GIMA
Geographical Information Management and Applications

Identify and visualize Dutch inland waterways vessel movement anomalies during low water levels

Master thesis

Chia-Yun, Chang

Student number: 6689361
Email: c.chang3@students.uu.nl

Supervisor: Martijn Meijers (TU Delft)
Fedor Baart (Deltares)
Solange van der Werff (TU Delft)

Responsible professor: Peter van Oosterom

February 2021

Universiteit Utrecht
TU Delft
UNIVERSITY OF TWENTE.
WAGENINGEN UR

ACKNOWLEDGEMENTS

Throughout the writing of this thesis, I have received a great deal of support and assistance.

First, I would like to thank my supervisors Martijn Meijers, Fedor Baart, Solange van der Werff, and Peter van Oosterom for their guidance and inspiring ideas throughout the whole research. The feedback and resources you provide have sharpen my skill and broaden my knowledge.

I would also like to thank Rijkswaterstaat for providing the dataset to make this research possible, and Jurjen de Jong from Deltares who gave me great insights and views from different perspectives. I would also like to acknowledge my fellow GIMA student Virginia for our daily study session and discussion.

Finally, I would like to thank Joakim for supporting me and helping me through this difficult time. Special thanks to my brother for his mathematical views and my parents for their encouragement and support.

ABSTRACT

This MSc thesis aims to develop a workflow analyzing inland vessel traffic anomalies during low water levels by using historical automated identification system (AIS) data. Low water level impacts inland waterway transports because cargo ships must carry less weight in order to stay safe and afloat. The consequences are the same amount of freight needs more trips to be transported comparing to normal water level times, and more trips lead to busier waterways and higher freight rates. Many inland transportation reports and news articles have discussed the impacts of low water levels on inland shipping, but not many studies have used AIS data and water level data to analyze the relationship. Thus, this research tries to fill in this gap in two aspects: validate the market observations and detect vessel interaction anomalies.

AIS is a system that broadcasts vessel status and information to other vessels in the same area, so vessels can know each other's whereabouts and act accordingly. The information contained in AIS data are categorized as dynamic and static information, the first changes depending on the position and movement of the vessel, while the latter is about vessel identity and voyage information such as vessel IDs, destination, vessel types, and vessel size. The spatial and temporal information in AIS data can provide researchers an opportunity to look into this topic from a different perspective. Therefore, the research question is: *to what extent can historic AIS data contribute to the analysis of the impacts brought by low water level in inland waterways?*

The research was divided into two phases. First, four statements were summarized based on news articles and market reports, and a workflow was developed to validate those statements with AIS data analysis. Second, the vessel interaction anomalies were narrowed down to starboard side encounter events, which are allowed under certain conditions in the study area. Ship encounters were detected by the distance to the other ship, then the encounter events were classified by which side of the ship they met.

The study results show that AIS data can be used to identify how busy the waterway was by extracting and counting the number of trips. However, analysis that uses draught value as the data source is not recommended since the draught values in AIS data are not reliable. The analysis on the speed and water levels shows no correlation between these two elements, thus the statement is not true. On the other hand, the results of vessel size changes indicate that there are no significant patterns regarding vessel size and water levels, which might be because of a wrong study area. Lastly, the analysis results of ship-encounter anomalies present that starboard side encounters are related to how busy is the waterway.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AIS Automatic Identification System

API Application Programming Interface

CCNR Central Commission for the Navigation of the Rhine

COG Course over Ground

DBMS Database Management System

ENI unique European vessel Identification Number

ETA Estimated Time for Arrival

GIS Geographical Information System

IMO International Maritime Organization

M^3 Massive Movement Model

MMSI Maritime Mobile Service Identity

RWS Rijkswaterstaat

ROT Rate over Turn

SFC Space Filling Curve

SOG Speed over Ground

UTC Universal Time Coordinated

WMS Web Map Service

1. INTRODUCTION

In the Netherlands, inland waterway transport takes up more than 43% of the modal share. Every year, vessels transport more than 46 billion tonne-kilometre (moving *n* tonne of cargo a distance of *n* km; 1 tonne is 1,000 kg) goods (CCNR, 2019b). Inland waterways, which include rivers, canals, and lakes, play an important role in the Netherlands' economy and development. Furthermore, one of the goals in EU and the Netherlands' policy: Green Deal, is to optimize the inland shipping logistic chain, and make it more sustainable and resilient to impacts (European Barge Union, 2020; Green Deal, 2019).

This chapter explains why and how the research topic was developed, and how is it relevant to inland waterway transport. Section 1.1 introduces the background of topic development, and section 1.2 explains the relevance and possible contribution of this thesis. Then sections 1.3 and 1.4 establish the research aim and research question. Section 1.5 illustrates the research scope that states the research area, period, and subject. Lastly, section 1.6 describes this research and its methodology procedures to give an overview of the thesis.

1.1 Background

Low water level limits cargo traffic and capacity, which threatens the supply chain of goods (Meijer, 2018). The direct consequence of low water level is that cargo vessels have to load less in order to not run aground and sail safely on water, which leads to more trips to transport the same weight of cargo compared to days with normal water levels (Barendregt, 2018; NOS NIEUWS, 2018).

In spite of the disturbance brought to supply chain and industrial sectors, studies indicate that the navigability of canals and impounded rivers is less affected by drought because the water level can be determined artificially, and waterways close to estuaries are also less affected due to the river-sea nature (Klein &Meißner, 2017). Moreover, a report (CCNR, 2019a) states that canal transport activities in the Netherlands, Belgium, France, and northern Germany were significantly less affected during the 2018 drought comparing to rivers in the same period. This report also claims that the freight rates in these regions still rose, possibly due to many smaller cargo vessels shifted their operation area to the middle or the upper Rhine to take advantage of the high freight rate there. To be more specific, these reports and studies show that despite the Netherlands' canals were less affected by the drought, the poor navigability in the middle or upper stream can still impact the Dutch inland waterways transport. Moreover, Verschuren (2020) uses IVS90 data to investigate impacts of 2018 drought on the traffic flow and traffic capacity in the Waal. The results support the statement about busier traffic during low water levels and low water discharge period.

Despite the impacts that happened, news and articles (Barendregt, 2018; NOS NIEUWS, 2018; Rijkswaterstaat, 2018) only mentioned there were impacts on shipping based on observation or past experiences (Jonkeren &Rietveld, 2009), or studies using the freight amount and shipping information (CCNR, 2019a; Verschuren, 2020). The analysis of vessels and traffic by using AIS data is lacking. As a result, this thesis topic was formed in an attempt to fill in the gap between economic perspectives and vessel movement patterns presented by the automated identification system (AIS) data. AIS was developed to improve communication between vessels, and between vessels and traffic posts. Since December 1, 2014, AIS devices are mandatory for most types of transport vessels on the Rhine (Rijkswaterstaat, n.d.-a).

1.2 Relevance

Many have studied the impacts of low water level on the inland waterways transport activities by analyzing freight rates, total transported cargo weight, transport modal split, and experiences of carriers and shippers (CCNR, 2019a; Jonkeren, Jourquin, &Rietveld, 2011; Jonkeren, Rietveld, &vanOmmeren, 2007; Jonkeren &Rietveld, 2009; Scholten, Rothstein, &Pistocchi, 2017). However, the impacts were rarely examined from the perspective of vessel behavior or distribution patterns. According to the Central Commission for the Navigation of the Rhine (2019), historic data show water level fluctuations that can affect navigational conditions do not necessarily relate to climate change and are probably going to reoccur in the future. The research results of this thesis may contribute to inland waterway management by presenting the relevant factors in AIS data that can affect navigation. Apart from the research outcomes, only Python and Jupyter Notebook were employed to conduct the analysis, so the possible variety of applications is wider.

1.3 Aim

The research aim is to develop a workflow that uses historical AIS data to analyze traffic anomalies on inland waterways during low water levels. Furthermore, this aim can be divided into two parts. First, to find an effective way to detect and visualize traffic anomalies by exploiting historic AIS data because it has rarely been studied before. Second, to understand to what extent can historic AIS data analysis contribute to the study of low water level impacts, unlike other studies that view the issue from an economic perspective.

The information on inland waterway traffic's spatial or temporal distribution during low water level can provide decision-makers more information on the aspect of vessel movement, so to be better prepared for the next potential low water levels.

1.4 Research Questions

To achieve the objectives, the following question will be answered:

To what extent can historic AIS data contribute to the analysis of the impacts brought by low water level in inland waterways?

To answer the main question, five sub-questions are developed:

- *What types of impacts are caused by low water levels in inland waterways? Amongst them, which can be detected by exploiting AIS data?*
- *What attributes of AIS data are relevant to identify low water level impacts on inland waterways?*
- *Which kinds of anomalies happened on the Waal near Nijmegen during low water levels?*
- *What are the differences in analysis results between using AIS data and former reports that did not use AIS data?*

1.5 Scope

The study area is focused on the Waal from Lobith at the Dutch-German border to the bridge Tacitusbrug bij Ewijk (see Figure 1.1). The characteristics of certain canals and rivers described

above exclude areas like the port of Rotterdam and most of the waterways in the Netherlands. Furthermore, the Rhine is a river network that takes up the majority of inland waterways transport in Europe, it flows from Germany to the Netherlands and splits into the Waal and the Lek after entering the Dutch border (see Figure 1.2). The bend of the Waal at Nijmegen is known as a bottleneck for shipping due to three cross-river bridges and the uneven, shallow river bottom caused by the river bend dynamics (De deltacommissaris, 2020; De Gelderlander, 2016). Thus, the Waal near Nijmegen has the characteristics of a dangerous bottleneck and crucial inland waterway route, suitable for a study about low water level impacts. To control variables like travel purpose and vessel type characteristics, this study is focusing on cargo and tanker vessels.

The four study periods are 2016-10-01 to 2016-10-31, 2016-12-01 to 2016-12-20, 2017-10-01 to 2017-10-31, and 2017-12-01 to 2017-12-31. The choice of time periods is mainly because the data availability and the time limits of this thesis. Since Rijkswaterstaat needs about one week to process and anonymize one month of data, it would take 12 weeks for one year's data, which is unrealistic regarding the time schedule of this thesis. The water level above N.A.P. (Normaal Amsterdams Peil or the normal water level in Amsterdam) in December is usually normal or higher than autumn, but 2016 was unusual since the last time the water level was low happened in December 1959 (Transport Online, 2016). December 2017 has a high water-level comparing to 2016, with a difference up to 5 meters (see Figure 1.3). While water level in 2016-10 is similar to 2016-12, which can provide more insights about low water level impacts. Note that the water level above N.A.P. is not the depth of water, which is called fairway depth in navigation.

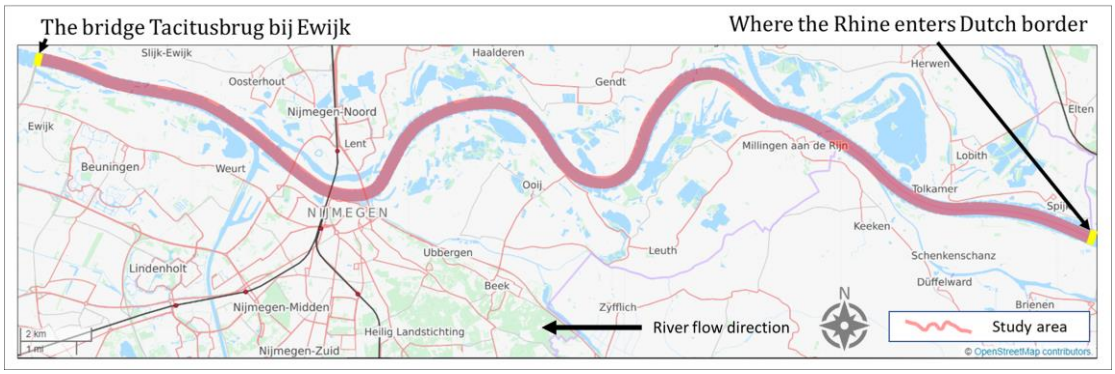


Figure 1.1 Study area. Reprinted from openstreetmap.org

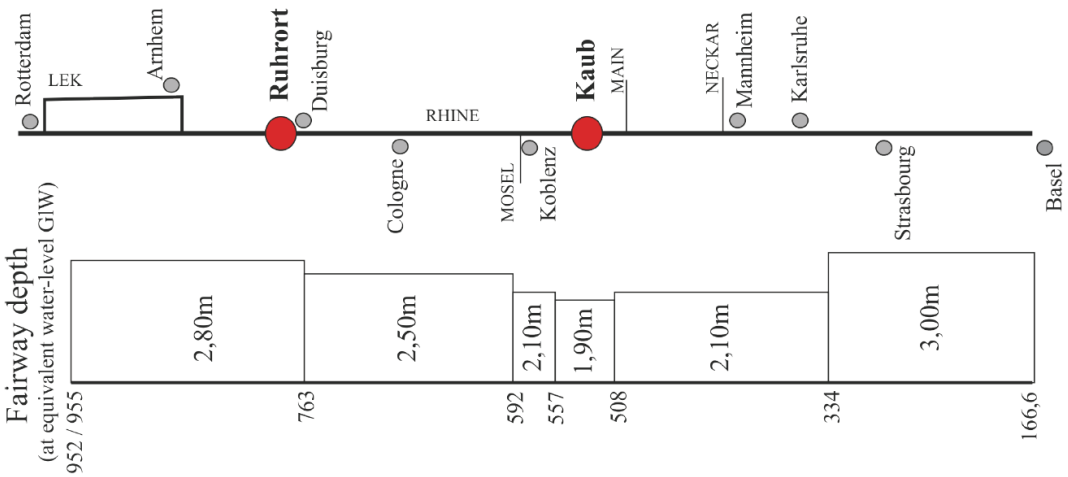


Figure 1.2 Fairway depth and bottlenecks (red dots) (Klein &Meißner, 2017)

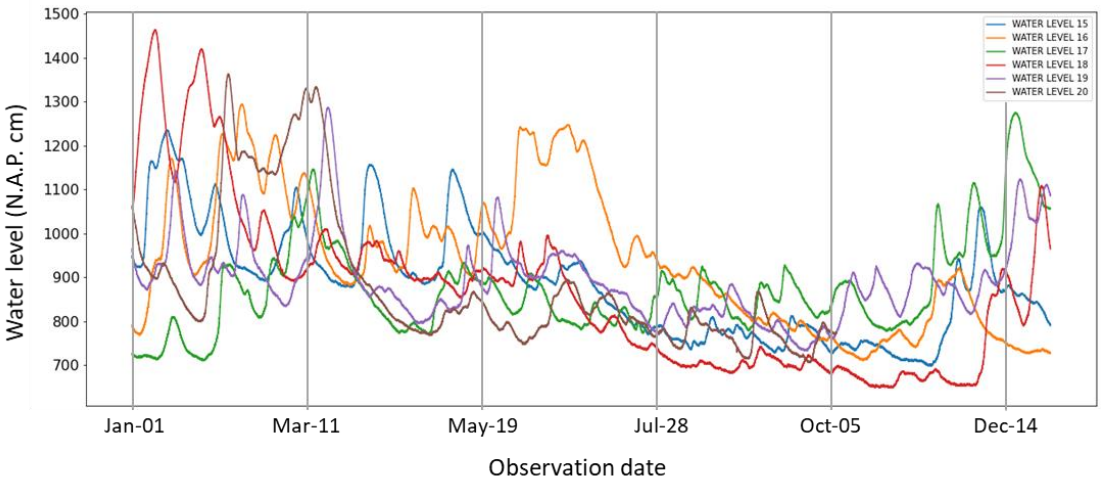


Figure 1.3 Water level N.A.P. between 2015 to 2020 at Lobith. Data retrieved from Rijkswaterstaat Waterinfo.

1.6 Thesis Outline

Chapter 2 reviews the related work to provide theoretical and practical knowledge for the thesis methodology. Research related to visualization of spatial-temporal data, analysis of moving objects, vessel interaction study, the nature of AIS data, and the influences of external factors are reviewed and discussed in this chapter. Chapter 3 explains the different stages of the research and the implementation of the methodology: clean and prepare the datasets, validate market observations by analyzing AIS elements that match the statements, and explore the outliers of vessel interactions. Chapter 4 presents the study results and concludes whether the market observations are true or not according to the AIS data analysis. Then Chapter 5 will answer the research questions by concluding on the research results presented in the previous chapter. Furthermore, this chapter discusses the relevance and possible future work directions.

2. Related Work

To conduct this research, there are several relevant subjects that need to be discussed. The following is an overview of the related works. Section 2.1 reviews how to extract and visualize moving objects. The review draws a big picture on movement data analysis and concludes a suitable method for this research. Section 2.2 focuses on methods that examine spatial-temporal data patterns. Sections 2.3 and 2.4 discuss factors that can represent or affect vessel behavior.

2.1 Visualization of Moving Objects

Trajectory data can be defined as a discrete time series of measured locations (Demšar et al., 2015). However, Graser (2019) considers the definition and analysis framework of movement data do not yet reach a consensus.

A common method to analyze massive movement data's trajectory is to use algorithms or train a model to learn the best way of grouping, aggregating, or clustering data (Sheng &Yin, 2018). Andrienko, Andrienko, and Wrobel (2007) introduce a framework to help analysts understand a movement dataset by visualizing massive movement data without changing its meaning or exceeding the human cognition limit. The framework is divided into 4 stages: data pre-processing, extraction of significant places, extraction of trips, and examination of trips. Andrienko and Andrienko (2011) present a method that consists of several algorithms to generalize and aggregate massive movement data. The study area is divided into several compartments and the trajectories are transformed and aggregated into groups of movements between source and destination compartments. Later, Andrienko and Andrienko (2013) present a survey of the state of the art visual analytics tools in massive movement data. They categorize related works into four types: looking at trajectories, looking inside trajectories, bird's-eye view on movement, investigating movement in context. The driving force of visual analytics is to exploit the massive volume of information and turn it into solutions or opportunities. The solutions are powerful because it combines the strengths of human and computer in pattern recognition and data processing. It has positive influences on knowledge improvement and decision-making (Keim et al., 2008).

Dodge (2015) categorizes movement research approaches into two fundamental areas: understanding movement processes, and modeling movement processes (see Figure 2.1). By looking into different aspects of the movement observation, the knowledge gained can calibrate the modeling processes. G.Andrienko, Andrienko, Chen, Maciejewski, and Zhao (2017) review recent technology development on visual analytics and visualization. They summarize relevant works in movement data and present the transformation relationship between different types of spatial-temporal data (see Figure 2.2).

Due to the technology development on data storage and computation, the capacity of analyzing massive datasets need to increase as well. Fiorini, Capata, and Bloisi (2016) proposed a workflow that uses only open source software to visualize massive movement data (90 million records). Graser, Widhalm, and Dragaschnig (2020b) present an incremental grid-based clustering algorithm called M^3 that can extract massive movement data (around 4 billion records) trajectory. Based on the M^3, Graser, Widhalm, and Dragaschnig (2020a) continue with the development of an algorithm that computes flows from massive movement datasets (350 million). It is worth noticing that Graser (2019) presents the work of MovingPandas, a Python library, as a solution for movement data analysis. Since MovingPandas is relatively new, not many have employed it in their studies. On the other hand, one paper uses Python libraries scikit-Mobility and MovingPandas to preprocess AIS data and to reconstruct the vessel trajectory to discover and explain the navigation situation before and after Covid-19 lockdown around the Venice Lagoon and its neighboring area

(Depellegrin, Bastianini, Fadini, &Menegon, 2020). Note that scikit-Mobility is aiming at the analysis of human mobility and flows, this research presents a possibility of application in maritime vessels movement (Pappalardo, Simini, Barlacchi, &Pellungrini, 2019).

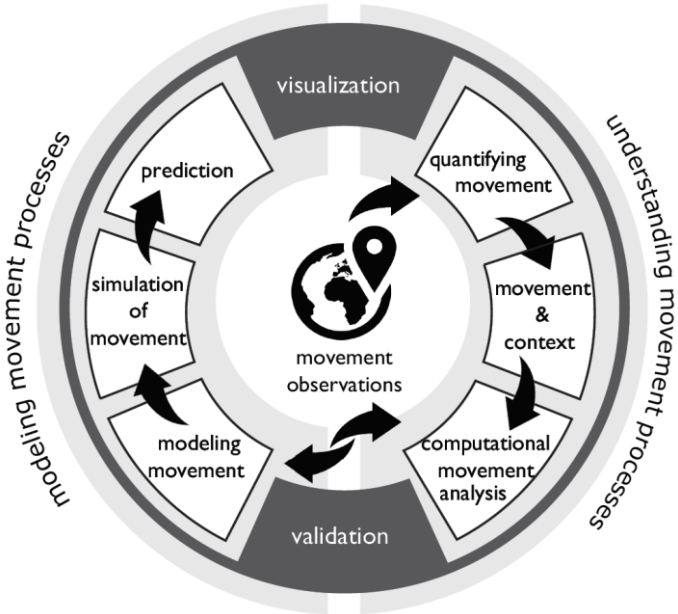


Figure 2.1 Movement research continuum (Dodge, 2015)

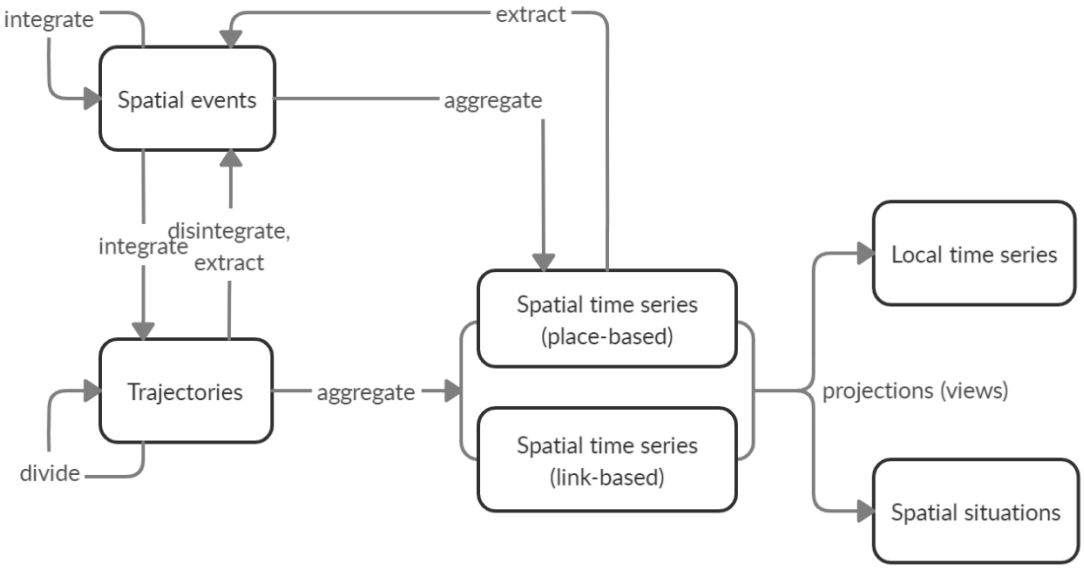


Figure 2.2 Movement data forms and the transformations between them (Andrienko, Andrienko, &Fuchs, 2016)

2.2 Spatial and temporal analysis of moving objects

The AIS data are point data, but the track of a ship sailed is line data. The differences between original data type and the best way to represent moving objects or events affect the choice of analysis method.

The first type of method to analyze vessel behavior patterns is by studying the AIS data points' temporal and spatial distribution. Each point contains static and dynamic information of the vessel

so by analyzing different attributes distribution in space or time, researchers can discover different patterns. According to Xiao, Ligteringen, VanGulijk, and Ale (2015), it is reasonable to use straightforward statistics to describe or analyze vessel behaviors. Meanwhile, Lee, Son, Lee, and Cho (2020) perform kernel density estimation on AIS data points to form a raster image of frequent-traveled path and then extract path boundary by applying image processing technique on the raster. The kernel density map shows an estimation of probability density function. The probability of a random point falls within a range of X and Y should be the integral under the surface (Smith, Goodchild, &Longley, 2018). It has a very similar function to histogram, but kernel density estimation provides a smooth estimation that is more valuable in 2-dimension data analysis (Węglarczyk, 2018).

The second type of method is by analyzing the trajectory lines formed by AIS data points. By clustering the trajectories, researchers can see spatial or temporal patterns of different attributes. Morris and Trivedi (2009) evaluate various trajectory clustering methods, which are: direct, agglomerative, divisive, hybrid, graph, spectral clustering. The evaluation results indicate that when the trajectory data are unsampled, the choice of clustering methods is not important. Nevertheless, Yuan, Sun, Zhao, Li and Wang (2017) summarize moving object trajectory clustering methods into three categories: trajectory lines, cluster point data, cluster sub-trajectories. Furthermore, they conclude that different clustering methods may be suitable in different situations depend on the research goals. But many studies are focusing on massive, large scale AIS dataset and finding their spatial or temporal patterns (Lee, Han &Whang, 2007; Rong, Teixeira &Guedes Soares, 2020; Yan et al., 2020), which is very different from the research scope of this thesis.

2.3 Vessels interaction analysis

The above sections reviewed analysis and visualization of moving objects' distribution, but the events representing interactions between moving objects should also be included. Many vessel interaction studies are focusing on the maritime collision risk assessment, which can be approached from various aspects like accident reports (Fan, Blanco-Davis, Yang, Zhang, &Yan, 2020; Jiang, Lu, Yang, &Li, 2020; Wang &Yin, 2020), historical AIS data analysis (Altan &Otay, 2018; Cucinotta, Guglielmino, &Sfravara, 2017; Rong et al., 2020) or machine learning and behavior prediction (Boztepe, 2019; Pallotta, Vespe, &Bryan, 2013). Amongst them, the method that uses historical AIS data to detect vessel encounters and domain violations can provide the knowledge of identifying vessels encounters in inland waterways. The definition of an encounter is when the ship domain around a ship is entered by another ship, while the ship domain consists of a safety area around a ship. The shape of a ship domain can vary between regions or different theories, from the oval, ellipse, to circle (Maritime Research Institute Netherlands, 2012). So, a ship domain violation detection algorithm can detect and flag encounter types by the position and movement of ships (Nordkvist, 2018). Arntsen (2019) verifies and improves Nordkvist's algorithm later. The algorithm takes processed data with a certain timestamp interval, then finds other ships that entered one's ship domain and classified the encounter types by the angle α to its sailing course, see Figure 2.3. Ship B will be classified as starboard side encounter, while ship C might be a head-on encounter if it enters ship A's domain, and ship E will be considered as met by the port side because it is at the back-left side of ship A, and if ship F stays at the current course it will be classified as port side encounter. However, note that this algorithm is considered as a brute-force method that has a high demand for computational power (Grossmann, 2019).

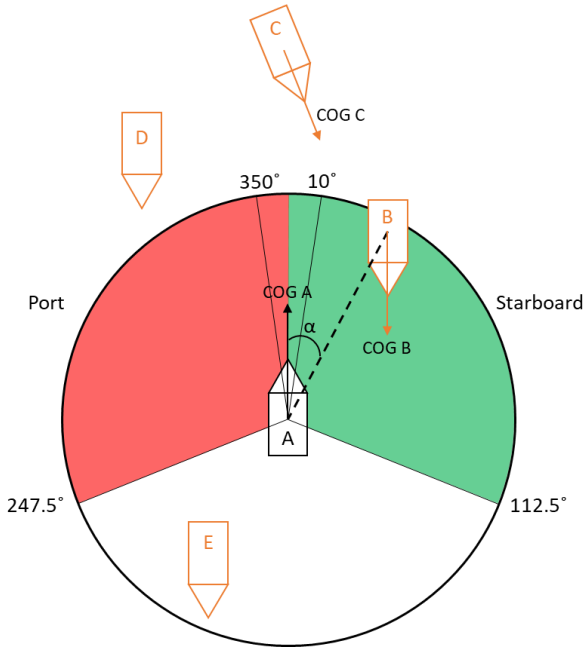


Figure 2.3 Diagram of ship encounters. Adapted from Arntsen (2019).

2.4 AIS data

Vessel conditions shown in AIS data can be divided into three categories: static, dynamic, and voyage-related data. Static and voyage related information is broadcasted every 6 minutes. Dynamic information is broadcasted every 3 minutes when a ship has stopped and between 2 and 10 seconds when a ship is sailing (CCNR, 2013). For navigation on the Rhine, it is mandatory to transmit the information marked with a star in Table 2.1 (CCNR, 2015). Static information is entered manually during AIS device installation. The dynamic information is updated automatically by the AIS device, while the voyage-related information is updated manually by the officer onboard (Zhou, Daamen, Vellinga &Hoogendoorn, 2019).

Table 2.1 AIS information types (CCNR, 2015; Zhou et al., 2019)

Elements	**Description**
Static	
MMSI*	"Maritime Mobile Service Identity", nine-digit number. User identity, that can be changed in case such as changing owner.
vessel name*	
vessel type*	Two digits, first digit represents general category of the vessel, while the second digit shows special sub-category.
ENI/IMO*	"unique European vessel Identification Number", or for seagoing vessels that have not been given an ENI number, the IMO number. Seven-digit number. Cannot be changed in any case.
length*	Overall length of the vessel or convoy, accurate to 0.1 m.
breadth*	Overall breadth of the vessel or convoy, accurate to 0.1 m.
tobow	In meter.
tostren	In meter.
toport	In meter.
tostarboard	In meter.
callsign	Radio call sign.
Dynamic	
position*	WGS 84 coordinate system.

Elements	Description
SOG*	Speed over ground.
COG*	Course over ground.
ROT*	Rate over turn.
timestamp*	Time (UTC) of the receiver or electronic location device.
heading	
headingValid	
navigational status	If available, e.g. 'Engaged in fishing', 'Under way using engine'
Voyage-related	
draught	In meter.
destination	
ETA at lock/bridge terminal	Estimated time for arrival.
hazardouscargo	Whether vessel is carrying hazardous cargo.
nationality	

2.5 Influence of external factors

External factors on restricted waterways can also bring influences on vessel traffic and safety. Shu, Daamen, Ligteringen and Hoogendoorn (2017) focus on vessel behavior around the port of Rotterdam by analyzing AIS data and external, environmental factors, including other vessels' behavior, fairway, visibility, wind, and current. The study found that wind, visibility, current, and encounters have significant impacts on the vessel speed and relative distance to starboard bank, while the current and encounters of other vessels have effects on vessel course. Moreover, the authors found that the vessels change behavior when encountering other vessels. A case study focusing on the narrow waterways of Sabine-Neches area (located southeast of Texas, USA) uses vessel domain to detect encountering by analyzing historic AIS data (Wu, Rahman, &Zaloom, 2018). However, the study results found out that the narrow waterways are not as congested as expected, and the authors conclude that this may be due to the piloting and navigation control posted by the waterways management system. Note that these studies focus on port or estuary areas, which have different nature than the thesis study area, so factors should be examined carefully before making conclusions.

3. METHODOLOGY

This chapter provides structural method and steps to achieve the research goals. The research methodology was divided into three phases in the order of steps: section 3.3 prepared the AIS dataset by cleaning and checking reliability, section 3.4 validated the market observations collected from news articles or reports, while section 3.5 identified anomalies of ship encounters. The analysis workflows were modified from the study of Andrienko and Andrienko (2011, 2013) and Graser (2020). Their workflow designs can be summarized into four stages: establish the overview of the dataset and its environment, generate trajectory or event objects from point data, explore the trajectories or events, analyze the outliers.

3.1 Data

The data used in this thesis is an anonymized historical recording of inland AIS provided by Rijkswaterstaat (see Table 3.1). The periods are 2016-10-01 to 2016-10-31, 2016-12-01 to 2016-12-20, 2017-10-01 to 2017-10-31 and 2017-12-01 to 2017-12-31. Figure 3.1 presents an overview of the AIS records of cargo and tanker vessels in the first period.

Table 3.1 List of datasets

	Availability	Description
AIS data during October 2018	×	RWS does not have AIS data between July and December 2018.
AIS data during December 2016	△	RWS only has 1st to 20th December AIS data.
AIS data during December 2017	○	RWS has the data.
AIS data during October 2016	○	RWS has the data.
AIS data during October 2017	○	RWS has the data.
Water level during research period	○	Data retrieved from RWS Waterinfo.

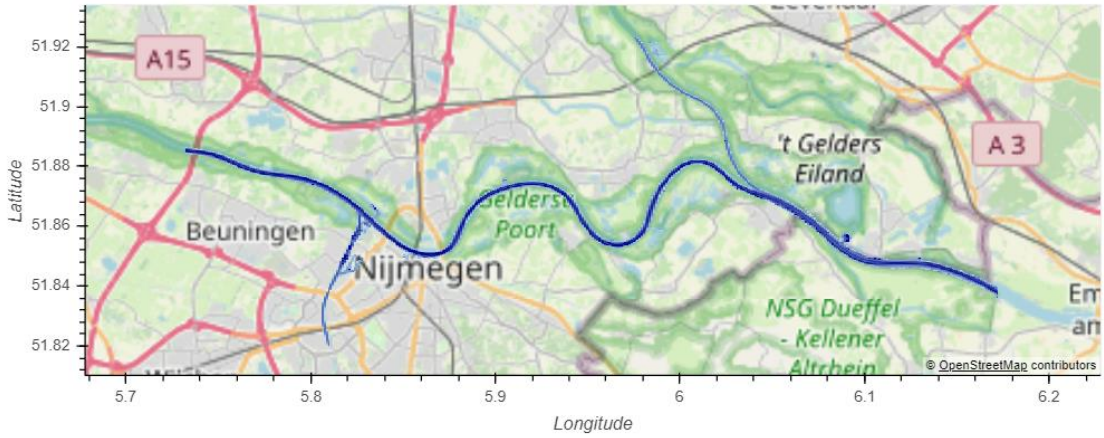


Figure 3.1 Overview of cargo and tanker vessels during December 2016

To clean and gain a basic understanding of the dataset reliability, it is necessary to exclude data that lacks essential information. Vessel type 0 is the default code for not available, and 1 to 19 are reserved for future use. Thus, it is not possible to classify vessels with vessel type codes below 20. We removed records without speed or vessel type, which results in datasets became about one third of the original size (see Figure 3.2). Secondly, since this study only focuses on cargo and tanker vessels, other types of vessels and their proportions are shown in Figure 3.3 and Figure 3.4 to provide an overview of the original dataset. For the table format of these data, see Appendix I.

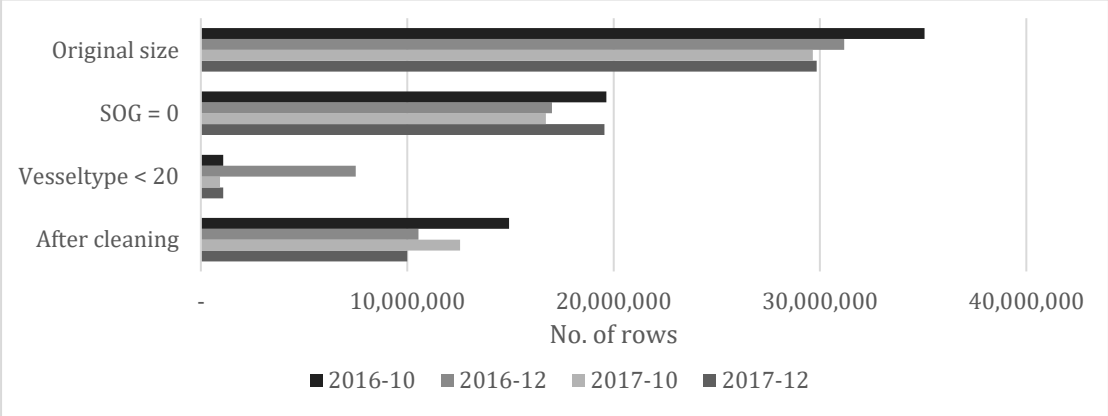


Figure 3.2 Excluded data

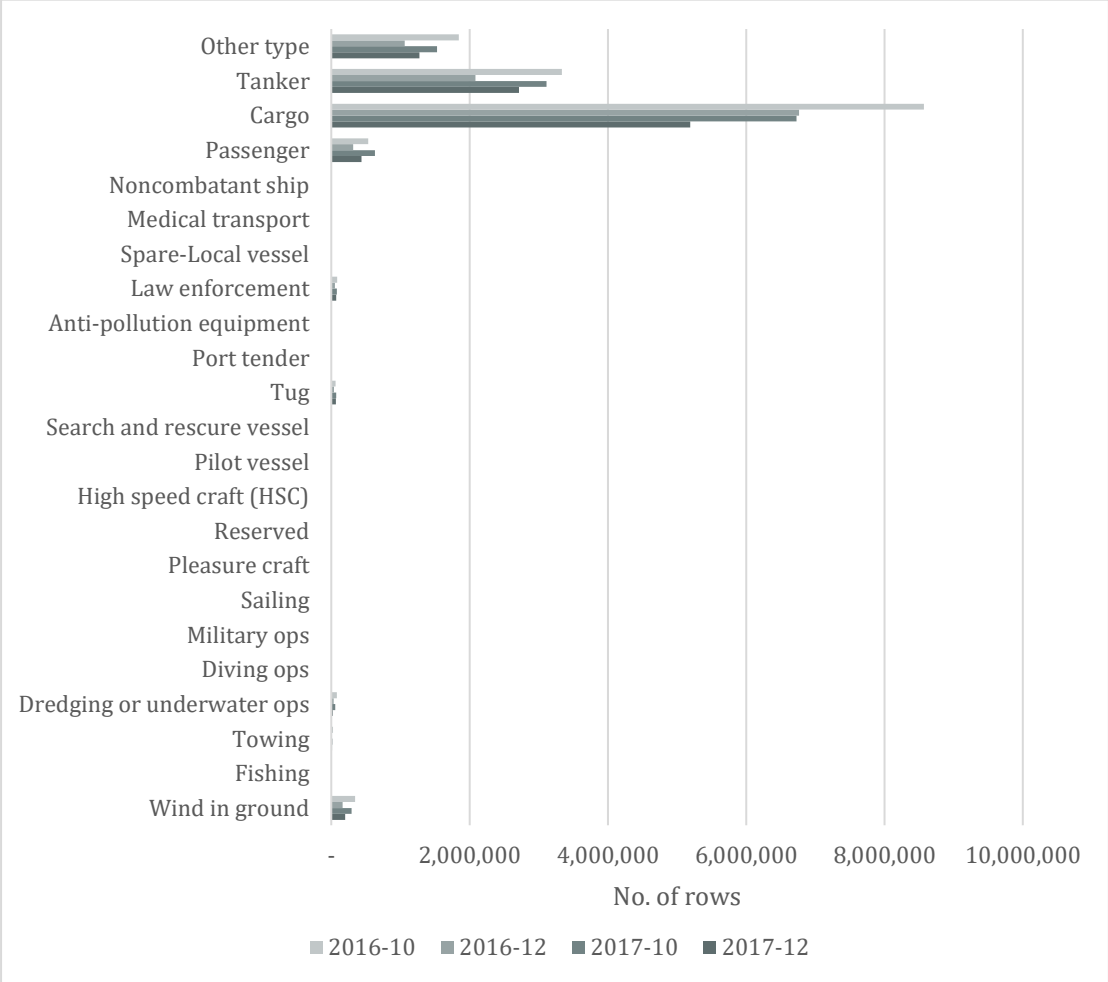


Figure 3.3 Number of each vessel type

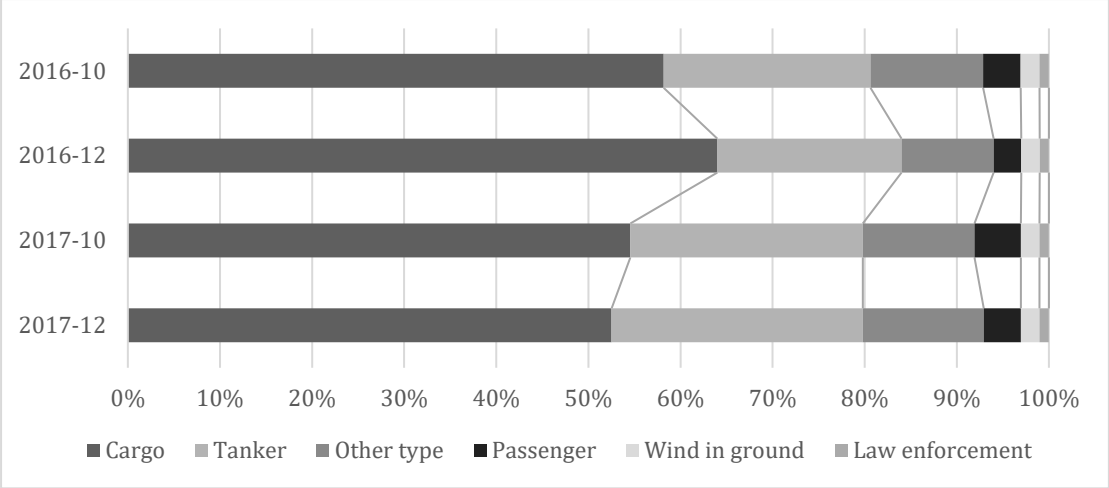


Figure 3.4 Percentage of major vessel types

3.2 Toolsets

The analysis was mainly conducted by using Python, its libraries and modules. Due to the privacy policy of Rijkswaterstaat, the AIS data are not allowed to be saved outside the Azure cloud storage, thus, the management of data was also carried out within the Jupyter Notebook environment. The toolsets used are listed in Table 3.2.

Table 3.2 List of tools

	Version	Description
Azure	-	The cloud environment that stores the AIS data provided by RWS. It is not allowed to export and save AIS data to local disk.
Anaconda	4.9.1	To load data temporarily to computer memory and conduct computation and analysis.
Python	3.7.7	
Jupyter Notebook	1.0.0	

The main Python libraries and modules that were employed to conduct analysis and visualization are listed in Table 3.3.

Table 3.3 Employed Python libraries and modules

	Version	Description
Pandas	1.1.4	A data analysis and manipulation tool.
GeoPandas	0.8.1	Extend the datatypes used by Pandas to allow spatial operations on geometric types.
hvPlot	0.6.0	A high-level plotting API that allows dynamic plotting and interactive figure display.
HoloViews	1.9.5	An open-source Python library for data analysis and visualization.
Plotly Express	0.4.1	A terse, consistent, high-level API for creating figures.
Shapely	1.7.1	Manipulation and analysis of geometric objects in the Cartesian plane. Note that AIS data are using a different coordinate system.
GeographicLib	1.50	Provides tools for solving the direct and inverse geodesic problems for an ellipsoid of revolution.

3.3 Data reliability and preparation

3.3.1 Conceptual model

AIS is widely considered as a tool that can improve navigation safety, but a machine that partially depends on human operation cannot prevent the occurrence of error. AIS data consist of automatically generated values and voyage information that relies on updating. Harati-mokhtari, Aharati-mokhtariljmuacuk, Wall, Brooks, and Wang (2004) investigate AIS data reliability and found that due to machine and human errors, AIS data is not reliable enough for skippers to entirely trust the device. For instance, the MMSI number might be incorrect due to wrong installation or equipment faults. These types of abnormal value need to be excluded during the dataset preparation stage (see Figure 3.5).

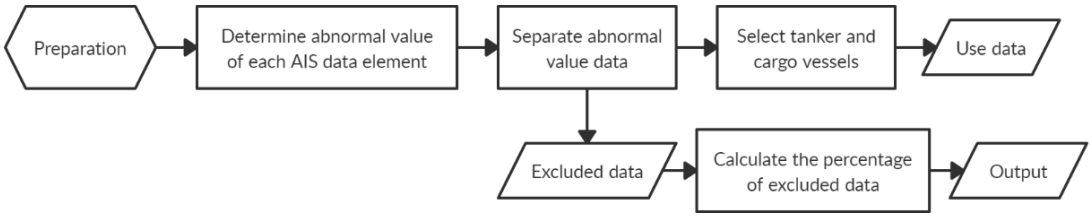


Figure 3.5 Preparation stage flowchart

3.3.2 Implementation

For both machine and human error or faults, it is possible that there are some abnormal values in the research datasets. According to Harati-mokhtari et al. (2004) and Zhou et al. (2019), common types of abnormal values that can happen in the research area are:

1. The timestamp of dynamic information has a maximum 3-minute interval, while static and voyage related information will be broadcasted in every 6 minutes.
2. MMSI must be a nine-digit number, and 999999999 does not exist.
3. Speed over ground should be smaller than 35, cargo and tanker vessels rarely exceed it.
4. Course over ground and heading should be between 0 and 360.
5. The width should be smaller than the length.
6. The draught should not exceed the fairway depth.

Note that the vessel identification numbers (e.g., MMSI, ENI, or IMO) were anonymized by RWS, thus the rule of numbering does not apply here, but instead the reliability of anonymization was examined. This was conducted by checking the outliers of each attribute and whether the distribution is possible or not. Then, we found out that the vessel ID is not unique as it should, around 10% of anonymized IDs are shared by more than one vessel. To avoid potential errors, new unique vessel IDs were assigned according to the anonymized ID, vessel size, and type. Moreover, we found out that at the timestamp 2016-12-20 18:19:56, there are around 1.6 million duplicate records. The duplicated 1.6 million records were removed from the dataset in order to prevent it from affecting the statistics of the dataset. The abnormal value counts are listed in Table 3.4, while Figure 3.6 shows after cleaning the datasets, how many AIS records a ship has sent and the distribution of record numbers. Namely, the chart tells how active or how long most ships staying within the study area.

Table 3.4 Abnormal values

	2016-10-01 to 2016-10-31	2016-12-01 to 2016-12-20	2017-10-01 to 2017-10-31	2017-12-01 to 2017-12-31
Original data counts				
Rows	11,907,144	8,848,795	9,841,820	7,907,469
Ship IDs	2,145	1,929	2,079	1,975
Abnormal data counts				
Ship IDs	256	231	233	238
SOG (rows)	672	6,543	915	432
COG (rows)	0	0	0	0
Width, length (ships)	0	0	1	0
Draught (ships)	61	40	38	40

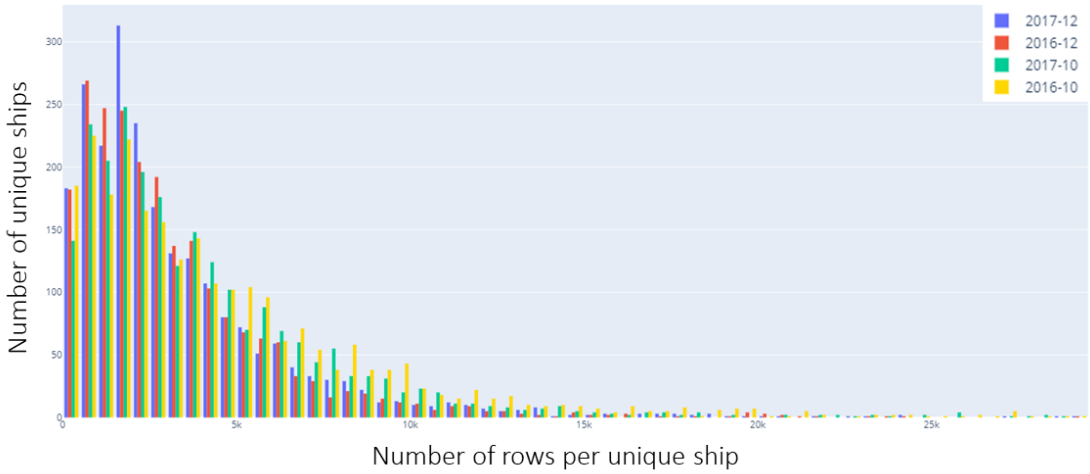


Figure 3.6 Distribution of AIS row numbers in each unique ship

3.4 Method to market observation validation

3.4.1 Conceptual model

This part is to check whether the observations match historic AIS data, and to what extent can those events being visualized or presented. According to Xiao, Ligteringen, VanGulijk and Ale (2015), it is reasonable to describe vessel behaviors by statistics distribution of the AIS data. Thus, the sub-questions were derived from statements made by reports or news articles about inland waterway transport activities, and then turned into questions that can be answered by AIS data. Next, trajectory IDs need to be assigned at the preparation stage in order to carry out the following steps of statement validation. The workflow and required data are shown in Figure 3.7.

Statement 1: ships have less cargo capacity (news articles).

→ *What are the average draught numbers in low water level and normal times?*

Statement 2: the waterway is busier than usual (news articles).

→ *Are there more ships on the waterway at the same time?*

→ *Does a ship sail more trips than usual?*

→ *Is the total number of operating vessels higher or lower?*

Statement 3: ships reduce speed to minimize the dynamic draft (Klein &Meißner, 2017).

→ *Is the average speed slower than usual?*

→ *Which area of the waterway do ships sail slower?*

Statement 4: smaller vessels shift operation area to the Rhine (CCNR, 2019a).

→ *Does the vessel size differ from normal times?*

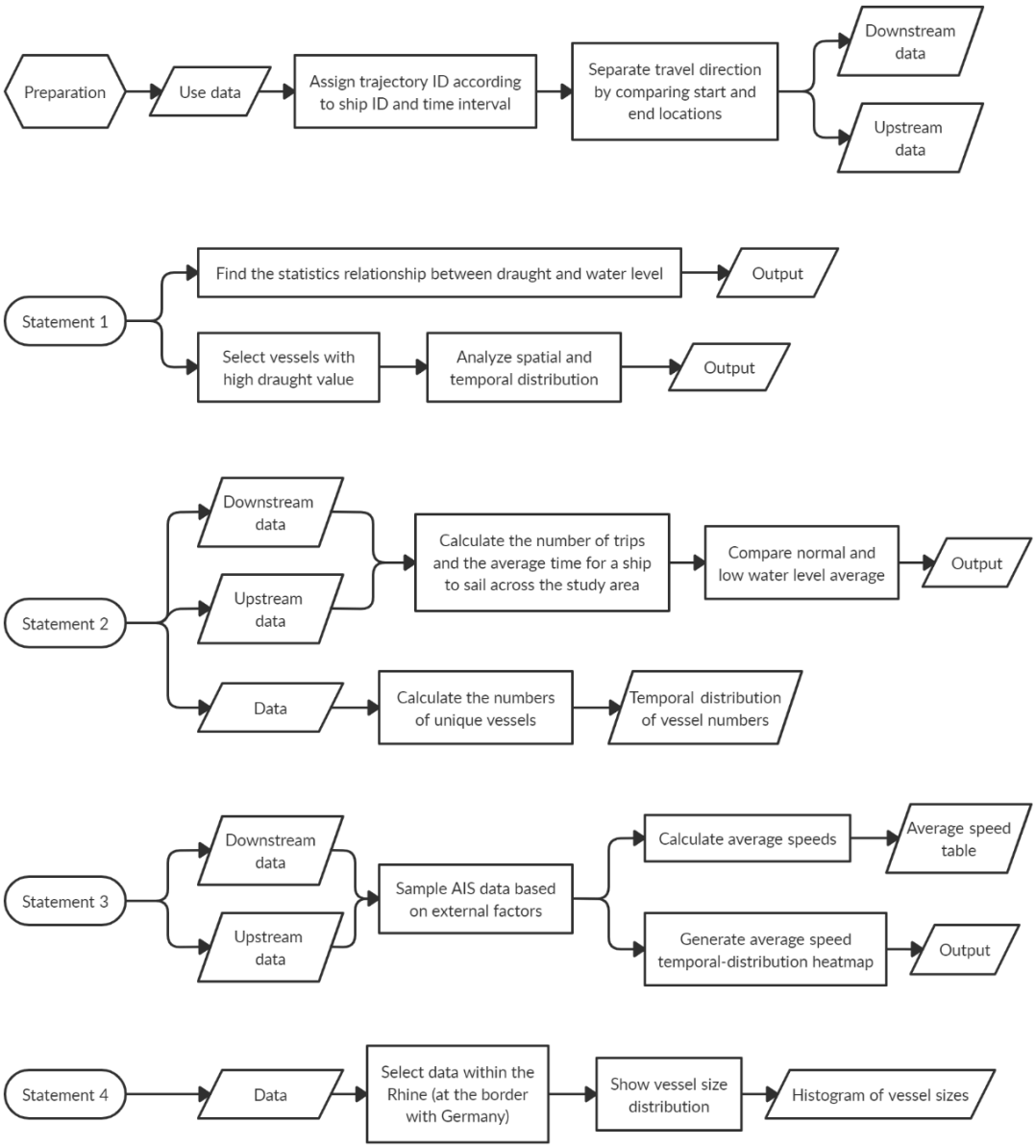


Figure 3.7 Market observation validation flowchart

3.4.2 Implementation

Before statement 1, rows were removed if the speed is smaller than 1 knot so idling or mooring vessels can be excluded, then trajectory ID were assigned to make later analysis possible. For each record of a unique ship ID, a trajectory ID was assigned to this row of record if this record and the next record are within 10 minutes. If not, the next record will be assigned a new ID number, and the procedure continues. After assigning trajectory ID, the direction of each trajectory was determined by comparing the start and end locations. Since this study area is relatively simple, the longitude of the start and end locations were used to tell if a ship was going west or going east.

It is worth noticing that another environmental factor also can have influence on inland waterway transports. Figure 3.8 and Figure 3.9 are water level data observed outside Lobith and the Nijmegen Port, while Figure 3.10 show that surface water flow rate outside Lobith. All three figures have almost the same shape throughout the year. Thus, if an element shows a pattern that matches the water levels, the influence of surface water flow rate cannot be eliminated. Furthermore, to make it easier to understand the water level data, Figure 3.11 and Figure 3.12 presents the underwater morphology and indicates that the river floor outside Lobith and the Nijmegen Port observation sites are 3.3 and 1.5 meter above N.A.P. Then, the fairway depth at these two sites can be calculated by having the water level values minus the river floor height above N.A.P.

Figure 3.8 Water level observed at Lobith. Data retrieved from Rijkswaterstaat Waterinfo.

Figure 3.9 Water level observed at Nijmegen Port. Data retrieved from Rijkswaterstaat Waterinfo.

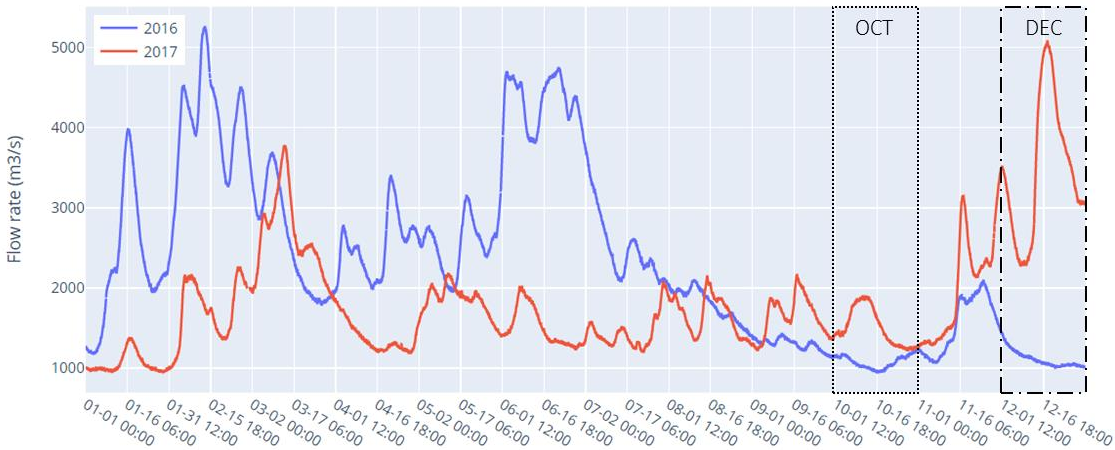


Figure 3.10 Surface water flow rate observed at Lobith. Data retrieved from Rijkswaterstaat Waterinfo.

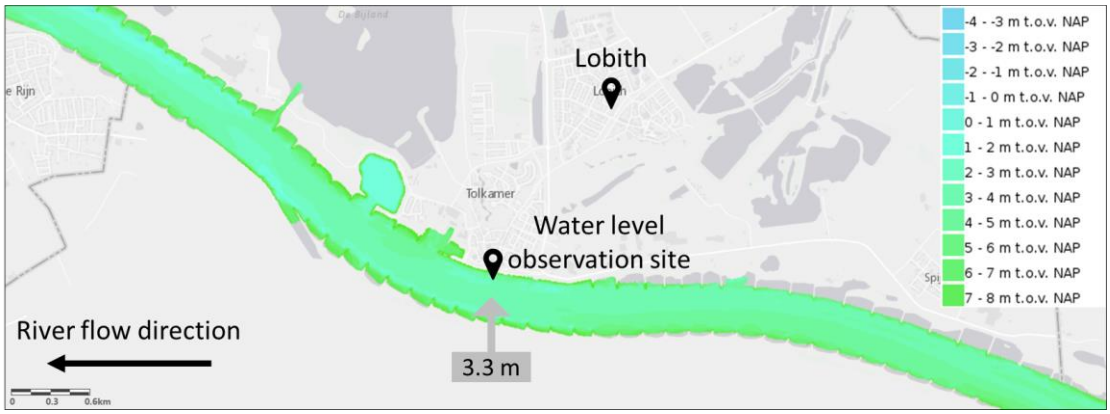


Figure 3.11 Bathymetric maps of Lobith. Reprinted from Rijkswaterstaat (2020a).

Figure 3.12 Bathymetric maps of Nijmegen Port. Reprinted from Rijkswaterstaat (2020a).

Statement 1
"ships have less cargo capacity when the water level is low"

The statement claims that cargo and tanker vessels carry less weight during low water levels, indicating that the vessel draught is affected by water levels. To validate this statement, the water level and flow rate data were analyzed with the vessel draught.

First, water level and flow rate data of 2016 and 2017 were selected and cleaned. Then line plots were created to display October and December in separate diagrams. Next, the vessel traveled through outside of Lobith were selected and only kept draught value of each trip to remove duplicate data, then the average draught was calculated in every 10 minutes to match the time interval of water level observations. However, the overview of the datasets shows no visible correlation between draught value and water level, see Figure 3.13 and Figure 3.14.

Figure 3.13 Comparsion of water level and draught outside Lobith (October)

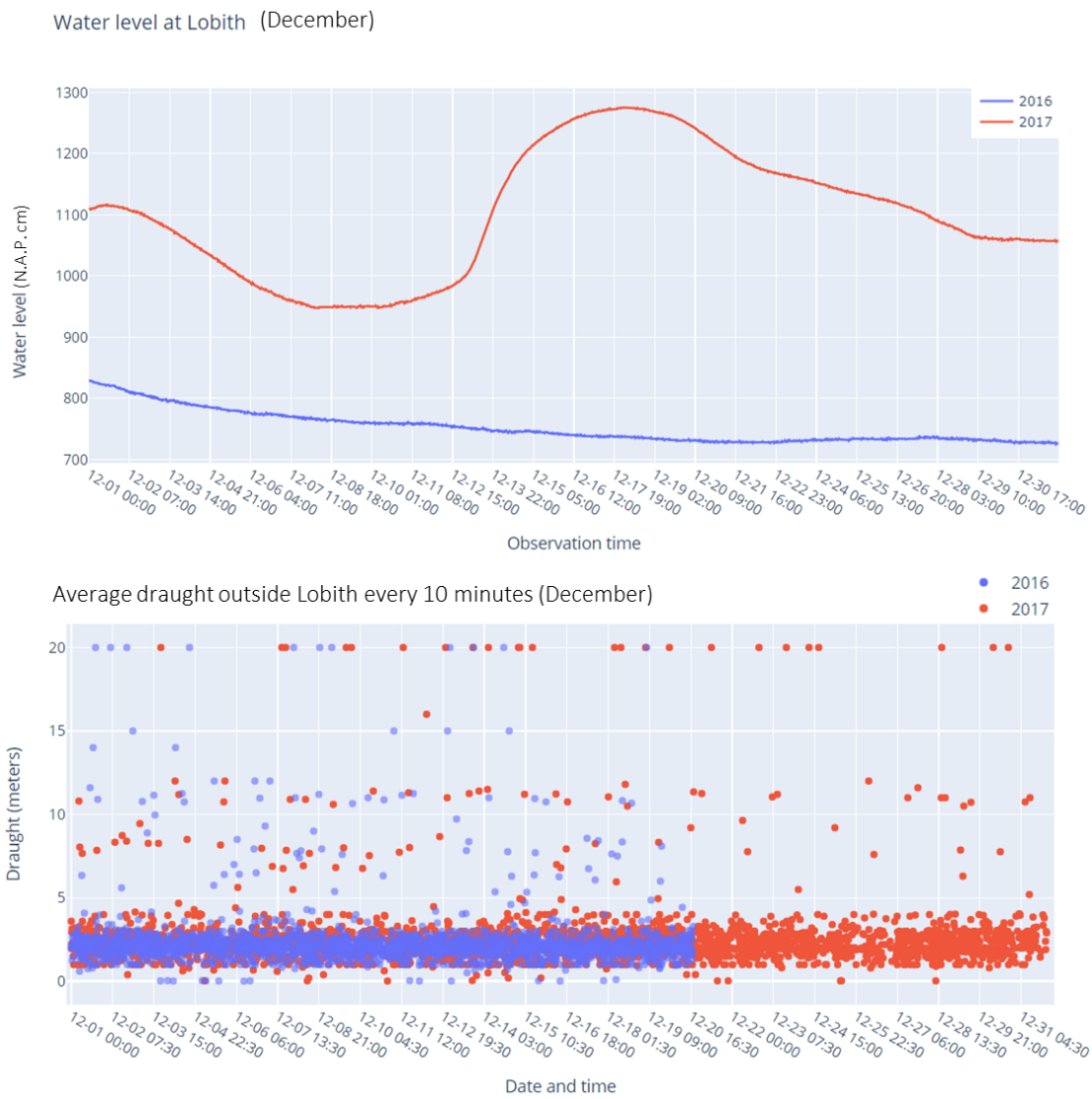


Figure 3.14 Comparsion of water level and draught outside Lobith (December)

According to the Lobith water level data, the highest water level is 12.76 meters above N.A.P. at 2017-12-18, while the lowest is 7.11 meters above N.A.P. at 2016-10-17. So, the highest and lowest fairway depth outside Lobith are 9.5 and 3.8 meters. Then, Figure 3.9 and Figure 3.12 present the water level and river floor height observed outside Nijmegen Port. The highest and lowest fairway depth outside of Lobith and Nijmegen Port are 8.9 and 3.4 meters at 2017-12-18 and 2016-10-17 respectively. Note that although the river floor is not even, deeper parts of the fairway may have close to 2 meters height difference, many parts of the fairway have flat river floor and ships cannot avoid them since the fairway is only 150 meter wide. Thus, the draught values above 5 meters inside the study area are extreme values considering the water levels in 2016 and 2017. In order to gain further understanding, the extreme draught values were extracted and analyzed together with the waterway conditions.

Statement 2
"the waterway is busier than usual"

Before responding to this statement, we need to understand the limitation brought by the AIS device. The AIS device transmits a signal containing its position every few seconds, which means counting vessels every second is not an appropriate way to find out how many ships were inside the study area at the same time. Furthermore, "busier" can be interpreted in different ways such as the number of ships or trips.

First, only the AIS data within the Waal were selected to exclude the variable of lock waiting time or ships turning to the Pannerdensch Kanaal. To minimize the influence from moored ships, AIS records that have speed smaller than 1 knot were also removed. Three 5-day time windows were decided for October and December based on water levels to represent the period of big water level difference, see Figure 3.13 and Figure 3.14 for water level comparisons. For October it is between 11th to 15th, while December has two time-windows which are 1st to 5th and 15th to 19th (see Figure 3.15). Another reason for using the time windows is to compare the statistics with the same length of time since 2016-12 dataset does not include data after 2016-12-20. Then we calculated the number of trips by counting the number of trajectory IDs in downstream and upstream categories.

October 2016							October 2017						
Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat
						1	1	2	3	4	5	6	7
2	3	4	5	6	7	8	8	9	10	11	12	13	14
9	10	11	12	13	14	15	15	16	17	18	19	20	21
16	17	18	19	20	21	22	22	23	24	25	26	27	28
23	24	25	26	27	28	29	29	30	31				
30	31												
December 2016							**December 2017**						
Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat
				1	2	3						1	2
4	5	6	7	8	9	10	3	4	5	6	7	8	9
11	12	13	14	15	16	17	10	11	12	13	14	15	16
18	19	20	21	22	23	24	17	18	19	20	21	22	23
25	26	27	28	29	30	31	24	25	26	27	28	29	30
							31						

Figure 3.15 Calendar of the study periods

Next, the average time for a vessel to sail across the Waal within the research area was needed because it is essential for determining time unit of counting ships. Thus, the first and last positions of a trajectory that have come close to the west and east side study area boundary were selected and then calculated the travel time from the position's timestamp. Then the average travel time was calculated each category. The shortest average travel time is 1 hour 44 minutes during 2017-12-15 to 2017-12-19 upstream, so the time unit for counting ships within the Waal was set to be one hour.

To answer the question of how many ships were inside the study area at the same time, the number of newly assigned ship IDs was counted every hour. The reason for using one hour as a unit is because the minimum average time is 1 hour 44 minutes, which means it is less likely to miss counting a ship if the time interval is one hour. Also, it is simpler to analyze the data for the researcher, and easier to understand the statistics for the reader.

Statement 3
"ships reduce speed to minimize the dynamic draft"

The statement claims that when the water level is so low, it is even necessary to reduce dynamic draft in order to keep the ship safe. A series of heatmaps were generated to present the temporal distribution of speed changes. Heatmaps were chosen because it can display value differences in two-dimensional diagram. To reduce the influences of various external factors such as turning to safe haven or mooring, six small areas were decided to present different conditions (see Figure 3.16):

A. Outside Vluchthaven Tolkamer: outside a safe haven, and on the Dutch-German border.
B. Outside Erlecom: river bend where downstream lane is on the slow flow side.
C. Outside Groenlanden: river bend where downstream lane is on the fast flow side.
D. East of Waalbrug: upstream of the Nijmegen river bend and bridges area.
E. Under De Oversteek bridge: under a bridge and outside a port.
F. Outside Ewijkse Plaat: relatively straight waterway.

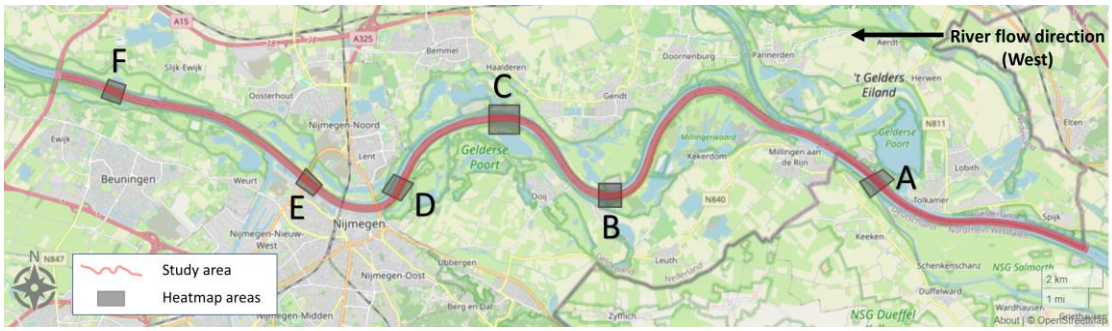


Figure 3.16 Heatmap areas

First, AIS data points were clipped by these six polygons separately, then the average speed was calculated in a 30-minute time interval. Also, a series of pixel-based average speed heatmaps were generated to show the speed distribution in every day of week (ZuchaoWang, Lu, Yuan, Zhang, &Wetering, 2013). The purpose of the week heatmaps is to see if there is a temporal pattern within every week. For instance, is the average speed in weekends faster or slower than weekdays? The x axis represents the time, while the y axis indicates the day of week. However, after removing the mooring ships within the datasets, the heatmaps show no visible pattern (see Appendix II). So, the y axis value was changed to date in order to find speed changes in detail. Then the heatmaps series were compared with water levels and holidays distribution at the time.

Statement 4
"smaller vessels shift operation area to the middle and upper Rhine"

The statement claims that because the fairway depth in the middle and upper Rhine is smaller and easier to be affected by the weather conditions. However, smaller ships usually have smaller draught without cargo, and during low water levels that can become an important factor. So, smaller ships in less affected regions tend to go to the middle and upper Rhine to take advantage of the high freight rate when the water level is low. The reasons of choosing this statement even though the study area is located at the lower Rhine are because this is where the Rhine flows into the Netherlands, and to investigate whether this behavior happens at the lower Rhine.

To limit the analysis on the Rhine, data along the Dutch-German border were selected. Then the data were reformed to show each trip's date, time, and the vessel's length and width. However, draught was not in the histograms due to its lacking reliability, as explained in statement 1. Note that the study area is actually part of the lower Rhine (see Figure 3.17), but the research was still carried out in order to test the capability of AIS data analysis.

Figure 3.17 Sections of the Rhine (Inland Navigation Europe, 2020)

The vessel length and width of every trip were first display as scatter plots which use a point to represent a vessel, see Figure 3.18 and Figure 3.19. However, the points are overlapping each other and hide the pattern of distribution. Thus, the vessel size data were then visualized into 2D histograms which can present the amount and intensity of groups in a contour shape. Then the diagrams were divided by year, month, and travel direction so reader can see and compare their differences at one glance.

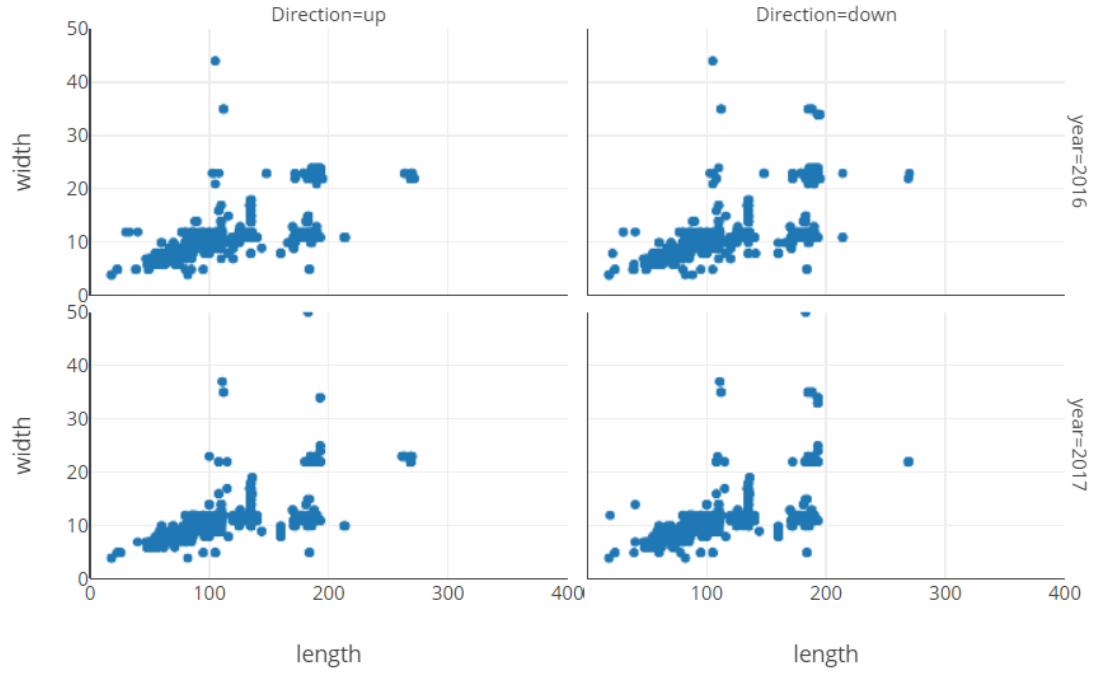


Figure 3.18 Scatter plots of ship size around the Rhine (October)

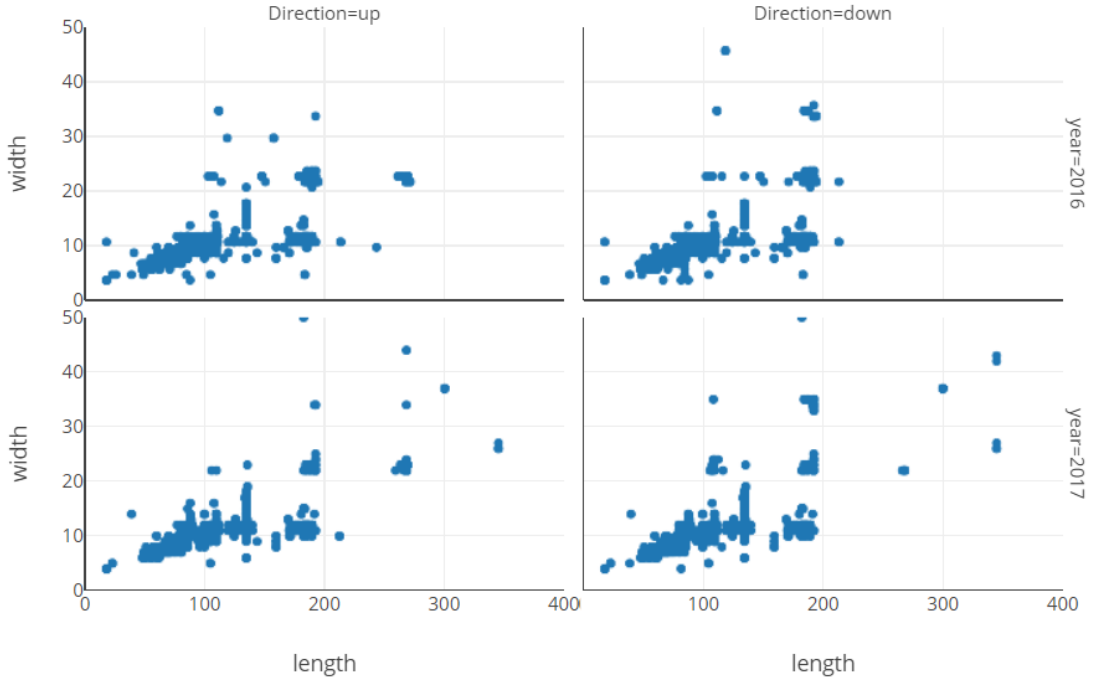


Figure 3.19 Scatter plots of ship size around the Rhine (December)

3.5 Method to encounter anomaly detection

3.5.1 Conceptual model

In the former stage, only the attributes of individual vessel are considered and discussed. While at this stage, we analyzed the interaction of vessels, specifically, the anomalous events of vessel encounters. The workflow and data are shown in Figure 3.20.

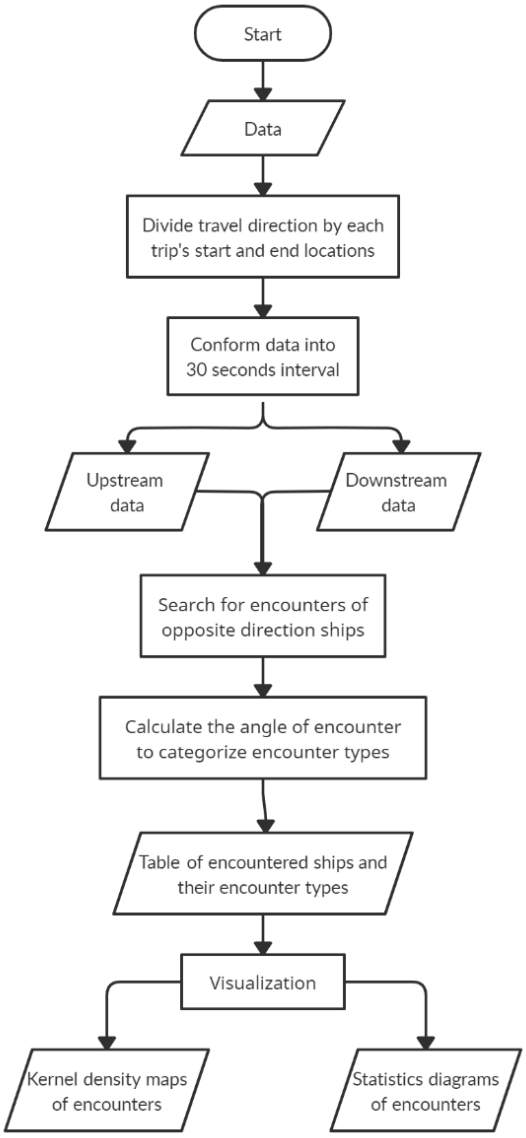


Figure 3.20 Ship encounter anomalies detection flowchart

Note that the navigation regulations varied in different waterways, it is important to investigate the relevant rules before defining what can be anomalous encounters. In this study area, there are two conflicting regulations for navigation on the Waal: The Inland Waterways Police Regulations (Binnenvaart Politieregelement, BPR) and the Rhine Navigation Police Regulations (RPR). The RPR ask small vessels (shorter than 20 m) to always give way to large vessels (longer than 20 m), but at the same time BPR states that vessels on the Waal must keep to starboard. The conflict can happen when a small vessel meets and gives way to large vessels, then the small vessel might need to sail to the port side. Since small vessel must give way to large vessel, the large vessel sail on port side needs to display a blue sign with a white blinking light in order to notify others (the Department of Waterways et al., n.d.). According to this document, large vessels passing a

waterway curve can sail on either side with the preferred current condition, while small vessels should stay on the right-hand side but change side if necessary.

3.5.2 Implementation

To eliminate the influence of locks, ports, and canals, only the AIS data within the Waal were used. Then the data were divided into upstream and downstream to exclude over-take events. Secondly, reduced the size of the datasets by conforming records into 30 seconds interval so to keep a certain precision for later analysis.

An algorithm adapted from Nordkvist (2018) ship domain violation detection model was applied for searching vessels that have met opposite direction vessels. In Nordvist's algorithm, ship domain is defined as an oval shape buffer from the center of the vessel, and true vessel center is calculated by the distance from GPS transmitter to bow, stern, port, and starboard. However, the distance information is not complete in the AIS datasets, and the nature of inland navigation makes the encounter event less complicated. As a result, we assume that the GPS position is the vessel center and used it to calculate the distance to other vessels at the same timestamp. Because the fairway width of the Waal is 150 meters (Rijkswaterstaat, n.d.-b), and the distance a ship could travel with maximum speed of 9 m/s in 30 seconds was 270 meters, we set 300 meters as the detecting radius considering the influence of river flow. Any ship that came closer than 300 meters are classified as encountering.

The calculation process first matched upstream trips with downstream trips by their timestamp. Then the distance between all paired points was calculated and the pair of upstream and downstream points that is closer than 300 meters was listed. The angle from course over ground (GOC) to the encountering ship was calculated to determine the side of the vessel encountering, see Figure 3.21. For the algorithm of detecting encounters, see Appendix III.

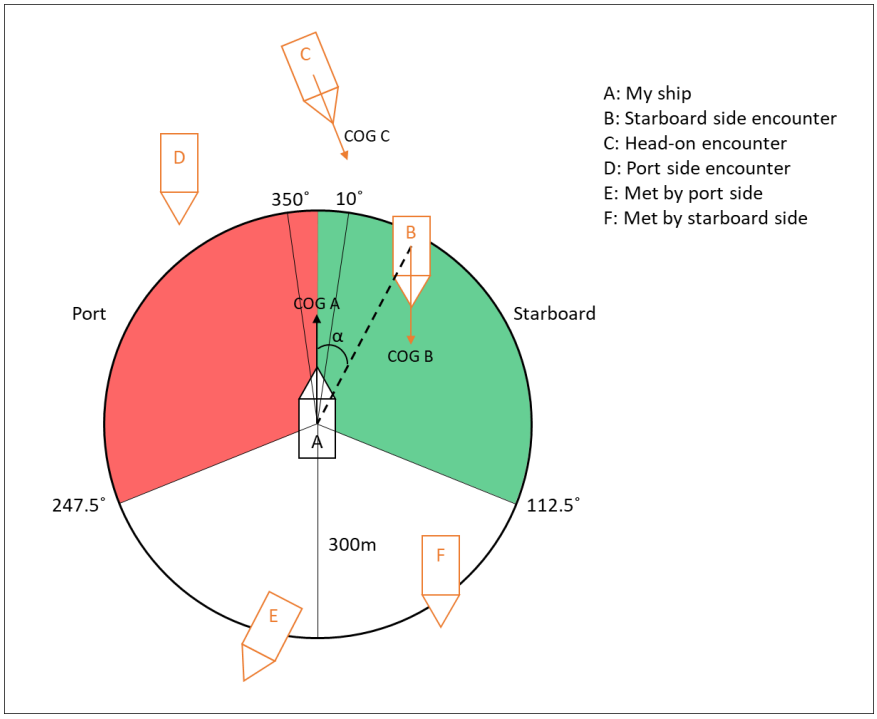


Figure 3.21 Types of encounter side. Adapted from Arntsen (2019).

The encounter events that happened on the starboard (right) side of the ship is considered as unusual. Even though it is allowed when small and large vessels meet on the Waal, this behavior can still increase the collision risk. Thus, the expected results are starboard side encounter hotspots and the statistic distribution of different encounters.

4. Results

This chapter analyzes the results gained and concludes the possibility of identifying low water level impacts from AIS data. From section 4.1 to 4.4, each statement and its results are explained and receive a verdict which concludes whether the statement is true or not. Then at section 4.5, the results of ship encounter anomalies are presented and explain what the possible relationship between water level and starboard side encounters could be.

4.1 Statement 1 – Lower cargo capacity

"ships have less cargo capacity when the water level is low"

It is worth noticing that the water levels in 2016 October and December were at similar levels, while there was a 5-meter water level difference between December of 2016 and 2017 (see Figure 4.1). This information is crucial for determining which external factors might have direct relationship with the impacts.

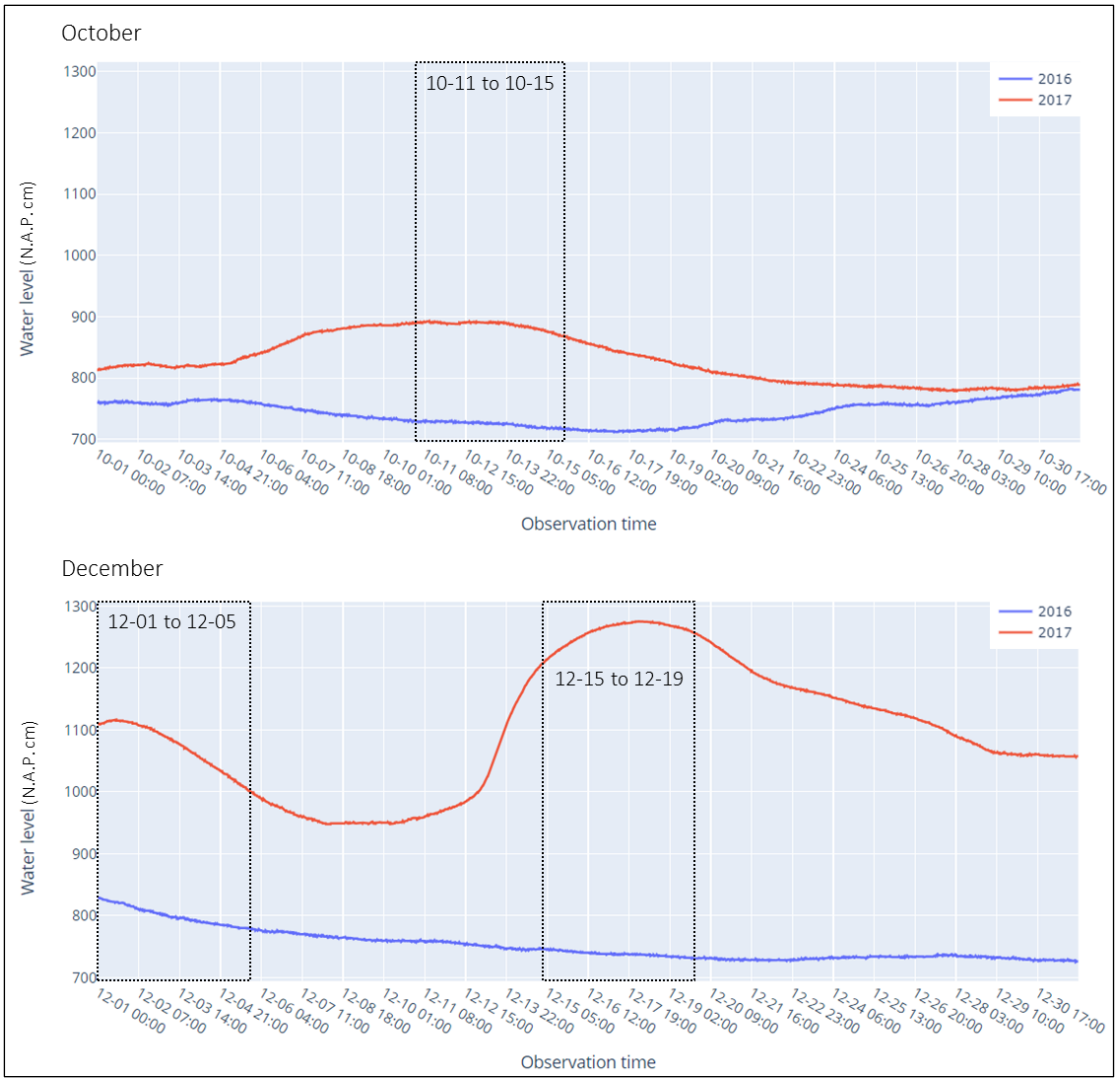


Figure 4.1 Three periods with water level plots

By looking at Figure 3.13 and Figure 3.14, the draught value distribution in different periods shows no pattern between high and low water levels. Instead, the figures display many extreme values that are impossible to exist considering the waterway conditions at the time. To be more specific, the river floor outside Lobith is 3.3 meters above N.A.P. according to the bathymetric map (Rijkswaterstaat, 2020a). So, the minimum fairway depth at the lowest water levels (7.11 meters above N.A.P. at 2016-10-17) should be about 3.8 meters, and the maximum fairway depth at the highest water levels (12.76 meters above N.A.P. at 2017-12-18) is around 9.5 meters. However, there was one vessel with a 20-meter draught that continued to sail along the study area regardless of the water levels. To present the temporal distribution of high draught values, trips that had draught higher than 6 meters were selected and made into histograms. Figure 4.2 presents the draught outliers in count. The results show no significant difference between four periods, moreover, the number of ships that have draught value bigger than 13 meters appeared in all four period regardless of the water level changes. Thus, we can conclude that the draught value in these AIS datasets has no correlation with water level or flow rate.

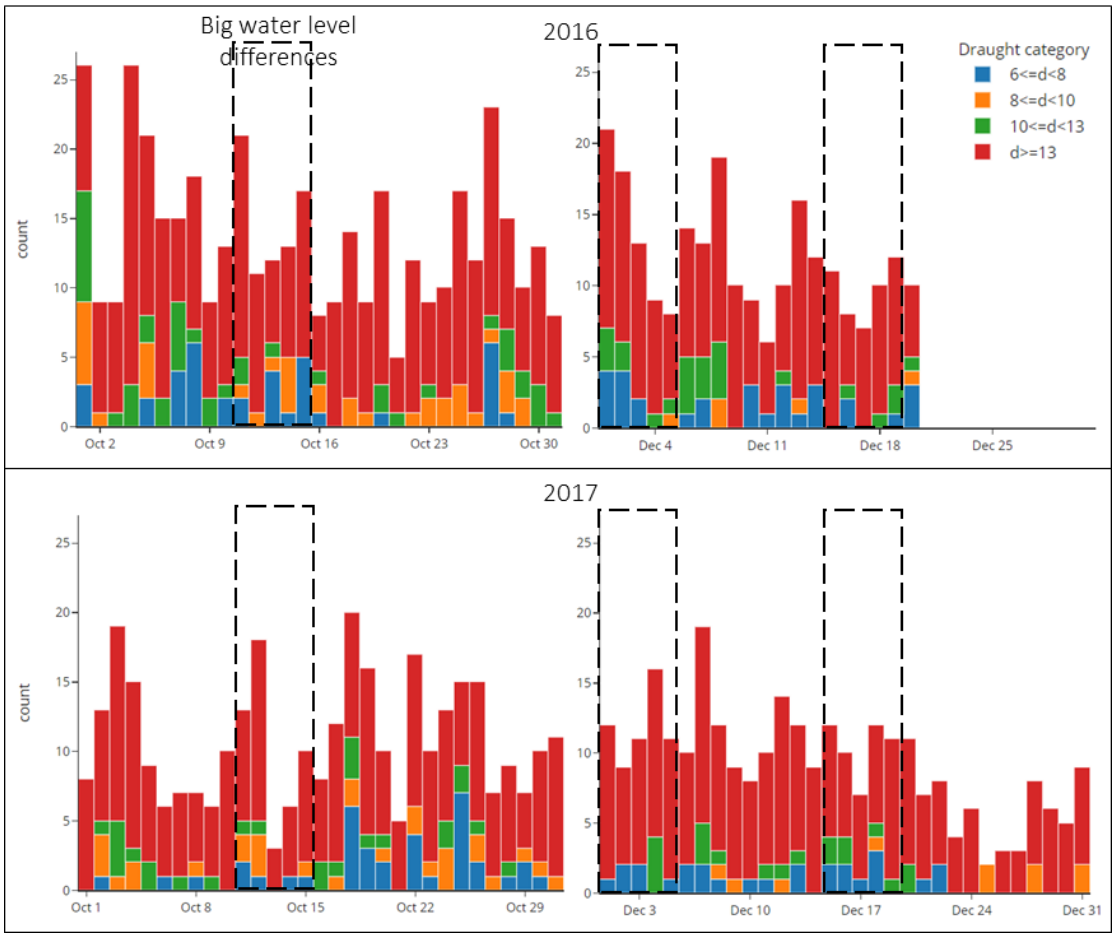


Figure 4.2 Number of ship's draught bigger than 6 meters

There are several possible reasons that might create the extreme draught values in the AIS datasets. Draught number relies on manual updating, which leads to the uncertainty of intentionally or unintentionally human errors. Considering the consistency of extreme draught value, it is hard to tell among not-extreme values which are reliable, and which are not. In summary, the fact that extreme draught values exist throughout the study periods makes the draught data unreliable, thus it is impossible to determine whether statement 1 is true or not.

Conclusion:
Lack reliable data to make decision.

4.2 Statement 2 – Busier waterway

"the waterway is busier than usual"

The statement is short and simple, but "busier" can be interpreted in many ways. One way is to judge it by the number of trips within the Waal. Figure 4.3 shows that during low water level times such as 2016-10 and 2016-12, more trips were made comparing to the same period of 2017. Moreover, October is busier than December as expected since December has Christmas and New Year's Eve holidays. Three time-windows in the chart are the period with big differences in water levels between 2016 and 2017 (see Figure 4.1). Another reason to analyze data in a time window is due to the missing data of 2016-12, a time window allows data comparison in the same duration of time.

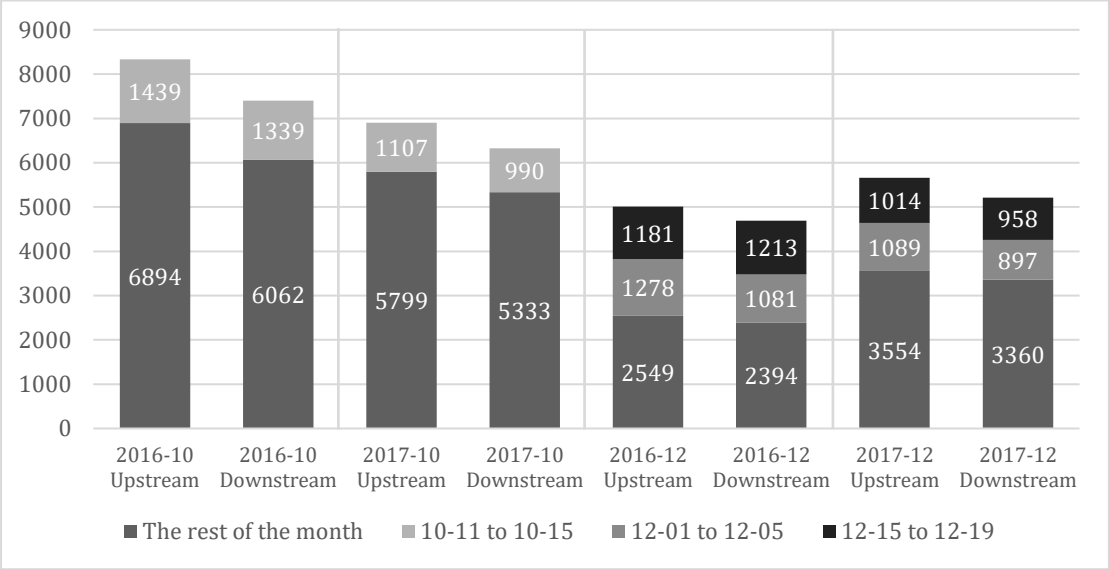


Figure 4.3 Trip numbers within the Waal

To further limit the influence of locks, canals, and mooring ships, only ships that traveled across the Waal were selected to calculate the average travel time between the east and west boundary of the study area. The results are shown in Figure 4.4. They show that there are more trips going downstream (west) than upstream (east), which is different from Figure 4.3. A possible reason for this might be there were more upstream ships turning to Pannerdensch Kanaal or Maas-Waalkanaal than downstream ships. Nevertheless, both figures agree that there are more trips in 2016 than in 2017.

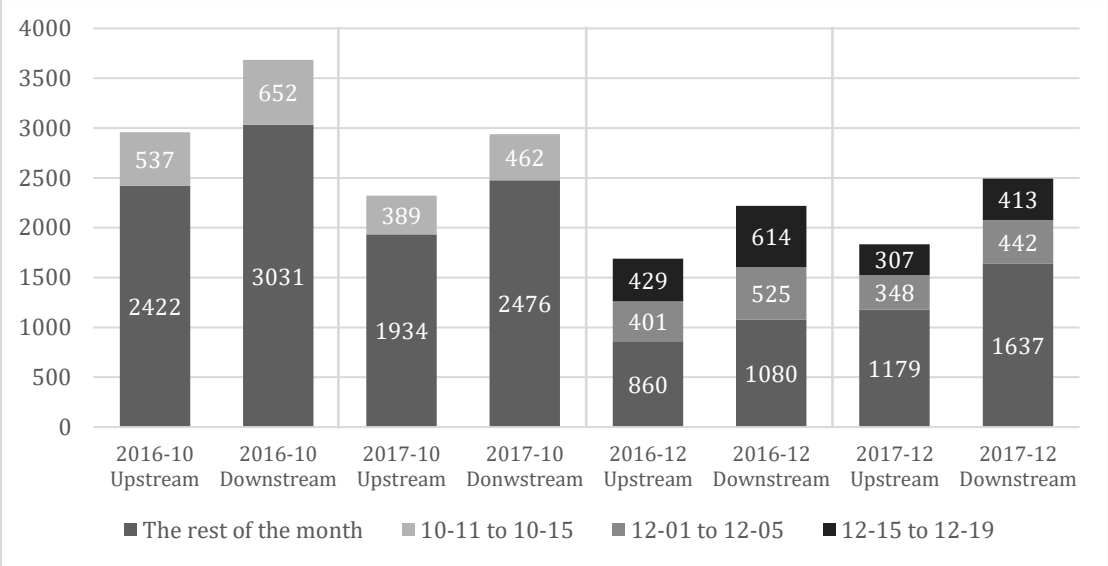


Figure 4.4 Trip numbers of traveling across the Waal

The average travel time is shown in Table 4.1 and the differences are as expected: going downstream took significantly less time than going upstream. Then Figure 4.5 and Figure 4.6 present the number of ships within the Waal in hour to provide a better view for readers to compare the numbers with water levels. In general, both figures show that there were more vessels traveling through the Waal in 2016 than 2017, while at the end of October, ship numbers in 2016 and 2017 became similar as the water levels trends (see Figure 4.1). Besides the patterns related to waterway conditions, the temporal pattern of weekday and weekend, daytime and nighttime can also be seen in Figure 4.5 and Figure 4.6 if compared to the calendar in Figure 4.7.

Table 4.1 Trip numbers and travel times for ships traveled across the Waal

Time period	Upstream/east travel time (h:mm)	Downstream/west travel time(h:mm)
2016-10-11 to 15	3:17	2:04
2016-10-01 to 31	3:20	2:03
2016-12-01 to 05	3:15	2:00
2016-12-15 to 19	3:20	2:03
2016-12-01 to 20	3:17	2:01
2017-10-11 to 15	3:18	1:56
2017-10-01 to 31	3:18	1:59
2017-12-01 to 05	3:25	1:49
2017-12-15 to 19	3:37	1:44
2017-12-01 to 31	3:28	1:48

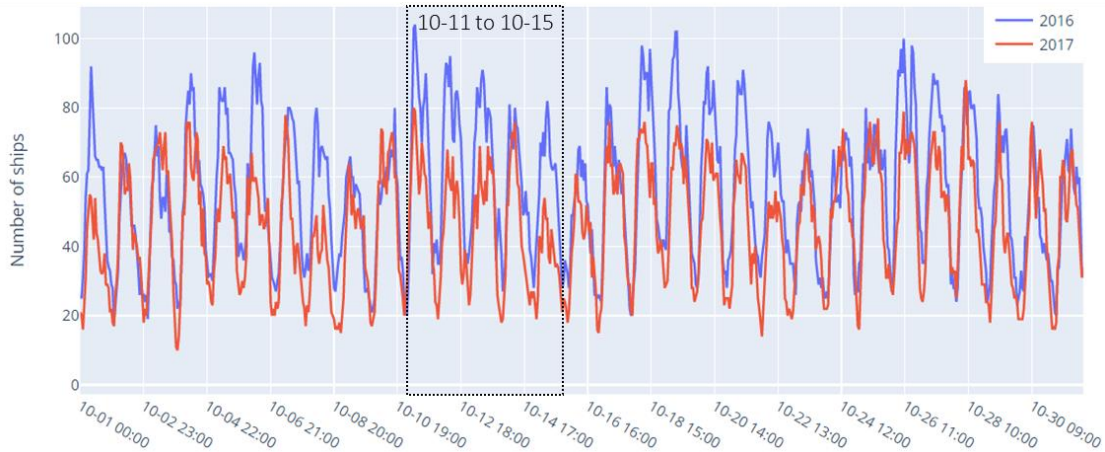


Figure 4.5 Number of ships within the Waal at each hour (October)

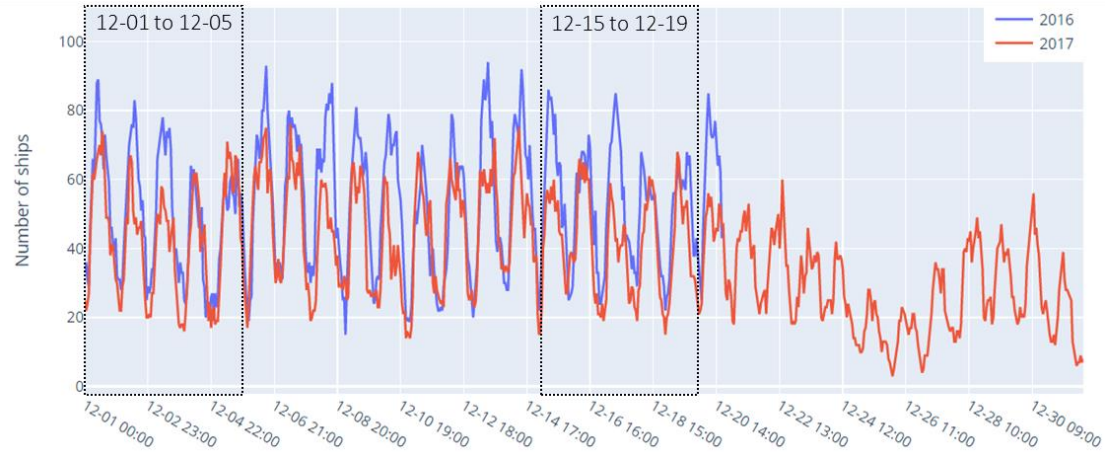


Figure 4.6 Number of ships within the Waal at each hour (December)

October 2016							October 2017						
Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat
						1	1	2	3	4	5	6	7
2	3	4	5	6	7	8	8	9	10	11	12	13	14
9	10	11	12	13	14	15	15	16	17	18	19	20	21
16	17	18	19	20	21	22	22	23	24	25	26	27	28
23	24	25	26	27	28	29	29	30	31				
30	31												
December 2016							**December 2017**						
Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat
				1	2	3						1	2
4	5	6	7	8	9	10	3	4	5	6	7	8	9
11	12	13	14	15	16	17	10	11	12	13	14	15	16
18	19	20	21	22	23	24	17	18	19	20	21	22	23
25	26	27	28	29	30	31	24	25	26	27	28	29	30
							31						

Figure 4.7 Calendar of the study periods

In summary, the results show that cargo and tanker vessels traveled more frequently, and there were more ships in the Waal during low water level times. Thus, the statement is true.

Conclusion:
Statement true.

4.3 Statement 3 – Reduce speed

"ships reduce speed to minimize the dynamic draft"

The speed of a vessel can be affected by many external factors, such as the connection to other waterways, the structure of a bridge, river bend shape, or a nearby port. As a result, it would be impossible to conclude which factor affected the average speed by comparing the whole study area. So, six small areas were chosen to represent different waterway conditions as mentioned in 3.4.2 Implementation, and the areas are shown in Figure 4.8. Then the average speed of each area was calculated and presented in Table 4.2, which shows two significant differences between low and high water level times. First, upstream ships were slower than downstream ships. Second, the average speed is related to water level and flow rate. December 2017 recorded unusual high water level numbers, and the river flow rate was also high at that time. During this time, downstream ships traveled faster, while upstream ships traveled slower. Upstream ships in December 2017 seems to be slightly affected by the water flow rate, but the average speed is not precise enough to conclude that it was related to water flow rate. Nevertheless, these results do not match the statement since flow rate seems to have more influence on vessel speed than the depth of the fairway.

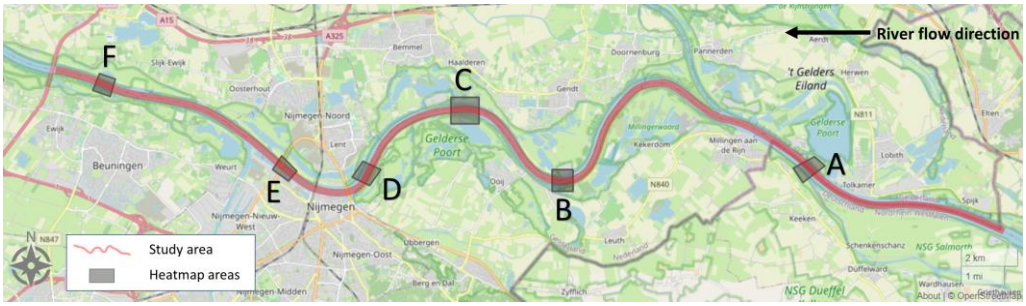


Figure 4.8 Heatmap areas

Table 4.2 Average SOG (knots)

Area		Upstream / Heading east				Downstream / Heading west			
		2016-10	2016-12	2017-10	2017-12	2016-10	2016-12	2017-10	2017-12
A	Outside Vluchthaven Tolkamer	5.89	5.94	5.82	4.99	8.83	8.98	9.31	10.61
B	Outside Erlecom	6.04	6.08	5.98	5.49	9.27	9.13	9.40	10.29
C	Outside Groenlanden	5.92	6.05	6.08	5.77	9.51	9.62	9.76	10.51
D	East of Waalbrug	5.85	5.95	5.90	5.71	9.43	9.59	9.83	10.82
E	Under De Oversteek bridge	5.76	5.84	5.77	5.56	9.25	9.45	9.72	10.78
F	Outside Ewijkse Plaat	5.99	6.08	5.99	5.89	9.35	9.53	9.74	10.64

To further investigate the temporal pattern of average speed, a series of heatmaps were generated. Every pixel in a heatmap represents a 30-minute timespan and the pixel's color shows the average speed within these 30 minutes. Green is the highest while red is the lowest speed. The x axis is the time axis, which starts from the left at 00:00:00, and ends at 24:00:00. The y axis represents the date, first day of the month is at the bottom of each heatmap while the top is the last day of the

month. (see Figure 4.9). Note that there are some white pixels which indicate that no vessel was in the area within these 30 minutes. Figure 4.10 and Figure 4.11 show that there is no significant temporal pattern of rush hour or congestion hotspot. However, the distribution of empty pixels shows that there were less vessels traveling during night times.

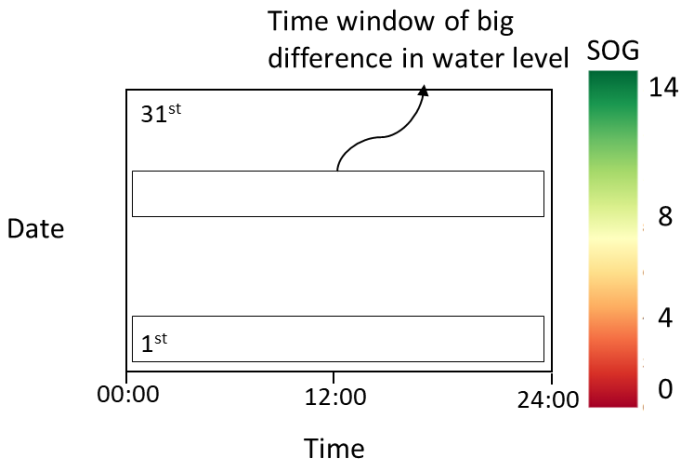


Figure 4.9 Heatmap reading guide

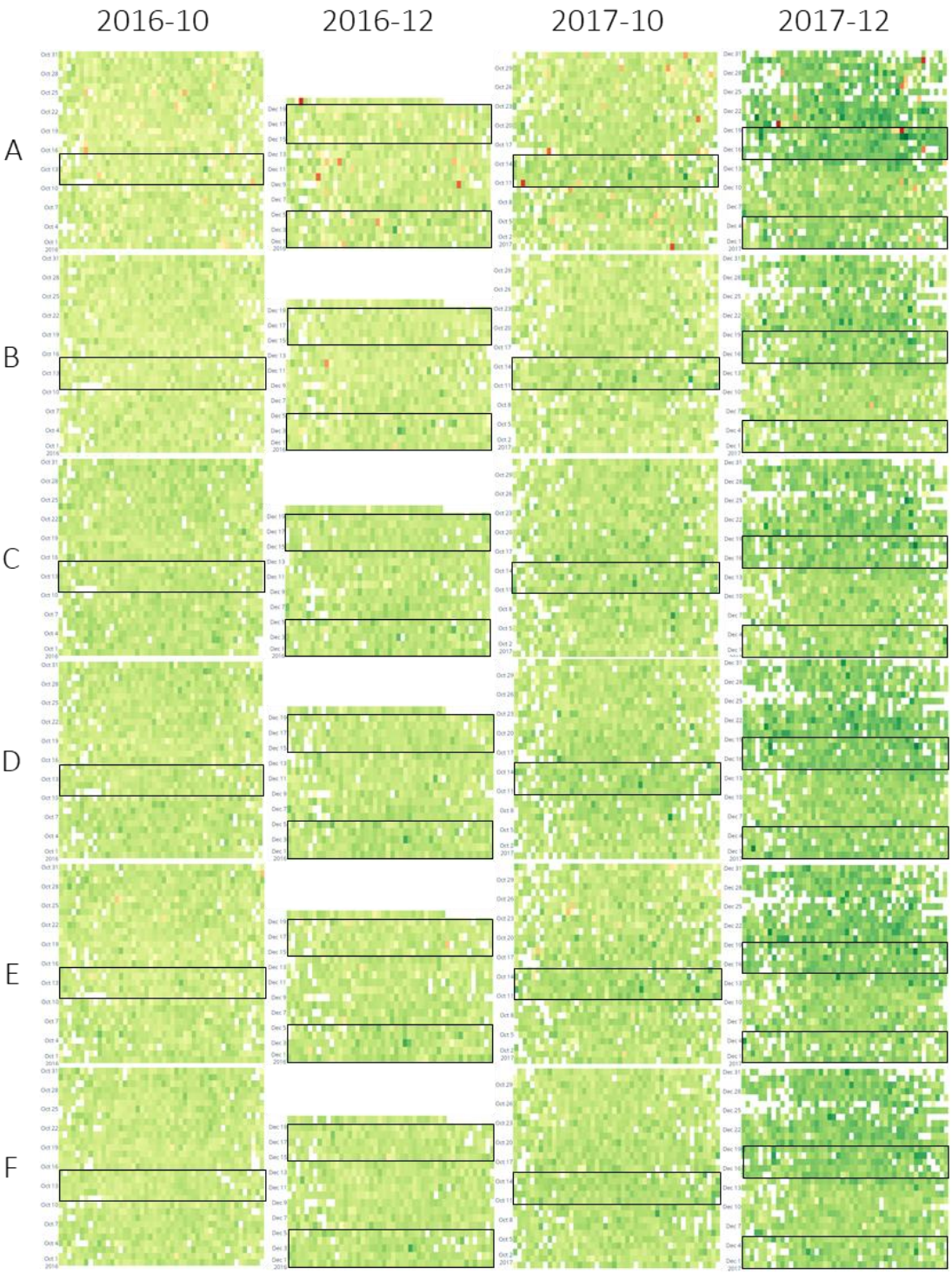


Figure 4.10 Temporal heatmaps of downstream mean SOG

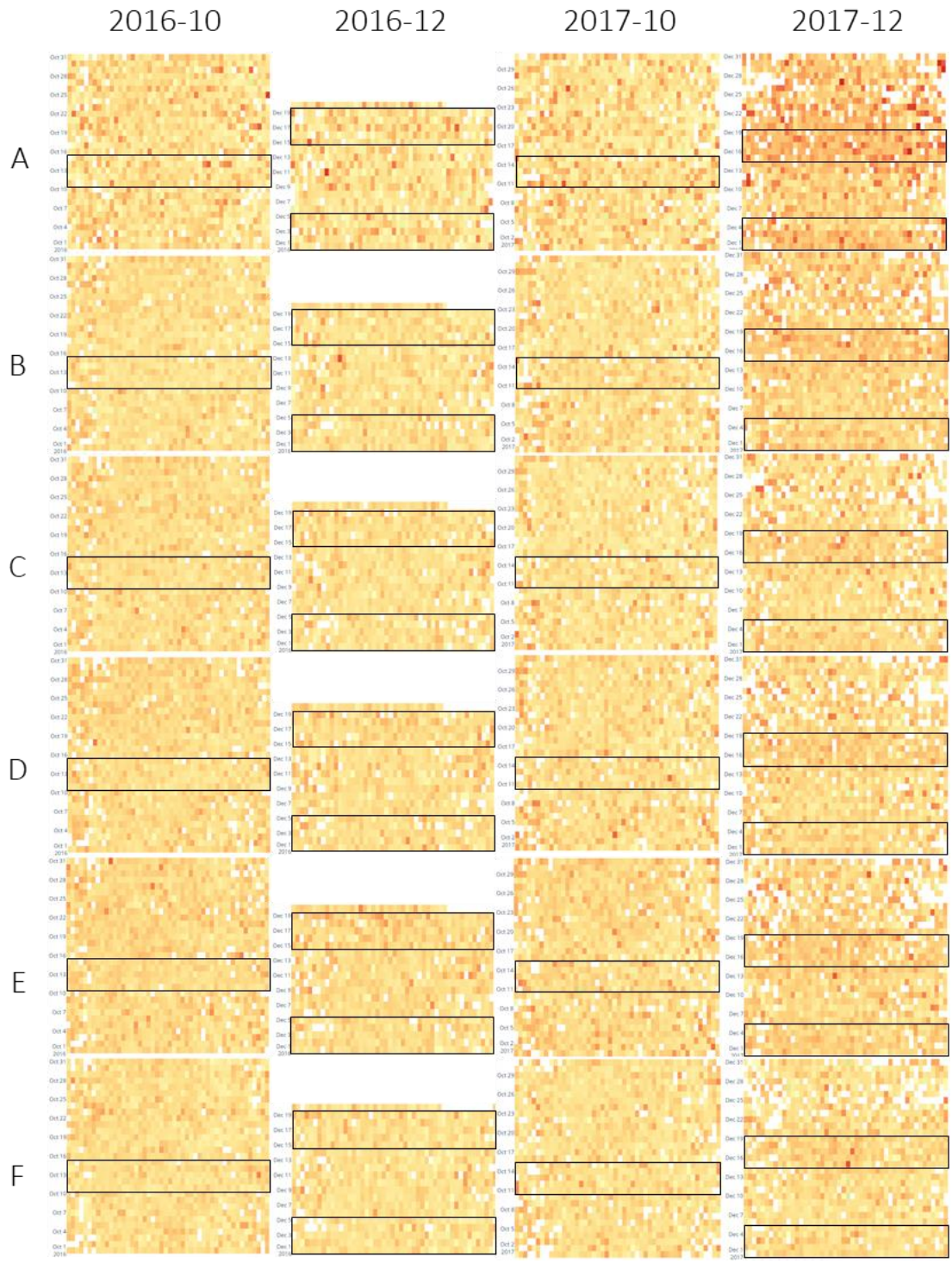


Figure 4.11 Temporal heatmaps of upstream mean SOG

In summary, the results are not enough to link low water level with low speed, instead high river flow rate is related to higher downstream speed and lower upstream speed. Thus, the statement is not true.

Conclusion:
Statement not true.

4.4 Statement 4 – Ship size changes

"smaller vessels shift operation area to the middle and upper Rhine"

To further verify the statement, AIS records within the Rhine along the Dutch border were selected. The data were then sorted by each trip and its vessel size. The results are displayed in a series of 2D histograms which generalize the data to resemble contour lines so that patterns can be clearly shown (comparing to Figure 3.13 and Figure 3.14). Figure 4.12 shows the vessel size distribution in October 2016 and 2017. The size groups are distributed similarly and show no significant difference.

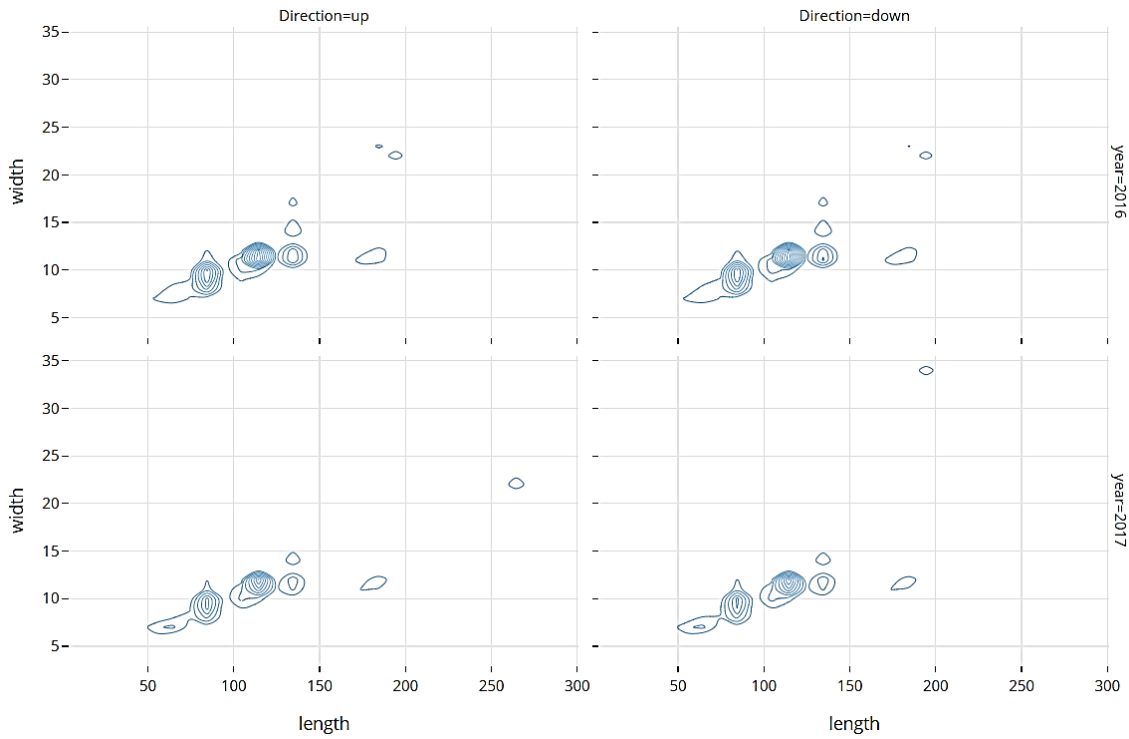


Figure 4.12 Ship sizes around the Rhine (October)

While Figure 4.13 shows that December's size distribution is different between 2016 and 2017, December 2017 ship sizes were more concentrated in two groups: 80×10 meters and 120×12 meters, but December 2016 has wider range of sizes (highlighted in red arrows).

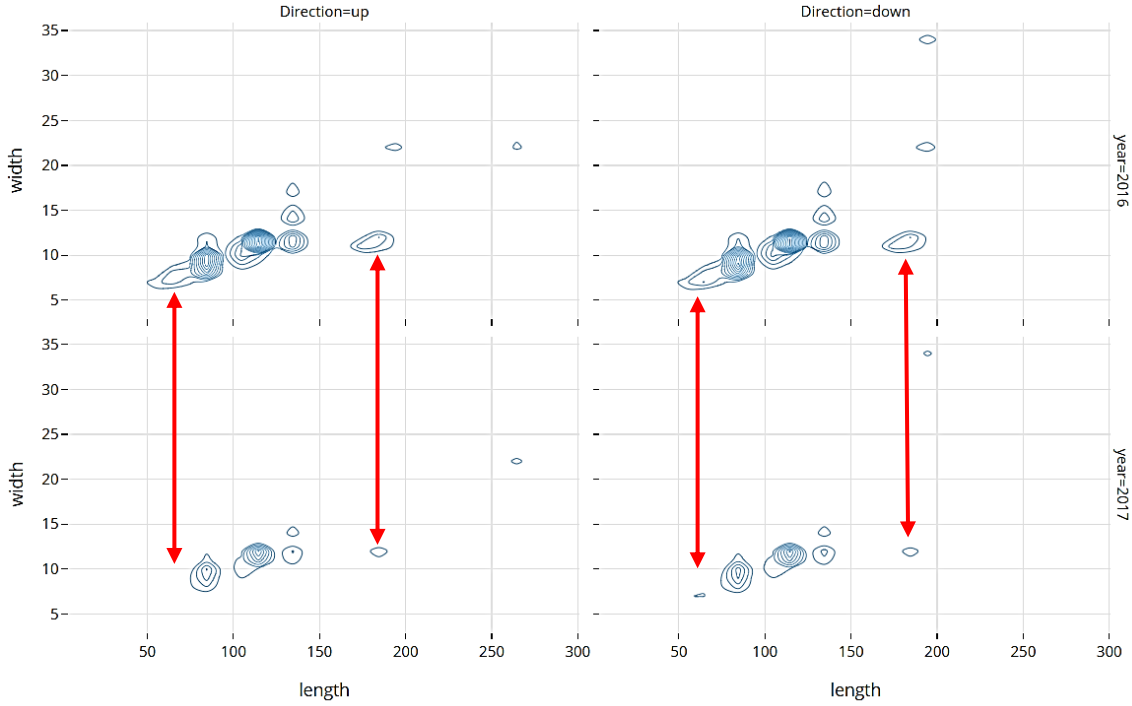


Figure 4.13 Ship sizes around the Rhine (December)

With all the results gained, it is important to understand that the study area is not part of the middle or upper Rhine. Thus, even the results do not verify the statement, it is too abrupt to conclude that the statement is not true. Moreover, even approaching the statement in other direction by asking whether there were fewer small vessels in the study area (located in the lower Rhine), the results still provide no answer to this question.

Conclusion:
Wrong study area. Cannot validate the statement.

4.5 Ship encounter anomalies detection

This section is aiming at discovering the interaction anomalies of moving objects, namely the vessel encounter anomalies. The algorithm applied was adapted from a ship domain violation detection model that is for assessing the collision and risk of open water navigation (Nordkvist, 2018). Considering the fundamental differences between open water and inland waterways, and to reduce the demand for calculation power, a simpler algorithm was developed to achieve the research goal.

The algorithm classifies the encounter types by which side it happened. Since the Waal adopts right-hand traffic on navigation, ships should encounter the opposite direction ship by the port side. However, the initial analysis found that sometimes ships met other ships by the starboard side, which is not normal. The types of encounters are presented in Figure 3.21.

The numbers of different encounters are displayed in Figure 4.14 as stacked histogram and Figure 4.15 as stacked percentage histogram. The results show that 2016-10 has the most encounter events, and both 2016-10 and 2016-12 have more encounter events than 2017. However, the percentage histogram indicates that the proportion of encounter types stay similar throughout the four study periods. The reason 2016 has more encounters than 2017 could be caused by the busier traffic in 2016. The encounters only happen when vessels came close to each other, so the chance of meeting another vessel is higher if there are more vessels in the waterway. As shown in Figure 4.1, 2016-10 and 2016-12 have similar water levels, and both have busier waterways than 2017-10 and 2017-12. More trips were made (see Figure 4.3 and Figure 4.4) and more ships were sailing inside the study area (see Figure 4.5 and Figure 4.6) in 2016, which increases the possibility of encountering other ships. Besides the encounter type temporal differences, Figure 4.14 also shows a pattern of weekday and weekend. The pattern of less encounter events on Sunday and Monday matches the results of statement 2, which was discussing about how busy the low water level times was (see Figure 4.5 and Figure 4.6).

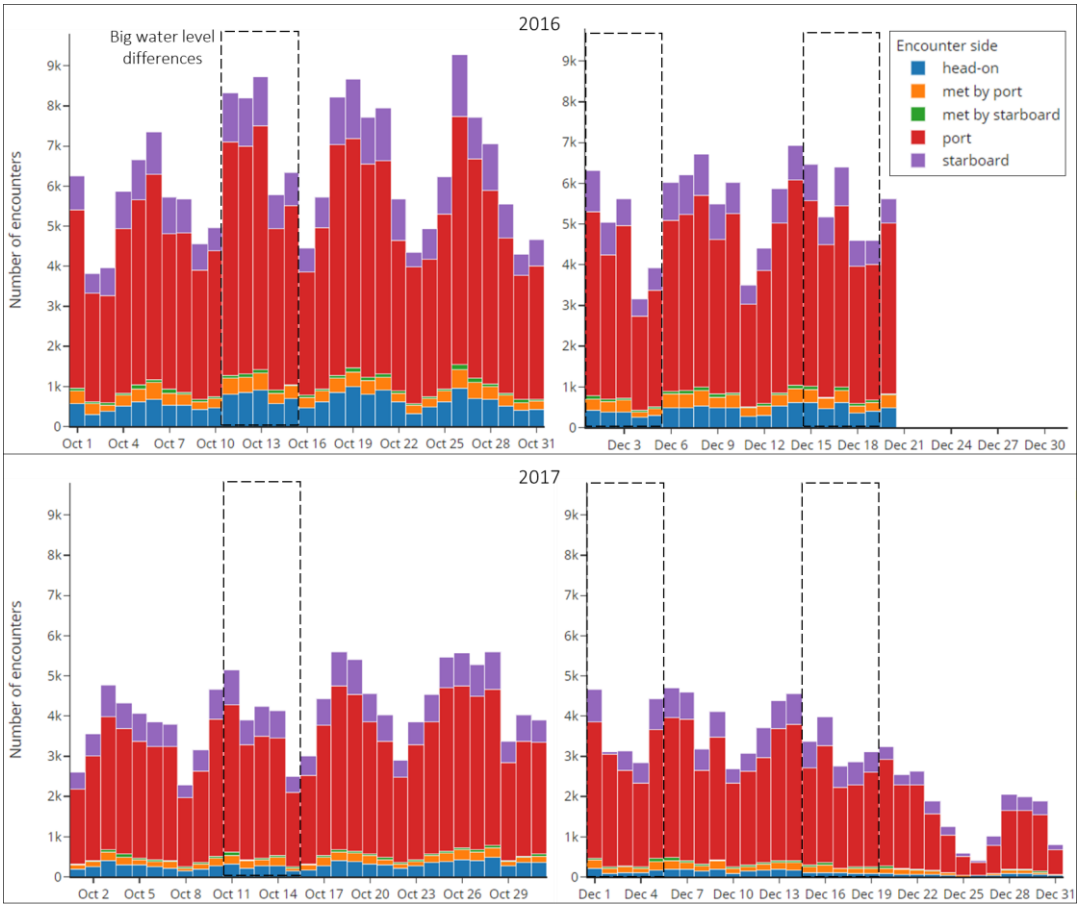


Figure 4.14 Histograms of encounter side counts

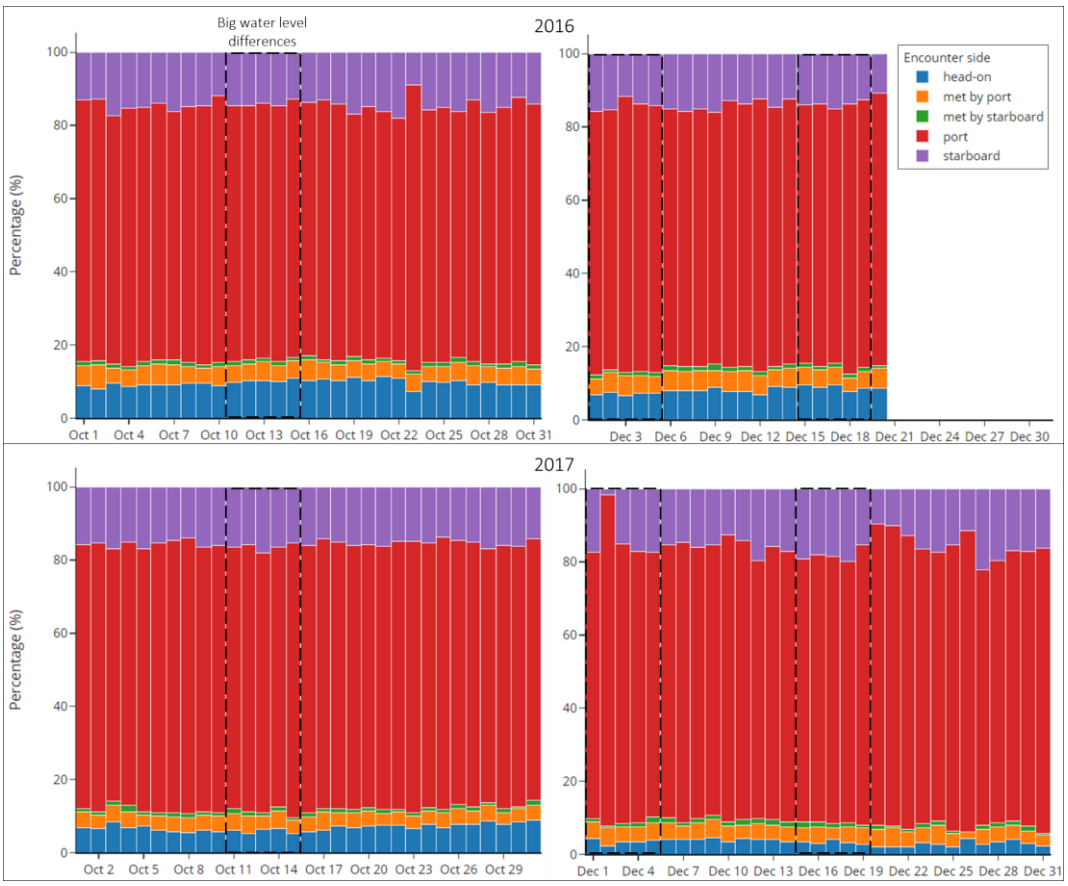


Figure 4.15 Histograms of encounter side percentage

To visualize the temporal intensity of encounter events, heatmap was chosen as the display method. The starboard encounter events were then selected and visualized by the number of events in each date. Other encounter types were also visualized but since they are not related to the research question, and not showing other significantly different patterns, those heatmaps are only included in Appendix IV. Figure 4.16 presents the temporal distribution of starboard encounters; each pixel represents one hour, and the white pixel means no data. The number of starboard encounter events are displayed by color: black is 0, while the light yellow is 195. Between the white lines are time windows when there was a big water level difference. Like Figure 4.5 and Figure 4.6 presented, the waterway is busier outside sleeping hours, and there is also a temporal pattern of weekday and weekend. More precisely, there were fewer ships and encounters before and after Sundays, which also explains why 2016-10-17 (Monday) has the lowest water level but 2016-10-19 (Wednesday) has the most encounters instead. More ships traveling increases the possibility of meeting other ships.

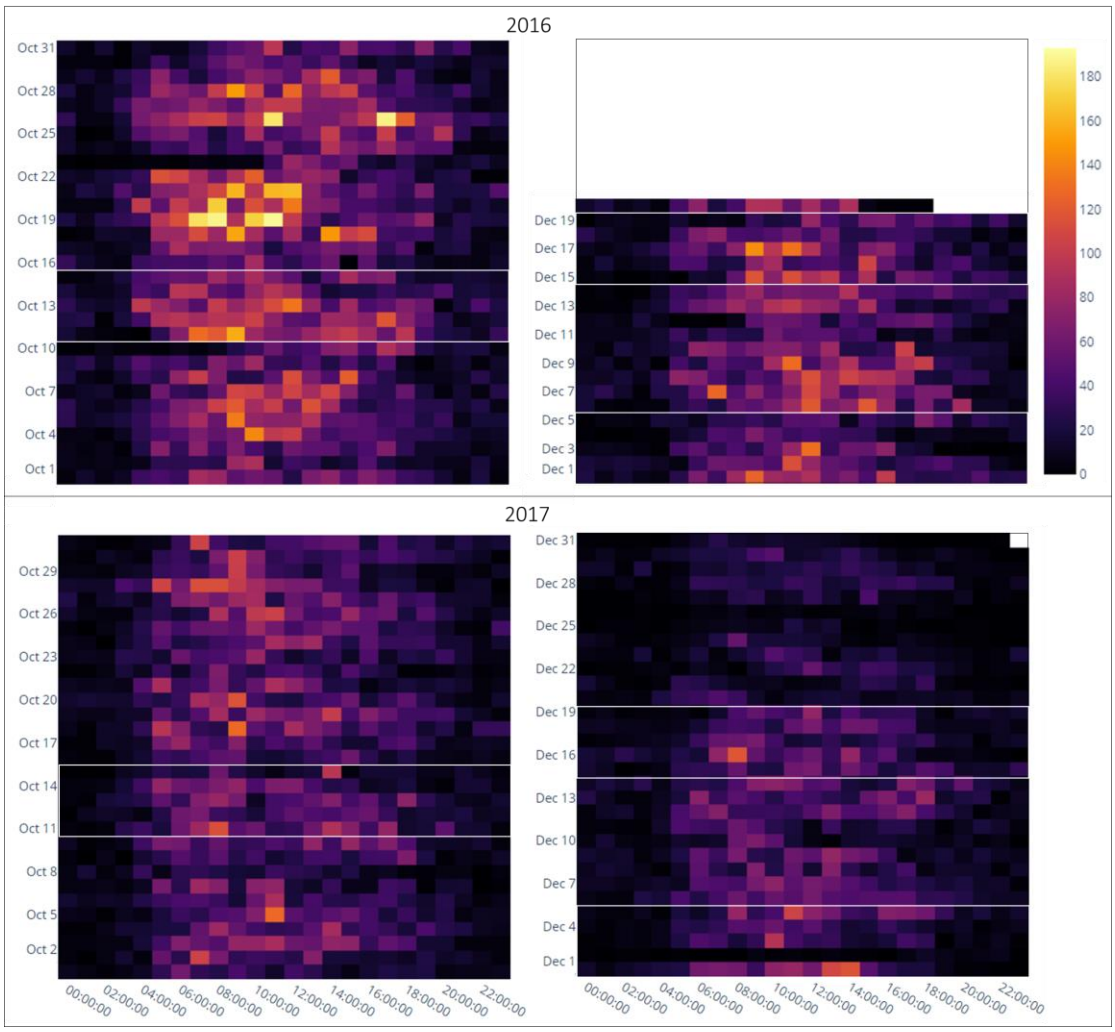


Figure 4.16 Temporal heatmaps of starboard-side encounter counts

According to above analysis, it is safe to conclude that during low water level times there are more ships and more encounters, while the percentage of starboard encounters was not affected by factors that vary with time. Thus, the question then turned into whether the events were affected by waterway characters like river bend shape, connection to other waterways, or outside a port. A series of kernel density maps were created to show the spatial distribution of starboard encounter events. A kernel density map is an estimation of probability density function. The probability of a random point from the dataset falls within a range of X and Y is the integral under the surface, see

Figure 4.17 (Portugués, 2021; Smith et al., 2018). The scale bar in Figure 4.18 is the estimation value to show the density and intensity pattern of a data distribution in space.

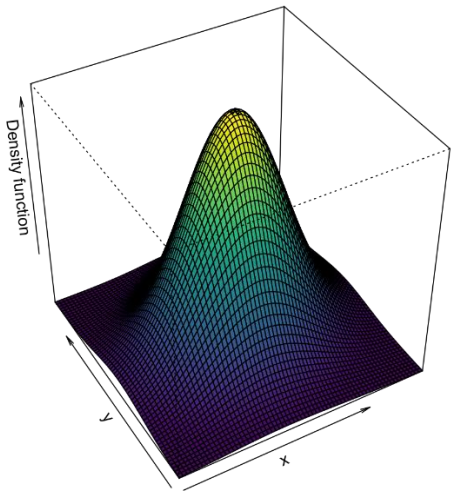


Figure 4.17 3D example of kernel density estimation (Portugués, 2021)

Figure 4.18 shows that the hotspots of starboard encounters are consistent in four different months (for the animation of daily density map, check http://doi.org/10.5281/zenodo.4531866). In the view of an upstream ship, the first hotspot occurs when approaching the entrance of Maas-Waalkanaal and Weurt lock. The geography of this place can be the cause of wrong side encounters since the waterway is relatively straight here, but ships might change course to the Maas-Waalkanaal and resulted in being detected as starboard encounters. This hotspot continues till the river bend at Nijmegen and three bridges above it. The river bend at Nijmegen is where downstream side of the waterway has a slower river flow, so many downstream ships take advantage of the faster flow by sailing on the left-hand side of the waterway. This also happens at the second river bend, where downstream ships sail on the left-hand side to follow the faster river flow. The next hotspot is close to the entrance of Pannerdensch Kanaal, as ships turn to or come out of the canal might meet another ship by the starboard side. The last hotspot is outside of Vluchthaven Tolkamer, where ships going into or coming out from the safehaven can meet ships from the opposite direction.

In conclusion, the frequency and percentage of starboard encounter events are not related to the water level changes, but instead factors such as river bend shape and waterway networks have more influence on their spatial distribution.

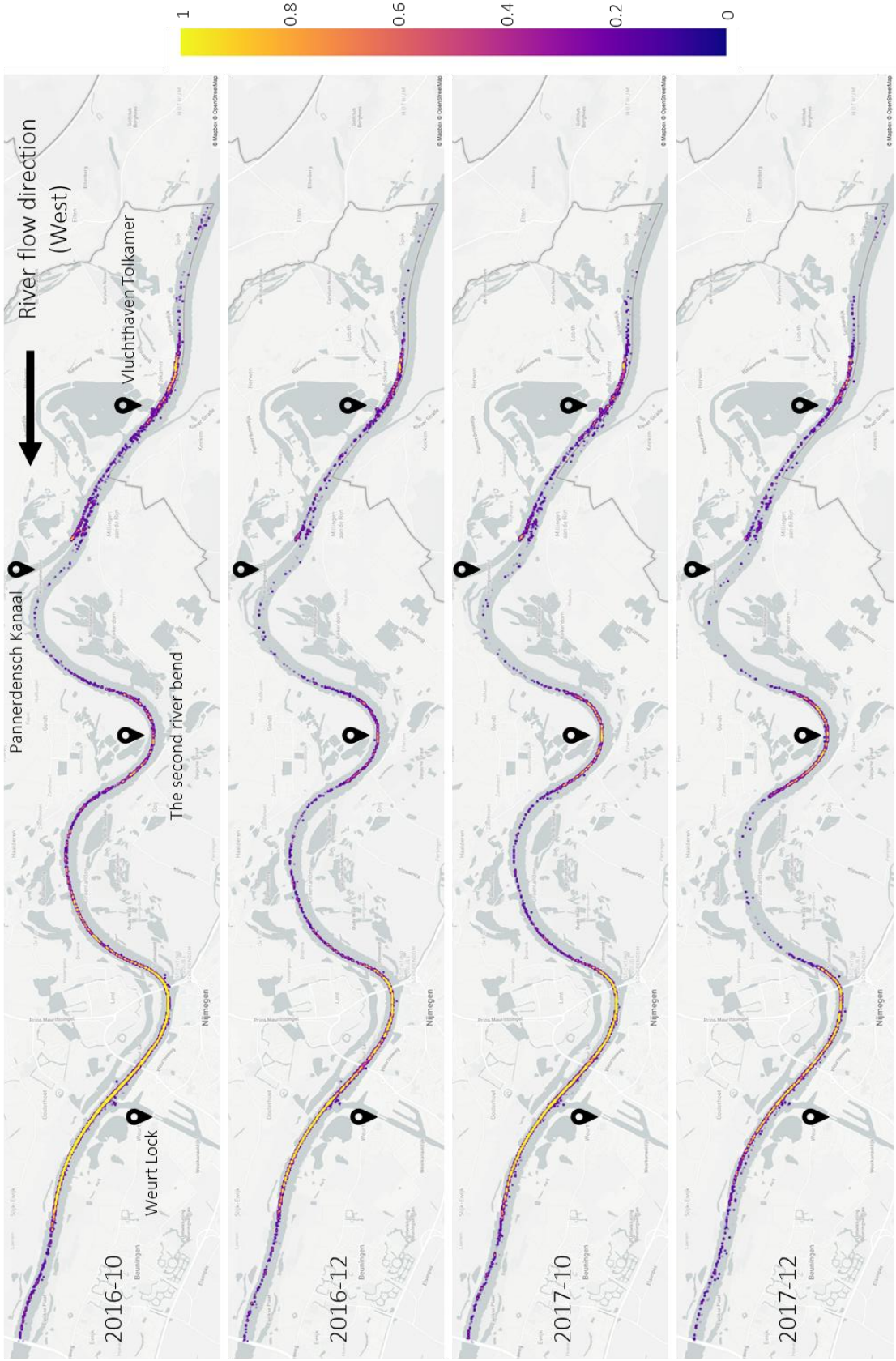


Figure 4.18 Kernel density maps of starboard-side encounters

5. Conclusion and Discussion

This chapter concludes the research results and discusses the limitations and future works. Section 5.1 summarizes the research questions and results. Section 5.2 reviews the thesis results and discusses the limitations. Section 5.3 brings up the possible future works that are either inspired or related to this research.

5.1 Conclusion

In the market observation validation, only statement 2 "the waterway is busier than usual" was concluded as true, the other 3 statements are either false, "wrong study area", or lacking reliable data. Nevertheless, the ship encounter anomalies detection results show that it is possible to identify and visualize vessel interactions by exploiting historical AIS data. Before answering the research question, the following five sub-questions need to be considered first.

What types of impacts are caused by low water levels in inland waterways? Amongst them, which can be detected by exploiting AIS data?

The first question was answered by those four statements which were summarized from reports and news articles, and the answer to the second question is the conclusions of the statements. Statement 1 states that low water level limits the cargo capacity of inland waterway transport, but it cannot be verified because the draught values in AIS data are not reliable, and draught is the only element related to the amount of cargo in AIS data. As a result, cargo amount changes cannot be detected by AIS data. Next, statement 2 claims that the waterways are busier than normal water level times because cargo ships had to travel more trips in order to transport the same weight of goods when the cargo capacity was limited. The statement was verified as true because the vessel and trip numbers varied with the water levels. Statement 3 points out that when the water level is extreme low, ships even have to reduce speed to minimize the dynamic draft. The statement was false according to the vessel speed distribution in different time and space. The downstream vessels did sail faster when the water level was high, but the speed shows no difference between normal and low water levels. Statement 4 says that smaller vessels move their area of operation to the middle or upper Rhine to take advantage of the high freight rate there. The analysis results do show differences between 2016-12 and 2017-12, but they are not significant enough to make a conclusion on the relationship between water level and ship size. Moreover, the statement was indicating the middle and upper Rhine, but the study area is located at the upstream of lower Rhine. Thus, it will be possible to detect ship size distribution on the right location, and the ship encounter anomaly detection shows that AIS data can present the ship interactions, though the starboard side encounters do not directly cause by low water levels. These statements and encounter analysis results are listed in Table 5.1.

What attributes of AIS data are relevant to identify low water level impacts on inland waterways?

Table 5.1 List of impacts and AIS elements

	Impacts	**Used AIS data elements**	**Conclusion**
Statement 1	Cargo capacity	draught	Lack of reliable data to make decision.
Statement 2	Busier waterway	timestamp, unique ship ID, longitude, latitude	True.
Statement 3	Speed changes	timestamp, speed over ground, longitude, latitude	False.
Statement 4	Ship sizes differences	unique ship ID, trajectory ID, longitude, latitude, length, width	Wrong study area. Cannot validate the statement.
Ship encounter anomalies	Starboard side encounters	timestamp, unique ship ID, longitude, latitude, course over ground,	Encounter events are related to how busy is the waterway, which can also be affected by low water levels.

Which kinds of anomalies happened on the Waal near Nijmegen during low water levels?

As stated earlier, the waterway was busier during low water levels, which increased the chance of encountering other vessels. Since the total number of meeting other ships increased, the events of starboard side encounters also increased. However, note that the percentage of encounter types shows that the ratio of starboard side encounters does not change during the four research periods. Further investigation also showed that locations for starboard encounters do not change with water levels.

What are the differences in analysis results between using AIS data and former reports that did not use AIS data?

The analysis that does not use AIS data usually choose market and freight statistics or data from other ship tracking systems, for instance IVS90 (Informatie en volgsysteem scheepvaart 1990). The fundamental differences are how the data was collected or generated at the beginning. AIS data are based on detectors that automatically generate data, as well as skippers to update data manually. While the market and freight statistics are from the business report of each shipping companies for inside usage, which guarantees a certain level of accuracy and quality. Data like IVS90 are collected and managed by government agency to secure inland traffic safety and make waterway management efficient and effective (Rijkswaterstaat, 2020b). As a result, different datasets have different strengths depends on data accuracy, quality, and reliability. In summary, AIS data are strong at precision of timestamp, geographical position data, and automatically generated data. Analysis that uses these elements will produce reliable results.

To summarize the answers above, AIS data are suitable for analyzing geo-spatial topics. Especially in combination with a database management system, a larger dataset can be analyzed and provides high quality results.

5.2 Discussion

The research goal of this thesis was to fill in the gap between the market report of inland shipping and AIS data analysis. So even though the findings of the low water level impacts on inland shipping are not groundbreaking, this thesis has tried to validate the observations with a different data type and concludes what is suitable and what is not. Furthermore, the analysis examines and presents

the AIS data elements' quality in the study of low water level impacts, which can provide management agency another tool on inland waterway navigation management. The practical relevance of this study is the tools used and their possible future applications. First, the entire analysis was done by using Python and Jupyter Notebook, which shows that it is possible to conduct this kind of analysis and visualization in any environment that can use Python and its libraries. Second, the analysis can be done without saving the AIS data to a local disk since the data were only kept temporarily in computer memory. It is possible to conduct the analysis virtually in a cloud environment, and with the support of a database management system, the selection, extraction, and management of the AIS data can be more efficient.

There are some limitations of this thesis that need to be discussed and hope to provide some recommendation for future works. First, this study only compared 2016-10, 2016-12, 2017-10, and 2017-12, which does not cover a wide range of time to make solid conclusion. The fundamental reason is because of the limited thesis research time and Rijkswaterstaat's long data-processing time. One month of AIS data within the study area takes one week to process and anonymize, so one year's AIS data will take at least 12 weeks of waiting time, which is unrealistic consider the thesis planning. However, a more precise research can include more data in different conditions such as during spring or summer times. The inland shipping transports goods from dry bulks like coal, metals, food products, to wet bulks like mineral oil products and chemicals. The supply and demand of different products can vary seasonally or even yearly, so the cargo and tanker vessel analysis might be different accordingly. Secondly, the draught values in these AIS datasets are not reliable, so it is reasonable to question the reliability of other static and voyage elements in the datasets (see Table 2.1). The static and voyage information is registered and updated manually to the AIS device, thus there is the possibility of human errors in device installation, data entry, or missing update. The third type of limitation is the reassigned vessel unique ID. New vessel ID was assigned to each unique vessel by its length, width, vessel type, and the anonymized IDs. This method works under the assumption of vessels having their own unique combination of length, width, vessel type, and the anonymized IDs. However, there is a possibility of assumption not true since how Rijkswaterstaat anonymized the original vessel IDs was unknown. It is also possible that the unique IDs assigned by this study did not fully separate each unique vessel, which would skew the results related to counting trips and unique vessels.

5.3 Future work

During the data cleaning and preparation stage, several errors in the data were discovered. Among them, two aspects are related to Rijkswaterstaat AIS data management and are possible to solve technically. First, the AIS database seems to be troubled by missing and duplicate data. The initial research goal of this thesis was to compare the second half of 2018 to a non-drought year, but Rijkswaterstaat does not have AIS data in that period. Then in the datasets of 2016-12, the last 1.6 million records are duplicates so the last timestamp of 2016-12 stops at 2016-12-20 18:19:56. Secondly, the AIS data provided by Rijkswaterstaat is anonymized, namely each vessel's IDs and its name were renamed to a random number by Rijkswaterstaat. The vessel ID is supposed to be unique and not shared by another vessel, but the datasets have around 10% of IDs shared by more than one vessel. This type of data error can skew the analysis on counting ship numbers or extracting vessel trajectories since the AIS data points of the same ID might appear in impossible positions and affect the results. Thus, a future work can focus on improving the database management of the Rijkswaterstaat's historical AIS database.

About the unreliability of some voyage information in the AIS data, a future study can test the possibility of complement AIS data with the IVS90 data. IVS90 stores basic information about the vessel and its voyage, even including real-time draught and cargo information (BICS, 2014).

However, a big obstacle lays in front of the study, that is the AIS data provided by Rijkswaterstaat have been anonymized. AIS data without real vessel IDs can hardly find and match the same vessel in IVS90 data. Since the AIS data were made anonymous to protect privacy, combing them with identified IVS90 data would be some kind of reverse engineering and totally ignoring the purpose of anonymization. Nevertheless, it is an idea for improving the analysis accuracy and quality in the future.

In section 4.5, the temporal and spatial distribution of starboard side encounters are analyzed, but this method only detects events where two vessels met. Another way to find out the spatial distribution of vessel sailing on the left bank of the waterway can be done by finding the frequent-traveled path. A further analysis on this topic can provide more information about where it happened and whether it meets the conditions of exception. The frequent-travel path can be extracted by applying kernel density function on the sampled AIS point data, and then transform the raster kernel density map into polygon based on the preferred raster value. Take this study area as an example, Figure 5.1 shows the kernel density estimation of downstream vessels in one day, the path is clear, while Figure 5.2 displays the same day of vessels but going upstream instead, the map shows that upstream vessels traveled in both sides of the waterway. The locations that have overlapping path can indicate the occurrence of unusual ship encounters. Moreover, a research has used this method on an open water case (Lee et al., 2020), but it would be interesting to see the effects on inland waterways as well. Since the inland waterway can be narrow comparing to open water or coastal area, the precision and strength can be examined and tested by analyzing the frequently traveled path within an existing fairway.

Figure 5.1 Kernel density map of downstream vessel in 2016-12-16

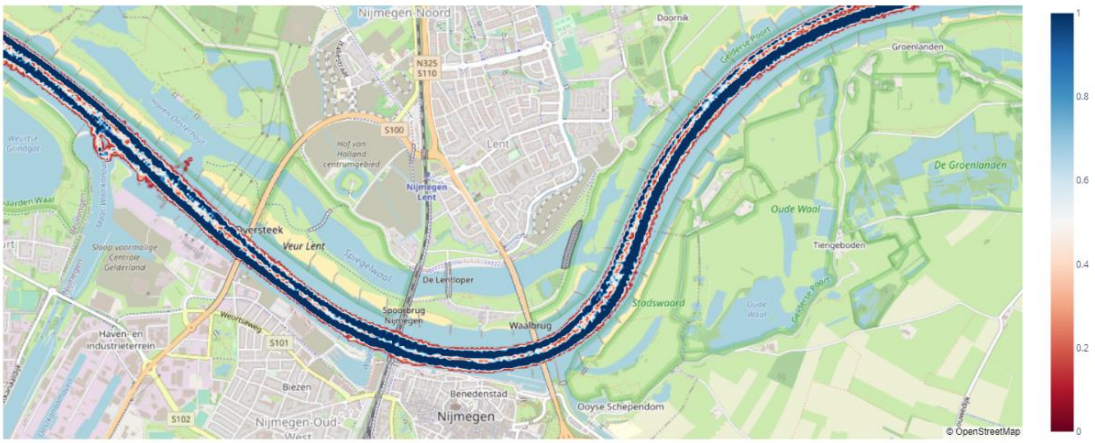


Figure 5.2 Kernel density map of upstream vessel in 2016-12-16

6. REFERENCES

Altan, Y. C., &Otay, E. N. (2018). Spatial mapping of encounter probability in congested waterways using AIS. *Ocean Engineering*, *164*(June), 263–271. https://doi.org/10.1016/j.oceaneng.2018.06.049

Andrienko, G., Andrienko, N., Chen, W., Maciejewski, R., &Zhao, Y. (2017). Visual Analytics for Transportation: State of the Art and Further Research Directions. *IEEE Transactions on Intelligent Transportation Systems*, *18*(8), 2232–2249. https://doi.org/10.1109/TITS.2017.2683539

Andrienko, G., Andrienko, N., &Fuchs, G. (2016). Understanding movement data quality. *Journal of Location Based Services*, *10*(1), 31–46. https://doi.org/10.1080/17489725.2016.1169322

Andrienko, G., Andrienko, N., &Wrobel, S. (2007). Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, *9*(2), 38–46. https://doi.org/10.1145/1345448.1345455

Andrienko, N., &Andrienko, G. (2011). Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics*, *17*(2), 205–219. https://doi.org/10.1109/TVCG.2010.44

Andrienko, N., &Andrienko, G. (2013). Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, *12*(1), 3–24. https://doi.org/10.1177/1473871612457601

Arntsen, R. (2019). *Analysis of AIS-data to identify near encounter situations between ships*. Norwegian University of Science and Technology.

Barendregt, L. (2018, August 22). Waal bij Nijmegen kan vandaag historisch laag waterpeil bereiken. *De Gelderlander*. Retrieved from https://www.gelderlander.nl/nijmegen/waal-bij-nijmegen-kan-vandaag-historisch-laag-waterpeil-bereiken~ab1063d0/

BICS. (2014). *Electronic Reporting* (pp. 1–10). pp. 1–10. Retrieved from https://www.wcpfc.int/electronic-reporting

Boztepe, G. (2019). *The Vessel Route Pattern Extraction and Anomaly Detection from AIS Data*. Middle East Technical University.

CCNR. (2013). *Vessel Tracking and Tracing Standard for Inland Navigation* (pp. 1–67). pp. 1–67. Central Commission for the Navigation of the Rhine.

CCNR. (2015). *Clarifications concerning the obligation of equipment with an Inland AIS and an Inland ECDIS device*. Central Commission for the Navigation of the Rhine.

CCNR. (2019a). Annual Report 2019 Inland Navigation in Europe Market Observation. In *Central Commission for the Navigation of the Rhine*. Retrieved from www.inland-navigation-market.org

CCNR. (2019b). Annual Report 2020 Inland Navigation in Europe Market Observation. In *Central Commission for the Navigation of the Rhine*. Retrieved from https://www.rtda.gov.rw/fileadmin/templates/publications/RWANDA_Annual_Report_2018-2019_SHARING.pdf,

Cucinotta, F., Guglielmino, E., &Sfravara, F. (2017). Frequency of Ship Collisions in the Strait of Messina through Regulatory and Environmental Constraints Assessment. *Journal of Navigation*, *70*(5), 1002–1022. https://doi.org/10.1017/S0373463317000157

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Weghe, N.de, …Weibel, R. (2015). Analysis and visualisation of movement: An interdisciplinary review. *Movement Ecology*, *3*(1), 1–24. https://doi.org/10.1186/s40462-015-0032-y

Depellegrin, D., Bastianini, M., Fadini, A., &Menegon, S. (2020). The effects of COVID-19 induced lockdown measures on maritime settings of a coastal region. *Science of the Total Environment*, *740*, 140123. https://doi.org/10.1016/j.scitotenv.2020.140123

Dodge, S. (2015). From Observation to Prediction: The Trajectory of Movement Research in GIScience. *Advancing Geographic Information Science: The Past and Next Twenty Years*, (February), 123–136.

European Barge Union. (2020). *STATEMENT BY EBU PRESIDENT PAUL GORIS AT THE IWT EVENT UNDER GERMAN EU PRESIDENCY , 20 NOVEMBER 2020*. Retrieved from https://mcusercontent.com/9a6d55ece9db05ec49eeb87e0/files/0eec36bd-fa2d-494a-b2dd-c85b4602e078/EBU_President_s_Statement_IWT_Conference_20202011.pdf

Fan, S., Blanco-Davis, E., Yang, Z., Zhang, J., &Yan, X. (2020). Incorporation of human factors into maritime accident analysis using a data-driven Bayesian network. *Reliability Engineering and System Safety*, *203*(June), 107070. https://doi.org/10.1016/j.ress.2020.107070

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

Graser, A. (2019). MovingPandas: Efficient Structures for Movement Data in Python. *GI_Forum – Journal of Geographic Information Science 2019*, (1), 54–68. https://doi.org/10.1553/giscience2019

Graser, A. (2020). Movement data in GIS #26: towards a template for exploring movement data. Retrieved October6, 2020, from Free and Open Source GIS Ramblings website: https://anitagraser.com/2020/01/03/movement-data-in-gis-26-towards-a-template-for-exploring-movement-data/

Graser, A., Widhalm, P., &Dragaschnig, M. (2020a). Extracting Patterns from Large Movement Datasets. *GI_Forum*, *1*(1), 153–163. https://doi.org/10.1553/giscience2020_01_s153

Graser, A., Widhalm, P., &Dragaschnig, M. (2020b). The M^3 massive movement model: a distributed incrementally updatable solution for big movement data exploration. *International Journal of Geographical Information Science*, *00*(00), 1–24. https://doi.org/10.1080/13658816.2020.1776293

Green Deal. (2019). *GD230 Green Deal on Maritime and inland shipping and Ports* (pp. 1–21). pp. 1–21. Retrieved from www.greendeals.nl/green-deals/green-deal-zeevaart-binnenvaart-en-havens

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

Harati-mokhtari, A., Aharati-mokhtariljmuacuk, E., Wall, A., Brooks, P., &Wang, J. (2004). Automatic Identification System (AIS): A Human Factors Approach. *Traffic*, (April).

Inland Navigation Europe. (2020). Annual Report 2019. *Inland Navigation Europe*. https://doi.org/10.3934/math.2020i

Jiang, M., Lu, J., Yang, Z., &Li, J. (2020). Risk analysis of maritime accidents along the main route of the Maritime Silk Road: a Bayesian network approach. *Maritime Policy and Management*, *47*(6), 815–832. https://doi.org/10.1080/03088839.2020.1730010

Jonkeren, O., Jourquin, B., &Rietveld, P. (2011). Modal-split effects of climate change: The effect of low water levels on the competitive position of inland waterway transport in the river Rhine area. *Transportation Research Part A: Policy and Practice*, *45*(10), 1007–1019. https://doi.org/10.1016/j.tra.2009.01.004

Jonkeren, O., &Rietveld, P. (2009). *Impacts of low and high water levels on reliability and transport costs in inland waterway transport : an interview based report*. Retrieved from

https://edepot.wur.nl/326455

Jonkeren, O., Rietveld, P., &vanOmmeren, J. (2007). Climate Change and Inland Waterway Transport: Welfare Effects of Low Water Levels on the River Rhine. *Journal of Transport Economics and Policy*, *41*(3), 387–411.

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., &GuyMelan¸con. (2008). Visual Analytics: Definition, Process, and Challenges. *Information Visualization*, 154–175. https://doi.org/10.1007/978-3-540-70956-5

Klein, B., &Meißner, D. (2017). *Vulnerability of Inland Waterway Transport and Waterway Management on Hydro-meteorological Extremes*.

Lee, J. G., Han, J., &Whang, K. Y. (2007). Trajectory clustering: A partition-and-group framework. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 593–604. https://doi.org/10.1145/1247480.1247546

Lee, J. S., Son, W. J., Lee, H. T., &Cho, I. S. (2020). Verification of novel maritime route extraction using kernel density estimation analysis with automatic identification system data. *Journal of Marine Science and Engineering*, *8*(5). https://doi.org/10.3390/JMSE8050375

Maritime Research Institute Netherlands. (2012). *Encounter model*. Retrieved from http://www.iala-aism.org/wiki/iwrap/images/1/1f/Encounter.pdf

Meijer, B. H. (2018, July 24). Long drought puts pressure on Dutch shipping sector. *REUTERS*. Retrieved from https://www.reuters.com/article/us-weather-netherlands-shipping/long-drought-puts-pressure-on-dutch-shipping-sector-idUSKBN1KE274

Morris, B., &Trivedi, M. (2009). Learning trajectory patterns by clustering: Experimental studies and comparative evaluation. *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, *2009 IEEE*, 312–319. https://doi.org/10.1109/CVPRW.2009.5206559

Nordkvist, H. A. (2018). *An advanced method for detecting exceptional vessel encounters in open waters from high resolution AIS data* (Norwegian University of Science and Technology). https://doi.org/10.3850/978-981-11-2724-3_0038-cd

NOS NIEUWS. (2018, August 11). Binnenvaart houdt nog weken last van laagwater in rivieren. *NOS*. Retrieved from https://nos.nl/artikel/2245692-binnenvaart-houdt-nog-weken-last-van-laagwater-in-rivieren.html

Pallotta, G., Vespe, M., &Bryan, K. (2013). Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. *Entropy*, *15*(6), 2218–2245. https://doi.org/10.3390/e15062218

Pappalardo, L., Simini, F., Barlacchi, G., &Pellungrini, R. (2019). *scikit-mobility: a Python library for the analysis, generation and risk assessment of mobility data*. Retrieved from http://arxiv.org/abs/1907.07062

Portugués, E. G. (2021). Multivariate kernel density estimation. Retrieved from https://bookdown.org/egarpor/NP-UC3M/kde-ii-mult.html

Rijkswaterstaat. (n.d.-a). Verkeersinformatie: AIS. Retrieved August8, 2020, from Rijkswaterstaat website: https://www.rijkswaterstaat.nl/zakelijk/verkeersmanagement/scheepvaart/scheepvaartverkeersbegeleiding/river-information-services/automatic-identification-system/index.aspx

Rijkswaterstaat. (n.d.-b). Waal. Retrieved February8, 2021, from https://www.rijkswaterstaat.nl/water/vaarwegenoverzicht/waal/index.aspx

Rijkswaterstaat. (2018). Nog steeds weinig aanvoer water in grote rivieren. Retrieved September21, 2020, from Rijkswaterstaat website: https://www.rijkswaterstaat.nl/nieuws/2018/10/nog-steeds-weinig-aanvoer-water-op-grote-rivieren.aspx

Rijkswaterstaat. (2020a). Bathymetrie Nederland. Retrieved February22, 2021, from https://maps.rijkswaterstaat.nl/geoweb55/index.html?viewer=Bathymetrie_Nederland

Rijkswaterstaat. (2020b). IVS Next. Retrieved from https://www.rijkswaterstaat.nl/zakelijk/verkeersmanagement/scheepvaart/scheepvaartverkeersbegeleiding/ivs-next/index.aspx

Rong, H., Teixeira, A. P., &Guedes Soares, C. (2020). Data mining approach to shipping route characterization and anomaly detection based on AIS data. *Ocean Engineering*, *198*(August 2019), 106936. https://doi.org/10.1016/j.oceaneng.2020.106936

Scholten, A., Rothstein, B., &Pistocchi, A. (2017). Navigation on the Danube - Limitations by low water levels and their impacts. In *Publications Office of the European Union*. https://doi.org/10.2788/236234

Sheng, P., &Yin, J. (2018). Extracting shipping route patterns by trajectory clustering model based on Automatic Identification System data. *Sustainability (Switzerland)*, *10*(7). https://doi.org/10.3390/su10072327

Shu, Y., Daamen, W., Ligteringen, H., &Hoogendoorn, S. P. (2017). Influence of external conditions and vessel encounters on vessel behavior in ports and waterways using Automatic Identification System data. *Ocean Engineering*, *131*(December 2016), 1–14. https://doi.org/10.1016/j.oceaneng.2016.12.027

Smith, M. J.de, Goodchild, M. F., &Longley, P. A. (2018). *Geospatial Analysis A Comprehensive Guide to Principles Techniques and Software Tools* (6th ed.). Retrieved from https://spatialanalysisonline.com/HTML/index.html

the Department of Waterways, the Provinces, Rotterdam Harbour, Groningen Seaports, Amsterdam Harbour, Koninklijke Schuttevaer, …SRN. (n.d.). *The blue sign*. Retrieved from https://www.varendoejesamen.nl/storage/app/media/downloads/EN_the-blue-sign.pdf

Transport Online. (2016, December 22). Lage waterstanden op grote rivieren. *Transport Online*. Retrieved from https://www.transport-online.nl/site/77589/lage-waterstanden-op-grote-rivieren/

Verschuren, D. J. J. (2020). *Effects of drought on the traffic capacity of the river Waal and the occurrence of congestion.* (Delft University of Technology). https://doi.org/https://doi.org/10.4121/12982232

Wang, Zhaochen, &Yin, J. (2020). Risk assessment of inland waterborne transportation using data mining. *Maritime Policy and Management*, *47*(5), 633–648. https://doi.org/10.1080/03088839.2020.1738582

Wang, Zuchao, Lu, M., Yuan, X., Zhang, J., &Wetering, H.Van De. (2013). Visual traffic jam analysis based on trajectory data. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2159–2168. https://doi.org/10.1109/TVCG.2013.228

Węglarczyk, S. (2018). Kernel density estimation and its application. *Proceeding of the International Conference on Electrical Power Quality and Utilisation, EPQU*, *00037*. https://doi.org/10.1109/EPQU.2011.6128915

Wu, X., Rahman, A., &Zaloom, V. A. (2018). Study of travel behavior of vessels in narrow waterways using AIS data – A case study in Sabine-Neches Waterways. *Ocean Engineering*, *147*(December 2016), 399–413. https://doi.org/10.1016/j.oceaneng.2017.10.049

Xiao, F., Ligteringen, H., VanGulijk, C., &Ale, B. (2015). Comparison study on AIS data of ship traffic behavior. *Ocean Engineering*, *95*, 84–93. https://doi.org/10.1016/j.oceaneng.2014.11.020

Yan, Z., Xiao, Y., Cheng, L., He, R., Ruan, X., Zhou, X., …Bin, R. (2020). Exploring AIS data for intelligent maritime routes extraction. *Applied Ocean Research*, *101*(May), 102271. https://doi.org/10.1016/j.apor.2020.102271

Yuan, G., Sun, P., Zhao, J., Li, D., &Wang, C. (2017). A review of moving object trajectory clustering algorithms. *Artificial Intelligence Review*, *47*(1), 123–144. https://doi.org/10.1007/s10462-016-9477-7

Zhou, Y., Daamen, W., Vellinga, T., &Hoogendoorn, S. P. (2019). Ship classification based on ship behavior clustering from AIS data. *Ocean Engineering*, *175*(February), 176–187. https://doi.org/10.1016/j.oceaneng.2019.02.005

7. APPENDIX I

Table 1: List of excluded data

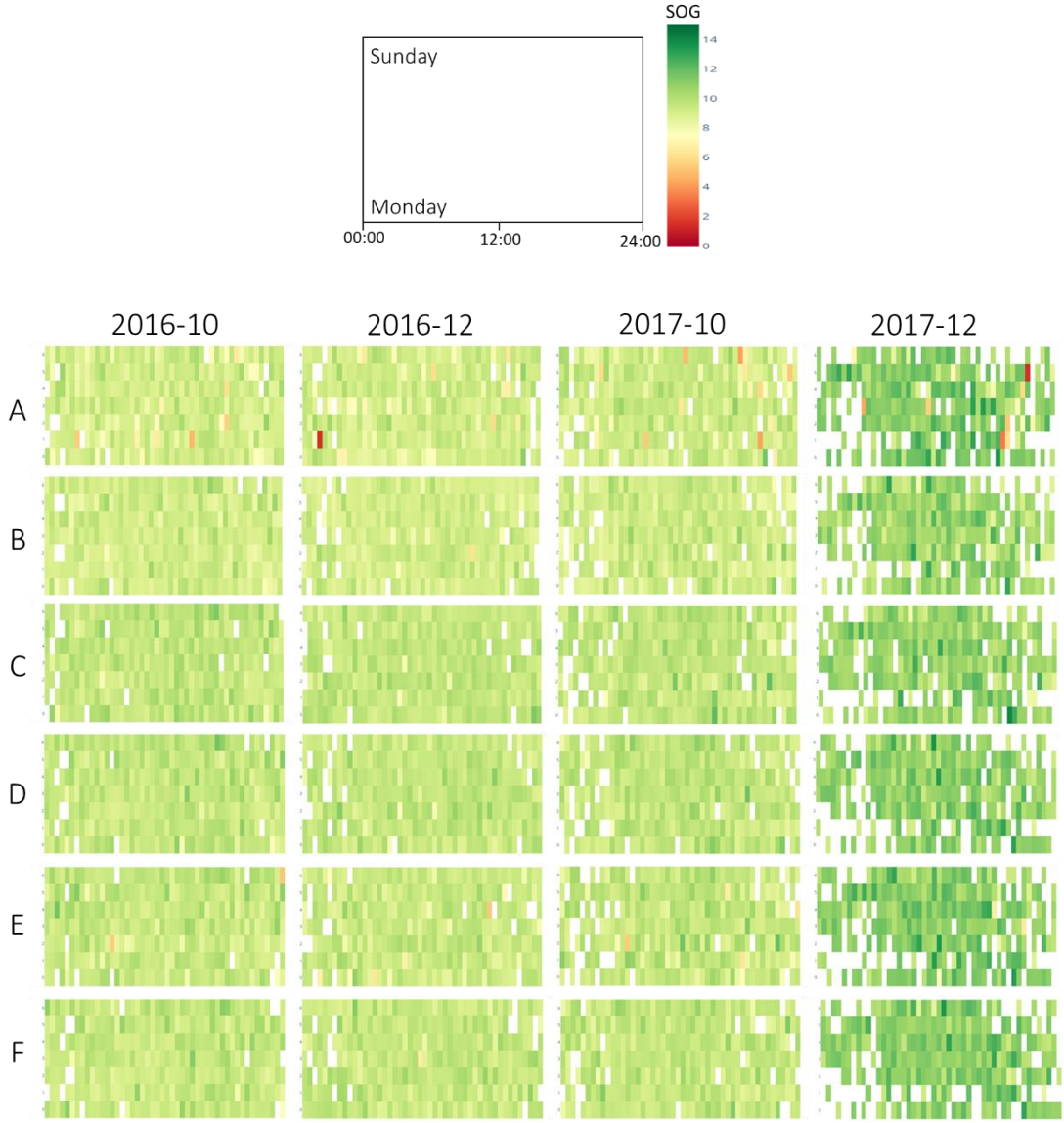
Size (rows)	2016-10-01 to 2016-10-31	2016-12-01 to 2016-12-20	2017-10-01 to 2017-10-31	2017-12-01 to 2017-12-31
Original size	35,064,275	31,179,509	29,656,997	29,840,546
SOG = 0	19,651,309	17,019,674	16,715,936	19,551,718
Vesseltype < 20	1,077,372	7,501,066	933,127	1,086,664
After cleaning	14,928,190	10,545,315	12,556,967	10,000,559
	42.57%	33.82%	42.34%	33.51%

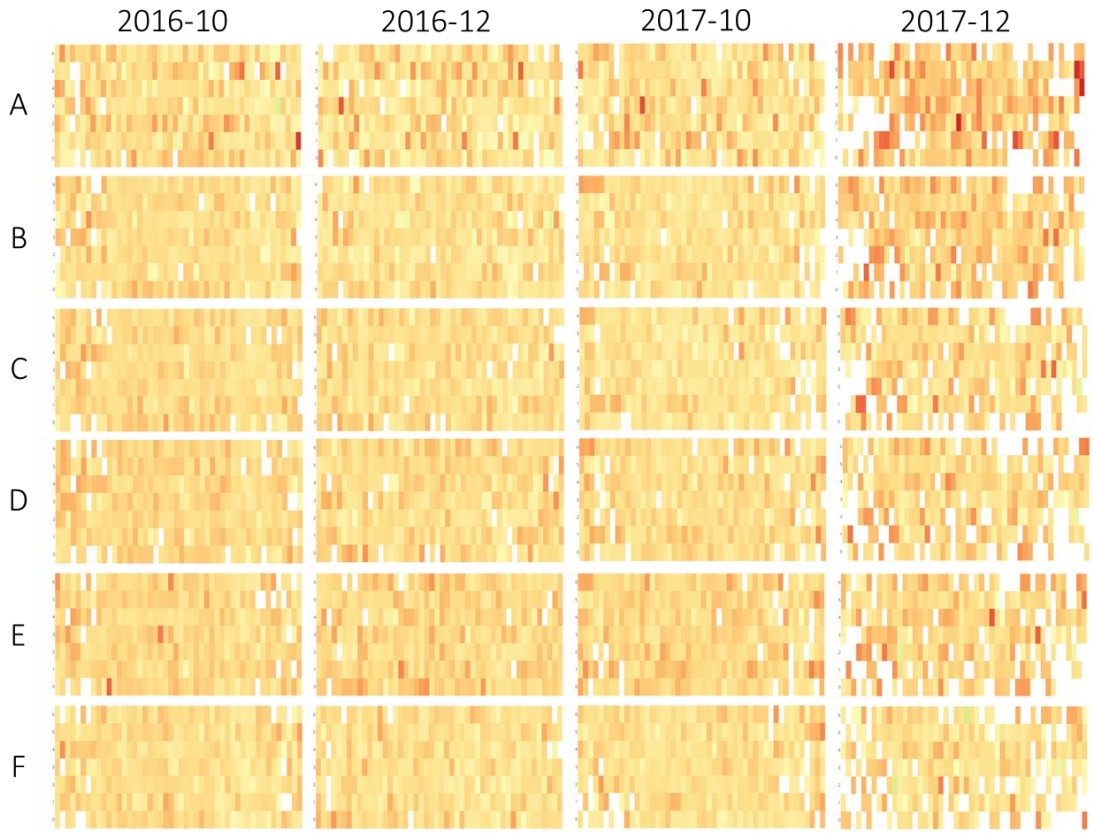
Table 2: Number of each vessel category

Vessel category	Vesseltype code	Period	No. of rows	Of original dataset
Wing in ground	20 – 29	2016-10	345,941	2.32%
		2016-12	162,538	1.54%
		2017-10	293,655	2.34%
		2017-12	202,181	2.02%
Fishing	30	2016-10	1,818	0.01%
		2016-12	34	0.00%
		2017-10	629	0.01%
		2017-12	4,315	0.04%
Towing	31, 32	2016-10	24,264	0.16%
		2016-12	6,294	0.06%
		2017-10	13,838	0.11%
		2017-12	11,301	0.11%
Dredging or underwater ops	33	2016-10	82,618	0.55%
		2016-12	35,547	0.34%
		2017-10	57,202	0.46%
		2017-12	18,877	0.19%
Diving ops	34	2016-10	0	0.00%
		2016-12	0	0.00%
		2017-10	3,603	0.03%
		2017-12	0	0.00%
Military ops	35	2016-10	4,348	0.03%
		2016-12	0	0.00%
		2017-10	29	0.00%
		2017-12	0	0.00%
Sailing	36	2016-10	175	0.00%
		2016-12	1,203	0.01%
		2017-10	0	0.00%
		2017-12	0	0.00%
Pleasure Craft	37	2016-10	5,358	0.04%
		2016-12	2,067	0.02%
		2017-10	2,409	0.02%
		2017-12	4,042	0.04%
Reserved	38, 39	2016-10	1,326	0.01%
		2016-12	0	0.00%
		2017-10	0	0.00%
		2017-12	0	0.00%
High speed craft (HSC)	40 – 49	2016-10	7,543	0.05%
		2016-12	533	0.01%
		2017-10	5,855	0.05%
		2017-12	671	0.01%
Pilot Vessel	50	2016-10	0	0.00%
		2016-12	0	0.00%
		2017-10	0	0.00%

Vessel category	Vesseltype code	Period	No. of rows	Of original dataset
		2017-12	0	0.00%
Search and Rescue vessel	51	2016-10	0	0.00%
		2016-12	0	0.00%
		2017-10	0	0.00%
		2017-12	0	0.00%
Tug	52	2016-10	63,324	0.42%
		2016-12	36,905	0.35%
		2017-10	73,065	0.58%
		2017-12	66,274	0.66%
Port Tender	53	2016-10	0	0.00%
		2016-12	0	0.00%
		2017-10	0	0.00%
		2017-12	0	0.00%
Anti-pollution equipment	54	2016-10	0	0.00%
		2016-12	3,301	0.03%
		2017-10	2,658	0.02%
		2017-12	1,784	0.02%
Law Enforcement	55	2016-10	85,286	0.57%
		2016-12	54,437	0.52%
		2017-10	80,450	0.64%
		2017-12	70,588	0.71%
Spare - Local Vessel	56 – 57	2016-10	2	0.00%
		2016-12	0	0.00%
		2017-10	0	0.00%
		2017-12	0	0.00%
Medical Transport	58	2016-10	4,900	0.03%
		2016-12	647	0.01%
		2017-10	2,652	0.02%
		2017-12	1,663	0.02%
Noncombatant ship	59	2016-10	0	0.00%
		2016-12	336	0.00%
		2017-10	3,373	0.03%
		2017-12	0	0.00%
Passenger	60 – 69	2016-10	534,777	3.58%
		2016-12	317,540	3.01%
		2017-10	629,891	5.02%
		2017-12	437,483	4.37%
Cargo	70 – 79	2016-10	8,571,369	57.42%
		2016-12	6,763,012	64.13%
		2017-10	6,728,616	53.58%
		2017-12	5,191,221	51.91%
Tanker	80 – 89	2016-10	3,335,775	22.35%
		2016-12	2,085,783	19.78%
		2017-10	3,113,204	24.79%
		2017-12	2,716,248	27.16%
Other type	90 – 99	2016-10	1,845,752	12.36%
		2016-12	1,063,639	10.09%
		2017-10	1,530,808	12.19%
		2017-12	1,273,813	12.74%

8. Appendix II




2016-10
2016-12
2017-10
2017-12
A
B
C
D
E
F

9. APPENDIX III

Python libraries and modules.

```
import urllib

import os

import numpy as np

import pandas as pd

from geographiclib.geodesic import Geodesic

from datetime import datetime, timedelta
```

Conform data to 30-second interval.

```
new_id = df['new_id'].unique().tolist()

cols = list(df)

list30s = []

df30s = pd.DataFrame(columns=cols)

df['t']= df['t'].astype('datetime64[s]')

date_index = pd.date_range(start='2016-10-01', end='2016-10-31
23:59:59', freq='30S')


for i in new_id:

    if len(df[df['new_id'] == i]) > 0:

        records = df[df['new_id'] == i]

        records = records.drop_duplicates(subset='t',keep='first')

        records.set_index('t', inplace=True)

        records = records.reindex(date_index, method='nearest',
tolerance=timedelta(seconds=5))

        records.dropna(subset=['sog', 'cog', 'longitude',
'latitude'],inplace=True)

        records.reset_index(inplace=True)

        # store DataFrame in list

        list30s.append(records)


df30s = df30s.append(list15s, True)
```

```
df30s.drop(columns='t', inplace=True)

df30s.rename(columns={'index':'t','timestamplast':'original_t'},
inplace=True)

df30s[['mmsi', 'vesseltype', 'new_id', 'traj_id']] = df15s[['mmsi',
'vesseltype', 'new_id', 'traj_id']].astype(int)

df30s['original_t'] = df30s['original_t'].astype('datetime64[s]')

df30s.sort_values(by='t', inplace=True)

df30s.reset_index(drop=True, inplace=True)
```

Detect encounter.

```
encounter = []

cols = ['t_0', 'original_t_0', 'original_t_1',

        'new_id_0', 'traj_id_0',

        'lat_0', 'lon_0',

        'cog_0',

        'new_id_1', 'traj_id_1',

        'lat_1', 'lon_1',

        'cog_1',

        'distance']


df = pd.DataFrame(columns=cols)


# The maximum distance to categorize ship encountering case

length_max = 300


for i in range(len(up)):

    df_1 = down[down['t'] == up['t'].iloc[i]]

    lat_0 = up['latitude'].iloc[i]

    lon_0 = up['longitude'].iloc[i]

    t_0 = up['t'].iloc[i]

    original_t_0 = up['original_t'].iloc[i]
```

```
    new_id_0 = up['new_id'].iloc[i]

    traj_id_0 = up['traj_id'].iloc[i]

    cog_0 = up['cog'].iloc[i]


    if len(df_1) == 0:

        continue


    for x in range(len(df_1)):

        original_t_1 = df_1['original_t'].iloc[x]

        lat_1 = df_1['latitude'].iloc[x]

        lon_1 = df_1['longitude'].iloc[x]

        new_id_1 = df_1['new_id'].iloc[x]

        traj_id_1 = df_1['traj_id'].iloc[x]

        cog_1 = df_1['cog'].iloc[x]

        distance = Geodesic.WGS84.Inverse(lat_0, lon_0, lat_1,
lon_1)['s12']

        if distance <= length_max:

            values = [t_0, original_t_0, original_t_1, new_id_0,
traj_id_0, lat_0, lon_0, cog_0,

                      new_id_1, traj_id_1, lat_1, lon_1, cog_1,
distance]

            new_row = dict(zip(cols, values))

            encounter.append(new_row)


df = df.append(encounter, True)
```

Classify encounter types.

```
# Angle from heading of own ship to position of other ship

def alpha(df):

    lat_0 = df['lat_0']

    lon_0 = df['lon_0']
```

```
    lat_1 = df['lat_1']
    lon_1 = df['lon_1']
    azi1 = Geodesic.WGS84.Inverse(lat_0, lon_0, lat_1, lon_1)['azi1']
    cog_0 = df['cog_0']
    if type(cog_0) == str:
        if azi1 < 0:
            return 360 + azi1
        else:
            return azi1
    else:
        if azi1 < 0:
            azi1 = 360 + azi1
        if cog_0 >= azi1:
            return 360 - (cog_0 - azi1)
        else:
            return azi1 - cog_0

# Decide which side of the ship were met by the other ships
def which_side(df):
    if (df['alpha'] >= 350) | (df['alpha'] < 10):
        return 'head-on'
    if (df['alpha'] >= 10) & (df['alpha'] <= 112.5):
        return 'starboard'
    if ((df['alpha'] > 112.5) & (df['alpha'] <= 180)):
        return 'met by starboard'
    if ((df['alpha'] > 180) & (df['alpha'] < 247.5)):
        return 'met by port'
    if (df['alpha'] >= 247.5) & (df['alpha'] <= 350):
        return 'port'
```

```
# Apply the function to the dataframe
df['alpha'] = df.apply(alpha, axis=1)
df['encounter_side'] = df.apply(which_side, axis=1)
```

10. APPENDIX IV

Port side encounter heatmaps. White boxes are time windows when there was a big water level difference.