Port of Rotterdam. The main conclusion is that AIS data can improve insight into individual vessel behaviour, where it was found that vessel size, current, wind and visibility all have significant impact on the vessel path and speed. This research is particularly interesting as it gives a very thorough description of AIS data, explaining that it can be divided in static, voyage related and dynamic data. It also includes a small literature review on errors in AIS data based on, among others, the papers of Bailey et al. (2008) and Harati-Mokhtari et al. (2007). It is mentioned that errors were found in almost all fields of the AIS data, including MMSI, call sign, location, name, length, draught, destination and course. However, Bailey et al. (2008) worked with data over three years, from which they were able to conclude that the number of errors decreased significantly over time. As the current thesis is based on data from 2016 and up, this conclusion is reassuring. Nevertheless, De Boer (2010) does state that more research would be necessary to support this statement with more evidence.

The research of Harati-Mokhtari et al. (2007) is completely focused on errors and inaccuracies in the AIS data. This paper is interesting since it gives us some extra awareness on which fields of the data are reliable and which fields should be handled with caution. The conclusion of the paper is that AIS data contains reasonably many flaws and therefore, navigators on sea should not solely rely on this data to avoid collisions. Especially the manually entered information such as ship type, cargo and destination have a high probability of containing mistakes and should therefore not be trusted blindly.

The latest research on AIS data found is Shu et al. (2017), which investigates the isolated effects of current, wind, visibility and vessel encounters on vessel behaviour based on AIS data and data on weather conditions from the Botlek area in the Port of Rotterdam. The paper finds significant effects of the factors on vessel speed and vessel location. To reduce the number of AIS observations, the authors make a cross section every 50 meters in the waterways, they interpolate two AIS observations on opposite sides of a cross section to obtain the location of the ship at the cross section. This is done for each ship and each cross section in the waterways, to create vessel paths with one observation at each cross section. This way of reducing the size of the data set can be useful in a research like the current data analysis if the AIS data set turns out to be too large for computations. In Shu et al. (2017) no issues with the accuracy of the AIS data are addressed.

3.3 Research gap

Although it is clear that both the main research question and the underlying research question of this research are in some way overlapping with previous research directions, we still find a clear research gap in the field. In general, there is no existing literature on the mathematical optimization of the structure of the nautical chain of a port. More specifically, the investigation of combining mathematical optimization of a facility location problem and detailed real world ship location data to improve the structure of the nautical chain of the port proposed in this research creates a new link between AIS data research and mathematical optimization research.

4 Problem description

The tugboat resting location problem regards the problem of deciding where in the Port of Rotterdam the resting locations of the tugboats should ideally be located. At the resting locations, the tugboat crews can rotate, take a short break or wait for their next assignment. Because more detailed data was not available, the current model formulation is simplified by the assumption that all tugboats belong to the same tugboat organization. The resting location problem can then be seen as a variant of the facility location problem.

There are, however, some deviations which make the problem a special case. First of all, usually in a facility location problem goods are being distributed from a warehouse. In the case of goods (disregarding any returns) distribution happens only once and warehouse stock decreases accordingly. However, in the current problem, services are distributed, which means that we are dealing with service times and returns of the tugboats, after which the same tugboat can be used again for a new task. The second factor which makes this problem different from a general facility location problem is the fact that the tugboats move during service. This means that the start and end locations of service are at a different place and thus that the distance before service from resting location j to a ship A requesting towing service does not equal the distance after service from ship A back to resting location j. Lastly, in the current problem, the Port of Rotterdam authority has set a fixed number of resting locations. According to the authority, the costs for building facilities at each of these locations are equal and negligible, and therefore do not need to be taken into consideration.

Because of the deviations described above, even though the model is based on facility location models, we make several changes for such a model to apply to the current situation. One change in definition which is made to accommodate these differences is that the tugboats travel from the facility to a demand point and back from a return point to the facility. In this way the distance from a facility to a demand/return point remains the same throughout the problem instead of having to work with the changing distance between the ship and the facility.

The mathematical model uses the following sets, decision variables and parameters.

Sets:

- $i \in I$: denotes the set of demand/return points;
- $j \in J$: denotes the set of potential facility locations;

Decision Variables:

- y_{ij}: number of tugboats delivered to demand point i from facility location j;
- w_{ij}: number of tugboats returned from return point i to facility location j;
- z_j: equals 1 if resting location j is opened and 0 otherwise;

Parameters:

- d_i: demand rate for tugboats from demand/return point i;
- r_i: return rate of tugboats from demand/return point i;
- k_{ij}: distance for tugboat to travel from resting location j to point i or vice versa;
- a: number of tugboat resting locations to be opened;
- M_1, M_2: value large enough to ensure that if a resting location is opened, all demand and return points can be satisfied from this location.

Using these sets, variables and parameters, the problem can be formulated as follows:

$$\min \quad \sum_{i \in I} \sum_{j \in J} k_{ij}(y_{ij} + w_{ij}) \tag{1}$$

$$\text{s.t.:} \quad \sum_{j \in J} y_{ij} = d_i \qquad \forall\, i \in I \tag{2}$$

$$\sum_{j \in J} w_{ij} = r_i \qquad \forall\, i \in I \tag{3}$$

$$\sum_{i \in I} y_{ij} \leq M_1 z_j \qquad \forall\, j \in J \tag{4}$$

$$\sum_{i \in I} w_{ij} \leq M_2 z_j \qquad \forall\, j \in J \tag{5}$$

$$\sum_{j \in J} z_j = a \tag{6}$$

$$y_{ij}, w_{ij} \in \mathbb{N} \qquad \forall\, i \in I, j \in J \tag{7}$$

$$z_j \in \mathbb{B} \qquad \forall\, j \in J. \tag{8}$$

The objective is to minimize the total cost of traveling from the resting locations to the demand points and vice versa. Constraints (2) ensure that the demand for service in each of the demand points is satisfied exactly. Constraints (3) ensure that all tugboats which are done with a service are returned to a resting point. Next, constraints (4) and (5) state that tugboats can only be placed at and taken from a resting location if that location has in fact been opened. Constraint (6) ensures that exactly the correct number of resting locations is opened. Finally, constraints (7)-(8) ensure that the variables are all non-negative integers or binary variables.

Even though this is an intuitive way of modeling the problem, and therefore easy to understand, it is not the most efficient mathematical representation of the model. This is mostly due to the so-called Big-M constraints (4) and (5). These constraints do not constrain the feasible region of the problem as much as is possible and could therefore be responsible for long calculation times. Therefore, we remodel the problem using the following reasoning. As we do not have any capacity constraints on the resting locations, in the optimal solution the full demand at demand point i will be taken from that resting location j for which transportation costs k_{ij} are smallest as this is the only way to minimize the objective. Following this reasoning, we know that in the optimal solution $y_{ij} \in \{0, d_i\}$ and $w_{ij} \in \{0, r_i\}$ always hold. Therefore, we can introduce continuous variables $u_{ij} \in [0, 1]$ and $v_{ij} \in [0, 1]$ and write $y_{ij} = d_i u_{ij}$ and $w_{ij} = r_i v_{ij}$. Finally, with these changes in variables we come to the following model.

$$
\begin{aligned}
\min \quad & \sum_{i \in I} \sum_{j \in J} k_{ij}(d_i u_{ij} + r_i v_{ij}) & & \\
\text{s.t.:} \quad & \sum_{j \in J} u_{ij} = 1 & \forall\, i \in I & \quad (9) \\
& \sum_{j \in J} v_{ij} = 1 & \forall\, i \in I & \quad (10) \\
& u_{ij} \leq z_j & \forall\, i \in I, j \in J & \quad (11) \\
& v_{ij} \leq z_j & \forall\, i \in I, j \in J & \quad (12) \\
& \sum_{j \in J} z_j = a & & \quad (13) \\
& u_{ij}, v_{ij} \in [0, 1] & \forall\, i \in I, j \in J & \quad (14) \\
& z_j \in \mathbb{B} & \forall\, j \in J. & \quad (15)
\end{aligned}
$$

In the new model, the objective is still to minimize the total cost of traveling between demand points and resting locations. Constraints (9) and (10) now state that each demand point should be covered by exactly one resting location for each time period. These constraints can be changed in this way because of the above derivation that demand will always be covered by exactly one resting location. In constraints (11) and (12) we can now work with binary and continuous variables and we no longer need to use the big-M constraints. The constraints now state that a demand/return point can only be covered from a resting location if that resting location has in fact been opened. Another change to the model is that now all variables are binary or continuous between zero and one.

For convenience, we write the model in a more compact notation, as this will be useful at a later stage. We define the parameter vectors k, d and r to be the traveling costs, demand and return rate, respectively, all having appropriate dimensions. Furthermore, D, R, F, G and A are appropriate matrices corresponding to equations (9)-(13) and 1 and 0 are vectors of the appropriate size containing all ones or zeros, respectively.

$$
\begin{aligned}
\min \quad & k(du + rv) \\
\text{s.t.:} \quad & Du = 1 && (16) \\
& Rv = 1 && (17) \\
& u \leq Fz && (18) \\
& v \leq Gz && (19) \\
& Az = a && (20) \\
& u, v \in [0, 1] && (21) \\
& z \in \mathbb{B}^{|J|}. && (22)
\end{aligned}
$$

The above model is a static model, however, the demand and return rates are in fact uncertain. Therefore, as a last step we need to rewrite the model as a two-stage stochastic model. In this model, the objective is to minimize the expected future transportation costs. We now work with the bold-face vector of random variables $\boldsymbol{\omega} = (\boldsymbol{d}, \boldsymbol{r})$. If we refer to a particular realization of these variables, we denote them as $\omega = (d, r)$ where $\omega \in \Omega$. The two-stage stochastic model is written out in (23)-(31).

$$
\begin{aligned}
\min \quad & \mathbb{E}[Q(z, \boldsymbol{\omega})] && (23) \\
\text{s.t.:} \quad & Az = a && (24) \\
& z \in \mathbb{B}^{|J|}. && (25)
\end{aligned}
$$

Where $Q(z, \boldsymbol{\omega})$ is the optimal value of the following problem:

$$
\begin{aligned}
\min \quad & k(du + rv) && (26) \\
\text{s.t.:} \quad & Du = 1 && (27) \\
& Rv = 1 && (28) \\
& u \leq Fz && (29) \\
& v \leq Gz && (30) \\
& u, v \in [0, 1]. && (31)
\end{aligned}
$$

In words, in the first stage we decide on which resting locations are opened. Then in the second stage, it is decided which demand and return points are satisfied from which resting locations. The objective is to minimize the expected total distance traveled between resting locations and demand and return points.

5 Methods

In this chapter we describe the methods used to calculate upper and lower bound on the objective value and a method to solve the two-stage stochastic resting location problem.

5.1 Upper bound and lower bounds

Solving a two-stage stochastic programming problem can be difficult and might need extensive computing power. Therefore it is often a good idea to first investigate easier to compute upper and lower bound values to determine whether the value of solving the stochastic problem outweighs the computation effort.

We compute a lower bound by solving the Wait-And-See problem. In this problem different values of z can be chosen for each scenario of the uncertain parameters, that is, for each scenario different resting locations can be chosen. As the distribution of the uncertain parameters is discrete, the Wait-and-See objective value can be found as the expected value over all scenarios ω of

$$\begin{aligned}
\min \quad & k(d^\omega u^\omega + r^\omega v^\omega) && (32)\\
\text{s.t.:} \quad & Du^\omega = 1 && (33)\\
& Rv^\omega = 1 && (34)\\
& u^\omega \leq Fz^\omega && (35)\\
& v^\omega \leq Gz^\omega && (36)\\
& Az^\omega = a && (37)\\
& u^\omega, v^\omega \in [0,1] && (38)\\
& z^\omega \in \mathbb{B}^{|J|}. && (39)
\end{aligned}$$

The objective value of the Wait-and-See problem (WS) is a lower bound on the objective value of the stochastic programming problem (SP) as the optimal solution to the stochastic programming problem is feasible for all realizations of the uncertain parameters, however it might not be the optimal solution for each of the scenarios. Therefore, it holds that

$$WS \leq SP. \tag{40}$$

Our second-stage problem is however a special case since it can be written as a fixed recourse problem. In a fixed recourse problem the constraint matrix and the objective coefficients are both known with certainty and the uncertain parameters are only in the right-hand side of the constraints. With the variables y and w as defined for problem definition (1)-(8) and D_1, R_1, M_1 and M_2 matrices of appropriate size, the second stage problem can be written as

$$\min \quad k(y+w) \tag{41}$$

$$\text{s.t.:} \quad D_1 y = d \tag{42}$$

$$R_1 w = r \tag{43}$$

$$y \leq M_1 z \tag{44}$$

$$v \leq M_2 z \tag{45}$$

$$y, w \in \mathbb{N}^{|I| \times |J|} \tag{46}$$

where the uncertain parameters d and r are only on the right-hand side of the constraints. Hence, the second-stage problem can be written as a fixed recourse problem. Moreover, the second-stage problem is a convex function of the uncertain parameters. Using these properties and the reasoning in Higle (2005) we therefore know that we can find a lower bound from the expected value problem, where the uncertain parameters are replaced by their expected values. We denote the expected values of the uncertain parameters by $\bar{d}$ and $\bar{r}$. The expected value problem is given by

$$\min \quad k(\bar{d}u + \bar{r}v)$$

$$\text{s.t.:} \quad \text{Constraints (16)-(22).}$$

If we denote the objective value of the expected value problem by EV, from Higle (2005) and the above reasoning we know that

$$EV \leq SP. \tag{47}$$

In the appendix we use an example to show that for the current problem the Wait-and-See problem does not by definition provide a tighter lower bound than the expected value problem.

Lastly, to calculate an upper bound on SP we fix the first-stage solution of the stochastic problem to the optimal solution z^{EV} of the expected value problem. The problem to be solved for each scenario ω is then given by

$$\min \quad k(d^\omega u^\omega + r^\omega v^\omega) \tag{48}$$

$$\text{s.t.:} \quad Du^\omega = 1 \tag{49}$$

$$Rv^\omega = 1 \tag{50}$$

$$u^\omega \leq Fz^{EV} \tag{51}$$

$$v^\omega \leq Gz^{EV} \tag{52}$$

$$u^\omega, v^\omega \in [0, 1]. \tag{53}$$

The first-stage solution to the expected value problem satisfies the constraints of the first-stage problem by definition. This makes the solution a feasible solution to the stochastic problem and therefore, denoting the objective value of this problem by EEV, which is obtained by taking the expected value over all scenarios, we can state

$$SP \leq EEV. \tag{54}$$

With these lower and upper bounds we can find the value of solving the two-stage stochastic problem and therefore we can determine whether solving this problem is worth the computation time and effort.

5.2 Rewriting the two-stage stochastic model

If it turns out that computing the solution to the real stochastic problem is necessary or valuable, we need a method to compute the solution to this problem. A commonly used approach to solve such problems with discrete distributions of the uncertain parameters is Benders decomposition. However, in this method in each iteration we need to solve the second-stage problem for all scenarios. As we expect to have many scenarios in our problem this method is likely to become quite slow for our problem. Therefore, instead, we use a rewriting approach shown in Carøe and Schultz (1999) to solve problem (23)-(31).

We first define the sets

$$Z := \{z : Az = a, z \in \mathbb{B}^{|J|}\} \tag{55}$$

and

$$W := \{(u, v) : Du = 1, Rv = 1, u, v \in [0, 1]\}. \tag{56}$$

If we define for each scenario ω the sets

$$S^\omega := \{(z, u^\omega, v^\omega) : z \in Z, u^\omega \leq Fz, v^\omega \leq Gz, (u^\omega, v^\omega) \in W\} \tag{57}$$

and denote by N the number of scenarios $\omega \in \Omega$, we can then write the two-stage stochastic problem (23)-(31) more compactly as

$$f = \min\{\frac{1}{N} \sum_{\omega \in \Omega} (k(d^\omega u^\omega + r^\omega v^\omega)) : (z, u^\omega, v^\omega) \in S^\omega, \forall \omega \in \Omega\}. \tag{58}$$

Next, we introduce copies z^ω for all $\omega \in \Omega$ of the first-stage variables z. With these variables we can rewrite problem (58) as

$$f = \min\{\frac{1}{N} \sum_{\omega \in \Omega} (k(d^\omega u^\omega + r^\omega v^\omega)) : (z^\omega, u^\omega, v^\omega) \in S^\omega, \forall \omega \in \Omega, z^{\omega_1} = z^{\omega_2}, \forall \omega_1, \omega_2 \in \Omega\}. \tag{59}$$

Here $z^{\omega_1} = z^{\omega_2}, \forall \omega_1, \omega_2 \in \Omega$ is called the non-anticipativity constraint. In this form the problem is no longer a two-stage problem and can therefore be solved with a regular MIP solver.

5.3 Simulation

To test to what extent the suggested optimal solutions make a positive difference in the total distance traveled by the tugboats, we simulate the tugboat dispatching process. We use a discrete-event simulation model in which four events can happen: a job start, a job end, a break start and a break end. Furthermore, interarrival times between jobs and start and end locations of jobs are randomly generated and the service times are calculated from the distance between start and end locations and the average tugboat speed.

At the start of the simulation the break start time of each of the 40 tugboats is chosen from a uniform distribution over $[100, 300]$ such that each of the tugboats starts its 15-minute break somewhere between the 100th and the 300th minute of the simulation unless its break had to be postponed due to an incoming job. This last case is discussed in more detail when we explain the dispatching algorithm used. Also, at the start of the simulation all tugboats are equally distributed over the five resting locations.

Now, let us look into the actions taken at each of the four events. First, at a job start event the status variables are updated such that the tugboat is now occupied and no longer available and the event is removed from the event list. Furthermore, a new job arrival event is generated from the interarrival time distribution, the start and end locations of this job are randomly generated and the service duration is calculated. The new job start and job end events are added to the event list. Now the greedy algorithm is called to dispatch this newly generated job to one of the tugboats. A detailed description of this algorithm can be found in Algorithm 1. If none of the tugboats are available at the generated arrival time, the start and end time of the job are delayed by 15 minutes and the greedy algorithm is again called. This process is repeated until an available tugboat is found.

At a job end event, again the status variables are updated such that the occupied tugboat is now available again and the event is removed from the event list. Moreover, if the tug's next event is a break time it is now decided at which resting location this break should take place. The model always chooses the resting location which is closest to the last job end location. Finally, we elaborate on the procedures for the break start and break end events. At a break start event, the status of the tugboat is set to unavailable and the event is removed from the event list. On the other hand, at a break end event the status of the tugboat is set back to

Algorithm 1 Greedy dispatching algorithm

Inputs: movStartTime, movEndTime, movStartLoc, movEndLoc
$travelTimeRestStart, travelTimeRestEnd \leftarrow$ travel times to nearest rest location from movStartLoc and movEndLoc
$movTime \leftarrow$ time between movStartTime and movEndTime
for each tugboat in the fleet **do**
  if break time of tugboat not in $movTime$ & assigned jobs not in $movTime$ **then**
    $prevJobAv \leftarrow True$ if tugboat can travel to mov from previous job before $movStart$, else $False$
    $nextJobAv \leftarrow True$ if tugboat can travel from mov to next job in time, else $False$
    if $prevJobAv$ & $nextJobAv$ **then**
      Add tugboat to list of available tugboats
      Store travel time from last location for tugboat
    end if
  else if delaying break time of tugboat makes tugboat available **then**
    $prevJobAv \leftarrow True$ if tugboat can travel to mov from previous job before $movStart$, else $False$
    $nextJobAv \leftarrow True$ if tugboat can travel from mov to next job in time, else $False$
    if $prevJobAv$ & $nextJobAv$ **then**
      Add tugboat to list of available tugboats
      Store travel time from last location for tugboat and store break delay time
    end if
  end if
end for
Choose the tugboat from available tugboats which has the smallest travel time to get to $movStart$
Add mov to tug's job list and if necessary change tug's break time

available and the event is removed from the event list. As a last action, at a break end event the next break start and break end times for this tugboat are calculated and these events are added to the event list.

The final aspect of this simulation model which is worth mentioning are the performance measures: kilometers traveled and jobs delayed. As kilometers traveled is the objective of the optimization model it is also an important result of the simulation. Moreover, as stated before, if a ship has to wait to be towed to its next destination this involves significant costs, therefore jobs delayed is another important performance measure. The simulation starts measuring these performance indicators after the warm-up period. The length of this warm-up period is determined by running the simulation model for multiple days and determining the number of days after which the daily performance measures are similar on subsequent days.

6 Data

To find results for the Port of Rotterdam, we use several different data sources. The two main data sets used are an AIS data set of the Port of Rotterdam area and a data set of all port calls in the Port of Rotterdam. Other data includes berth locations and a mapping between MMSI numbers and ship names.

6.1 AIS data

The largest data set used in this research consists of AIS data, AIS stands for Automatic Identification System. An AIS can detect other AIS equipped vessels or stations and provides information about those ships to avoid collision hazard. As such, this data was originally adopted to improve safety at sea. Maritime radar detection is based on Super High Frequency (SHF) signals, which have difficulties with detection behind hills and in foggy or rainy circumstances. These are situations where the Very High Frequency (VHF) based AIS signals perform much better, which is why these signals were created.

As mentioned in Section 3.2.2, AIS data contains roughly three categories of messages: static, voyage related and dynamic data. The static message contains information on the ship's characteristics, such as MMSI number, vessel type and length. This data is registered at installation of the AIS equipment. The second type of data is voyage related data. This data is entered by the vessel crew, it is information such as destination, the vessel's draught and estimated time of arrival. Since this information has to be updated by the crew, it is prone to errors and therefore not used in this research. The last type of data in the AIS message is dynamic data, containing information on the vessel's position, heading and rate of turn. Both the static and the dynamic data are used in this research. The exact data fields which are used can be found in Table 1.

Our AIS data set is split into three parts: information on cargoships, tankerships and service ships. In this categorization cargoships are defined as RoRo vessels or vessels which carry containers or dry bulk and tankerships are defined as ships which carry liquid bulk. Furthermore, service ships are defined as tugboats and pilot boats. Over the chosen four month

Data field	Description
MMSI number	Unique identification number for ships, groups and stations consisting of nine digits
UTS	Time and date of an observation
sog	Speed over ground of vessel
x	Latitude (location of the vessel)
y	Longitude (location of the vessel)

Table 1: AIS data fields which are relevant for this research.

period in 2016 we have several millions of observations for each category in the AIS data set.

6.2 Port call data

In addition to the AIS data, a data set which contains port call information, such as movement type, berth name and number, ship name, arrival and departure time at berth, ship details, etc over the same period of four months in 2016 was made available by the Port of Rotterdam Authority. Also in this data set, not all data fields are relevant for this research. Please find an overview of the relevant data fields in Table 2.

Data field	Description
Visit ID	Unique ID for a visit consisting of an arrival movement, a departure movement and optional shifts
Movement ID	Unique ID for each movement (arrival/departure/shift) in the harbour
Ship name	Name of the ship which belongs to this specific movement
Movement type	Arrival, departure or shift
Berth name Arrival/Departure	The name of the berth where the ship arrives/departs
Length of ship	Length of the ship associated with the movement
Pilot start/end DT	Date and time when piloting activities start/end for this movement (manually entered)

Table 2: Port call data fields which are relevant for this research.

6.3 Additional data

Additionally, a database with ship names and MMSI numbers is used in order to link the MMSI numbers in the AIS data to the ship names in the port call data. As a last piece of data, the port of Rotterdam provides longitude and latitude coordinates of the berths in the Port of Rotterdam.

7 Data analysis

The data analysis is mainly focused on the AIS data, however some parts of the analysis require the port call data. The data analysis can roughly be split into two segments: identifying events in raw data and defining statistical distributions. In this chapter we explain the approaches taken to tackle these challenges and we elaborate on our findings in the data.

7.1 Missing data

While working with the AIS location data of the ships we realized that this data was not complete. That is, for 20.6% of the movements which were present in the port call data over the period of June 1st 2016 to October 31st 2016 we could not find AIS data over a time range longer than 30 minutes for the deep-sea vessel in a wide range around the Port of Rotterdam. We choose to consider less than 30 minutes of data as missing data since this time range is not nearly enough to be able to do any event detection. In Table 3 we give an overview of the distribution of this missing data over the different classes of vessels. These classes of vessels are made based on the length and the draft of the ship where we exclude ships smaller than 120 meters in length. This choice is made because ships smaller than 120 meters in general do not need the service of tugboats and pilots and are therefore not of interest. From the table we conclude that there is no significant difference between the percentages of data missing in the different classes.

Moreover, Figure 1 gives an overview of the percentage of missing movements in the AIS data over the days in the given timerange. As can be seen, on most days no more than 30% of the data is missing. However, there are a few days on which a large part or almost all of the data is missing. For example, on the 9th of June 93.5% of the movement data is missing, on the 8th of June 73.7% of the movement data is missing and on the 29th of August 66.7% of the movement data is missing. This observation is unexpected and somewhat worrisome since the data set was assumed to be complete. However, since the missing of the data is not based on a clear pattern in time or type of ship, we assume that the available AIS data is an unbiased sample of the complete ship location data. Therefore, the quality

Class	Length (m)	Draft (m)	Total	Missing	% Missing
2	120-200		3458	746	21.6%
3a	200-300	<14.3	820	160	19.5%
3b	200-300	>14.3	1034	200	19.3%
4	>300	<14.3	21	3	14.3%
5	>300	14.3-17.4	334	58	17.4%
6	>300	>17.4	127	23	18.1%
Total			5794	1190	20.5%

Table 3: The vessel classes used in this thesis corresponding to the length and draft measurements of the vessel in meters. The right side of the table shows the number of port movements as well as the number of port movements missing from the AIS data for each class in the time range from June 1st 2016 to October 31st 2016. In the last column the percentage of port movements missing from the AIS data is shown.

of the statistical distributions in this research is not immediately affected by this missing data.

Nevertheless, the reason why such an extensive part of the port movements seems to be missing from the AIS data should be further investigated. Furthermore, the fact that this data is missing should also be a warning that the same could hold for the AIS data of tugboats and pilot tenders. This can lead to worse quality of the statistical results as no data for a tugboat which is serving means that the service is not noted in the research and therefore that the workload is estimated lower in the research than the actual workload would be. However, since we do not have any data on which tugboat or pilot tender should be in operation at which times, we cannot investigate whether this data is in fact also missing and if so how much of the data is missing.

7.2 Event identification

We now identify events in the raw data of ship locations which are needed to define statistical distributions regarding the berthing and unberthing process in the Port of Rotterdam. The events that we are interested in are for example the start and end of towing service for a vessel. A full list of the events which need to be identified is given in Table 4. Additionally, this table shows the patterns in AIS data which should be found at the occurrence of these events. One important remark is that the time at which the events are detected from the data cannot be accurate down to a single minute, because the location data does not allow

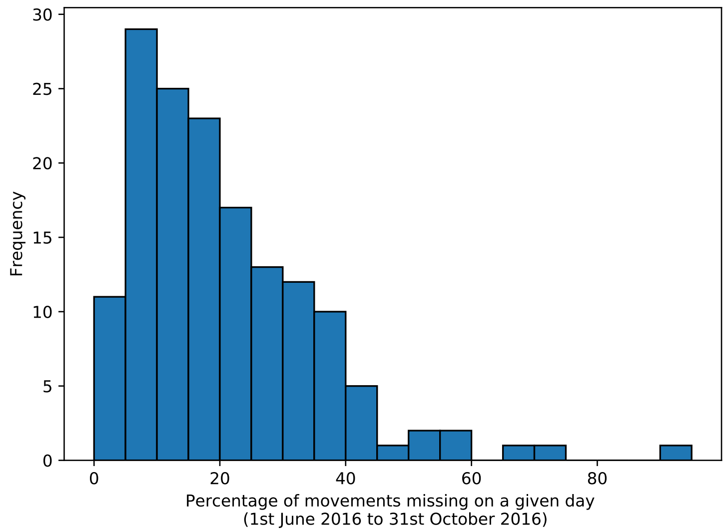


Figure 1: Distribution of days from June 1st 2016 to October 31st 2016 over the percentage of missing port movements in AIS data.

for this much precision. However for the purposes of this research and given that the total duration of the berthing and unberthing processes is generally at least 1.5 hours, such a high level of precision is not necessary.

As stated in Table 4, we can use information on pilot start and end times from the port call data set provided by the Port of Rotterdam to identify some of the events. However, these start and end times are manually added to the data set by the pilots and are therefore not assumed to be very accurate. Therefore, these times are a good starting point from where to search in the AIS data, but should not be used as the actual event time. Knowing this, we try to find which service ships are serving which deep-sea vessels at which time, because if we know this we can easily identify the events afterwards.

We look for a threshold which separates the service boats which are providing service at some point in time from those who are not. One observation from Table 4 is that the patterns identifying each of the six events involve both the distance and the difference in speed between the pilot tender or tugboat and the deep-sea vessel. Therefore, we look for such a threshold in these two parts of the data. We first apply this approach to the part of the process which involves tugboats, leaving pilot tenders out of account for now.

Event	Pattern in data	Data source
Arrival at the pilot meeting location *(arrival movement)*	Pilot meeting time estimate	Port call data
	Small distance pilot tender and vessel	AIS data
Start of tugboat service *(arrival movement)*	Small distance tugboat and vessel	AIS data
	Vessel speed equal to tugboat speed	AIS data
End of tugboat and pilot service *(arrival movement)*, this equals the start of berthing time	Pilot end time estimate	Port call data
	Large difference in vessel/tugboat speed	AIS data
	Increasing distance vessel and tugboat	AIS data
Start of tugboat and pilot service *(shift movement)*, this equals the end of berthing time	Pilot start time estimate	Port call data
	Vessel speed equal to tugboat speed	AIS data
	Small distance tugboat and vessel	AIS data
End of tugboat and pilot service *(shift movement)*, this equals the start of berthing time	Pilot end time estimate	Port call data
	Large difference in vessel/tugboat speed	AIS data
	Increasing distance vessel and tugboat	AIS data
Start of tugboat and pilot service *(departure movement)*, this equals the end of berthing time	Pilot start time estimate	Port call data
	Vessel speed equal to tugboat speed	AIS data
	Small distance tugboat and vessel	AIS data
End of tugboat service *(departure movement)*	Increasing distance vessel and tugboat	AIS data
	Large difference in vessel/tugboat speed	AIS data
End of pilot service *(departure movement)*	Pilot end time estimate	Port call data
	Small distance pilot tender and vessel	AIS data

Table 4: Overview of events to be identified and patterns which describe these events. (In this table vessel refers to deep-sea vessel.)

7.2.1 Tugboats

To identify a service relationship between a tugboat and a deep-sea vessel, we first need to decide on a timeframe in which to search. The only estimate we have for this comes from the pilot start time ($\tilde{s}_p$) and pilot end time ($\tilde{e}_p$) estimates in the port call data. With these start and end time estimates, we define the searching timeframe as follows. We distinguish three types of movements: arrival, departure and shift. For an arrival, the search frame is [$\tilde{s}_p-$ 60 minutes, $\tilde{e}_p+$ 30 minutes], for a departure, the search frame is [$\tilde{s}_p-$ 30 minutes, $\tilde{e}_p+$ 60 minutes] and lastly for a shift, the search frame is [$\tilde{s}_p-$ 30 minutes, $\tilde{e}_p+$ 30 minutes]. We choose to take 60 minutes of extra time outside of the port, because we expect the process to be less predictable there. Inside of the port, the process should be more predictable and therefore 30 minutes of extra time should suffice.

In the given search frames we define a group of candidate tugboats by applying a k-d tree

(Bentley, 1975) to find the 5 nearest neighbours at each minute in the search frame. In this case, the leaves of the tree are the geographic locations of all tugboats and the vessel at some minute m in the timeframe. The tree splits the geographic plane into neighbourhoods which are used to find the nearest neighbour of the vessel. Then, the candidate tugboats are those which were a nearest neighbour to the deep-sea vessel for at least 3 minutes within the search frame. For all candidate tugboats, the median of the geographic distance and absolute difference in speed between tugboat and vessel over all 15 minute intervals in the search frame is plotted for the month of October 2016. We choose to use the median because this is less sensitive to outliers (and therefore to mistakes in the data) than the mean. The speed is calculated from the AIS locations of the ships as previous research has shown that the AIS speeds are not accurate in many cases. As some of the nearest neighbour tugboats are up to 80 kilometers away, we zoom to the most relevant part of the plot which is shown in Figure 2. As Figure 2 shows, the threshold for a tugboat to be serving is at 150 meters of distance from the front or back of the deep-sea vessel and at 0.4 meters per second difference in speed.

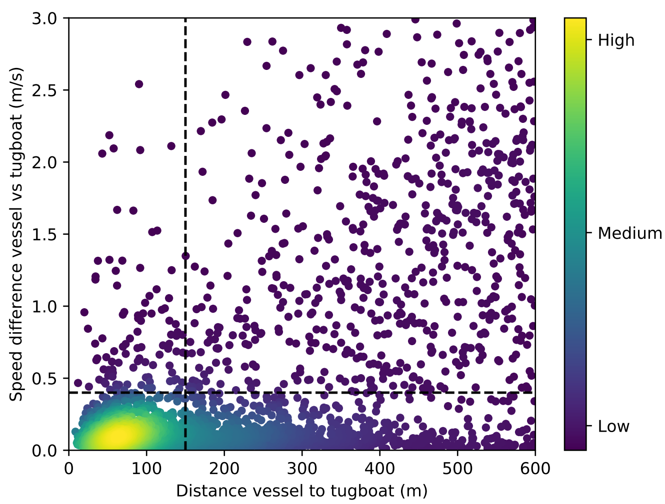


Figure 2: Scatter plot used to identify boundaries on distance and difference in speed. Each point represents the median over 15 minutes for a vessel and tugboat combination. Higher density of points in the plots is represented by a lighter color (yellow), lower density is represented by a darker color (blue) in both graphs. The dotted lines represent the final boundaries used to identify whether a tugboat is providing service or not.

Using these boundaries we identify the tugboats in service and times at which they are towing a vessel. As an extra check, we compare the number of tugboats used per vessel that we get from the AIS analysis with the number of tugboats used per vessel that was found in Verduijn (2017) based on different data of the Port of Rotterdam over the same year. The results can be found in Table 5. From the table we can see that the results are not perfectly the same, however the defined thresholds seem reasonable. Part of the reason that the results of the tugboat matching are somewhat off might be the relatively large percentage of AIS data which is missing and was therefore not included in the results shown on the right hand side of the table. However, as we do not know the data source used by Verduijn (2017), we also do not know the quality of this data source. This means that the comparison in Table 5 can only be used as a rough indication of whether the values are similar, in fact we cannot draw any definitive conclusions on the quality of the current work based on this comparison.

	Verduijn (2017)		**AIS analysis**	
#Tugboats	#Movements	%	#Movements	%
0	1777	59%	1269	65%
1	352	12%	486	25%
2	784	26%	185	9%
3	90	3%	10	1%
4	25	1%	3	0%

Table 5: Comparison of percentages of vessel movements which used 0, 1, 2, 3 or 4 tugboats. Left, the results of the research of Verduijn (2017) are shown for a two-month period, on the right the results of the current AIS analysis are shown over a period of two months.

7.2.2 Pilots

For the pilots we use the same approach at heart, with just a few differences. First, as pilots are only delivered or retrieved from the vessel by a pilot tender if the vessel is at sea, we only focus on those parts of the process. That is, our search frame is now set to [$\tilde{s}_p-$ 60 minutes, $\tilde{s}_p+$ 60 minutes] for arrival movements and [$\tilde{e}_p-$ 60 minutes, $\tilde{e}_p+$ 60 minutes] for departure movements. Shift movements are discarded in this case, as these go from one berth to another and in general no pilot tender is used to bring a pilot to and from a berth, instead pilots travel to and from berths by car.

Secondly, as the pilot tender only meets the vessel for a very short moment instead of staying with the vessel during service, we do not use the median over 15 minutes in the scatter plot. Instead, we set each minute to be a separate point in the scatter plot. With these two adjustments, we find the scatter plot shown in Figure 3. In this plot the thresholds are not as clear as in the case of the tugboats as the cluster has a center but also a long tail to the right. We suspect that the actual thresholds are at a distance of roughly 200 meters from the vessel and at roughly 2 meters per second difference in speed. However, as these thresholds are not as clear, we also test the case where we include the tail of the cluster and set the thresholds at a distance of 600 meters and 7.5 meters per second difference in speed. After investigating the geographical positions of these potential pilot meeting points we see that the points in the tail of the cluster are not actual pilot meeting moments. Therefore, the final thresholds are set at a distance of 200 meters and 2 meters per second speed difference.

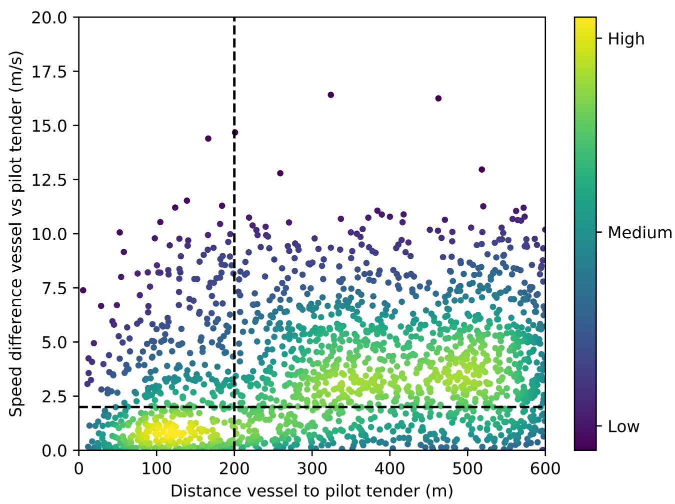
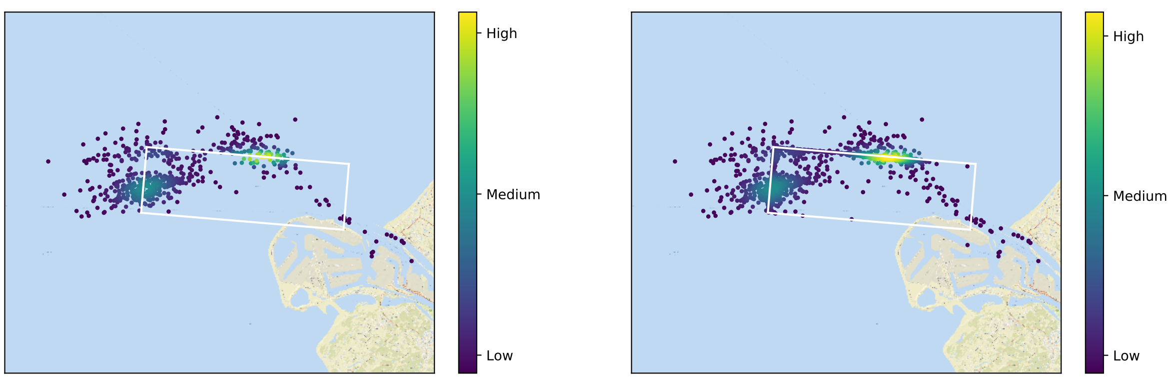


Figure 3: Scatter plot for pilots to identify boundaries on distance and difference in speed. Each point in this plot represents an observation in the AIS data. Higher density of points in the plots is represented by a lighter color (yellow), lower density is represented by a darker color (blue) in both graphs. The dotted lines illustrate the chosen thresholds between pilot tenders meeting a vessel and other pilot tenders.

Using these thresholds we find the corresponding pilot and the pilot start or end time for 568 vessel movements. However, there are also 2632 movements for which a pilot was necessary according to the port call data and AIS data was available, but no pilot meeting was found with the described method. Therefore, we extend the method in the following way. We plot the locations of the pilot meeting points which were already found. These locations are shown in Figure 4a and we can immediately see areas where most pilot meetings take place. With this information we find that we can estimate the start of the pilot service time by taking it as the time when the vessel crosses the borders at N52.03, N51.98 or E3.80, as shown by the white rectangle in Figure 4b. Finally, the east border of the white rectangle is used to make sure that a vessel crossing the north and south boundaries to the east of this border is not registered as a pilot meeting point.

With this method we find 2365 pilot meeting points out of the 3200 movements which should have a pilot according to the port call data. Although the estimates made with this method are probably not completely accurate, they should give a reasonable indication of the pilot service start or end times which can be used for optimization and simulation purposes.



(a) Initial meeting points

(b) Meeting points using boundary method

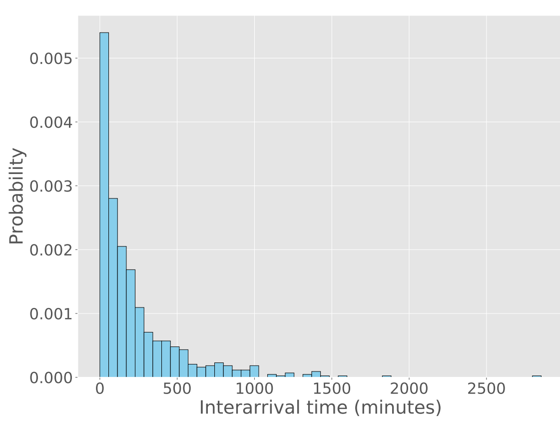
Figure 4: Meeting locations between vessels and pilots over the course of four months. Higher density of points in the plots is represented by a lighter color (yellow), lower density is represented by a darker color (blue) in both graphs. The north, west and south sides of the white rectangle represent the boundaries used for the estimated pilot meeting points.

7.3 Distributions

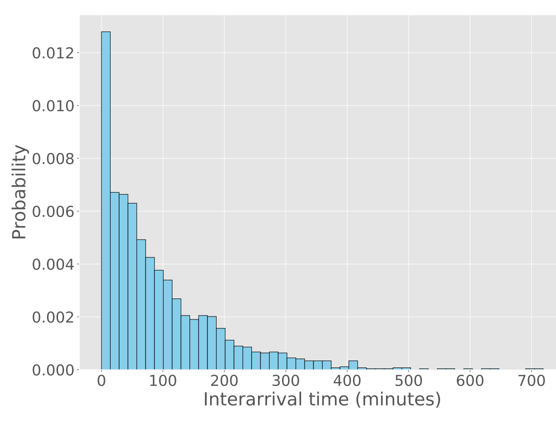
Once we have found the necessary events in the AIS data, we accumulate these events into empirical interarrival time and service time distributions. These distributions are used for the optimization problems as well as for the Swarmport simulation model. The following distributions are extracted from the data:

- Interarrival time at pilot meeting point;
- Service time of pilots in the arrival, departure and shift processes;
- Service time of tugboats in the arrival, departure and shift processes;
- Distribution of number of tugboats needed per request (See Table 5).

The resulting distributions are shown in Figures 5, 6 and 7. We now briefly discuss the found distributions. Figure 5a shows the interarrival times at the pilot meeting location found from the AIS data analysis. These interarrival times are longer than expected from the knowledge we have of the Port of Rotterdam. This is likely to be caused by the missing data in this AIS data set. Therefore, in the remainder of this research we decide to use the interarrival time distribution from the port call data estimates which does include all movements which are missing in the AIS data. This interarrival time distribution is shown in Figure 5b and seems much more representative of the expected interarrival times. In a similar way, if we would be interested in the distribution of the duration of terminal activities, a better result is reached if we use the port call data to create this empirical distribution.



(a) EPDF interarrival times from AIS data



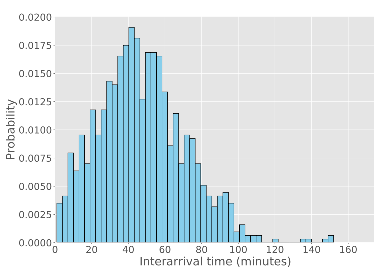
(b) EPDF interarrival times from port call data

Figure 5: Empirical probability density function of interarrival times in minutes from different data sources: AIS data and port call data.

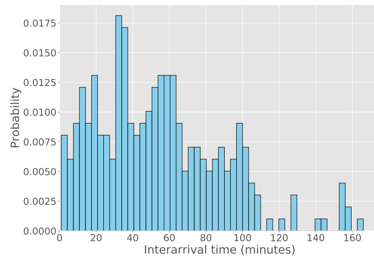
The distributions of the tugboat and pilot service times are shown in Figures 6 and 7. The accuracy of all six of these distributions was confirmed by the Port of Rotterdam authority.

From this information we conclude that the missing data does not seem to influence the tugboat and pilot service time distributions. More specifically, this observation supports our earlier assumption that the AIS data set is an unbiased sample of the ship location data.

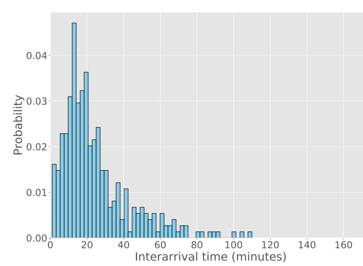
If we compare the arrival and departure tugboat service time distributions in Figures 6a and 6c, the most notable difference is that the average duration of a departure service is significantly shorter than the average duration of an arrival service. This can be explained by the fact that connecting the towing lines is more difficult and therefore takes longer when outside of the port compared to when at the berth. As shown in Figure 6b, the service durations for shift movements are more uniformly distributed. An explanation for this is given by the fact that shifts can cross the full port, but they can also just be a move of 100 meters to the other side of the waterway or any distance in between. Thus, a more uniform distribution is an intuitive result.



(a) Service times on arrival of a vessel



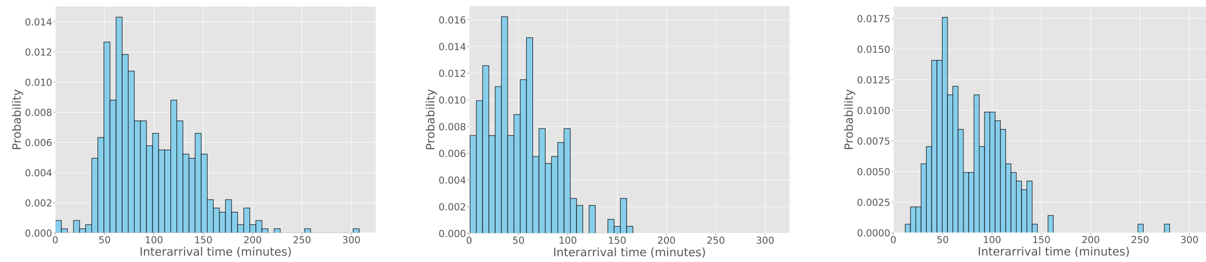
(b) Service times on shift of a vessel



(c) Service times on departure of a vessel

Figure 6: Empirical probability density function of tugboat service times for arrival, shift and departure movements. All service times in minutes.

Lastly, for the pilot service duration distributions shown in Figure 7 we again see that the average departure service time is a little shorter than the average arrival service time. Moreover, in both the arrival service time distribution (Figure 7a) and the departure service time distribution (Figure 7c) we see two peaks. These two peaks can be explained by the fact that there are two main destinations for ships in the port, it is either the Maasvlakte and Europoort area close to the port entrance, or the Botlek and City Centre area further away from the port entrance. Therefore, the pilots either serve a ship only for a short period of



(a) Service times on arrival of a vessel

(b) Service times on shift of a vessel

(c) Service times on departure of a vessel

Figure 7: Empirical probability density function of pilot service times for arrival, shift and departure movements. All service times in minutes.

time or for a time which is significantly longer. This difference is not visible in the tugboat service time distributions as in the case of a destination more inwards of the port, the tugboats connect further into the canal instead of at the port entrance. The pilot shift service time distributions in Figure 7b are based on the tugboat start and end times as explained earlier and are therefore the same as the tugboat shift service time distributions.

8 Results

Unless it is stated otherwise, the computations in this research were carried out on a high performance computer with an Intel Xeon E5-2699 processor with 512GB installed memory and programmed in Python. For the mathematical optimization problems the Gurobi solver was used.

8.1 Model instance

Given that the Port of Rotterdam has not suggested any specific locations to fill the set of potential resting locations, we decide to split the Port of Rotterdam into 32 areas based on port infrastructure and area size as shown in Figure 8. The sets of demand and return points are also filled with these 32 areas. To calculate the distance traveled between two areas we take the average over the distance between all combinations of berths in the two areas. This means that the distance traveled within an area is in general not zero since it is common for multiple berths to be in one area. Figure 8 also shows the current locations of the resting facilities, these are located in areas 4, 6, 15, 25 and 28. Because there is currently a total of five resting locations in the port, the parameter a in our model is set to five.

The other parameter settings are as follows, the transportation costs between areas are set equal to the distance between the areas. As the missing AIS data prevents us from getting the demand and return rates from the AIS data, the demand and return rates per day are taken from discrete empirical distributions fitted to the port call data. In choosing our method we assumed that the uncertain parameters are correlated as the demand and return rates are in practice connected with each other by the towing service of the tugboats. However, before applying the proposed methods to the practical case we test for correlation between the parameters to ensure that our intuition is in fact correct.

We perform the Pearson's r test for correlation with a significance level of $\alpha = 0.05$ for each of the pairs of parameters. We set the null hypothesis to be zero correlation, meaning that the parameters are independent. We test this against the alternative hypothesis that param-

Figure 8: The 32 areas in the Port of Rotterdam which occupy both the set of potential resting locations and the set of demand and return points. The red crosses indicate the current locations of the resting facilities.

eters are dependent and therefore the correlation is not zero. Figure 9 shows the heatmap of the correlation, where a lighter color corresponds with a larger correlation. If an area is not included in the heatmap, the demand and return rates in that area are zero.

From this we get an indication that at least for a few pairs of parameters there is some correlation. In particular, it is clearly visible that the demand and return rates of region 1 are correlated with the demand and return rates of most of the other regions. This is intuitive since many arrival and departure movements start or end at region 1 and move to or from one of the regions inside the port. From the results of the Pearson's r tests, we do in fact find that for 38 parameter pairs the p-value of the test is smaller than α and therefore the correlation is significant. In conclusion, as expected we cannot treat the uncertain demand and return parameters as being independent. We are aware that performing many statistical tests simultaneously in this way poses the risk of negatively influencing the quality of the conclusion. However, as our conclusion supports a more general solution approach which is also appropriate for uncorrelated parameters, if the conclusion turns out to be incorrect this has no further consequences for the problem solution.

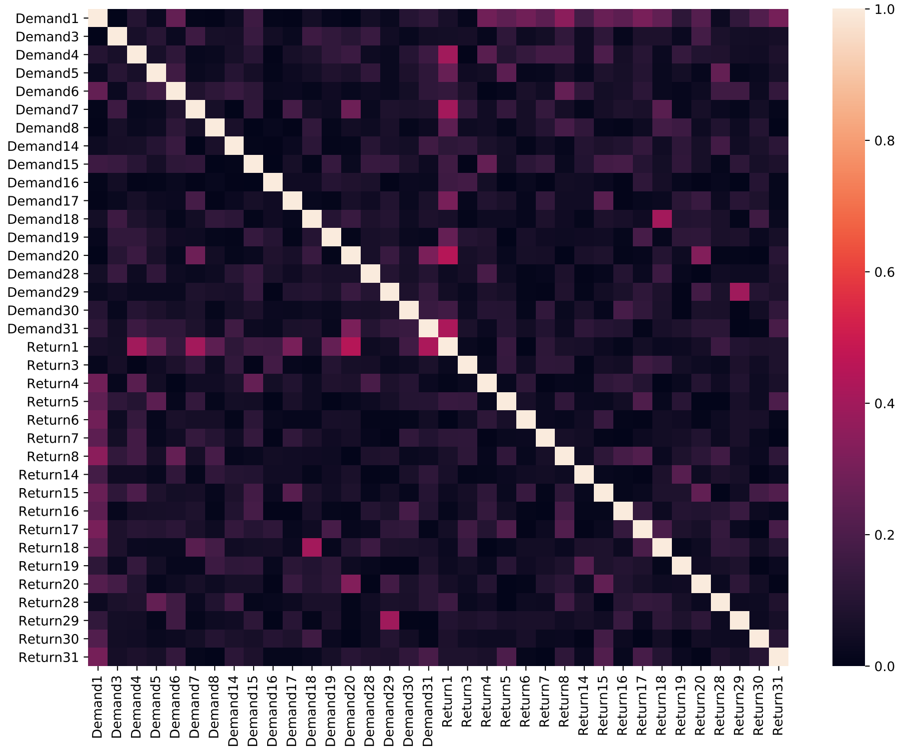


Figure 9: Heatmap of correlation between uncertain demand and return parameters. A lighter color indicates a larger degree of correlation.

8.2 Upper bound, lower bounds and stochastic model solution

Next, we calculate the upper bound EEV and the lower bound WS for the stochastic programming problem. The objective values are given in Table 6. Using these two bounds, the optimality gap has a value of 12.42, sometimes referred to as the value of solving the stochastic problem. As the problem can be rewritten to have fixed recourse and to be convex in the uncertain parameters, the lower bound can be tightened to the objective value of the expected-value problem EV. As shown in Table 6, the value of this lower bound equals the value of the upper bound EEV and therefore we can conclude that the objective value and the solution of the real stochastic problem should be equal to those of the upper bound. This optimal solution is equal to positioning the resting locations in areas $6, 15, 17, 20$ and 30.

Even though this is not necessary, for completeness we solve the real stochastic problem as well. Again, the results are shown in Table 6. As expected, the objective value of the real problem equals those of both the lower and the upper bound. Furthermore, we notice that solving the real problem requires a larger computation time, which is again an intuitive result and the main reason to first calculate an upper and a lower bound. We also see that the expected value problem solves extremely fast compared to the three other problems. This is caused by the fact that this is the only formulation where not all scenarios are used and instead the average over the scenarios is taken.

	$\boldsymbol{WS}$	$\boldsymbol{EV}$	$\boldsymbol{SP}$	$\boldsymbol{EEV}$
Objective value (km)	281.45	293.87	293.87	293.87
Run time (s)	167.16	0.55	195.92	118.87

Table 6: Objective values (km) and run times (s) for respectively the wait-and-see problem (WS), the expected value problem (EV), the actual stochastic problem (SP) and the upper bound (EEV) given by using the first-stage solution to the expected value in the stochastic problem. Computations were carried out on a on an Intel Core i5 processor with 8GB installed memory.

If we compare the found optimal resting locations with the current resting locations as shown in Figure 8, this found solution is in general relatively close to the current situation. There are however two significant changes, namely the location in area 4 is removed and replaced by a location in area 17. Furthermore, the resting location from area 25 is moved to area 20. Both these changes are rather intuitive as they make the distribution of locations over the port more even and therefore ensure less kilometers traveled.

8.3 Comparison by simulation

We compare the solution of the stochastic programming problem with the current resting locations by implementing a discrete-event simulation model. To ensure fair comparison of all computational results, we set the same seed for the uncertain parameters for each run of the simulation model. We first need to decide on a good length for the warm-up period of the simulation. We therefore run the simulation model for 20 days to discover at which point the performance measures per day do not change anymore. The results of this investigation can be found in Figure 10. It is quite surprising from these results that the simulation does not seem to need any warm-up period as the kilometers traveled and jobs delayed are not increasing or decreasing in the first few days. Nevertheless, a warm-up period is a safe choice to ensure that the performance measures are not influenced by the simulation start up time. For this reason, to be certain we set a warm-up period of five days.

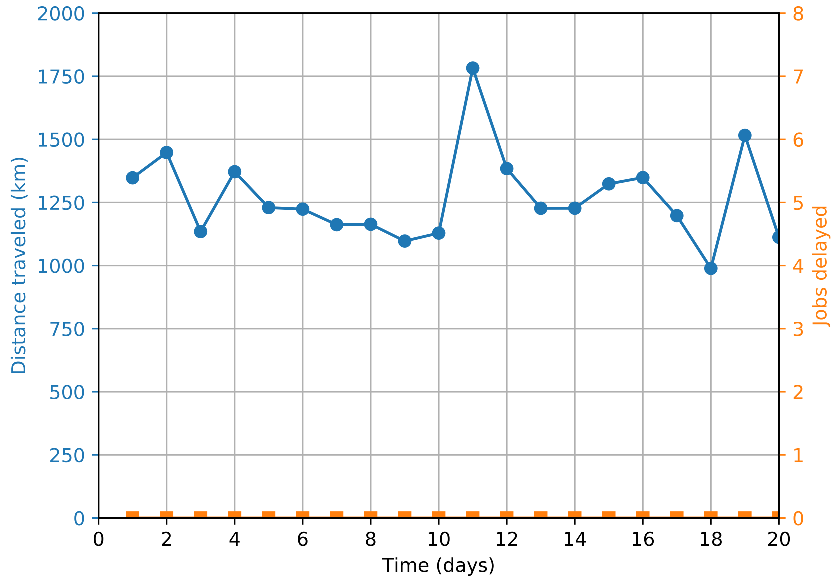


Figure 10: Performance measures (distance traveled and jobs delayed) per day in the first 20 days of the simulation. The blue line (circles) represents distance traveled, the orange line (squares) represents jobs delayed.

	Current locations	**Optimized locations**	**Saving**
Areas	4, 6, 15, 25, 28	6, 15, 17, 20, 30	
Jobs delayed	4	0	4
Kilometers traveled	38,029.35 km	37,213.59 km	815.76 km
Kilometers traveled transit	22,286.29 km	21,444.22 km	842.07 km
Transit fuel consumption (l)	276,951.73l	266,487.32l	10,464.41l
Transit fuel cost (€)	€141,522.33	€136,175.02	€5,347.31
Percentage saving			3.8%

Table 7: Comparison of monthly simulation results between the current tugboat resting locations and the resting locations optimized with the stochastic model. The two are compared on jobs delayed and kilometers traveled and kilometers traveled in transit, which translates into estimated fuel consumption and cost for the tugboat companies. Finally, the percentage saving is given.

We run the simulation for both the current and the suggested resting locations. From the simulation model we then get the number of jobs delayed and the number of kilometers traveled by all tugboats both in total and in transit. The term in transit refers to the tugboats traveling when not towing any vessel. The results of this comparison by simulation are shown in Table 7. To emphasize the effects of this change in resting locations, we also give an estimate of the savings in fuel consumption and fuel costs based on fuel consumption of a commonly used tugboat type and average tugboat speed in transit of 7.6 knots.

In the current situation we had to delay four jobs in the simulation over 30 days, whereas in the optimized situation it was possible to perform all jobs on time. Moreover, changing to the new situation in the simulation corresponds to traveling almost 850 kilometers less, consuming an estimated 10464 liters of fuel less and saving over €5,000 per month which corresponds to 3.8% of the total transit fuel cost. That is, on a yearly basis the tugboat companies can achieve an estimated saving of almost €65,000.

9 Discussion

In this thesis research we found several results on using AIS data to describe the nautical chain, we were able to optimize the tugboat resting locations and we compared the current and optimized locations with a simulation model. This chapter discusses the found results, their implications for research and practice, their limitations and several possibilities for future research in the field.

AIS data is generally portrayed as the answer to describing any maritime problem regarding visualization of a process and forecasting ship movements. Also for this research the expectation was that AIS data would give us all the information necessary to properly map the pilot and tugboat processes of the Port of Rotterdam. However, as we soon discovered, the AIS data set is not as clean and complete as expected. As stated, roughly 20 percent of the port movements is missing from the data without a clear reason or pattern. Because of this randomness, the most likely cause of the missing data are errors in the transmission or the receiving of the location data to the data set.

The fact that this data is missing has consequences for the results of this research in multiple ways. For the event detection, the black holes in the location data of the deep-sea vessels make it more difficult to match a tugboat or pilot tender to the vessel, possibly causing errors in the matching values. Moreover, there might be black holes in the tugboat and pilot AIS data, which means that a service vessel might not be visible in the data set at the moment when it is providing service to a vessel. This might again cause less accurate matching values, which also affects the number of tugboats that we can match to the target vessels. This can be an explanation for the difference in number of tugboats per vessel in Table 5. Lastly, because of the missing data we were unfortunately not able to optimize the resting locations based on AIS data. Instead, we had to rely on the manually entered port call data set. Therefore, if the Port of Rotterdam want to work on more research based on the AIS data, it would be advisable to first investigate the exact cause of the missing data and, if possible, resolve these issues. Still, based on the global investigation of the missing data done in this thesis research, it seems reasonable to assume that the information gaps are randomly distributed over the data and therefore they should not affect the statistical findings in this research significantly. This remark is confirmed by the affirmation of the port authority that the statistical distributions of both locations and service duration are as expected.

As stated, the second important finding in this thesis research are the statistical results extracted from the AIS data. These results include, but are not limited to, the meeting locations of deep-sea vessels and pilot tenders, start and end locations of tugging service for the different destinations inside the port and service duration distributions for both pilots and tugboats. These results are especially relevant to the further research in the Swarmport project to map the nautical chain. That is, with these results the other researchers are able to move from a highly unrealistic simulation model with parameters based on guessed values to a much more realistic model with parameters based on real data values. This realistic simulation model can then be applied to test the effect of certain measures on the performance of the chain and to show these effects to the actors in the chain in order to get them on board with the project and possible measures being taken.

Another request from the port authority was to investigate what the best locations in the port would be to position resting locations for tugboats. We found that some of the current locations are already in a good position while some others are best to be located differently. This result is based purely on optimizing the kilometers traveled between the start or end locations of a job and the nearest resting location. This criterion was chosen based on the very limited information we had on the tugboat process. Our data did not include any information on the fleets of the three competing tugboat companies in the port, neither did it tell us which of the resting locations are rented out to which of the tugboat companies. If the port authority is willing to provide this information in the future, this could improve the practical relevance of the optimization significantly as we can then take into account additional constraints to prevent boats from other companies to rest at the resting location of one company. Additionally, we could add constraints to make sure that the locations of competing companies are positioned in such a way that they have equal chances of being chosen by a vessel operator to perform the towing job. On the other hand we know that shared resting locations between companies are currently already implemented at some locations and the port might consider extending this concept to all resting locations in the future. If so, the suggested optimized locations are a good proposal for the port authority to consider.

Lastly, in this thesis research we created a discrete-event simulation model to compare the jobs delayed, kilometers traveled and fuel consumption between different tugboat locations. Using this simulation model we find that a switch from the current resting locations to those found by the optimization results in a saving of 842.07 kilometers and an estimated 10,464.41 liters of fuel per month. This saving in liters of fuel equals a saving in costs for the tugboat companies of €5,347.31. As stated, the saving in liters of fuel is an estimate based on the average fuel consumption of a commonly used tugboat type in the Port of Rotterdam. In the simulation model, a greedy dispatching algorithm is used to assign tugboats to vessel movements. Because of economical considerations of the tugboat companies they have no interest in providing us with any information on the dispatching strategies used by them. Therefore, we cannot determine whether or not the greedy dispatching algorithm is in fact representative of the strategies used in practice.

This uncertainty provides two possible directions for future research. First, a future research could be focused on identifying the real dispatching strategies used by the tugboat companies by studying the AIS data even closer. Second, the effect of the dispatching strategy on the savings results of the simulation could be investigated by implementing one or more complex dispatching strategy such as a dynamic tabu search algorithm and comparing the savings results for each of the strategies. As for the first research it has to be noted that given the 20% of missing data and some errors in the data, it is questionable whether enough data remains to accurately identify the dispatching strategy used. Therefore, it might be preferable to first investigate the effects of changing the dispatching strategy on simulation results before embarking on the first suggested research.

10 Conclusion

This thesis research was involved with two research questions. In the main research question we investigated the most efficient positioning of tugboat resting locations in the Port of Rotterdam. We planned to answer this question using real ship location data, which leads us to the underlying research question of this thesis where we analyzed whether Automatic Identification System (AIS) data can accurately describe the nautical chain. In the AIS data analysis, we found that roughly 20% of the AIS data is missing, which was an unforeseen finding that influenced the further course of this research. Nevertheless, we were able to find tugboat start and end locations and service durations as well as pilot meeting locations and service durations from the AIS data, all of which were found to be plausible by the port authority of the Port of Rotterdam. This confirms that the methods used to match the location data of vessel movements to tugboats and pilot tenders are accurate for answering our underlying research question regarding AIS data.

For the main research question we were able to find areas to position the tugboat resting locations such that the distance they need to travel between these resting locations and their jobs was minimized using a two-stage stochastic programming model. Based on the AIS data analysis we decided to use empirical distributions for the demand and return rates of tugboats retrieved from the port call data instead of the AIS data. We calculated upper and lower bounds on the optimal objective value of the stochastic program and as a solution method we rewrote the program to a one stage stochastic programming problem. We found that the lower bound, upper bound and optimal objective value of the problem instance are all equal. Finally, we investigated to what extent the suggested locations lead to a decrease in kilometers traveled and jobs delayed by the tugboats through a discrete-event simulation using a greedy dispatching algorithm. From this comparison we found that the suggested locations yield less delayed jobs and a saving in kilometers traveled of roughly 4% compared to the current locations, which is equivalent to an estimated fuel cost saving of almost €65,000 per year.

This is the first research where AIS data is used as a tool to better describe and analyze the nautical chain inside a port. Therefore, even with the limitations of missing data, it provides important new insights into the applicability of AIS data to analyze processes inside the Port of Rotterdam. Additionally, the distributions that were found from the AIS data are valuable information for the future research in the Swarmport project. The project team will be able to create a more accurate representation of the real nautical chain processes in the Port of Rotterdam, which can help to improve the cooperation between the different actors in the nautical chain. Moreover, we investigated the possibilities for using AIS data as input for mathematical optimization of the nautical chain, which opened the door for possible future research in this area.

The limitations of this research offer several directions for future research. Firstly, this thesis only scratches the surface of the possible applications of AIS data within the port. By using more complex machine learning techniques one might for example be able to identify the tugboat dispatching strategies currently used by the tugboat companies or one might be able to identify bottlenecks in the nautical chain. Secondly, to make the optimization model an even more accurate representation of reality, one might consider taking into account the restrictions which are specific to the tugboat companies, such as which resting locations are theirs and restrictions on choosing locations in a more competitive manner. Lastly, to make the comparison by simulation more representative it would be good to investigate the effects of changing the dispatching strategy from a greedy algorithm to more complex strategies such as a dynamic tabu search algorithm.

To conclude, we found that AIS data can be used to describe service locations and times for both tugboat and pilot processes rather well, but because of missing data it cannot be used to accurately describe interarrival times. Furthermore, we established an improved assignment of tugboat resting locations in the Port of Rotterdam by applying a stochastic two-stage programming model. Additionally, we created a discrete-event simulation model to easily compare different solutions to the tugboat resting location problem. We found that the optimized locations result in a decrease in delayed jobs and a significantly lower number of kilometers traveled and as a consequence lower fuel costs for the tugboat operators.

References

Amin, S. H., Zhang, G., 2013. A multi-objective facility location model for closed-loop supply chain network under uncertain demand and return. Applied Mathematical Modelling 37 (6), 4165–4176.

Bailey, N. J., Ellis, N., Sampson, H., 2008. Training and technology onboard ship: how seafarers learned to use the shipboard automatic identification system (AIS). Seafarers International Research Centre (SIRC), Cardiff University.

Bentley, J. L., 1975. Multidimensional binary search trees used for associative searching. Communications of the ACM 18 (9), 509–517.

Carøe, C. C., Schultz, R., 1999. Dual decomposition in stochastic integer programming. Operations Research Letters 24 (1-2), 37–45.

De Boer, T., 2010. An analysis of vessel behaviour based on ais data. Master's thesis, Delft University of Technology.

Dobrkovic, A., Iacob, M.-E., van Hillegersberg, J., Mes, M. R., Glandrup, M., 2016. Towards an approach for long term ais-based prediction of vessel arrival times. In: Logistics and Supply Chain Innovation. Springer, pp. 281–294.

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.

Higle, J. L., 2005. Stochastic programming: optimization when uncertainty matters. In: Emerging Theory, Methods, and Applications. Informs, pp. 30–53.

Madansky, A., 1960. Inequalities for stochastic linear programming problems. Management science 6 (2), 197–204.

Melo, M. T., Nickel, S., Saldanha-Da-Gama, F., 2009. Facility location and supply chain management–a review. European journal of operational research 196 (2), 401–412.

Mou, J. M., Van Der Tak, C., Ligteringen, H., 2010. Study on collision avoidance in busy waterways by using ais data. Ocean Engineering 37 (5-6), 483–490.

Murray, W., 2015. Economies of scale in container ships costs. United States Kings Point Merchant Marine Academy.

Nikghadam, S., Rezaei, J., Tavasszy, L. A., 2019. Relationships between actors of the nautical chain: Strength and symmetry. The Nordic Logistics Research Network conference. Oslo.

Pallotta, G., Vespe, M., Bryan, K., 2013. Vessel pattern knowledge discovery from ais data: A framework for anomaly detection and route prediction. Entropy 15 (6), 2218–2245.

Robards, M., Silber, G., Adams, J., Arroyo, J., Lorenzini, D., Schwehr, K., Amos, J., 2016. Conservation science and policy applications of the marine vessel automatic identification system (ais)—a review. Bulletin of Marine Science 92 (1), 75–103.

Santoso, T., Ahmed, S., Goetschalckx, M., Shapiro, A., 2005. A stochastic programming approach for supply chain network design under uncertainty. European Journal of Operational Research 167 (1), 96–115.

Shelmerdine, R. L., 2015. Teasing out the detail: How our understanding of marine ais data can better inform industries, developments, and planning. Marine Policy 54, 17–25.

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.

Tavasszy, L., Minderhoud, M., Perrin, J.-F., Notteboom, T., 2011. A strategic network choice model for global container flows: specification, estimation and application. Journal of Transport Geography 19 (6), 1163–1172.

Verduijn, S., 2017. Identifying the relations between and mapping the processes of the nautical service providers in the port of rotterdam. Master's thesis, Delft University of Technology.

11 Appendix

Several papers, such as Madansky (1960) have proven in the past that under specific conditions, the Wait-and-See problem provides a tighter lower bound on the optimal objective value than the expected value problem. Here, we give a counterexample to show that for the current problem the Wait-and-See problem does not by definition give a tighter lower bound than the expected value problem. In the example we use the notation from Chapters 4 and 5.

We start by setting $I = J = \{1, 2\}$ and for simplicity we work with two scenarios α and β with equal probability. We choose the demand and return parameters to be $d_1^\alpha = r_1^\alpha = 4$, $d_2^\alpha = r_2^\alpha = 0$, $d_1^\beta = r_1^\beta = 0$ and $d_2^\beta = r_2^\beta = 8$. From which we know that the expected values of the demand and return parameters are $\bar{d}_1 = \bar{r}_1 = 2$ and $\bar{d}_2 = \bar{r}_2 = 4$. Moreover, the distance parameters are set to $k_{11} = 1$, $k_{12} = 7$, $k_{21} = 7$ and $k_{22} = 1$. Lastly, we set the number of facilities to be opened to $a = 1$.

With these parameter settings, we can write the Wait-and-See problem in the following way for scenario α

$$\begin{aligned}
\min \quad & 4u_{11}^\alpha + 28u_{12}^\alpha + 4v_{11}^\alpha + 28v_{12}^\alpha && (60)\\
\text{s.t.:} \quad & u_{11}^\alpha + u_{12}^\alpha = 1 && (61)\\
& u_{21}^\alpha + u_{22}^\alpha = 1 && (62)\\
& v_{11}^\alpha + v_{12}^\alpha = 1 && (63)\\
& v_{21}^\alpha + v_{22}^\alpha = 1 && (64)\\
& u_{11}^\alpha, u_{21}^\alpha \le z_1^\alpha && (65)\\
& u_{12}^\alpha, u_{22}^\alpha \le z_2^\alpha && (66)\\
& v_{11}^\alpha, v_{21}^\alpha \le z_1^\alpha && (67)\\
& v_{12}^\alpha, v_{22}^\alpha \le z_2^\alpha && (68)\\
& z_1 + z_2 = 1 && (69)\\
& u_{11}^\alpha, u_{21}^\alpha, u_{12}^\alpha, u_{22}^\alpha, v_{11}^\alpha, v_{21}^\alpha, v_{12}^\alpha, v_{22}^\alpha \in \mathbb{B} && (70)\\
& z_1^\alpha, z_2^\alpha \in \mathbb{B}. && (71)
\end{aligned}$$

We can write the Wait-and-See problem for scenario β in a very similar way. These problems yield the optimal objective values $WS_\alpha = 8$ and $WS_\beta = 16$, from which we know that the optimal objective value of the Wait-and-See problem is $WS = 12$. On the other hand, the expected value problem can be written as

$$
\begin{aligned}
\min \quad & 2u_{11} + 14u_{12} + 28u_{21} + 4u_{22} + 2v_{11} + 14v_{12} + 28v_{21} + 4v_{22} && (72)\\
\text{s.t.:} \quad & u_{11} + u_{12} = 1 && (73)\\
& u_{21} + u_{22} = 1 && (74)\\
& v_{11} + v_{12} = 1 && (75)\\
& v_{21} + v_{22} = 1 && (76)\\
& u_{11}, u_{21} \leq z_1 && (77)\\
& u_{12}, u_{22} \leq z_2 && (78)\\
& v_{11}, v_{21} \leq z_1 && (79)\\
& v_{12}, v_{22} \leq z_2 && (80)\\
& z_1 + z_2 = 1 && (81)\\
& u_{11}, u_{21}, u_{12}, u_{22}, v_{11}, v_{21}, v_{12}, v_{22} \in \mathbb{B} && (82)\\
& z_1, z_2 \in \mathbb{B}. && (83)
\end{aligned}
$$

The optimal objective value of this expected value problem is then $EV = 36$. As this value is larger than the optimal objective value of the Wait-and-See problem, we have shown that there is at least one instance of the problem where $EV > WS$ holds and subsequently that $EV \leq WS$ does not hold for the current optimization problem.