Master Thesis

Port Call Efficiency Optimization, Using Data Analysis, Process Mining and Discrete Event Simulation

MSc Programme Engineering and Policy Analysis - TU Delft

Matti Mašović
27-8-2019

Port Call Efficiency Optimization, Using Data Analysis, Process Mining and Discrete Event Simulation

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in Engineering and Policy Analysis

Faculty of Technology, Policy and Management

by

Matti Mašović

Student number: 4231317

To be defended in public on August 27th 2019

Graduation committee

Chairperson	: Prof.dr.ir. A. Verbraeck, Multi-Actor Systems (Policy analysis)
First Supervisor	: Dr.ir. I. Lefter, Multi-Actor Systems (Systems Engineering)
Second Supervisor	: Prof.dr.ir. A. Verbraeck, Multi-Actor Systems (Policy analysis)
External Supervisor	: Dr.ir. A. M. P. de Leege, Port of Rotterdam (Senior Data Scientist)

Acknowledgements

A master thesis is the final requirement for the MSc degree in Engineering and Policy analysis of the faculty of Technology, Policy and Management at the TU Delft. I have worked on this thesis for almost six months, during the second semester of the academic year 2018-2019. I am proud to present to you my final work.

I would first like to thank my thesis supervisors Dr.ir. I. Lefter and Prof.dr.ir. A. Verbraeck, for their valuable feedback and guidance. They steered me in the right direction throughout the entire process of writing this thesis. Secondly, I would like to express my sincere appreciation towards the entire Digital Business and Solutions (DBS) team of the Port of Rotterdam. I was offered to do an internship at DBS, where I was allowed to freely work on this research and use all their resources and expertise. Not only, did this help in writing my thesis, it also gave me a good insight of an office environment. A special thanks goes out to Dr.ir. A. De Leege, who was my external supervisor from the Port of Rotterdam. He was my mentor during the internship and invested a lot of time in me.

I hope you will enjoy reading this,

Rotterdam, August 27th, 2019

Matti Mašović

Disclaimer

Prior to conducting this research, a confidentiality agreement has been signed by the writer, with the goal to prevent sensitive information from leaking. This resulted in two reports: an uncensored private version, meant for the supervisors and the Port of Rotterdam and a censored public version. This is the public version. All the information that is deemed to be confidential is removed from this report. Therefore, note that some figures, tables and parts of the text are missing.

Executive Summary

The Port of Rotterdam (PoR) is the fourth biggest port in the world and the biggest in Europe. In order to maintain its leading position, the PoR strives to constantly improve its attractiveness. Port call efficiency (PCE) is a combination of the duration of a ship's visit in a port, the quality of cargo handling and the quality of services. This efficiency has a significant influence on the attractiveness of a port and thus is a determining factor in the port choice of shipping lines. PCE optimization is already a huge theme in the Port of Rotterdam. The goal of this research is to contribute to the current work and knowledge by applying relative new techniques on port call related data. A combination of process mining (PM) and discrete event simulation (DEVS) is explored, to determine how they can contribute to identifying and assessing policies that improve port call efficiency. However, first, a qualitative analysis, based on expert consultation and a more traditional data analysis of the event data is performed. This is done with the goal to get a clear picture of the port's operations and the actors that are involved.

The idea of process mining is to discover, monitor and improve real processes by extracting knowledge from event logs. The logs, that are analysed in this research, contain records of over 300.000 port call related events that occurred in the PoR in 2018. The logs are imported in the process mining software Disco, which automatically generates a model that provides a large variety of insights. To explore the applications of these insights in a structured way, a framework, created by a IEEE taskforce for process mining, is used. It is concluded that there are multiple ways to benefit from process mining when analysing event data in the maritime transport domain. First of all, process mining makes it possible to map and monitor a port's behaviour with less effort than through a more traditional data analysis. The clear empirical insights that PM provides, makes it a smooth process to spot bottlenecks in the port, from which policies can be derived, with the potential to improve the PCE. After comparison with the port call process map, that resulted from a qualitative analysis, it is concluded that the quantitative PM method is successful in finding a similar result. Thus, local expertise is not essential to get a grip on the process of a port call, when using this method. Furthermore, PM has proven to be an effective technique for conformance checking of the event engine. PM is an excellent tool for creating a good overview, seeing how the events relate to each other and spotting shortfalls in this complex system. However, when implementing PM, a high level of alertness is required. The reliability of the insights that process mining provides is directly related to the quality of the data. Therefore, being aware of this quality is essential.

To find a way to assess the identified measures that have the potential to improve the port call efficiency, discrete event simulation is explored. More concretely, a simulation model is built of the Port of Rotterdam, which simulates the port and all the seagoing vessels that visited it in 2018. This model simulates terminals and the use of nautical services like tugboats and pilots. This research shows and concludes that it is possible to create a simulation model based on event data and a model structure derived from PM, that is valid enough to perform relative comparisons of scenarios. After identifying a policy based on inefficiencies observed in the PM phase, it is implemented in the model and compared to the current situation. A policy is tested in which the two tug companies, that are active in the PoR, operate as one fleet instead of two. However, other policies can already be implemented in the model as well. It is concluded that the model is able to simulate scenarios that include a change in the number of vessels, terminals and pilots. The model is built in a modular way, which makes it possible to extend the model even further which enables the simulation of other policies that are not yet included. Well defined key performance indicators show the significant effects of a policy on the port. For example, in the tested policy, the number of anchorage visits in one year dropped from 2313 to 1994. This shows that the policy, where all the tugboats in the port operate as one fleet, has potential to improve the port call efficiency. This way a measure can be assessed through simulation in a quantitative and very structured way, before it is implemented in the real world, even though the modelled system is a complex environment.

This research identifies multiple applications of process mining. Some are explored in a hands on quantitative way, which provided useful insights, for other applications it is only described how they may provide value. Further research is needed to determine the exact benefits of this second group of process mining applications. More concretely, it is recommended to explore the use of process mining as a validation tool for the simulation model.

A simulation model is a simplification of the reality. These simplifications are a direct source of inspiration for future work. The next logical step is to increase the complexity of the simulation by including the simulation of critical services at berth, like cargo handling and bunkering. This will increase the number of policies that can be tested in the model. Next, a research that explores more complex ways of mapping the behaviour and operations of pilots is recommended. The exact number of pilots and what permits they have, was an unknown fact in this research. Furthermore, it is suggested to extend the model so that it is possible for a vessel to visit different terminals in one port call.

Next, it is recommended to keep optimizing the event data. All the quantitative analyses in this research are event data driven. The reliability of these methods is directly related to the quality of data. Expanding the number of unique events is also suggested. More concretely, it is recommended to explore the possibilities of identifying events that specify when a delay begins and ends. Now it is rather complex to identify the delays of vessels based on the data, especially the non-container vessels. Having events of delays would enable a more precise calibration of the simulation model.

Table of contents

List of Figures

List of Tables

List of Abbreviations

AHP	Analytic Hierarchy Process
AIS	Automatic Identification System
ATA	Actual Time of Arrival
ATD	Actual Time of Departure
DEVS	Discrete Event System Specification
ETA	Expected Time of Arrival
ETD	Expected Time of Departure
IMO	International Maritime Organization
KPI	Key Performance Indicator
MMSI	Maritime Mobile Service Identity
PCC	Port Coordination Centre
PCE	Port Call Efficiency
PM	Process Mining
PoR	Port of Rotterdam
PTD	Planned Time of Departure
UCRN	Unique Port Call Reference Number
VTS	Vessel Traffic Service

1. Introduction

This chapter shows the relevance of this research and describes how this research is conducted. First the case study is introduced which leads to the main research question. This is followed by the sub-questions, that together answer the main research question. Lastly the methods, used to answer these questions are discussed and the structure of this study is shown.

1.1 Introduction to case study

The Port of Rotterdam (PoR) is the fourth biggest port in the world and the biggest port in Europe (Kiprop, 2018). In 2017, the Port of Rotterdam processed over 467 million tonnes of dry bulk, liquid bulk, breakbulk and containers (Port of Rotterdam, 2018), which shows that they are a significant player in the world market for sea cargo transportation.

The PoR is not the only port in Western Europe that strives to be the most successful one. This position is being contested continuously. The two competitors are the Port of Antwerp, which is the second biggest port of Europe, with a throughput of 214 million tonnes (Port of Antwerp, 2018) and the Port of Hamburg, with a throughput of 136.5 million tons (Port of Hamburg, 2018). Therefore, it is necessary for the Port of Rotterdam to constantly improve its attractiveness for cargo shipment companies, in order to maintain its leading position.

Port efficiency has a significant influence on the attractiveness of a port and thus is a determining factor for shipping lines when choosing which port to operate in. This is shown in multiple studies that each identify efficiency related factors as significant factors in the port choice process: Brooks & Schellinck (2015) mention incidence of delay, Magala & Sammons (2008) mention efficiency and Piresda Cruz, Ferreira & Garrido Azevedo (2013) mention vessel turnaround time. Currently, it is difficult to define port efficiency, due to the non-universal definition of what indicates an efficient port (De Monie, 2009). The following definition by Kennedy et al. (2011) will be used in this research: The efficiency of a sea-port is determined by the duration of a ship's stay in a port, by the quality of cargo handling and by the quality of services. Optimizing the first component of the definition, the duration of a ship's stay in a port, is referred to as port call efficiency (PCE) and is the focus of this study. The Port of Rotterdam is used as a case study in this research.

A rising economy accompanied by a growing volume of mega-ships, increasing transhipment and a fragmented, complex feeder network, all contribute to Rotterdam's ongoing congestion problem (Knowler, 2018a). This leads to unreliable planning and additional costs (Parthenon. 2018), which does not contribute to a more attractive port. Up to five mega-ships, each handling more than 8.000 TEU, can call at the same time in the port (Knowler, 2018b). These peaks put huge pressure on the resources of the Port of Rotterdam, like tugboats, pilots and terminals. This can lead to delays up to 72 hours (Knowler, 2018a). Even if a big vessel is delayed for a few hours, it can result in delays of over a week, later on in the schedule (Jumelet & Van Strien, 2019). In order to keep growing and to maintain its strong position in the world market for sea cargo, the Port of Rotterdam identified port call efficiency optimization as one of the improvement areas of the port.

Port call efficiency optimization is already a huge theme in the Port of Rotterdam. The PoR has multiple departments that focus on PCE optimization. Many of those departments work on event data driven projects. Vast amounts of data is recorded and gathered. The data consists of events with a timestamp, that happen during a port call, for example the time of arrival at the terminal or anchorage. The event data is used to provide insights and to create products that contribute to a higher PCE. This research contributes to the current work and knowledge, by discovering the applications of the combination of process mining and discrete event simulation in a major port. It is concluded from the literature study (Appendix I) that this exact combination of methods is an unexplored area in this domain.

Therefore, the main goal of this this research is to explore how the port call efficiency in the Port of Rotterdam can be improved, through the applications of process mining and discrete event simulation in this domain. This will be done by first analysing the port call process and the involved actors in a qualitative way. This is followed by quantitative phase. Event data that describes all the port calls of 2018 in the Port of Rotterdam, is used as an input for all the quantitative parts of this research. The first one is the data analysis phase, which provides quantitative insights of the port's behaviour. Hereafter, the potentially beneficial applications of process mining in in this domain are explored. Also, process mining is used to extract a model structure from the event data. Next, a modular simulation model of the sea-side of the port is built, in Simio. The model structure, identified during the process mining phase and quantitative inputs, extracted during the data analysis phase, are used to create this simulation model. This model is capable of testing the effects of different policies, under different scenarios, on the performance of the port. To demonstrate this, a policy is identified and simulated and compared to the simulation of the current situation, to predict the effects on the performances of the port. The scenario that is tested, describes a situation where the tug companies, that are active in the PoR, collaborate in order to increase the efficiency of the entire tug fleet. Now, the tug companies do not collaborate with each other. This new policy is expected to reduce the delays caused by the lack of available tugboats, and thus improve the PCE. Finally, the possible implementation of this policy in the real world is discussed, taking the multi-actor environment into account, which is mapped in the first qualitative phase of this research. In order to achieve all this, the following main research question in formulated:

How can data analysis, process mining and discrete event simulation contribute to identifying and assessing policies that improve port call efficiency?

The next section discusses the identified sub-research questions, which together answer the main research question.

1.2 Sub-questions

This research will use the congestion of seagoing vessels on the water side of the Port of Rotterdam as a case study. The Port of Rotterdam main goal is to enhance the port's competitive position as a logistics hub and world-class industrial complex (Port of Rotterdam, 2019a). Even though this study is focussed on the Port of Rotterdam, this approach can be applied on any major port, with the condition that automatic identification system (AIS) derived event data is available.

First of all, before it is possible to analyse any processes in the port, it is essential to understand how the port functions. More specific, it is necessary to identify the main building blocks the port is made out of and to identify the most important resources the port has, regarding a port call. Furthermore, the exact process of how the cargo travels through the port and how the actors in the port interact must be clear. To get a grip on how the port works, the following question is answered:

What are the steps of a port call and which actors are involved?

The previous question is answered in a quantitative way (literature review and expert consultation) and in a quantitative way (data analysis). A more detailed description of the methods and tools can be found in chapter 1.3. A comparison of these two approaches, with a special focus on the process mining, will illustrate the added value of using process mining techniques with the purpose of improving the port efficiency. Further possible applications of process mining are explored following the framework of constructed by a process mining IEEE taskforce (Van der Aalst et al, 2011). This together answers the following question:

What is the added value of process mining in port call efficiency optimization?

Based on the identified processes in the port and quantitative insights on the behaviour of the harbour, a discrete event simulation model of the wet-side of the port is built, with the goal to demonstrate how to assess policies in this domain. In order to do so, a policy measure, that is expected to reduce vessel delays in the port and thus is expected to improve the port efficiency, is formulated. In this identified scenario, multiple tug companies assist each other when one of the companies is dealing with an undercapacity of available tugboats. This policy is then tested in the simulation model. The exact methods used for the construction of the simulation model are discussed in chapter 1.3. This what-if analysis will answer the following question, which shows how PCE improving policies in the harbour can be assessed through simulation:

What is the effect of the collaboration of tug companies on the delays of vessels in a port?

1.3 Methodology and tools

The methodology that will be used in the upcoming research can be divided into four segments: literature study, empirical data gathering and analysis, process mining and discrete event simulation. These will be individually discussed in this paragraph.

Literature review

A literature study is a search and evaluation of the available literature in the given subject (Royal Literary Fund, 2019). The purpose is to give a clear picture of the current, relevant knowledge in the described topic area. An effective review creates a firm foundation for advancing knowledge (Watson & Webster, 2002). The literature study is essential for setting the foundation for a scientifically relevant research, as has been done in Appendix I and for partly answering the first sub question. In order to understand the port call process in general and the processes Port of Rotterdam in specific, literature on this matter needs to be analysed. More specifically, knowledge on the relevant players and how they interact, is required. Furthermore, insights on the building blocks of the port and the internal processes is vital for the research as well. The information, gathered here, will set the rules of the environment, that this study will be set in. This literature will be gathered by searching platforms like Google Scholar (2019) and by expert consultation.

Empirical data gathering and analysis

In order to perform the data driven simulation, which is part of the quantitative analysis, and answer all the associated questions, a certain set of empirical data is required. Empirical data is information about something, gathered through observation (Roundy, 2019). The data required for this study needs to date back multiple months, preferably an entire year. Therefore, the data will be requested from the client, Port of Rotterdam Authority, who gathers AIS data, processes it and stores it in a cloud.

After the data is retrieved from the cloud using queries, it will be cleaned and filtered, to finally get the information needed for this research. This will be done using Python. Python is programming language, that can function as a good tool for data cleaning and analysis (Idris, 2014). This process is explained in more detail in chapter 3. This methodology will, together with the literature research, answer sub-question 1.

Process mining

Process mining is a technique that is used during the quantitative part of the research. The idea of process mining is to discover, monitor and improve real processes by extracting knowledge from event logs (Van der Aals et al., 2011). The more general possibilities of process mining in this research scope will be explored. Furthermore, it will be used to extract the model structure from the data for the simulation model. The process mining software, Disco (Fluxicon, 2019), will be used to execute this part of the analysis.

The exploration of this method will directly answer sub-question two. The result of this methodology will also be used as an input for the *Discrete Event System Specification* (DEVS) method, described below, which will directly answer sub question 3.

DEVS

The output from the empirical data analysis will be used as an input to build a valid discrete model that can simulate the effects of various scenarios and policies on the port efficiency. The methodology that will be used to build this model is DEVS (Hild, 2000). This methodology runs on a continuous time basis, however the variables in the model are subject to discrete changes in value. This means that the variables stay constant for a period of time and then jump to a new value. These changes in states are called events. The model specifies these events and the shifts of the states they cause. A simulator is used to execute the model (Hild, 2000). Given these properties, it can be concluded that this method is effective for simulating queues and delays and thus is fit for this project. The simulator that will be used to execute the model is called Simio. This tool follows the rules of the DEVS methodology and makes it possible to create a valid simulation of the system behaviour that can be used for performance optimization and decision-making (Simio, 2019). The result of the simulation will be a set of scenarios that influence port efficiency and a set of policies which will reduce the delays and thus improve the port efficiency. This output will directly give answers to sub-question 3.

1.4 Research Structure

Figure 1.1 shows the research flow diagram for this study. The research flow diagram is a visual representation of the internal structure of the research as described in this chapter. This report has the following structure: After introducing the problem, a literature research identifies the knowledge gaps, which lead to the main research question and its sub-questions. This is followed by the introduction of the methodology, where after a qualitative analysis is performed to get a clear picture of the processes and actors in the Port of Rotterdam. A data gathering and cleaning phase will be followed up by actual data analysis and the process mining phase. This will respectively result in inputs for the simulation model and a structure for the simulation model. At this point it is possible to answer the first two sub-questions. After the model conceptualisation, specification, creation, verification and validation, the policy is formulated. After testing the policy with collaborating tug companies in the model, the results are statistically compared and discussed. Next, the results from the statistical comparison, in combination with the previously executed actor analysis will lead to a policy advice in the recommendations. This policy advice will then not only be technically feasible, it will also take the multi-actor environment into account. Each phase of the research is followed by a discussion that answers the corresponding sub-question. The last part of this research shows a more generalized discussion and conclusion, which answers the main research question.

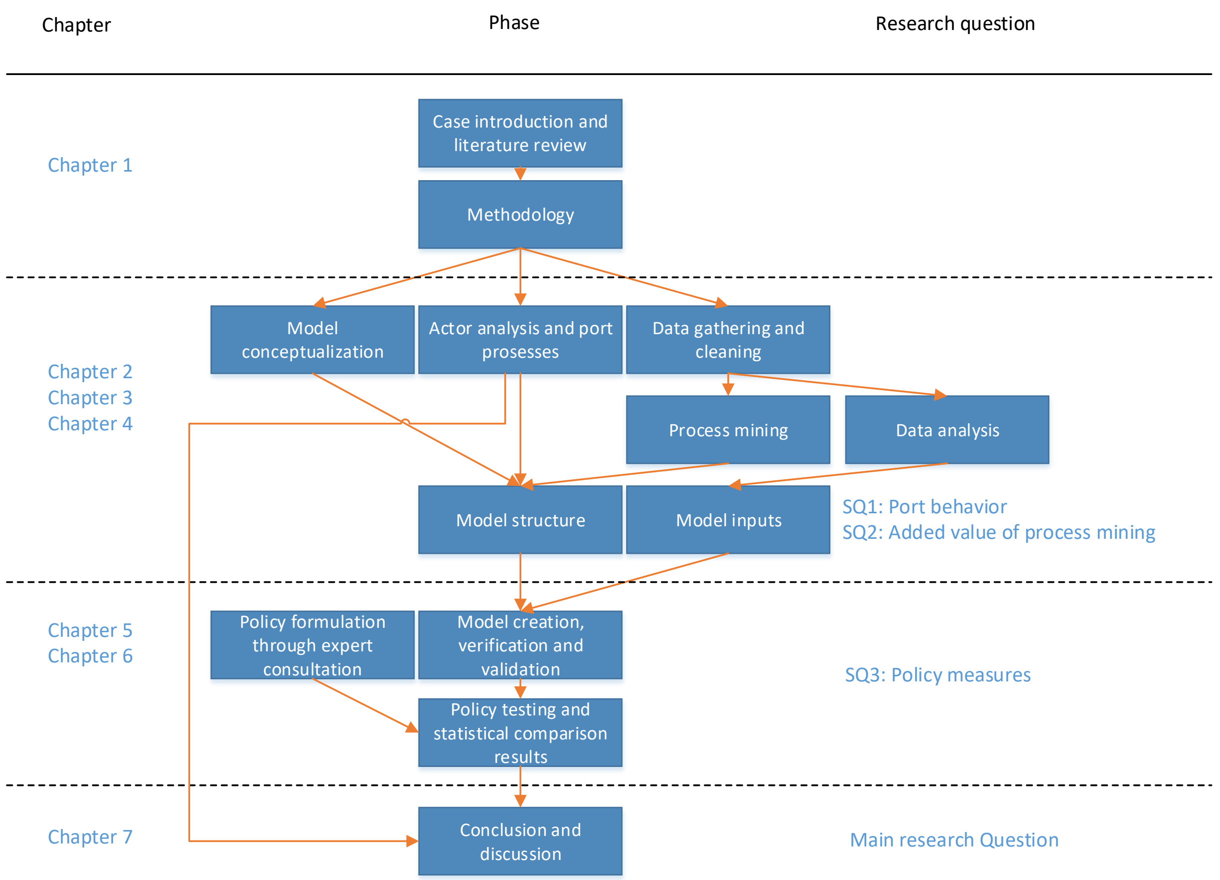

Figure 1.1: Research flow diagram

2. Port call and actor analysis

A port visit from a cargo ship's perspective is a complex process, set in a multi-actor environment. To get the exact scope of what elements need to be modelled, in order to simulate the problem, it is first essential to understand the bigger picture. More concretely, this chapter will present the complete list of actors involved in a port visit during the operational part of the process. Furthermore, the port visit process will be described and lastly the formal relationships between the actors is described. The information used in this chapter is mainly obtained from the Port Call Optimization Process Handbook, which is created by the International Taskforce Port Call Optimization (2018). Furthermore, information is gained through expert consultation.

2.1 List of actors in the Port of Rotterdam

This paragraph shows the list of actors that are directly involved in the port visit process. Furthermore, important characteristics and main tasks are also discussed for each actor. The actors are listed in alphabetic order.

Bunker ship

A Bunker ship is a vessel that operates in the port itself. The vessels are comparable in size and shape to Barges. However, its main function is to supply other vessels with fuel for their own use (Maritime and Port Authority of Singapore, 2009). The bunker ship will park next to the vessel and start pumping fuel to the fuel tank. This process mostly takes place at the dock. Bunker ships get the fuel from terminals and then service one or more ships before they go back to refuel again.

Figure 2.1: Bunker ship (Source: The Noun Project)

Captain

Figure 2.2: Captain (Source: The Noun Project)

The sea captain is the highest commander on board of a vessel. It is the captain's responsibility that the ship is handled in a safe and efficient way. All interactions between the ship and other parties need to be approved first by the Captain. These interactions could be the bunkering process or the process of getting a pilot on board, which will be explained later on. The captain acts as a commercial representative of the vessel's shipping company.

Consignor and Consignee

The consignor is the party that is selling and sending a shipment, in this case by a container- or bulk carrier. The Consignee, on the other hand, is the buyer and receiver of the goods. These are both formal roles and the exact specifications are included in a contract.

Figure 2.3: Consignor and Consignee (Source: The Noun Project)

Harbour Master's Division

The Harbour Master's Division is an authoritarian organ that strives to create a safe, efficient, clean and secure port environment. It is often either a governmental or quasi-governmental department. Depending on the size of the port and the internal structure, the exact tasks of the Harbour Master may differ from port to port. The description of the Harbour Master's Division here will specify how this actor functions in the Port of Rotterdam.

The Harbour Master's Division is an organisation that answers to the municipality of Rotterdam and the Dutch state, who are the only two stakeholders of the Port of Rotterdam. The Harbour Master is the head of the Harbour Master's Division. He is responsible for managing the four departments as visualized in Figure 2.4.

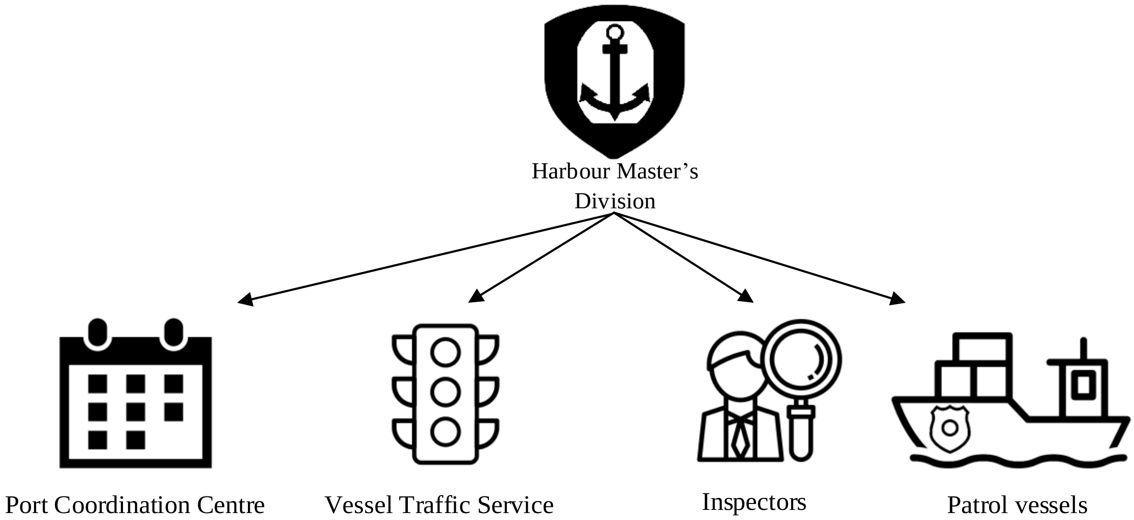

Figure 2.4: Harbour Master's Division: (Source Symbols: The Noun Project)

The Port Coordination Centre (PCC) is responsible for the waterway planning. Vessels that plan to visit the port are required to announce their arrival to the Port Coordination Centre at least 24 hours in advance. Based on the eta's, amount and type of visits, the PCC then makes an estimation of the number of pilots, tugboats and lineman needed to meet the demand. This estimation is then communicated to these actors, who together form the Nautical Services.

The Vessel Traffic Service (VTS) focusses on the real-time traffic handling in and around the Port of Rotterdam. Their primary goal is to make sure that the traffic arrives and leaves the port on schedule in a safe manner, which enhances the port's competitive position (Port of Rotterdam, 2019c). The VTS monitors the traffic in the area that surrounds the Port of Rotterdam. More specific, an area of 60 km into the sea and 40 km into the land. The area is split into two regions, region Europort and region Rotterdam. The VTS grands access to incoming vessels and assigns them safe routes to travel on. This is directly communicated to the vessels through the VHF radio.

The inspectors have the responsibility to make sure that the vessels satisfy national and international safety, security and environmental regulations (Port of Rotterdam, 2019d). If there is any incorrect behaviour, the inspectors make sure that measures are taken. Furthermore, they carry out systematic checks, to inspect if the shipping companies and agents comply with the statutory administrative reporting.

The patrol vessels are responsible for inspection and enforcement on the water in the port (Port of Rotterdam, 2019d). If needed, they supervise and guide incoming and outgoing cargo ships. Furthermore, they regularly are on patrol in the harbour and they assist the inspectors.

An important side note is that in reality the term 'Port of Rotterdam Authority' is often used for the actor 'Harbour Master's Division' as well. However, in this report the term 'Port of Rotterdam Authority' will only be used when referred to the profit oriented landlord company and not the authoritarian organ.

Figure 2.5: Linesman (Source: The Noun Project)

Linesmen

The main responsibility of linesmen is to secure every vessel at the quays, in all weather conditions at every moment of the day (Port of Rotterdam, 2019b). They are experts in the berthing and unberthing of ships. In the Port of Rotterdam, the linesmen of the Royal Dutch Boatmen's Association are distributed across 13 locations. Vessels longer than 75 meters are obligated by the Port Authority to use the services of the linesmen (Port of Rotterdam, 2019b). The lineman are part of the Nautical Services, offered by the Port of Rotterdam.

Port of Rotterdam Authority

The Port of Rotterdam Authority is an unlisted public limited company, with only two shareholders (Port of Rotterdam, 2019e). The Municipality of Rotterdam is the biggest one, with a 70% share. The Dutch government holds a share of 30%. The Port of Rotterdam is a commercial company that is profit orientated. With a turnover of €710 million and 1.200 employees, the Port of Rotterdam Authority is an major player in the world cargo business. The main tasks are the development, construction, management and operation of the port and industrial areas in Rotterdam. The Port of Rotterdam is a landlord port, which means that it only owns the basic infrastructure. This infrastructure is rented to operators, often on a long-term concession basis. The generated revenue from rental is the primary cash flow for the Port of Rotterdam Authority. The secondary cash flow is generated by port dues (Port of Rotterdam, 2019f). These are dues that a seagoing vessel has to pay to the Authority when they visit the Port of Rotterdam. Depending on the use of the port, the gross tonnage of the vessel ant the quantity of cargo, the amount of the dues is determined. The third and final cashflow is generated by selling services and products, that contribute to the safety and efficiency of cargo handling, to third parties.

Figure 2.6: Port of Rotterdam Authority (Source: The Port of Rotterdam)

An important side note is that in reality the term ‘Port of Rotterdam Authority’ is often used for the actor ‘Harbour Master’s Division’ as well. However, in this report the term ‘Port of Rotterdam Authority’ will only be used when referred to the profit oriented landlord company and not the authoritarian organ.

Pilot

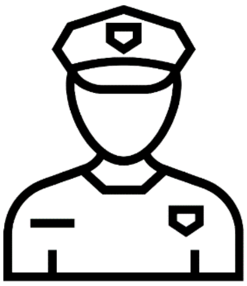

Figure 2.7: Pilot (Source: The Noun Project)

The main task of a pilot is to assist the master of a ship in navigation when entering or leaving a port, in a safe and efficient manner (Seine Maritime, 2018). This is often required in ports with a complex infrastructure. In the Port of Rotterdam pilots are required for ships longer than 75 meters, which is almost every sea cargo ship. In the Port of Rotterdam, the pilots are organized through the Rotterdam-Rijnmond Pilotage Service. Approximately 20 pilots are working simultaneously in the Port of Rotterdam (Port of Rotterdam, 2019b). A pilot comes aboard the vessel before it enters the port. He then manoeuvres the vessel to its berth spot. When the ship has finished its business in the port, the pilot has to safely manoeuvre the ship outside of the port. Vessels need to request a pilot at least two hours in advance. The pilots are part of the Nautical Services of the Port of Rotterdam.

Shipping Agent

The shipping agent is the local representative of a shipping company who is planning to visit a port with one of their vessels (MI News Network, 2018). Some of their main tasks are to book a vessel in and out of the port and to get all the right paper work for the visit (John Good Shipping, 2019). The agent is responsible for communicating the estimated time of arrival to the port authorities. Furthermore, they supervise the loading and discharge of cargo, advise the local customs of the ship’s arrival, submit information on crew to local immigration authorities and handle ship services like fuel, repairs and maintenance. An agent in the Port of Rotterdam can fulfil most of his obligations in Portbase. This is an digital environment that connects all the actors, involved in port logistics, by facilitating the data-sharing between the parties (Portbase, 2019).

Figure 2.8: Shipping Agent (Source: The Noun Project)

Shipping Line

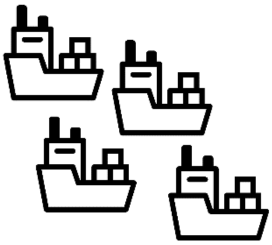

Figure 2.9: Shipping line (Source: The Noun Project)

A company that transports cargo using vessels is called a shipping line. Shipping lines set up contracts with terminal operators and other service providers in a port. This results in fixed routes for each of their vessels, during the duration of the contracts. The contracts often have a length of six months to a year, after this period it is possible to switch to another port. In the literature research in Appendix I, it is concluded that one of the most important factors that influence the port choice is port efficiency. Once the contracts are signed, most of the communication with the service providers in the port happens through a local shipping agent, who acts as the representative of the shipping line. Some of the big shipping lines have their own terminals or are stakeholders of terminals.

The biggest container shipping lines, not only in the Port of Rotterdam, but also worldwide, are A.P. Moller – Maerks Group, Mediterranean Shipping Company (MSC) and China Ocean Shipping Company (COSCO) (Alphaliner, 2019). The dry- and liquid bulk carriers are more fragmented.

Terminal Operator

A terminal operator runs a terminal. A terminal is a facility where the cargo is transferred from vessels to other vehicle types and the other way around. This process is often automated. There are different types of terminals in the Port of Rotterdam. Each one serves a specific cargo type like containers, dry bulk or liquid bulk. The terminals in the Port of Rotterdam have direct deep-sea-, feeder-, inland waterway-, road- and rail connections. Terminal operators are private companies in the Port of Rotterdam, as this port is a landlord port. There are about 150 terminals in the Port of Rotterdam.

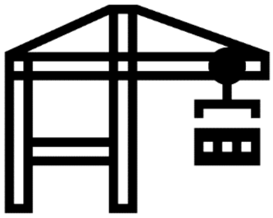

Figure 2.10: Terminal Operator (Source: The Noun Project)

Tugboat Operator

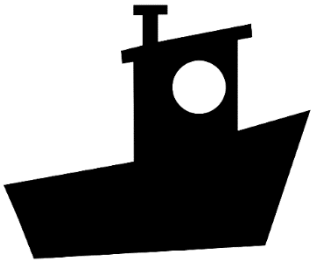

Figure 2.11: Tugboat Operator (Source: The Noun Project)

A tugboat is a strongly built powerful boat, that is used for towing and pushing (Merriam-Webster, 2019). In the Port of Rotterdam, the tugboats are used to manoeuvre ships mainly larger than 75 meter in and out of the harbour. Each big vessel needs between two and four tugboats. This depends on the requirements of the shipping line, captain and Harbour Master. Furthermore, it depends on the weather conditions and vessel size. There are two main private tugboat operators in the port: Fairplay and Kotug Smit. Together they operate around 40 tugboats and are part of the nautical services, offered by the harbour. The Shipping agent is responsible for booking a tugboat. The tugboats are part of the Nautical Services offered by the Port of Rotterdam.

Vessel Service Providers

Besides the bunkering- and Nautical Services there are many other logistical services that can be offered to an incoming vessel. These services include supplying a ship with food, fresh water and medicines, but also the collection of waste water and maintenance and repairs on the ship. These parties that offer vessel services are bundled into one actor, the vessel service providers. All the services need to be first approved by the captain and by the terminal, in the case that the service comes over land.

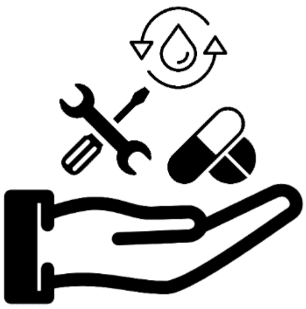

Figure 2.12: Vessel Service Providers (Source: The Noun Project)

2.2 Port actor relationships

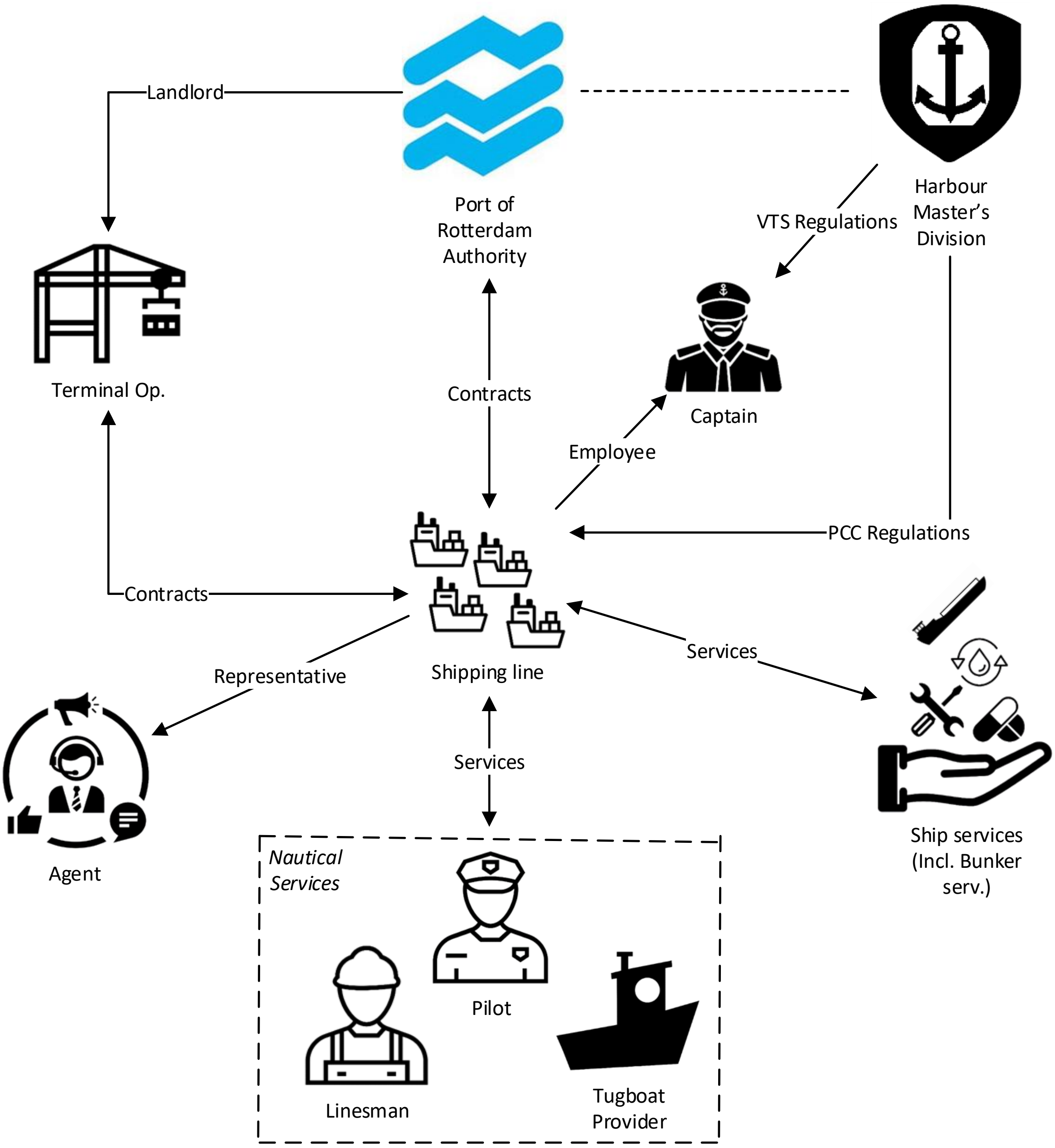

Figure 2.13: Relationships between actors involved in a port call

The formal map (Figure 2.13) shows the formal relationships between the described actors. Furthermore, it shows the type of formal relationship and the direction of the arrow shows hierarchical structure. For example, the arrow from shipping line to agent shows that the agent acts as a representative for the shipping line and that the shipping line has hierarchical power over the agent.

The Port of Rotterdam Authority and the Harbour Master's Division are formally independent. The Port of Rotterdam is a commercial company and the HMD is an organization that answers to the municipality of Rotterdam and the Dutch state. However, these two actors still work closely together in practice. This is why the line between the Port of Rotterdam Authority and the Harbour Master's Division is dotted.

2.3 Port call process

The previous paragraph gave individual descriptions of all the actors involved in the operational part of a port visit. This paragraph will describe the port call process itself and what role the actors play in this process. The goal is to get a clear image of which actors are involved in which step of the process. Furthermore, this paragraph will give a first impression of where delays could happen in the process. The process will be described from the perspective of a cargo vessel. Also, the potential identified delays are delays experienced by the cargo vessels, visiting the port.

The description of the port call process will be based on the port call map and its documentation (ITPCO, 2018), created by the International Taskforce Port Call Optimization (ITPCO). Appendix II shows this map. This map succeeds in describing the port call process in a very structured way. However, it does not fit the scope of this study entirely in terms of timeline and level of detail. The timeline used by the ITPCO begins already three months in advance of the actual port visit with the contractual phase. In this phase, the cargo traders and shipping lines draw up contracts regarding the vessels, terminals and cargo. This research, however, focuses on the operational part of the port call. Therefore, the contractual phase will not be included. Furthermore, the level of detail used by the ITPCO is in certain sections not high enough. For example, the map in Appendix II identifies the Nautical Service Providers as one actor. In this research, this actor is split into three actors: the tugboat companies, the pilots and the linesmen.

The port call process identified through expert consultation is visualized in figure 2.14. The figure is read from left to right. Three different levels of aggregation are used to describe the process in steps. The top one is the most aggregated, while the bottom one has the highest level of detail. The boxes that have a dotted outline are steps that are not part of every port call. For example, not all vessels have to wait before they can access the port. That is why the box with ‘Anchorage’ is dotted. Also, some vessels need to visit multiple destinations, which explains why the second ‘Berth’ box is dotted. The bottom section shows which actors, described in chapter 2.1, are directly involved in what steps of a port call. Figure 2.14 is followed by a more detailed explanation of the operational part of a port call.

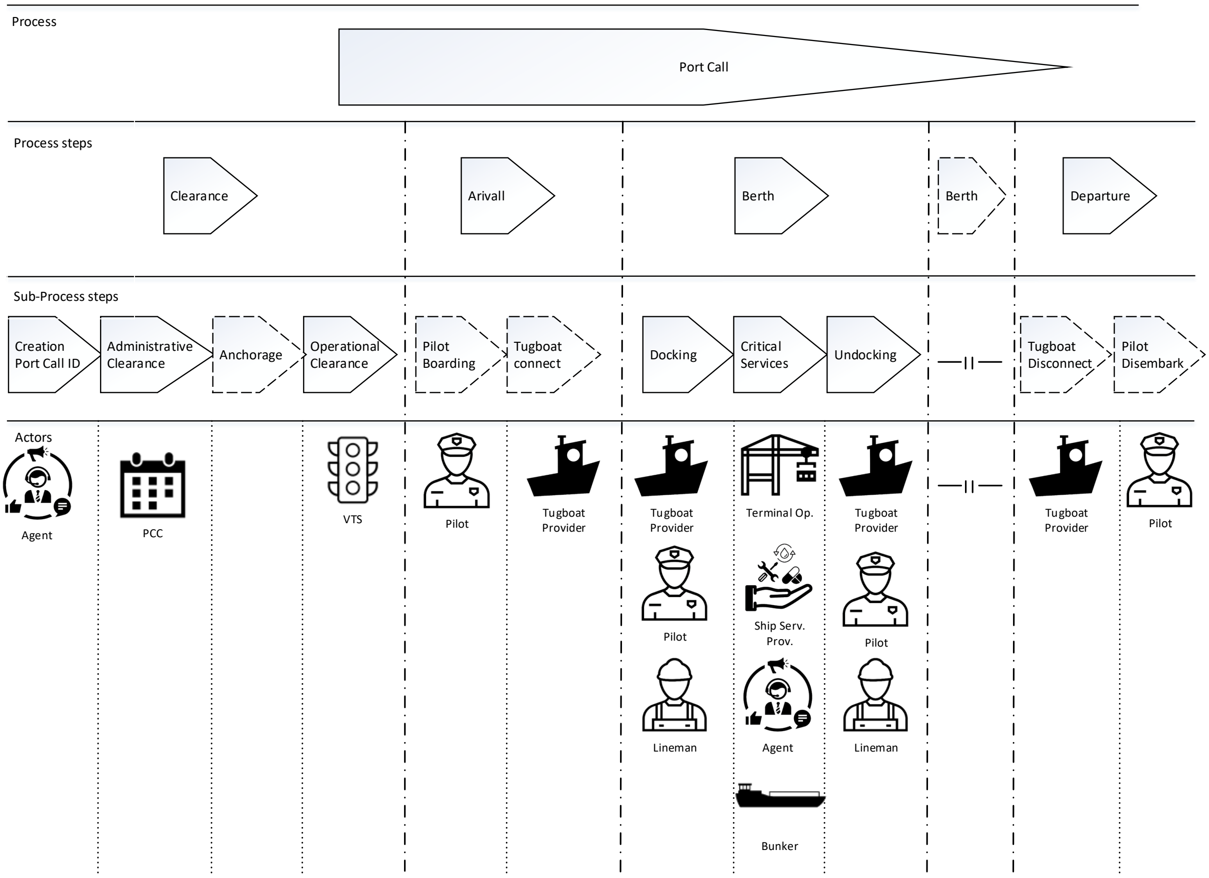

*Figure 2.14: Port call process map (*Critical services are not worked out in the same level of detail as the other steps)*

1. Approximately 24 hours in advance, the captain sends the Expected Time of Arrival (ETA) of the vessel at berth to its agent, who represents the shipping line. The agent then communicates the ETA at berth to Portbase, a digital data-sharing environment. Portbase automatically generates a Unique Call Reference Number (UCRN) for the visit of the incoming vessel. This event is identified as the beginning of the operational part of a port visit. All the involved actors can access this information in the Portbase environment and can start preparing for the ship's arrival. The agent will update the ETA at berth more frequently as the vessel gets closer to its destination.
2. Soon after the UCRN is generated, the Port Coordination Centre will review all the paperwork required for a legitimate visit and will give administrative clearance to the vessel to enter the port. Access is of course only granted in the case that the paperwork matches all the requirements. In actual practice, the amount of times that the Port Coordination Centre does not grant access is negligible.
3. Besides the administrative clearance, the vessel also needs operational clearance. This is granted by the Vessel Traffic Service if the requirements are met. In order to get clearance, the ship needs to be on the VTS radar, which covers an area of 60 km from the Port of Rotterdam into the sea and also an radio connection must be established between the captain of the vessel and the VTS. Furthermore, there must be available linesmen and a free berthing location at the designated terminal. Also, if required, the right number of tugboats from the right company and a pilot must be available. Whether a pilot or tugs are needed depends on the type and sizes of the vessel, the weather, the requirements of the port and the permits of the captain. If one or more of these requirements are not met, access will not be granted. In this case, the vessel will either loiter near the port, or the vessel will drop anchor in one of the anchorage areas until

operational clearance is given. The anchor areas are marked with polygons in figure 2.15. This is where most of the delays occur. Identifying the cause of each delay is challenging, because the location of where a vessel experiences the delay, has no connection with the origin of the delay.

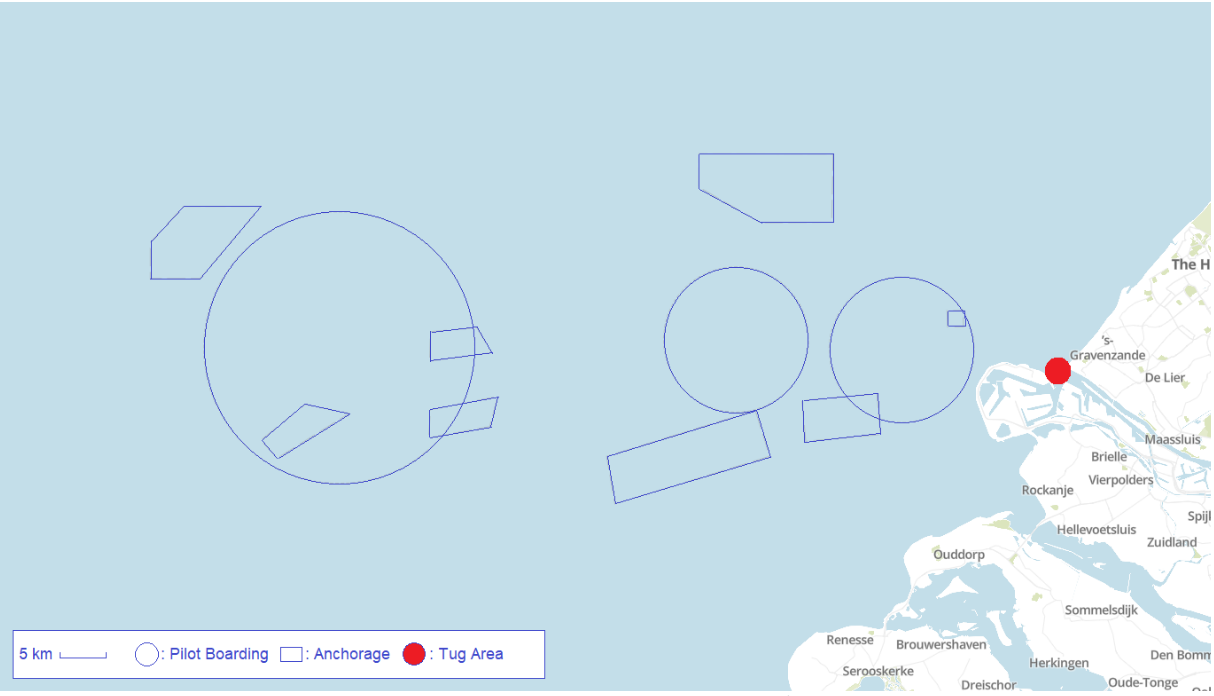

Figure 2.15: Nautical Services Rotterdam (Map Source: Pronto, Data source: Port of Rotterdam)

4. After getting clearance to enter the port, a vessel proceeds to the pilot boarding place. These are marked in figure 2.15 with hollow circles. The circle closest to the port is called Maascenter and is the most frequently used area for pilot boarding. The second area is called Pilot Area LNG and is used for bulk carriers that transfer liquified natural gas. These vessels have their own area to reduce the risk of incidents. The area that is the furthest away from the Port of Rotterdam is called Rendez Vous and is dedicated for vessels that have a draft of over 17,39 meters. Rendez Vous is located at the start of the fairways, which lead to the entrance of the port. These narrow fairways need to be followed by the deep ships (draft > 17,39 m). Typically, the pilots get transported to the incoming vessel by a dedicated pilot ship. However, sometimes helicopters are used.
5. Now that the pilot is onboard of the vessel, it can proceed to the entrance of the port. This is where the tug area is located (figure 2.15). The tug area is where the tugboats meet the incoming vessel and connect to the vessel if necessary. Typically, up to four tugboats are required. It depends on the size of the vessel, type of cargo, weather conditions and the policies of the shipping line.
6. The destination in the port is a contractually predetermined terminal that handles the same type of cargo that the vessel is carrying. Different areas in the port specialize in handling different types of cargo, like containers, dry bulk and liquid bulk. These areas and the unique terminals in them are visualized in figure 2.16, in order to get a better feel of the port. Once arrived at the berthing location, the ship begins the docking process. The pilot, tugboats and lineman work in a coordinated way, to get the vessel secured at the dock. When the ship is docked, the captain of the ship reports the Actual Time of Arrival (ATA), this is also documented in the event data.
7. The duration of the stay at the dock is determined by the critical services and range from a couple of hours to a couple of days. Critical services are the services that need to be completed before departure. In contrast to non-critical services, which can be rendered in the next port.

There are two types of critical services: cargo services and ship services. Cargo services are carried out at the terminal. The cargo is loaded onto the vessel or conveyed to the terminal from the ship. The speed of (un)loading the cargo depends on the quality and quantity of the equipment used by the terminal operator. For example, in the case of containers, the speed is directly related to the amount and capacity of cranes and AGV’s (automated guided vehicles). The ship services are ordered by the agent. Refuelling the ship often takes the longest time of all the ship services. The refuelling process is most of the time done using a bunker vessel and takes multiple hours to complete. Other ship service activities include waste collection, the supply of food and ship maintenance and -repairs. All the involved parties are motivated not to get behind the agreed-on schedule, to avoid financial consequences.

8. When the critical services are completed and the agent has ordered nautical services, like a pilot, tugs and linesmen, the ship executes the undocking procedure. The actors involved in this procedure strive to get the ship undocked on the Planned Time of Departure (PTD) at Berth, which is set by the VTS and based on a prediction made by the ship. The Actual Time of Departure (ATD) at Berth is clocked when the last line that connects the vessel to the quay is released by the linemen.

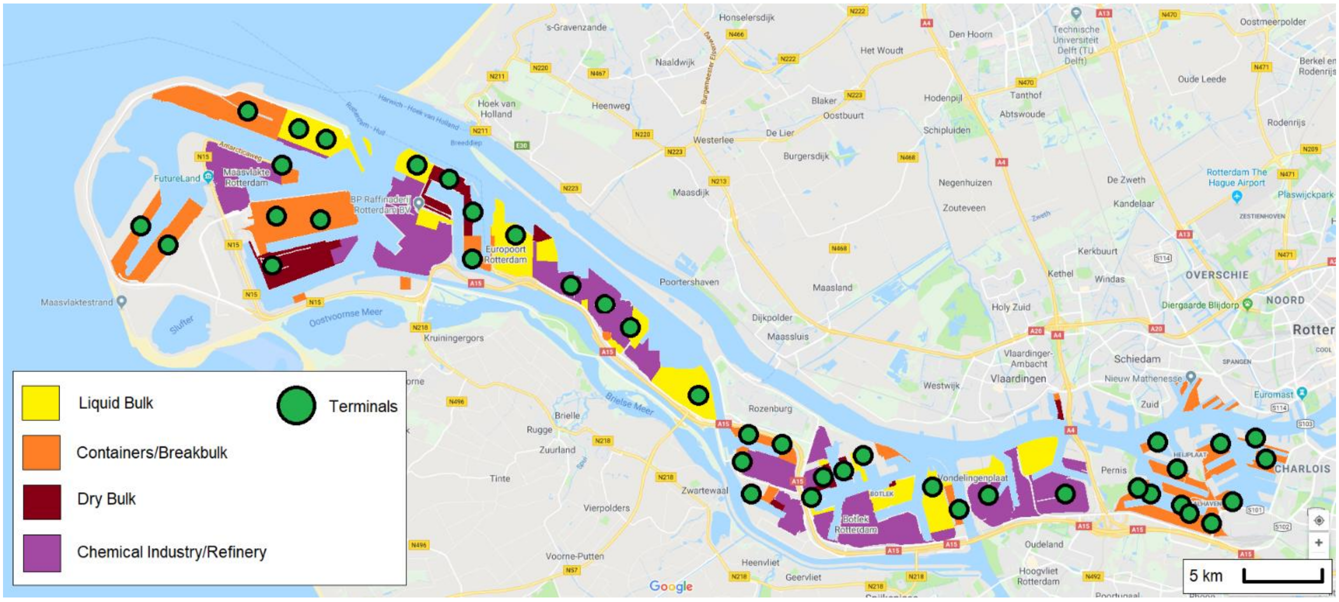

Figure 2.16: Terminals and cargo type areas (Map Source: Google Maps, Data Source: Port of Rotterdam)

9. If the vessel needs to visit another terminal in the port (figure 2.16), it will be escorted to this destination and the process will loop back to step 6. These possible other visits are already agreed on in the contractual phase and given permission for during administrative- and operational clearance. It is important to note that in some cases, when the two terminals are too far away, the tugboats wont escort the vessel the entire route. They will detach right after undocking and attach again before docking at the new terminal. If there are no other destinations in the port, the tugboats will disconnect from the ship, shortly after the ship is undocked and the ship will proceed to exit the port. However, there are exceptions: during bad weather conditions and in case of special cargo, the tugboats escort the vessel to the exit of the port.
10. Before the vessel starts a new trip to the following port, it disembarks the pilot at the designated pilot boarding areas (figure 2.15). This is the last step of the operational part of a port call from the ship’s perspective.

3. Data analysis for quantitative insight of vessel behaviour

The main goal of this chapter is to report on the data analysis in Python, in which AIS event data from 2018 is analysed in order to provide quantitative insights for the steps of a port call. First, the required data and its origin is discussed and how it is gathered. Next the quality of the data is assessed. Lastly the most important insights, which followed from this analysis, are discussed.

The data analysis phase in Python produced two more products. The first is the simulation input schedule, which is described in the specification of the simulation model (Appendix V-II) and the second one is the process mining input schedule, which is described in the process mining chapter (4).

3.1 Data Gathering

To get the desired insights, data about the vessel category, number of berth visits, destination in the port, vessel arrivals, pilot usage and number of tugboats and their company is needed. Also, data about the berthing- and anchorage durations is required for input distributions and validation purposes. Lastly, the input data that is used for the process mining phase in chapter 4 has to meet certain criteria. It needs to represent each event of the operational part of a port call, as visualized in figure 2.14, for every vessel over time. More concretely, the data needs to exist out of rows that each specify the unique port calls, the current event and a timestamp for that event. Data in this exact structure with these properties is not available. However Automatic Identification System (AIS) data (IEC, 2001) shows potential to be transformed into this format. Each sea-going vessel emits AIS data. This data shows the Maritime Mobile Service Identity (MMSI) number of a ship, which is a unique identification code, the coordinates of the ship and a timestamp. Furthermore, it holds information about the vessel’s heading, size, cargo, ship type, destination and estimated time of arrival (IEC, 2001). An AIS signal is emitted every two to ten seconds by a vessel. The reporting interval is directly related to the speed of a ship (IEC, 2001).

This paragraph describes the origin, paths and transformations of the data that forms the event database, which is used as an input for the data analysis. These paths are visualized in figure 3.1. There is not one source that collects the AIS data. Multiple parties receive and store the raw data. Antennas on the coast are used to receive data from vessels that are near the shore and satellites are used to collect data from vessels that are deep in the open sea. This raw data collection is done by organizations like LuxSpace and M.M.S.I.S., who then sell the data to third parties or make it available for free. The Port of Rotterdam Authority, buys and collects this data. Self-developed tools from the Port of Rotterdam are then used to identify different events regarding a port call. The collective name for these self-developed tools is the event engine. The identified events from the AIS data are then fed to the event database. However, there are more events than just the events identified through the event engine. Multiple parties, that somehow are involved in a port call, feed this database with events as well. The agents, for example communicate the number of tugboats that the vessel will request, or they estimate the expected time of arrival at berth. The terminal operators, on the other hand, give their estimated- and actual time of arrival for a particular vessel at their terminal. These events and many more are also added to the event database. This database is stored in a cloud environment. Access to this database is granted for this research by the Port of Rotterdam Authority.

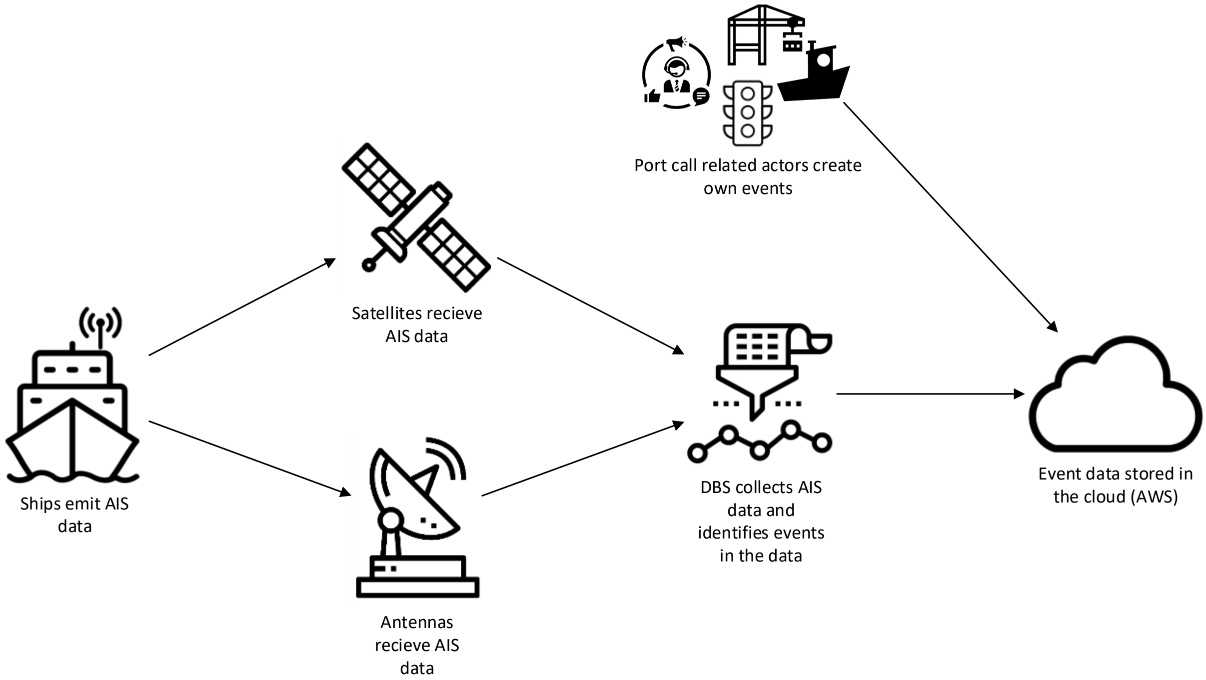

Figure 3.1: AIS Data Path (Symbol Source: The Noun Project)

For this research data is used from the year 2018 for the Port of Rotterdam. Data subsets are queried from the AWS database and converted from a nested json structure to a pandas data frame in Jupyter Notebook. The data contains details about each vessel that visits the Port of Rotterdam, like the Unique Port Call Reference Number (UCRN), International Maritime Organization (IMO) number, the Maritime Mobile Service Identity (MMSI), the ship type code and the dimensions of the ship. The ship type code, which is a standardized code that described the ship type and cargo type, is used to label every ship with one of the six seagoing vessel types, identified in Appendix V-I. Furthermore, the data shows 41 different types of events (Appendix III) which are directly linked to a specific vessel. These events include, among other things, the ETA's, ETD's, ATA's and ATD's, for different steps in a port call, like pilot boarding or docking. Some of the 41 events are recorder by different sources. The Pronto team uses an algorithm (Appendix III) to determine the most accurate source. The events from the most accurate sources are stored in the data under 'mostAccurates'. Only the events from the sources that are identified as most accurate by the Pronto team, are used for this research.

It is important to note that the Port of Rotterdam processed the AIS event data following the GS1 standards (GS1, 2019). This enables the exchange of the event data with other parties. Furthermore, all the scripts used for data transformation and analysis are coded in such a way that they are generalizable. It should be a smooth process to query the data and perform the same data analysis on any other port. Therefore, all the ingredients are available to perform this exact analysis on any port in the world.

3.2 Data quality

In order to perform a valid data analysis and produce input for the simulation model and for the process mining model it is essential to be aware of the quality of the data. Big data brings a lot of opportunities, but it also brings some challenges. Due to the massive size, multiple sources and high dimensionality it is sensitive to errors (Fan et al, 2014). Therefore, missing values and outliers are inevitable. The better the data quality, the more trustworthy the results are. The way that is dealt with the noise in the data, directly influences the liability of the outcome. This chapter discusses the way this issue is handled.

During the data analysis in Python, a significant amount of outliers and missing data was identified. It is essential to know if there are any patterns in the noise, for example, an entire week of data that is missing or one type of vessel or event that is showing false information. Three main reasons are identified for the noise. The first is human errors. Some predictions in the data are made by agents in the port. This can lead to mistakes in the data like typo's or missing values. The second is measurement equipment failure. Tracking devices make errors as well. There have been examples of vessels that, according to the data, travel thousands of kilometres in a matter of seconds. This is of course due to an measurement error. The third are Port of Rotterdam tools that wrongly identify events from the AIS data. In general, the tools perform really well. However, there is always room for improvement. An example of an error from this category is that, when two vessels are docked right next to each other, a tug connect event is wrongfully identified for one vessel, when a tug is actually connecting to the other vessel. All the noise seems to be randomly spread over time and over many events.

For this research, it is chosen to drop all the rows with dirty data. The amount of data that has been taken out of the three data analysis products, which are the quantitative port call insights, the simulation input schedule and the process mining input schedule, is best illustrated when looking at the simulation input schedule (Appendix V-II-I). The simulation input schedule is a table that contains the processed and cleaned data for almost 25.000 port calls in 2018. The actual amount of port calls should be almost 30.000 (Port of Rotterdam, 2019g). This means that approximately 5.000 (16.7%) of the port calls did not end up in the final data sets. This is not compensated during the presentation of the insights (section 3.3), because it is assumed that the noise is randomly spread and therefore will not affect the shape of the empirical distributions. For the process mining input schedule (chapter 4) also no action is taken. Because of the random spread, it is not expected that the process mining model behaviour will be influenced significantly, except for the number of observations. However, the lack of 16.7% of the total port calls in the simulation input schedule needs to be compensated. Otherwise, around 5.000 vessels will not be simulated. This will result in less pressure on resources in the model like pilots and tugboats. This means that the simulation model, that is designed to test delay reducing policies, will never be valid. The delays in the model would be lower than in real life. In order to compensate for this, 5.000 vessels are added, equally spread over the simulation run. The attributes, except for the arrival time, are sampled from the simulation input schedule.

3.3 Data analysis and visualisation

This chapter discusses the results from the AIS event data analysis in python. First, the data quality is discussed. Next, a visual representation of the most important findings is meant to give an insight in the behaviour of different vessels during a port call. Notable or remarkable findings are briefly discussed. The final deliverables for this project include the actual data sets, empirical distributions and the scripts that are used. It is important to note that each visualisation of events from the AIS data is vessel type specific. Lastly, how each discussed element, from the data analysis, is used in the following parts of the research is described as well.

Vessel arrival schedule

In the AIS event data, the events 'ATA 120 nmi zone', 'ATA 60 nmi zone' and 'ATA 12 nmi zone' describe when a vessel that is sailing to the Port of Rotterdam is 120, 60 or 12 nautical miles away from the port. These are used as the entry points for the simulation model. Incoming vessels that come from a port more than 120 nmi away will enter the port at the 120nmi entry point. Vessels that come from a port less than 120 nmi away but more than 60 nmi will enter the model at the 60 nmi entry point. Vessels that come from a port less than 60 nmi away will enter the model at the 12nmi entry point. Figure 3.2 shows the distribution between the three entry points as

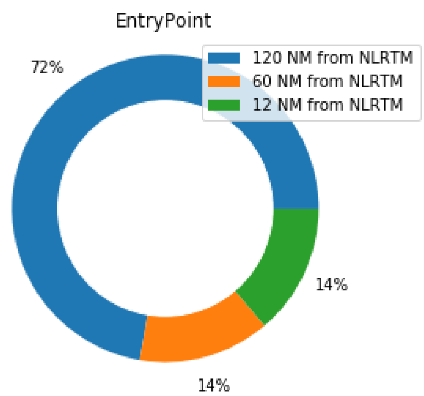

Figure 3.2: Entry point distribution

observed from the data. Most of the vessels (72%) will enter the simulation at the 120 nmi point. The entry point and entry time form one element of the simulation input schedule (Table V-1). Simulating the last part of the journey to the harbour makes it possible to experiment with policies that include behavioural changes of incoming ships on the open sea.

Berth visits per port call

The number of berth visits per port call are extracted from the data by counting the number of unique berth ID's per port call. A visualization of the findings is shown in figure 3.3. It is notable that the large container vessels and passenger vessels have only one destination in the Port in almost all cases. Regular container vessels have by far the most berth visits per port call. The number of berth visits is incorporated in the simulation input schedule (Table V-1) for each port call separately.

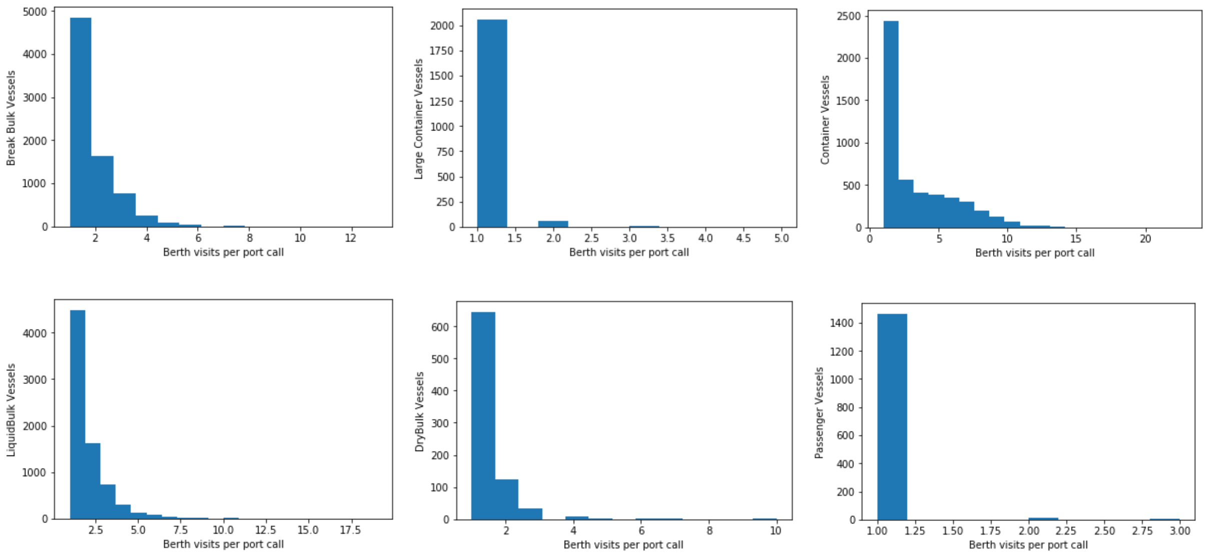

Figure 3.3: Berth visits per port call per vessel category

Number of tugboats

For determining the number of tugs per berth visit, a distinction is made between the number of tugs used to a berth spot and the number of tugs used from a berth spot. The number of tugs used per vessel are extracted from the data by looking at the 'NumberOfTugsToBerth' and the 'NumberOfTugsFromBerth'. These specify the requested number of tugs to and from a berth per berth visit by the agent of the vessel. This does not always match the actual number of tugs used, however it gives a good indication. The actual number of tugs used per berth visit is not recorded in the data available for this research. The number of tugs to and from a berth spot per vessel category is visualized in figure 3.4. It is notable that breakbulk-, regular container- and passenger vessels use zero tugs in most of the cases. It is likely that the cases where these categories use tugs, are cases with extraordinary cargo or bad weather conditions. Most of the large container vessels and dry bulk vessels use two tugboats. The liquid bulk vessels have a more scattered distribution between zero and two tugs. This could be explained due to the fact that there is no distinguish being made between relatively small and relatively big liquid bulk vessels. The number of tugs to- and from a berth are incorporated in the simulation input schedule (Table V-1) for each port call separately.

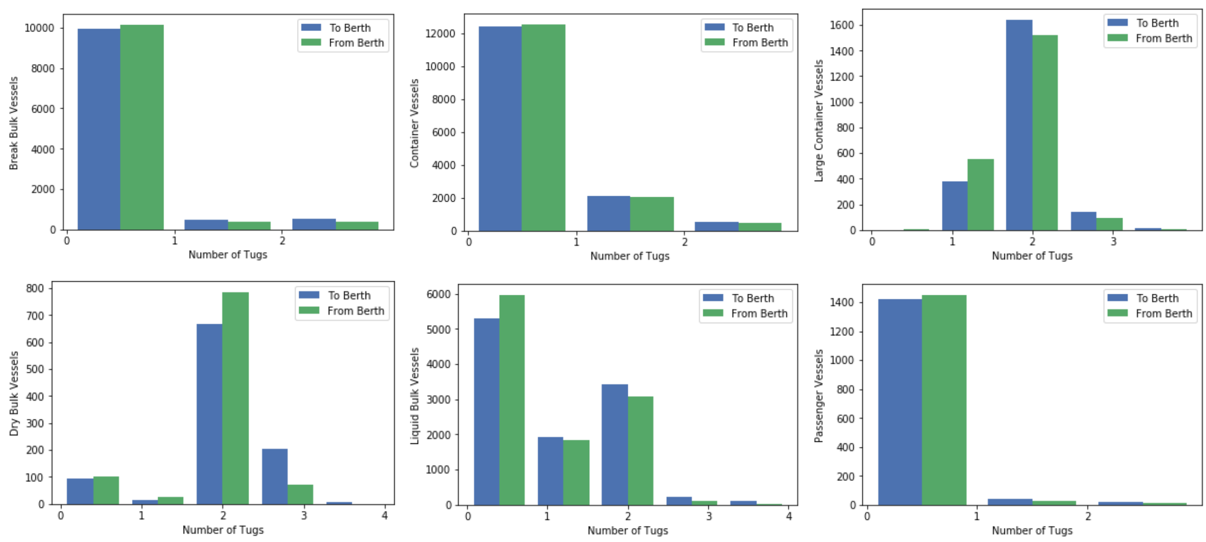

Figure 3.4: Number of tugs to and from berth, per vessel, for each vessel category

Anchorage visits

The percentage of vessels per category that drop anchor in the anchor area is determined by calculating the percentage of port calls that have the event 'ATAAnchorArea' in them. It is concluded that every port call that has this event, visited an anchor area. The percentage per category is presented in table 3.1. The Container vessels show a relatively low percentage compared to the liquid bulk and dry bulk vessels. This can be explained due to their just in time-schedule which is described in appendix V-II-II.

Vessel category	**Percentage to anchorage**
Liquid bulk	46.96%
Breakbulk	8.35%
Dry bulk	43.13%
Container	18.15%
Large container	23.04%
Passenger	0.07%

Table 3.1: Percentage of vessels that visit the anchorage per vessel category

The time spent at the anchorage is calculated using the 'ATAAnchorArea' event and the 'ATDAnchorArea' event for each vessel category. It is only calculated for vessels that in fact visit the anchorage. The empirical distributions are shown in figure 3.5. It is notable that for the entire year of 2018 there is only one passenger vessel that entered the anchorage area. In the simulation, each vessel will either visit the anchorage or not, based on the percentages in table 3.1. This distribution (3.5) can be used for validating the simulation model, as the delay at the anchorage is simulated and not inputted in the simulation. However, this has not been done, due to the limited amount of time for this research.

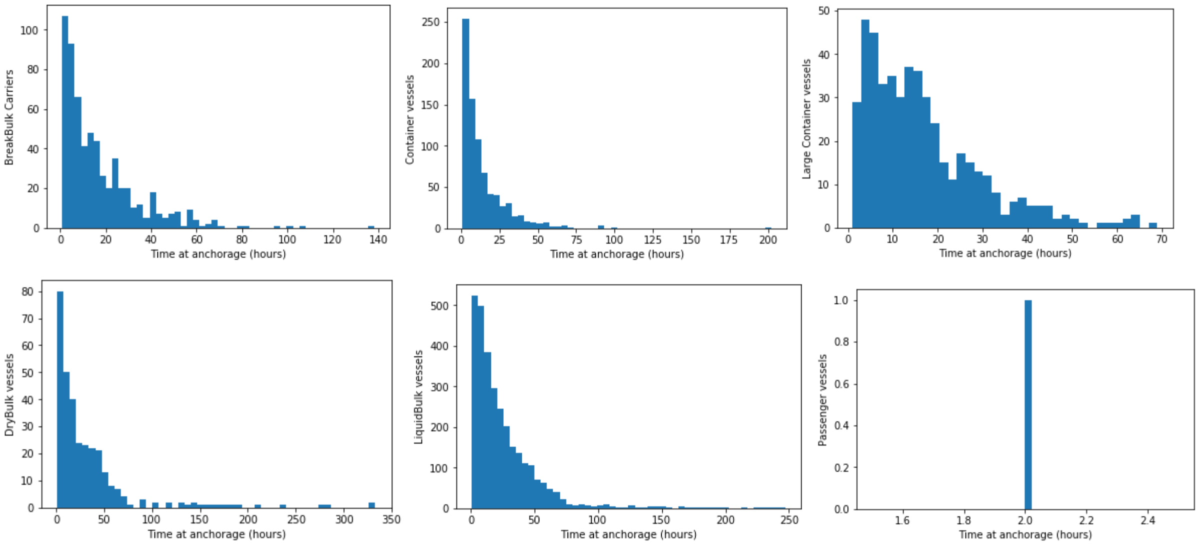

Figure 3.5: Average time spent at anchorage per vessel category

Vessels pilot usage and pilot boarding place

For each vessel, it is determined whether it used a pilot by looking at the ‘ATPilotOnBoard’ event. In the case that a pilot is used, the pilot boarding place was extracted from the ‘ATAPilotBoardingPlace’ event. The results are shown in figure 3.6. It is notable that the pilot boarding place Maascenter is by far the most used one. Furthermore, almost all the large container vessels use a pilot, where almost all the passenger vessels do not use a pilot. Whether a vessel uses a pilot and if so, at what pilot boarding place, is incorporated in the simulation input schedule (table V-1) for each port call separately.

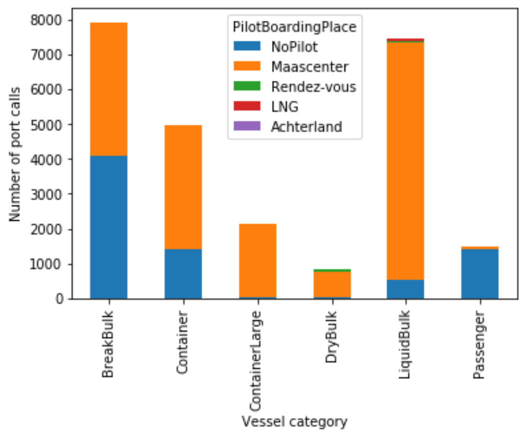

Figure 3.6: Pilot usage and pilot boarding place

Berthing Durations

The berthing durations are calculated using the most accurate berth ATA and berth ATD per berth visit. This is done for each vessel category separately. The empirical distributions are visualized in figure 3.7. A vessel in the simulation will draw a value from the empirical distribution corresponding to its vessel type to determine the length of its stay at a berth. It is chosen to sample this and not to give the actual berth duration for each port call, to add variance to the simulation, to test the robustness of the model.

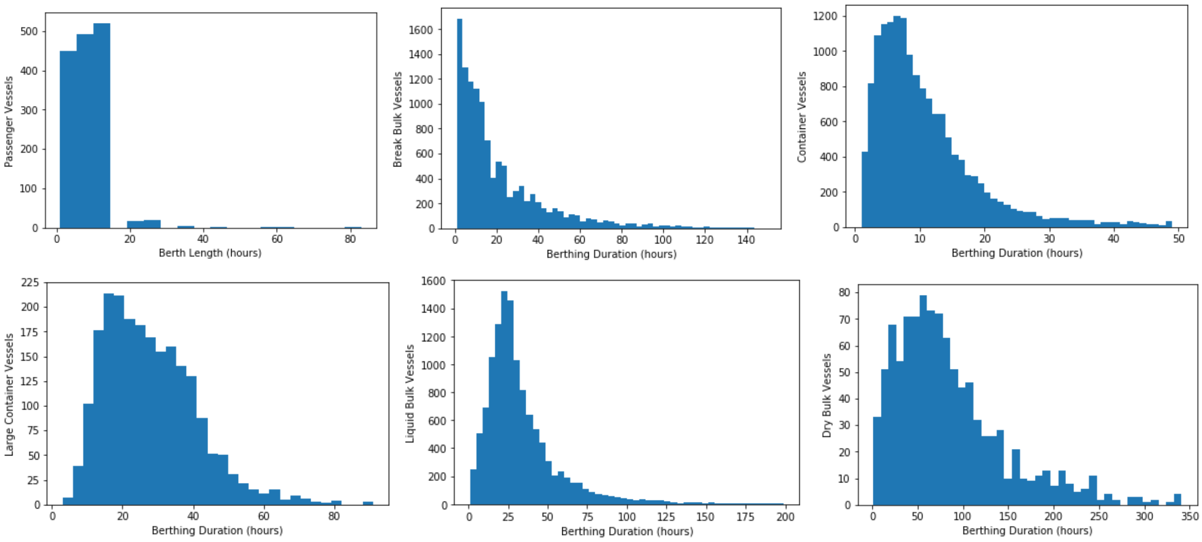

Figure 3.7: Berthing duration per vessel category

3.4 Discussion port call analysis

The first objective of this research is to get a clear picture of a port call process and the actors involved. This is essential for research phases that follow this phase. It is not possible to improve a process without having a fine understanding of what exactly is happening. Furthermore, in order to assess policies that potentially improve the port call efficiency, one first has to identify these measures. When having a clear picture of the entire process, bottlenecks can be pointed out. By reviewing these bottlenecks, policies are identified with the potential to improve the port call efficiency. Lastly this phase also produced the inputs for the process mining phase and the simulation phase and it assists in validating the findings that result from these phases. The following sub-question was formulated in the beginning of this research to achieve the above.

What are the steps of a port call and which actors are involved?

The knowledge needed to answer this question resulted from a combination of a qualitative- and quantitative analysis. The qualitative part includes expert interview and literature review (appendix I) and the quantitative part exists out of an event data analysis in Python (chapter 3).

Figure 2.13 shows the formal relationships of the actors that are involved with a port call. Figure 2.14 shows the steps of a port call and which actors are involved in which step. These steps are described in different levels of aggregation. This figure alone forms a huge part of the answer for this sub-question. The figure shows all the steps from a vessel's perspective that visits the Port of Rotterdam, which is the case-study of this research. However, this figure is also generalizable and is applicable to other huge sea ports. This figure is not enough to fully understand the behaviour of a port from a vessel's perspective. Quantitative insights are missing. This gap has been filled with the event data analysis in chapter 3. This analysis quantifies the findings from figure 2.14, in terms of frequencies of steps and duration of steps.

During the data analysis, event data of 25.000 of the total 30.000 port calls is collected and used to provide insights. The remaining 5.000 port calls are lost along the way. It is assumed that these are equally spread over the data and that they are missing due to a combination of measurement errors, shortfalls of the event detect engines. It is expected that the missing data does not have a significant impact on the patterns that arise from the data analysis.

During the data analysis, it became clear that different types of vessels show different behaviour. For this reason, it is chosen to divide the port calls in six categories, which are medium sized container vessels, large container vessels, liquid bulk vessels, breakbulk vessels, dry bulk vessels and passenger vessels. By adding these categoryies, a more detailed image is created of the vessels' behaviour in the port. For example, 3.1 shows that almost no passenger vessels visit the anchorage, while almost 50% of the liquid bulk carriers do. Approximately 18% of the small container vessels visit the anchorage while 22% of the large container vessels visit the anchorage. These last two categories only visit the anchorage when they are delayed, because they are the only ones that travel on a just-in-time schedule. Therefore, these percentages give a good indication of the amount of delayed container vessels. In combination with figure 3.5, which shows the length of the anchorage visits per category, the length of the delays of the container vessels can be calculated. This is very useful information for the simulation model (chapter 5), where the delays of the vessels are simulated. Figure 3.6 provides a clear image of the pilot usage. Almost all the pilots board a vessel at the Maascenter, which is the closest pilot boarding place to the port (figure 2.15). From figure 3.6 it is concluded that almost no passenger vessels use pilots. On the other hand, almost all liquid bulk, dry bulk and large container vessels do use pilots. Breakbulk vessels use a pilot roughly 50% of the time. Figure 3.4 provides insights about the tugboat usage. This shows that the vessels use up to four tugboats at once. Typically, large container vessels and dry bulk vessels use two tugs. Small container vessels, breakbulk vessels and passenger vessels, most of the time do not use tugs. For liquid bulk vessels, the data is more scattered, between zero and two. This information about the usage of the nautical services is essential for chapter 5, where a model is presented which, among other things, simulates a port call that includes these services. The number of berth visits per port call is shown in figure 3.3. This is used in the simulation model as well. The most remarkable finding is that large container vessels typically only visit one terminal. However, small container vessels have up to fourteen berth visits. Lastly, figure 3.7 shows the berthing durations. The passenger vessels, on average have the shortest stay, roughly between zero and eighteen hours. The longest visits, up to 350 hours, are observed in the dry bulk category. However, most of the visits are approximately between 10 and 140 hours. Large container vessels typically have visits between 10 and 40 hours, while small container vessels have berth visits between 5 and 20 hours. These berthing durations are an essential ingredient for the simulation model.

Together, the port call process map (figure 2.14) and the qualitative insights answer the sub-question. The quantitative part cannot be generalized to other ports as easily as the qualitative part. This is because it is too port and time specific. However, the queries and scripts that resulted in these findings, are written in such a manner that they are easily applicable to other ports and timeframes. It should be a relative smooth process to get the same insights for any other port, for any timeframe, if the event data is available.

4. Process mining event data

The idea of process mining is to discover, monitor and improve real processes by extracting knowledge from event logs (Van der Aals et al., 2011). First, this chapter discusses how the input for the process mining model is prepared. This is followed by the two main objectives of this chapter. The first one is to find a model structure for the simulation model by process mining the transformed event data. The second objective of this chapter is to explore the more general applications and benefits that the process mining of event data can bring in the port domain. The process mining software, Disco (Fluxicon, 2019) is used to create the process mining models. The main principles for process mining, defined in the process mining manifesto, written by an IEEE taskforce, are used as a guideline in this chapter (Van der Aalst et al, 2011). These principles are discussed in more detail in Appendix I (literature study). All the models and datasets are included in the final deliverables for this project. The last part of this chapter will summarize and discuss all the findings in order to answer the following sub-question:

What is the added value of process mining in port call efficiency optimization?

4.1 Process mining input table

Data preparation

Before the dataset is used as an input for the process mining phase, it is cleaned and transformed into a structure that is fit for Disco, the process mining software. The data is then trimmed down to just the events that fit the scope of this study. The final dataset consists of more than 300.000 rows with events. A more detailed description of the desired data structure and how the data is transformed into this structure and trimmed down, is provided in appendix IV.

The entire process mining data set is directly inputted in Disco, which in a matter of minutes produced the process mining model.

4.2 Process mining to find a model structure for simulation

This part discusses the extracted model structure, through process mining, that will be used for the simulation model. During the literature review (Appendix I), no other studies were found that use this method to find a model structure for a simulation model, let alone studies that do this using AIS event data. The findings are compared to the model of the port call process defined by expert consultation in figure 2.14. This will assess the effectiveness of process mining for this purpose.

The model that results from process mining uses over 300.000 records of ten unique events to reveal over 25.000 unique port calls that took almost 7.000 unique paths through the model. The structure is visualized in figure 4.1, appendix figure IV-1 and appendix figure IV-2, with respectively, high-, medium- and low level of aggregation. This means that figure 4.1 shows the most commonly observed paths, where appendix figure IV-2 shows all the observed paths. The values in the figures show the frequencies of vessels that past those paths or events. The Disco model with the actual structure and the process mining input dataset are part of the final deliverables for this project. This chapter will discuss the most common paths and compare them to the structure identified by expert opinion and statistics found in the Python data analysis. Not all paths will be discussed, because there are almost 7.000 unique paths, however the (also less common) deviations will be highlighted.

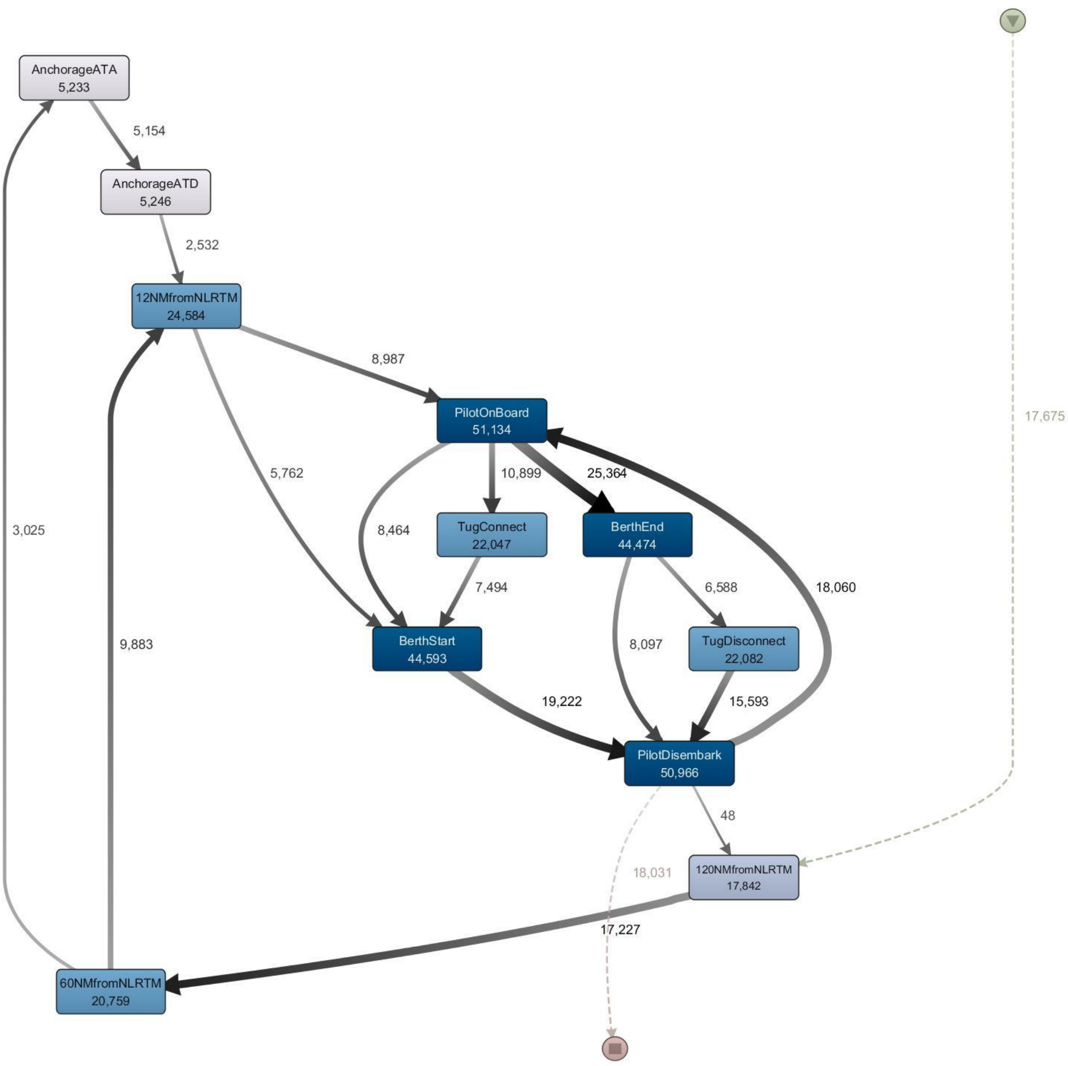

Figure 4.1: Process mining structure (level of aggregation: high)

The Disco model shows that most of the vessels enter the model at the 120 nmi, a lot enter at 60 nmi and a smaller part at the 12 nmi point. This matches the findings from figure 3.2. However, there are also vessels entering the model at the anchorage and at the pilot boarding place and then proceed to the 12 nmi point. This is explained by anchorage- and boarding places which are located further than the 12 nmi zone, as shown in figure 2.15. The 12 nmi point will still act as one of the entry points in the simulation model, in order to also simulate the vessels that come from a destination less than 60 nmi away. These particular cases will visit the anchorage- and boarding places in the simulation model, if necessary, after the 12 nmi zone.

After entering the model, a part of the vessels visits the anchorage while the other part does not (figure 4.1), which is in line with the model structure derived from expert consultation (figure 2.14). Next, the vessels that do not make use of tugs and pilots, proceed to BerthStart and then to BerthEnd. Vessels that do make use of one of these nautical services, go to the corresponding events (PilotOnBoard and TugConnect). After berth start, the PilotOfBoard and the TugDisconnect events take place. This is again followed by the PilotOnBoard and TugConnect, which are needed to fire the BerthEnd event. Feedback loops (figure 4.1) illustrate the port calls with multiple berth visits. This matches figure 2.14. Figure 4.1 is pretty hard to read for the nautical services and berth related events, because there are multiple unique paths with many loops visualized in one map. For this reason, the most commonly taken paths are discussed separately below. When these paths are combined, they result into the structure visualized in figure IV-3 or, on a more aggregated level, figure 4.1.

Three different categories of paths have been identified. The first category is a port call without the usage of pilots and tugs. The most frequently taken path in this category is shown in figure 4.2. This path is observed 1757 times, which is 6.9% of all the cases. The second most observed path (5.9%) is similar to figure 4.2, expect for the first event, which is missing. These are vessels that came from a port, less than 120 nmi away from the port of Rotterdam. Other variants in this category have multiple berth visits. This category is in line with previous findings (figure 2.14).

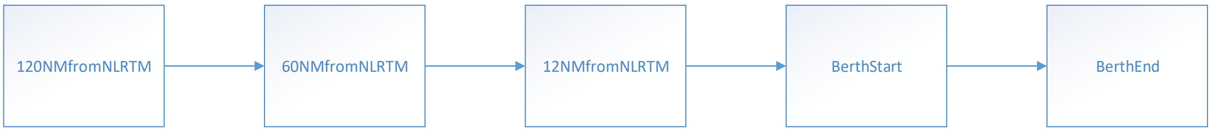

Figure 4.2: Process mining category 1: no pilots & no tugs

The second category of port calls uses pilots but does not use tugboats. The most frequently taken path in this category is shown in figure 4.3. This path is observed 829 times, which is 3.2% of all the cases. The second most observed path (3.1%) is the same as the path in figure 4.3, except for the 12NMfromNLRTM event, which has moved to the third place in the chain. Other variants in this category just show multiple berth visits. This category is in line with the expectations for almost all the events. The event that does not fit the expectations is the 12NMfromNLRTM event. The documentation made by Port of Rotterdam Authority for the events (Appendix III) shows that the 12NMfromNLRTM should follow the 120NMfromNLRTM and 60NMfromNLRTM events, however this is not the case. Many other paths that were identified with process mining also show the 12NMfromNLRTM event after the final BerthEnd event and before the PilotDisembark event. An explanation for this is that the PilotDisembark events are recorded after the vessels exit the 12 nmi zone and that for some reason the event identifiers wrongfully identify the 12NMfromNLRTM event. This should only be recorded for incoming vessels that enter the 12 nmi zone (Appendix III). The 12NMfromNLRTM event is the only event that in some cases shows odd behaviour. Therefore, there is reason enough to conclude that this deviation is a measurement error and that the model structure found using process mining is still in line with the previous identified structure (figure 2.14).

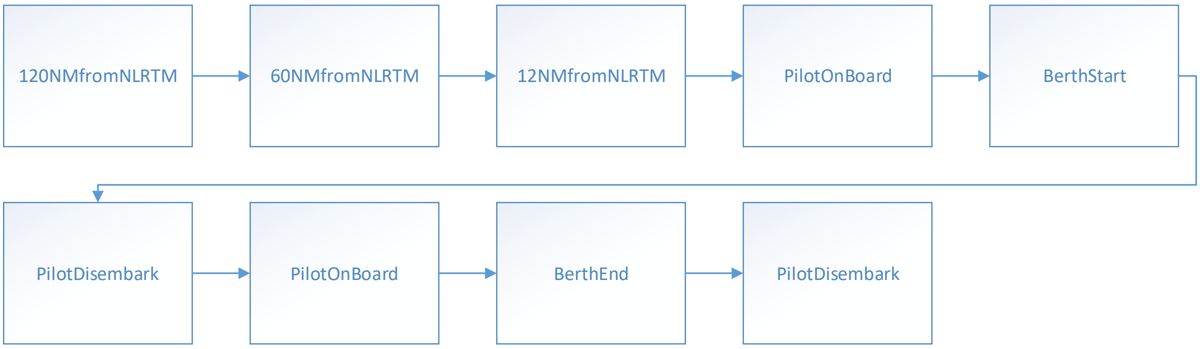

Figure 4.3: Process mining category 2: pilots & no tugs

The third and final category consists of port calls that make use of pilots and tugboats. The most frequently taken path in this category is shown in figure 4.4. It is observed in 479 cases which is 1.9% of the total. The second most observed path in this category (1.4% of all the cases) is almost the same as figure 4.4, only the second PilotOnBoard event and the second TugConnect event switched places. This is explained by the fact that these events take place at almost the same time. When looking at cases from this category, it is seen that these two events during the docking phase of a ship are just a few minutes apart from each other. Therefore, these findings are still in line with the expectations (figure

2.14). Other port calls from this category just vary in the number of berth visits, as is expected (figure 2.14).

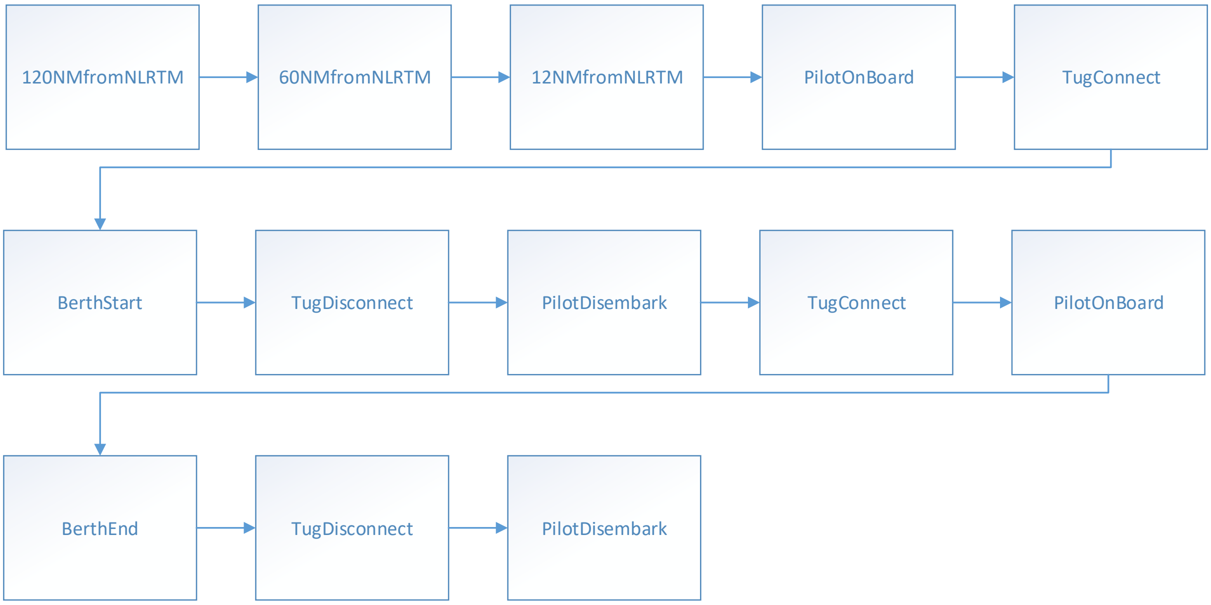

Figure 4.4: Process mining category 3: pilots & tugs

The process mining has resulted in a model structure that is broadly the same as the structure identified through expert consultation. The only event that shows odd behaviour is the 12NMfromNLRTM event, which in many cases is recorded near the end of the visit. In other words, it is recorded when a vessel exits the 12 nmi zone instead of when it enters it. In this case, it is clear that vessels do not actually enter the 12nmi zone (which is what the event should show) at the end of the port call, but that this is a measurement error. Even if this study did not include expert consultation, this error would probably be spotted. However, it still displays the potential weak point of process mining with the goal to identify model structure: dirty data. Other cases, where the errors in the data are more subtle, are definitely conceivable. The errors may not be always easy to spot and could pass as the truth. Therefore, a high level of alertness when interpreting the results is essential. Other than that, it is concluded that this technique shows high potential for successfully identifying model structures.

4.3 Possible applications of process mining AIS event data

Chapter 4.2 assessed only one specific application of process mining AIS event data. However, there are more applications. This part discusses these other applications that have the potential to in the end improve the port call process. This will be done for each type of process mining separately, as defined by the IEEE taskforce (figure 4.5) for process mining.

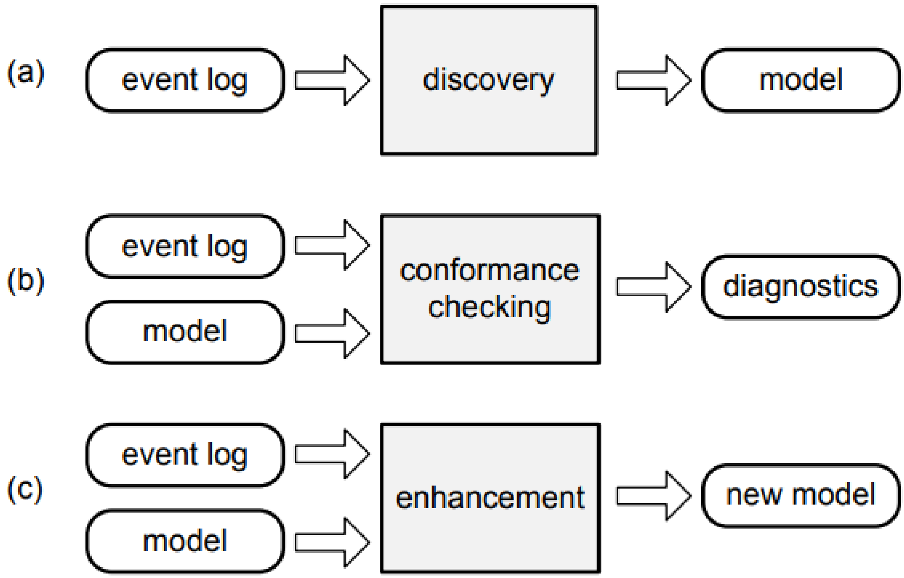

Figure 4.5: The three basic types of process mining explained in terms of input and output: (a) discovery, (b) conformance checking, and (c) enhancement (Van der Aalst et al, 2011).

The first type of process mining is discovery. This means process mining the AIS event data without any prior knowledge to create a model that can provide helpful information. Finding a model structure for simulation can be categorised as discovery. This is not the only application in this category. The model that results from process mining the AIS event data can be used as a tool to monitor a port's behaviour. With very little effort, it is possible extract statistical insights like time spent at the berth or anchorage for each vessel or for all the vessels together. This is visualized in the model over time, so different periods can be analysed separately to spot new trends. Furthermore, this monitoring property of the model makes it possible to quickly identify bottlenecks in the port. Especially, when looking at a new port without any prior knowledge, it could stand out that for example a certain type of vessel spends way more time at a certain step in the process than the other ones. This application provides similar results as the data analysis in Python in chapter 3, however it is a much smoother and cleaner process and less prior knowledge is required, for example, one does not need to be fluent in Python or R.

The second type of process mining is conformance checking. This is when an existing model is compared to the process mining model of the same process and vice versa. One application of conformance checking is already done in 7.1 where the process mining model is compared to the port call process model which was defined through expert consultation, to assess if the process mining model is suitable for directly extracting the model structure. Another application in this category is to check the performance of the event engine. The event engine is the collective name for the tools used to identify and extract the events from the raw AIS data. This process is described in more detail in chapter 3.1. This application of PM is already demonstrated in 7.1, where it was concluded that the event 12NMfromNLRTM was not behaving as expected. The event engine wrongly identified this event for a significant amount of port calls when a vessel was leaving the 12nmi zone instead of only when the vessel is entering it. These diagnostics contain very useful information because it shows the short comes of the event engine. The identified shortfalls can then be further investigated with the goal to improve the event engine.

The third type of process mining is model enhancement. The goal is to improve already existing models using information about the actual process from event logs. There are two main applications for the AIS event data identified in this category. The first one is to use the process mining model as a validation tool for an already existing model. The model behaviour can be compared to the process mining model and tweaked where necessary. This application can be used for the simulation model described in chapter

5. The second application in this category is the extension of already existing models, for example the simulation model built for this research (chapter 5). The simulation model is built in such a way that it is easily expandable. For example, including bunkering events in the model would be a logical next step. By first including bunkering events in the process mining model, which are already recorded and available in the AIS event data, the modeller would get good insights in how these events fit in the model structure and how they behave. The process of adding this to the simulation model would be guided by the process mining model and should be smoother.

4.4 Discussion process mining

The second objective of this research is to explore the possibilities of process mining the AIS event data. Process mining is a relative new technique that makes it easy to analyse big chunks of log data. One of the end products of the Python data analysis is the process mining input schedule, which holds over 300.000 records of 11 unique events that directly relate to the events described in the port call process map (figure 2.14). The following sub-question was formulated in the beginning of this research to achieve the above.

What is the added value of process mining in port call efficiency optimization?

This question is answered by putting the process mining technique into practice using the PM software Disco on the process mining input schedule. Additional possible PM implications in this domain are further theoretically explored following the PM guidelines setup by the IEEE PM taskforce, who composed a process mining manifesto (Van Der Aalst, 2011).

The first component of the answer to this sub-question comes from the Disco model. The prime product of the Disco model is a model structure that is used for the simulation model. To assess whether this technique is suitable for extracting a system structure for log data, it is compared to the process map which resulted from expert interview. It is concluded that the two structures are very similar and that they result in a similar simulation model. This makes it possible to perform this research on any other port of which AIS event data is tracked, without the need for expert consultation or on the spot exploration. This potentially holds a lot of value for the PoR, whose main goal it is to create a new revenue stream by selling products to third parties like other ports, carriers and terminals.

The second implication of the PM of AIS event data is to use it to monitor a ports behaviour. PM has the potential to quickly provide statistical insights over time, furthermore, bottlenecks and inefficiencies can be spotted this way. However, this implication of PM is not explored in Disco. Like how the model structure from Disco (section 4.2) is compared to the port call process map (figure 2.14), the statistical insights from the Python data analysis (chapter 3.3) could be compared to the statistical insights provided by the Disco model. The reason why the actual comparison is not done in this research is directly related to the limited amount of time and the size of the project team, which is one person. However, it is definitely an aspect of process mining that should be investigated more thoroughly.

The third aspect of process mining which potentially adds value in the process of port call efficiency optimization is the conformance checking of the event engine. During the process mining phase one flaw of the event engine was detected. The Disco model made it easy to spot that the events of vessels entering the 12nmi zone of the Port of Rotterdam are clearly often in the wrong chronological place in relation to the other events. When looking at the most common paths taken by vessels, which is an insight that the Disco model automatically provides, it is seen that that particular event often comes at the end of the port call, which contradicts with the definition of the event (12nmi upon entering the port). The detection of this flaw already proves the potential of this implication of PM.

The fourth aspect of the PM of AIS event data, which has the potential to contribute to a better port call efficiency is using the PM model to validate and extend other models like the simulation model build for this research. Additional events can easily be added to the PM model to investigate how they fit in the bigger picture. And outputs of the simulation model can be used as an input for a new PM model which then can be compared to the original PM model, this will give a good indication of how valid the simulation model is. However, this implication of PM is not explored in this research, due to the limited amount of available time for this research. If the PoR decides to proceed with making and using simulation models, it is recommended to further explore this aspect of process mining.

The application specific limitations of PM are already discussed above. However, there are some limitations that are more general and that the PoR should take into account when implementing this technique to any extend. The biggest limitation is that the completeness and accuracy of the data directly affects the outcomes of the process mining. Being aware of the data quality and an high level of alertness when interpreting results from process mining is essential. However, this property of PM can also be used to the PoR's advantage, when using this technique for conformance checking, which is previously discussed in this chapter.

5. Simulation model

This chapter focusses on the discrete event simulation model in Simio, that will be used as a tool to test what-if scenarios that are expected to improve the port call efficiency.

This chapter highlights the most important parts of the steps that were needed to create the simulation model. A more detailed description, that includes the entire model conceptualization, specification, creation, verification and validation, can be found in Appendix V.

5.1 Model Ingredients

The conceptualisation (Appendix V-I) sets the scope of the model and the key performance indicators (KPI's). Furthermore, the specification (Appendix V-II) describes the model reductions and assumptions, that justify the scope. The simulation model is built with all these elements and the model ingredients, described below in this section.

The data analysis (chapter 3) is carried out, while keeping the modelling scope in mind. This resulted in the simulation input schedule. The simulation input schedule (Appendix V-II) forms the input of the deterministic part of the model. Each row represents an actual port call from 2018 with corresponding attributes. This schedule is a crucial ingredient for the model. It is essential to simulate all the port calls with their corresponding attributes as they actually happened. A easier solution would be to calculate averages or distributions for each of these attributes and set an inter arrival time for vessel arrivals in the model based on the amount of total visits in 2018. However, this would never lead to a valid model, as these attributes are dependent. With a simulation input schedule, it is possible to simulate peaks of demand. For examples days where more large container vessels arrived than on average, with bad weather conditions, which increases the demand for tugboats, would not be simulated, if averages and distributions would be used for these attributes. These days are the most interesting days, as the short comes of the current port call efficiency is most visible in days with peaks of demand.

The second ingredient that resulted from the data analysis (chapter 3) are the empirical distributions of the berthing durations of vessels. These distributions form the stochastic part of the simulation model. Together with the simulation input schedule, they form all the empirical input data that is directly linked to the model.

The third ingredient is the number of pilots in the model and number of tugs. The number of pilots is set to a fixed number of 30 pilots, which are available for the entire simulation run. The reasoning behind this is discussed in more detail in the model reductions (Appendix V-II). The number of tugs is set to 8 for tug company A and 7 for tug company B. This is based on the number of unique tugs that are active in one day, which is concluded from the AIS event data analysis, in combination with expert consultation. The assumptions made, regarding the simulated tugs can be found in Appendix V-II.

The fourth ingredient is the model structure. This is a product of the process mining phase, discussed in chapter 4 and is based on the event data. The model structure describes the path that the vessels take during a port call. It shows in what order different steps of a port call process are arranged and what exact paths are the most common paths. This model structure is backed up by the port call process map in figure 2.14, that is derived from expert consultation and literature review.

The last ingredient is the speeds for the vessels. This is not based on empirical data, as already explained in the model reductions (Appendix V-II). All the vessels, including tugs and pilots, sail with a speed of 10 knots in the port. The port starts at the tug area. Outside the port, the speeds depend on the vessel type. The uniform distribution based on the guidelines set by Marine Insight (2019) for the speeds outside the port are provided in table V-2.

5.2 Model creation

To get a first broad idea of how the model looks in the front view, a snippet is shown of the model while running in figure 5.1. It shows all the objects except for the entry points, which are further into the sea and the small container terminal which is further inland. Figure 5.1 shows terminals with berthing sports, the anchorage, operational clearance point, pilot boarding place, tug area and the paths that the vessels can take.

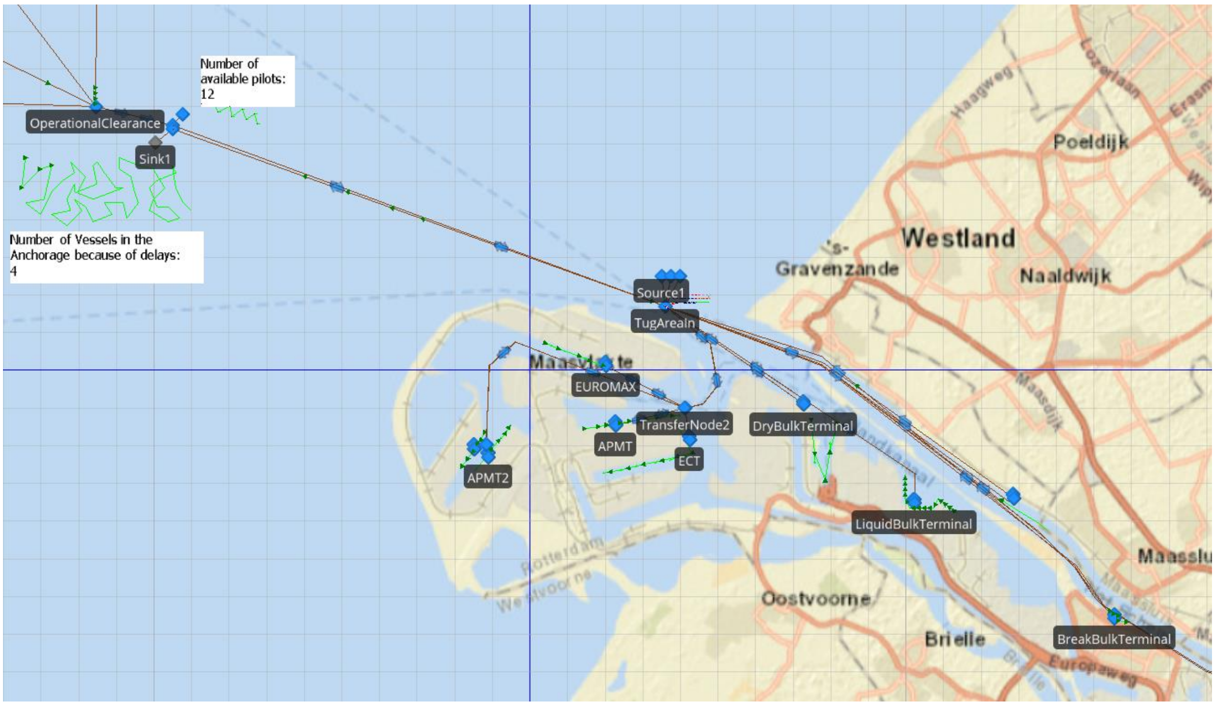

Figure 5.1: Snippet of front view of Simio model (entry points and small container terminal are missing in snippet)

For a more detailed description of all the model elements consult appendix V-III. This appendix also explains the modelling logic behind the most complex elements of the model.

5.3 Model verification and validation

During the verification- and validation phase, it is tested if the model is accurate enough and behaves in a credible manner. This is essential for the next phase of the research, which is scenario testing in the simulation model. If the model behaves inaccurately, it is not possible to translate the conclusions of the experiments to policy in the real world.

The verification is done by testing if the attributes are correctly included in the model. Next, simulation properties are checked and through a structural verification test, it is tested if the Simio model is consistent with the conceptualisation in appendix V-I.

The simulation model is validated in two ways. First, in a replicative validation test, the model outcomes are compared to empirical data. After that, an expert validation is performed, where a port call expert takes a detailed look at all the elements, steps and processes in the model, to conclude if the model performs accurately.

The verification and validation tests are performed and described in more detail in appendix V-IV. Some limitations are identified that should be taken into account during the interpretation of the experiment results in chapter 6. From the tests, it is concluded that the model is valid enough for simulation, especially for a relative comparison between two scenario's.

5.4 Discussion simulation model

The third objective of this research is to create a simulation model that can be used as a what-if analysis tool to test potential port efficiency improving measures. From every phase of this research, elements are used to create the model. The model is build based on the structure found in the process mining phase. The inputs for the model are provided in a form of an empirical simulation input schedule and empirical distributions which resulted from the data analysis in Python. The model is built in Simio and custom modelling logic is added to achieve the desired behaviour. This section discusses the simulation model, which is used in chapter 6 for scenario testing.

Once all the essential ingredients were gathered for the model, it was a relatively smooth process to create it. Nonetheless, a model is a simplification of reality and can always be made more complex with the goal to increase the validity or the scope of the model. Because of the limited amount of time, assumptions and simplifications are made when creating the model (Appendix V-II-II). The priority to improve certain aspects of the model is higher than other aspects. The ones with the highest priority are discussed here. It is recommended to improve the simulation of the simulation of the pilots. They correctly manoeuvre the vessels in and out of the port, so the modelling logic is deemed valid enough for simulation. However, for the number of pilots in the model, reductions and assumptions had to be made. In the simulation, every pilot services any ship that requests a piloting service. In reality, all the pilots have a set of permits which enables them to service only certain types of vessels. This was too complex to model, also the information needed to model this was lacking. Adding this layer of complexity is recommended before using the model to explore scenario's that directly involve the pilots. Secondly, the time a vessel spends at berth is simulated by sampling from a empirical distribution. There are no actual cargo services and ship services, like containers being unloaded or bunkering events, included in the simulation. This would be the next step when making the model more complex. It would make it possible to simulate scenario's that involve these critical services. Lastly, the amount of tugs active at each moment in time in the model is a fixed number. It is expected that this is not the case in reality, however, this research failed to extract that information from the AIS event data. Unlike with the cargo vessels, only averages are taken. Although the tugs are modelled in a way deemed valid enough for scenario testing, it is recommended to model this more accurately.

It is also beneficial to be aware of strong sides of the simulation model, when reviewing the results of the experiments. These are highlighted in this section. First of all, the simulation of the cargo vessels and their attributes over time is concluded to be very accurate in the validation (Appendix V-IV). Instead of using distributions to simulate the entrance of vessels in the model and their attributes, *actuals* are used. This way peaks in demand for various resources are simulated accurately. These peaks happen for example during bad weather, when multiple vessels request a higher number of tugboats than usually. These properties would get lost if distributions were used. In other words, the model successfully simulates multiple properties that are dependent on each other. A second strong point of the model is the interaction between vessels and the nautical services like tugboats and pilots which directly influences the port clearance. This modelling logic, created to simulate these aspects, is the most complex. There are still some flaws, which are discussed in more detail in the validation, but they are not expected to impact the model behaviour significantly. Next, vessels enter the model up to 120nmi away from the Port of Rotterdam. This makes it possible to test policies that include activities on sea. For example, lowering the speed when there is an expected delay. It is expected that this would save the vessels fuel. This scenario is not tested, but the model already has all the elements needed to simulate it. The last highlighted point is the fact that the modeller strived to build the model in such a way, that it is easily adaptable to other ports. This is not only done for the model itself, but also for the ingredients needed for the model. It would be a much smoother process to create the exact same simulation model for a different port, than the creation of this original model.

Taking all these weak and strong sides into consideration, it is concluded, during the validation, that the simulation model is valid enough for a relative comparison of two scenario. This is done in chapter 6.

6. Scenarios and experiments

This chapter focusses on policy testing, using the what-if analysis tool made in Simio (chapter 5). The Simio model is adjusted to simulate a scenario, that is expected to increase the port call efficiency. The results are statistically compared to the simulation results of the current situation.

First, a description of the scenario and an explanation, of how the measure is implemented in the simulation model, is given. Next, the configuration of the experiment is presented, followed by an overview of the results of the current situation and the results of the port efficiency increasing scenario. The results of these two models are then statistically compared.

This chapter answers the following sub-question:

What is the effect of the collaboration of tug companies on the delays of vessels in a port?

6.1 Scenario description

The policy, that is tested, results from analysing the port call process described in chapter 2. Potential bottlenecks and weaknesses in the process are identified and after expert consultation with a port call expert and data scientist a scenario is composed.

The nautical services of the Port of Rotterdam consist of the pilots, tugboats and linesman. A more detailed description of their exact roles can be found in chapter 2. By looking at the port call process map (figure 2.14), it is easily concluded that these actors play a huge role in the port call process and thus in its efficiency. The nautical services form critical resources for vessels that visit the port. At moments of peaks in demand for their services, delays can occur, due to the limited number of pilots, linesman or tugboats. An improvement of the availability of one of these services, is expected to result in less delays for incoming vessels and thus lead to a higher port call efficiency. Therefore, a policy is designed that improves the availability of a nautical service.

When comparing the governance of the three services, the tugboats stand out. The tug service is the only one that consists of multiple companies. There are situations where vessels are waiting for tugs, even though there are tugs available in the port, they just happen to be from a different company than that the vessel has a contract with. Therefore, it is interesting to see what would happen with the delays if hypothetically the tug companies worked together. This scenario, where all the tugs operate as one fleet, is tested in the simulation tool.

The process to implement this policy is as follows: In the simulation model, the initial number of tugboats of the two companies is summed. One of the two companies gets this summed value as their new initial number of tugboats and the other one gets zero as their new value. Next, all the add-on processes in the model that have tug company related steps are adjusted so that only the branches of the processes are reached that are related to the tug company that now represents the two companies combined. This way there is only one big company in the model, which simulates the collaboration of the original two companies.

The model of the new scenario is part of the final deliverables of this project.

6.2 Expected results

This section discusses the expected results of the experiments with the current situation and the new scenario. This is done by comparing the expectations per key performance indicator. The exact meaning of the KPI's is discussed in appendix VI-I. If there are no significant differences found in the statistical comparison (6.4) for KPI's where there is no difference expected, the reliability of the model increases. This works the same the other way around.

Throughput time operational port call

The average throughput time for the container vessels for the operational part of the port call is expected to decrease significantly in the new scenario. The scenario is designed with the goal to reduce delays. If this would be the case, the total throughput time would decrease as well, which improves the port call efficiency.

Total number of anchorage visits

The total number of anchorage visits of the container vessels is expected to decrease significantly. There are three possible reasons why a container vessel visits the anchorage: no available berth spot at their destination, no available pilot, no available tugs. The scenario is expected to increase the availability of tugs, therefore is expected to reduce the total number of anchorage visits.

Total average delay at anchorage

It is unknown whether the average delay at the anchorage will increase or decrease. Only container vessels that actually visit the anchorage are included in the calculations. Furthermore, it is unknown which of the three possible reasons to visit the anchorage, causes the most delay. Therefore, it could happen that the average delay will increase, even though the total delay of all the vessels combined decreases, because less container vessels are expected to be delayed by a lack of available tugboats. However, it is not possible to predict this with the current knowledge.

Total average delay at berth

The total average delay at berth for container vessels is expected to decrease in the new scenario. The scenario is designed with the goal to reduce delays caused by the lack of available tugboats. This is one of the reasons that a delay occurs. Therefore, it is easy to conclude that the total average delay at berth is expected to decrease.

Availability of the pilots

The availability of the pilots is not expected to increase or decrease significantly. The pilots are not directly affected by the implemented scenario.

Availability of the tugboats

The availability of the tugboats is not expected to increase or decrease significantly. The two scenarios have the same amount of port calls that need to be escorted by tugboats. It is only expected that the workload is more equally spread over the tugboats, however this would not influence the availability.

6.3 Results

The Simio model successfully ran the experiments. The simulation time was approximately 5 minutes per replication. With 40 replications for each scenario, this resulted in a total simulation time of less than 7 hours, which is more than acceptable. This section shows a summary of the results of the experiment of the current scenario and the new scenario. The differences between the scenarios, per KPI are presented as well, with the purpose to give a first impression of the effects of the new policy. Also, a percentual difference relative to the results of the current situation is given. Whether the scenarios have significant differences is tested in 6.4. Table 6.1 shows a summary of the results of the experiments, per KPI the average of the 40 replications is shown. A detailed output of the Simio model is provided as one of the deliverables for project.

Key performance indicator	Current scenario	Collaborating tug scenario	Mean difference	Percentual difference
Throughput time operational port call (hours)	Mean: 50.65 St. dev.: 0.24	Mean: 50.65 St. dev.: 0.21	+0.00	+0%
Total number of anchorage visits (#)	Mean: 2313 St. dev.: 78.65	Mean: 1994 St. dev.: 69.47	-319	-13.79%
Total average delay at anchorage (hours) (only vessels that actually visit the anchorage)	Mean: 15.98 St. dev.: 1.53	Mean: 17.70 St. dev.: 1.81	+1.72	+10.76%
Total average delay at berth (hours)	Mean: 0.03 St. dev.: 0.00	Mean: 0.02 St. dev.: 0.00	-0.01	-33.33%
Availability of the pilots (%)	Mean: 54.74% St. dev.: 0.10	Mean: 54.91% St. dev.: 0.11	+0.17	+0.31%
Availability of the tugboats (%)	Mean: 70.26% St. dev.: 0.17	Mean: 70.91% St. dev.: 0.14	+0.65	+0.93%

Table 6.1: Summary experiment results

6.4 Statistical analysis

In order to be able to draw conclusions from the experiments, the results of the simulations are statistically analysed and compared. This is shown in this section.

First the method and approach for the statistical analysis is described. This contains all the potentially needed statistical tests and how they relate to each other. A more detailed description of the tests that are used can be found in appendix VI-I. Next, the most important findings are presented. This is followed by a discussion of the results. For a overview of all the performed statistical tests, see appendix VI-II.

Method and approach

Figure 6.1 shows a chart that assists in determining the best fitting statistical tests. First Shapiro-Wilk tests are performed to conclude whether it is feasible that the datasets originate from a population with a normal distribution. This is necessary for determining what test to use when comparing the two scenarios. When the Shapiro-Wilk test concludes that a KPI is not normally distributed, the Wilcoxon rank test is used to compare that KPI for each scenario. From this it is concluded whether there is a significant difference between the two scenarios. When the Shapiro-Wilk test concludes that a KPI is normally distributed, a Levene’s test is performed to check for equal variance. In the case of equal variance, a Student t-test is performed, otherwise a Welch t-test is performed, both with the goal to test whether the two scenarios have a significant difference for a KPI. For all tests, a confidence interval is chosen of 95%. All the analysis are executed using IBM SPSS Statistics. This is a software for analysing data and performing statistical tests. It includes all the tests mentioned in figure 6.1.

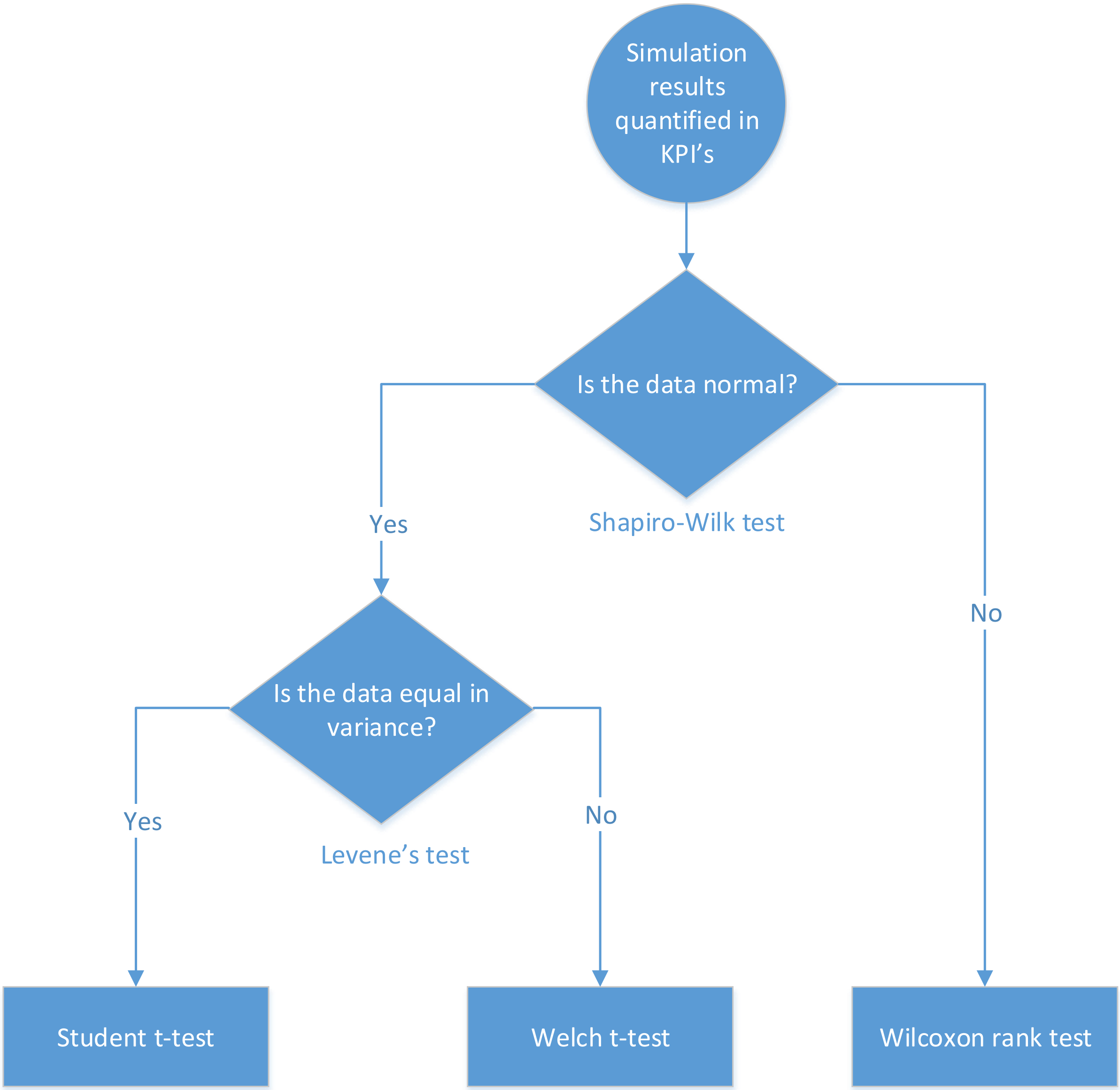

Figure 6.1: Chart for statistical tests

Results statistical analysis

The results for each KPI are analysed separately (appendix VI-II). This section reports the most important findings from the analysis

Throughput time operational port call

Figure 6.2 shows plots of the results of the throughput time of container vessels. The independent sample t-test is performed to conclude whether there is a significant difference. The results are shown in table 6.2. The p-value is greater than 0.05. This means that the null hypotheses is accepted. In other words, there is not a significant difference in the throughput time of the operational part of a port call for the two scenarios.

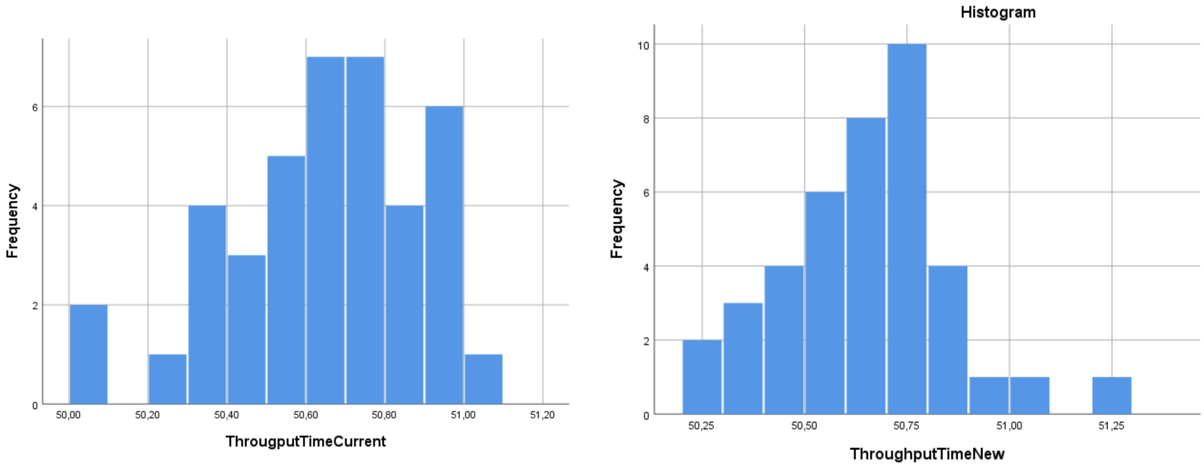

Figure 6.2: Histogram throughput operational port call

	t-test for equality of means		
	t	df	Sig.
Through put time	-0.080	78	***0.973***

Table 6.2: Independent samples t-test throughput time

Total number of anchorage visits

A visual representation of total anchorage visits of container vessels for the two scenarios is shown in figure 6.3. The independent sample t-test is performed. The results are shown in table 6.3. The p-value is smaller than 0.05. This means that the null hypothesis is rejected. In other words, there is a significant difference in anchorage visits for the two scenarios.

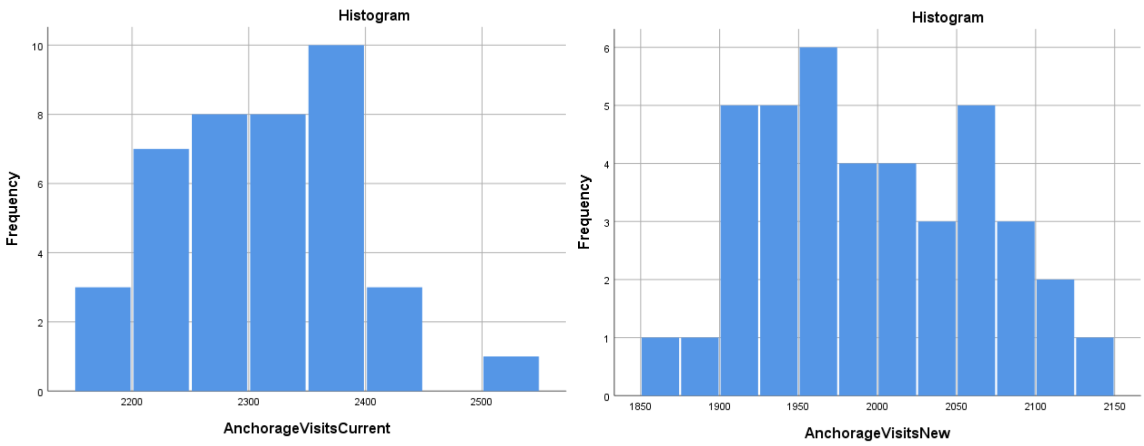

Figure 6.3: Histogram total number of anchorage visits

	t-test for equality of means		
	t	df	Sig.
Anchorage visits	19.225	78	***0.000***

Table 6.3: Independent samples t-test anchorage visits

Total average delay at anchorage

A visual representation of the delay at the anchorage for the two scenarios is shown in figure 6.4. The Wilcoxon signed ranks test is performed and the results are shown in table 6.4 and 6.5. The test determined that in 11 replications the delays are bigger in the current scenario and in 29 replications that they are bigger in the new scenario. This indicates a significant difference in mean. This suspicion is confirmed in table 6.5. A p-value smaller than 0.05 is calculated, which means that the null hypotheses is rejected. It is concluded that there is a significant difference between the current and new scenario when it comes to delays at the anchorage.

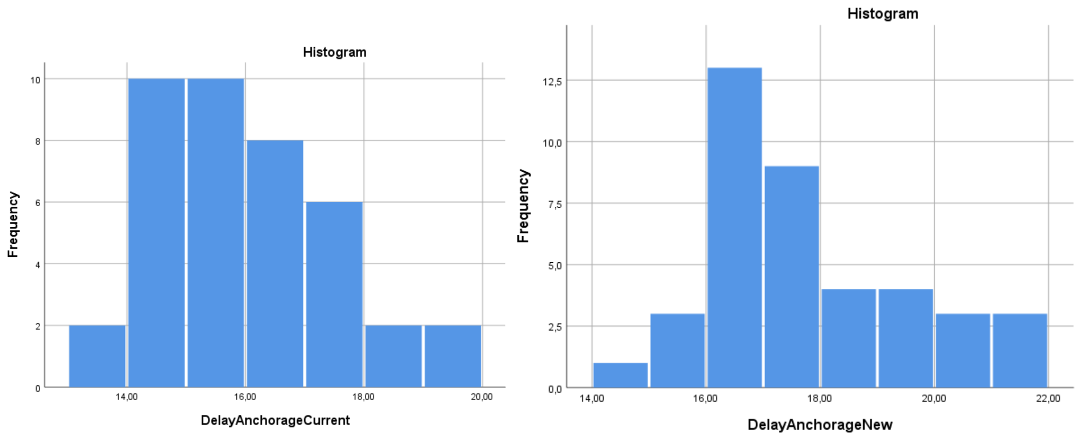

Figure 6.4: Histogram delay at anchorage

		N	Mean Rank	Sum of Ranks
DelayAnchorageNew - DelayAnchorageCurrent	Negative Ranks	11[a]	11,45	126,00
	Positive Ranks	29[b]	23,93	694,00

a. DelayAnchorageNew < DelayAnchorageCurrent

b. DelayAnchorageNew > DelayAnchorageCurrent

Table 6.4: Wilcoxon signed ranks test delay anchorage

	DelayAnchorageNew - DelayAnchorageCurrent
Z	-3,817
Asymp. Sig. (2-tailed)	,000

Table 6.5: Wilcoxon test statistics delay anchorage

Total average delay at berth

A visual representation of the delay at berth for the two scenarios is shown in figure 6.5. The Wilcoxon signed ranks test is performed and the results are shown in table 6.6 and 6.7. The test determined that in 40 replications the delays are bigger in the current scenario. This indicates a significant difference in mean. This suspicion is confirmed in table 6.7. A p-value smaller than 0.05 is calculated, which means that the null hypotheses is rejected. It is concluded that there is a significant difference between the current and new scenario when it comes to delays at the berth. Table 6.1 suspects a decrease in delay. In combination with the proved significant difference, it is assumed that the delay at berth will decrease as a result of collaboration of tug companies.

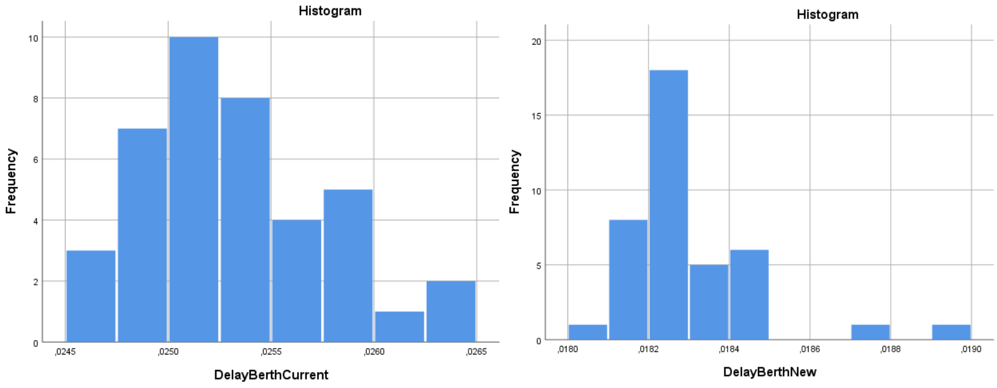

Figure 6.5: Histogram average delay at berth

		N	Mean Rank	Sum of Ranks
DelayBerthNew - DelayBerthCurrent	Negative Ranks	40[a]	20,50	820,00
	Positive Ranks	0[b]	,00	,00

a. DelayBerthNew < DelayBerthCurrent

b. DelayBerthNew > DelayBerthCurrent

Table 6.6: Wilcoxon signed ranks test delay at berth

	DelayBerthNew - DelayBerthCurrent
Z	-5,511
Asymp. Sig. (2-tailed)	,000

Table 6.7: Wilcoxon test statistics delay at berth

Availability of the pilots

A visual representation of the availability of the pilots for the two scenarios is shown in figure 6.6. The independent sample t-test is performed. The results are shown in table 6.8. The p-value is smaller than 0.05. This means that the null hypotheses is rejected. In other words, there is a significant difference in the availability of pilots for the two scenarios.

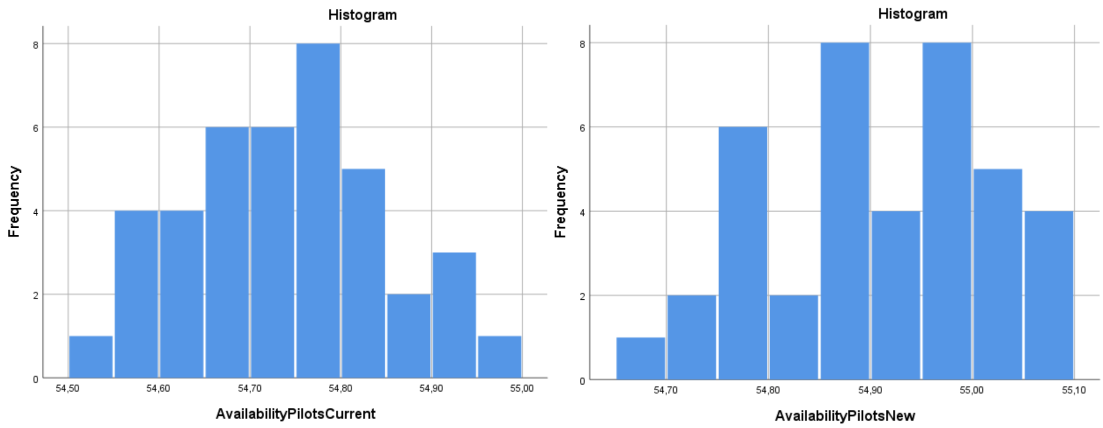

Figure 6.6: Histogram availability pilots

	t-test for equality of means		
	t	df	Sig.
Anchorage visits	-7.114	78	***0.000***

Table 6.8: Independent samples t-test availability pilots

Availability of the tugboats

A visual representation of the availability of the tugs for the two scenarios is shown in figure 6.7. The independent sample t-test is performed. The results are shown in table 6.9. The p-value is smaller than 0.05. This means that the null hypotheses is rejected. In other words, there is a significant difference in the availability of pilots for the two scenarios.

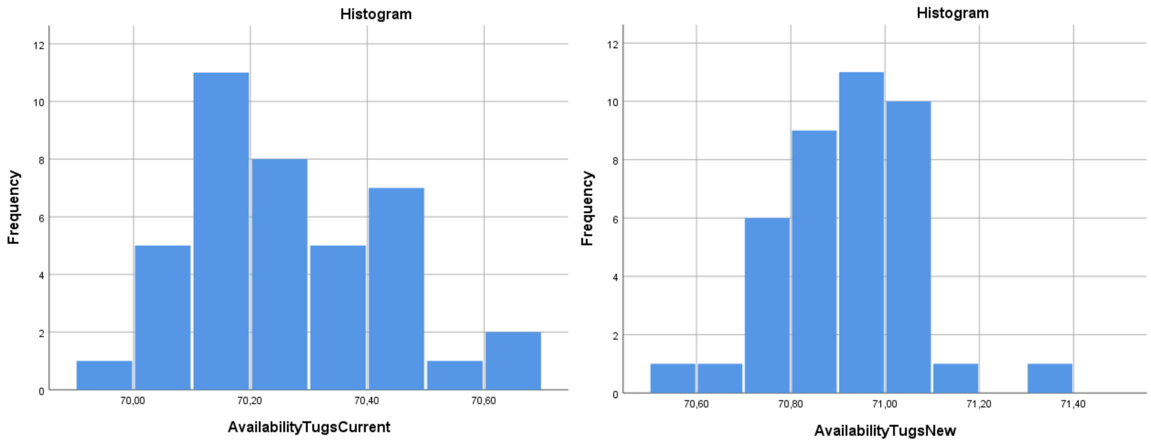

Figure 6.7: Histogram availability tugs

	t-test for equality of means		
	t	df	Sig.
Anchorage visits	-18,784	78	***0.000***

Table 6.9: Independent samples t-test availability tugboats

6.5 Discussion experiments

The third objective of this research is to create a simulation model that can be used as a what-if analysis tool to test potential port efficiency improving measures. To illustrate how this tool can be used a scenario is tested. The scenario that is tested resulted from identifying potential bottlenecks in the port call process defined by expert consultation. To demonstrate the implementation of the model, it is tested what the effects on the port call efficiency would be in the Port of Rotterdam if the two tug companies would start working together. This section answers the following sub-question, by discussing the interpretation of the experiments and whether it is feasible to implement the measure in the real world:

What is the effect of the collaboration of tug companies on the delays of vessels in a port?

Interpretation of experiment results

This paragraph discusses how to interpret the results from the experiments in the Simio model and how to translate them to real world actions. An overview of the results and the statistical analysis of the KPI’s are presented in chapter 6.3 and 6.4. First the insights provided by each KPI are discussed, which also includes a reflection on the expected results in chapter 6.2.

From the statistical analysis, it is concluded that there is not a significant difference between the current scenario and the new scenario with collaborating tug companies for the throughput time of the operational part of a port call of container vessels. This is not in line with the expectations (chapter 6.2). This could be explained by the fact that the largest part of the of the port calls are not delayed (table 3.1). This policy only affects the part of the delayed port calls, that are delayed by a lack of available tugboats and this KPI shows the throughput time of all the port calls. Apparently, it does not affect the total container vessel population that significantly that the throughput time for all the port calls decreases as well. It is still expected that the tug collaborating policy has a significant effect on the part of the port calls that is delayed.

The total average delay at the anchorage of container vessels has a significant difference between the two scenarios. In combination with an average increase of 1.72 hours, which is a percentual increase of 10.76%, it is concluded that the average delay at the anchorage for container vessels increases in the scenario with collaborating tugs. This is explained by the fact that only vessels that go to the anchorage are included in the calculations of this KPI. Apparently, the vessels that first were delayed, because of a lack of tugboat availability, but are not delayed anymore in the new scenario, had on average a smaller delay than the average delay of all the vessels in the anchorage. To give a better explanation of why this is happening the following simplified example is provided: Vessel A is delayed because of the lack of a berth duration for 15 minutes and vessel B is delayed because of the lack of available tugboats for 5 minutes. The average delay of vessels that visit the anchorage is (15+5) / 2 = 10 minutes. Now imagine that, by implementing a policy that increases the availability of tugs, vessel B is not delayed anymore and does not visit the anchorage. The average delay at the anchorage is (15) / 1 = 15. The average delay increases, even though the total delay decreases from 20 to 15. Therefore, it is concluded that this KPI on its own can create a misleading image of the situation. To get a clear picture of the effects of tugs collaborating, this KPI has to be viewed in combination with the KPI that describes the total number of anchorage visits of container vessels.

From the statistical analysis, it is concluded that there is a significant difference in anchorage visits for the two scenarios. Table 6.3 shows that, on average, there are almost 14% less anchorage visits in the new scenario. The combination of a significant difference and the negative difference in mean confirms the expectations. It is concluded that the policy with the collaborating tugs significantly lowers the number of anchorage visits for container vessels. Table 6.1 shows, for the current situation, that there are on average 2313 visits to the anchorage and that the visits last for 15.95 hours on average. In the new scenario, there is an average 1994 visits which last 17.70 hours on average. The total time spent at the anchorage for all the container vessels combined is 15.95*2313=36892 hours for the current scenario and 17.70*1994=35294 hours for the new scenario. This shows an average decrease of 1598.55 hours (4.33%) of the total anchorage time of container vessels. This is a promising result and shows that the collaboration of tugboats has the potential to improve the port call efficiency.

The statistical analysis concludes that there is a significant difference between the current scenario and the new one, when it comes to average delay at berth of container vessels. Table 6.1 suspects a decrease in delay of 33%. In combination with the proved significant difference, it is assumed that the delay at berth will decrease as a result of the collaboration of tug companies. The percentual decrease of the delay at berth is bigger than the percentual decrease of the delay in the anchorage. This has two reasons. The first is that vessels at berth, that plan to leave the port, cannot be delayed by the lack of a berth spot, because they do not need one. The second reason is a shortfall in the modelling logic caused by a model reduction. This model reduction lets vessels visit the same terminal multiple times, if they have multiple destinations in the port. This results in an unforeseen consequence: vessels at berth are never delayed because of the lack of a berth spot. Therefore, the percentual difference in average delay at berth in the new scenario is likely to be lower in reality, because the delays of the vessels at berth, caused by the lack of a berth spot in their next port destination are missing in the simulation. However, it is still believed that the delays at berth will decrease if this scenario with collaborating tug companies is implemented, because the delay decreases as well in the anchorage.

The statistical analysis concluded that there is a significant difference in the availability of pilots. This is not in line with the expectations in chapter 6.2. It was not expected that the scenario would significantly influence the pilot usage. Table 6.1 shows that the percentual difference of availability of pilots is only 0.31%. Therefore, it is assumed that the difference is significant but has low impact on the entire system. In combination with the fact that there are major assumptions made for simulation of the pilots, this result should not be taken to heavily.

It is concluded from the statistical analysis that there is a significant difference in the availability of tugboats in the new scenario. This is not in line with the expectations (chapter 6.2). It was expected that the occupation rate will be more evenly spread over the tugboats, but that the total availability will stay the same. Table 6.1 predicts an average increase of almost 1% in availability. It is expected that this predicted increase would only have a positive effect on the port call efficiency from a cargo vessel's perspective.

Implementation of measure in real world

From the interpretation of the results of the experiments is concluded that the scenario with collaborating tug companies shows potential. Even though it does not significantly decrease the total throughput time of container vessels, it does significantly impact the delayed vessels. The total waiting time in the anchorage illustrates the potential benefits of this new scenario, as it is predicted to decrease with 4.33%. For a port that handles millions of tonnes of cargo every year (Port of Rotterdam, 2019h), a measure that could reduce the delays of container vessels for a couple of percent is definitely worth looking into. However, the question arises, whether it is feasible to actually implement this measure in the real world. A collaboration or the union of other nautical services is not a rare phenomenon in the Port of Rotterdam. The pilots and linesmen are both organized as one private party. This suggests that it could also be done with the tug companies or that they at least could collaborate. This was actually the case in the past. The tug companies used to work together with the main goal of increasing the efficiency of their joint fleet (Mackor, 2014). However, the Dutch Authority for Consumers and Markets, who is an independent regulator that champions the rights of consumers and businesses (ACM, 2019), blamed the companies of forming a cartel which prevented other potential parties of joining the PoR tug market. This was not only for their activities in the Port of Rotterdam, but also for their activities in the Port of Hamburg. This resulted in prohibition of cooperation and fines of in total more than € 17 million, spread over multiple tug companies (Lin, 2018). Therefore, it is concluded that a collaboration in the current system is not realistic.

7. Conclusions and recommendations

The goal of this chapter is to draw the main conclusions of this research and to state the recommendations for future work. The most relevant elements of the discussion sections of the previous phases (3.4, 4.4, 5.4 and 6.5) are combined to provide a more general answer to the main research question:

How can data analysis, process mining and discrete event simulation contribute to identifying and assessing policies that improve port call efficiency?

7.1 Answer to main research Question

After the qualitative analysis and a data analysis, which are successful in mapping the behaviour of the vessels in a large sea harbour, a combination of two methods is explored: process mining and discrete event simulation. From the literature research, it is concluded that the implementation of these two methods in the maritime transport domain was uncharted and showed potential for contributing to a higher port call efficiency.

It is concluded from this research that there are multiple ways to benefit from process mining when analysing event data in the maritime transport domain. First of all, process mining makes it possible to map and monitor a ports behaviour with less effort than through a more traditional data analysis as is done in chapter 3. The clear empirical insights that PM provides, makes it a smooth process to spot bottlenecks in the port, from which policies can be derived, with the potential to improve the PCE. After comparison with the port call process map, that resulted from a qualitative analysis, it is concluded that the quantitative PM method is successful in finding a similar result. Thus, expert knowledge is not essential to get a grip on the process of a port call, when using this method. Furthermore, PM has proven to be an effective technique for conformance checking of the event engine. A major port has thousands of port calls a year, which each generate dozens of events from multiple sources. PM is an excellent tool for creating a good overview, seeing how the events relate to each other and spotting shortfalls in this complex system. However, when implementing PM, a high level of alertness is required. The reliability of the insights that process mining provides is directly related to the quality of the data. Therefore, being aware of this quality is essential.

To find a way to assess the identified measures that have the potential to improve the port call efficiency, discrete event simulation is explored. This research shows and concludes that it is possible to create a simulation model based on event data and a model structure derived from PM, that is valid enough to perform relative comparisons of scenarios. After identifying a policy based on inefficiencies observed in the PM phase, it is implemented in the model and compared to the current situation. This research tested a policy in which the tug companies collaborate. However, other policies can already be implemented in the model as well. It is concluded that the model is able to simulate policies that include a change in the number of vessels, terminals and pilots. The model is built in a modular way, which makes it possible to extend the model even further which enables the simulation of other policies that are not yet included. Well defined key performance indicators show the significant effects of a policy on the port. For example, in the tested policy, the number of anchorage visits in one year dropped from 2313 to 1994. This shows that the policy has potential to improve the port call efficiency. This way a measure can be assessed through simulation in a quantitative and very structured way, before it is implemented in the real world, even though the modelled system is a complex environment. Lastly, it is concluded that it is essential to be aware of the weak and strong sides of this simulation tool when interpreting the results. This increases the reliability of the results.

During the translation of the simulation results to real world decision making, it is concluded that an actor analysis is an effective way for establishing whether it is possible to implement the tested policy in the real world. Therefore, a good understanding of the complex multi-actor environment, in combination with a method that predicts the quantitative effect of policies on the performance of the port, improves the quality of decision making regarding port call efficiency measures.

7.2 Further research

This research identifies multiple applications of process mining. Some are explored in a hands on quantitative way, which provided useful insights, for other applications it is only described how they may provide value. Further research is needed to determine the exact benefits of this second group of process mining applications. More concretely, it is recommended to explore the use of process mining as a validation tool for the simulation model. It is also recommended to compare the statistical insights that various process mining tools provide with the statistical insights that data analysts of major ports use for their analyses. This will further establish how PM can contribute to a better PCE.

A simulation model is a simplification of the reality. These simplifications are a direct source of inspiration for future work. The next logical step is to increase the complexity of the simulation by including the simulation of critical services at berth, like cargo handling and bunkering. This will increase the number of policies that can be tested in the model. Next, a research that explores more complex ways of mapping the behaviour and operations of pilots is recommended. The exact number of pilots and what permits they have, was an unknown fact in this research. Furthermore, it is suggested to extend the model so that it is possible for a vessel to visit different terminals in one port call.

Next, it is recommended to keep optimizing the event data. All the quantitative analysis in this research are event data driven. The reliability of these methods is directly related to the quality of data. Expanding the number of unique events is also suggested. More concretely, it is recommended to explore the possibilities of identifying events that specify when a delay begins and ends. Now it is rather complex to identify the delays of vessels based on the data, especially the non-container vessels. Having events of delays would enable a more precise calibration of the simulation model.

References

ACM. (2019). *Our Organization.* Retrieved from https://www.acm.nl/en/about-acm/our-organization/the-netherlands-authority-for-consumers-and-markets

Almaz, O. A., & Altiok, T. (2012). Simulation modeling of the vessel traffic in Delaware River: Impact of deepening on port performance. Simulation Modelling Practice and Theory, 22, 146-165.

Alphaliner. (2019). *Alphaliner TOP 200 / 26 Mar 2019.* Retrieved from https://alphaliner.axsmarine.com/PublicTop100/

Bichou, K. Gray, R. (2004) *A logistics and supply chain management approach to port performance measurement, Maritime Policy & Management,* 31:1, 47-67.

Brooks, M. R., & Schellinck, T. (2015). Measuring Port Effectiveness: Does supply chain partner performance measurement matter? Transportation Research Record (2479), 42-48.

DBS. (2019). *AIS Event Data.* Available from AWS Database.

De Monie, G. (2009). "Measuring and Evaluating Port Performance and Productivity," CEPAL Review, No. 99, 2009, p.173.

Davidrajuh, R., & Lin, B. (2011). Exploring airport traffic capability using Petri net based model. Expert Systems with Applications, 38(9), 10923-10931.

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. National science review, 1(2), 293-314.

Fluxicon. (2019). *Discover your processes.* Retrieved from https://fluxicon.com/disco/

Google Scholar. (2019). https://scholar.google.nl/

GS1. (2019). https://www.gs1.org/standards

Herty, M., & Klar, A. (2003). Modeling, simulation, and optimization of traffic flow networks. SIAM Journal on Scientific Computing, 25(3), 1066-1087.

Hild, R. (2000). DISCRETE EVENT SYSTEM SPECIFICATION (DEVS) DISTRIBUTED OBJECT COMPUTING (DOC) MODELING AND SIMULATION. A Dissertation Submitted to the Faculty of the DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING.

Idris, I. (2014). Python data analysis. Packt Publishing Ltd.
IEC. (2001). 61993-2 Clause 6.2

International Taskforce Port Call Optimization. (2018). *Appendix to Port Call Process.* Retrieved from https://portcalloptimization.org/

John Good Shipping. (2019). *What is a Shipping Agency or Shipping Agent?* Retrieved from https://www.johngood.co.uk/2018/03/15/shipping-agency-definition/

Jumelet, M. Van Strien, A. (2019). *Rotterdam and Maersk Partner for Port Call Optimization.* Retrieved from https://www.porttechnology.org/news/rotterdam_and_maersk_partner_for_port_call_optimization

Kennedy, O. R., Lin, K., Yang, H., & Ruth, B. (2011). Sea-port operational efficiency: an evaluation of five Asian ports using stochastic frontier production function model. Journal of Service Science and Management, 4(3), 391.

Kim, C. J., Akinbodunse, D. A., & Nwakamma, C. (2005, November). Modeling Arrival Flight Traffic Using Arena®. In CAINE (pp. 9-14).

Kiprop, V. (2018). *The Busiest Cargo Ports In Europe*. Retrieved from https://www.worldatlas.com/articles/the-busiest-cargo -ports-in-europe.html

Knowler, G. (2018a). *Rotterdam barge congestion still causing up to 48-hours delays.* Retrieved from https://www.joc.com/port-news/rotterdam-barge-congestion-still-causing-48-hour-delays_20180418.html

Knowler, G. (2018b). *Europe's busiest ports search for answers to barge congestion.* Retrieved from https://www.joc.com/port-news/european-ports/port-rotterdam/europe%E2%80%99s-busiest-ports-search-answers-barge-congestion_20180927.html

Laganá, D., Legato, P., Pisacane, O., & Vocaturo, F. (2006). Solving simulation optimization problems on grid computing systems. Parallel Computing, 32(9), 688-700.

Legato, P., Canonaco, P., & Mazza, R. M. (2009). Yard crane management by simulation and optimisation. Maritime Economics & Logistics, 11(1), 36-57.

Levene, H. (1960). Contributions to probability and statistics. Essays in honor of Harold Hotelling, 278-292.

Lin, L. (2018). *Havensleper Kotug schikt voor miljoenen euro's in kartelzaak.* Retrieved from https://www.ad.nl/rotterdam/havensleper-kotug-schikt-voor-miljoenen-euro-s-in-kartelzaak~a5da0289/

LV, B., & NIU, H. M. (2011). Setting Condition and Effect Analysis of Bus Lanes at Signalized Intersection [J]. Journal of Lanzhou Jiaotong University, 6.

Mackor, R. (2014). *Kartelonderzoek slepers in Rotterdam en Hamburg.* Retrieved from https://www.nieuwsbladtransport.nl/scheepvaart/2014/12/03/kartelonderzoek-slepers-in-rotterdam-en-hamburg/

Marine Insight. (2019). *What is The Speed of a Ship at Sea?* Retrieved from https://www.marineinsight.com/guidelines/speed-of-a-ship-at-sea/

MarineTraffic. (2019). MarineTraffic. Retrieved from https://www.marinetraffic.com/en/ais/home/centerx:4.027/centery:52.029/zoom:12

Maritime and Port Authority of Singapore. (2009). *Bunkering.* Retrieved from https://web.archive.org/web/20150107064443/http://www.mpa.gov.sg/sites/port_and_shipping/port/bunkering/bunkering.page

Maruster, L., Weijters, A. T., Van der Aalst, W. W., & van den Bosch, A. (2002, November). Process mining: Discovering direct successors in process logs. In International Conference on Discovery Science (pp. 364-373). Springer, Berlin, Heidelberg.

Magala, M., & Sammons, A. (2008). A New Approach to Port Choice Modelling. Maritime Economics & Logistics (10), 9-34.

Merriam-Webster. (2019). *Tugboat.* Retrieved from https://www.merriam-webster.com/dictionary/tugboat

MI News Network. (2018). *Who is a Shipping Agent?.* Retrieved from https://www.marineinsight.com/careers-2/who-is-a-shipping-agent/

Nash, A., & Huerlimann, D. (2004). Railroad simulation using OpenTrack. WIT Transactions on The Built Environment, 74.

Ng, K. (2006) Assessing the attractiveness of ports in the North European container transhipment market: An agenda for future research in port competition. Maritime Economics and Logistics 8 (3): 234–241.

Parthenon. (2018). *Fact-based study on the container barge supply chain in the Port of Rotterdam*. Retrieved from *http://www.fenex.nl/Documents/Study%20container%20barge%20supply%20chain%20Port%20of%20Rotterdam.pdf*

Pires da Cruz, M. R., Ferreira, J. J., & Garrido Azevedo, S. (2013). Key factors of seaport competitiveness based on the stakeholder perspective: An Analytic Hierarchy Process (AHP) model. Maritime Economics & Logistics (15), 416-443.

Port of Antwerp. (2017). *Facts & Figures*. Retrieved from https://www.portofantwerp.com/

Port of Rotterdam. (2018). *Facts & Figures. A WEALTH OF INFORMATION.* Retrieved from https://www.portofrotterdam.com/sites/default/files/facts-and-figures-port-of-rotterdam_0.pdf

Port of Rotterdam. (2019a). *Port of Rotterdam Authority.* Retrieved from https://www.portofrotterdam.com/en/port-of-rotterdam-authority

Port of Rotterdam. (2019b). *Nautical services.* Retrieved from https://www.portofrotterdam.com/en/doing-business/services/service-range/nautical-services

Port of Rotterdam. (2019c). *Veilig. Efficiënt. Duurzaam.* Retrieved from https://www.portofrotterdam.com/sites/default/files/downloads/vts-nl.pdf?token=vLtIT8Eb

Port of Rotterdam. (2019d). *Harbour Master.* Retrieved from https://www.portofrotterdam.com/en/port-authority/about-the-port-authority/organisation/our-organisation/harbour-master

Port of Rotterdam. (2019e). *About the Port Authority.* Retrieved from https://www.portofrotterdam.com/en/port-authority/about-the-port-authority

Port of Rotterdam. (2019f). *Seaport dues.* Retrieved from https://www.portofrotterdam.com/en/shipping/sea-shipping/port-dues/seaport-dues

Port of Rotterdam. (2019g). *Port Call Optimisation.* Retrieved from https://www.portofrotterdam.com/nl/scheepvaart/zeevaart/overig/port-call-optimisation

Port of Rotterdam. (2019h). *Containers stuwen groei overslag haven Rotterdam.* Retrieved from https://www.portofrotterdam.com/nl/nieuws-en-persberichten/containers-stuwen-groei-overslag-haven-rotterdam

Port of Hamburg. (2017). *Facts and figures.* Retrieved from https://www.hafen-hamburg.de/en/statistics/tradearea

Portbase. (2009). *Over Ons.* Retrieved from https://www.portbase.com/over-ons/over-ons/

R. (2019). *Introduction to R.* Retrieved from https://www.r-project.org/about.html

Rahimikelarijani, B., Abedi, A., Hamidi, M., & Cho, J. (2018). Simulation modeling of Houston Ship

Channel vessel traffic for optimal closure scheduling. Simulation Modelling Practice and Theory, 80, 89-103.

Roundy, L. (2019). *Empirical Data: Definition & Example.* Retrieved from https://study.com/academy/lesson/empirical-data-definition-example.html

Royal Literary Fund. (2019). *What is a literature review?.* Retrieved from https://www.rlf.org.uk/resources/what-is-a-literature-review/

Saeed, N. (2009) An analysis of carrier selection criteria when choosing container terminals in Pakistan. Maritime Economics and Logistics 11 (3): 270–288.

Seine Maritime. (2018). *Glossary of Port and Shipping terms.* Retrieved from http://www.seinemaritime.net/suports/uploads/files/Glossary%20of%20Port%20and%20Shipping%20Terms.pdf

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3/4), 591-611.

Student. (1908). The probable error of a mean. Biometrika, 1-25.

Sturrock, D. T., & Pegden, C. D. (2011). Recent innovations in Simio. In Proceedings of the 2011 Winter Simulation Conference (WSC) (pp. 52-62). IEEE.

Thiers, G. F., & Janssens, G. K. (1998). A Port Simulation Model as a Permanent Decision Instrument.

SIMULATION, 71(2), 117–125. https://doi.org/10.1177/003754979807100206Tiwari, P., Itoh, H., & Doi, M. (2003). Shippers' port and carrier selection behaviour in China: a discrete choice analysis. Maritime Economics & Logistics, 5(1), 23-39.

Tongzon, J. (2002). Port choice determinants in a competitive environment. The International Association of Maritime Economists Annual Conference 2002 Conference Proceedings Panama.

Tongzon, J. (2009) Port choice and freight forwarders. Transportation Research Part E 45 (1): 186–195.

Ugboma, C., Ugboma, O. and Ogwude, I. (2006) An analytic hierarchy process (AHP) approach to port selection decisions – Empirical evidence from Nigerian ports. Maritime Economics and Logistics 8 (3): pp. 251–266.

Van Der Aalst, W. et al. (2011). Process mining manifesto. In International Conference on usiness Process Management (pp. 169-194). Springer, Berlin, Heidelberg.

Vieira, A., Dias, L. S., Pereira, G. B., & Oliveira, J. A. (2014, June). Micro simulation to evaluate the impact of introducing pre-signals in traffic intersections. In *International Conference on Computational Science and Its Applications* (pp. 722-745). Springer, Cham.

Vieira, A., Dias, L., Pereira, G., & Oliveira, J. A. (2014). Comparison of SIMIO and ARENA simulation tools. In 12th Annual Industrial Simulation Conference (ISC2014) (pp. 5-13). EUROSIS-ETI.

Simio. (2019). About Simio. http://www.simio.com

Waldrop, M. M. (2016). The chips are down for Moore's law. Nature News, 530(7589), 144.

Watson, R, T. Webster, J. (2002). Analyzing the past to prepare for the future: Writing a literature review. Mis Quarterly. VOL.26, NO.2, pp. 8-18.

Welch, B. L. (1947). The generalization of student's' problem when several different population variances are involved. Biometrika, 34(1/2), 28-35.

Wellens, A., & Mota, M. M. (2017, December). Modeling approach for managing the demand in congested airport networks: The case of Mexico city airport. In 2017 Winter Simulation Conference (WSC) (pp. 4535-4537). IEEE.

Wilcoxon, F. (1945). *Individual Comparisons by Ranking Methods* .Biometrics Bulletin, Vol. 1, No. 6. pp. 80-83.

Wilson, N., Fourie, C. J., & Delmistro, R. (2016). Mathematical and simulation techniques for modelling urban train networks. South African Journal of Industrial Engineering, 27(2), 109-119.

Wolfe, S. R., Jarvis, P. A., Enomoto, F. Y., Sierhuis, M., & van Putten, B. J. (2009). A multi-agent simulation of collaborative air traffic flow management. In Multi-Agent Systems for traffic and transportation engineering (pp. 357-381). IGI Global.

Yilmaz, Taylor, Fujimoto, Darema. "Panel: the future of research in modeling & simulation". Proceedings of the 2014 Winter Simulation Conference, IEEE, 2014, pp. 2797-2811.

Appendices

Appendix I – Literature study

The main goal of this literature study is to identify knowledge gaps in current literature relevant to the topic of this research: improvement of port call efficiency. More concretely, this will be done by pointing out the strengths of the studies and identifying the shortcomings and weaknesses. This way all the aspects required to perform a relevant research are gathered, which leads to the formulation of the main research question and the methods.

The literature study focuses on gathering the existing, available knowledge that shows the importance of port efficiency. Furthermore, relevant studies on port call efficiency, that have the aim to identify effective and supported solutions for congestion problems, will be discussed. A special focus will lay on the simulation aspect of these studies. Traffic simulation studies in other domains than the marine sector will also be examined. Lastly, process mining techniques will be discussed, in order to conclude if these techniques could contribute in a research that strives to optimize port call efficiency.

I-I Port efficiency relevance

Various factors are of significant influence for the port choice of shipping companies. A research by Ugboma (2016) used an Analytic Hierarchy Process (AHP) approach to investigate the port selection decisions of shippers for four major Nigerian ports. The conclusion provides an empirical support that port efficiency is the most important factor for port choice. This is later confirmed as well in a study using the same method by Pires da Cruz et al. (2013). The main goal of this study was to identify these factors for port choice using an AHP model. It is stated that the vessel turnaround time is considered as the most influential factor for competitiveness, from the shippers' perspective. The turnaround time is often used as a measure of port efficiency (Seine Maritime, 2018). The vessel turnaround time is defined as the time it takes between the arrival of a vessel and its departure from a port. Furthermore, a paper by Ng (2006) describes the results of a Likert-style questionnaire, held under 30 shipping lines. It showed that after monetary costs, time efficiency is the most important factor for port choice. A different study by Tongzon (2002) investigated the port choice process of Malaysian and Thailand shippers. It was concluded, among a list of factors like adequate infrastructure and frequency, that port efficiency was one of the most important factors. Lastly, a paper (Tiwari et al., 2003) uses a discrete choice model, where Chinese shippers have to make a choice out of 14 ports, based on shipper- and port characteristics. This paper shows that the most important factors for a shipper's port choice are the distance of the shipper to the port, the distance to the destination, distance from origin, port congestion and shipping line 's fleet size. This shows that port efficiency plays an important role in this research as well, as the factor port congestions directly relates to the port efficiency.

These discussed studies confirm that port efficiency has a significant influence on the attractiveness of a port and thus is a determining factor for shipping lines when choosing which port to operate in. Currently, it is difficult to exactly define port efficiency, due to the non-universal definition of what indicates an efficient port (De Monie, 2009). The following definition by Kennedy et al. (2011) will be used in this research: The efficiency of a sea-port is determined by the duration of the stay of a ship in a port, by the quality of cargo handling and by the quality of services. Optimizing the first component of the definition, the duration of a ship's stay in a port, is the focus of this study and will be referred to as port call efficiency. The Key Performance Indicators (KPI's), which will be defined and described in a later part of this report, will make it possible to measure the port call efficiency in a quantitative way.

I-II Port optimization through analytical models

The previous paragraph introduced the concept of port call efficiency and showed its scientific relevance. The next paragraphs will discuss previous and current port call efficiency optimization studies, not only related to the Port of Rotterdam, but also to other major ports. This way, a broad picture will be created of the current available and unavailable knowledge in this field.

A study by Parthenon (2018) investigated the container chain of the Port of Rotterdam. This study strived to identify the causes and possible solutions for the increasing congestion in the port. It showed that finding effective solutions is very difficult, because there are often multiple potential causes and distinguishing facts from opinions is not always possible. The container logistic chain is increasingly getting more complex, with a large number of actors involved. These actors need to work together but often do not have a contractual relationship, this leads to a lack of binding agreements on an operational level. Also, the system often has enough capacity to handle regular demand, however congestion arises during periods of increased demand. Parthenon (2018) suggests to look into potential solutions that reduce relative peak demand, that reduce the complexity of the chain and that align incentives to improve modal split. However, concrete solutions seem to be lacking. Also, the scope of this study is reduced to only container vessels. Dry bulk, liquid bulk and breakbulk cargo is not discussed. Lastly, it notable that the finding in this research are based on a data analysis, however even the possibilities of implementing simulation techniques, which could follow up the data analysis phase, are not discussed. The case is the same for a different study by Bichou and Gray (2004), which also strives to optimize a port's performance through conceptualizing ports from a logistics and supply chain management approach. It is stated that this will contribute to a higher port efficiency. The disuse of simulation models seems like an untapped potential.

I-III Port optimization through simulation models

The availability of data is increasing, this is accompanied by a wider use of analytics. These developments facilitate the expansion of our modelling and simulation capabilities (Yilmaz et al, 2014). Data driven simulation is used in a large variety of fields, where the problems are too complex to use analytical models (Almaz and Altiok, 2012), such as in the area of weather. For example, weather forecasts based on model predictions, have become highly reliable (Yilmaz et al, 2014). During the literature study, there were no data driven simulation studies found, which are specifically developed for improving the port efficiency in the Port of Rotterdam. However, simulation of certain port processes is definitely used in Rotterdam and other ports. The most relevant and interesting simulation studies are discussed in the next paragraphs.

There are three categories of simulation studies in the maritime transportation domain (Almaz and Altiok, 2012). The first category is applications on port/terminal logistics. The second one is modelling of vessel traffic on waterways for scenario and policy analysis. The third one is using simulation platforms as a tool to evaluate accident probabilities. The simulation models from the second category are often used as a starting point for studies in the third category.

There are many researches that use simulation models to analyse and optimize terminal logistics in ports. For example, Lagana et al. (2006), who looked into the optimal assignment of berth slots and cranes to shipping services, following a simulation optimization approach using a computational grid. A similar study (Legato et al., 2009) used a simulation model to test which policy of assigning yard cranes to yard blocks performs the best. However, these kinds of studies are outside of the scope of this research.

Marine traffic simulations, which lay in the scope of this study, are much rarer. A research by Thiers and Janssens (1998) looked into the consequences of building a container quay on the river outside the Port of Antwerp. It was expected that this quay would cause congestion problems on the river, as access to the Port of Antwerp is provided through locks connecting the river with the docks. A simulation model was built in order to estimate the increase in turnaround manoeuvre hindrance. However, this was not its only purpose. The other objective of the research was to develop a more general modular model that simulates the marine traffic. This way the model could serve as a reusable and expandable tool. This tool could then be implemented in the future to simulate other possible port infrastructure enhancements or variations in traffic streams. Since this study has been conducted, huge steps have been made regarding our modelling and simulation capabilities (Yilmaz et al, 2014). It is possible to make much bigger and more complex models using more input data. This is where the model made by Thiers and Janssens seems a little outdated. Furthermore, it uses a discrete time basis and a 2D interface (Thiers and Janssens, 1998). The visualisation aspect of the model does not come close to the current possibilities regarding visualization in models. Even though this model is deprecated to some extent, the concept of building a modular simulation model of marine traffic that is not built for one specific problem, but is adjustable to test different policies under different scenarios, can be of great value for this research.

A different marine traffic simulation study focussed on simulation and analysis of the Houston Ship Channel vessel traffic and operation (Rahimikelarijani et al., 2018). It was already decided that the Sam Houston bridge would be constructed over the river. The construction would cause hindrance for the marine traffic. The main focus of the study was to simulate which closing schedule would lead to the least waiting time for vessels. One recommendation made in the conclusions for further research is to include resources like tugboats, which can cause significant delays for other vessels (Rahimikelarijani et al., 2018). Another interesting aspect of this study is how the output of the simulations of different scenarios is statistically compared. The Fisher pairwise comparison method, which outperforms the ANOVA method (Rahimikelarijani et al., 2018), was used to test for significant differences. The exact method and tool used for this simulation was discrete event modelling in Arena. This is a simulation software which is suitable for modelling marine traffic, whilst meeting the current day expectations of a discrete event simulation tool. However, there is a different software, Simio, which also is a discrete event simulation tool and arguably outperforms Arena. A study by Vieira et al. (2014) modelled different case studies in both Simio and Arena. This research shows that it is possible to model the same model in both tools. Nevertheless, it is less complex and time-consuming to achieve the same results in Simio. Furthermore, Simio is less abstract as it is easier to mimic the actual shape of, for example, a port and its traffic. In combination with the fact that Simio has direct access to Google Warehouse, a library of graphic symbols for 3D animation (Sturrock and Pegden, 2011), it is concluded that the visualisation aspect of a model has more potential when built with Simio. This is of added value when presenting a study to a non-technical audience.

The last study on marine traffic simulation which will be discussed looks at the impact of deepening the Delaware River on the navigational efficiency in the River (Almaz and Altiok, 2012). Similar to the previous discussed research by Rahimikelarijani et al. (2018), the model was built using discrete event simulation in Arena. Different scenarios were set up and compared. Besides giving insights in the impact of deepening the river, a final product of the project was the simulation model itself. It can be used in the future decision-making processes by testing different policies and scenarios (Almaz and Altiok, 2012). This strong property of the research was also observed in a research by Thiers and Janssens (1998). A shortcoming of the study is the many made assumptions and simplifications regarding the model structure. It is stated by Almaz and Altiok that the process of extracting the model structure from historical data is open to improvements. A method that can improve this process is process mining. This is discussed in appendix I-V.

I-IV Traffic simulation models in other domains

Simulation studies in other domains than marine traffic like air-, rail- and road traffic can also be of added value for this research. This paragraph will focus on these projects.

A study by Wellens and and Mota (2017) introduced a discrete event simulation approach to assess flight demand for a congested airport. The simulation tool Simio is used for the model. The model is set up for Mexico City Airport. However, it is stated that the flexibility of the model makes it easy to adapt the model to airports in other regions. The model makes it possible to assess the effectiveness of current policies and test alternative policies. A different study by Kim et al. (2005) carried out a discrete event simulation study, using Arena, which simulated the arrivals of airplanes at an airport. 2000 replications were generated, which contributed to the robustness of this study. The study recommends for future work to include the aircraft types, fuel costs, flight origin and weather influence. Wolfe et al. (2009) states that the air traffic management system is not expected to scale to the projected increase in traffic. A simulation model was made to analyse different scenarios. It showed that, even though the model was still in an early phase, the results have relevant and interesting properties. This illustrates that simulation studies can already show promising results in early development stages. Davidrajuh and Lin (2011) performed a simulation study where a novel Petri net simulator was developed, in order to explore air traffic management capability of Evenes airport in Norway. Three characteristics of the simulation model where highlighted. The model should be flexible, to enable easy integration with other libraries and tools. The model should be extensible, to make it possible for users to add their own modelling extensions. Finally, the model should be easy to use, to make it accessible for non-technical users. These are strong properties that can be used in simulation studies in the marine traffic sector.

Wilson et al. (2016) compared different ways of modelling train networks. Three categories of models were identified: analytical-, heuristic- and simulation models. Wilson says that analytical models are very limited in terms of scope and complexity. Furthermore, the research states that heuristic models are effective for optimising large complex networks, however they lack the visualization aspect and scenario testing aspect. This is where simulation models outperform the other two categories. This is in line with the findings in the previously discussed literature regarding marine traffic studies. A research by Nash and Huerlimann (2004) developed a railroad simulation program. It uses a mixed discrete and continuous simulation process that calculates the train motions. The model functions as a rail road laboratory, which enables users to test different policies under different scenarios. A strong property of this model is that it automatically generates a lot of output data including graphs, diagrams and tables.

Herty and Klar (2003) introduced a traffic flow network simulation model, which is able to simulate multilane junctions in a detailed way. After comparing the model with already existing models, the model was used to simulate and compare multiple scenarios. In order to establish the right input parameters for each scenario, a optimization procedure with control parameters was initialized. This is a strong trait of this research. The input variables were all constant. It was recommended for future studies to use time-dependent parameters. Viera et al (2014) developed a traffic micro simulation model using Simio, with the goal to improve the efficiency of an intersection. Scenarios were implemented that include the use of pre-signals. These were compered by looking at the waiting times and queue sizes. This study is a good example of a very well demarcated research using discrete simulation in Simio. Lv and Niu (2011) did a similar study. A simulation model was built in Simio to investigate the effects of bus lanes on a three-phase signalized intersection. Different scenarios in terms of traffic demand and different policies in terms of bus lanes and signal timing were implemented to predict and compare the average vehicle delays and queue length. Again, a well demarcated simulation. The model that will be built for this research will have more components than just one intersection. It is essential to be aware of this higher degree of complexity, in order to set the right scope and implement it in the actual simulation.

I-V Process mining

Process mining techniques make it possible to gain knowledge in a structured way from event logs (Van der Aalst et al, 2011). In other words, process mining is a tool that makes it relatively easy to extract the structure of the system from historical data about that system. Process mining is a relative new technique of analysing big chunks of log data. The interest in- and use of process mining techniques is increasing as the quality of data is rising and the computing power and process mining algorithms are improving. The most relevant studies on this subject are discussed in order to get a clear picture of how far these techniques are developed and how they could contribute to this research which aims to improve port call efficiency.

Maruster, Van der Aalst and Wijters (2002) identified an issue when it comes to enacting a given workflow process: In order to so, it is necessary to have explicit process models that help to understand the system behaviour. The research states that such a qualitative workflow design is often incomplete, subjective and time consuming. A learning method is introduced that uses the workflow log, which holds data about the processes as it is actually being executed, to identify the system behaviour in a quantitative way. The introduced method uses a logistic regression model to discover direct connections between recorded events of a process. This is the oldest study found on this subject during this literature study (2002) and thus marks the birth of process mining. However, the use of this technique was still very limited. This process mining technique fails to identify the process as a whole. It only shows the relationship between two events in the process. Furthermore, log data often contains a timestamp, this is not incorporated into this model, which seems like a missed opportunity. Lastly, only a relative small number of events is tested in the model (a couple of thousands). It is unknown how this model would behave when the number of events would increase drastically, especially back in 2002 when the computing power was significantly lower than now (Waldrop, 2016). Even though this process mining technique has many short comes and does not seem to fit well into this research, the concept of process mining shows potential for the use in ports, because of the AIS event data that is being recorded. Chapter 3.1 describes what AIS event data exactly is.

In 2011 a process mining manifesto was written by an IEEE taskforce (Van der Aalst et al, 2011). The goal of the manifesto was to define a set of principles and to list important challenges for process mining. The task force states that the start of every process mining technique is an event log, that includes recorded events, that each refer to a certain activity, with a timestamp and optionally additional information like resources and quantities. This manifesto identified three main purposes for process mining. The first is process discovery. The event log of a process is taken to construct a new model without any prior knowledge. The second is process conformance. An already existing model is compared to the event log of the same process and vice versa. The third type is process enhancement. The goal is to improve already existing models using information about the actual process from event logs. Figure I-1 shows these three types of process mining in terms of input and output.

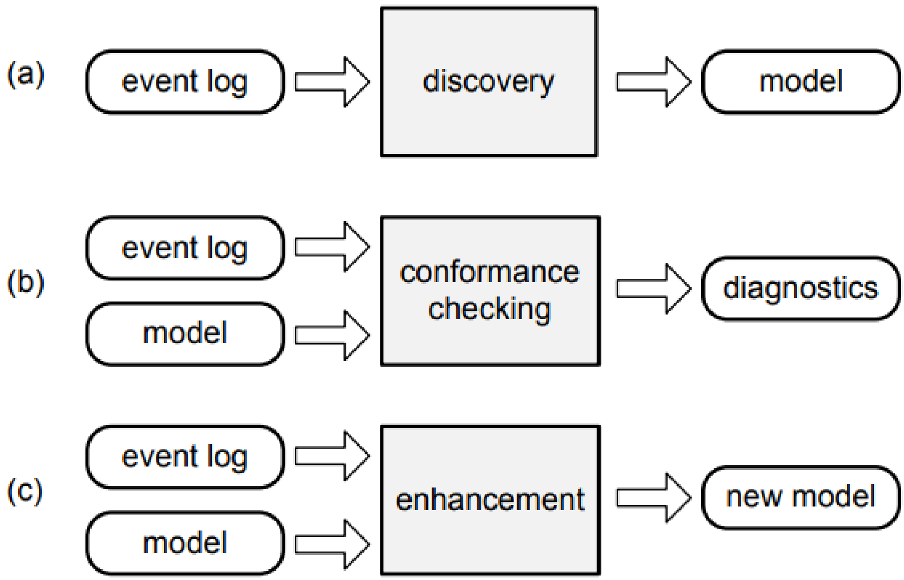

Figure I-1: The three basic types of process mining explained in terms of input and output: (a) discovery, (b) conformance checking, and (c) enhancement (Van der Aalst et al, 2011).

Since the first process mining models (Maruster, Van der Aalst and Wijters, 2002) a lot has improved. Significantly increased computing power (Waldrop, 2016) and improved process mining algorithms, that are based on the principles defined in the process mining manifesto, have made it possible to successfully implement these techniques on huge chunks of log data. However, the IEEE taskforce concluded that there are still challenges that have to be overcome in order to execute these analysis in a valid way. The biggest one is finding, merging and cleaning event data. More concretely, the data may be distributed over different sources, event data may be incomplete, an event log may contain outliers and logs may contain events at different level of granularity.

No process mining studies have been found that use the port call event data that has been made available for this research (chapter 3.1). It is still unknown in what ways a port could benefit from process mining. Therefore, exploring the exact possibilities of process mining in this domain has the potential to reveal a lot of benefits.

I-VI Main research question

The following can be concluded from the above: It is essential to always strive for a higher attractiveness of a port in order to be a key player in the world market for sea cargo transportation. A high port call efficiency has significant impact on the port efficiency and thus on the attractiveness of a port, and therefore must constantly be improved. The problem seems too complex to use analytical models. A data driven discrete event simulation study seems the most suitable for this issue. Creating a modular simulation model with Simio will not only contribute to solving the current discussed issue, but has also the potential to be used in future research as a tool for policy and scenario testing. A good data analysis is essential, in order to build a well-structured and clearly demarcated model. Process mining techniques show potential for improving the process of extracting the model structure from the log data for the simulation model. However, the exact possible implementations, benefits and limitations of this technique in this research domain are still unknown. Assessing these techniques and discovering the potential implementations of process mining in this research domain are also part of this study.

A study with the components described above, using the Port of Rotterdam as a case study, will be conducted in order to improve the port efficiency, to produce a more general, flexible and extendable tool for policy and scenario testing and to explore the possibilities of process mining in this domain. This leads to sub-questions in section 1.2 and the following main research question below:

How can data analysis, process mining and discrete event simulation contribute to identifying and assessing policies that improve port call efficiency?

Appendix II – Port Call Optimization Map (ITPCO)

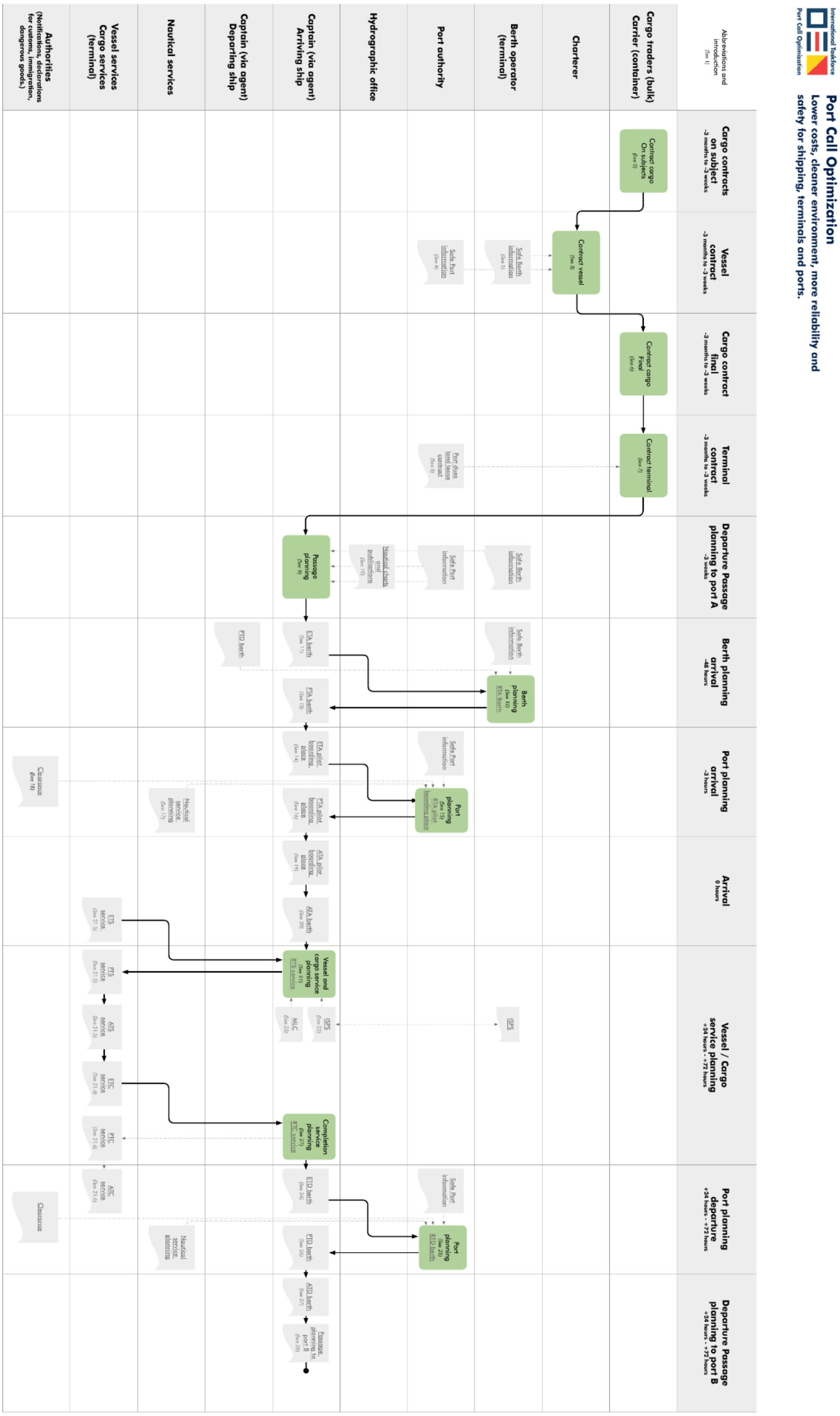

Figure II-1: Port Call Map (Source: International Taskforce Port Call Optimization)

Appendix III – Overview of standardized events

(Censored)

Figure III-1: Overview of standardized events

Appendix IV – Process mining model

IV-I Data preparation

Before the dataset is used as an input for the process mining phase, it is cleaned and transformed into a structure that is fit for Disco, the process mining software. The elements that the dataset is required to have are the following:

- ID: a column that shows the id of each vessel
- Activity: a column that specifies what event is occurring
- Time stamp: two columns that specify the start and the end of an event
- Attribute: This element is optional. It is possible to add an infinite number of columns that each specify an attribute of the ID (vessel)

Next, dataset is reduced to only the events that are within the scope of the model as described during the conceptualization (Appendix V- I). The following events remain (Source: Appendix III – Overview of standardized events):

- ATA 120 nmi Zone: Actual arrival at the 120 nautical miles zone
- ATA 60 nmi Zone: Actual arrival at the 60 nautical miles zone
- ATA 12 nmi Zone: Actual arrival at the 12 nautical miles zone
- ATA Anchorage: Actual arrival at (one of) the anchorage area(-s)
- ATD Anchorage: Actual departure from (one of) the anchorage area(-s)
- Tugs Stand By: Actual Time the tug(-s) are available to assist the vessel
- Tugs No More Stand By: Actual time the tug(-s) are no more available to assist the vessel
- Pilot On Board: Actual time the pilot physically embarked the vessel to be piloted
- Pilot Disembarked: Actual time the pilot physically disembarked the vessel that has been piloted
- ATA Berth: Actual arrival at the berth
- ATD Berth: Actual departure at the berth

A snippet of the transformed and cleaned dataset with only the relevant events is shown in table IV-1. The actual dataset that is inputted in the process mining software Disco, contains 303.714 rows and thus 303.714 recorded events.

UCRN	Event	Start	End	VesselType
	60NMfromNLRTM	26-1-2018 01:55	26-1-2018 01:55	BreakBulk
	12NMfromNLRTM	9-1-2018 07:24	9-1-2018 07:24	BreakBulk
	TugConnect	12-1-2018 11:54	12-1-2018 11:54	ContainerLarge
	TugConnect	13-1-2018 17:58	13-1-2018 17:58	ContainerLarge
	TugDisconnect	12-1-2018 13:10	12-1-2018 13:10	ContainerLarge
	TugDisconnect	13-1-2018 19:31	13-1-2018 19:31	ContainerLarge
	PilotDisembark	12-1-2018 13:33	12-1-2018 13:33	ContainerLarge
	PilotDisembark	13-1-2018 20:12	13-1-2018 20:12	ContainerLarge
	PilotOnBoard	12-1-2018 10:58	12-1-2018 10:58	ContainerLarge
	PilotOnBoard	13-1-2018 18:15	13-1-2018 18:15	ContainerLarge

Table IV-1: Snippet of process mining input schedule

The entire process mining data set is directly inputted in Disco, which in a matter of minutes produced the process mining model.

IV-II Process mining model structures

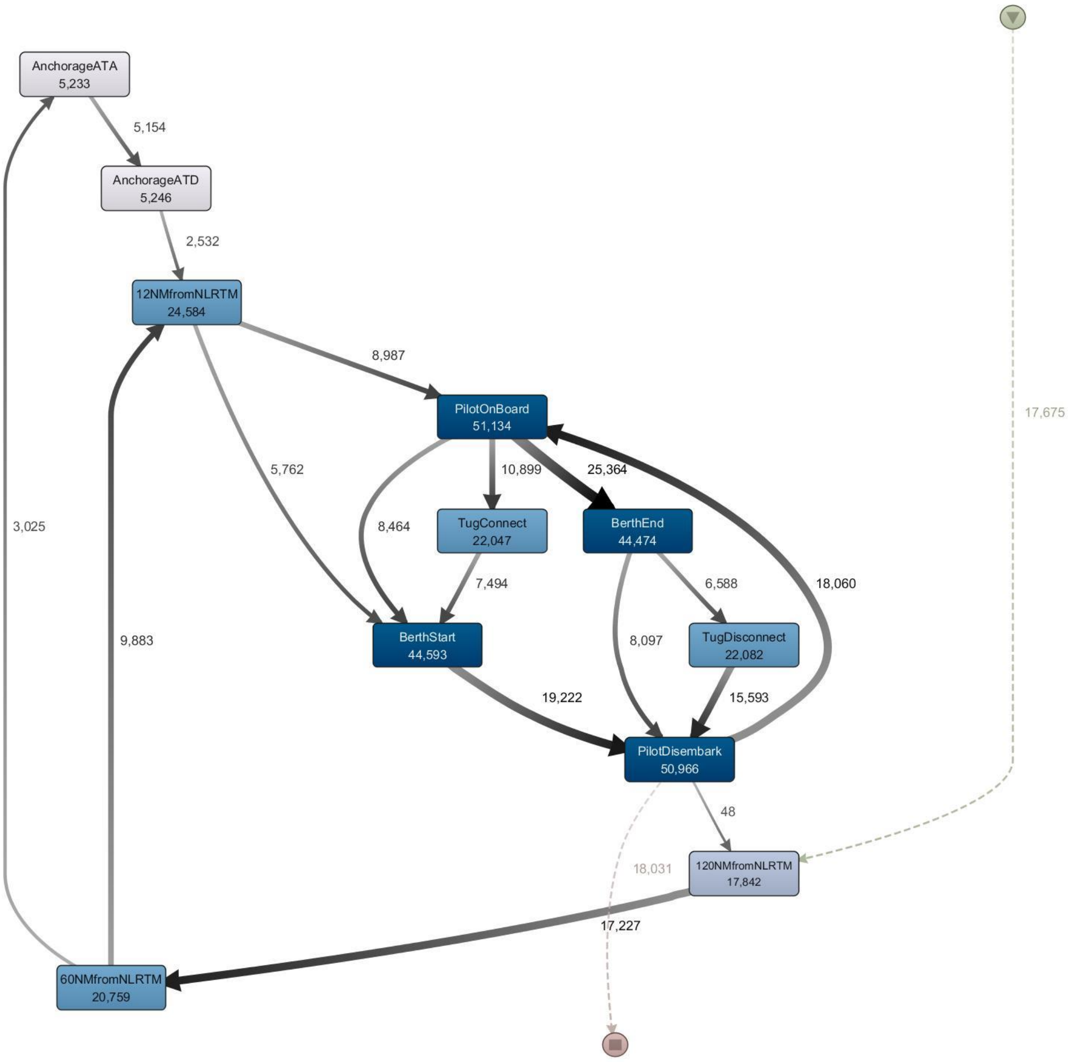

Figure IV-1: Process mining structure (level of aggregation: high)

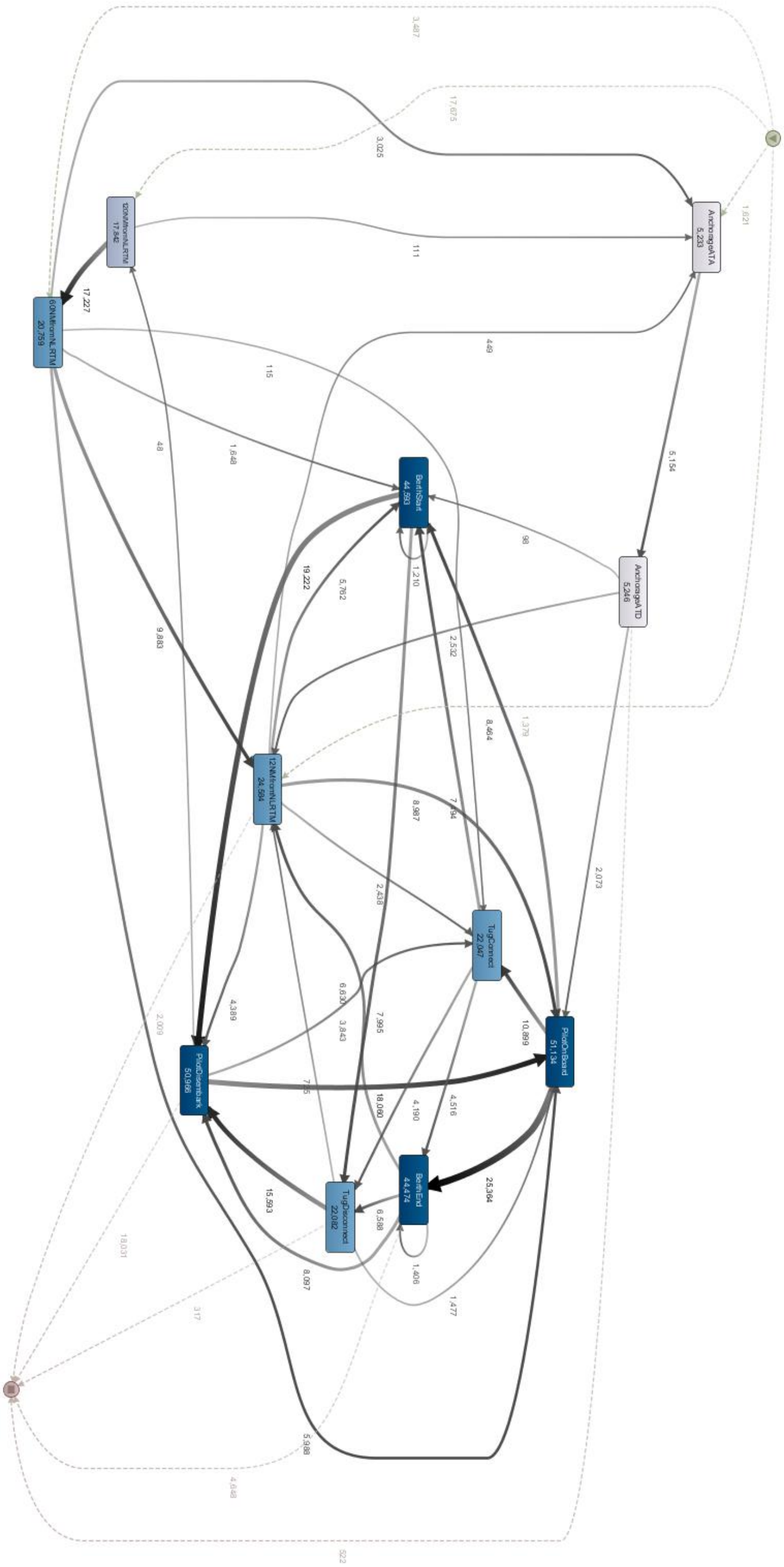

Figure IV-2: Process mining structure (level of aggregation: medium)

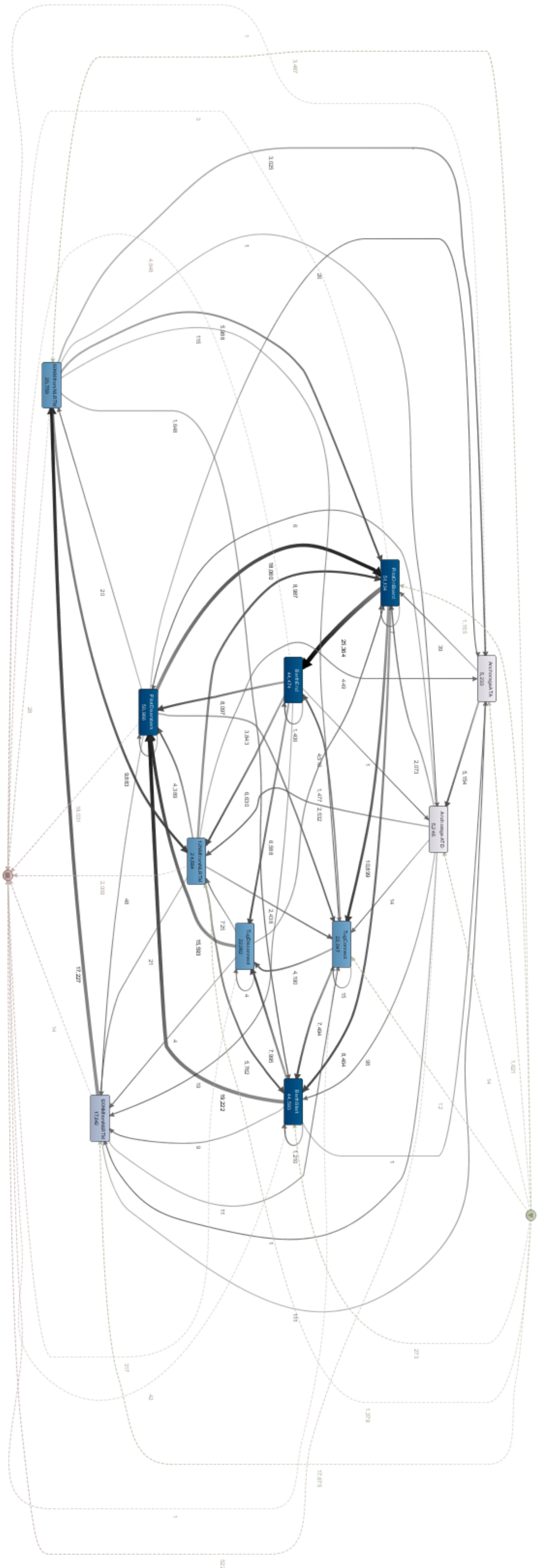

Figure IV-3: Process mining structure (level of aggregation: low)

Appendix V – Simulation model

This appendix describes the process of creating the what if analysis tool in the form of a simulation model. This starts with the conceptualisation, where the conceptual building blocks and their attributes are identified. This is followed by the specification, where the input for the model is specified and the model assumptions and reductions are discussed. Next, a description is provided of how the actual Simio model is created. Lastly the model if verified and validated. From this it is concluded whether the model is reliable enough to use for simulation and to translate the results back to the real world.

V-I Conceptualisation

The goal of this section is to identify the conceptual building blocks, based on the provided information in chapter 2, which will be used to build the model. Also the parameters of each building block are classified. Furthermore, the input parameters, which make it possible to simulate different policies and scenarios, and the output variables, which show the performance of the simulated policies and scenarios, are determined. How these elements relate to each other is visualized in figure V-1.

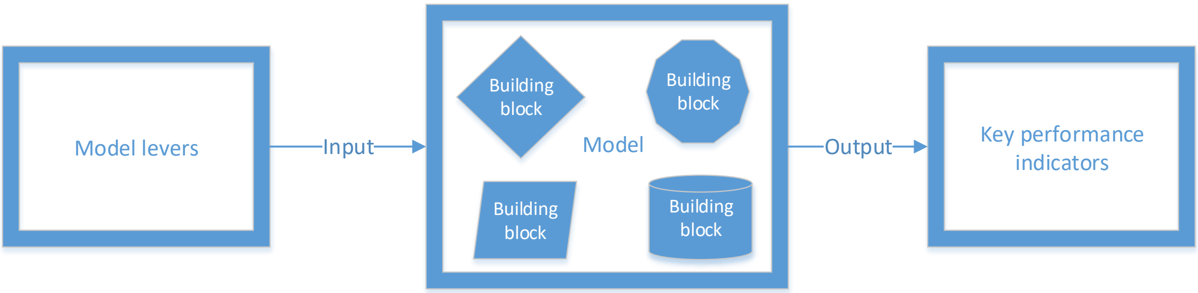

Figure V-1: Metamodel

V-I-I Building blocks

Cargo/passenger vessels

For this research, it is chosen to simulate six different types of cargo/passenger vessels. These six different types are chosen during the data analysis because it was seen from the data that each type shows its own unique behaviour in the port. This behaviour is expressed in terms of cargo, (number of) destinations, berthing durations, number of tugboats, pilot usage, speeds and sizes. Each vessel holds a UCRN and MMSI number. The destination of the vessel is at least one terminal that matches its cargo type. Each ship has a certain length, which needs to be smaller than the available quay length of a terminal. Every vessel has a number of required tugboats to and from a berth (between zero and four). Each container vessel keeps track of the amount of delay that it experienced. Lastly, it is tracked for every vessel whether it has operational clearance to enter the port or not. Operational clearance is given when there is a berth spot available and, if required, a pilot and tugboats.

The six types are:

- Regular container vessels (length < 250 m)
- Large container vessels (length > 250 m)
- Dry bulk vessels
- Liquid bulk vessels
- Breakbulk vessels
- Passenger vessels

The properties that each vessel has:

- Unique port call registration number (UCRN)
- Maritime mobile service identity number (MMSI)
- Speed (knots)

- Length (m)
- Model entry point (location)
- Model entry time (time stamp)
- Pilot usage (yes/no)
- Pilot boarding place (location)
- Number of tugboats to berth (#tugboats)
- Number of tugboats from berth (#tugboats)
- Tug company (name)
- Destination (location)
- Total berth visits (#visits)
- Remaining berth visits (#visits)
- Delay (hours)
- Operational clearance (yes or no)

Infrastructure

Waterway

Waterways are the paths that the ships travel on. Each path is unidirectional and has a speed limit. The waterways connect every node in the model, from the place where the vessels enter the model to all the destinations in the port.

- Speed limit (knots)
- Length (m)

Tug area

The tug area is the area where the tugboats connect with the incoming seagoing vessels.

- Number of available tugs (number of available tugboats)

Terminal

The terminal is the place where the cargo of the seagoing vessels is loaded and unloaded. Each terminal has a specific cargo type, which in this model is either container or bulk. For each deep-sea container terminal, the length of the quay and the length of the unoccupied quay is specified as well. The length of the unoccupied quay is calculated by first subtracting the length of every vessel that is docked in at that specific quay from the total quay length. Then, for each docked vessel an additional 40m is subtracted in order to compensate for the minimal distance between two docked vessels.

- Type (one of six vessels)
- Total length quay (m)
- Length unoccupied quay (m)

Anchorage

The anchorage is the designated area where incoming seagoing vessels drop anchor until operational clearance is given by the VTS. These areas are visualized for the Port of Rotterdam in figure 2.15.

- Number of vessels in anchorage (#)

Pilot boarding area

The pilot boarding area is the designated area where the pilots board and disembark an incoming seagoing vessel.

- Number of available pilots (number of available pilots)

- Total number of pilots

Services

Tugboat

A tugboat is a vessel that manoeuvres bulk carries and container vessels from the tug area to their destination in the port. Furthermore, tugboats assist in the undocking process. Their speed is measured in knots and they are either available or unavailable. Every tugboat belongs to a company, which is specified as well. This makes it possible to experiment with scenario's in which tugboat companies are working together and serving each other's customers.

- Speed (knots)
- Available (yes or no)
- Company (name)

Pilot

A pilot boards incoming seagoing vessels, which require a pilot, in the designated pilot boarding area and manoeuvres the vessel to its destinations. The pilot leaves the ship during its time at berth. Before the vessel leaves, a pilot gets onboard again to assist the captain of the vessel in safely manoeuvring the ship outside of the port.

- Available (yes or no)

VTS

The vessel traffic service provides an operational clearance to the incoming seagoing vessels if all requirements are met. An operational clearance enables the vessels to enter the port, otherwise they have to wait at the anchorage until clearance is given. The requirements are that there is at least one available pilot, that there are more available tugboats than the number of required tugboats for that vessel and that the length of the incoming ship is smaller than the length of the unoccupied quay of the terminal where the ship is headed.

V-I-II Model levers

Number of tugboats

This lever sets the initial number of tugboats is the system.

Number of tugboat companies

This lever sets the number of tug boat companies in the system.

Number of pilots

This lever sets the initial number of pilots in the system.

Number of vessel arrivals

This lever sets the arrival rate of vessels in the system.

V-I-III Key performance indicators

This section discusses the key performance indicators (KPI's). It is important to note that all the key performance indicators are either container vessel specific or pilot or tug specific.

Throughput time operational port call

This KPI shows the average duration of the operational part of the port call for all container vessels. The operational part of the port call starts when a vessel requests operational clearance and ends when it leaves the pilot boarding place, after visiting the port. The average duration is measured in hours.

Total number of anchorage visits

This KPI shows the total number of container vessels that visit the anchorage.

Total average delay at anchorage

This is the total average waiting time for container vessels, as they wait for operational clearance. This KPI is measured in hours. Container vessels that do not visit the anchorage are not included in this KPI.

Total average delay at berth

This KPI shows the total average waiting time that container vessels experience at a terminal. The delay at berth only occurs when a vessel requests nautical services, but they are not available. This KPI is measured in hours.

Availability of the pilots

This KPI shows the average occupation of the pilots for their entire population (%).

Availability of the tugboats

This KPI shows the average occupation of the tugboats for their entire population (%).

V-II Specification

This section describes the inputs for the model and the model reductions and assumptions

V-II-I Simulation input data

The simulation input schedule which is directly linked to the model in Simio, forms a large part of fixed simulation inputs. The properties of this schedule are based on the properties of the vessels in the model, described in the conceptualization phase (appendix V-I-I). Table V-1 shows a snippet of the actual schedule, which has 24291 rows. Each row represents a unique port call, which means that 24291 unique port visits are simulated in the simulation model. The remaining 5.000 are sampled from this table. The schedule shows the MMSI for identification purposes. The column ‘Ship Length’ specifies the length for each vessel in the model. This is used to calculate if there is still room at the quay of the vessel’s berth location (= length quay – length ships at quay). ‘Ship Arrival’ and ‘Entry Point’ specify the time when and the source where a vessel enters the model. ‘Vessel Type’ shows to which category a vessel belongs. This attribute makes it possible to measure certain KPI’s per category. The ‘Berth Name’ tells the destination of each vessel. Five deep sea terminals at the Maasvlakte are included as separate destinations in the model. The other terminals are bundled per cargo type, in order to reduce complexity. ‘Tugs to Berth’, ‘Tugs from Berth’, and ‘Tug Company’ show for each incoming vessel how many tugboats they require and to which company these tugboats belong. Finally, the ‘Pilot Boarding Place’ gives the pilot boarding location for each vessel and in the case of no pilot usage it will say ‘NoPilot’.

(Censored)

Table V-1: Snippet simulation input schedule

The other data inputs for the model are the empirical distributions that describe the berthing durations of each vessel category. These are visualized in figure 3.7.

For the speeds of the vessels outside of the port, the distributions in table V-2 are used. The distributions are based on findings of Marine Insight (2019), who strives to provide information on various aspects of the marine world. It is chosen to use uniform distributions, because there is no reason not to believe that every outcome is equally likely to occore.

Vessel Type	Speed outside the port
Large container vessels	Random.Uniform(16, 24)
Small container vessels	Random.Uniform(16, 24)
Liquid bulk vessels	Random.Uniform(13, 17)
Dry bulk vessels	Random.Uniform(13, 15)
Breakbulk vessels	Random.Uniform(13, 15)
Passenger vessels	Random.Uniform(16, 25)

Table V-2: Speed distributions outside port (source: Marine Insight)

V-II-II Model reductions and assumptions

A model is a simplification of the reality. Because of the restricted time available for this project and a project team of only one person, it is necessary to make model reductions. Elements and actors that are not relevant enough or would make the model too complex are not included in the simulation. It is important to note that this chapter has been written in an iterative way. New reductions and assumptions were constantly added during different phases of the research, especially before and after the data analysis phase, process mining phase and modelling phase. These reductions and assumptions are explained in this chapter.

Cargo vessels

- Only the delays of the container vessels are measured in the model. It is concluded from expert consultation that all the container vessels travel according to a 'just in time'-schedule. This means that their aim is to arrive at the port on the agreed time. These 'just in time'-agreements are included in the contracts between shipping lines and consignor. However, not all dry-, liquid- and breakbulk vessels, travel according to a 'just in time'-schedule, as this is not included in their contracts. Moreover, often an minimum speed for the vessel is determined. This leads to situations where vessels arrive at the port multiple days too early. These vessels then have to drop anchor at the anchorage areas (figure 2.15) and wait. This is not an delay from the vessel's perspective, as these vessels arrived too early. However, when analysing the log-data of vessels, which is done in chapter 3, it is impossible to distinguish vessels that actually experience delays at the anchorage from vessels that arrived too early at the anchorage. Therefore, delays cannot be measured for dry-, liquid- and breakbulk vessels. This is not the case for container vessels, because they are the only vessel type that always has ‘just in time’-agreements. This is why the identified KPI’s only focus on container vessels.

- The cargo vessels in the model also function as a visual representation of the captain. The captains will not be modelled separately.
- Speeds of the vessels will not be based on empirical data, but based on literature review and expert consultation. However, it is expected that this will not impact the performance of the model significantly, as the KPI’s are not directly dependent on the speed of the vessels.

Infrastructure

- Only the pilot boarding place ‘Maascenter’ is simulated. In chapter 3 it is concluded that the share of the other two pilot boarding places is insignificant. In other words, almost all the vessels that use a pilot, let the pilot board at Maascenter. Simulating the other pilot boarding places adds very little value to the simulation model.
- Not all the terminals are modelled separately. The focus of this study lays on container vessel. For this reason, the five deep sea container terminals at the Maasvlakte are modelled separately. However, the other terminals are bundled into one terminal per vessel type. The bundled terminals have an infinite capacity, because it is impossible to bundle their capacities in a valid manner. The deep-sea container terminals have a capacity which is equal to the length of their quay. These values are extracted from Google Earth. For each deep-sea terminal, only the part of the quay that has cranes suitable for large deep sea container vessels is measured. The following terminals are simulated separately in the model:
    - APM Terminal (APMT) (1.500m)
    - APM Terminal 2 (APMT2) (1.000m)
    - Euromax Terminal Rotterdam (EUROMAX) (1.700m)
    - Rotterdam World Gateway (RWG) (1.100m)
    - Europe Container Terminals Rotterdam (ECT) (3.500m)
    - SmallContainerTerminal
    - LiquidBulkTerminal
    - BreakBulkterminal
    - DryBulkTerminal
    - PassengerTerminal

Services

- The critical services at the terminal (cargo services and ship services) are not modelled separately. The time a ship spends at a berth spot is directly taken from an ship category specific empirical distribution. This means that there are also no bunkering events in the model. This restricts the model to scenarios where the entire duration at a berth is a parameter. However, it is still possible to higher the level of detail in the future.
- The nautical services are represented in the model by tugboats and pilots. However, the linesmen are not visually present in the model, because the level of detail would be too high for this project. Also the Port of Rotterdam is one of the only ports that uses linesmen. Other ports use workers at the terminals to moor a ship. Not modelling the linesmen makes the model more applicable to other ports.
- The pilot transfer vehicles (vessels and helicopters), that are used to transfer the pilots to the incoming cargo ships at the pilot boarding areas, are not be modelled. The level of detail would be too high for this project. The pilot entity symbolises not only a pilot but also its vehicle, when it is not actively piloting an incoming vessel.

- Agents are not modelled. The processes of agents booking nautical services and communicating with the terminal and port authority will not be simulated. The level of detail would be too high for this project.
- The administrative clearance, given by the Port Coordination Centre is not modelled. This would be too complex and adds very little value. According to a port call expert, the amount of times that the administrative clearance is not given is negligible. Therefore, delays caused by this are not significant enough to model them.
- When a pilot or tugboat is requested in the model at a terminal and it is available, it is transferred in one simulation step to that terminal. This is not realistic and is a modelling restriction. In the available time, it was too complex to model a logic that makes the pilot or tugboat travel to the specific node where the vessel is waiting.
- Through expert consultation of a data scientist, the following information is gathered: There are 220 pilots in the port of Rotterdam. All those pilots have a different set of permits, which each enables them to only board a certain set of vessels. This is a complex process to model. Therefore it chosen to simulate pilots, that work 24/7 and that can board any ship. Furthermore, the expert stated that a delay is almost never caused because of pilots. That is why it is decided to set the number of pilots in the model on a relative high number, so that they will not cause any delays in the simulation.
- From expert-consultation it is concluded that the tugboats are available 24 hours a day and 7 days a week. It is assumed that each tugboat is operated by three crews. Each crew does an 8-hour shift. However, because of port regulations, they are not allowed to work 8 hours straight. In each shift, each crew needs to take breaks. However, the duration of those breaks is an unknown factor in this research. It is assumed that these breaks take 20% of the total time. When looking at the AIS event data, it is concluded that tug company A has on average 10 unique tugs and tug company B 9 unique tugs, active in the port, each day. In order to compensate for the breaks, 8 company A tugs and 7 company B tugs are simulated that are active 24/7.

V-III Model mechanisms

This part explains the most imported model mechanisms. In other words, it is explained how the model is created and how it functions. This is done by first addressing the different types of entities that are incorporated in the model. Next all the objects, that are visible in the front view of the model, are addressed separately and it is explained for each object how this is modelled in the background and what happens when each of the three types of entities (cargo vessels, pilots and tugboats) pass through this object. The front view of the model consists of a map of the world with the objects, that are within the scope of this research, placed on the map. The objects are linked with paths, that simulate the waterways. The entities move over theses paths from object to object. Most of the background modelling logic is implemented through add-on processes, which allows the modeller to add custom logic to the model. The objects that will be discussed are:

- Model entry points
- Operational clearance & anchorage
- Pilot boarding place
- Tug Area
- Terminals

To get a first broad idea of how the model looks in the front view, a snippet is shown of the model while running in figure V-2. It shows all the objects except for the entry points, which are further into the sea and the small container terminal which is further inland.

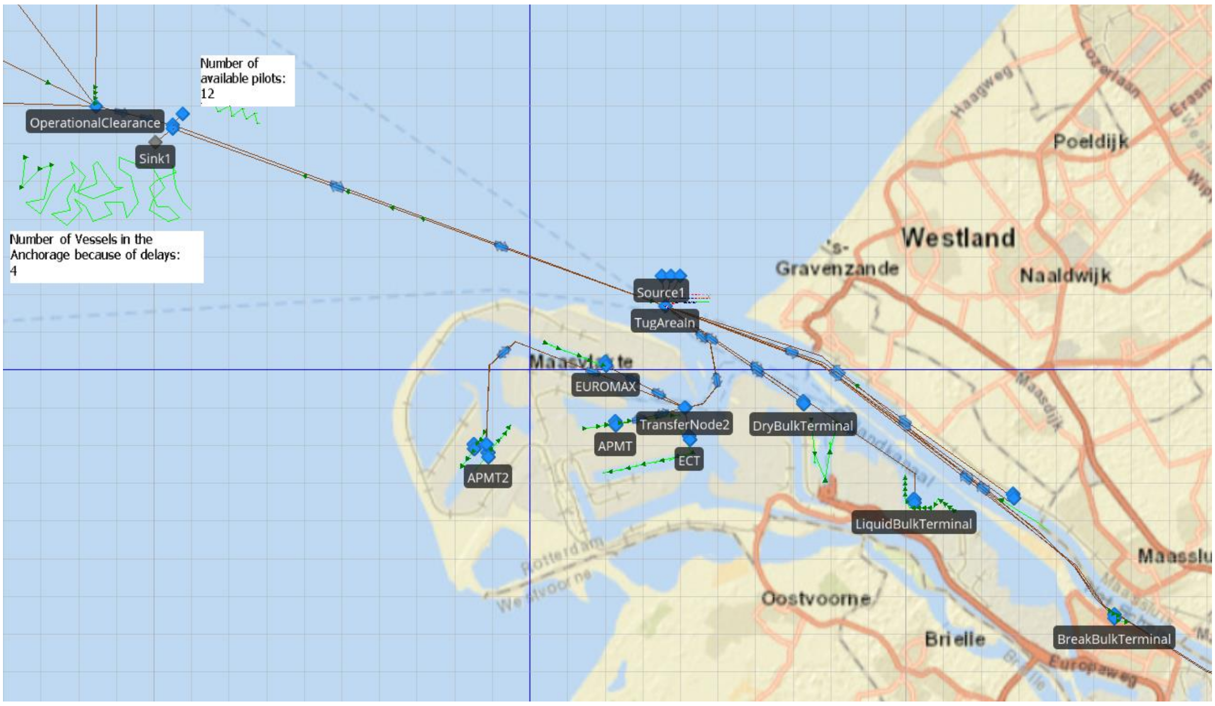

Figure V-2: Snippet of front view of Simio model (entry points and small container terminal are missing in snippet)

Model entities

Vessel

The vessel entity group simulates all the cargo and passenger vessels that visit a terminal in the port of Rotterdam. Figure V-3 shows a 2d representation of what this entity type looks like in the model. An entity of this group will enter the model in one of the four entry points, discussed later in this chapter. Its attributes, described in appendix V-I, are assigned through state assignments that lookup the corresponding values in the simulation input schedule. Each vessel gets the right destination assigned, through an add-on process. After the vessels visit the port they leave the model through a sink located just outside of the port.

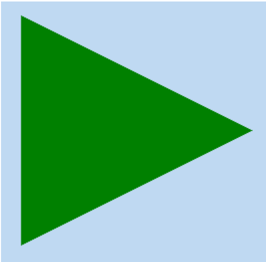

Figure V-3: Vessel

Pilots

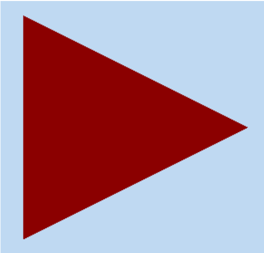

Figure V-4: Pilots

The pilots entity group simulates the pilots and their pilot vessels and pilot helicopters that they use to go to the pilot boarding place. Figure V-4 shows a 2d representation of what this entity type looks like in the model. A fixed number of pilots is created at the start of the run in the pilot source. Their routing logic is as follows: if they are attached to a vessel entity, they follow that vessel's routing logic. If they are not attached they return to the pilot boarding place or wait at the pilot boarding place in the available-pilots-queue. This entity type does not exit the model during the simulation.

Tugs

The tugs entity group simulates the tugboats that tow vessels in the Port of Rotterdam. There are two subcategories in this entity group because there are two different types of tug companies. Figure V-5 shows a 2d representation of the two subcategories. A fixed amount of each sub-type is created in Source1 and Source2. Their routing logic is as follows: if they are attached to a vessel entity, they follow that vessel's routing logic. If they are not attached they return to the tug area or they wait at the tug area in the available-tug-queue. This entity type does not exit the model during the simulation.

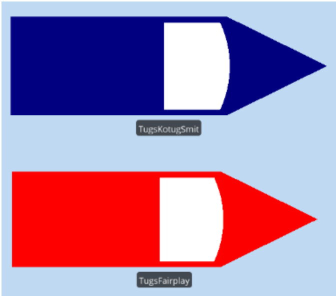

Figure V-5: tugs

Model objects

Model entry points

The model has four points where the vessel entity group enters the model. All are located on the open sea.

Operational clearance & anchorage

This object simulates the operational clearance and the potential delays outside the port for the vessel entity group. The other entity groups do not visit this location in the simulation. When a vessel enters the operational clearance node, an add-on process is triggered. The add on-process checks whether the vessel needs a pilot, needs tugboats and from which company and checks whether there is enough capacity available in the terminal where the vessel is headed. It then checks if these required resources are available at that moment in time in the simulation model. If this is not the case, the vessel is sent to the anchorage, which is simulated by a parking station. There, another add-on process is triggered, that checks for the availability of these resources every minute. The moment that these resources are available in the model for a specific vessel, it is released from the parking station and it can continue its journey to the port.

Pilot boarding place

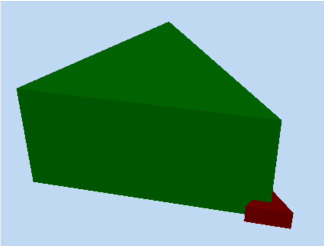

Figure V-6: Vessel with pilot

This node simulates the pilot boarding place. The two entity types that pass through this object in the model are the vessel group and the pilot group. When an entity enters this object, an add-on process is triggered. The add-on process checks whether the entity is a pilot or a vessel. If it is a pilot, it is placed in a queue, where it waits until it is requested by a vessel that needs a pilot. If it is a vessel, the add-on process checks whether the vessel requires a pilot. If this is not the case, the vessel continues its journey to the port. If this is the case, a pilot is released from the queue and batches with the vessel and they then proceed with their journey. An 3d view of a batched vessel and pilot is shown in figure V-6.

Tug Area

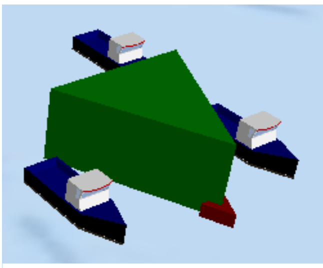

Figure V-7: Vessel with tugs and pilot

This object is located just outside of the entrance of the port. It simulates the tug area where vessels are linked to the right number of tugs from the corresponding tug company. When one of the three entity types pass this point, an add-on process is triggered. This process first checks of what type the passing entity is. If a pilot passes this point, nothing happens and it continues it journey back to the pilot boarding place. If a tug passes this point it is put in the available tugs queue, where it waits until it is requested by a Vessel. If a vessel passes, the add-on process checks whether the vessel needs a tug, how many tugboats are needed and from what tug company. If no tugs are needed, the vessel continues. If tugs are needed, the vessel will batch with the right number of tugs from the corresponding company and continue its journey to the port. Figure V-7 shows an 3d view of a vessel with a pilot and three tugs.

Terminals

There are multiple terminals that all have the same modelling logic. In order to prevent unnecessary double work, a sub-model is created that can be used multiple times in the main model for each terminal. The sub-model keeps track of the number of vessels that is inside and keeps track of the occupation of its quays (%). This is necessary in order to be able to simulate the operational clearance outside the port. When a vessel enters a terminal, an add-on process is triggered. This process checks whether the vessel has batch members (tugboats or a pilot). If this is the case, the members are unbatched and sent to the exit node of the sub-model, this way they directly enter the main model and travel back to the tug area and pilot boarding place. The vessel enters a server which simulates the critical services at the dock. An attribute of the vessel, number of remaining berth visits, gets a new value (=old value -1) and the processing time (berthing duration) is calculated. This is done by sampling a value from a corresponding empirical distribution and subtracting the average time that a vessel gets delayed at berth. This deduction is essential in order to prevent double punishment. It would be invalid to sample the berthing durations from an empirical distribution, which already includes the delays at berth, and then also simulate the berth delays. The time that gets deducted from the berthing durations is calculated by running the model and recording the simulated delay length at berth. After the berthing duration has passed a new add-on process is triggered. This process checks whether the vessel needs tugs and from which company or a pilot. If this is the case the vessel will wait until the required tugs and pilot become available. The moment that they become available, they batch with the vessel. Next, the add-on process checks whether the attribute, remaining berth visits, is zero. If this is the case the vessel and potential batch members leave the sub-model and continue their journey towards the exit of the port. Is this is not the case, the vessel and its batch members re-enter the sub-model and the entire terminal process starts again. An example of a terminal is shown in 2d snapshot in figure V-8. Three vessels were docked at the moment the snapshot was taken.

Figure V-8: Snapshot of Euromax terminal with three docked vessels.

V-IV Verification and validation

During the verification- and validation phase, it is tested if the model is accurate enough and behaves in a credible manner. This is essential for the next phase of the research, which is scenario testing in the simulation model. If the model behaves inaccurately, it is not possible to translate the conclusions of the experiments to policy in the real world.

The verification is done by testing if the attributes are correctly included in the model. Next, simulation properties will be checked and through a structural verification test, it is tested if the Simio model is consistent with the conceptualisation in appendix V-I.

The simulation model is validated in two ways. First, in a replicative validation test, the model outcomes are compared to empirical data. After that, an expert validation is performed, where an port call expert takes a detailed look at all the elements, steps and processes in the model, to conclude if the model performs accurately.

Model property verification

This section discusses and verifies the simulation setup that will be used to run the scenarios in chapter 6. The line of reasoning used to define the configuration is explained and the configuration is verified in the Simio model. The discussed elements are the warming-up period, the run length and the number of replications.

Warming-up period

In order to get the correct warming-up period, it is essential to know whether the model simulates an ending or a never-ending system. An example of a ending system are the customers in a supermarket that opens every morning and closes every night. The number of customers may vary during the day, but each day the supermarket starts with zero customers inside. An example of a never-ending system are the products in that same supermarket. The supermarket does not start with zero products every morning, it is always stocked, the system is always running, 24 hours a day, 365 days a year. Port calls in the Port of Rotterdam are a never-ending system. Vessels and their cargo are constantly being processed and the port never closes.

The warming-up period in the model is the period of time that the never-ending system needs to be simulated to reach its stable state. The simulation model will start with an empty port. This is not truthful. In order to avoid this, the first two weeks of the simulation, will be set as the warming-up period. During this period, no statistics are recorded. It is expected that this warming-up period is long enough, as it is more than ten times longer than the average time in system of all the vessels. Furthermore, during those two weeks, more than a thousand vessels enter the model.

The warming-up period of two weeks is correctly configurated in the Simio model. This has been double checked.

Run length

The run length is the period of time that is simulated in the model. This should be a minimum of three times the cycle that is modelled. In the case of this project one port call stands for one cycle. It is concluded from the process mining phase that most of the port calls take a handful of days. A run length of two to three weeks should already be sufficient. However, the analysed AIS event data, that is used as an input for the simulation model, originates from a period of exactly one year. This means, that all the ingredients are there to simulate an entire year, which would make the model more robust than when only two to three weeks are simulated. The only constraint to this would be the amount of time it would take to simulate a year, as this would demand more computing power. Fortunately, the model, when performing on the highest speed possible, needs less than two minutes to run one replication of a year. This is more than acceptable for this research. Therefore, the run length is set to one year, from the 1st of January 2018 till the 1st of January 2019.

The run length of one year is correctly configurated in the Simio model. This has been double checked.

Number of replications

The simulation model is not entirely deterministic. The berthing durations are sampled from vessel type specific empirical distributions. This results in different model outputs during different simulations. One replication would therefore be insufficient to draw conclusions from. Therefore it is chosen to simulate 40 times with a different seed per scenario.

The number of replications is correctly configurated in the Simio model. This has been double checked.

Structural verification

During the structural verification, the consistency between the simulation model and the conceptualisation (Appendix V-I) is checked. First, it is verified whether the different entity groups, go through the modelled process correctly. Then, it is verified whether the Simio-model is consistent with the building blocks described in Appendix V-I.

Model entities

There are three entity groups modelled separately. Per group, multiple entities are tracked throughout the model to check whether they behave in a way that is in line with the conceptualisation. Based on the degree of complexity, a few points in the model are identified as high-risk model elements. There is a special focus on the following points when checking the behaviour of the entities: Operational clearance, anchorage, pilot boarding place, tug area and terminal. Also, the entry and exit points of different entities and their path and speed is verified. From this verification method, it is concluded that the tracked entities are behaving correctly throughout the biggest part of the simulation. However, there are some remarks that are discussed in this section.

When a vessel requests an available tugboat or an available pilot that is located in the tug area or pilot boarding place, the pilot and the tugboat teleport from their location to the vessel's location in one simulation step. This is not how these entities should behave, as they should sail to their destination. The reason that this is happening is that the modelling logic did not take vessels into account that request these nautical services from a different location than the pilot boarding place and the tug area. An easy and quick fix was not found and it was decided to leave it like this for now. The benefits do not outweigh the drawbacks, as there is only a limited amount of time for this research and this modelling error is not expected to affect the outcome of the simulation runs significantly. The biggest impact it has is on the visualisation. Other than this, the tugboats and pilots behave in line with appendix V-I.

The second remark is about small container vessels visiting one of the deep-sea terminals in combination with the model reduction that a vessel with multiple destinations in the port will visit its first destination multiple times, instead of proceeding to the other actual destinations (appendix V-II). This causes an unforeseen consequence. Small container vessels often have more destinations in the port than large container vessels and are more likely to visit terminals that are not deep-sea terminals. Their berthing durations are also shorter on average. Every vessel that visits a deep-sea container terminal, samples the berthing duration from a empirical distribution of berthing durations of large container vessels. When a small container vessel, that happens to have multiple destinations, visits a deep-sea container terminal as its first location, it samples from the wrong distribution multiple times. This can lead to the deep-sea container terminals being more busy than they are in reality.

The final remark has to do with the visualization. The entities are not visualized in correct proportions on the map. The vessel entity group and the pilot entity group are both visualized with a triangle, which may lead to confusion, even though they are different sizes and colour. Furthermore, the vessel entity group has an attribute 'ship length', which is correctly assigned to the vessels in the model. However, it does not affect the size of the vessels. This seems like a missed opportunity for a better visualization. It was chosen during the modelling phase, not to focus too much on the visual aspect of the model as it does not affect the outcome.

Model objects

The building blocks and their attributes (Appendix V-I) are compared to the objects in the Simio-model. Furthermore, the model is also checked for additional model objects that have not been specified in the conceptualization. It is concluded that the model objects are in line with the conceptual building blocks.

Replicative validation

The replicative validation will assess whether the simulation model produces accurate data. Model behaviour, expressed in values, is compared to historical data of port calls from the AIS event data. The results are presented in tables. All cells in the tables that display a valid value are shown in green. All cells that show an invalid value are shown in red. In all the cases where the validity is difficult to dis(prove) are shown in yellow. The results in yellow and red are discussed in detail.

Nautical services and berth destinations

First, it is checked whether the model simulates six randomly picked actual port calls from 2018. Furthermore, for these port calls it is also validated if they use the correct number of tugboats, correct pilot usage and go to the correct destination, on the correct time.

Source	UCRN	MMSI	Tugs to first berth	Pilot	First destination	Time of arrival at first berth
Historical			1	Yes	RST NZ	25-01 03:04
Simulated			1	Yes	Small container terminal	25-01 01:12

Table V-3: Validation vessel A

Source	UCRN	MMSI	Tugs to first berth	Pilot	First destination	Time of arrival at first berth
Historical			2	Yes	APM WZ	08-03 16:26
Simulated			2	Yes	APMT	08-03 14:51

Table V-4: Validation vessel B

Source	UCRN	MMSI	Tugs to first berth	Pilot	First destination	Time of arrival at first berth
Historical			1	Yes	METAAL TRANSPORT NZ	16-05 00:44
Simulated			1	Yes	Breakbulk t	15-05 19:05

Table V-5: Validation vessel C

Source	UCRN	MMSI	Tugs to first berth	Pilot	First destination	Time of arrival at first berth
Historical			0	No	STENALINE 1	09-23 07:04
Simulated			0	No	Passenger terminal	09-23 05:07

Table V-6: Validation vessel D

Source	UCRN	MMSI	Tugs to first berth	Pilot	First destination	Time of arrival at first berth
Historical			0	Yes	LYONDELL 1	05-11 00:49
Simulated			0	Yes	Liquid bulk terminal	04-11 23:19

Table V-7: Validation vessel E

Source	UCRN	MMSI	Tugs to first berth	Pilot	First destination	Time of arrival at first berth
Historical			2	No	P&O OZ Z	08-12 14:25
Simulated			2	No	Passenger terminal	08-12 06:39

Table V-8: Validation vessel F

When looking at the results in table V-3 to V-8 it can be concluded that all the randomly picked port calls are included in the simulation. Furthermore, the incoming vessels all use the correct combination of nautical services. The destinations are correctly simulated as well, even though it sometimes seemingly gives two different terminals. For example the vessel from table V-8 went to the P&O OZ Z terminal, while in the simulation model it went to the Passenger terminal. However, P&O OZ Z is a ferry terminal and is thus part of the Passenger terminal in the model. Some terminals that serve the same vessel type are grouped in the model, in order to reduce complexity. This model reduction is explained in more detail in appendix V-II. The only values that are not exactly the same are the times of arrival at the first berth location in the port. This has multiple reasons. The time of arrival is dependent on stochastic elements in the simulation, like the speed distributions (table V-2) and the empirical distributions of the berthing durations. These elements influence the time of arrival. Another factor is that the model loses accuracy because of certain model reductions (Appendix V-II). For example, some terminals are bundled into one large terminal. The distances to those bundled terminals are all the same in the model, while in reality they are not. However, the difference seems to be relatively small. All the checked vessels arrive within a few hours of the actual time of arrival. Therefore, it is concluded that when it comes to the model elements compared above, the model is valid enough for simulation.

Expert validation

Two separate expert validations have been performed. The first one is performed incrementally with a data scientist and port call expert. During the creation of the model, every week a consultation session was held with this expert, with the main goal to create a valid model. In these sessions, an updated version of the simulation model was presented and in return, feedback was given by the expert, which was processed every time.

The second expert validation is performed by a data scientist and port call expert. This session was held after the model was created on the 17th of July 2019. The expert concluded that the model is valid enough for scenario testing in chapter 6. However, there are still some flaws in the model that could be improved, to increase the validity. These flaws are discussed in this section.

When a tugboat successfully escorts a vessel to its berth location, the vessel starts with the critical services at berth and the tugboat becomes available again for other towage requests. The expert stated that in this situation, it often happens that a tug is requested by another vessel nearby, to escort the vessel outside of the port. This is not simulated in the Simio-model. In the model a tug, that finishes a towing service, first returns to the tug area, before it becomes available for other requests. The availability of the tugboats in the simulation model is lower compared to real life, because this modelling logic is missing. It is too complex to add this element to the model in the time that is left for this research.

When a tugboat is requested by a vessel at berth, it often happens that the tugboat only assists in pulling the vessel away from the quay and then releases itself from the vessel. The vessel continues its journey on its own and the tugboat becomes available again for other towage requests. In the simulation model, all the vessels that need a tug to assist them, as they leave their berth location, get escorted all the way to the tug area (at the exit of the port). This would negatively influence the availability of the tug service. However, as long as the previously discussed flaw is not fixed, this flaw will not influence the simulation, because the tugboats return to the tug area anyway, with or without the vessel.

The final flaw, is a flaw that is already discussed in the structural verification of entities (Appendix IV-IV). In short, it is an unforeseen effect of the model reduction that states that the port calls with multiple berth visits, are all simulated at the first berth of that port call. In other words, if a vessel has ten destinations in the port and the Euromax terminal is the first one, it will visit the Euromax ten times in the simulation. An already discussed effect is that smaller container vessels, that first visit one of the deep-sea container terminals, will put extra pressure on those terminals, because they will visit that terminal multiple times. Another consequence, that is not discussed before and is identified during the expert validation, is that the tugboats, that are requested by these vessels will only have to escort the vessel to the same terminal, again and again. This causes them to be less occupied in the model as the distance that they need to tow the vessel is zero. In other words, when this situation occurs and a tug is requested by a vessel has multiple visits in the port, the tugs go to the terminal of the vessel, connect to the vessel, and immediately disconnect (because the destination is immediately reached) and return to the tug area. In reality, they would first have to escort the vessel to a different terminal, or at least pull it away from the quay. This will positively influence the availability of the tugboats in the simulation model.

As mentioned above, the expert concludes that the model is valid enough for the proposed what-if analysis in chapter 6, however the invalid elements discussed above need to be taken into account when drawing conclusions and writing the recommendations. Fortunately, one flaw influences the availability of tugboats positively and the other one negatively. From this, it can be concluded that these two flaws compensate each other to an unknown degree. None the less, these are flaws that should be fixed and need to be taken into account.

Appendix VI – Statistical comparison of the simulation model results

In order to be able to draw conclusions from the experiments, the results of the simulations are statistically analysed and compared. This is discussed in this section.

First the method and approach for the statistical analysis is described. This contains all the potentially needed statistical tests and how they relate to each other. This also includes a brief summary of the tests. This is followed by the implementation and the results of the statistical analysis and comparison per KPI. How to interpret the results from the statistical analysis is discussed in chapter 6.5.

VI-I Method and approach

Figure VI-1 shows a chart that assists in determining the best fitting statistical tests. First Shapiro-Wilk tests are performed to conclude whether it is feasible that the datasets originate from a population with a normal distribution. This is necessary for determining what test to use when comparing the two scenarios. When the Shapiro-Wilk test concludes that a KPI is not normally distributed, the Wilcoxon rank test is used to compare that KPI for each scenario. When the Shapiro-Wilk test concludes that a KPI is normally distributed, a Levene's test is performed to check for equal variance. In the case of equal variance, a Student t-test is performed, otherwise a Welch t-test is performed. For all tests, a confidence interval is chosen of 95%. All the analyses are executed using IBM SPSS Statistics. This is a software for analysing data and performing statistical tests. It includes all the tests mentioned in figure VI-1.

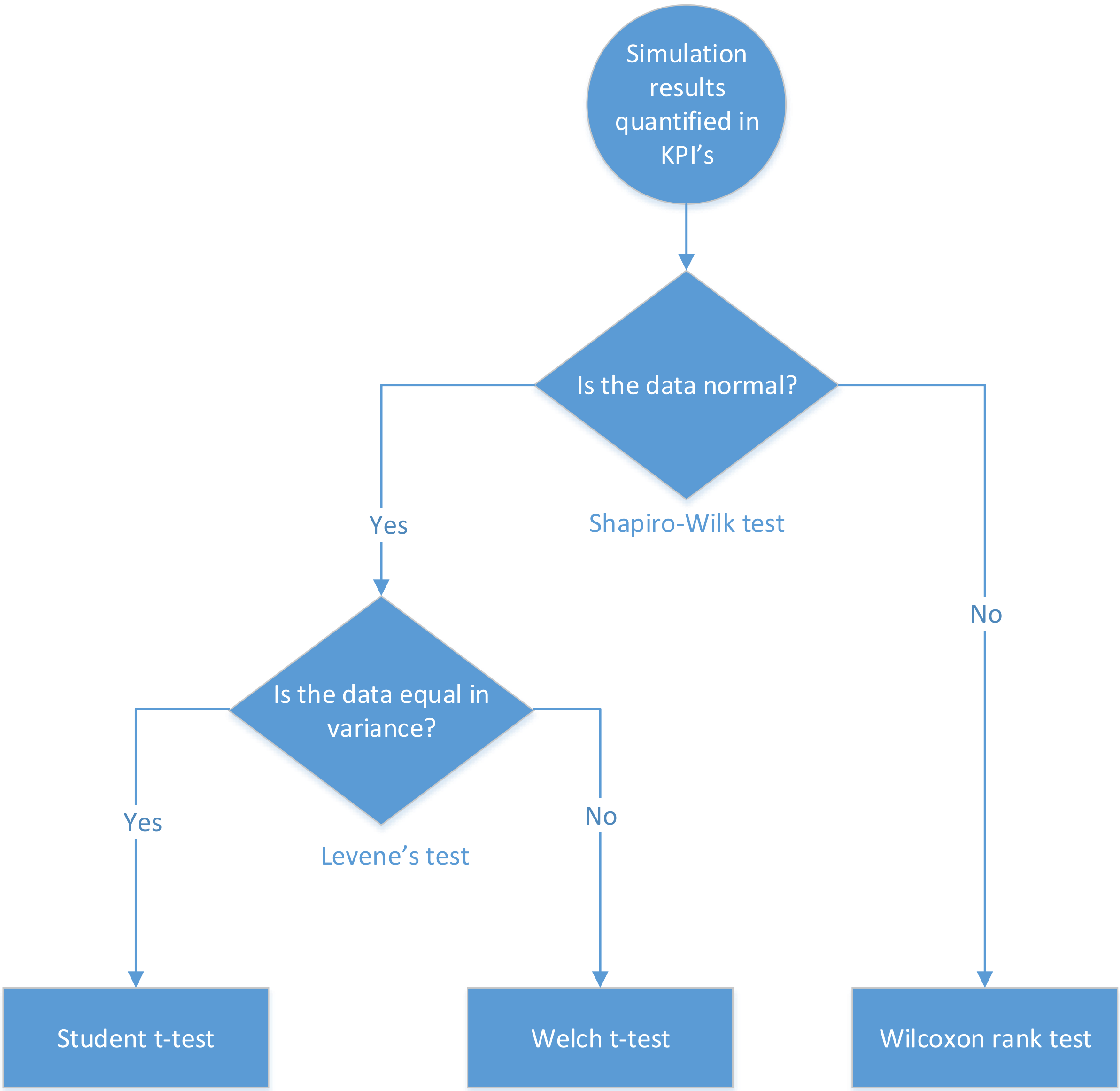

Figure VI-1: Chart for statistical tests

Shapiro-Wilk test

The Shapiro-Wilk test is introduced by S. S. Shapiro and M. B. Wilk (1965) and is designed to test a sample for normality. The test is executed by dividing the square of an linear combination of the sample order statistics by the usual symmetric estimate of variance. This ratio is invariant in both scale and origin and is therefore fit for to test the hypothesis of normality. The null hypothesis and the equation are as follows:

The population of the sampled data is normally distributed

$$W = \left(\frac{\left(\sum_{i=1}^{n} a_i y_i\right)^2}{\sum_{i=1}^{n} (y_1 - \bar{y})^2}\right)$$

Levene's test

The Levene's test is introduced by H. Levene (1960), with the main purpose to determine whether the variances of samples of two groups is equal. Levene's test can be used as a main test to tackle an issue. However, in this case it is used before an comparison of means, to assess whether a test that presumes homogeneity of variance is applicable, or that more general test is needed that does not presume this. The null hypothesis and the equation are as follows:

The variance of the population of two or more sample groups are equal

$$W = \frac{(N-k)}{(k-1)} * \frac{\sum_{i=1}^{k} N_i (Z_{i.} - Z_{..})^2}{\sum_{i=1}^{k} \sum_{j-1}^{Ni} (Z_{ij} - Z_{i.})^2}$$

Independent two-sample t-test

The independent two-sample t-test is first introduced in 1908 (Gosset), with the goal to assess whether the average difference between two groups is significant. This test is especially necessary for two reasons, when comparing two groups with relative small sample sizes. The two sample groups may seem to originate from two different populations as a result of the error of random sampling. Furthermore, the two sample groups may seem from two different populations because their distributions vary, as a result of their small sizes. This test assists in assessing if this is the case. The null hypothesis and the equation are as follows:

There is no significant difference in mean between the two groups

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{((n_1 - 1)s_{x_1}^2 + (n_2 - 1)s_{x_2}^2)/(n_1 + n_2 - 2)}}$$

Welch t-test

The Welch t-test is first introduced by B. L. Welch (1947). The test is used to check whether two groups have equal means. In other words, it assesses is there are any significant differences between the two sample groups. This test is more reliable than the independent two-sample t-test in the case of unequal variance or unequal sample size. However, the assumption of normality remains. The null hypothesis and the equation are as follows:

There is no significant difference in mean between the two groups

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{(s_1^2/n_1 + s_2^2/n_2)}}$$

Wilcoxon signed-rank test

The Wilcoxon signed-rank test is introduced by F. Wilcoxon (1945). It is used to test the significance of the difference of the means of two non-parametric samples. In other words, it is used for samples of populations of which it cannot be assumed that they are normally distributed. The null hypothesis is as follows:

There is no significant difference in mean between the two groups

VI-II Results statistical analysis

The results for each KPI are analysed separately. This section reports the results from the statistical comparisons.

Throughput time operational port call

A visual representation of throughput time of the operational part of the port call for container vessels for the two scenarios is shown in figure VI-2. Looking at these histograms, it seems that the two samples originate from a normally distributed population. The significance of the Shapiro-Will test for the normality for this KPI confirms this. The p-values are greater than 0.05 for both scenarios (table VI-1). From this it is concluded that the null hypothesis remains, so it can be assumed that both samples originate from a normal distribution.

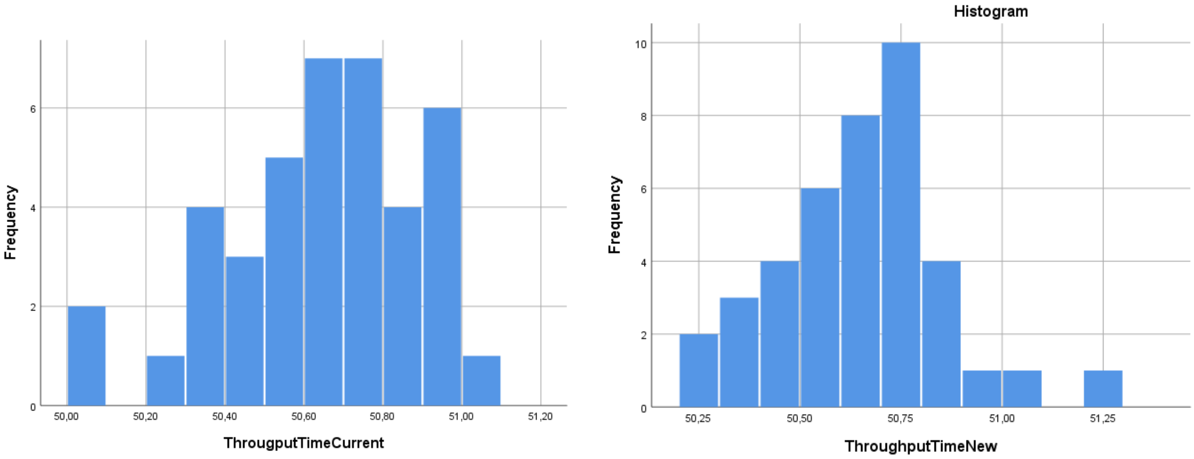

Figure VI-2: Histogram throughput operational port call

	Shapiro-Wilk		
	Statistic	df	Sig.
Current scenario	,959	40	***,153***
New scenario	,981	40	***,742***

Table VI-1: Shapiro-Wilk test Throughput time operational port call

The next step is to test whether there is a homogeneity of variances between the current situation and the scenario with collaborating tug companies for this KPI. Levene's test shows, for all methods, p-values greater than 0.05 (table VI-2). This means that the null hypotheses cannot be rejected, and thus it is assumed that the variances are equal and that an independent two-sample t-test is the best fitting test for these samples.

		Levene Statistic	df1	df2	Sig.
Throughput Time	Based on Mean	,910	1	78	***,343***
	Based on Median	,678	1	78	***,413***
	Based on Median and with adjusted df	,678	1	76,508	***,413***
	Based on trimmed mean	,803	1	78	***,373***

Table VI-2: Levene's test Throughput Time

The independent sample t-test is performed. The results are shown in table VI-3. The p-value is greater than 0.05. This means that the null hypotheses is accepted. In other words, there is not a significant difference in the throughput time of the operational part of a port call for the two scenarios.

	t-test for equality of means		
	t	df	Sig.
Through put time	-0.080	78	***0.973***

Table VI-3: Independent samples t-test throughput time

Total number of anchorage visits

A visual representation of total anchorage visits of container vessels for the two scenarios is shown in figure VI-3. Looking at these histograms, it seems that the two samples originate from a normally distributed population. The significance of the Shapiro-Will test for the normality for this KPI confirms this. The p-values, shown in green, are greater than 0.05 for both scenarios (table VI-4). From this it is concluded that the null hypothesis remains, so it can be assumed that both samples originate from a normal distribution.

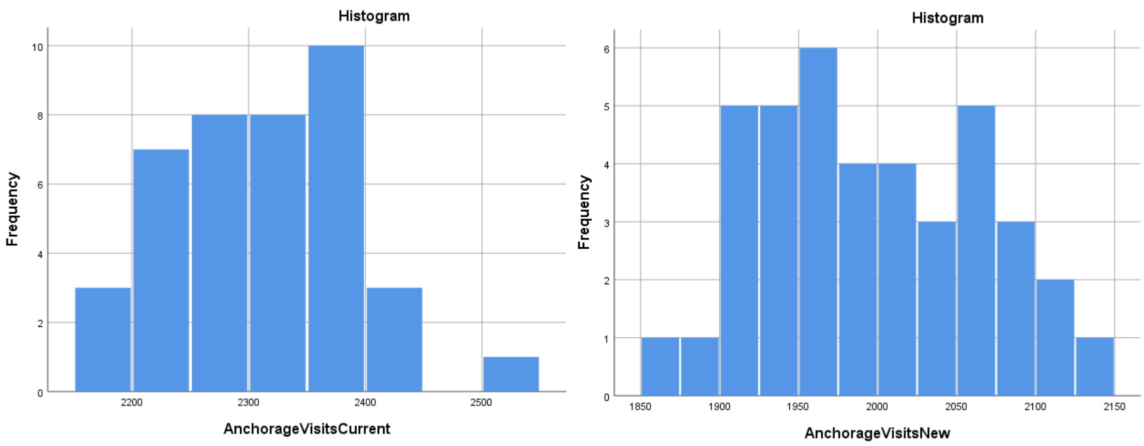

Figure VI-3: Histogram total number of anchorage visits

	Shapiro-Wilk		
	Statistic	df	Sig.
Current scenario	,983	40	***,813***
New scenario	,965	40	***,249***

Table VI-4: Shapiro-Wilk test anchorage visits

The next step is to test whether there is a homogeneity of variances between the current situation and the scenario with collaborating tug companies for this KPI. Levene's test shows, for all methods, p-values greater than 0.05 (table VI-5). This means that the null hypotheses cannot be rejected, and thus it is assumed that the variances are equal and that an independent two-sample t-test is the best fitting test for these samples.

		Levene Statistic	df1	df2	Sig.
Anchorage Visits	Based on Mean	,661	1	78	***,419***
	Based on Median	,738	1	78	***,393***
	Based on Median and with adjusted df	,738	1	77,406	***,393***
	Based on trimmed mean	,676	1	78	***,414***

Table VI-5: Levene's test anchorage visits

The independent sample t-test is performed. The results are shown in table VI-6. The p-value is smaller than 0.05. This means that the null hypotheses is rejected. In other words, there is a significant difference in anchorage visits for the two scenarios.

	t-test for equality of means		
	t	df	Sig.
Anchorage visits	19.225	78	***0.000***

Table VI-6: Independent samples t-test anchorage visits

Total average delay at anchorage

A visual representation of the delay at the anchorage for the two scenarios is shown in figure VI-4. Looking at these histograms, it seems that the two samples originate from a normally distributed population. The significance of the Shapiro-Will test for the normality for this KPI does not confirm this. The p-value of the current scenario is greater than 0.05, therefore it is concluded that it is normally distributed. However, the p-value of the new scenario is lower than 0.05, which means that it cannot be assumed that its population is normally distributed (table VI-7). The null hypotheses is rejected.

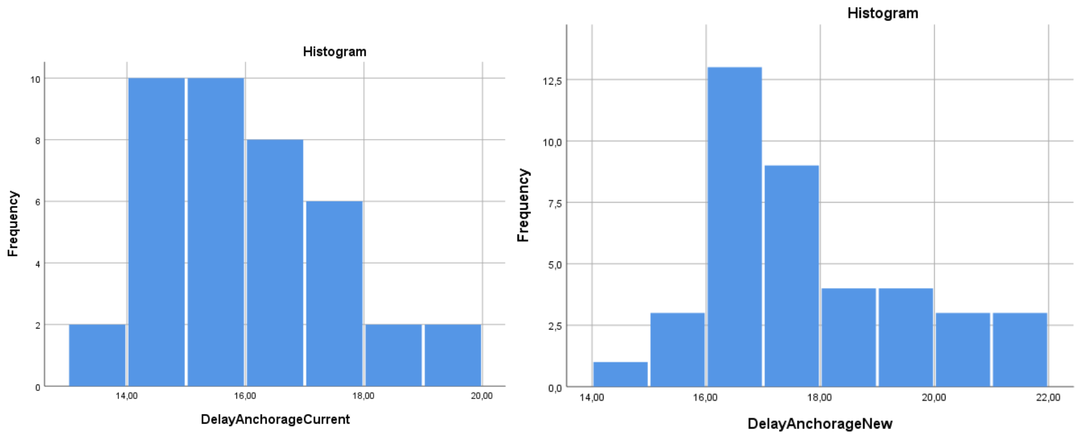

Figure VI-4: Histogram delay at anchorage

	Shapiro-Wilk		
	Statistic	df	Sig.
Current scenario	,967	40	***,279***
New scenario	,921	40	***,008***

Table VI-7: Shapiro-Wilk test delay at anchorage

The Wilcoxon signed ranks test is performed and the results are shown in table VI-8 and VI-9. The test determined that in 11 replications the delays are bigger in the current scenario and in 29 replications that they are bigger in the new scenario. This indicates a significant difference in mean. This suspicion is confirmed in table VI-9. A p-value smaller than 0.05 is calculated, which means that the null hypotheses is rejected. It is concluded that there is a significant difference between the current and new scenario when it comes to delays at the anchorage.

		N	Mean Rank	Sum of Ranks
DelayAnchorageNew - DelayAnchorageCurrent	Negative Ranks	11[a]	11,45	126,00
	Positive Ranks	29[b]	23,93	694,00

a. DelayAnchorageNew < DelayAnchorageCurrent

b. DelayAnchorageNew > DelayAnchorageCurrent

Table VI-8: Wilcoxon signed ranks test delay anchorage

	DelayAnchorageNew - DelayAnchorageCurrent
Z	-3,817
Asymp. Sig. (2-tailed)	,000

Table VI-9: Wilcoxon test statistics delay anchorage

Total average delay at berth

A visual representation of the delay at berth for the two scenarios is shown in figure VI-5. Looking at these histograms, it seems that the two samples originate from a normally distributed population. The significance of the Shapiro-Will test for the normality for this KPI does not confirm this. The p-value of the current scenario is greater than 0.05, therefore it is concluded that it is normally distributed. However, the p-value of the new scenario is lower than 0.05, which means that it cannot be assumed that its population is normally distributed (table VI-10). The null hypothesis is rejected.

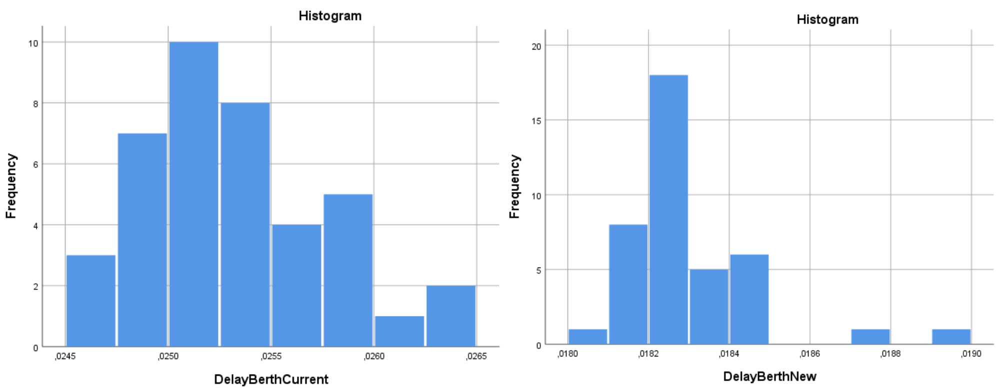

Figure VI-5: Histogram average delay at berth

	Shapiro-Wilk		
	Statistic	df	Sig.
Current scenario	,975	40	***,518***
New scenario	,815	40	***,000***

Table VI-10: Shapiro-Wilk test delay at berth

The Wilcoxon signed ranks test is performed and the results are shown in table VI-11 and VI-12. The test determined that in 40 replications the delays are bigger in the current scenario. This indicates a significant difference in mean. This suspicion is confirmed in table VI-12. A p-value smaller than 0.05 is calculated, which means that the null hypotheses is rejected. It is concluded that there is a significant difference between the current and new scenario when it comes to delays at the berth. Table 6.1 suspects a decrease in delay. In combination with the proved significant difference, it is assumed that the delay at berth will decrease as a result of collaboration of tug companies.

		N	Mean Rank	Sum of Ranks
DelayBerthNew - DelayBerthCurrent	Negative Ranks	40[a]	20,50	820,00
	Positive Ranks	0[b]	,00	,00

a. DelayBerthNew < DelayBerthCurrent

b. DelayBerthNew > DelayBerthCurrent

Table VI-11: Wilcoxon signed ranks test delay at berth

	DelayBerthNew - DelayBerthCurrent
Z	-5,511
Asymp. Sig. (2-tailed)	,000

Table VI-12: Wilcoxon test statistics delay at berth

Availability of the pilots

A visual representation of the availability of the pilots for the two scenarios is shown in figure VI-6. Looking at these histograms, it seems that the two samples originate from a normally distributed population. The significance of the Shapiro-Will test for the normality for this KPI confirms this. The p-values, shown in green, are greater than 0.05 for both scenarios (table VI-13). From this it is concluded that the null hypothesis remains, so it can be assumed that both samples originate from a normal distribution.

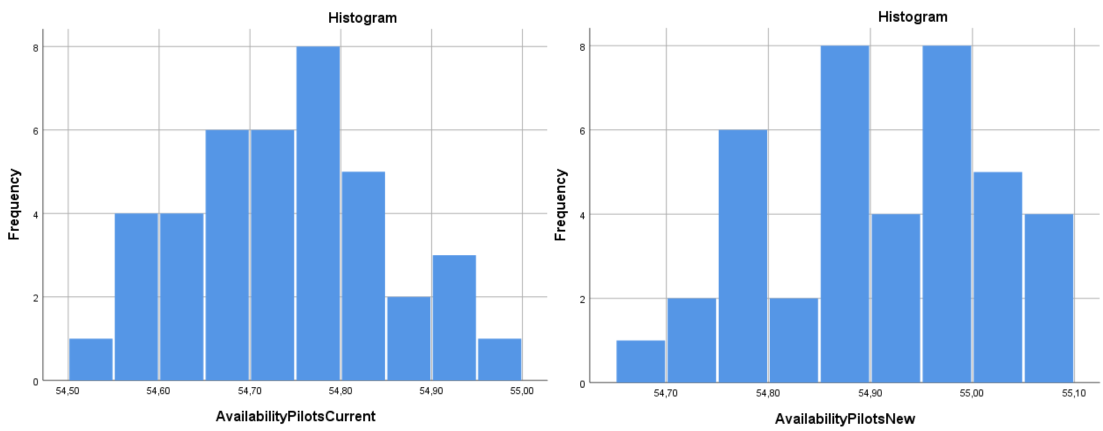

Figure VI-6: Histogram availability pilots

	Shapiro-Wilk		
	Statistic	df	Sig.
Current scenario	,981	40	***,772***
New scenario	,975	40	***,466***

Table VI-13: Shapiro-Wilk test availability pilots

The next step is to test whether there is a homogeneity of variances between the current situation and the scenario with collaborating tug companies for this KPI. Levene's test shows, for all methods, p-values greater than 0.05 (table VI-14). This means that the null hypotheses cannot be rejected, and thus it is assumed that the variances are equal and that an independent two-sample t-test is the best fitting test for these samples.

		Levene Statistic	df1	df2	Sig.
Availability Pilots	Based on Mean	,256	1	78	***,614***
	Based on Median	,222	1	78	***,639***
	Based on Median and with adjusted df	,222	1	77,998	***,639***
	Based on trimmed mean	,241	1	78	***,625***

Table VI-14: Levene's test availability pilots

The independent sample t-test is performed. The results are shown in table VI-15. The p-value is smaller than 0.05. This means that the null hypotheses is rejected. In other words, there is a significant difference in the availability of pilots for the two scenarios.

	t-test for equality of means		
	t	df	Sig.
Anchorage visits	-7.114	78	***0.000***

Table VI-15: Independent samples t-test availability pilots

Availability of the tugboats

A visual representation of the availability of the tugs for the two scenarios is shown in figure VI-7. Looking at these histograms, it seems that the two samples originate from a normally distributed population. The significance of the Shapiro-Will test for the normality for this KPI confirms this. The p-values, shown in green, are greater than 0.05 for both scenarios (table VI-16). From this it is concluded that the null hypothesis remains, so it can be assumed that both samples originate from a normal distribution.

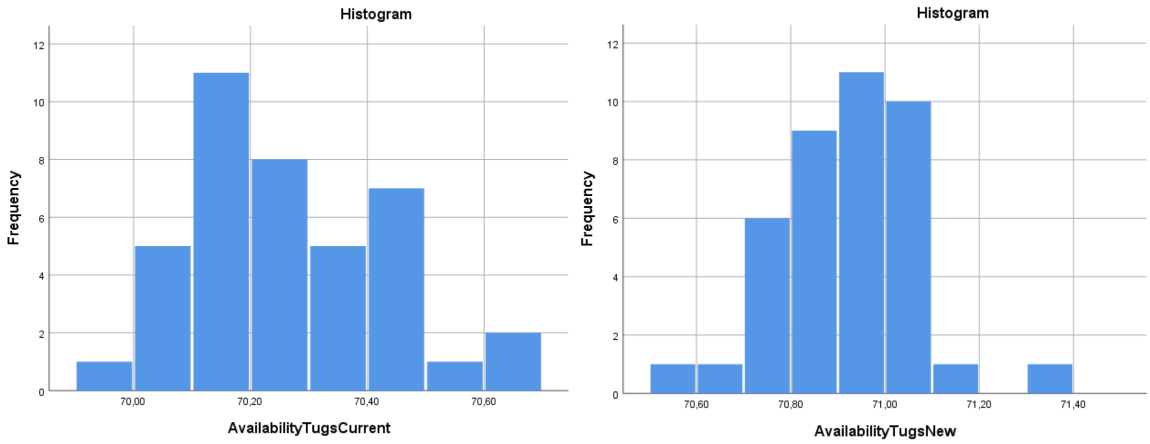

Figure VI-7: Histogram availability tugs

	Shapiro-Wilk		
	Statistic	df	Sig.
Current scenario	,961	40	***,180***
New scenario	,962	40	***,197***

Table VI-16: Shapiro-Wilk test availability tugs

The next step is to test whether there is a homogeneity of variances between the current situation and the scenario with collaborating tug companies for this KPI. Levene's test shows, for all methods, p-values greater than 0.05 (table VI-17). This means that the null hypotheses cannot be rejected, and thus it is assumed that the variances are equal and that an independent two-sample t-test is the best fitting test for these samples.

		Levene Statistic	df1	df2	Sig.
Availability Tugs	Based on Mean	1,921	1	78	***,170***
	Based on Median	1,841	1	78	***,179***
	Based on Median and with adjusted df	1,841	1	77,311	***,179***
	Based on trimmed mean	1,875	1	78	***,175***

Table VI-17: Levene's test availability tugboats

The independent sample t-test is performed. The results are shown in table VI-18. The p-value is smaller than 0.05. This means that the null hypotheses is rejected. In other words, there is a significant difference in the availability of tugs for the two scenarios.

	t-test for equality of means		
	t	df	Sig.
Anchorage visits	-18,784	78	***0.000***

Table VI-18: Independent samples t-test availability tugboats