A Study of Satellite AIS Data and the Global Ship Traffic Through the Singapore Strait

Bjørnar Brende Smestad

Preface

This is a master thesis from the Department of Marine Technology at the Norwegian University of Science and Technology, and is a part of my Master Degree in Engineering & ICT with a specialisation in Marine Design & Logistics. It was carried out during the spring semester of 2015. Although I utilise my background in ICT in this thesis, it is directed towards a general technical audience.

<div align="center">

Trondheim, 2015-06-10

Bjørnar Brende Smestad

</div>

Acknowledgment

First and foremost, I would like to thank my supervisor Bjørn Egil Asbjørnslett and my co-supervisor Ørnulf Jan Rødseth for all the effort and invaluable guidance throughout the work with this thesis.

Åsmund Tjora and Eirik Røsvik have read through this thesis, and provided useful feedback.

Lastly, Anne Line Stensjøen and my office mates has provided support, and been at great help during this work.

B.B.S.

Abstract

This thesis investigates the quality and utility of Automatic Identification System (AIS) data. AIS is primarily used as a tracking system for ships, but with the launch of satellites to collect these data new and previously untested possibilities are emerging. It is outlined how heuristics can be used to determine specific ship types on the basis of Satellite AIS data. This enables studies with AIS data to be conducted without the previously necessary use of commercial ship databases. The heuristics are used to determine the specific ship type of vessels trafficking the Singapore Strait, which shows how these methods can be employed. The ships sailing in the Singapore Strait are also tracked back to their origin, which shows a broader range for how Satellite AIS data can be employed. The thesis suggests that further work should use this type of data to, for instance do emission analyses for the ship traffic in Singapore Strait through a combination of the heuristics in this data and Satellite AIS data. A global or local study on the prevalence of slow steaming could also be conducted with Satellite AIS data, using the same methods used to analyse the traffic in the Singapore Strait. Heuristics can also be developed for a wider range of vessels and be tested on data from a longer time period.

Sammendrag

Denne masteroppgåva undersøkjer kvaliteten til, og gjev døme på bruk av Automatisk Identifikasjons System (AIS)-data. AIS blir først og fremst brukt som eit system for å spore skip, men no når det har blitt starta fleire prosjekt der satellittar blir nytta til å samle inn AIS-data, har det opna seg mange nye, og hittil u-utnytta moglegheiter for bruk av desse data. I oppgåva blir det vist korleis heurestikkar kan bli brukt til å fastsetje eit skip si skipsklasse frå desse Satellitt AIS-data. Denne innsikta opnar for at fleire studiar, med Satellitt AIS-data som basis, kan bli sett i gong, utan tilgong på kommersielle skipsdatabasar. Desse heurestikkane blir her brukt til å bestemme kva skip som segla gjennom Singaporestredet i ein gitt periode. Skipa som segla gjennom stredet blir også spora tilbake til opprinnelsestaden sin, ved hjelp av Satellitt AIS-data. Det blir foreslått fleire mulegheiter for framtidige studiar. Ein av disse er ei utslippsanalyse for skipa som seglar gjennom Singaporestredet. Dette kan bli gjort ved å nytte heurestikkane og det same datagrunnlaget som i denne oppgåva. Det kan også bli utført ein global, eller lokal, studie på utbreiinga av såkalla 'slow-steaming' med basis i Satellitt AIS-data, ved bruk av same metodar som blir nytta i analysane av trafikken i Singaporestredet. Utvikling av fleire heurestikkar, samt testing av eksisterande heurestikkar på eit enda større datagrunnlag, kan og være svært nyttig.

Contents

V Appendices   107

List of Figures

List of Tables

Acronyms

AIS   Automatic Identification System

COG   Course Over Ground

COLREG   The International Regulations for Preventing Collisions at Sea

DG   Dangerous goods

DWT   Deadweight Tonnage

ETA   Estimated Time of Arrival

GHG   Greenhouse Gas

HS   Harmful substances

IALA   International Association of Marine Aids to Navigation and
               Lighthouse Authorities

IMO   International Maritime Organization

ITU   International Telecommunication Union

LNG   Liquefied Natural Gas

LPG   Liquefied Petroleum Gas

MMSI   Maritime Mobile Service Identity

MP   Marine Pollutants

NTNU   Norwegian University of Science and Technology

NUC	Not Under Command
OOW	Officer Of the Watch
RIATM	Restricted In Ability To Manoeuvre
ROT	Rate Of Turn
RORO	Roll on/roll off
SOG	Speed Over Ground
SOLAS	International Convention for the Safety of Life at Sea
SQL	Structured Query Language
ULCC	Ultra Large Crude Carriers
TEU	Twenty foot Equivalent Units
UTC	Coordinated Universal Time
VHF	Very High Frequency
VLCC	Very Large Crude Carriers
VTS	Vessel Traffic Service
WIG	Wing In Ground

Part I

Background

Chapter 1

Introduction

The Automatic Identification System (hereafter AIS) was introduced to monitor ship traffic and avoid accidents, especially in highly trafficked waterways. Lately, satellites have been launched to receive AIS data from all over the world in order to extend the coverage area beyond what was possible with the exclusively land based stations. This is a new development, which provides new opportunities for monitoring and analysing global ship traffic.

1.1 Objectives

This master thesis will investigate whether the large amount of Satellite AIS data is reliable, and compare this type of data to land based AIS data. Two large studies that have used these data as a base of their research will be presented, as well as other relevant research concerning AIS and Satellite AIS data. The dissertation will provide its own examples on how Satellite AIS data can be utilised for analysis of ship traffic. As a case study, the traffic through Singapore Strait will be analysed. It will investigate whether Satellite AIS data can be used to track ships entering Singapore Strait back to their port of origin.

Heuristics that can identify the specific ship type of a ship from its AIS data will be developed. This specific ship type is not specified in the AIS data, and is only available by looking up each ship in commercial databases. The specific ship type will differentiate between for example Liquid Natural Gas (LNG) Vessels and Oil Tankers, both solely registered as tankers in the AIS data. This exploration into AIS data and Satellite AIS

data done in this thesis will be beneficial for both ship operation analyses and emission analyses.

1.2 Approach

Raw Satellite AIS (hereafter S-AIS) data spanning the time period 1st May to 15th September, 2014 in addition to, AIS data from the Singapore Strait, spanning 25th May to 25th November have been retrieved through the Norwegian Coastal Service and the SESAME Straits project[1]. These data were decoded and indexed in a SQLite[2] database for easier access. This was done using a Python[3] program found in Appendix C. This program also utilised a library for decoding AIS messages called AIS Parser SDK v1.10[4].

The data was later inspected using small Python scripts, to find obviously erroneous data. These data were either deleted from the database, or omitted from further analyses by restriction clauses in the code.

The heuristics for identifying a ship's specific ship type were developed using these data. By using template groups of ships inside a known ship type, length and breadth intervals were generalised. Maximum or minimum draught or change of draught, as well as a limit for either maximum or minimum speed was developed. This was done using an incremental process, where these constraints were gradually increased and decreased to identify the maximum number of vessels, with the lowest margin of error. These characteristics, along with the general AIS shiptype formed the heuristic. These heuristics were then used to identify the specific ship types from the S-AIS data. Manual lookup in online ship databases was performed to inspect the precision of the heuristic, which was defined as the number of correctly identified vessels versus the number of identified

---

[1]SESAME Straits (Secure, Efficient and Safe maritime traffic Management in the Straits of Malacca and Singapore) is a part of a multinational project to develop and implement innovative new traffic management strategies for congested waterways worldwide. The SESAME Straits project intends to change traditional Vessel Traffic Services (VTS), which today only provide monitoring and advisory services locally.

[2]SQL, Structured Query Language is a programming language specifically made to retrieve data from databases. Its development is controlled by the International Electrotechnical Comission and the International Organization for Standarization, ISO. SQLite is a free software library that powers databases that use SQL (`http://www.sqlite.org`).

[3]Python is a programming language that can be found at `https://www.python.org/`.

[4]This is a software library in a various of programming languages, notably Python, C and Java that provides methods for decompiling AIS messages. It is distributed with openly, with a BSD license, which are licenses with a minimum of restrictions on the use. See `https://github.com/bcl/aisparser`, (Lane, 2006).

vessels inside each ship type and size category.

1.3 Literature Survey

There have been relatively few studies that have used AIS data, and especially S-AIS data on a global level. Using satellites to receive AIS data is still a relatively new concept. The launch of the first Norwegian AIS Satellite happened as recently as in 2010. This is discussed further in Section 2.2. However, there have been two large, and thorough studies that have demonstrated some of the potential of S-AIS data, which is described into detail in the following sections. Other relevant studies will be cited and discussed later in the thesis, where it is appropriate.

Third Greenhouse Gas Study, International Maritime Organization

The Third Greenhouse Gas (GHG) study by Smith et al. (2014) built on and updated the conclusions of the International Maritime Organization's, IMO, Second IMO Greenhouse Gas study by Buhaug et al. (2009). It was conducted by a consortium led by University College London, with a range of international partners. The aim for the study was to give updated figures on shipping emissions during the time period of 2007 to 2012, to identify fuel use trends and break these down to specific ship types, as well as to give a foresight on different future scenarios for the maritime greenhouse gas emissions. Unlike the second IMO GHG study, this latter study could utilise Satellite AIS data to produce better estimates on shipping emissions. These data were used to get precise activity measures and better emissions estimates for each ship in service for their data period. This was aggregated to find the total emissions for each ship type. In the previous study, emissions were estimated by using annual average activities numbers for different ship types, and it thus had a lower precision rate than the latter study.

In the most recent study, it was estimated that international shipping emitted 2.6% of global CO_2 emissions, while shipping in total emitted 3.1% of the global CO_2 emissions over the years of 2007-2012. The breakdown of emissions from each year shows how it varied, with a slight downturn in emissions from shipping compared to total emissions during the period (from 2.8% of the global CO_2 emissions in 2007 to 2.2% in 2012, and 2.6% of the global GHG emissions in 2007 to 2.4% in 2012). The decrease of the share of

global emissions from both local and global shipping was comparable.

The main ship types contributing a high volume of the fuel consumption, and thus the emissions, were oil tankers, container ships and bulk carriers. The decrease of emissions could be credited to a higher use of slow steaming and a low market activity following the financial crisis.

The average reduction in at-sea speed relative to design speed was 12%, which in turn contributed to an average reduction in daily fuel consumption of 27%. The reductions in daily fuel consumption were up to 50% for some oil tankers, while some container ships reduced it by more than 70%.

It must be noted that the great reduction in daily fuel consumption does not necessarily contribute to an overall decline in emissions from shipping, since for the same amount of goods to be shipped, more days at sea are required due to the slow speed. This, however, will heavily depend on the type of goods being shipped, and the individual fleet's gain from slow steaming.

Assessment of Shipping's Efficiency Using Satellite AIS Data

Smith et al. (2014) prepared a report as a part of the World Shipping Efficiency Indices project funded by the International Council on Clean Transportation called 'Assessment of Shipping's Efficiency Using Satellite AIS Data'. The objective of this report was to generate new knowledge and insight on the subject of shipping's technical and operational energy efficiency. The report was credited to be the first to use S-AIS data to analyse the energy efficiency of the world's shipping fleet. The study combined global S-AIS data from 2011 with technical ship data from sources like Clarksons World Fleet Register, and the Second IMO Greenhouse Gas Study (Buhaug et al., 2009).

Amongst the key findings in this report was that the average operating speeds in the world bulk fleet was approximately 10-15% lower, while the average operating speed for the container ship fleet was up to 25% lower than the average speeds presented in the Second Greenhouse Gas Study Buhaug et al. (2009) conducted with data from 2007. This speed reduction was attributed to the use of 'slow steaming', which has become more widespread since 2007.

This report concluded that S-AIS data should be regarded as a valuable and reliable source for knowledge. Such data is especially valuable when it comes to characterizing

heterogeneity and variability of a fleet's operation parameters as for instance operating speed, which can be used to calculate the operational energy efficiency.

The report presents several topics suitable for further exploration. One of these topics involves investigating whether variables such as charter type, nature of the fixture, operator or customer preferences can explain why the range of operational efficiency within a fleet is fairly high, which in turn can help to find incentives to get a more homogeneously efficient ship fleet.

Another topic pointed out in the report as suitable for further study is to do a longitudinal analysis in a quest to explain the influence fuel prices and freight rates can have on a fleet's operational efficiency. Although the study found that there had been more use of measures such as slow steaming to achieve a greater operational efficiency the latter years it covered, it did not go to the lengths of investigating whether the main factor was down to a difference in fuel price, freight rates or the increased focus on improving operational efficiency. However, to do such a extensive study it would be necessary to have data spanning several years, while this thesis only has analysed data from one specific year.

1.4 Organisation of the Thesis

This thesis is divided into four parts, with eight chapters.

Part I - Background

Chapter 1 introduces the objectives of this thesis, as well as the approach and the literature survey. Relevant literature is cited throughout this thesis, but the two main studies concerning Satellite AIS data are introduced here.

Chapter 2 presents the relevant theory behind this thesis. This includes the guidelines for use of AIS data and the data content and frequency of different AIS message types. This part is a continuation of work done in the Project Thesis that lead up to this Master Thesis. An introduction of Satellite AIS data, as well as the two AIS satellites, is also given here. Finally, a brief presentation of the ships that constitutes the world's fleet is given.

Part II - Analysis of Raw AIS Data

Chapter 3 gives an overview of the S-AIS data that forms the basis for much of the analyses in Part III. A discussion on the trustworthiness of S-AIS data is also carried out.

Chapter 4 presents AIS Data acquired from base stations in Singapore. As with the S-AIS data, an overview over the data content is provided.

Part III - Examples of Use of AIS Data

Chapter 5 forms a detailed example on how S-AIS data can be used to derive new knowledge. Heuristics for estimating ship and cargo type are presented for four main ship classes.

Chapter 6 illustrates how S-AIS data can be combined with AIS data from base stations, and be used to analyse the traffic travelling through the Singapore Strait.

Part IV - Discussion and Conclusion

Chapter 7 gives a discussion of the work done, results attained and choices made in this thesis. Note that there is discussion throughout this thesis, so this chapter will summarise the most important points.

Chapter 8 concludes the work done in this thesis in the light of the objectives.

Appendices

Appendix A provides the project description of this master thesis.

Appendix B list the AIS Data Contents.

Appendix C provides the python program to decode the AIS messages.

Appendix D shows how one month of AIS positions is plotted.

Appendix E shows how the maximum speed constraint is implemented.

Appendix F provides the program to find erroneous IMO numbers.

Appendix G includes all scatter plots found in this thesis, in full size, to improve the readability.

Chapter 2

Theory

In this chapter, the theory that underlies the work in this thesis is presented. In Section 2.1, AIS data, and the data content is presented. The concept of collecting AIS data using satellites is discussed in Section 2.2. Lastly, the world's fleet, and different ship and size classifications, are presented in Section 2.3.

2.1 AIS Data

Automatic Identification System (AIS) is a communication system that uses the maritime Very High Frequency (VHF) system. It consists of a protocol for communication that specifies the information that shall be transmitted, as well as the technological equipment that utilises this protocol. AIS enables automatic exchange of information from the vessel. This includes static data such as navigational details, dynamic data from the ships sensors, such as speed, and voyage-related information such as draught, destination and Estimated Time of Arrival (ETA). A typical use of AIS is to exchange information between vessels that are in the same area, to avoid high risk situations. It is also used in traffic management between a on shore station and the vessels. AIS data is gathered by AIS receivers, which can be found on board ships, on buoys and on land IALA (2011). Recently there have been launched satellites that gathers AIS data, and this data is labelled S-AIS data, see Section 2.2. The development of AIS is a joint project between the International Maritime Organization (IMO) and the International Association of Marine Aids to Navigation and Lighthouse Authorities (IALA) among others. The work to develop the Automatic Identification System was first initiated in 1994. In 1998, IMO

amended regulations about use of AIS to the International Convention for the Safety of Life at Sea (SOLAS) (IMO, 1974).

2.1.1 Guidelines for Use of AIS

SOLAS states that all ships of 300 gross tonnage and upwards engaged in international voyages, cargo ships of 500 gross tonnage and upwards not engaged on international voyages, as well as all passenger ships built after 2002, or operated after 2008[5], should have an Automatic Identification System (IMO, 2002).

2.1.2 Message Types

The International Telecommunication Union (ITU) have defined 27 different message types in their 'Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile frequency band' (ITU, 2014). Each message is labelled with its message ID. The most common message types is included in Table 2.1. Message type 1, 2, 3 and 4 will be referred to as Dynamic Messages, while message type 5 will be referred to as Static Messages.

Table 2.1: Most common AIS message types.

Message ID	Description (ITU, 2014)
1 and 2	Scheduled position report and assigned scheduled position report
3	Special position report and response to interrogation
4	Position, UTC, date, current slot number of base station and related vessel data report

2.1.3 AIS Data Content

The AIS data content is given in Table B.1, which can be found in Appendix B. The data content is given by the 'Guidelines for the onboard operational use of shipborne AIS' by IMO (IMO, 2002). In this thesis the following AIS data content is mainly used:

- Maritime Mobile Service Identity (MMSI)

- IMO number - see below

---

[5]This was amended in 2002 to be in 2004 instead of 2008 (SOLAS, 2011).

- Main dimensions - length and breadth of ship

- AIS ship type - see below

- Position - latitude and longitude

- Speed over ground - the ships speed relative to the ground as opposed to the water

- Current ship draught

- Ship name

MMSI

The Maritime Mobile Service Identity (MMSI) is a unique number that signifies the ship station that transmits the AIS messages. This number is only changed if there is a change of ownership for the ship in question.

IMO number

The IMO Resolution A.600(15) IMO (1987), which was made mandatory through SOLAS regulation XI/3 IMO (1974), regulates the use of IMO number as a ships identification. All ships over 100 gross tonnage, with exception to vessels solely engaged in fishing, ships without mechanical means of propulsion, pleasure yachts, ships engaged on special service, hopper barges, hydrofoils and hovercrafts, floating docks, ships of war and wooden ships, should be identified with an IMO number. This is a number that is assigned to the hull itself, and should follow the ships hull through its lifespan. The IMO number can be found in Static AIS Messages, and is a seven digit number. The validity of IMO numbers can be verified by its check digit. This is done by multiplying each digit with its position from right to left, with exception from its first rightmost digit. This is called the check digit. The results of these multiplications shall be summed, and the last digit in that sum shall correspond to the check digit in the IMO number. So for the IMO number 9652806, $9x7 + 6x6 + 5x5 + 2x4 + 8x3 + 0x2 = 156$, the rightmost digit of this sum is equal to the rightmost digit of the IMO number, and thus the IMO number is valid.

AIS Ship Type

The AIS ship type is reported as a double digit number between 10 and 99, where the first digit signifies the ship type, while the second digit signifies whether the cargo is dangerous, hazardous or pollutant. The AIS ship types is listed in Table 2.2.

Table 2.2: Ship types signified from first digit in AIS data, see full table in USCG (2012).

First digit	Ship type
1	Reserved for future use
2	WIG (Wing In Ground)
3	Other vessels
4	High speed carrier,
	or vessels < 100 Gross Tonnes
5	Special craft
6	Passenger ships > 100 Gross Tonnes
7	Cargo ships
8	Tankers
9	Other types of ship

2.1.4 Data Frequency

According to the 'Guidelines for the onboard operational use of shipborne AIS' by IMO (IMO, 2002) the data is sent in different rates. The static and voyage-related data is sent every 6 minutes or upon request, while the dynamic data is sent in different intervals according to the ship speed and status, as seen in Table 2.3.

Table 2.3: Ship status and general reporting interval (IALA, 2011).

Ship status	General reporting interval
Ship at anchor	3 min
Ship at 0-14 knots	12 sec
Ship at 0-14 knots and changing course	4 sec
Ship at 14-23 knots	6 sec
Ship at 14-23 knots and changing course	2 sec
Ship at >23 knots	3 sec
Ship at >23 knots and changing course	2 sec

2.2   S-AIS Data

Land based AIS receivers can detect AIS messages up to 40-50 nautical miles[6] off-shore (Skauen et al., 2013). Messages from ships outside of this zone will not be received. As an effort to extend Norwegian authorities maritime situational awareness, feasibility studies on a Satellite based AIS system were initiated. In 2005, researchers from The Norwegian Defence Research Establishment[7] published the first study investigating whether satellites could be used to gather AIS signals (Wahl et al., 2005). In 2007, a follow up study by Hoye et al. (2008), found that AIS signals could be detected by satellite based AIS receivers positioned in altitudes of up to 1000 km. However, since the AIS system initially was not designed for space based receivers, but rather to be a ship to ship communication based system, there were some problems. A satellite will have a much larger coverage area than AIS receivers were designed for, which could lead to interference problems between the AIS signals and the satellite. According to the study, the result could be that some AIS messages from ships would not be detected by the satellite. In Figure 2.1, the problem is illustrated by displaying the maximum number of vessels inside a single area that the early satellite systems could handle. This shows that high traffic areas such as Singapore Strait could be a problem. Figure 2.2 shows the predicted detection performance for AISSat-1, the first Norwegian AIS satellite. Most of the world has a high detection performance, but North European waters, south-east Asia and the western parts of the United States have far poorer predicted performance. This is due to the high amount of traffic, in combination with a low satellite passing rate in these areas.

In 2010, AISSat-1 was launched. This satellite is in a sun-synchronous polar orbit[8] at 630 km altitude (Eriksen et al., 2010). The satellite forwards its AIS messages to Svalbard Ground Station at each passing. Eriksen et al. (2010) states that over a time span of 24 hours, areas along the equator is covered two to three times, while the High North and South is covered up to 15 times. The path of one of the satellites over a 25 hour time period is visualised in Figure 2.3. In 2013, AISSat-2 was launched to give extended coverage. This gave a higher update rate to the Svalbard Ground Station, as well as a

---

[6] 40-50 nautical miles is approximately 74-92 km.
[7] Forsvarets Forskningsinstitutt in Norwegian.
[8] A polar orbit is an orbit where the satellite travels from the north to the South Pole. A sun-synchronous polar orbit means that the satellite passes the same area at approximately the same time of the day, each day.

Ship reporting interval	Max number of ships
2 s	150
6 s	450
10 s	750
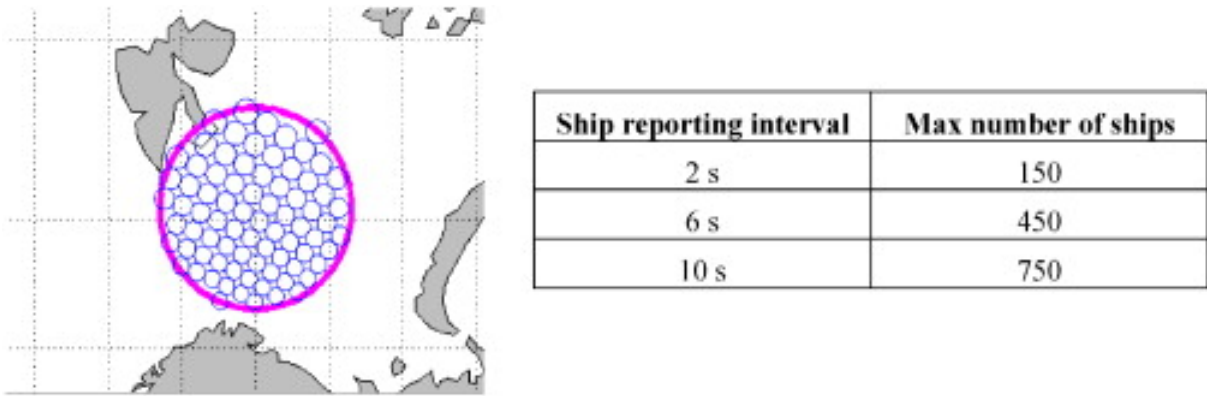
Figure 2.1: The left figure shows the example of a field of view (the purple circle), with several organised areas (the blue circles). The table shows the maximum number of ships the AIS system can handle inside each of those organised areas, depending on the AIS reporting interval. This figure was presented in Hoye et al. (2008)
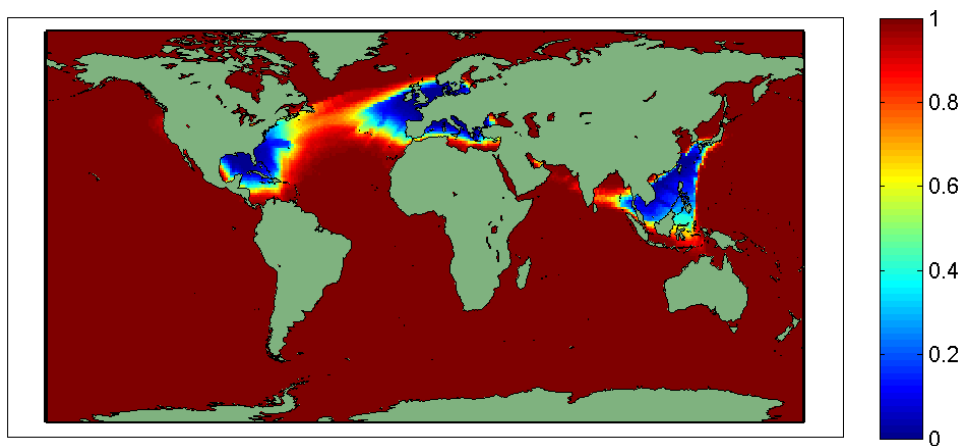
Figure 2.2: Predicted detection performance for AISSat-1, for an AIS message update interval of one day. Red means 100% detection probability, while blue signifies 0% detection probability. Eriksen et al. (2010)

higher global detection rate.

In Figure 2.4, which is copied from Helleren et al. (2012), all observations for September 2011 from AISSat-1 are plotted. In Figure 2.5 all observations for August[9] 2014 from AISSat-1 and AISSat-2 are plotted. When plotting this, the data had to be divided into two parts, as the full amount was too big to be plotted at the same time. The total number of points in this plot was above 48 million, which exceeds the amount a normal computer can handle at once. The code for this plot is shown in Appendix D. In the Indian Ocean and off the east coast of Africa, there is a visible increase of observations between these two images. One can possibly attribute the low traffic in 2011 to the

---

[9]As we only had S-AIS data for 1st May to 15th September 2014, we chose to plot all observations in August.
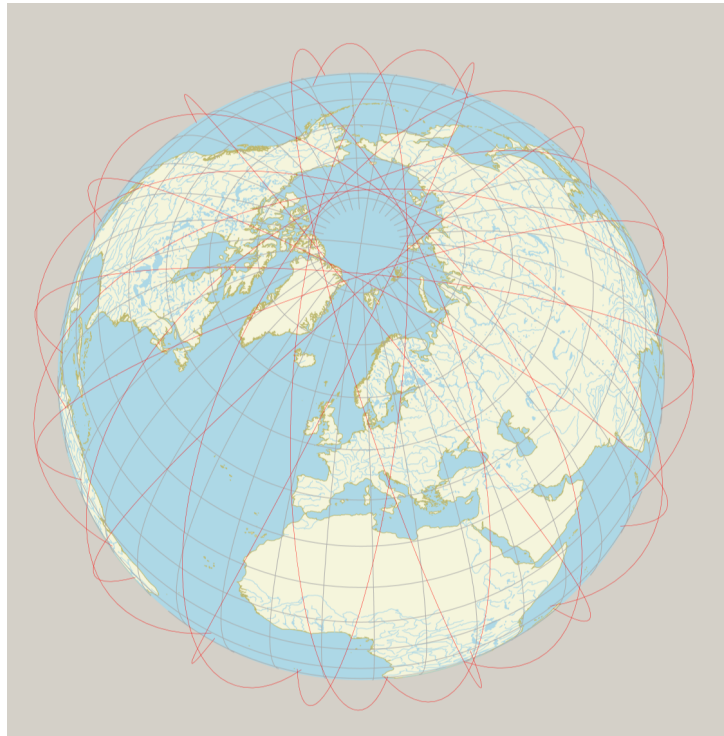
Figure 2.3: The red line indicates the travel path for one of the satellites inside a 24 hour time period. This figure was produced by Ørnulf Jan Rødseth.

relative high number of pirate attacks of the East African coast. Since the plots have data from two different months, August and September, some seasonal variety can also be an explanation. The most likely explanation, however, is that the two satellites have increased the coverage in this area. There is also a larger amount of observations across the pacific region. As described earlier, the satellites can have interference problems in areas with a high density of ships. This can be observed from these plots. There are few observations in the southern part of the North Sea, as well as in the South and East China sea and the western part of the Gulf of Mexico in both figures, even though it is expected that these regions to be highly trafficked. This corresponds with Figure 2.2, which predicted that the detection performance for AISSat-1 would be the least in these regions. Note that Figure 2.4 and Figure 2.5 are probably plotted using different tools, so there can be some differences due to different line widths etc.
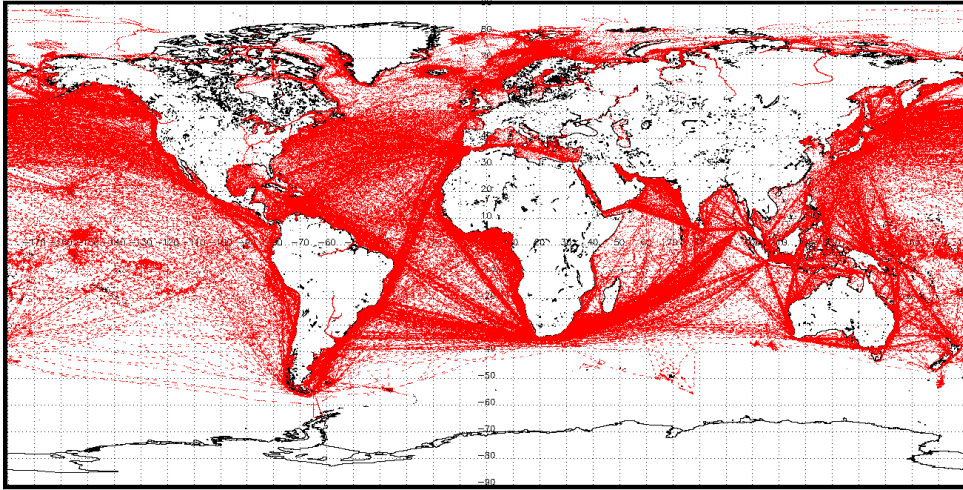
Figure 2.4: Global AIS observations from AISSat-1 for September 2011 (Helleren et al., 2012).
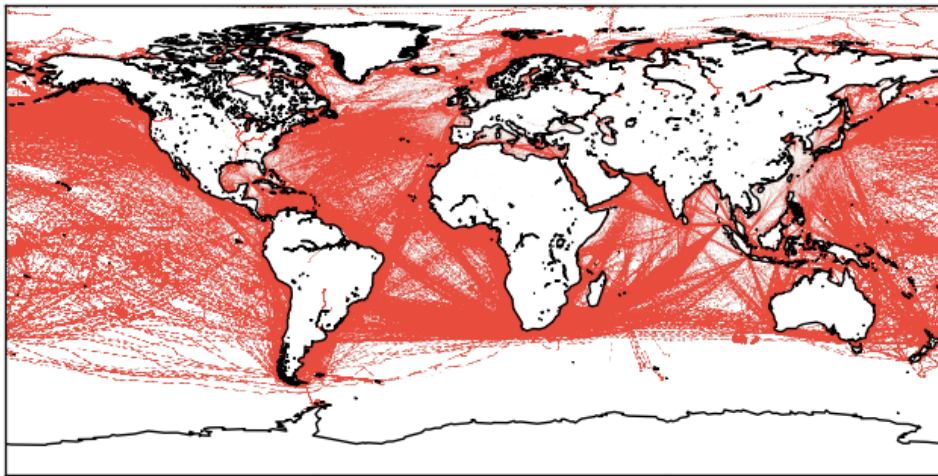
Figure 2.5: Global AIS observations from AISSat-1 and AISSat-2 for August 2014.

2.3 The World Fleet

To quantify the distribution of the world fleet, data from Shipping Intelligence Weekly, which is published by Clarkson Research Services Limited (Mantell et al., 2014), is utilised. The publication is from 2nd May 2014. This corresponds well with the time period of our S-AIS data, where the earliest data is from 1st May 2014. Note that there is a continuous flow of new vessels into the world fleet and old vessels out of it. Clarkson defined the world cargo fleet as the total fleet of tankers, cargo and offshore ships.

Table 2.4: Number of vessels in the world fleet 2. May 2014 (Mantell et al., 2014) p.14.

World Cargo fleet	57,829
Total World fleet	88,483

The world cargo fleet constitutes about 65% of the world fleet, where the remaining is constituted by vessels like passenger ships, tugs and other work boats, pleasure boats etc. In Table 2.5 the different vessel categories that constitutes the World Cargo Fleet is shown. The largest category is multi purpose- and general cargo ships. Multi purpose vessels are essentially cargo vessels that can carry a wide range of cargo types, and they often carry different kinds of cargo at the same time. Because of this wide range of cargo types and use, it is hard to distinguish them as one group. The general cargo ships are cargo ships that are not included in the other cargo categories. This is typically smaller vessels. The next largest category is Bulk Carriers. These can be divided into sub categories, Handysize, Handymax, Panamax and Capesize. This is outlined in the next section.

2.3.1 Size Classifications

Cargo ships and tankers are typically classified into different size categories, which can correspond to the maximum dimensions of important seaways and ports, such as the Panama Canal[10] and the Suez Canal[11]. Note that these size classifications are a arbitrary way to group ships, and not necessarily technical standards. These dimensions should only act as guidance to the approximate length and breadth for ships inside the different categories.

Oil Tankers

According to Smith et al. (2014), oil tankers can be divided into seven size categories, with corresponding capacity ranges. These are presented, together with the average length and breadth dimensions for each dead weight (DWT) category which is acquired from ClarksonsGroup (2015) in Table 2.6. Note that the different ship categories are presented

---

[10]The Panamal Canal acts as a link between the Pacific Ocean and Atlantic Canal. At the narrowest it is 33.5 meter wide. The maximum allowed draught is 12.5 meters.

[11]The Suez Canal connects the Mediterranean Sea and the Red Sea. The maximum allowed draught is 20.1 meters, there is no breadth restriction.

Table 2.5: World cargo fleet by vessel type (Mantell et al., 2014). Ordered by number of vessels.

Ship type	Number of vessels
Multi-purpose and general cargo ships	18,303
Bulk Carriers	10,053
Handysize	3,095
Handymax	3,008
Panamax	2,405
Capesize	1,590
Oil Tankers (<10,000 dwt)	7,456
Oil Tankers (>10,000 dwt)	5,830
Sub panamax	3,401
Aframax	884
UL&VLCC	624
Suezmax	495
Panamax	416
Offshore (AHTS/PSV)	5,129
Containerships	5,102
Sub panamax	3,019
Post Panamax	1,208
Panamax	875
Reefers	1,438
Ro-Ro vessels	1,311
LPG Carriers	1,258
LNG Carriers	388
Others	1,561

with the average length and breadth for ships inside this capacity range, so there could be some variation of these dimensions inside the same size category.

Table 2.6: Capacity range, size category and average length and breadth for oil tankers (ClarksonsGroup, 2015).

Capacity range (DWT)	Size cateogry	Average Length	Average Breadth
< 10,000	Small	< 138.1	< 21.6
10,000-19,999	Handysize	138.1	21.6
20,000-59,999	Handymax	165.6-183.4	26.1-32.2
60,000-79,999	Panamax	227.2	32.7
80,000-119,999	Aframax	244.6	42.6
120-199,999	Suezmax	273.8	47.9
200,000+	VLCC, ULCC	332.3+	59.5+

Container Ships

Container ship capacity is typically measured in TEU (Twenty foot Equivalent Units) where twenty foot is the standard length of a freight container. Rodrigue and Ashar (2012) has categorised these ships as seen in Table 2.7. Note that in Table 2.5, Mantell et al. (2014) grouped container ships into only three size categories, sub-, post- and panamax. Possibly, all the categories above panamax are grouped as the same post panamax category.

Table 2.7: Capacity range, size category and average length and breadth for container ships.

Capacity range (TEU)	Size cateogry	Length	Breadth
3,000-4,500	Panamax	250-290	32
4,000-8,000	Post Panamax	285-300	40-43
12,500	New Panamax	366	49
15,000	Post New Panamax	397	56
18,000	Triple E	400	59

Bulk Carriers

According to Smith et al. (2014), Bulk Carriers are often divided into six main categories. As with the oil tankers these categories are coupled this with average length and breadth

for each ship class from ClarksonsGroup (2015). The results can be seen in Table 2.8.

Table 2.8: Capacity range, size category and average length and breadth for Bulk Carriers.

Capacity range (DWT)	Size cateogry	Length	Breadth
10,000-39,999[12]	Handysize	139.4-181.6	21.7-28.5
40,000-59,999	Handymax	190.6-192.4	31.2-32.3
60,000-99,999	Panamax	225.1-230.1	32.4-34.7
100,000+	Capesize	>230.2	>34.7

---

[12]The upper bound of this interval was increased from 34,999 as in Smith et al. (2014) to 39,999 to correspond with ClarksonsGroup (2015).

Part II

Analysis of Raw AIS Data

Chapter 3

S-AIS Data

In this chapter global Satellite AIS data, collected by AISSAT-1 and AISSAT-2, is presented. To demonstrate the size of this large amount of data, the amount of messages, and the distribution of these messages between different message types, are presented in Section 3.2. Section 3.3 presents the total number of distinct ships present in the S-AIS data, and the different ship classes these ships belong to is presented in Section 3.4. The geographical distribution of the S-AIS data can be found in Section 3.5. Lastly, in Section 3.6, a discussion is made on the possibility of erroneous data.

3.1 Introduction

The S-AIS data was supplied by the Norwegian Coastal Administration, and included global data from a four month period, from 1st May to 15th September, 2014. The data had been received from both AISSAT-1 and AISSAT-2. As a part of this thesis, the AIS data was decoded and indexed in a SQLite database. This was done using a Python program, which is included in Appendix C. The program also utilised a library for decoding AIS messages called AIS Parser SDK v1.10 (Lane, 2006). Decoding and indexing the AIS data could take several hours, because the high amount of raw data. The raw satellite AIS data amounted to just below 35 GB. This can arguably come under the definition Big Data[13]. Extra care was given to reduce the code complexity, to facilitate analyses on this large amount of data. These precautions would be unnecessary in cases

---

[13]There is no common definition on Big Data, however, Big Data can be defined as a combination of different data, in such an amount that it has to be computationally processed and where a domain knowledge should be utilized to provide useful information.

with a smaller dataset.

3.2 Distribution Over Message Types

The total number of messages was 197,538,966, with a distribution between dynamic- (message type 1, 2, 3 and 4) and static- (message type 5) messages as seen in Table 3.1. As expected there were a much higher number of dynamic messages, as these are sent with shorter reporting intervals.

Table 3.1: Distribution of S-AIS messages across the different message types.

Message Type	Number of messages
Dynamic	194,171,846
Static	3,367,120

3.3 Distinct Ships

The number of distinct ships in the S-AIS data can be found by counting the number of distinct MMSI numbers across all message types. Using this method a total of 85,108 distinct ships was found in the S-AIS data. Mantell et al. (2014) stated that the total world fleet consisted of 88,483 ships on 2nd May 2014. This means that approximately 95.05% of the total the world fleet can be accounted for with the S-AIS data. The last 4.95% of the world fleet may be ships that have been inactive in the data period. Another option is that these ships have exclusively operated in extremely high trafficked areas where they have not been detected by the satellites. Compared to the steady state number of ships reported in Mantell et al. (2014), the S-AIS data, which is from 1st May to 15th September, probably has some in and outflow of ships throughout the period due to new buildings and ship demolition.

The number of distinct MMSI numbers in each of the different message types can be seen in Table 3.2. The relatively high number of MMSI numbers in the static message type is favourable, as the messages contain destination, estimated time of arrival, ship types and key dimensions about the ship. There is more MMSI numbers in the dynamic messages, this is probably caused by the much higher volume of messages in this category.

Table 3.2: Distribution of distinct MMSI across the different message types in the S-AIS data.

Message Type	Number of distinct MMSI
Dynamic	81,690
Static	47,089
Union between static and dynamic	43,671

3.4 Ship Types in the S-AIS Data

Static messages contain information about the ship type, and this can be used to show which ship types that are present in the S-AIS data. As can be seen from Table 3.2 there are static data for 55.3% of the ships in the combined static and dynamic data. For the 47,089 ships in the static messages, 46,202 ships had valid information about their ship type, while the rest lacks this information. The distribution between AIS ship types for the remaining 46,202 ships can be seen in in Figure 3.1. Cargo ships (AIS ship type 70-79) constitute the majority, while tankers (AIS ship type 80-89) is the next largest ship group. The category "other ship types" comprises everything from tugs, military vessels, diving vessels, sailing-boats, pleasure crafts to semi-submersible rigs.
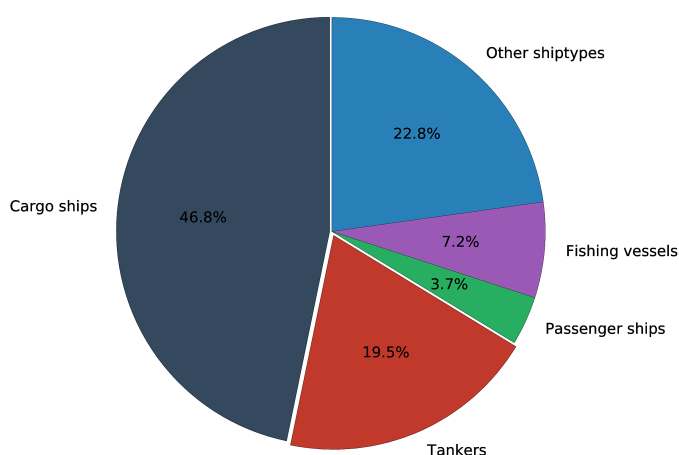
Figure 3.1: Distribution of ship types for 46,202 different ships.

The world cargo fleet, divided into vessel types, according to Mantell et al. (2014), as

seen in Table 2.5, is more refined than in Figure 3.1. If all oil tankers, LPG and LNG carriers from Table 2.5 is combined, these represent 16.8% of the total world fleet. In the S-AIS data, these ship types are labelled as tankers, and represent 19.5% of the total number of ships. According to Mantell et al. (2014), all cargo ships represent 31% of the world fleet. All cargo ships from the S-AIS data represent 46.8% of the total number of ships. Since the ship type only can be found in the static AIS messages, this distribution does not need to be representative for all ships in the S-AIS data. However, this indicates that tankers and cargo ships are slightly higher represented in the static S-AIS messages compared to Mantell et al. (2014). One reason for this can be that tankers and cargo ships often are ocean going ships, where the satellites reception is higher. Coastal vessels, such as passenger ferries, tugs etc. can be under-represented in the S-AIS data. Another explanation is that Mantell et al. (2014) includes ships that are not obliged to emit AIS data.

3.4.1   Cargo Ships

There was a total of 19,740 cargo ships with available length and breadth statistics.

As can be seen from Figure 3.2, the largest length category for cargo ships is between 180 and 210 m, with about 4,750 ships. Figure 3.3 shows that just shy of 7,000 ships had a breadth between 30 and 35 m, making this the largest category. These were most probably Panamax vessels.

The relation between the length and breadth of all cargo ships in the S-AIS data can be seen in the length and breadth scatter plot shown in Figure 3.4. This is particularly interesting, since it appears that most of these ships fall into a fairly distinct pattern. This could be used to validate the data, as a high deviation from a regression line could point to erroneous data. The high amount of ships with a breadth between 30 and 35 m is also visible in this scatter plot.
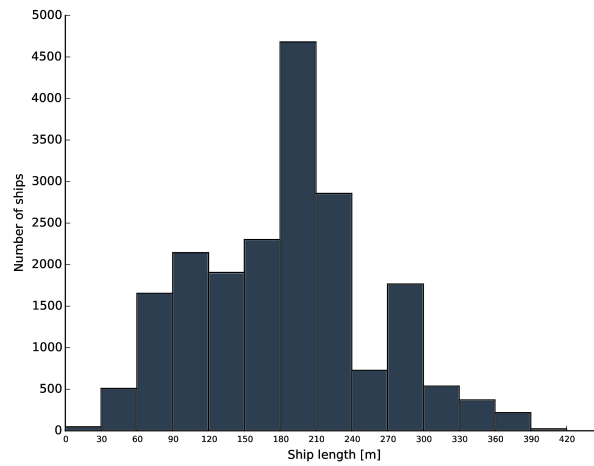
Figure 3.2: Distribution of length for all cargo ships.
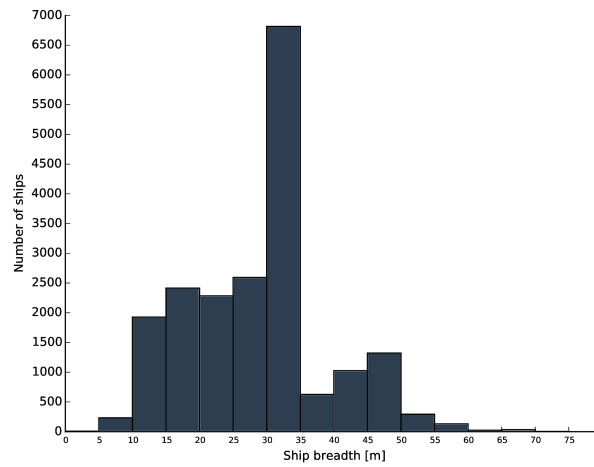
Figure 3.3: Distribution of breadth for all cargo ships.
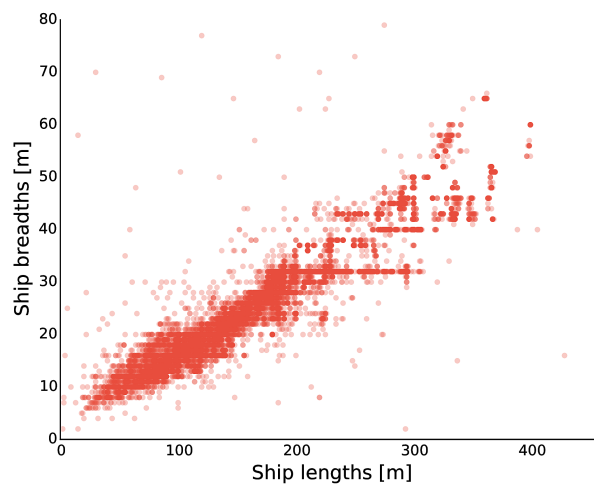
Figure 3.4: Scatter plot over length and breadth for all cargo ships.

29

3.4.2   Tankers

There was a total of 7849 tankers with available information on length and breadth. As can be seen from Figure 3.5, the most common length was between 180 and 210 m, with just under 1500 vessels. The length category between 90 and 120 m is also fairly common. As with the cargo vessels, the most common breadth for tankers was between 30 and 35 m, this can be seen in Figure 3.6.
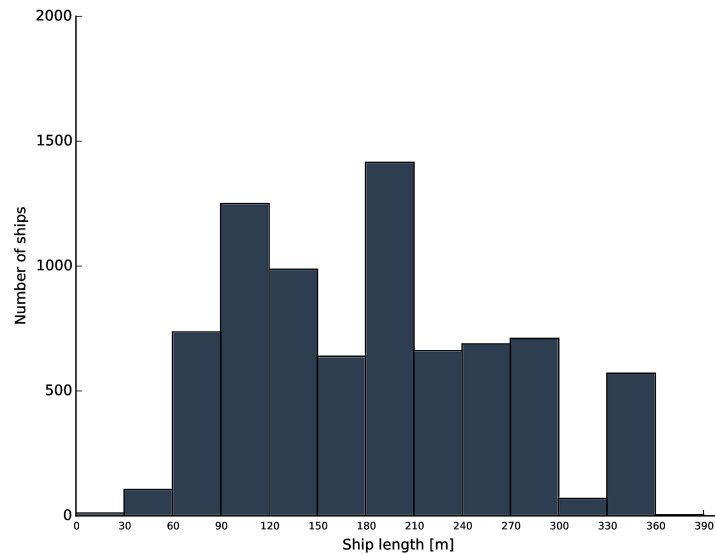
Figure 3.5: Distribution of length for all tankers.
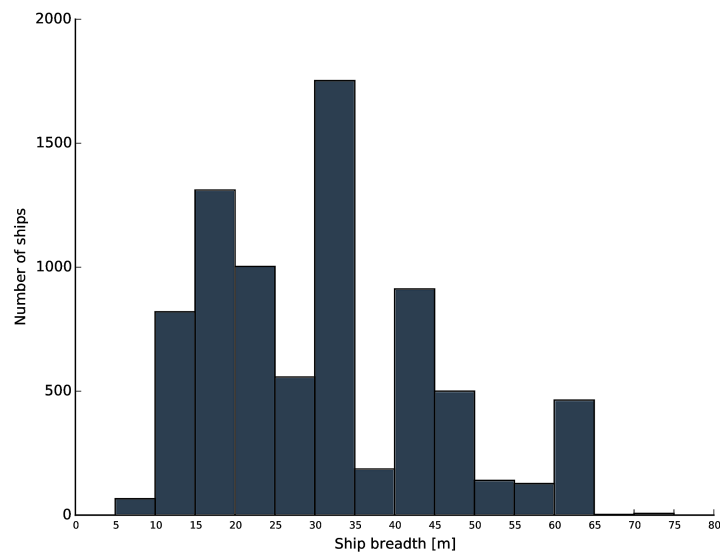
Figure 3.6: Distribution of breadth for all tankers.

The scatter plot over breadth and length in Figure 3.4 shows, as with the cargo ships, a distinct pattern. However, the tankers have a bit more uniform distribution, and a lot less outliers.
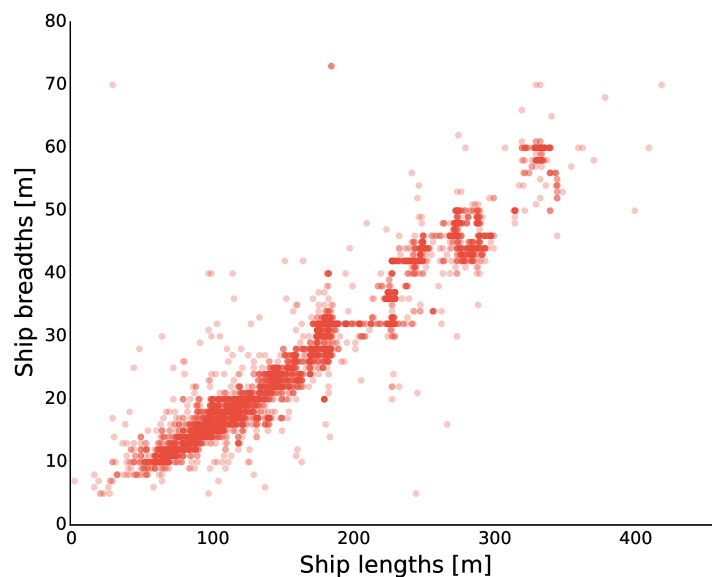
Figure 3.7: Scatter plot over length and breadth for all tankers.

3.5 Geographical Message Distribution

As described in Section 2.2, the two satellites collecting AIS data have a much higher passing frequency, and thus better coverage, in the far north and south. This is shown by plotting the average number of AIS messages received per ship, for each degree latitude and longitude. This is done by simply counting the number of messages, and the number of ships for each degree latitude, then plotting the number of total messages divided by the number of ships. As we can see in Figure 3.8a, there is an extremely high density at -64 degrees latitude. Upon investigation, there were only two vessels in this area, where one was the source of the bulk of the messages.

This was the research and survey vessel RV Laurence M.Gould. This is an icebreaker used for research in the Southern Ocean by the United States National Science Foundation. The vessel had a high activity inside an area with few other ships as well as a high satellite passing rate, which explains the extremely high density. In Figure 3.8b the messages from the research vessel is omitted. A pattern of increased density over the southern and northern most areas comes clear. This is as expected, as the passing rate of the satellites is higher over these areas. Another factor that can explain the higher density of messages is that the signal from AIS messages can be suppressed when there are a large number of vessels inside the same area, as explained in Section 2.2. To show this, the number of distinct ships across each degree latitude is plotted in Figure 3.9. As can be seen by

31
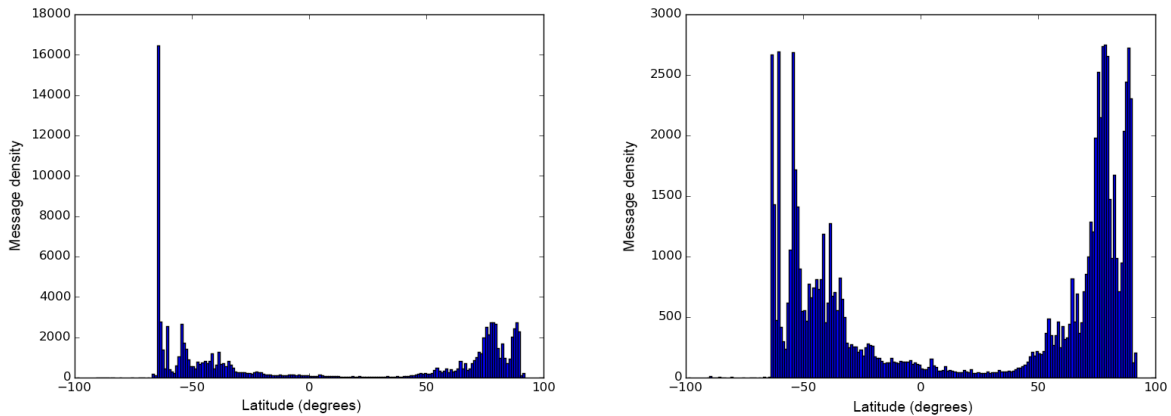
(a) Message density across latitude, with extreme point at -64 degrees.

(b) Message density across latitude, with extreme point at -64 degrees omitted.

Figure 3.8: Message density across latitude. Density is defined as the number of messages divided by the number of ships.

comparing Figure 3.9 and Figure 3.8b, there is a seemingly negative correlation between the number of ships present, and the message density. A co-factor is of course the higher passing rate of the satellites far north and south, which will increase the already good detection rate. The highest amount of distinct ships in the figure is at 40 degrees north. This corresponds with ship traffic in the Mediterranean and the East China Sea. The amount of ships is by far largest all along zero to 50 degrees latitude. The large amount of ships over 90 degree latitude is discussed in Section 3.6.
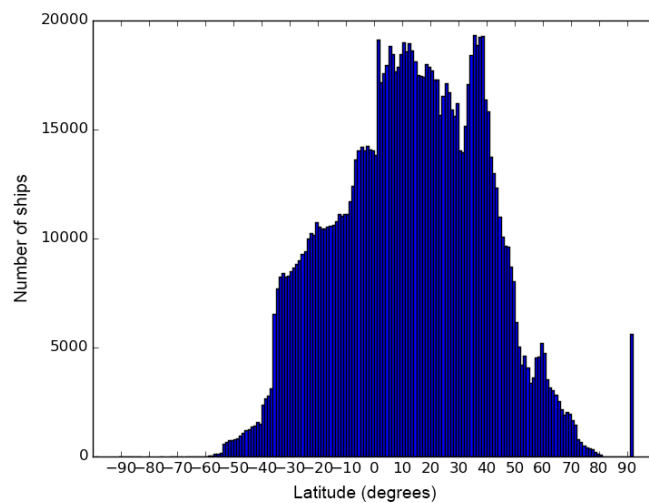
Figure 3.9: The total number of ships across each degree latitude.

3.6 Erroneous S-AIS Data

3.6.1 Erroneous IMO Numbers

To assert the validity of the IMO numbers in the S-AIS data the checksum verification described in Section 2.1.3 was performed on all IMO numbers. The code for this can be seen in Appendix F. It was assumed that all ships not required to have an IMO number will report a blank instead of an erroneous IMO number. Of the total 34,894 IMO numbers, 1,183 IMO numbers were erroneous.

Note that only static messages contain the IMO number, so the total number of distinct IMO numbers does not reflect the total number of vessels present in the S-AIS data.

3.6.2 Erroneous MMSI

Since the data period was from a single year, it was assumed that the number of ships that changed owners is negligible. This means that each ship should have only one MMSI number and correspondingly that each MMSI should belong to one, and only one ship. Verifying this can be done by matching the MMSI with the ships IMO number. Any MMSI with several different IMO numbers attached to it is not valid.

There was 385 different MMSI numbers which belonged to two or more distinct ships, signified by different IMO numbers. Since the IMO number only can be found in the static messages, this method only applies for the ships with static messages.

3.6.3 Erroneous Ship Dimensions

There was a total of 42 distinct ships that had a ship length above 460 m, which is more than the length of the longest ship ever built, Seawise Giant. This is not that probable, and the length could either be manually corrected, or the vessels could be excluded from the database. Since these 42 ships constituted such a small part of the total number of ships, the latter was chosen.

There was a total of 4,304 ships where the length or breadth was reported as 0 in one or more AIS messages. 1,062 of these ships had reported lengths and breadths longer than 0 in at least one AIS message. Only the static messages where the length or breadth was set to 0 was deleted. This means that the ships with other length and breadth data still

33

could be included, while the remaining 3,242 ships was excluded from further studies.

3.6.4 Erroneous Ship Positions

As can be seen from Figure 3.9 there is a high number of ships at over 90 degrees north, which must be erroneous messages since latitude only ranges from -90 to 90 degrees. Upon further inspection, 5,638 vessels had records of a position either over 90 degree latitude or under -90 degree latitude. 5,647 vessels had a position either over 180 degree longitude or under -180 degree longitude. 5,625 of these vessels had both erroneous longitudinal and lateral positions. This can be an error with either the AIS equipment or the decoding. These positional messages should be excluded if any global analysis is to be conducted. However these, was not deleted from the database, but rather excluded by constraints in the code, as the rest of the information in the AIS messages could be right.

3.6.5 Discussion

There were several thousand vessels with at least some erroneous data. However, this constitutes a fairly low proportion of our total number of ships, since there were 85,108 different ships in the S-AIS data. Also, in the cases with erroneous ship positions, only a share of the positions for each ship is in fact erroneous, a large part of the information is still valid. These data was removed from further analysis by enforcing maximum and minimum longitude and latitude.

Chapter 4

AIS Data from VTS Singapore

In this chapter the AIS data from Vessel Traffic Service (VTS) Singapore is introduced and analysed. The distribution of these data between the different message types can be seen in Section 4.2. Section 4.3 show the number of distinct ships in these data, and Section 4.4 show which AIS ship type these ships belong to.

4.1 Introduction

To compare S-AIS data with traditional AIS data, as well as to combine these two data sources, AIS from the AIS network in Singapore Straits was acquired. This was collected by land-based base stations and then processed by Vessel Traffic Service (VTS) in Singapore. The VTS-AIS data was acquired through the SESAME project. The data was processed by Ørnulf Jan Rødseth in the SESAME project - to remove local ship traffic from the dynamic messages. The data period was from 25. May to 25. November, 2014.

4.2 Distribution over Message Types

As with the S-AIS data, these VTS-AIS data was divided into two types of messages, static and dynamic. The distribution between these two message types can be seen in Table 4.1. The number of messages is a lot higher than the S-AIS data. This is

35

explained by the fact that the AIS network consist of a large number of base stations[14], which gives a very high reception rate of AIS messages, compared to the AIS satellites.

Table 4.1: Distribution of VTS-AIS messages across the different message types.

Message Type	Number of messages
Dynamic	526,203,165
Static	1,084,652

4.3 Distinct Ships

As with the S-AIS data, the number of distinct ships can be found by counting the number of distinct MMSI numbers. There was in total 17,026 distinct ships, signified by distinct MMSI, in the VTS-AIS data. The full distribution of distinct MMSI numbers across the different message types can be seen in Table 4.2.

Table 4.2: Distribution of distinct MMSI across the different message types in the VTS-AIS data.

Message type	Number of distinct MMSI
Dynamic	16,980
Static	15,821
Union between static and dynamic	15,775

4.4 Ship Types

To do a survey over the different ship types represented in the VTS-AIS data from the Singapore Straight, the static message type can be used, these contains information on ship- and cargo type.

---

[14]"Base stations are designed for use by Competent authorities to manage the VDL and enable effective ship to shore / shore to ship transmission of information. They are the core of any AIS Service and can be networked to provide broad VTS or Coastal Surveillance coverage and overall maritime domain awareness." (IALA, 2011)

Table 4.3: Distribution of MMSI on ship type, from VTS-AIS data

Ship type	Number of ships	Percentage of total
Cargo ships	8,665	52.5%
Tankers	4,158	25.2%
Others	3,648	22.1%
Total	16,487	100%

The total number of ships was 16,487, which is higher than the total number of ships in the static data from Table 4.2. This is probably due to some ships being registered with several ship types. This can be due to a ship being registered in several main categories (tanker, cargo ship and others). The bulk of the traffic comes from cargo ships and tankers, while categories such as tugs, passenger traffic and private vessels, which constitute the other category, comprise a fifth of the ships. There was a slightly higher percentage of cargo ships and tankers in the VTS-AIS data from Singapore Strait compared to the global numbers from the S-AIS data. This means that the Singapore Strait is an area with a higher share of cargo ships and tankers compared to the global S-AIS data. Compared to the world fleet from Mantell et al. (2014), there is a higher share of cargo ships.

Further analysis showed that 51.5% of the static messages from the VTS-AIS data come from either cargo ships or tankers. 51.7% of the messages in the dynamic data come from either cargo ships or tankers. This is found by selecting all the MMSIs for the tankers and cargo ships in the static messages, and counting the number of messages with these MMSIs in the dynamic messages. As international traffic is of interest, the VTS-AIS data was narrowed down to just tankers and cargo ships, meaning that the number of static and dynamic messages will be cut in half. This decrease of data volume helped to speed up the following analyses.

Part III

Examples of Use of AIS Data

Chapter 5

Heuristics for Determining Specific Ship Type from S-AIS Data

5.1 Introduction

Earlier in this thesis, the width of information that can be derived from the S-AIS data, as well as the amount of data that is found in the S-AIS data is presented. In this chapter an example on how these data can readily be used to give more knowledge about global or local ship traffic is given. Heuristics, which in this context will be a blueprint of ship characteristics for specific ship types, is presented. In Sections 5.2 to 5.5 heuristics for Gas Carriers, Container Vessels, Bulk Carriers and Oil Tankers can be found. The results, and the accuracy of the heuristic is summarised and discussed in Section 5.6.

5.1.1 Application of Heuristics

The static messages contain information on the main dimensions of the ship: length, breadth, as well as the current draught of the ship. The static messages also contain crude information on the ship type, where it groups the different vessel types into tankers, cargo ships, fishing vessels etc. However, there is no information that directly tells us whether a tanker is an oil tanker, a chemical tanker or an Liquefied Natural Gas (LNG) carrier, or whether a cargo ship is a ro-ro (roll on, roll of) vessel, a container vessel or a bulk carrier. Classifying ships by its ship and cargo type is of interest, as these different vessels are constructed with different optimal speeds. Knowing the optimal speed is of relevance for

analyses on emissions from ships, where the operating speed is compared to the optimal speed. The optimal speed is amongst others a factor of the block coefficient[16] of the ship. This information is of essential knowledge if any analyses on the fuel consumption, the CO2- and CO2 equivalents emissions or the use of slow steaming for a large group of vessels are to be made. Previous studies, such as Smith et al. (2014), have used commercial vessel databases to retrieve this ship specific information. The MMSI, IMO number or the vessel name can be used to retrieve additional ship information. However, it may be extremely time consuming to manually look up every ship and these databases can be an expensive and unnecessary middle man if these heuristics can provide the same information. An application of these heuristics is given in Chapter 6.

5.1.2 Approach

To develop these heuristics, a template group of ships for each ship type was needed to derive common characteristics for each ship type. Ship databases such as the ClarksonsGroup (2015), provides data-sheets where a large number of ships of each ship type is listed by its ship name. These data sheets simply provide the registered names of ships in each ship type, but are not exhaustive. The ship names were used to retrieve the ship dimensions as well as draught and ship speed from the S-AIS data. This was used to form a template group, which was inspected for any correlations between the size, speed and draught parameters. In addition to the largest draught, the difference between maximum and minimum draught can be a good metric to show what cargo a ship holds. Ships that contain a high mass of either densely packed cargo, or cargo with a high mass, such as oil tankers and bulk ships, can be expected to have a large difference in draught between their fully loaded and unloaded state. Meanwhile, ships with a low-density cargo, like LNG vessels, can be expected to have a smaller difference in draught. Information like this was ultimately used to make a heuristics to identify a ships specific ship type from S-AIS data alone. Experiences from the development of the first heuristic, on LNG vessels, was subsequently used to simplify the process of making new heuristics. This process is

---

[16]Coefficients is used as simple metrics to describe a ship hulls shape. Amongst these coefficients is the block coefficient. This describes the ships fullness, compared to a cube, beneath the waterline. A high block coefficient represents a typical U shaped ship, designed for lower speeds. A low block coefficient represents a typical V shaped ship, designed for higher speed. The formula for the block coefficient is $\frac{\nabla}{L*B*T}$ where L is the length, B is the breadth and T is the draught of the ship. ∇ is the volume of the ship (Amdahl et al., 2011).

outlined in the following sections.

Maximum speed constraint

Early testing of the first heuristic, the LNG heuristic, showed that even though it generally performed well, it misidentified a lot of Oil Tankers as LNG vessels in certain breadth categories. As an example, the 274 m long and 48 m broad Oil Tanker Sestrea had a maximum recorded speed of 20 knots. Because of the relatively high speed, this vessel was identified as an LNG Carrier. However, how often the different speeds are reported for the specific vessel gives a picture on how representative this maximum recorded speed was. To do this, all reported speeds were bucketed in one knot intervals. Out of 165 speed reports for this vessel, only one was for the maximum recorded speed of 20 knots. The maximum speed among the speeds with most occurrences was 14 knots with 10 recordings. This may be more representative as a top speed for this vessel. To avoid these rare occurrences of high speed, a new constraint was put on the LNG heuristic; for a maximum speed to be valid, it should have 10 or more recordings. This constraint was subsequently implemented in all of the heuristics. The code to find the maximum speed under this constraint is included in Appendix E. It is hard to point to one specific reason why some ships can get a few recordings of extremely high speeds. This can be due to erroneous data, or a combination of abnormally high speed and favourable current and wind conditions. Speed recordings from AIS data is, as previously mentioned, speed over ground, and not the speed relative to the water, see Table B.1 in the Appendix.

5.1.3 Ship types

The following heuristics will later in this thesis be used to identify the different ship types frequenting Singapore Strait, so the aim is to develop heuristics primarily for the most common ship types in the Singapore Strait. Bateman et al. (2007) presents the number of vessels registered, as well as average dead weight, of each ship type that used the Malacca Strait in 2004. The Malacca strait is positioned as a north-western inlet to Singapore Strait. The four most common categories from this study, in terms of dead weight, are presented in Table 5.1. Note that these numbers do not show the number of unique vessels.

As the approach is similar across the different ship types, these heuristics will serve as a proof of concept for developing heuristics for all other ship types.

Table 5.1: Number of registered vessels and total tonnage of vessels using the Malacca Strait by Ship type in 2004 (Bateman et al., 2007)

Type	Number of ships	Total DWT
Oil Tanker	22,995	1,857,067
Container Vessel	29,672	1,013,1552
Bulk Carrier	13,599	772,555
Gas Carrier	3,933	138,560

5.2 Gas Carriers

Liquefied Natural Gas (LNG) Carriers are tank ships that are designed to carry liquefied natural gas. Along with Liquefied Petroleum Gas (LPG) carriers, they form the general group gas carriers. These vessels are designed to both hold liquefied gases, and to retain them as liquefied. Construction of both these vessel types has to follow the International Code for the Construction and Equipment of Ships Carrying Liquefied Gases in Bulk, IMO (1993), so they share the most common characteristics. Because of this, the term LNG Carriers will from this point encompass both LNG and LPG Tankers. Using a dataset accessed from Clarksons ship database, containing a total of 403 LNG carriers, we could obtain dimensions for a total of 322 LNG carriers from our S-AIS data.

Length and breadth

As can be seen from the scatter plot in Figure 5.1, there is no clear linear correlation between length and breadth. From the figure it is obvious that the vessels are divided into different breadth intervals. This can also be seen by the ship breadth histogram in Figure 5.2. The vessels can be generalised into three clusters based on length. The main cluster between 273 m and 300 m, one cluster between 310 m and 320 m and a final cluster around 345 m. These two last clusters correspond to the Q-Flex and Q-Max classes. The latter is a ship class signifying vessels that have the maximum allowed dimensions to be able to dock at the LNG terminals in Qatar. These are the largest LNG vessels built to day. The Q-Flex category had the largest vessels until the construction of the Q-Max

vessels. The economy of scale and the increasing demand for LNG indicates that the number of ships in these classes will grow.
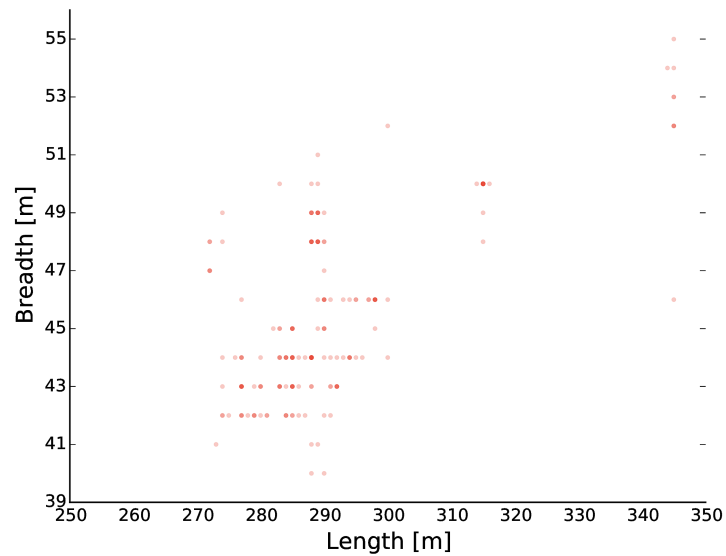
Figure 5.1: Length and breadth of LNG Carriers.
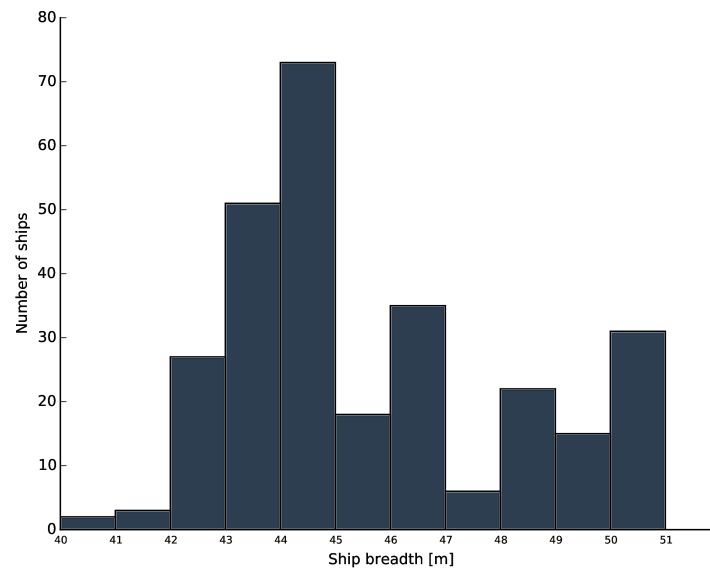
Figure 5.2: Breadth of LNG Carriers.

There is a significant division on breadth inside the general clustering. The Q-Max and Q-Flex will be put in their own groups. This generalisation can be seen in Table 5.2. Note that 23 vessels was omitted as these either had breadth below 35 m, or a length below 200 m, and thus were clear outliers from the general group. This indicates that they can have been misidentified as LNG carriers.

45

Table 5.2: Generalisation of LNG Carriers based on breadth and length

Breadth (m)	Length (m)	Number of ships in template group
40-42	273-294	32
43	274-292	52
44	274-300	75
45	282-298	16
46-47	272-300	40
48	272-290	19
49	274-290	13
50-52	283-300	9
Q-Flex		
48-50	314-316	26
Q-Max		
46-54	344-345	10

Draught

To extend the ability to identify LNG Carriers, a scatter plot over the breadth and max draught was made. Max draught is the maximum reported draught in the data period. This can be seen in Figure 5.3. There is considerable variance in maximum draught inside each breadth category. Most of the draughts are between 8.5m and 13m.

The differences in max draught can be due to a number of reasons:

- The ships inside each breadth class have different lengths, which can mean a larger draught.

- Not every vessel had necessarily been at full capacity in the data period.

- The practices in reporting draught can be different, as this is entered manually.

- Different containment systems determine how large amount of LNG a ship can hold.

For example, Spherical Moss Tanks is not as volume effective as other solutions[17]. Vessels with cubical shaped membrane tanks will typically load 8% more cargo than spherical ones (Moon et al., 2005). The Q-Max and Q-Flex vessel categories have tanks of the latter type. As an example, Høegh owned Arctic Princess, a LNG vessel with spherical tanks is 288 m long and has a breadth of 49 m. This vessel has a cargo capacity of 147,980m^3. Independence, which also is a Høegh owned LNG Carrier, is 294 m long and 46 m broad. It is a vessel with Membrane containment system, and has a cargo capacity of 170,132m^3. Both vessels have the same designed max draught (summer draught).

LNG Carriers are expected to have a small difference between maximum and minimum draught compared to oil tankers as the LNG fluid has relative low density[18]. The difference in draught for the general group is plotted in Figure 5.4. Most vessels have a change in draught between 0 and 3.5 m. There are three outliers which can be attributed to erroneous data, and possibly due to manual plotting. Draught is set manually by the ships operators.

The change of draught in the Q-Flex group was ranging between 0 and 3 m, with one outlier at 9.29 m. This was the vessel Al Oraiq, which had five reported draughts at 9.4 m and one at 18.7 m. The vessels in the largest group, Q-Max, had a change of draught between 0 and 3.3 m. Since most of the changes

---

[17]The volume of a sphere is $\frac{4}{3}\pi r^3$, but a cube with the same length, breadth and height as the diameter of the sphere will have a volume of $8r^3$, and will thus be more space-effective in a ship.

[18]Crude oil typically has a density between 816 $\frac{kg}{m^3}$ to over 1000 $\frac{kg}{m^3}$, while LNG typically has a density between 410 $\frac{kg}{m^3}$ and 500 $\frac{kg}{m^3}$.
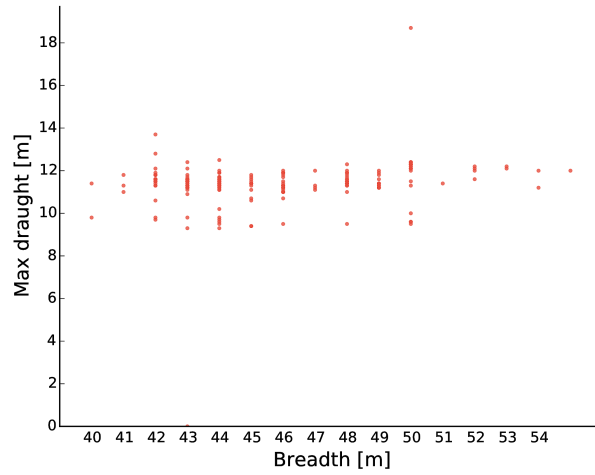
Figure 5.3: Max draught of LNG Carriers inside different breadth categories.
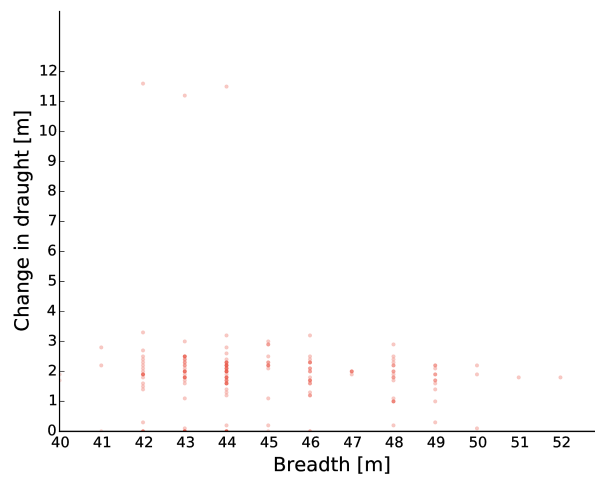
Figure 5.4: Max change in draught for LNG Carriers excluding Q-flex and Q-max vessels.

Speed

The last metric that can be applied to determine ship type is maximum speed. Note that due to erroneous data, some vessels had a maximum speed reported to be 102.3 knots which obviously is wrong. Prior to using the maximum speed constraint found in Appendix E, all speed reports for each vessel that was over 50 knots, which is well over the expected maximum speed for any large vessels, were removed. This was done to lower the running time of the code. Figures 5.5 to 5.7 shows scatter plots of maximum speed and breadth for the selection of vessels. It is clearly no linear relation between breadth of the vessel and the maximum vessel speed. However, the maximum speeds for the different vessel- and breadth categories can be categorised, see Table 5.3, where the lowest, the highest and the mean recorded maximum speed is given.

Table 5.3: Lowest-, mean- and largest maximum recorded speed for LNG Carriers in different breadth categories.

Breadth	Min maximum speed	Mean maximum speed	Highest maximum speed	Standard deviation
(m)	(kn)	(kn)	(kn)	(kn)
40-42	14.1	19.43	23.5	2.09
43	0.2	20.4	32.4	4.34
44	16.4	20.68	23.4	1.35
45	10.9	20.43	23.2	2.89
46-47	16.1	20.9	23.4	1.40
48	17.9	19.93	22.3	1.35
49	18.8	22.66	35.2	4.21
50-52	18.9	21.42	22.2	1.27
Q-Flex				
48-50	19.3	21.12	23.1	0.96
Q-Max				
46-54	20.1	21.25	22.1	0.63
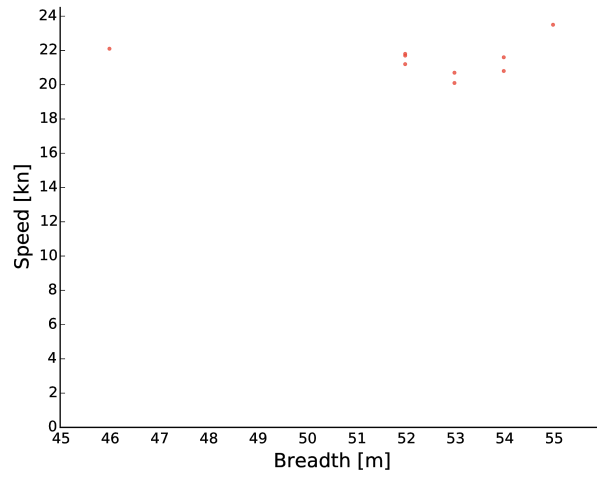
Figure 5.5: Max speed and breadth plot, Q-Max vessels.
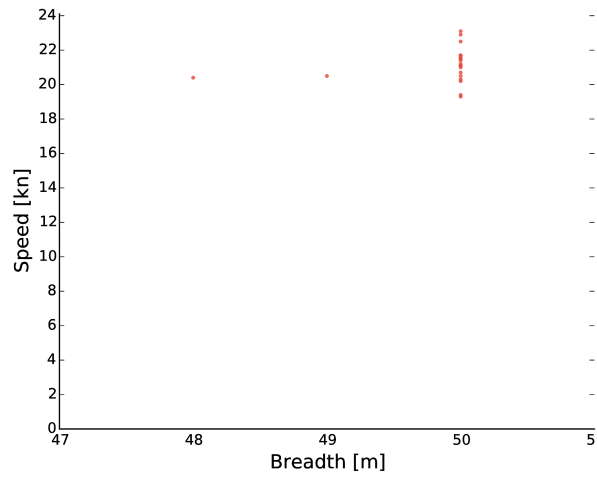
Figure 5.6: Max speed and breadth plot, Q-Flex vessels.
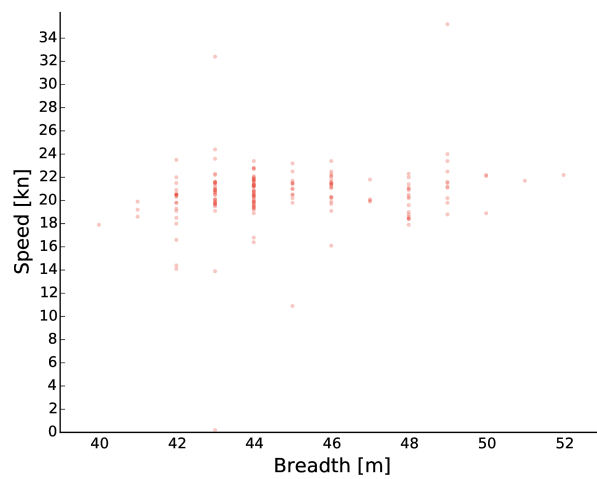
Figure 5.7: Max speed and breadth plot, LNG vessels in General Group.

5.2.1 Heuristic

Using the ideal group of confirmed LNG vessels, a number of metrics that can be used to identify a vessel as a LNG Carriers from its AIS data is identified. Analysing key dimensions, such as length and breadth, have shown that these LNG vessels can primarily be divided into three size categories: general, Q-Flex and Q-Max. The general category could be divided into different breadth categories. Each breadth category has a distinct maximum and minimum length, as well as a minimum-, mean- and maximum speed. A heuristic for identifying LNG vessels was made using these metrics, along with the maximum draught and the difference between minimum and maximum draught. The complete heuristic is presented in Table 5.4.

Table 5.4: Key dimensions, change in draughts as well as min, mean and max speed for the groups of LNG Carriers.

Maximum draught				13 m		
Maximum change in draught				3.5 m		
AIS ship type				80-89 (tanker)		
Breadth	**Length**	**Number of ships**	**Min top speed**	**Mean top speed**	**Max top speed**	**Standard deviation**
(m)	(m)		(kn)	(kn)	(kn)	(kn)
40-42	273-294	32	14.1	19.43	23.5	2.09
43	274-292	52	0.2	20.4	32.4	4.34
44	274-300	74	16.4	20.675	23.4	1.35
45	282-298	16	10.9	20.43	23.2	2.89
46-47	272-300	40	16.1	20.9	23.4	1.40
48	272-290	19	17.9	19.93	22.3	1.35
49	274-290	13	18.8	22.66	35.2	4.21
50-52	283-300	9	18.9	21.42	22.2	1.27
Q-Flex						
48-50	314-316	26	19.3	21.12	23.1	0.96
Q-Max						
46-54	344-345	10	20.1	21.25	22.1	0.63

Heuristic testing

The heuristic in Table 5.4 as well as the maximum speed criteria for the different vessel classes was applied on the Satellite AIS data. It was required that the AIS ship type was a tanker, that the draught was within 14 and 7 m, that the maximum change in draught was 3.5 m and that the maximum speed either over the minimum recorded maximum speed minus two standard deviations, or over 16 knots, which ever was the highest. Maximum speed was defined as in the maximum speed constraint. Maximum and minimum length and breadth was set according to the heuristic.

Breadth 40-42m

This group had a relatively low minimum top speed at 14.1 knots. This means that the lowest allowed maximum speed is set to 16 knots. Using the heuristic and the maximum speed criteria 27 vessels from the template group of 32 vessels was rediscovered. One vessel outside the template group was identified as a LNG Carrier, this was the LNG Carrier Milaha Ras Laffan.

Breadth 43m

This group had one ship with a maximum recorded speed just over zero. This ship was omitted as the minimum allowed maximum speed was set to 16 knots. This resulted in that 45 of 52 ships in the template group was rediscovered, while finding four new LNG Carriers: K.Jasmine, K.Acacia, K.Mugungwha and Tangguh Palung.

Breadth 44m

The ships with a breadth of 44 m were the largest template group with 74 vessels. 74 different vessels from the S-AIS data matched the heuristic. Seven of these vessels were not present in the template group. A manual check of the ships not present in the template group yielded the following results, presented in Table 5.5.

Table 5.5: Ships with 44 m breadth outside the template group, identified by the heuristic.

MMSI	Name	Vessel Type
205506000	S/S EXPLORER	LPG Carrier
310510000	MET. JANE ELIZABETH	LPG Carrier
310535000	M.SHIRLEY ELISABETH	LPG Carrier
310540000	METH ALISON VICTORIA	LPG Carrier
352645000	K.FREESIA	LNG Carrier
533035000	LNGC PUTERI NILAM	LPG Carrier
538004980	LNG/C GOLAR SEAL	LNG Carrier

A lot of LPG tankers is identified as LNG Carriers. However as previously stated, construction of both these vessel types has to follow the International Code for the Construction and Equipment of Ships Carrying Liquefied Gases in Bulk (IMO, 1993), so they can be expected to be in the same group. Seven vessels from the ideal group is not identified, due to low speeds. This could be explained by a lack of activity inside this time period, and it can be expected to find a larger amount of ships with data from a longer time period.

Breadth 45m

Using the heuristic for vessels with a breadth of 45 m, a total of 17 vessels was identified. In addition to the 15 of the vessels in the template group, two new LNG Carriers, Meth Patricia Camila and M.Mickie Harper were found.

Breadth 46-47m

This was the third largest template group, with 40 ships. Using the standard heuristic, 36 vessels from the heuristic were rediscovered, and seven other vessels were identified as LNG Carriers. Five of these vessels were LNG Carriers, while two were oil tankers.

Table 5.6: Ships with breadth between 46-47 m outside the template group, identified by the heuristic.

MMSI	Name	Vessel Type
215739000	CASTILLO-SANTISTEBAN	LNG Carrier
259082000	RIBERA DUERO KNUTSEN	LNG Carrier
310171000	NW SHEARWATER	LNG Carrier
367067110	POLAR ENTERPRISE	Oil Tanker
369272000	POLAR DISCOVERY	Oil Tanker
431015000	NORTHWEST SWALLOW	LNG Carrier
503009000	NW STORMPETREL	LNG Carrier

Breadth 48m

This was the least accurate class by far, with 169 ships being identified as LNG Carriers. As the template group only comprised of 19 vessels, this indicated that a lot of ships was misidentified as LNG Carriers. Investigation of the reported speeds for the vessels outside the template group made it apparent that a lot of oil tankers are identified as LNG Carriers, due to short bursts of high speeds. This was the reason for the development of the maximum speed constraint. With the maximum speed constraint, only three ships outside the template group was identified as LNG Carriers. These ships was Taitar No.4, Taitar No.1 and Golar Viking which are all LNG Carriers. The maximum speed constraint was used on all following breadth groups and ship types.

Breadth 49m

In this breadth group, there were 13 different vessels in the template group. Ten of these were rediscovered, while two additional LNG vessels were identified. These were Taitar No.2 and Taitar No.3.

Breadth 50-52m

In this breadth category, eight of the nine ships in the template group were identified.

Q-Max and Q-Flex

Using the heuristic for the Q-Flex class we rediscovered all ships in the template group while additional two LNG Carriers, Al Shamal and AL Sahla, were found. In the Q-Max class all the vessels in the template group were rediscovered with no additional vessels being identified.

Discussion

The performance of the heuristic was over all quite good, especially after the maximum speed constraint was implemented. However, this heuristic is pretty detailed, and will be impractical to put to use in practice. A simplified heuristic is presented and tested in the following sections.

5.2.2   Simplified Heuristic

In the same way the detailed heuristic was tested, a simplified heuristic with fewer breadth classes can be tested. A simplified heuristic will be faster, and easier to use in practice. This simplified heuristic had only three ship groups based on their dimensions. These three groups were the general group, Q-Flex and the Q-Max group. The general group had a breadth between 40 and 52 m, and a length between 270 and 300 m. The Q-Flex group had a breadth between 48 and 50 m, and a length between 314 and 316 m. Lastly, the Q-Max group had a breadth between 46 and 54 m and a length between 344 and 345 m.

Table 5.7: Simplified heuristic, including max draught, max change in draught, maximum speed, AIS ship type and key dimensions for three groups of LNG Carriers.

	Maximum draught:		13 m
	Maximum change in draught:		3.5 m
	Maximum speed:		>= 16 kn
	AIS ship type:		80-89 (tanker)

	Breadth (m)	**Length** (m)	**Number** **of ships in template group**
General group	40-52	270-300	255
Q-Flex	48-50	314-316	26
Q-Max	46-54	344-345	10

5.2.3   Testing of Simplified Heuristic

In the general group a total of 255 ships was present in the template group. Out of these, 223 were identified with the heuristic. 26 ships outside the template group were also rightfully identified as LNG Carriers, an increase from 22 in the detailed heuristic. This increase can be due to ships that fell outside the length interval that had been defined for each breadth category. Only two ships were misidentified as LNG Carriers, this was Polar Enterprise and Polar Discovery. In the Q-Flex and Q-Max groups no changes were made, and the results naturally stayed the same. This means that the performance of the simplified heuristic was good compared to the detailed heuristic. The experiences from the LNG-heuristics, including the maximum speed constraint and the performance of the simplified heuristic, led us to develop a simplified form of all of the following heuristics.

5.3   Container Vessels

Container Vessels are cargo vessels that carry their cargo in containers loaded onto the main deck. These ships are usually measured in Twenty-foot equivalent units (TEU) where twenty foot is the standard size of a container. The container can hold anything from cars, to clothes and food.

5.3.1   Roll On-Roll Off and Sub Panamax Container Vessels

This section show an example of how two ship types was indistinguishable from each other. The cargo of both these ship types will typically have a low density compared to oil, ore or similar. Because of this low density, the difference in draught between a fully loaded and an unloaded state will be relatively small. When these parameters are similar between the ship types, the main dimensions as well as the speed profile are also expected to be similar.

The Sub Panamax ships have dimensions below Panamax, which is the ship class that has the maximum size to be travelling through the Panama canal. Sub Panamax vessels typically have a capacity of 1,000 to 2,999 TEU. From the S-AIS data it was possible to retrieve dimensions for 1163 of the 1868 Sub Panamax ships from the Clarksons ship database (ClarksonsGroup, 2015). Roll on/roll off (RORO) ships are ships with cargo that is rolled on and rolled of the ship. These cargo types can be heavy machinery, automobiles, railroad cars etc. Some of these vessels can be of type RO-PAX which is car ferries that does a combination of passenger and vehicle transportation. We will focus on ships that strictly transport cargo, and no passengers. Using a dataset retrieved from Clarksons it was possible to find dimensions for 361 of the in total 828 vessels in the vessel sheets, from the S-AIS data.

Length and Breadth

The minimum length was set to be 150 m, as it is expected that vessel types that has a length below 150 m will be very hard to distinguish from each other. This reduced the number of Sub Panamax vessels to 976, and the number of RORO vessels to 150. From the plot in Figure 5.8 it was evident that these two ship types have near identical length and breadth relation as a population, and that it would be very hard to distinguish them from each other based on these two parameters. The RORO vessels are plotted in a blue colour, while the Sub Panamax vessels are plotted in a red colour. The linear regression line of breadth and length for each ship type is also plotted in the figure, in the same colour as the respective ship types. These regression lines was overlapping.
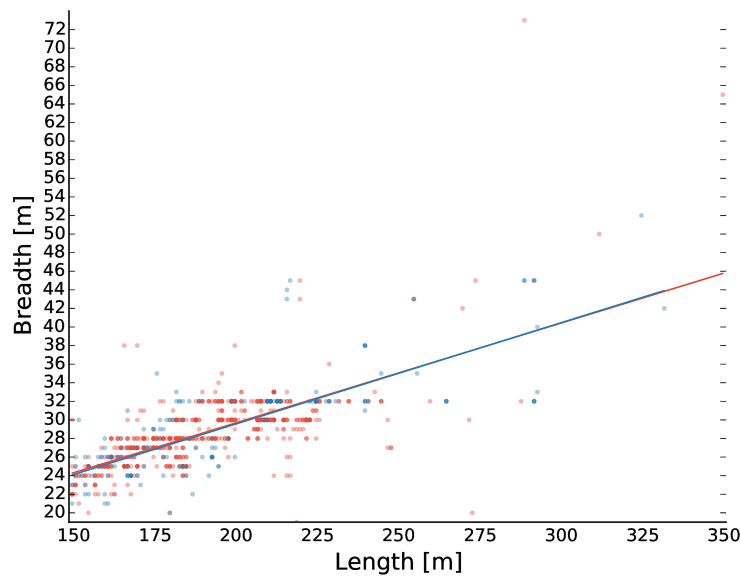
Figure 5.8: Length and breadth for Ro Ro vessels (blue) and Sub Panamax ships (red).

Draught

The breadth and maximum draught for these two ship types was plotted to investigate if these parameters could be applied to distinguish them from each other. As can be seen in Figure 5.9 there were some differences, as the RORO vessels had a lower draught up to about 44 m breadth, according to the linear regression line. Note that the Pearson Correlation Coefficient for the Sub Panamax Container Vessels was 0.61 while it was 0.78 for the RORO vessels, indicating a stronger correlation for the latter. The maximum draught for these two vessel types were almost identical, and there was no clear way to distinguish them from each other.

When investigating the relation between the maximum change in draught and the maximum breatdh for these two ship types in Figure 5.10there still was no clear difference. After plotting maximum draught and minimum draught in the same scatter plot in Figure 5.11, there was still no visible difference between these two groups.
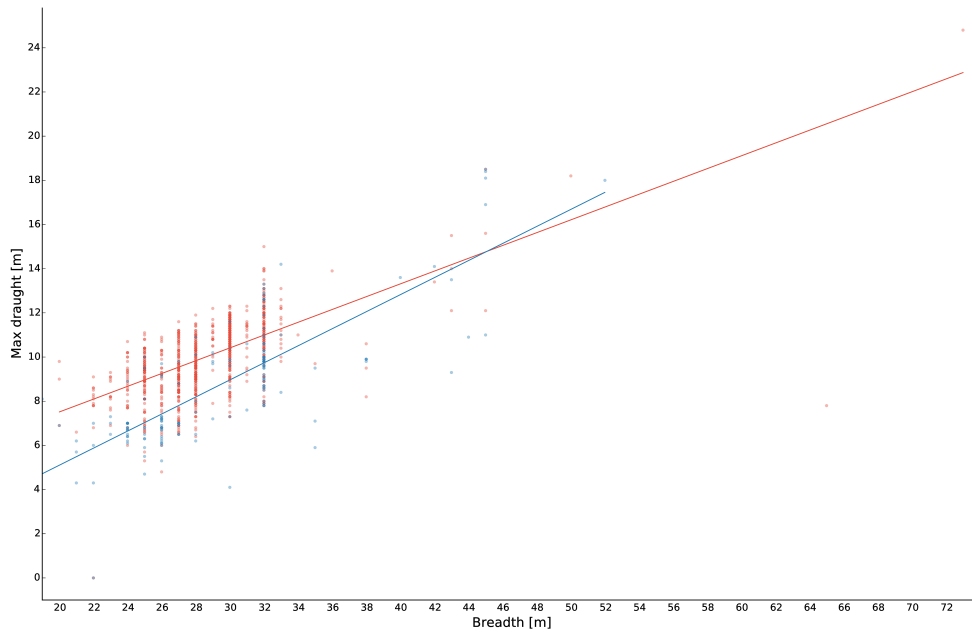
Figure 5.9: Maximum depth and breadth for RoRo vessels (blue) and Sub Panamax ships (red).
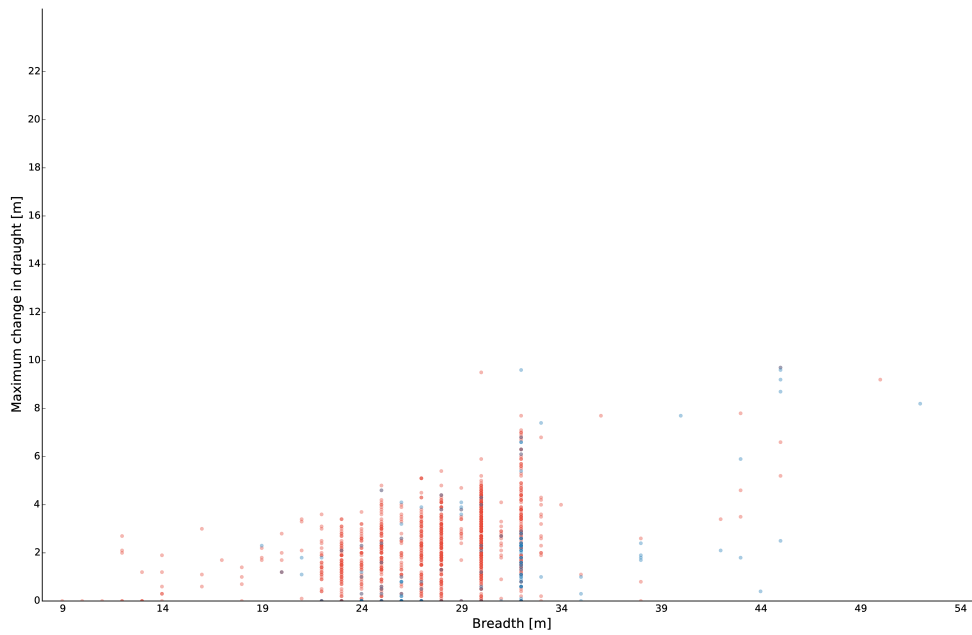
Figure 5.10: Change in draught and breadth. Roro vessels in blue, Sub Panamax in red
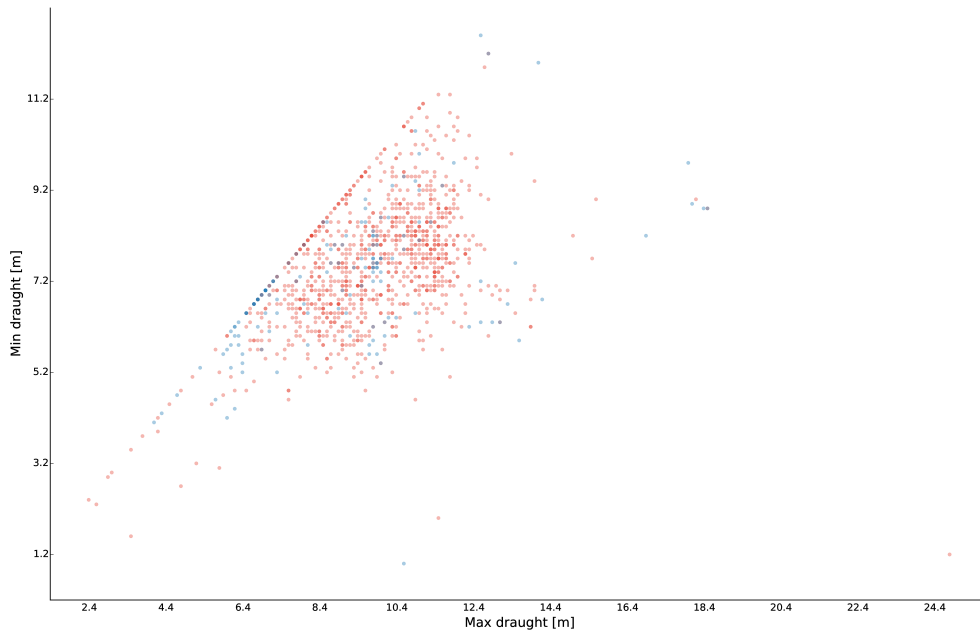
Figure 5.11: Maximum and minimum draught. RoRo vessels in blue, Sub Panamax in red.

Speed

The last metric that could be applied to differentiate between these two vessel types was speed. As seen in Table 5.8, the lowest, median and highest maximum speed were virtually identical for the two groups.

Table 5.8: Maximum speed characteristics for RORO and Sub Panamax ships

Vessel category	Lowest maximum speed (kn)	Median maximum speed (kn)	Highest maximum speed (kn)
Sub Panamax	5.5	16.9	24.9
RORO	5.5	17.9	24.9

Conclusion

As both the dimensions as well as the operating pattern (maximum and minimum speed) were nearly identical between these two ship classes, it was impossible to differentiate between them using this information. However, one could argue that since these ship types are so similar, they should maybe not be differentiated when doing emission studies anyway. Because of the heuristic's lack of ability to differentiate them from each other, the heuristic will not be used in further analyses.

5.3.2 Panamax Container Vessels

Length and Breadth

Out of the 861 Panamax Container Vessels in the vessel sheet from Clarksons, dimensions for 695 ships were found in the S-AIS data. The selection was then limited to only the most common length and breadth category, as shown in the left part of Figure 5.12. The length limitation can be seen as the vertical line at 210 m, and the breadth limitation can be seen as the horizontal line at 34 m. The final selection, which is the the lower right quadrant, comprised of 677 vessels, and is shown in the right part of Figure 5.12.

As can be seen from the figure, most of these vessels had a breadth of 32 m. Out of the 677 vessels, only 57 had a breadth deviating from 32 m. If one meter deviation in both directions from 32 m was allowed, there were only 12 vessels not fitting the criteria. This simplified the heuristic greatly, and reduced the group further to 665 vessels. The
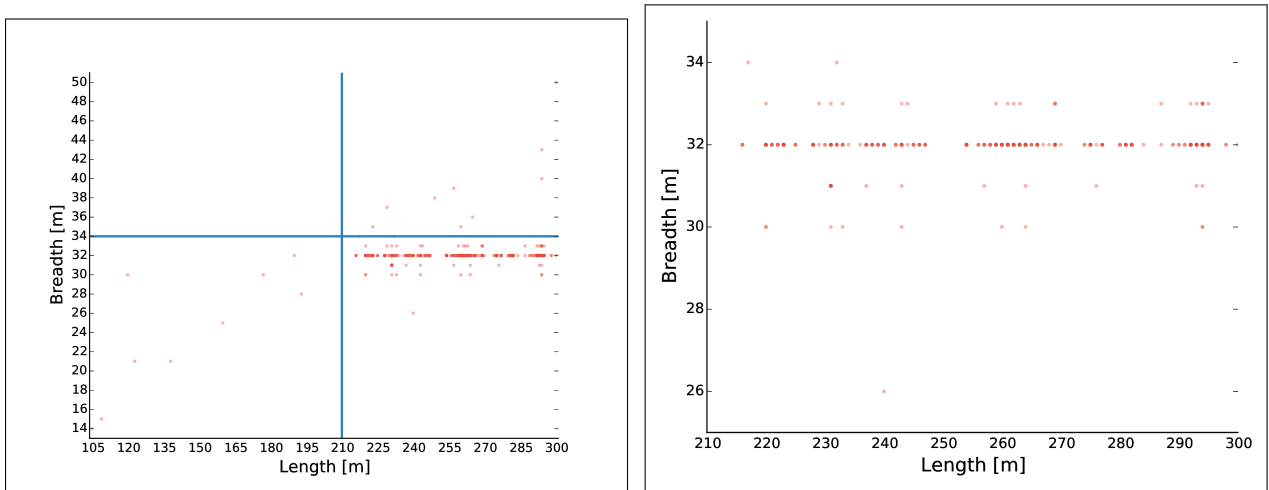
Figure 5.12: **Left:** Breadth and length for container vessels in the Panamax category. **Right**: Close up of lower right quadrant of left figure.

minimum length in this group was 217 m, while the maximum was 300 m.

Draught

Since the vessels seemed to group into length categories rather than breadth categories, it was inspected whether there was a relation between draught and length. In the left plot in Figure 5.13 it can bee seen that the ships form two major clusters, based on the length of the vessels. One group was between 290 and 300 m, and one between 250 and 270 m. However, if clustered by draught, roughly all vessels from 270 to 300 m seem to have similarities when it comes to draught, and all vessels below 270 m also seem to have somewhat similar draught profiles. The draught profiles for these two groups can be found in Table 5.9.

Table 5.9: Maximum draught for two groups of Panamax Container Vessels.

Length (m)	Number of vessels	Median maximum draught (m)
210-269.9	421	12.0
270-300	244	12.5

If the lowest allowed maximum draught was set to be 11 m, and the highest allowable to be 14 m, which was the median +- 1.5 m, 95% of all vessels in the 270-300 m group
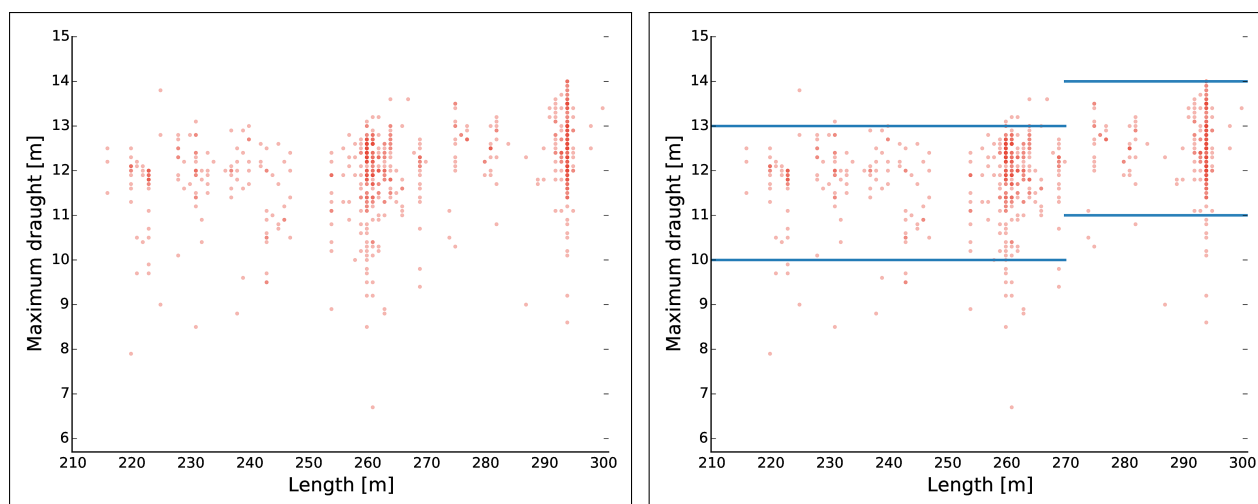
Figure 5.13: **Left:** Maximum draught and length for Panamax Vessels. **Right:** Lines indicating the lowest and highest maximum draught for the heuristic for the two groups

was included. Requiring that the maximum draught should be over 10 m and under 13 m, this included 94.4% of the other group. These seem like good constraints, as the selection was narrowed down, and outliers, which could reduce the quality of the heuristic, were removed. These constraints can be seen as blue lines in the right plot in Figure 5.13. There was a clear difference in maximum draught between these two selections, which comprised of 95% of the population in the two groups. This difference could most probably be accounted to differences in length.

The change in draught was more uniformly distributed over these two groups. 97.4% of the vessels had a change of draught of 5.5 m or less, indicated by the blue horizontal line in Figure 5.14.

Speed

The maximum recorded speed for the ships in this selection ranged from 12.9 knots to 24.9 knots. There was no clear difference between the two groups, the median highest recorded speed were the same although there was more variance in the 210-269.9 m length group. 92% of the vessels had a maximum speed at 15.9 knots or over, which was set as an constraint in the heuristic.
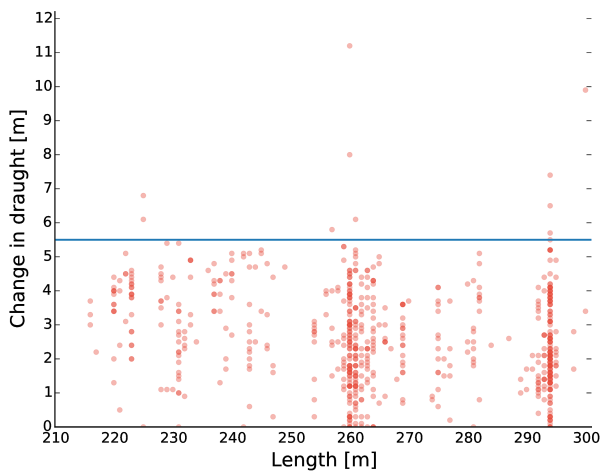
Figure 5.14: Change in draught for Panamax Container Vessels. 97.4% of the vessels had a change of draught less than 5.5 m, indicated by the blue line.
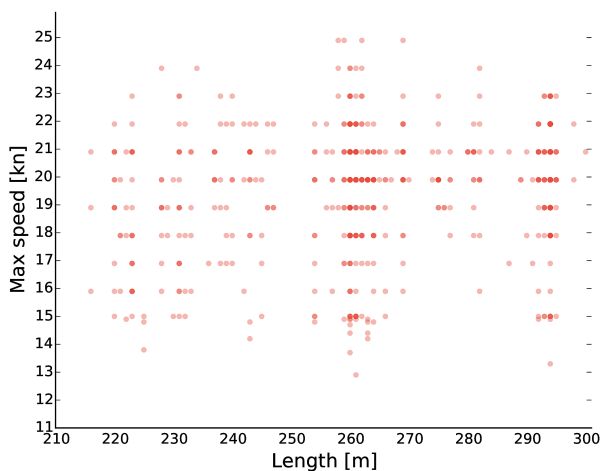
Figure 5.15: Maximum recorded speed for Panamax Container Vessels

Heuristic

Based on the template group and the proceeding analyses, the final heuristic for Panamax Container Vessels can be found in Table 5.10.

Table 5.10: Heuristic for identifying Panamax Container Vessels

| **Maximum speed:** | $>= 15.9$ kn |
| **AIS ship type:** | 70-79 (cargo ship) |

Length (m)	Breadth (m)	Number of ships	Max draught (m)	Max change of draught (m)
210-269.9	31-33	421	13	5.5
270-300	31-33	244	14	5.5

Heuristic testing and discussion

Out of the 244 ships in the template group for the longest vessels, 227 vessels were rediscovered with the heuristic. In addition, 52 vessels outside of this group were found. Out of these 52 vessels, only one was misidentified as a container vessel. This was the vehicle carrier ALGOL. This is an American military vehicle carrier, and thus a fairly uncommon ship. Of the 421 ships in the shortest group, 376 vessels from the ideal group were rediscovered. In addition, 152 ships outside of the template group were identified as Container Vessels. However, 77 of these vessels were misidentified. Most of these vessels were either Bulk Carriers or Vehicle Carriers. This means that it gets increasingly difficult to differentiate between vessels when the size is decreased. Bulk Carriers, Vehicle Carriers and Container Vessels seemingly have similar dimensions and operating patterns at this size.

5.3.3 Post-, New Panamax and Triple E Container Vessels

Using the heuristics in Table 5.10 in combination with the dimensions from Table 2.7, the heuristics for these three ship classes were defined. As Table 2.7 defines the average length and breadth inside each category, the heuristic was set with a somewhat arbitrary interval over that average. Since these size categories are somewhat arbitrary by themselves, a slightly wrong size classification could be tolerated. The speed requirement was carried over from the heuristic for Panamax Container Vessels.

Table 5.11: Heuristic for Post-, New-, Post New Panamax and Triple E container vessel.

Size category	Length (m)	Breadth (m)	Maximum change of draught (m)	Maximum speed (kn)
Post Panamax	270-315	40-43	5.5	>=15.9
New Panamax	320-370	46-52	5.5	>=15.9
Post New Panamax	380-397	54-58	5.5	>=15.9
Triple E	397-401	58-61	5.5	>=15.9

Heuristic testing and discussion

The heuristics performed well, with a very high level of accuracy. However, it seemed like the heuristic could have had somewhat less strict restrictions. In total 710 vessels were correctly identified. Mantell et al. (2014) categorised all these ship types in the Post Panamax category, and the total world fleet in this category consisted of 1,208 vessels, see Table 2.5). Some accuracy could have been sacrificed for better results. However, after inspection, there was no clearly better results when discarding the speed requirement, which seemed like the most viable option to identify more ships.

Table 5.12: Heuristic results for Post-, New-, Post New Panamax and Triple E container vessel.

Size category	Ships identified	Ships correctly identified	
Post Panamax	418	410	98%
New Panamax	284	284	100%
Post New Panamax	4	4	100%
Triple E	10	10	100%

5.4   Bulk Carriers

As seen in Table 2.5, the second most common ship type in the world is Bulk Carriers. Bulk Carriers typically carries unpacked dry cargo. The cargo is often coal, ore, grains, sugar or cement. Bulk Carriers have a high utilisation of their volume, as the cargo is held in several transverse cargo holds over the full ship breadth (Amdahl et al., 2011). Because of this high utilisation of the ships volume, a high maximum draught, as well

as a large difference between maximum and minimum draught, are expected. It is also expected that the maximum speed is lower compared to Container Vessels. The AIS ship types of Bulk Carriers and Container Vessels are both ship type 70-79.

5.4.1 Panamax Bulk Carriers

Out of the 2459 ships in the vessel sheets from Clarksons, 2200 vessels were present in the S-AIS data. This is a quite high number compared to the other ship classes. A possible explanation for this can be that it could have been a high activity for nearly all Panamax sized Bulk Carriers in the data period, compared to other ship classes.

Length and breadth

As with the Panamax sized Container Vessels, the breadth was required to be less than 34 m, as the maxmimum width of the Panama Canal is 33.5 m (the ships was allowed to be 0.5 m above this, because some ships could have been registered with a width of 34 m due to a rounding error). This is illustrated by the top horizontal line in the left plot in Figure 5.16. There was a lot of vessels over this breadth, which may seem strange as these are supposed to be Panamax sized ships. However, the ships in the vessel sheets were only registered with their name, so these dimensions could belong to ships having the same name in the S-AIS data. Because of the large number of vessels in the template group, it could be expected that a larger number of these vessels either would have wrong dimensions or were being mistaken for other vessels with the same name compared to smaller sized template groups. To ensure that only Panamax vessels were present in the selection, an additional breadth requirement of minimum 30 m was added. This can be seen by the bottom horizontal line. These breadth requirements reduced the selection down to 1668 ships. The final selection can be seen in the right plot in Figure 5.16.

From the plot it can be seen that three vessels had a length above 250 m. These three vessels were manually inspected in open ship databases. The longest ship, Vishva Anand, actually had a length of 229m, not the 332m it was registered with in the S-AIS data. The second longest ship was a container vessel misidentified as the Bulk carrier Santa Regina,
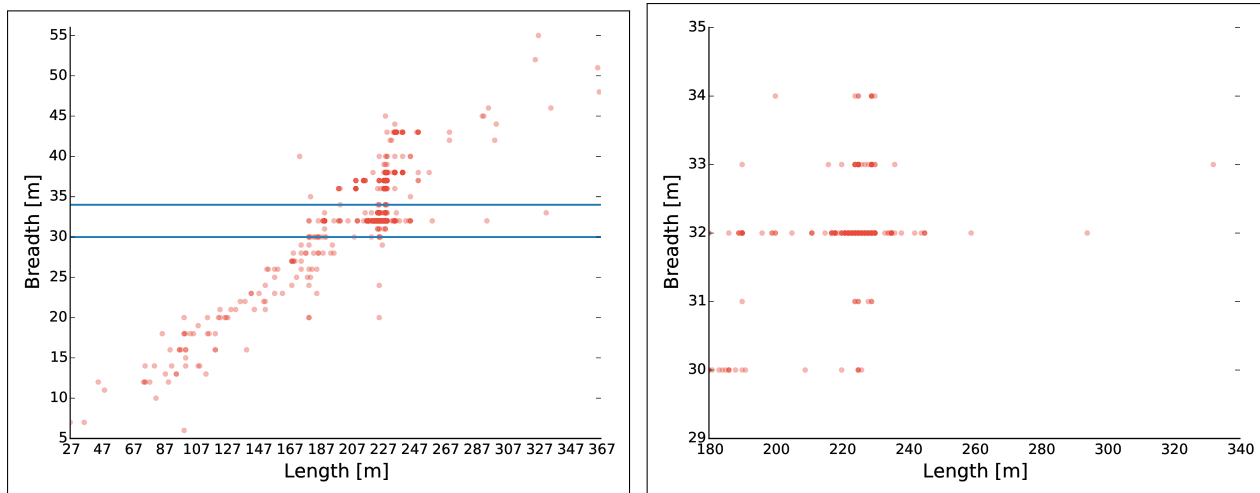
Figure 5.16: **Left:** Length and breadth for the ideal group of Panamax sized Bulk Carriers. The horizontal lines indicates the maximum and minimum allowed breadth. **Right:** Vessels inside the breadth requirements.

as they shared name. The last vessel is the 259 m long Bulk carrier Orissa. This is an exceptionally long bulk carrier, with a breadth of only 32 m. Since these three vessels either were wrongly registered or exceptionally large, all of them were excluded. This reduced the ideal group down to 1665 vessels. The length and breadth for this selection can be seen in Figure 5.17.
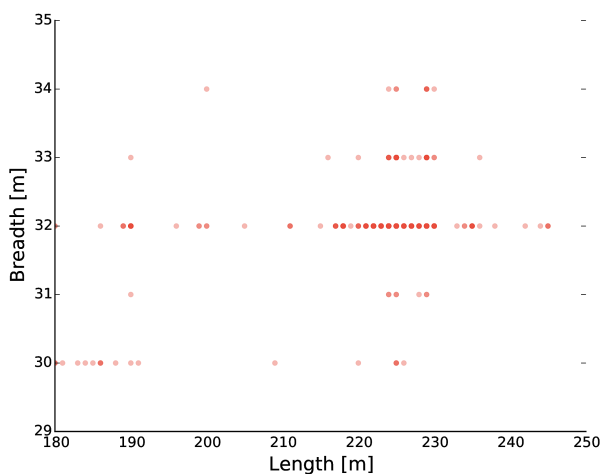
Figure 5.17: Length and breadth for the final selection of Bulk Carriers, where the breadth is between 30 and 34 m, and the length is under 250m.

68

Draught

From the draught plots in Figure 5.18 it can be seen that there were variations of the maximum draught across both lengths and breadths, and it was hard to find any correlation between these variables.
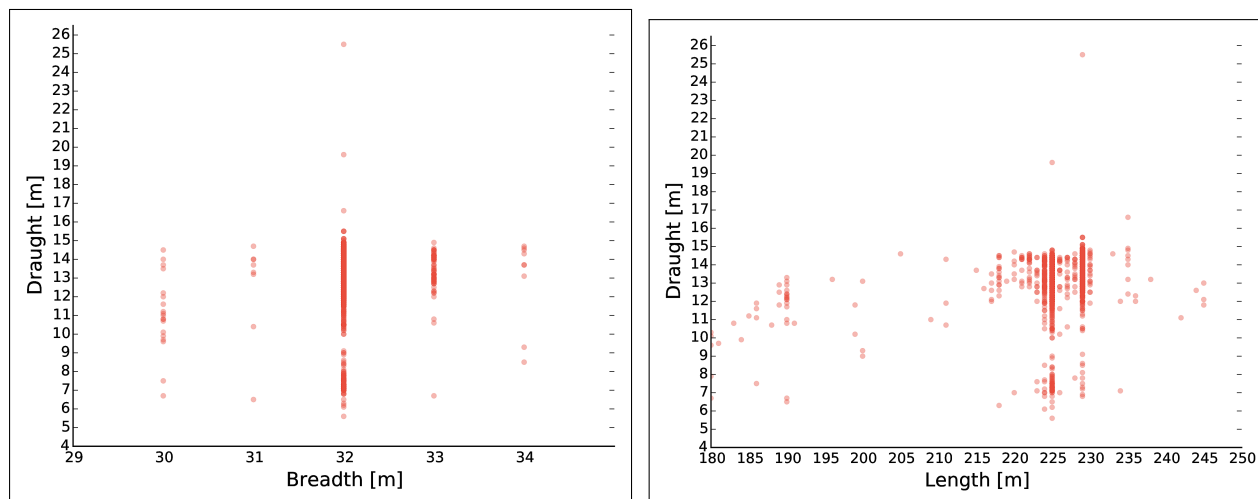
Figure 5.18: Draught, breadth and length of Panamax Bulk Carriers. **Left:** Maximum draught and breadth. **Right:** Maximum draught and length.

There were also a high variation of the maximum change in draught within the different length and breadth groups. The change in draught ranged from zero to just shy of 19 m. 99% of the vessels had a change of draught less than 9.5 m. However, as we wanted to distinguish the Bulk Carriers from the Cargo ships inside the same AIS ship type, we were more interested in establishing a lower limit for maximum change of draught, rather than a maximum limit. From the data Figure 5.14 was based on, we found that 97.4% of the Cargo Vessels had a change of draught of less than 5.5 m. When the lower limit for the maximum change of draught was set to 5.5 m, 27.4% of the Bulk Carriers were excluded, reducing the group to 1210 vessles. This seems like the only viable option to distinguish Bulk Carriers from Container Vessels based on their draught.

Speed

As many as 92% of the Container Vessels had a maximum speed of 15.9 knots or over, while 92% of the Panamax Bulk Carriers had a maximum speed of 15 knots or less. There

was a group having speeds up to 18 knots, which can be seen in Figure 5.19. This is most likely due to either specially favourable conditions, or other ships being misidentified as Bulk Carriers, as most Bulk Carriers are designed to have a speed of 15 knots (Amdahl et al., 2011).
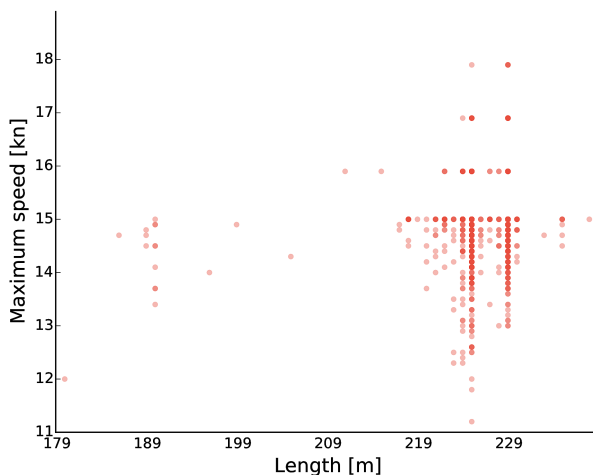
Figure 5.19: Maximum speed and length for the template Bulk Carrier group.

Heuristic

The heuristic for Panamax Bulk Carriers, based on the analyses above, is provided in Table 5.13.

Table 5.13: Heuristic for identifying Panamax Bulk Carriers

Maximum speed:		$<= 15$ kn	
AIS ship type:		70-79 (cargo ship)	
Length (m)	**Breadth** (m)	**Minimum draught** (m)	**Minimum change of draught** (m)
180-250	30-34	5	5.5

Heuristic testing

When testing the heuristic on the S-AIS data, a total of 6,024 vessels matched the dimensions in the heuristic. After applying the minimum change in draught and the maximum speed constraint, 2,346 vessels remained. This shows how important the draught and

speed constraint are for this ship type. 1,240 of these ships were in the template group, while 1,106 vessels were identified as Panamax Bulk Carriers outside the template group. 43 of the vessels outside the ideal group were misidentified as Bulk Carriers. Out of the misidentified ships, 38 were general cargo ships and three were Offshore Support Vessels. In addition, a Container Vessel and a Vehicles Carrier were amongst these ships. The rest of the ships outside the template group were correctly identified. The resulting accuracy of the heuristic, which is the number of ships correctly identified compared to the number of ships identified, was 98%. This is quite good for such a large group of vessels.

5.4.2 Capesize, Handymax and Handysize

Using the same methods as demonstrated for the Panamax Bulk Carriers, the heuristics for Capesize, Handymax and Handysize vessels were derived. This can be seen in Table 5.14. There was probably some overlap between Panamax vessels and Handymax vessels. To decrease the potential overlap of vessels, the required length was set somewhat short of the average length, as shown in Table 2.8.

Table 5.14: Heuristic for Capesize, Handymax and Handysize Bulk Carriers

Size category	Length (m)	Breadth (m)	Minimum change of draught (m)	Maximum speed (kn)
Capesize	230-320	36-50	5	15
Handymax	160-180	29-33	5	15
Handysize	130-180	20-29	5	15

Heuristic testing and discussion

Since these vessels vary in size, there may be some variations of draught for each size category. Preliminary testing of the heuristic showed that if the maximum allowed draught was decreased, the number of ships identified increased. At the same time, the number of wrongly identified vessels also increased, however this change was not linear. This means that there was a trade-off between having less strict constraints and having a decrease in accuracy from the heuristic. The results of the testing, with different maximum draught constraints, are shown in the following tables. Since the decrease in accuracy was so small for the Capesize and Handymax

Bulk Carriers, this process was only performed for two different draught constraints.

Table 5.15: Number of vessels identified, and correctly identified as Capesize Bulk Carriers

Minimum change in draught	5	0
Vessels identified	1,351	1,488
Vessels correctly identified	1,332 (98.6%)	1,429 (96%)

Table 5.16: Number of vessels identified, and correctly identified as Handymax Bulk Carriers

Minimum change in draught	5	0
Vessels identified	39	465
Vessels correctly identified	37(94.9%)	420 (91.3%)

Table 5.17: Number of vessels identified, and correctly identified as Handysize Bulk Carriers with different minimum change in draught requirements.

Minimum change in draught	5	4	3	2	1	0
Vessels identified	128	763	1,263	1,523	1,720	2,441
Vessels correctly identified	109	653	970	1036	1089	1298
%	(85.2%)	(85.6%)	(76%)	(69.8%)	(63.3%)	(53.2%)

The heuristic for the Capesize and Handymax Bulk Carriers had overall a pretty good performance, with a slight decrease in accuracy if the draught constraint was lowered. This lead to a slightly larger amount of ships identified in the Capesize class, and a drastically higher detection rate for the Handymax class. For the latter, the high increase of vessels identified is arguably worth the small decrease in accuracy. In the Capesize class, the decrease in accuracy, relative to the number of extra vessels identified, may not be worth it. In the Handysize class, the accuracy was at best 85.2%, which was lower than the other two classes. This could be due to the relative small size of these vessels, as there are both container and Ro-Ro vessels with similar dimensions. The number of ships identified increased greatly if the maximum draught constraint was decreased, however, the decrease in accuracy was also relatively high.

5.5   Oil Tankers

5.5.1   UL/VLCC Tankers

Very Large Crude Carriers (VLCC), as well as Ultra Large Crude Carriers (ULCC), has the largest dimensions out of all ships categorised as Oil Tankers. They are used primarily for so called long hauls, with transportation of crude oil from the Middle East to America or Asia. The data sheets with vessel names from Clarksons provided us with the name of 642 different vessels. 465 of these vessels were also identified in the S-AIS data. It should be noted that the Tanker market has been suffering through extremely high volatility in charter rates, and gone through periods where a lot of vessels have been either re-purposed or scrapped. As the Clarksons dataset contained vessels from as far back as 1993, this can explain some of the lacking data. As this is two extremely large ship classes, it was expected to be fairly easy to identify them, few ships have similar dimensions.

Length and breadth

Plotting the length and breadth of the VLCC/ULCC template group showed that these were extremely large vessels, with lengths between 308 and 378 m and breadths between 50 and 70 m. As can be seen in Figure 5.20, there were two breadth categories, 58 m and 60 m.
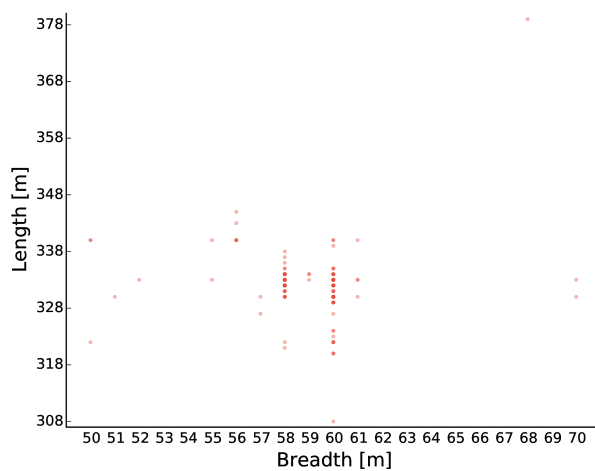
Figure 5.20: Breadth and length for ULCC and VLCC.

Draught

From the plot of length and maximum draught in Figure 5.21, it was evident that there was no relation between these two parameters. It rather seemed like the draught varied between 10 and 24 m, regardless of length. The vessels seemed to form two groups based on draught, one below 12 m and one over 18 m.
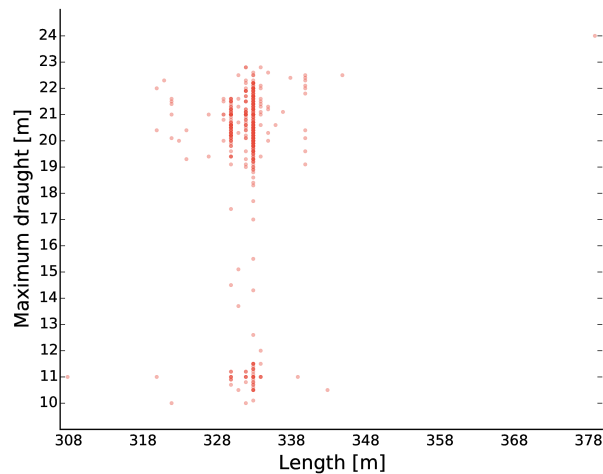
Figure 5.21: Maximum draught and length for ULCC and VLCC.

There was some variation in breadth inside different length groups, as seen in Figure 5.20. To investigate if this could explain the draught-grouping, a plot of breadth and draught was made, see From Figure 5.22. It did not seem like there was any correlation between breadth and draught, the same two distinct breadth groups was still evident.
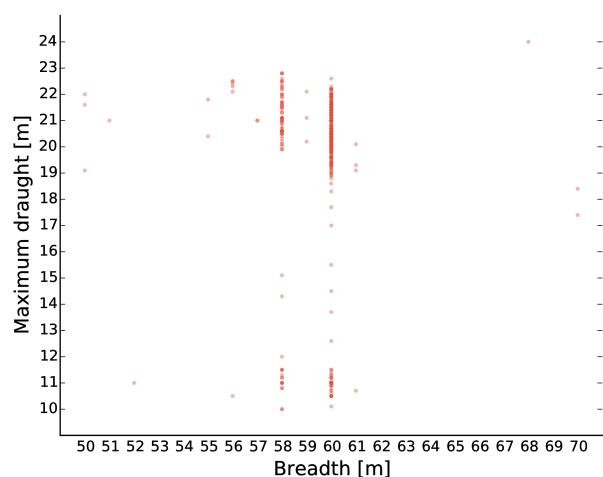
Figure 5.22: Maximum draught and breadth for ULCC and VLCC.

The grouping of maximum draught inside the same length and breadth categories could be explained by different load states. Some vessels may not have been fully loaded in

this period, and thus the maximum draught is lower. This was further supported by the minimum draught plot in Figure 5.23. Most vessels had a minimum draught between 10 and 12 m in this period. This means that the group with a maximum draught of less than 12 m, probably was unloaded throughout the time period.
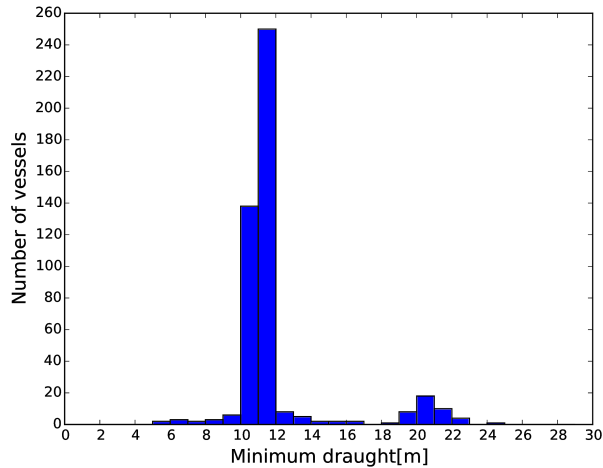
Figure 5.23: Minimum draught for ULCC and VLCC.

Looking at the change in draught, i.e the difference between maximum and minimum draught, plotted in Figure 5.24, there were two large groups. A high number of vessels had a change in draught between zero and one meters, which clearly means that these ships had not been close to fully loaded, and fitted into the previously discussed groups. Most ships had a change in draught between 8 and 12 m. If only the vessels with a change of draught over 2 m was counted, i.e. the vessels that were expected to be in operation, the maximum change in draught was 15 m and the median change in draught was 9.7 m with a standard deviation of 1.41 m.
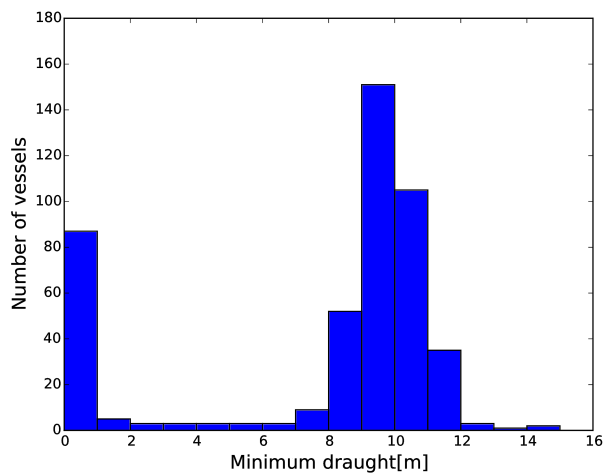
Figure 5.24: Change in draught for ULCC and VLCC.

Speed

As described earlier in this chapter, a maximum speed constraint was used to ensure that the maximum speed had enough recordings to be reliable. However, some of the Oil Tankers had maximum speeds inside somewhat reasonable intervals, even though they violated the maximum speed constraint. In this relatively small category, these were still included in the heuristic. The limit for what could be accepted, even with fewer than ten recordings of the maximum speed, was set to be a speed of 11 knots. 463 vessels had a maximum speed under 11 knots or at least 10 reports of a higher speed than 11 knots. The median maximum speed was 13.9 knots, while the highest maximum speed was 17.9 knots. A majority of the vessels had a maximum speed between 12 and 16 knots.
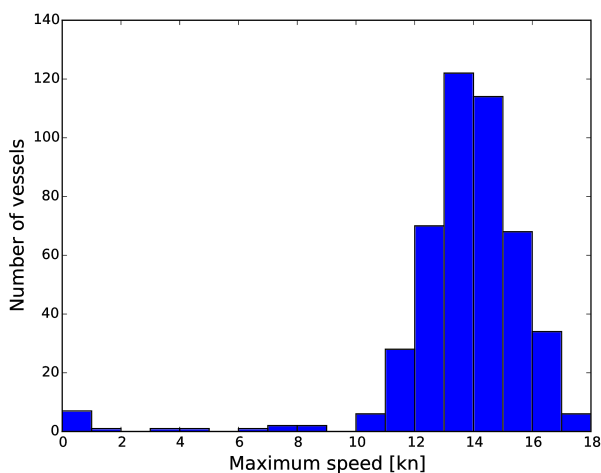
Figure 5.25: Maximum speed for ULCC and VLCC.

Heuristic

For LNG Carriers, the simplified heuristic performed better than the expanded heuristic, the heuristic would not be divided into different breadth and length categories. The heuristic for ULCC and VLCC vessels is summarised in Table 5.18.

Table 5.18: Heuristic for categorising ships as ULCC and VLCC, including minimum and maximum draught, minimum change in draught, maximum speed, AIS ship type and key dimensions.

Maximum draught:	25 m
Minimum draught:	10 m
Minimum change in draught:	8 m
Maximum speed:	$<= 16$ kn
AIS ship type:	80-89 (tanker)

Breadth (m)	**Length (m)**	**Number of ships in template group**
50-70	320-400	476

Heuristic testing and discussion

Using the heuristic in Table 5.18, 309 of the 463 vessels in the template group were identified. This decline was due to a minimum change in draught constraint of 8 m, and a maximum speed limit of 16 knots or less. A higher number of ships would be expected to be identified if the S-AIS data came from a longer time period. This is because the chance of a ship going from a minimum to a maximum loaded state is higher inside a longer time period. In addition to the 309 rediscovered vessels, 65 of the vessels discovered were not present in the template group. Out of these 65 vessels, only two ships were not Oil Tankers. This gave a 0.53% error margin, which means that the results obtained using this heuristic are accountable. The two misidentified ships were a barge and a high speed craft. These vessels had different dimensions in the S-AIS database and in the online open access vessel databases. One explanation for this might be that the dimensions in the S-AIS data were wrong. Another possible explanation is that the ship was registered with a wrong MMSI, which was used for the matching with the online databases.

5.5.2 Panamax, Aframax and Suezmax

Using the same methods demonstrated for the ULCC and VLCC, we defined heuristics for the Suezmax, Aframax and Panamax vessels. These are presented in the following table.

Table 5.19: Heuristic for Suezmax, Aframax and Panamax vessels

Size category	Length (m)	Breadth (m)	Minimum change in draught change in draught (m)	Maximum speed (kn)
Suezmax	265-320	45-50	5 (max 20 m draught)	16
Aframax	235-265	38-44	0	16
Panamax	200-235	30-33.5	4	16

Heuristic testing and discussion

Initially it was required that the maximum change in draught for the Aframax vessels had to be at least 3 m. Upon testing it was discovered that more vessels were found if this requirement was removed, with just a small increase in vessels misidentified as Oil Tankers. If there is a poor freight market, it can be expected that more Oil Tankers is idle, and thus do not have a change in draught over the period. Removing this requirement assured that these idle ships were identified. After the removal of this constraint, the accuracy for Suezmax and Panamax vessels dropped from above 99% to around 95%. This can be seen in the following tables. Again, there was a trade-off between having a large group or having a very high accuracy. However, an accuracy over 95% can still be satisfactory for most applications.

Table 5.20: Number of vessels identified, and correctly identified as Aframax Oil Tankers

Minimum change in draught	3	0
Vessels identified	443	645
Vessels correctly identified	441 (99.5%)	634 (98.3%)

Table 5.21: Number of vessels identified, and correctly identified as Suezmax Oil Tankers

Minimum change in draught	5	0
Vessels identified	356	433
Vessels correctly identified	354 (99.4%)	413 (95.4%)

Table 5.22: Number of vessels identified, and correctly identified as Panamax Oil Tankers

Minimum change in draught	4	0
Vessels identified	246	382
Vessels correctly identified	238 (96.7%)	357 (93.4%)

5.6 Heuristic Results and Discussion

Previously in this chapter, heuristics for the four most prevalent ship types in Malacca Strait are presented (Bateman et al., 2007). The heuristic for one size class for each ship type is meticulously explained, and the heuristics for the rest of the size classes for that ship type are derived. However, for Oil Tankers and Container Vessels, heuristics for the Sub Panamax vessels are not developed. This was omitted because the heuristics for the remaining ship classes is used as a proof of concept of this method, because of the time limit for this thesis. The number of vessels in the world fleet, the number of vessels identified by the heuristics in the S-AIS data, as well as the number of vessels correctly identified by the heuristic are given in Table 5.23. The number of vessels correctly identified are also reflected by the accuracy percentages. Note that for the ship classes where a range of results was given, depending on constraints such as the maximum change of draught, the result in the table comes from a trade-off between most ships identified and the highest accuracy. However, depending on the purpose of the studies where these heuristics can be used, the two properties can be given different importance. The reader should therefore consult the sections where the results for different constraints are outlined.

The heuristics for most ship classes performed well, where the accuracy for Bulk Carriers in total was 95.5%. The accuracy for Container Vessels was 94%, 96.8% for Oil Tankers over 10,000 DWT and 99% for LNG Carriers. The heuristics performed best in the large ship size categories, such as the LNG classes Q-Flex, Q-Max, and the Container vessel class Triple E. This is most likely because these large ships are quite extraordinary compared to other ships inside the same AIS ship type. The poorest performing size categories inside these ship classes were generally the smallest categories, such as Handysize Bulk Carriers, with as low accuracy as 85.6%. This is most likely due to the large range of different ships inside this ship class, with a range of different cargo types. These might even have a mix of different cargo types, making the ship types less distinct. When analysing the global ship traffic, one can argue that the large ship categories are of the highest interest, since they obviously tend to be involved in more global, ocean crossing traffic. Under this assumption, the heuristics presented here can give enough information for most usages.

In Table 5.23 the size of the world fleet, found in Table 2.5, is presented alongside the number of correctly identified vessels. The number of correctly identified vessels will

represent an upper limit of ships possible to identify with these heuristics using our S-AIS data. This must be taken into account when using the heuristics to perform analyses, such as in Chapter 6. However, with data for a longer time period, the number of vessels identified compared to the world fleet, is expected to rise. A longer period of time and more data means a higher possibility that ships exceed constraints such as minimum recorded top speed, or minimum change in draught. This means that when S-AIS data from a longer time period is obtained, new analyses of the performance of these heuristics should be performed.

The ship type where the heuristics discovered the smallest number of ships, compared to the world fleet, was with Bulk Carriers. Here the number of correctly identified vessels comprised only 46.8% of the world fleet. This was mostly due to the two smallest size classes, Handysize and Handymax, which also were the two largest classes in terms of the number of ships. As stated before, there is a trade-off in the heuristics between accuracy, and the number of ships identified. However, it may be possible to develop these heuristics further to achieve both high accuracy and a large quantity of identified vessels using other parameters or more refined techniques.

The rest of the different ship classes; LNG Carriers, Oil Tankers, Container Vessels and Bulk Carriers, the vessels identified by the heuristic covered over 73% of the world fleet, which is good.

Table 5.23: The number of vessels in the world fleet, and the performance of the different heuristics.

type	AIS ship ship type	Vessels in world fleet	Identified in S-AIS data	Correctly identified	Accuracy
LNG Carriers	Tankers	388	289	287	99%
General LNG Carriers			251	249	99%
Q-Flex			28	28	100%
Q-Max			10	10	100%
Oil Tankers[19]	Tankers	2,429	1,834	1,776	96.8%
Aframax		884	645	634	98.3%
UL&VLCC		624	374	372	99.4%
Suezmax		495	433	413	95.4%
Panamax		416	382	357	93.4%
Container Vessels[20]	Cargo	2,083	1,623	1,537	94%
Panamax		875	807	729	90.3%
Post panamax[21]		1,208			
Post Panamax			418	410	98%
New Panamax			384	384	100%
Post New Panamax			4	4	100%
Triple E			10	10	100%
Bulk Carriers	Cargo	10,053	4,926	4,708	95.5%
Handysize		3,095	763	653	85.6%
Handymax		3,008	465	420	91.3%
Panamax		2,405	2,346	2,303	98%
Capesize		1,590	1,352	1,332	98.6%

---

[18]All Oil Tankers over >10,000 dwt, exclusive Sub Panamax vessels.

[19]Exclusive Sub Panamax vessels.

[20]Mantell et al. (2014) counted all ship sizes larger than panamax as Post panamax, while Rodrigue and Ashar (2012) distinguished between these in their size classification which was a guidance for the heuristic.

Chapter 6

Traffic Through the Singapore Strait

6.1 Intersection of VTS-AIS and S-AIS Data

As described in Section 4.3, there was 17,026 distinct ships in the Singapore VTS-AIS data. 83.3% of these ships were also found in the S-AIS data. All the dynamic and static messages for these ships were put into its own database, for easier access. To analyse the global ship traffic into Singapore Strait, we are interested in S-AIS data from the time period prior to when they are discovered by the VTS-AIS data. There are two methods to retrieve this information. One method is to specifically look for ships that have Singapore set as their destination in the static messages in the S-AIS data. The other is to trace back ships from the Singapore VTS data day for day in the S-AIS data.

Since we are primarily interested in freight transport, we will focus on cargo ships and tankers in the following analyses. The resulting distribution between the ship types of the remaining 12,192 tankers and cargo ships that are both present in VTS-AIS and S-AIS data can be seen in Figure 6.1. As we can see there is a higher proportion of cargo ships compared to tankers.

6.1.1 Destination Singapore

The destination that is reported through the static message type 5 is manually set, as we can see from Appendix B. This means that there are a lot of missing, cryptic, coded or erroneous destination messages. There were fourteen different destinations that could, with high confidence, be attributed as Singapore. That was either just Singapore or Singapore
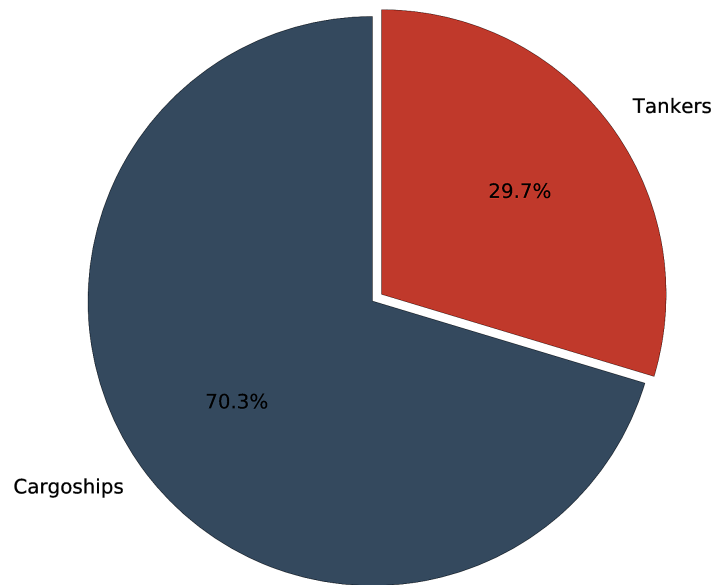
Figure 6.1: The distribution between tankers and cargo ships for the remaining vessels.

with the following suffixes: WBGA, ARMED GUA, EBG A, ABGA, ABGB, EPBG A, PEBG B, PEBGB, PJSB, WBGA, WPGB-A, WPBGA, -EBGB. To find the number of messages that contains Singapore with any of these suffixes we executed the following SQLite statement: 'SELECT COUNT(*) from staticMessage WHERE destination LIKE "%SINGAPORE%" '. This resulted in 102,746 messages from 3,397 different tankers and cargo ships from the S-AIS data. In comparison, the numbers of tankers and cargo ships in the VTS data that had set Singapore as destination, was 8,551. These ships can be of interest, as they also have reported an estimated time of arrival to Singapore, which can be matched with the actual time of arrival. Note that there probably is a lot of ships in the Singapore VTS data that is only passing through the Singapore Strait, and thus have not plotted Singapore as its destination. These ships that are in transit through Singapore Strait is also of interest.

6.1.2 Backtracking from VTS Data

To do an analysis of the ships arriving into Singapore Strait, it is of interest to have satellite data on the ships for as much of their entire voyage before entering Singapore Strait, as possible. Here the key strength of the S-AIS data comes to play, as it holds

global data for a long time period. These data can provide us with important information such as speed and position for the vessels in their voyage to Singapore Strait. To do this, a simple query to the database is made to find whether the S-AIS data contains information for the ships that are registered to arrive inside the Singapore area. To find the data directly linked to the ships that are arriving into Singapore strait, we query specifically after messages that are received by the satellite before they enter Singapore Strait. There was 145,092 arrivals of tankers and cargo ships into Singapore Strait over the whole VTS-AIS data period. To give a metric on how many of these arrivals that could be backtracked from the VTS-AIS data into the S-AIS data, we have inspected if we were able to find the ship that arrived into Singapore Strait inside a timeslot of 24 hours n days before arrival. Table 6.1 shows the result of this analysis, with n ranging from 1 to 15. As can be seen, the number of ships found increases slightly the first days upon day three. This is probably partly due to that Singapore Strait is a high traffic zone, where it is a low coverage rate. This phenomena is shown in Figure 2.2 in Section 2.2. As the time before arrival increases, more ships are expected to be outside of the high traffic zone of Singapore Strait. After the initial increase in vessels the first days, the number of ships rediscovered in the S-AIS data stays pretty contstant between 42,000 and 43,000, which is about 30% of the total arrivals at Singapore. From this it can be inferred that a lot of the traffic is local traffic, which never comes out of the high traffic zone with low detection rate. Note that some ships can have several arrivals at Singapore Strait inside this time period of 15 days, and thus can have been registered several times.

Table 6.1: Number of arriving vessels rediscovered in the S-AIS data at the n'th day before arrival at Singapore.

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
rediscovered	38,154	39,121	40,511	42,171	43,017	42,827	42,261	42,124	42,268	42,500	42,488	42,713	42,760	42,759	42,918

6.1.3 Tracing the Voyages to Singapore Strait

To follow the full voyages from Singapore Strait to their port of origin the following method was employed. Firstly, the MMSI for the ship and the arrival time for each arrival to Singapore was selected. Then all positional data from the satellite AIS data previous to the arrival time was retrieved. To determine what defined a journey's origin, the data item

navigational status was used. This is found in the dynamic AIS messages. Navigational status gives information on whether a ship is under way, anchored, moored, aground etc., see Appendix B. Using the navigational status, there was two different options to determine whether a ship was at the start of a voyage. These two are navigational status anchored and navigational status moored. Some ships can lie anchored awaiting freight, rather than moored. This could be a viable option. On the other hand, some ships can anchor up, to wait for passage either through the Suez or Panama Canal, or to await entry to port. In these cases a large part of a vessel's potential voyages can be lost. Instead, a combination of the navigational status moored and anchored was used. In the cases where the vessels was never moored, navigational status as anchored was selected. Using this method a total of 64,112 voyages from Singapore Strait was traced to their presumed origin. The median voyage length was 470 hours, just shy of 20 days, and the average number of received AIS messages per hour was 0.2. This is shown in a scatter plot in Figure 6.2. Note that the latter metric is skewed as a result of the binary nature of satellite detection. Every time one of the satellites pass over the ship it will receive a large number of messages, while there will be a period without any messages between every passing. Both a longer sailing duration, as well as a large number of dynamic messages received per hour from the ship, gives more data points, and is of higher statistical quality.
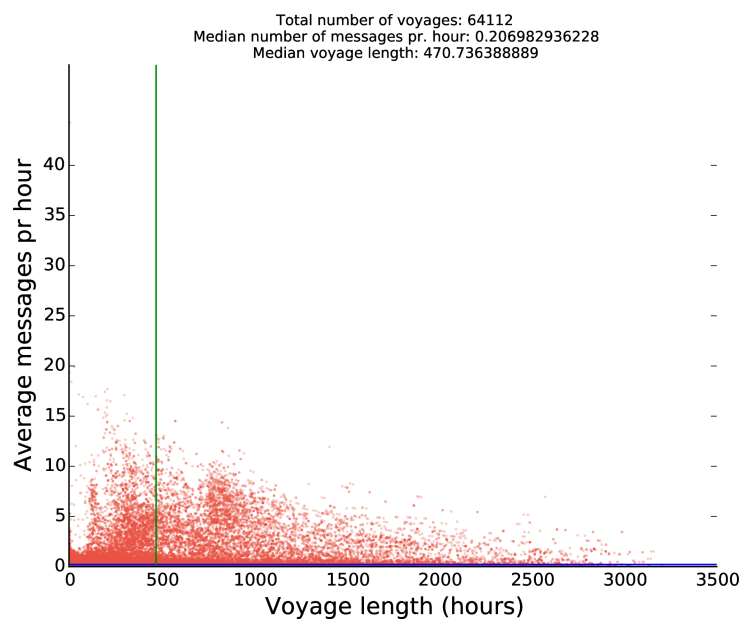

Figure 6.2: Scatterplot over voyage length and average number of messages per hour for each voyage. The horizontal line signifies the median number of messages per hour, while the vertical line signifies the median voyage length.

A selection of ships tracked to the port of origin is visualised by plotting their origin before the ships travelled to Singapore. This can be seen in Figure 6.3. Tankers are marked with a red dot, and cargo ships are marked with a blue dot. As can be seen from the figure, there are quite a lot of ships that have their origin in the area around Singapore. This leads to the assumption that there is a lot of local traffic, with relatively short sailing distances.
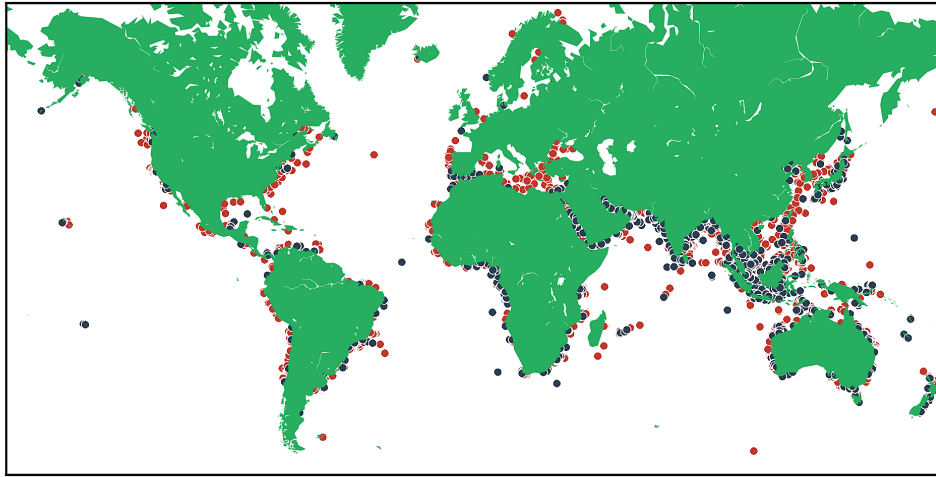
Figure 6.3: The origin of a selection of the tankers (blue) and cargo vessels (red) that could be backtracked from their entry to Singapore. These also have at least 6 dynamic messages per hour on average, a requirement made to easier visualise their voyage in an animation.

To further investigate the origins of the traffic to Singapore we have plotted the longitudinal coordinate for every origin in a histogram. This can be seen in Figure 6.4 for cargo ships, and in Figure 6.5 for tankers.

Singapore is positioned at 1.28 degree latitude and 103.83 longitude. The high frequency of ships inside the 100-110 longitude supports the theory that a lot of the traffic is from Singapore and the area around. As can be seen from the tanker data in the figure,

Figure 6.4: Histogram over longitudinal coordinates for the origin for every Cargo ship in our selection.

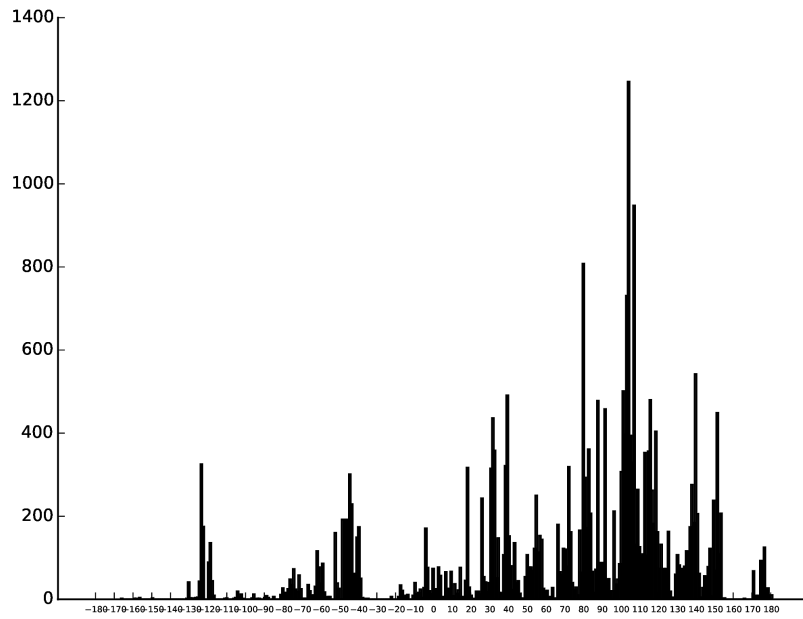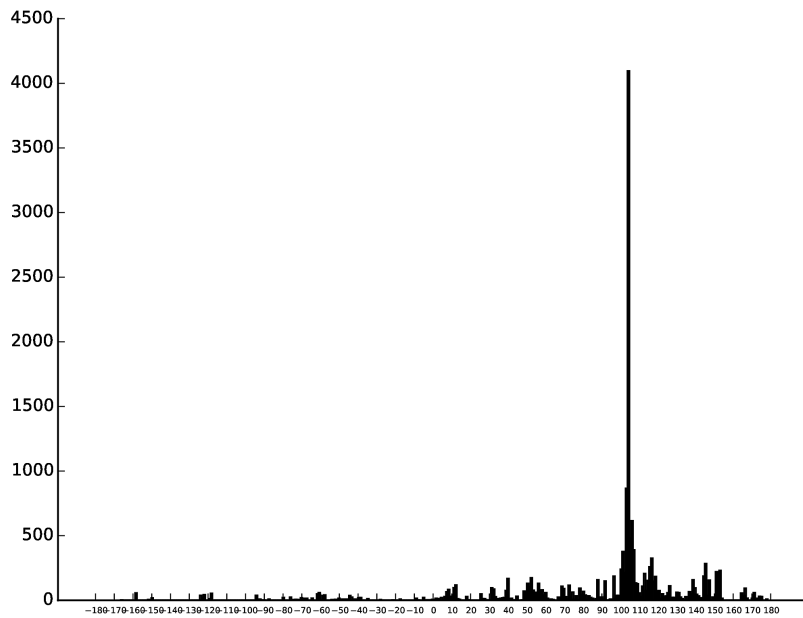Figure 6.5: Histogram over longitudinal coordinates for the origin for every tanker in our selection.

the vast majority of the tanker traffic was to the nearby areas, while the trend wasn't that clear for the cargo traffic. In Figure 6.4 the peaks around -130 to -120 degree longitude corresponds to the American north-western coast. The peak between -50 and -40 degree latitude corresponds to the Brazilian coast. The peaks between 30 and 40 degree latitude corresponds both to the western coast of Africa, as well as the eastern part of the Mediterranean and the Suez Canal. It is probable that a lot of ships are registered with the Suez Canal as the place of origin for the voyage, when in reality they have moored at, or nearby the canal as they wait for free passage through the canal. Other major peaks can be found at 80- to 90 degree longitude. This corresponds to the western Indian coast and Bangladesh. 130- and 150 degree longitude also has a high number of origins. This geographical area corresponds to Japan, South Korea, as well as the Indonesian- and the Australian coast.

6.2 Ship Types in the Singapore Strait

In Chapter 5, heuristics to identify the specific ship type of a vessel, based on its AIS data were defined. By using these heuristics the ship traffic in Singapore Strait can be determined down to the specific ship type. Sources like Bateman et al. (2007) count the number of voyages through the strait for each ship type, and not the number of distinct ships inside each ship type. There are no other good sources for this information. By applying the heuristics from Chapter 5 on the S-AIS data, the number of vessels of each specific ship type are found. As these heuristics had some margin of error, most below 5%[21], this must be accounted for if these numbers are to be be used in any further studies. Table 6.2 shows the number of distinct ships inside the specific ship types identified using our heuristic, as well as the total number of tankers and cargo ships in the Singapore Strait found in the S-AIS data, identified by the main ship type set in the AIS data. 38.7% of the tankers was identified by the heuristics. From this we can infer that a high proportion of the tankers either was chemical- or product tankers or relatively small oil tankers. This is supported by the histogram over longitudinal coordinates for the origin for the tankers in Figure 6.5. This figure indicated that there was a lot of local ship traffic, as there was a high number of ships inside the same longitudinal coordinates as Singapore.

---

[21] The accurate margin of error for each heuristic is found in Table 5.23.

The local ship traffic is, by logical reasoning, smaller ships using Singapore as a hub to further distribute oil and other goods to surrounding ports. The same argument hold for Cargo ships, where 40.6% of the vessels was identified by the heuristics, and the same high frequency inside these longitudinal coordinates was found. The poor performance for the smaller ship classes, shown in Section 5.4 contributed to the low proportion of detected ships for the cargo ships.

Although few large Container Vessels and LNG Carriers were identified by the heuristics, this is not a sign of poor performance, but rather a consequence of the relative small fleet in these size categories. It seems like the previous argument that large ships traffic Singapore Strait holds true, as 51.8% of the world's UL&VLCC fleet was identified in this analysis. This was actually 87% of all identified UL&VLCC in our global S-AIS data. In addition, all of the Triple E and Post New Panamax, as well as 64.5% of the New Panamax ships in the global S-AIS data was found in Singapore Strait.

Table 6.2: Specific ship types in Singapore Strait, identified using heuristics from Chapter 5.

Ship type	Number of vessels
Tankers in total in Singapore Strait	2,978
Tankers identified by heuristic	1,152
LNG and LPG carriers	
General group	150
Q-Flex	25
Q-Max	10
Oil Tankers	
Aframax	326
VLC & ULCC	323
Suezmax	185
Panamax	133
Cargo ships in total in Singapore Strait	7,508
Cargo ships identified by heuristic	3,051
Bulk carriers	
Panamax	1469
Capesize	793
Handysize	36
Handymax	13
Container vessels	
Panamax	262
New Panamax	248
Post Panamax	216
Triple E	10
Post New Panamax	4

92

Part IV

Discussion and Conclusion

Chapter 7

Discussion

As stated in the introduction, the results have been discussed when they were presented. However, there are some points that are valid for the thesis as a whole.

7.1 Representative Data

The heuristics developed in Chapter 5 are partly based on information in the S-AIS data such as draught and speed. In particular, it should be noted that the maximum and minimum draught, as well as the maximum speed, are products of the operating conditions of the vessels. These operating conditions can be affected by both seasonal micro-variations, as well as yearly or even annual macro-variations. Corbett et al. (2009) has showed that the average speed of the shipping fleet can be influenced by fuel cost, which in this study takes the form of fuel taxes. Maximum draught is influenced by the loading condition of the ship, which is influenced by the market the ship operates in. A strong market means a large quantity of goods are transported, and the maximum draught recorded is therefore expected to be higher compared to a weak market. In a strong market, ships can be expected to have a different operating speed as opposed to a weak market as well. As the S-AIS data spans just short of a five month period within a year, this may result in the heuristics performing differently with AIS data from another time period. However, these heuristics are also partly based on static information such as ships dimensions as well as a large number of ships, and this can negate some expected variations. The high number of AIS messages containing speed data for each ship, 2,281 on average, is so large that it can be assumed with reasonable confidence that

the maximum speed found when designing the heuristics here is close to the maximum speed under other circumstances.

7.2 Quality of S-AIS Data

As shown in Section 3.6, S-AIS data can be erroneous in several ways. Even though this was considered while performing analyses throughout the thesis, erroneous data may still be present. Erroneous data that are not blatantly erroneous, such as erroneous MMSI or IMO numbers, or extremely unlikely dimensions and ship positions, may only be discovered if all ships' dimension data are checked against ship registries. However, there are safety in numbers, and with approximately 95.05% of the world fleet covered by the AIS data, we expect the amount of correct information to outweigh the possible erroneous information.

7.3 Heuristic for Identifying Specific Ship Type from S-AIS Data

Heuristics for the most common ship types, in terms of dead weight, in the Singapore Strait were developed. However, heuristics were not developed for the full range of ship sizes, as the largest ship sizes were prioritised. This decision was made because these ships are more prevalent in international traffic, and to limit the size of this thesis. On the other hand, these relatively small sized ships constitutes a large share of the world's ship fleet, and heuristics should be developed in the future. The heuristics for the other ship classes are a proof of concept in developing these heuristics.

In the heuristics, precision and detection rate had an inverse relationship. Precision had to be sacrificed for a greater quantity of ships to be identified by the heuristic, and vice versa. In the heuristic results, what we saw as a middle ground between these two was presented. However, depending on the purpose of the study where these heuristics should be used, the prioritisation between these two properties can vary, so the reader should console the chapter where the results for different constraints are outlined.

These heuristics rely heavily on information found in the static AIS messages, but not

every ship in the AIS data had this information. This means that the number of ships found in the static AIS messages poses a limit on the number of ships that can be used as a basis for these heuristics, and conversely the number of ships that can be identified by the heuristic. However the number of ships identified by the heuristic constituted a large part of the world fleet in most larger ship sizes.

Another limitation is that when template groups were used to develop the initial heuristics, the only way to retrieve information for these vessels from the Satellite AIS data was through the ship's name. This was the case because the ships were only identified by their name in the vessel sheets that formed the template groups. As a ship's name is not a unique property, such as the MMSI or IMO number, it is expected that a quantity of the ships in the template group can be from another ship type, class or size. To avoid this, the template groups themselves were refined through a manual process, where ships not abiding the expected dimensions were sorted out. The development of the heuristic was in all an iterative process, where experience based constraints was put in place to ensure a best possible template group. Nonetheless, the heuristic results proved the proficiency of this method.

7.4 Different Data Periods

In Section 6.1 the intersection between the VTS-AIS data and S-AIS data was analysed, and this was used to analyse the traffic through Singapore Strait. The S-AIS data spanned the time period from 1st May to 15th September 2014, while the VTS-AIS data spanned the time period from 25th May to 25th September. This means that the ship traffic in the Singapore Strait after 15th September would not be found in the S-AIS data until up to 10 days earlier. It is worth stressing that this can explain some of the variance displayed in the number of vessels rediscovered in Table 6.1. As mentioned in Section 6.1.2, the low detection rate in Singapore, as shown in Figure 2.2 in Section 2.2, is probably also a factor, as some ships might have remained undiscovered by the Satellites.

7.5 Traffic Through the Singapore Strait

When the traffic through the Singapore Strait was examined in Chapter 6, there were two ways to retrieve this information. One was to select vessels that had Singapore set as their destination, which could be found in the static messages. However, there were only a limited number of vessels that had Singapore set as their destination. In addition, only vessels that specifically were set to go to Singapore would have this information, which therefore means that vessels travelling through the strait to other destinations would not be covered by the method. The other method, which was used in this thesis, was to match the vessels registered in the VTS-AIS data from the Singapore Strait, with the vessels in the S-AIS data. With this method, a lot more vessels were covered. However, about half of the ships registered in the VTS-AIS data could not be found in the satellite AIS data. This shows the weakness of the S-AIS data, as local and inshore traffic often cannot be detected due to interference. To study the voyages into the Singapore Strait, which was the travel a ship had from its port of origin into the Singapore Strait, the origin was defined as the last place a ship had its navigational status set as moored. In the cases where we had no record of the ship's navigational status set as moored, the origin was instead defined as the last place a ship had the status set as anchored. However, this method also have some weaknesses. In the cases where a ship moores to wait for passing through the Suez or Panama canal, or in the cases where a ship anchors up just off Singapore Strait, the voyages are cut short from reality. In the cases where a ship does not moore between two voyages, a voyage may be extended. This method also relies on the ship changing its navigational status when it should. As this is done manually, it does not need to be the case.

7.6 Recommendations for Further Work

There is a wide range of uses of Satellite AIS data, however there are some certain studies that would be natural extension of this thesis. Emission analyses for the ship traffic in Singapore Strait, through a combination of the heuristics presented and Satellite AIS data can readily be conducted, as this has not previously been possible without access of commercial databases. A global or local study on the prevalence of slow steaming could

also be run, using Satellite AIS data and the same methods used to analyse the traffic in the Singapore Strait. Heuristics can also be developed for a wider range of vessels, especially for the smaller size classes, and be tested on data from a longer time period. With the specific ship type derived from these heuristics, together with speed data and positional data, it is possible to give an estimate on the greenhouse gas emissions from different types of ships, ship types or shipping lanes. This means that with historical satellite AIS data, it is possible to monitor whether measures to reduce greenhouse gas emissions such as slow steaming is more or less prevalent than before.

Chapter 8

Conclusion

One of the aims of this thesis was to investigate whether Satellite AIS data was a reliable source of information. Analyses showed that these data can be erroneous. Ship dimensions, ship ID (IMO and MMSI numbers), geographical positions and especially speed recordings were erroneous in some data. This means that when working with AIS data, one have to be cautious. Sorting out these erroneous data is a fairly easy process. This should be done before any analyses on the basis of Satellite AIS data are to be carried out, or else the erroneous data has to be taken into account when the results are interpreted. Satellite AIS data is collected by satellites that have a certain passing rate and a certain detection probability. This means that some geographical positions can be overrepresented in the global data. This should also be put into consideration before any studies using AIS data are conducted.

To show the potential use of S-AIS data, heuristics were developed. These heuristics identified the specific ship type of a vessel, based on the information in the S-AIS data. This makes it possible to do refined studies, where specific ship type is a factor, without having access to vessel databases. This is especially helpful when working with a larger number of vessels. The heuristics showed a high accuracy, and covered a fairly large proportion of the world fleet. These heuristics is unprecedented in literature, and may be a very important finding.

Another aim for this thesis was to investigate whether Satellite AIS data could be used to track ships entering the Singapore Strait. Using a combination of Satellite AIS data and AIS data from the VTS in Singapore, a large number of vessels could be tracked back to

their port of origin. A lot of the local traffic was however not tracked back to its origin, as this traffic did not leave the highly trafficked Singapore Strait and the surrounding regions. When vessels are tracked back to their origin, the data over their voyage can be used in a number of studies, such as emission analyses and safety research.

Using the previously mentioned heuristics, the ship traffic into Singapore Strait was differentiated into specific ship types. There are few published studies that give this information, especially at that level of detail. This illustrates the feasibility of using heuristics in practical studies. A combination of the two latter findings can provide information, such as speed, prevalence of slow steaming etc.,which permits detailed emission and operational studies on a level previously only achieved by a few large organisations such as IMO.

Bibliography

Amdahl, J., Endal, A., Fuglerud, G., Hult, L. R., Minsaas, K., Rasmussen, M., Sillerud, B., Sortland, B., and Valland, H. (2011). *TMR4105 - Marin Teknikk Intro, Kompendium.* Marin Teknisk Senter, NTNU.

Bateman, W., Ho, J., and Mathai, M. (2007). Shipping patterns in the malacca and singapore straits: an assessment of the risks to different types of vessel. *Contemporary Southeast Asia: A Journal of International and Strategic Affairs*, 29(2):309–332.

Buhaug, O., Corbett, J., Endresen, O., Eyring, V., Faber, J., Hanayama, S., Lee, D., Lindstad, H., Mjelde, A., Palsson, C., et al. (2009). Second imo greenhouse gas study. *International Maritime Organization, London.*

ClarksonsGroup (2015). Clarkson research, shipping intelligence network. `https://sin.clarksons.net/`.

Corbett, J. J., Wang, H., and Winebrake, J. J. (2009). The effectiveness and costs of speed reductions on emissions from international shipping. *Transportation Research Part D: Transport and Environment*, 14(8):593–598.

Eriksen, T., Skauen, A. N., Narheim, B., Helleren, O., Olsen, Ø., and Olsen, R. B. (2010). Tracking ship traffic with space-based ais: experience gained in first months of operations. In *Waterside Security Conference (WSS), 2010 International*, pages 1–8. IEEE.

Helleren, Ø., Olsen, Ø., Narheim, B., Skauen, A., and Olsen, R. (2012). Aissat-1–2 years of service. European Space Agency Small Satellite Systems and Services Symposium.

Hoye, G. K., Eriksen, T., Meland, B. J., and Narheim, B. T. (2008). Space-based {AIS} for global maritime traffic monitoring. *Acta Astronautica*, 62(2-3):240 – 245.

IALA (2011). Guideline no. 1082 on an overview of ais.

IMO (1974). International convention for the safety of life at sea (solas), 1974.

IMO (1987). Resolution a.600(15).

IMO (1993). International code for the construction and equipment of ships carrying liquefied gases in bulk.

IMO (2002). Guidelines for the onboard operational use of shipborne automatic identification systems (ais),resolution a.917(22).

ITU (2014). *M.1371-5:Technical characteristics for an automatic identification system using time-division multiple access in the VHF maritime mobile band.* International Telecommunication Union.

Lane, B. C. (2006). Ais parser sdk v1.10. `https://github.com/bcl/aisparser`. Accessed 10.09.2014.

Mantell, C., Benson, R., Stopford, M., Crowe, T., and Gordon, S. (2014). Shipping intelligence weekly, 02 may 2014. Technical report, Clarkson Research Services Limited.

Moon, K., Chang, D., Lee, D., Kim, M.-B., Ahn, H.-J., and Ha, J.-P. (2005). Comparison of spherical and membrane large lng carriers in terms of cargo handling.

Rodrigue, J. P. and Ashar, A. (2012). Evolution of containerships.

Skauen, A. N., Helleren, O., Olsen, O., and Olsen, R. (2013). Operator and user perspective of fractionated ais satellite systems. In *27 th Annual AIAA/USU Conference on Small Satellites.*

Smith, D. T., Jalkanen, D. J.-P., Anderson, B., Corbett, D. J. J., Faber, D. J., Hanayama, S., O'Keeffe, E., Parker, S., Johansson, L., Aldous, L., Raucci, C., Traut, M., Ettinger, S., Nelissen, D., Lee, P. D., Ng, S., Agrawal, A., Winebrake, D. J. J., Hoen, M., Chesworth, S., and Pandey, A. (2014). Third imo ghg study 2014. Technical report, International Maritime Organization (IMO) London, UK, London, UK.

SOLAS (2011). Solas 1974: Brief history - list of amendments to date and where to find them.

USCG (2012). Automatic identification system encoding guide.

Wahl, T., Hoye, G., Lyngvi, A., and Narheim, B. T. (2005). New possible roles of small satellites in maritime surveillance. *Acta Astronautica*.

Part V

Appendices

Appendix A

Project description

Automatic Identification System (AIS) is a system that automatically sends positional, voyage related and ship identification data that usually is collected by base stations. Recently, it has been launched satellites that collect these data from all over the world. As this is a fairly new development and there has not been a lot of studies on the quality and potential use of these data. This is for interest, as it is important to know if the Satellite AIS data is a reliable source of information. In addition, it can be investigated if ordinary AIS data, from Singapore Strait combined with S-AIS data can be combined to give insight of the traffic through Singapore Strait.

Previous studies, which has used Satellite AIS data to perform emission estimates have used commercial ship databases to find the specific ship type of the ships in the study. Since AIS data contains information on a ship's dimensions, as well as draught and speed, it may be possible to determine the specific ship type based on these data alone.

The student shall give an example on how such large amounts of data can be used to give meaningful insight in both global and regional ship traffic by doing the following.

1. Collect, process and present Satellite AIS data.

2. Discuss the quality of Satellite AIS data, both in terms of accuracy and the amount of information given.

3. Show how these data can be combined with AIS data to give further insight.

4. Present new uses of Satellite AIS data by developing heuristics to identify the specific ship type

Appendix B

AIS Data Contents

Table B.1: Information item and information type, generation and quality, IMO (2002).

Information item	Information generation, type and quality of information
Static	
Maritime Mobile Service Identity (MMSI)	Set on installation
Call sign and name	Set on installation
IMO Number	Set on installation
Length and beam	Set on installation or if changed
Type of ship	Select from pre-installed list
Location of position fixing antenna	Set on installation
Dynamic	
Ship's position with accuracy indication and integrity status	Automatically updated from the position sensor connected to AIS. The accuracy indication is for better or worse than 10 m.
Position time stamp in UTC	Automatically updated from the position sensor connected to AIS

Course over ground (COG)	Automatically updated from ship's main position sensor connected to AIS, if that sensor calculates COG This information might not be available
Speed over ground (SOG)	Automatically updated from the position sensor connected to AIS This information might not be available
Heading	Automatically updated from the ship's heading sensor connected to AIS
Navigational status	Navigational status information has to be manually entered by the OOW[1] and changed as necessary, for example: -underway by engines -at anchor -not under command (NUC) -restricted in ability to manoeuvre (RIATM) -moored -constrained by draught -aground -engaged in fishing -underway by sail In practice, since all these relate to the COLREGs[2], any change that is needed could be undertaken at the same time that the lights or shapes were changed
Rate of turn (ROT)	Automatically updated from the ship's ROT sensor or derived from the gyro This information might not be available

---

[1]Officer Of the Watch
[2]The International Regulations for Preventing Collisions at Sea

Voyage-related	
Ship's draught	To be manually entered at the start of the voyage using the maximum draft for the voyage and amended as required (e.g.– result of de-ballasting prior to port entry)
Hazardous cargo (type)	To be manually entered at the start of the voyage confirming whether or not hazardous cargo is being carried, namely: DG (Dangerous goods) HS (Harmful substances) MP (Marine pollutants) Indications of quantities are not required
Destination and ETA[3]	To be manually entered at the start of the voyage and kept up to date as necessary

---

[3]Estimated Time of Arrival

Appendix C

AISDecode.py

```python
#!/usr/bin/python
#    Project: Master Thesis
#    Author: Bjornar Brende Smestad
#    Created: Spring 2015


import aisparser
import sqlite3 as lite
import sys
import os


def extractMessages(filepath):
    global messageType1
    global messageType2
    global messageType3
    global messageType4
    global messageType5
    global timeStamps1
    global timeStamps2
    global timeStamps3
    global timeStamps4
    global timeStamps5
    messageType1 = []
    messageType2 = []
    messageType3 = []
```

```python
27        messageType4 = []
28        messageType5 = []
29        timeStamps1 = []
30        timeStamps2 = []
31        timeStamps3 = []
32        timeStamps4 = []
33        timeStamps5 = []
34        s = []
35        i = 0
36        f = open(filepath, 'r')
37        for line in f:
38            s.append('c:'+line.split('c:')[1].split('*')[0]+'!BSVDM'+line.split
    ('!BSVDM')[1])
39        ais_state = aisparser.ais_state()
40        for p in s:
41
42            result = aisparser.assemble_vdm( ais_state, p )
43            if( result == 0):
44                timestamp = p.split('c:')[1].split('*')[0].split('!')[0]
45                ais_state.msgid = aisparser.get_6bit( ais_state.six_state, 6 )
46                i = i+1
47                if ais_state.msgid == 1:
48                    msg = aisparser.aismsg_1()
49                    aisparser.parse_ais_1( ais_state, msg )
50                    timeStamps1.append(timestamp)
51                    messageType1.append(msg)
52                    print msg
53                elif ais_state.msgid == 2:
54                    msg = aisparser.aismsg_2()
55                    aisparser.parse_ais_2( ais_state, msg )
56                    timeStamps2.append(timestamp)
57                    messageType2.append(msg)
58                    print msg
59                elif ais_state.msgid == 3:
60                    msg = aisparser.aismsg_3()
61                    aisparser.parse_ais_3( ais_state, msg )
62                    timeStamps3.append(timestamp)
63                    messageType3.append(msg)
```

VIII

```python
64                    print msg
65                elif ais_state.msgid == 4:
66                    msg = aisparser.aismsg_4()
67                    aisparser.parse_ais_4( ais_state , msg )
68                    (status ,lat_dd ,long_ddd) = aisparser.pos2ddd(msg.latitude ,
        msg.longitude)
69                    timeStamps4.append(timestamp)
70                    messageType4.append(msg)
71                    print msg
72                elif ais_state.msgid == 5:
73                    msg = aisparser.aismsg_5()
74                    aisparser.parse_ais_5( ais_state , msg )
75                    timeStamps5.append(timestamp)
76                    messageType5.append(msg)
77                    print msg
78            if i > 1:
79                break
80
81  def createDatabase(databasepath):
82
83      con = lite.connect(databasepath)
84
85      with con:
86
87          cur = con.cursor()
88          cur.execute("CREATE TABLE MessageType1(unixtime int , cog INT,
        latitude INT, longitude INT, msgid INT, nav_status INT, pos_acc INT,
        raim INT, regional INT, repeat INT, rot INT, slot_timeout INT, sog INT,
        spare INT, sub_message INT, sync_state INT, true INT, userid INT,
        utc_sec INT)")
89          cur.execute("CREATE TABLE MessageType2(unixtime int , cog INT,
        latitude INT, longitude INT, msgid INT, nav_status INT, pos_acc INT,
        raim INT, regional INT, repeat INT, rot INT, slot_timeout INT, sog INT,
        spare INT, sub_message INT, sync_state INT, true INT, userid INT,
        utc_sec INT)")
90          cur.execute("CREATE TABLE MessageType3(unixtime int , cog int ,keep
        INT,latitude INT,longitude INT,msgid INT,nav_status INT,num_slots INT,
        pos_acc INT,raim INT,regional INT,repeat INT,rot INT,slot_increment int ,
```

```
           sog int , spare INT, sync_state INT, true int , userid INT, utc_sec INT)")
91         cur . execute ("CREATE TABLE MessageType4(unixtime int , latitude INT,
           longitude INT, msgid INT, pos_acc INT, pos_type INT, raim INT, repeat INT,
           slot_timeout INT, spare int , sub_message int , sync_state INT, userid INT,
           utc_day INT, utc_hour INT, utc_minute INT, utc_month INT, utc_second INT,
           utc_year int )")
92         cur . execute ("CREATE TABLE MessageType5(unixtime int , callsign
           string , dest string , dim_bow int , dim_port INT, dim_starboard INT, dim_stern
           int , draught INT, dte INT, eta INT, imo INT, msgid INT, name text , pos_type INT
           , repeat INT, ship_type INT, spare INT , userid INT, version INT)")
93         cur . execute ("CREATE INDEX userid_index ON MessageType1 ( userid )")
94         cur . execute ("CREATE INDEX userid_index2 ON MessageType2 ( userid )")
95         cur . execute ("CREATE INDEX userid_index3 ON MessageType3 ( userid )")
96         cur . execute ("CREATE INDEX userid_index4 ON MessageType4 ( userid )")
97         cur . execute ("CREATE INDEX userid_index5  ON MessageType5 ( userid )"
           )
98         cur . execute ("CREATE INDEX unixtime_index ON MessageType1 ( unixtime )
           ")
99         cur . execute ("CREATE INDEX unixtime_index2 ON MessageType2 ( unixtime
           )")
100        cur . execute ("CREATE INDEX unixtime_index3 ON MessageType3 ( unixtime
           )")
101        cur . execute ("CREATE INDEX unixtime_index4 ON MessageType4 ( unixtime
           )")
102        cur . execute ("CREATE INDEX unixtime_index5  ON MessageType5 (
           unixtime )")
103
104 def writeToDatabase ( databasepath ):
105     con = lite . connect ( databasepath )
106     con . isolation_level = None
107
108     with con :
109         cur = con . cursor ()
110         cur . execute ( 'BEGIN TRANSACTION' )
111         for i in range (0 , len ( messageType1 )):
112             ( status , lat_dd , long_ddd ) = aisparser . pos2ddd ( messageType1 [ i ] .
           latitude , messageType1 [ i ] . longitude )
113                 cur . execute ("INSERT OR IGNORE INTO MessageType1(unixtime , cog ,
```

```
            latitude , longitude , msgid , nav_status , pos_acc , raim , regional ,
            repeat , rot , slot_timeout , sog, spare , sub_message , sync_state , true
            , userid , utc_sec ) VALUES(?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?)",
114                     (timeStamps1[i], messageType1[i].cog, lat_dd , long_ddd , ord
            (messageType1[i].msgid),ord(messageType1[i].nav_status),
115                         ord(messageType1[i].pos_acc), ord(messageType1[i].raim)
            , ord(messageType1[i].regional), ord(messageType1[i].repeat),
            messageType1[i].rot ,
116                         ord(messageType1[i].slot_timeout),messageType1[i].sog,
            ord(messageType1[i].spare), messageType1[i].sub_message , ord(
            messageType1[i].sync_state),
117                         messageType1[i].true, messageType1[i].userid , ord(
            messageType1[i].utc_sec)))
118         for i in range(0, len(messageType2)):
119             (status,lat_dd,long_ddd) = aisparser.pos2ddd(messageType2[i].
            latitude , messageType2[i].longitude)
120             cur.execute("INSERT OR IGNORE INTO MessageType2(unixtime , cog ,
            latitude , longitude , msgid , nav_status , pos_acc , raim , regional ,
            repeat , rot , slot_timeout , sog, spare , sub_message , sync_state , true
            , userid , utc_sec ) VALUES(?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?)",
121                     (timeStamps2[i], messageType2[i].cog, lat_dd , long_ddd , ord
            (messageType2[i].msgid),ord(messageType2[i].nav_status),
122                         ord(messageType2[i].pos_acc), ord(messageType2[i].raim)
            , ord(messageType2[i].regional), ord(messageType2[i].repeat),
            messageType2[i].rot ,
123                         ord(messageType2[i].slot_timeout),messageType2[i].sog,
            ord(messageType2[i].spare), messageType2[i].sub_message , ord(
            messageType2[i].sync_state),
124                         messageType2[i].true, messageType2[i].userid , ord(
            messageType2[i].utc_sec)))
125         for i in range(0, len(messageType3)):
126             (status,lat_dd,long_ddd) = aisparser.pos2ddd(messageType3[i].
            latitude , messageType3[i].longitude)
127             cur.execute("INSERT OR IGNORE INTO MessageType3(unixtime , cog ,
            keep,latitude ,longitude ,msgid ,nav_status ,num_slots ,pos_acc ,raim ,regional
            ,repeat ,rot ,slot_increment ,sog ,spare ,sync_state , true ,userid , utc_sec )
            VALUES(?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?)", (timeStamps3[i],
            messageType3[i].cog, ord(messageType3[i].keep) ,
```

```python
128                        lat_dd , long_ddd , ord(messageType3[i].msgid) , ord(
     messageType3[i].nav_status) , ord(messageType3[i].num_slots) , ord(
     messageType3[i].pos_acc) ,
129                        ord(messageType3[i].raim) , ord(messageType3[i].regional) ,
      ord(messageType3[i].repeat) , messageType3[i].rot , messageType3[i].
     slot_increment , messageType3[i].sog ,
130                        ord(messageType3[i].spare) , ord(messageType3[i].sync_state
     ) ,messageType3[i].true , messageType3[i].userid , ord(messageType3[i].
     utc_sec)))
131         for i in range(0, len(messageType4)):
132             (status ,lat_dd ,long_ddd) = aisparser.pos2ddd(messageType4[i].
     latitude , messageType4[i].longitude)
133             cur.execute("INSERT OR IGNORE INTO MessageType4(unixtime ,
     latitude ,longitude ,msgid ,pos_acc ,pos_type ,raim ,repeat ,slot_timeout ,spare
     ,sub_message ,sync_state ,userid ,utc_day ,utc_hour ,utc_minute ,utc_month ,
     utc_second ,utc_year) VALUES(?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?)", (
     timeStamps4[i], lat_dd , long_ddd , ord(messageType4[i].msgid) , ord(
     messageType4[i].pos_acc) ,
134                 ord(messageType4[i].pos_type) , ord(messageType4[i].raim) ,
      ord(messageType4[i].repeat) , ord(messageType4[i].slot_timeout) ,
     messageType4[i].spare , messageType4[i].sub_message ,
135                 ord(messageType4[i].sync_state) ,  messageType4[i].userid ,
     ord(messageType4[i].utc_day) , ord(messageType4[i].utc_hour) , ord(
     messageType4[i].utc_minute) , ord(messageType4[i].utc_month) ,
136                 ord(messageType4[i].utc_second) , messageType4[i].utc_year)
     )
137         for i in range(0, len(messageType5)):
138             messageType5[i].dest = messageType5[i].dest.replace(" ","")
139             messageType5[i].dest = messageType5[i].dest.replace("@","")
140             messageType5[i].callsign = messageType5[i].callsign.replace(" "
     ,"")
141             messageType5[i].callsign = messageType5[i].callsign.replace("@"
     ,"")
142             messageType5[i].name = messageType5[i].name.replace(" ","")
143             messageType5[i].name = messageType5[i].name.replace("@","")
144             cur.execute("INSERT OR IGNORE INTO MessageType5(unixtime ,
     callsign , dest , dim_bow , dim_port , dim_starboard , dim_stern , draught ,
     dte , eta , imo , msgid , name , pos_type , repeat , ship_type , spare  , userid ,
```

```
       version ) VALUES(?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?)", (timeStamps5[i
       ], messageType5[i].callsign , messageType5[i].dest, messageType5[i].
       dim_bow, ord(messageType5[i].dim_port) ,
145                ord(messageType5[i].dim_starboard) , messageType5[i].
       dim_stern, messageType5[i].draught, ord(messageType5[i].dte) ,
       messageType5[i].eta, messageType5[i].imo, ord(messageType5[i].msgid) ,
       messageType5[i].name  ,
146                ord(messageType5[i].pos_type) , ord(messageType5[i].repeat)
       , messageType5[i].ship_type , ord(messageType5[i].spare) , messageType5[
       i].userid , ord(messageType5[i].version)))
147        cur.execute('COMMIT')
148
149 if __name__ == "__main__":
150      databasepath = ""
151      createDatabase(databasepath)
152      for foldername in os.listdir('../../../../../media/bjornasm/My Passport
       /AIS Rawdata/Kystverket/'):
153          for filename in os.listdir('../../../../../media/bjornasm/My
       Passport/AIS Rawdata/Kystverket/'+foldername):
154              extractMessages('../../../../../media/bjornasm/My Passport/AIS
       Rawdata/Kystverket/'+foldername+'/'+filename)
155              writeToDatabase(databasepath)
```

Appendix D

plotOneMonthGlobal.py

```python
#!/usr/bin/python
#    Project: Master Thesis
#    Author: Bjornar Brende Smestad
#    Created: Spring 2015

import sqlite3 as lite
import sys
import os
import numpy as np
import datetime
import matplotlib.pyplot as plt
from mpl_toolkits.basemap import Basemap

if __name__ == "__main__":
    map = Basemap(projection='cyl',
    resolution = 'c', area_thresh = 40,
    llcrnrlon=-180, llcrnrlat=-90,
    urcrnrlon=180, urcrnrlat=90)

    with lite.connect('database/s/SAISREAL.db') as con:
        start = 1406851200
        end = 1409529600
        middle = start + (end-start)/2
        cur = con.cursor()
        cur.execute('SELECT latitude, longitude FROM dynamicMessage WHERE
        unixtime >= {start} AND unixtime < {middle}'.format(start = start,
```

```
         middle = middle))
26       data = cur.fetchall()
27       y,x = zip(*data)
28       x,y = map(x,y)
29       plt.scatter(x,y, s=0.02, alpha=0.4, color="#e74c3c", edgecolors='none')
30       cur.execute('SELECT latitude, longitude FROM dynamicMessage WHERE
         unixtime >= {middle} AND unixtime < {end}'.format(middle = middle, end =
          end))
31       data = cur.fetchall()
32       y,x = zip(*data)
33       x,y = map(x,y)
34       plt.scatter(x,y, s=0.02, alpha=0.4, color="#e74c3c", edgecolors='none')
35       plt.savefig('AugustGlobalPlot.png', bbox_inches='tight')
36       plt.clf()
```

Appendix E

findMaxSpeed.py

```python
#!/usr/bin/python
#    Project:  Master  Thesis
#    Author:  Bjornar  Brende  Smestad
#    Created:  Spring  2015

def f(speeds):
    rev = [speed for speed in speeds if speed < 30]
    rev.sort(reverse=True)
    c = Counter((int(v) for v in rev))
    for speed in rev:
        if speed > 15 and c[int(speed)] > 9:
            return speed
        elif 5 <= speed <= 15:
            return speed
    return False

def getMaxSpeed(databasepath, mmsi):
    from collections import Counter
    con = lite.connect(databasepath)
    speeds = []
    with con:
        cur = con.execute("SELECT distinct sog FROM dynamicMessage where
    userid = {userid} and sog != '102.30' ".format(userid = mmsi))
    speed = [x[0] for x in cur]
    for i in range(0,len(speed)):
        speeds.append(float(speed[i]))
```

```
26        return f(speeds)
```

Appendix F

erroneousIMOnumber.py

```python
#!/ usr / bin / python
#    Project : Master  Thesis
#    Author : Bjornar  Brende  Smestad
#    Created : Spring  2015

import sqlite3 as lite

def getImoNumbers ( databasepath ,  table ) :
    con = lite . connect ( databasepath )

    with con :
        cur = con . execute ( "SELECT distinct imo FROM {table}" . format ( table =
        table ) )
        con . commit ( )
        imonumbers = [ int ( x [ 0 ] ) for x in cur ]
    con . close ( )
    return imonumbers

def checkImoNumber ( number ) :
  number = str ( number )
  checksum = 0
  if len ( number ) != 7 :
    return False
  if len ( number ) == 0 :
    return True
  for i in range ( 0 , 6 ) :
```

```python
26         checksum += int(number[i])*(7-i)
27     if str(checksum)[-1] != number[-1]:
28         return False
29     return True
30
31 if __name__ == "__main__":
32     numtrues = 0
33     numfalse = 0
34     imonumbers = getImoNumbers('database/s/SAISREAL.db', 'staticMessage')
35     for num in imonumbers:
36         if checkImoNumber(num):
37             numtrues += 1
38         else:
39             numfalse += 1
40             print num
41     print "Total IMOs"
42     print numtrues+numfalse
43     print "Number of true IMOs"
44     print numtrues
45     print "Number of false IMOs"
46     print numfalse
```

Appendix G

Scatterplots in Full Size
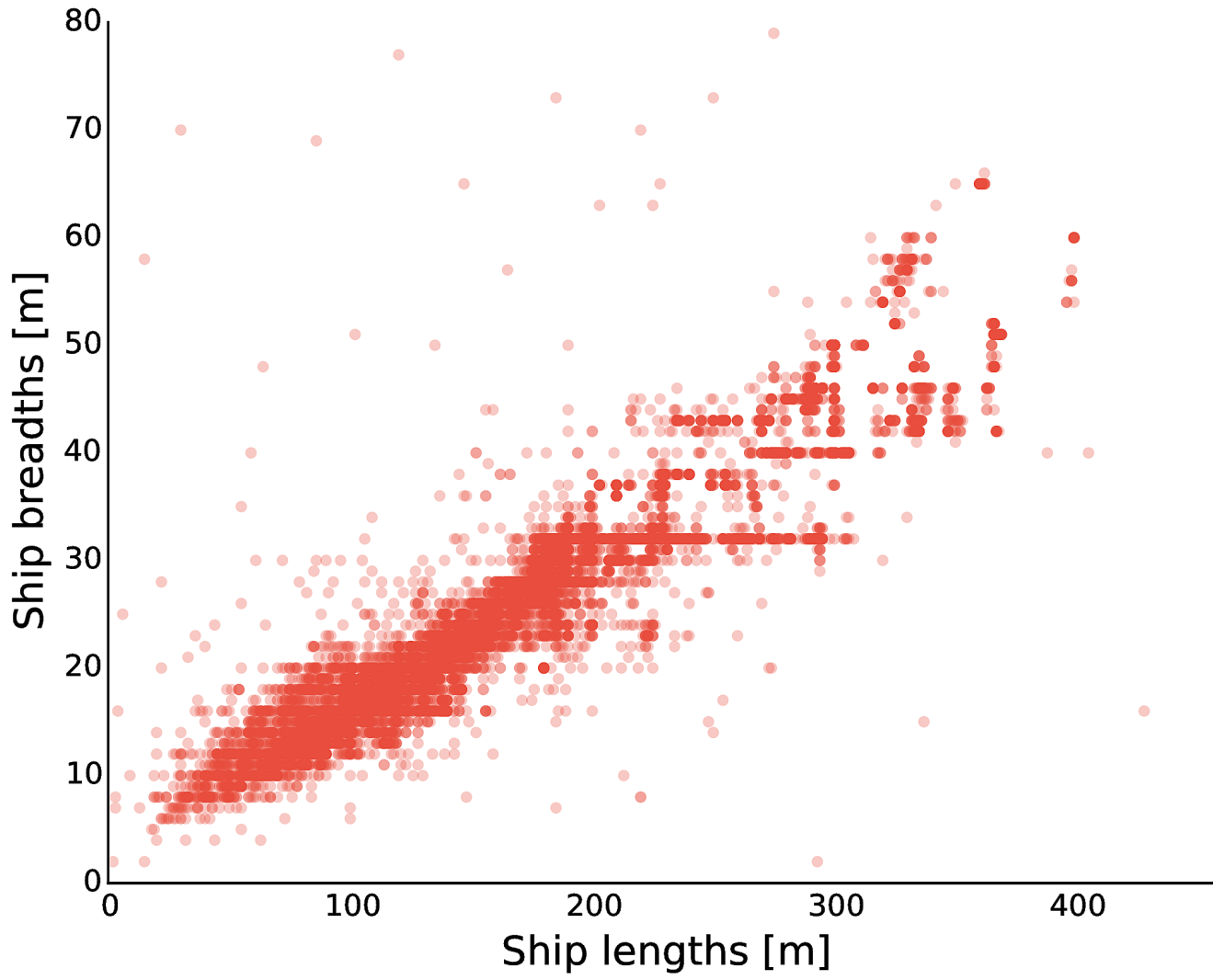
Table G.1: Scatter plot over length and breadth for all cargo ships.
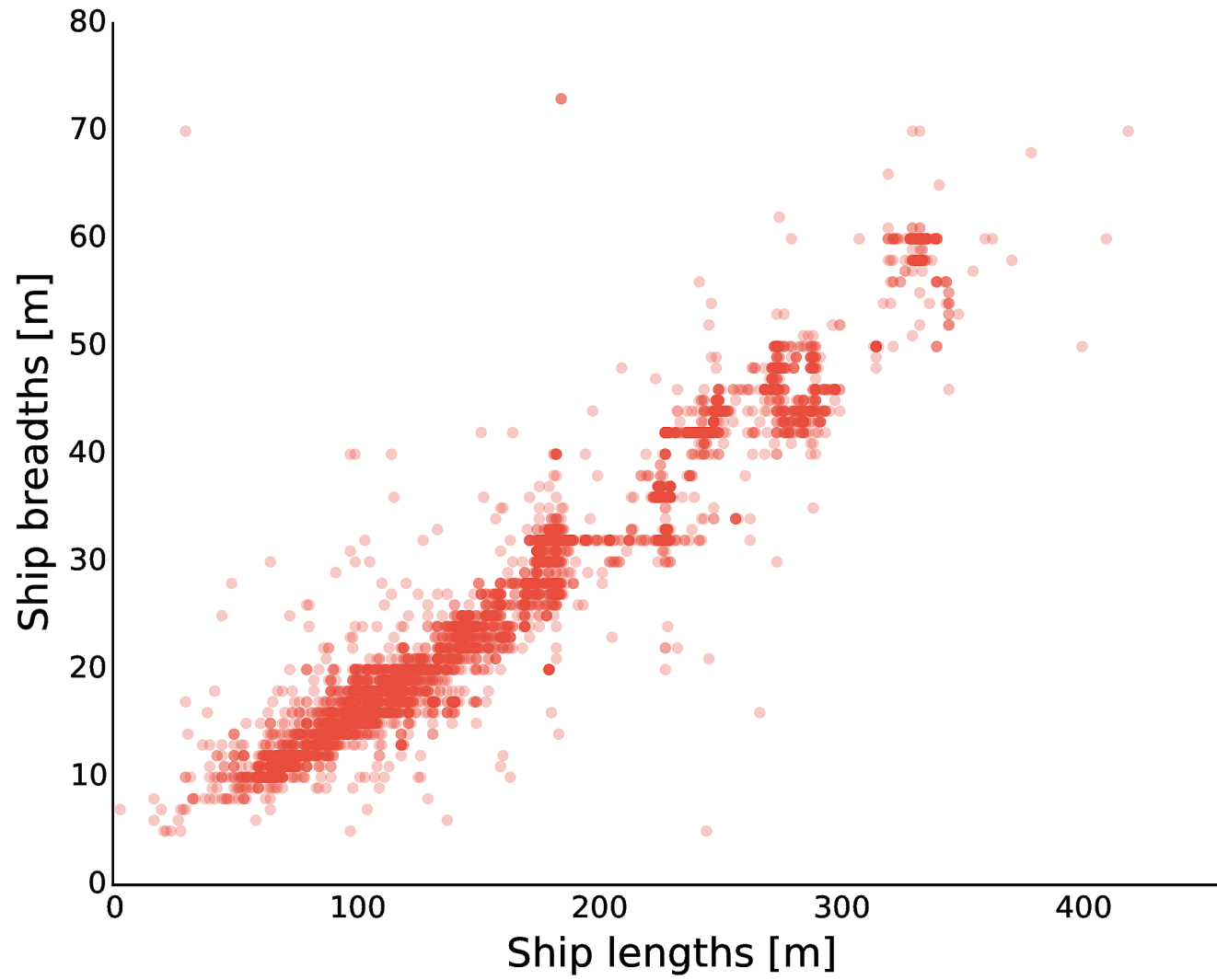
Table G.2: Scatter plot over length and breadth for all tankers.
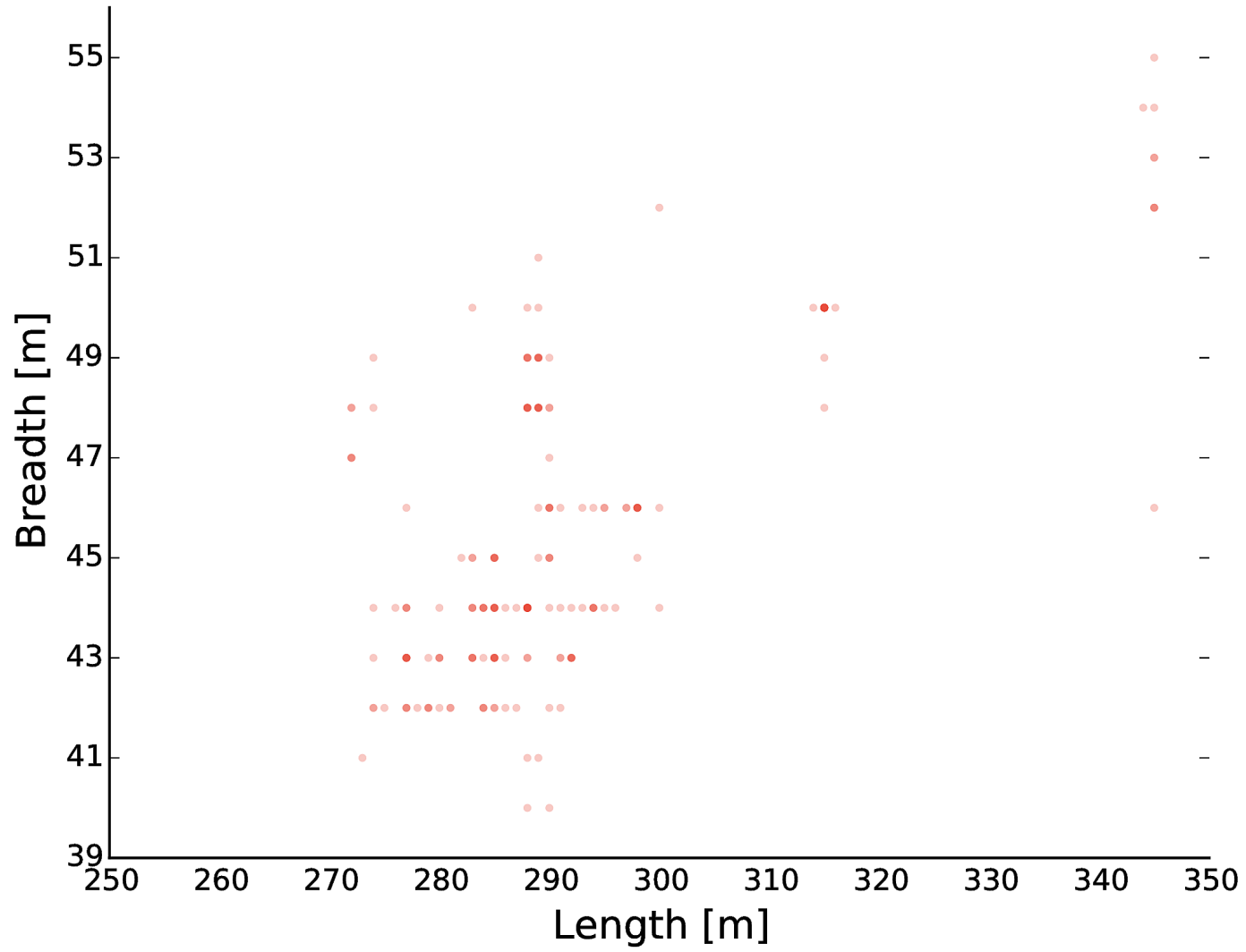
Table G.3: Length and breadth of LNG tankers.
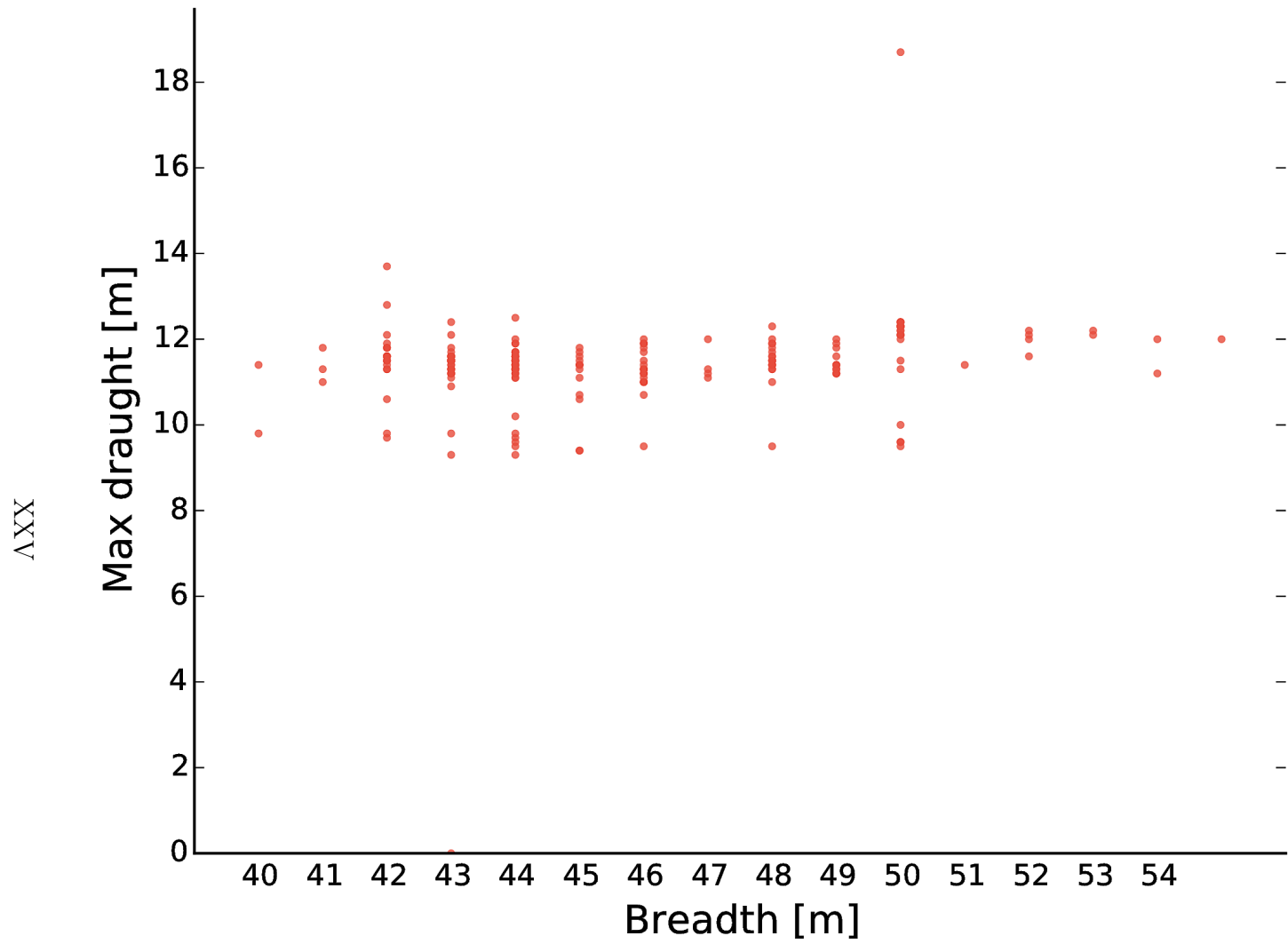
Table G.4: Max draught of LNG tankers inside different breadth categories.
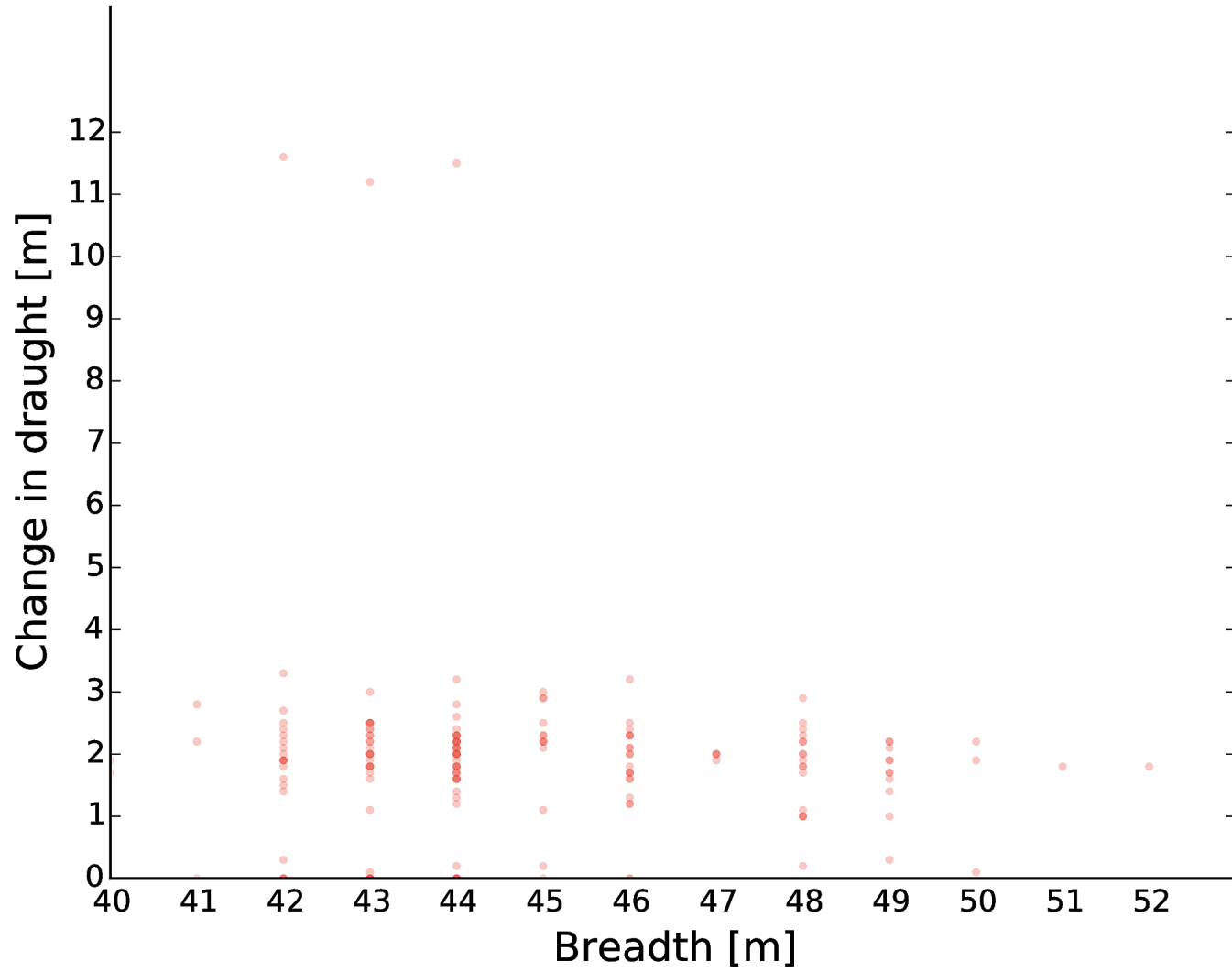
Table G.5: Max change in draught for LNG tankers excluding Q-flex and Q-max vessels.
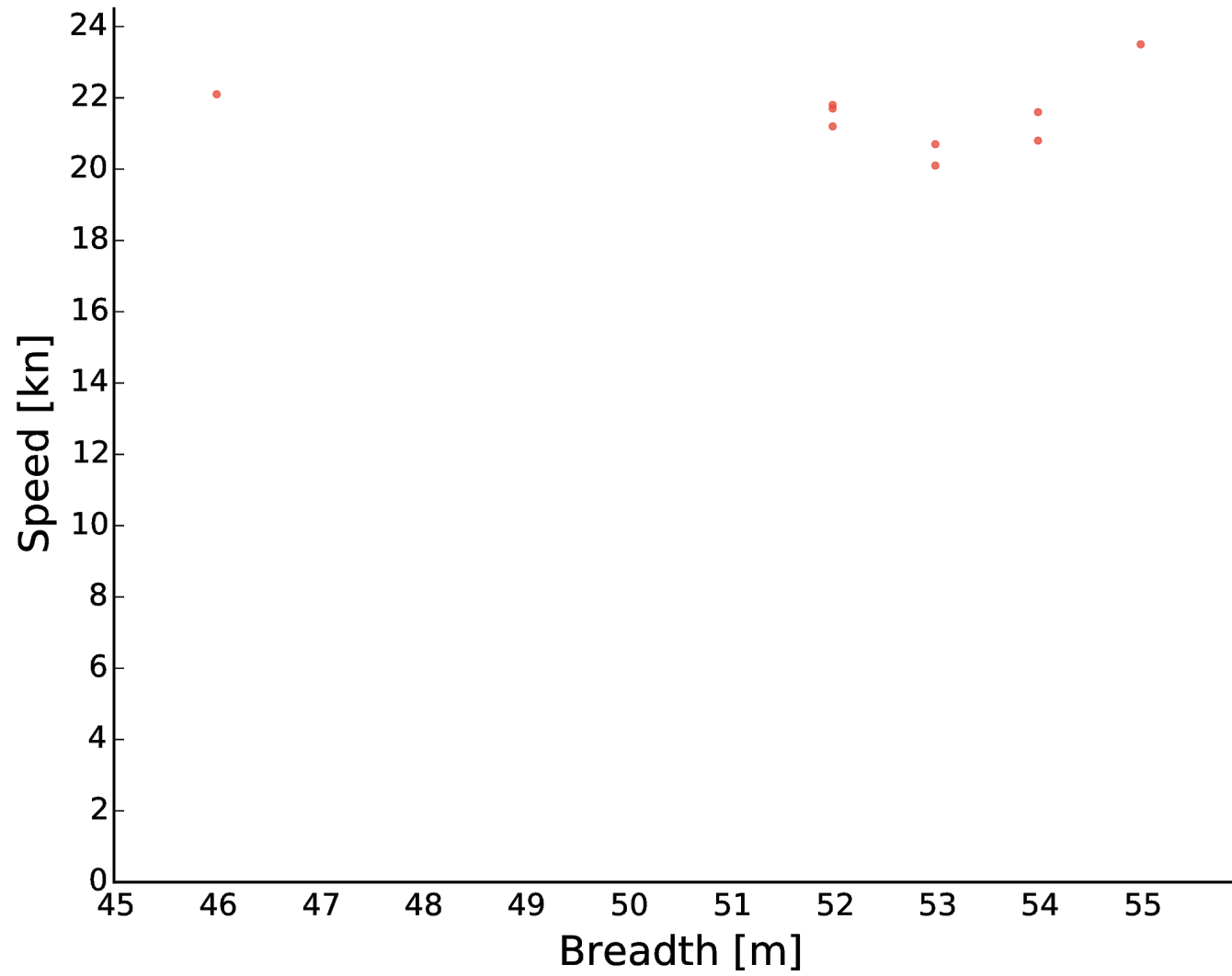
Table G.6: Max speed and breadth plot, Q-Max vessels.
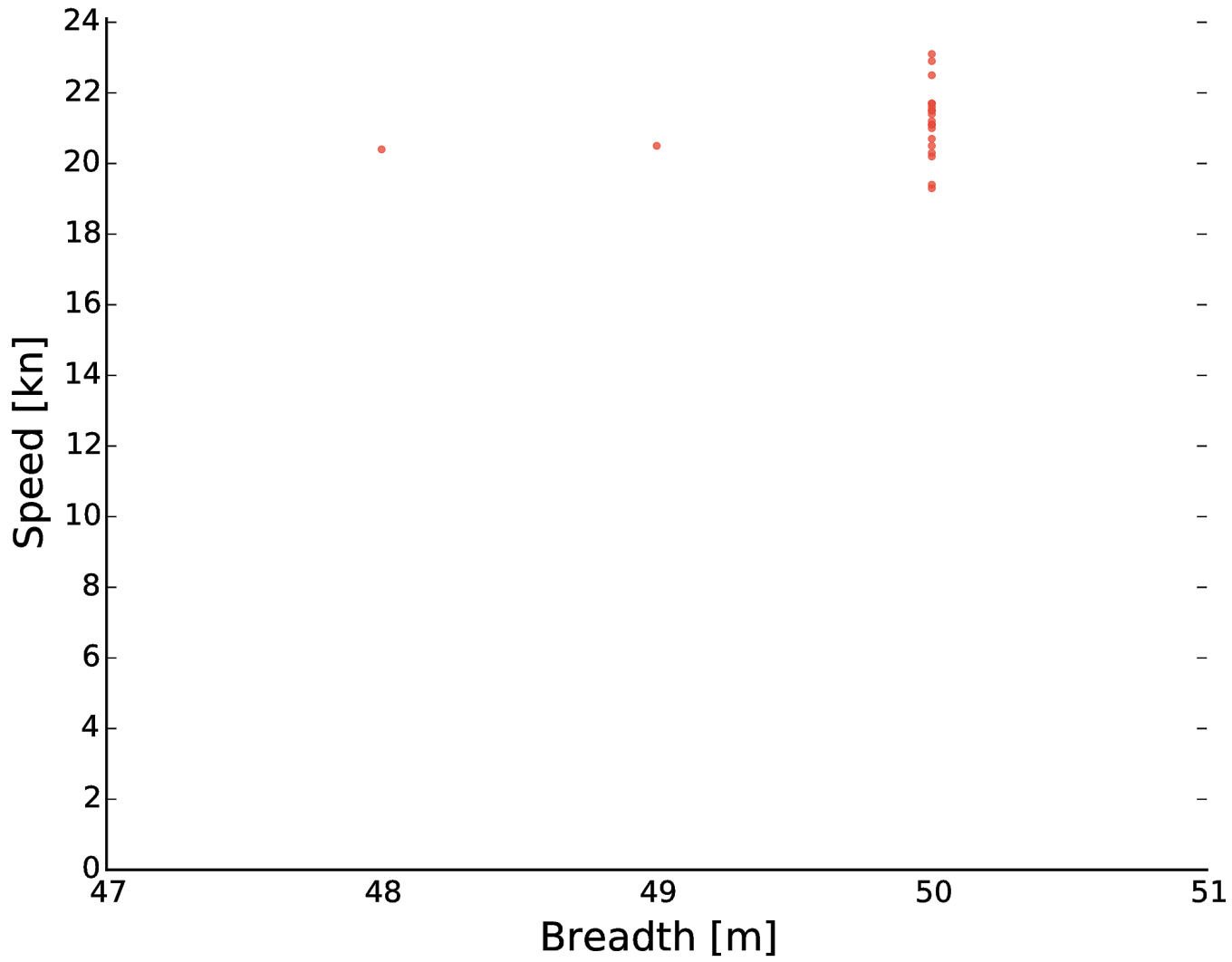
Table G.7: Max speed and breadth plot, Q-Flex vessels.
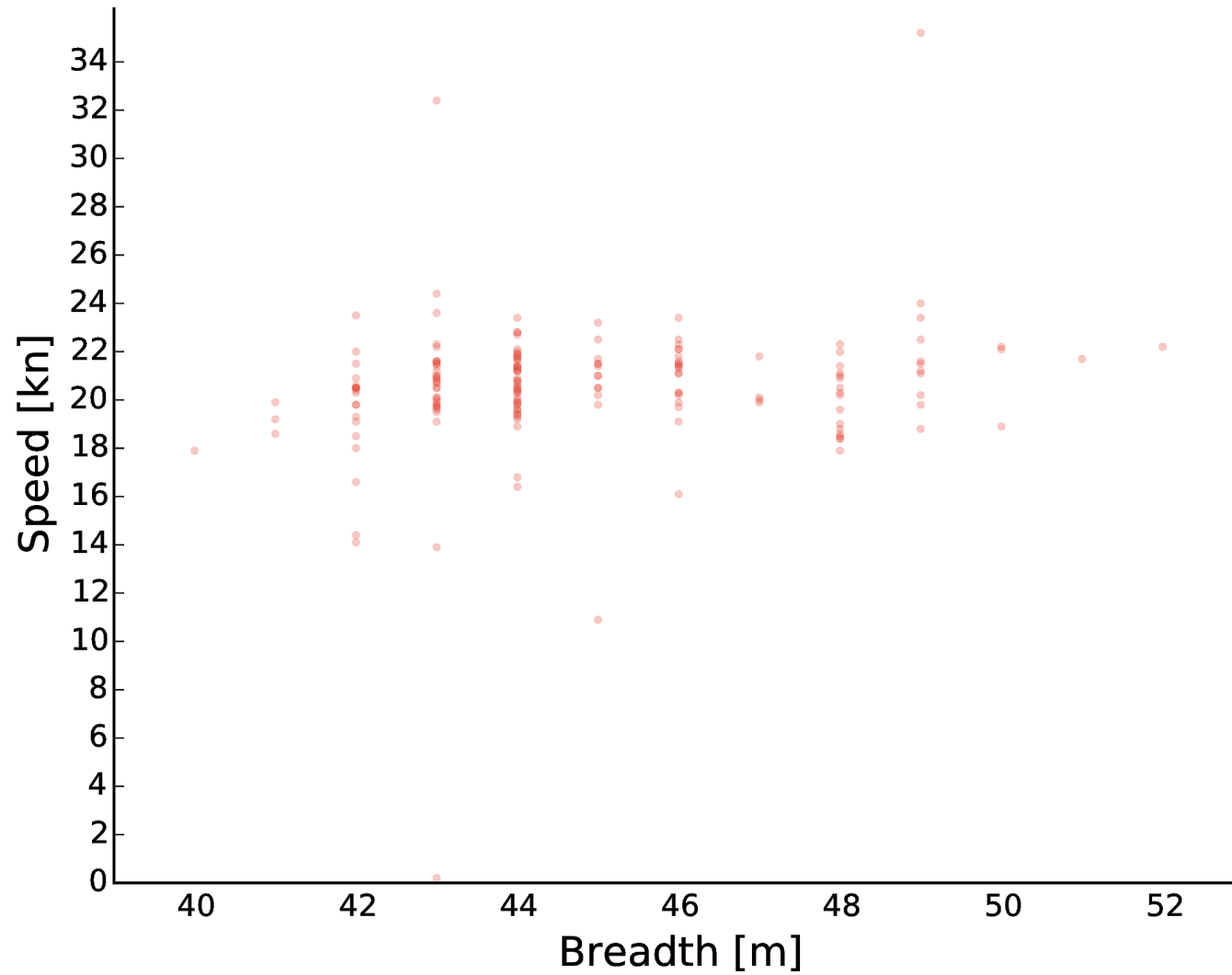
Table G.8: Max speed and breadth plot, LNG vessels in General Group.
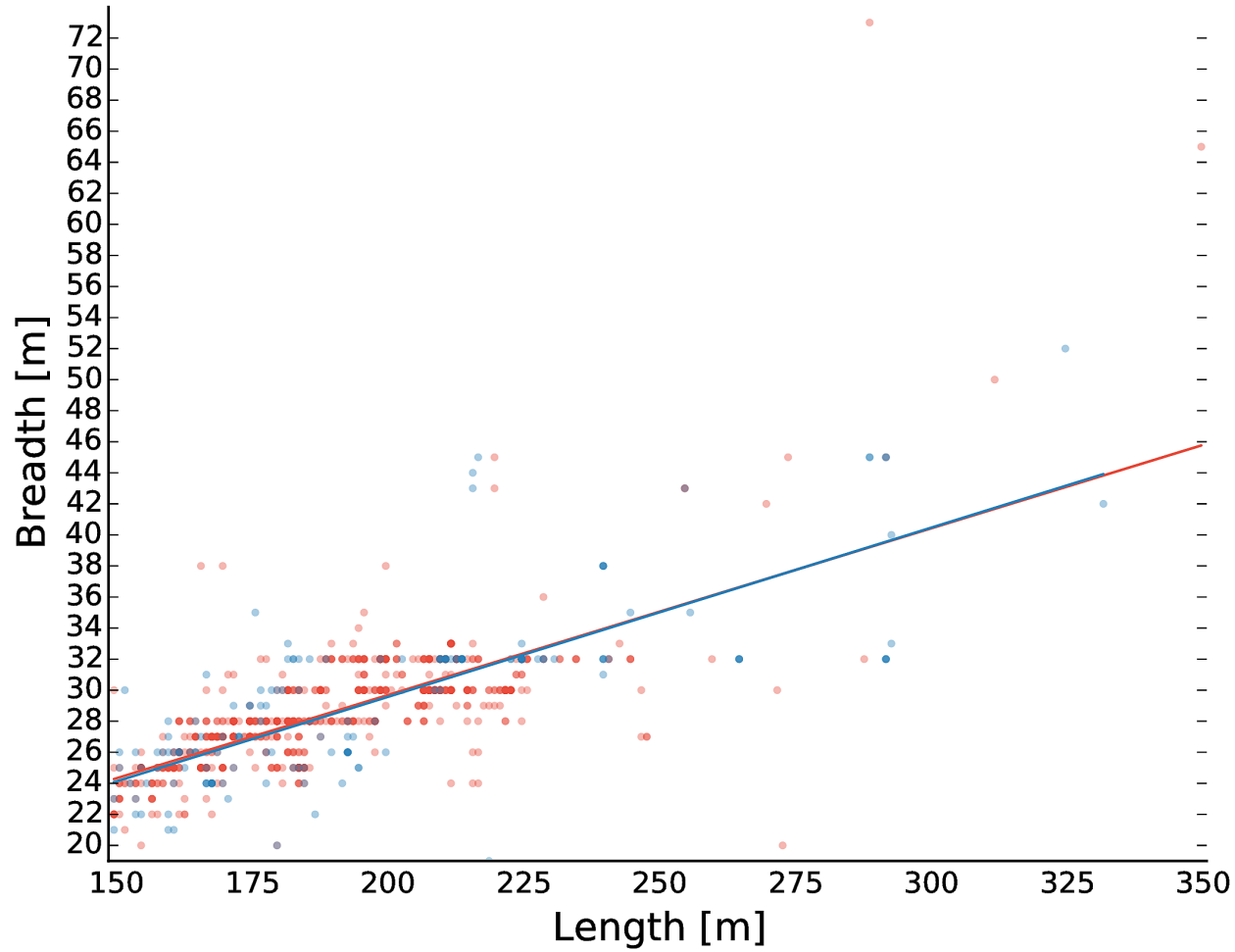
Table G.9: Length and breadth for Ro Ro vessels (blue) and Sub Panamax ships (red).
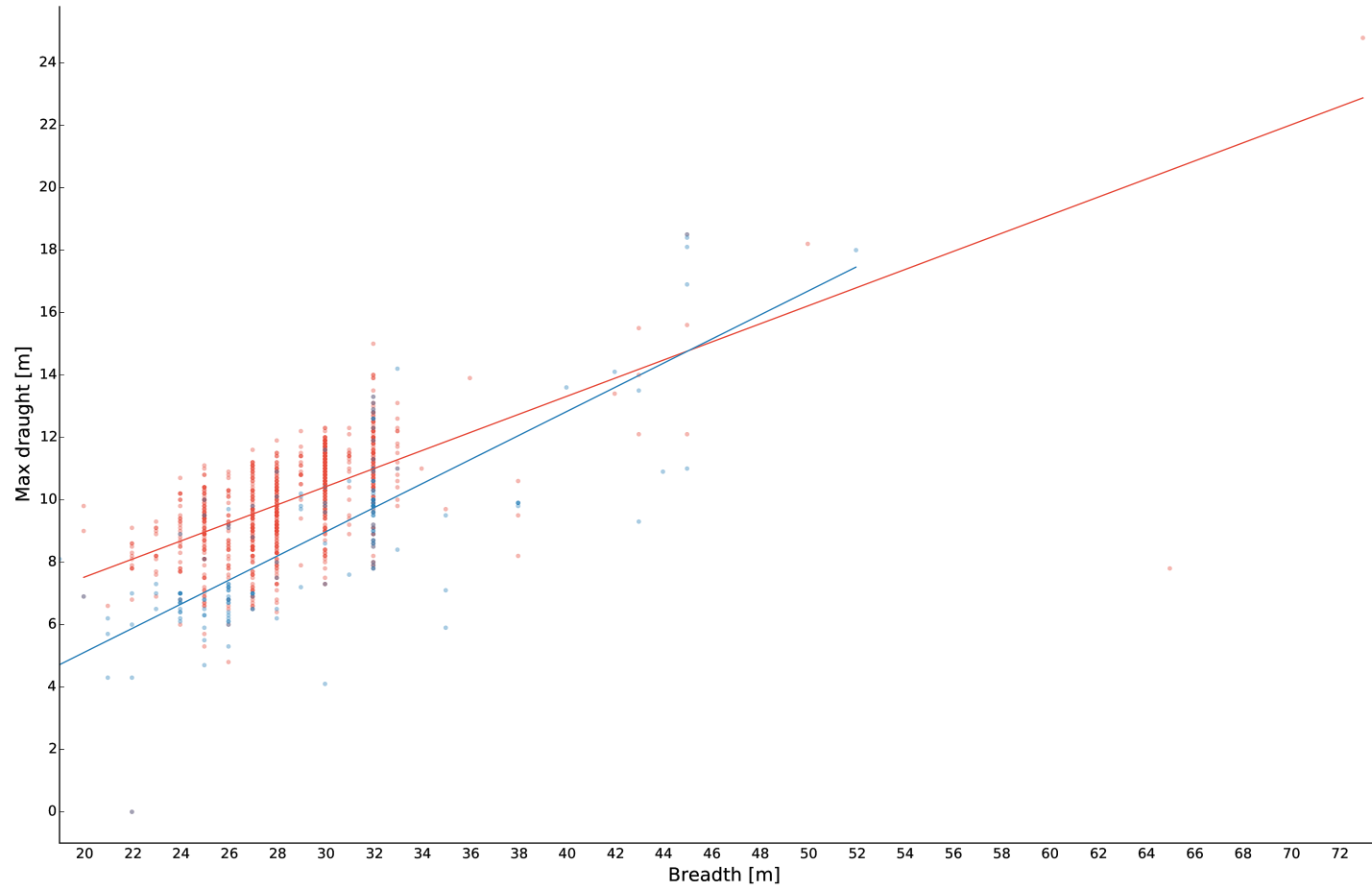
Table G.10: Maximum depth and breadth for RoRo vessels (blue) and Sub Panamax ships (red).
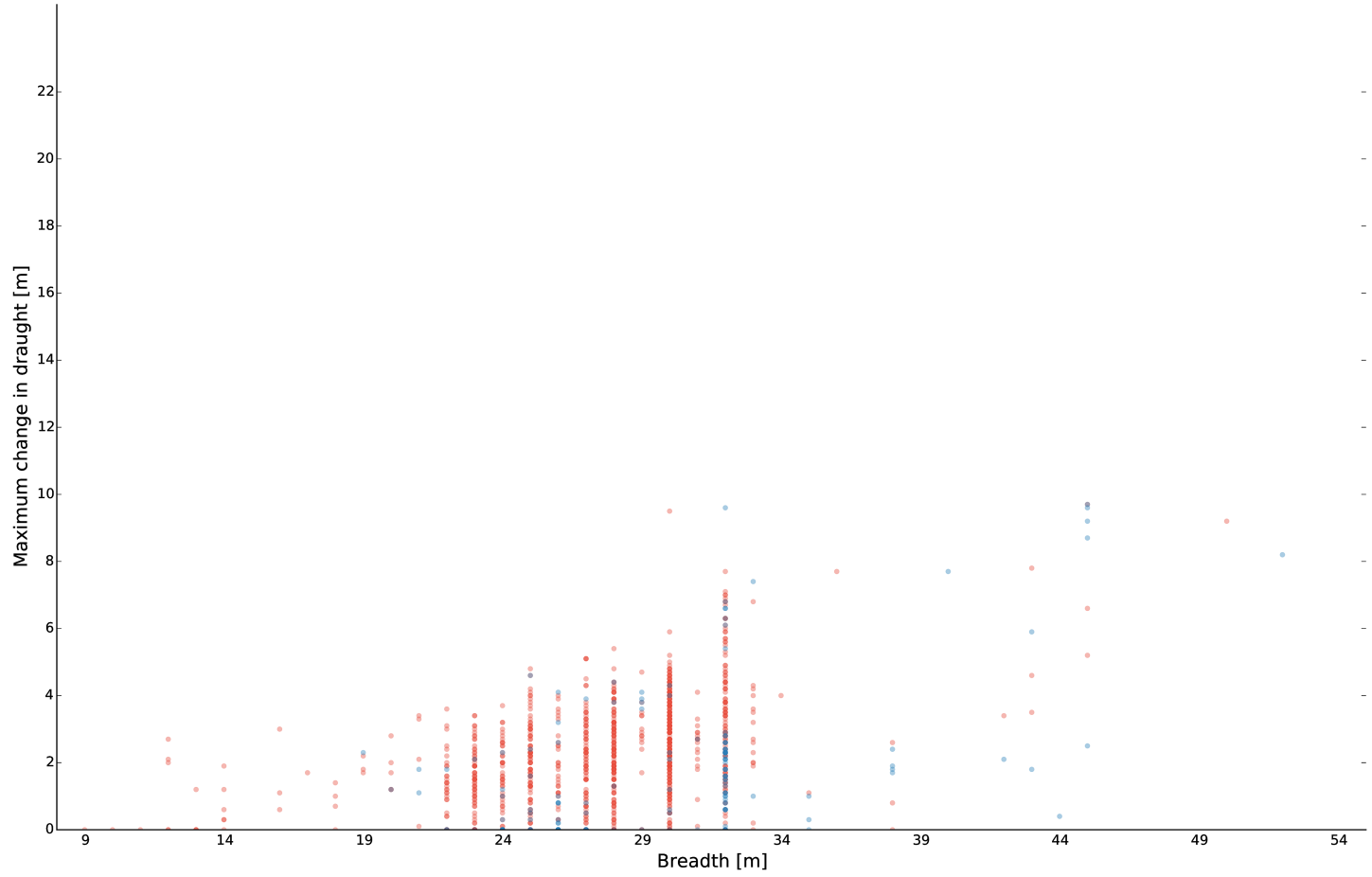
Table G.11: Change in draught and breadth. Roro vessels in blue, Sub Panamax in red
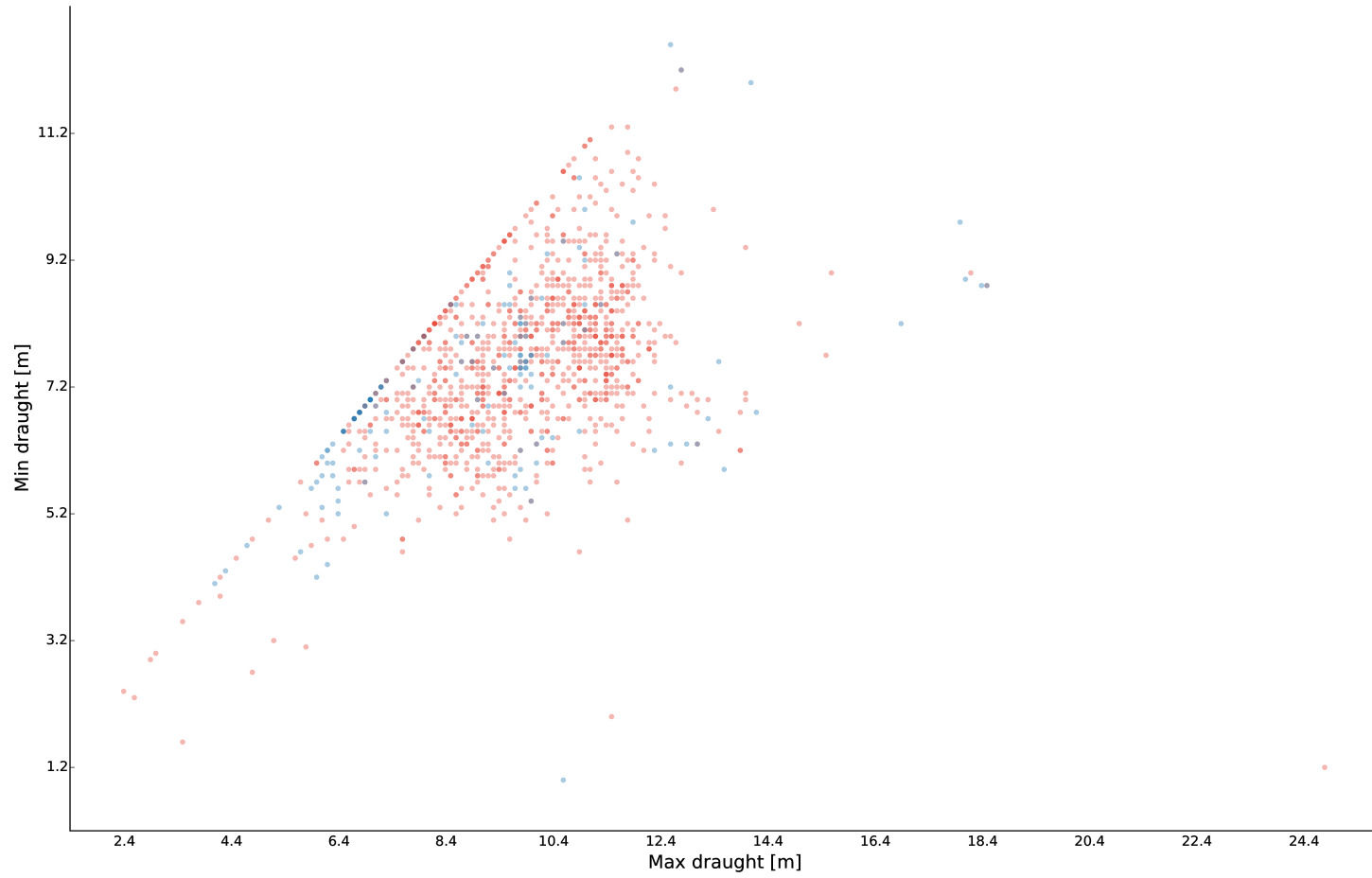
Table G.12: Maximum and minimum draught. RoRo vessels in blue, Sub Panamax in red.
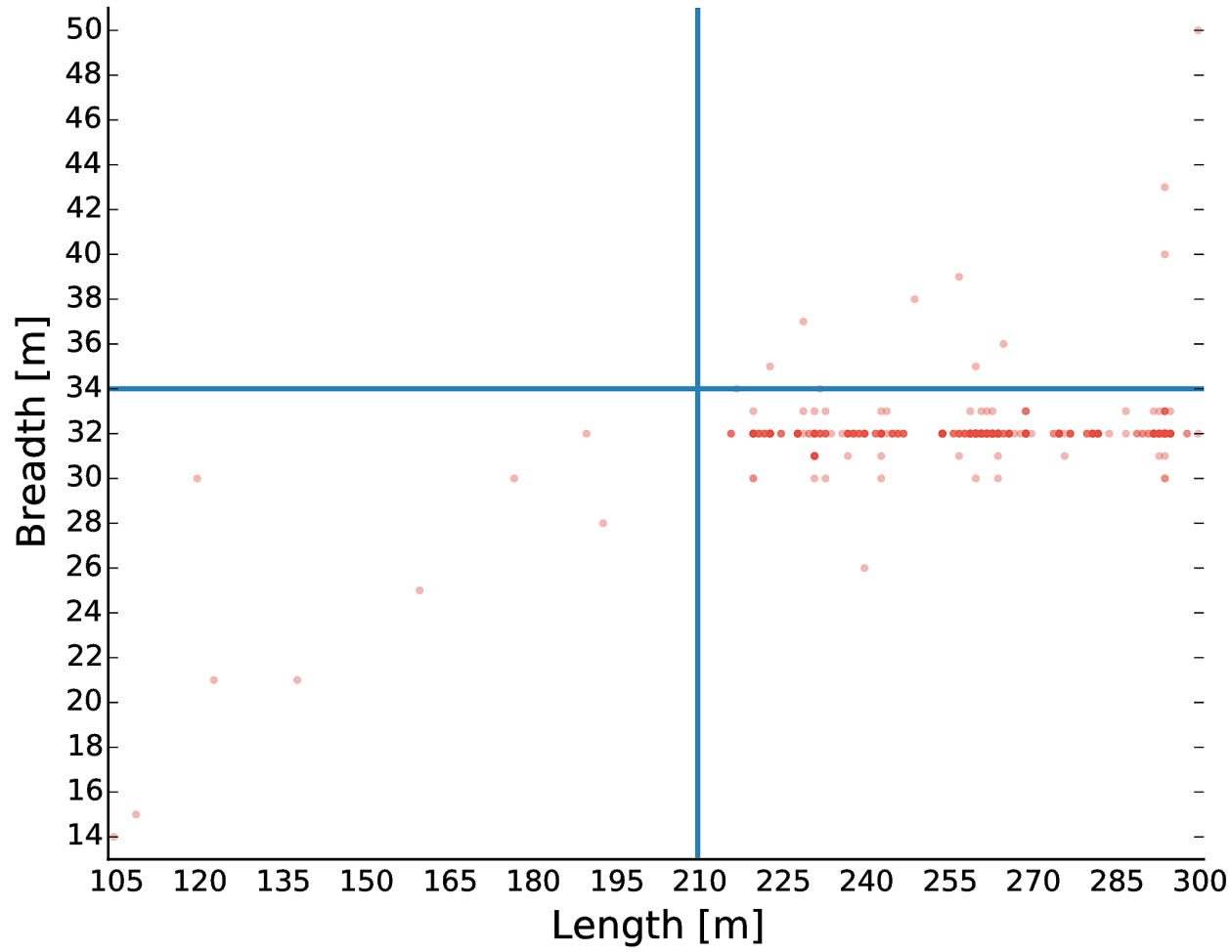
Table G.13: Breadth and length for container vessels in the Panamax category.
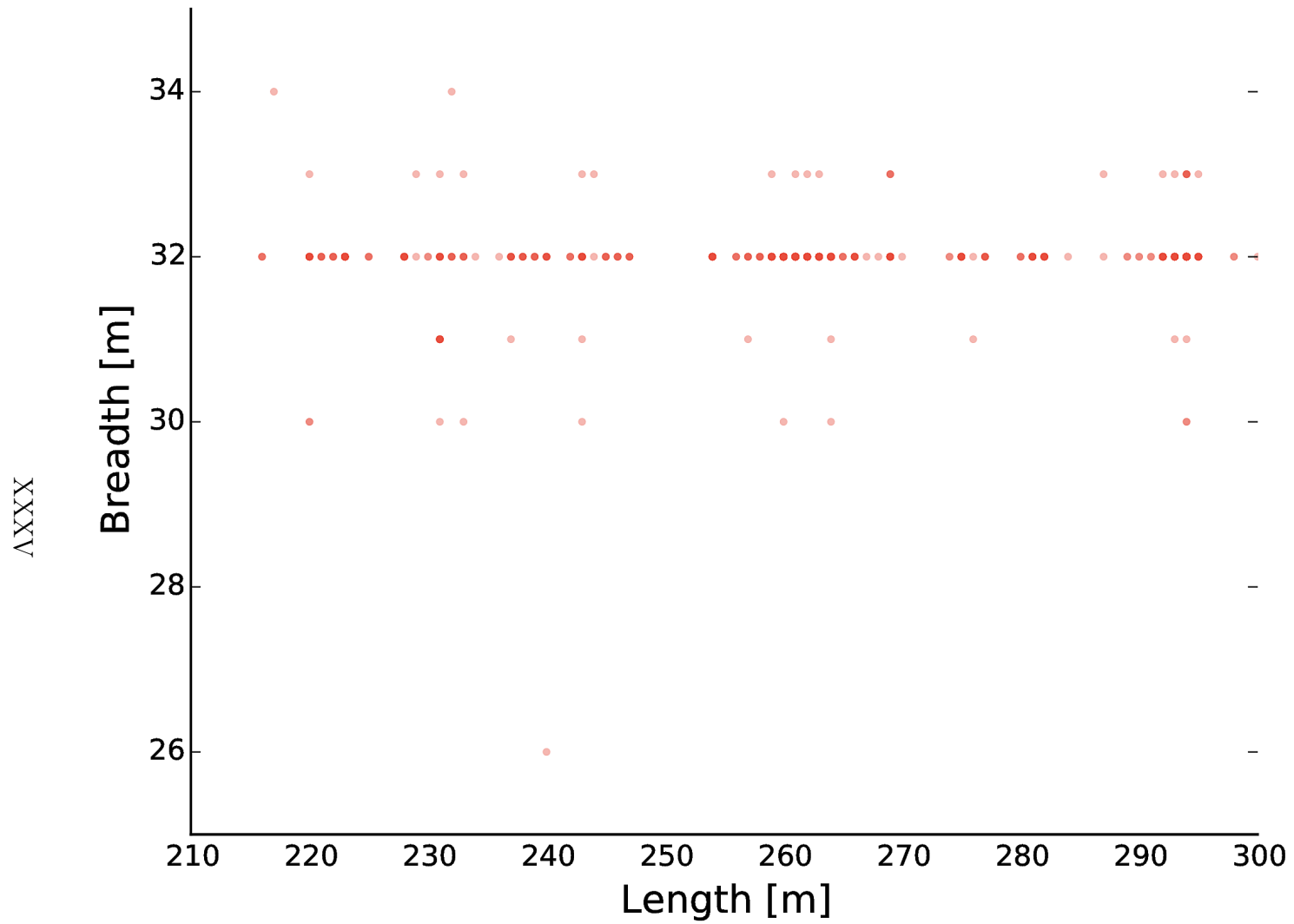
Table G.14: Close up of lower right quadrant of left figure.
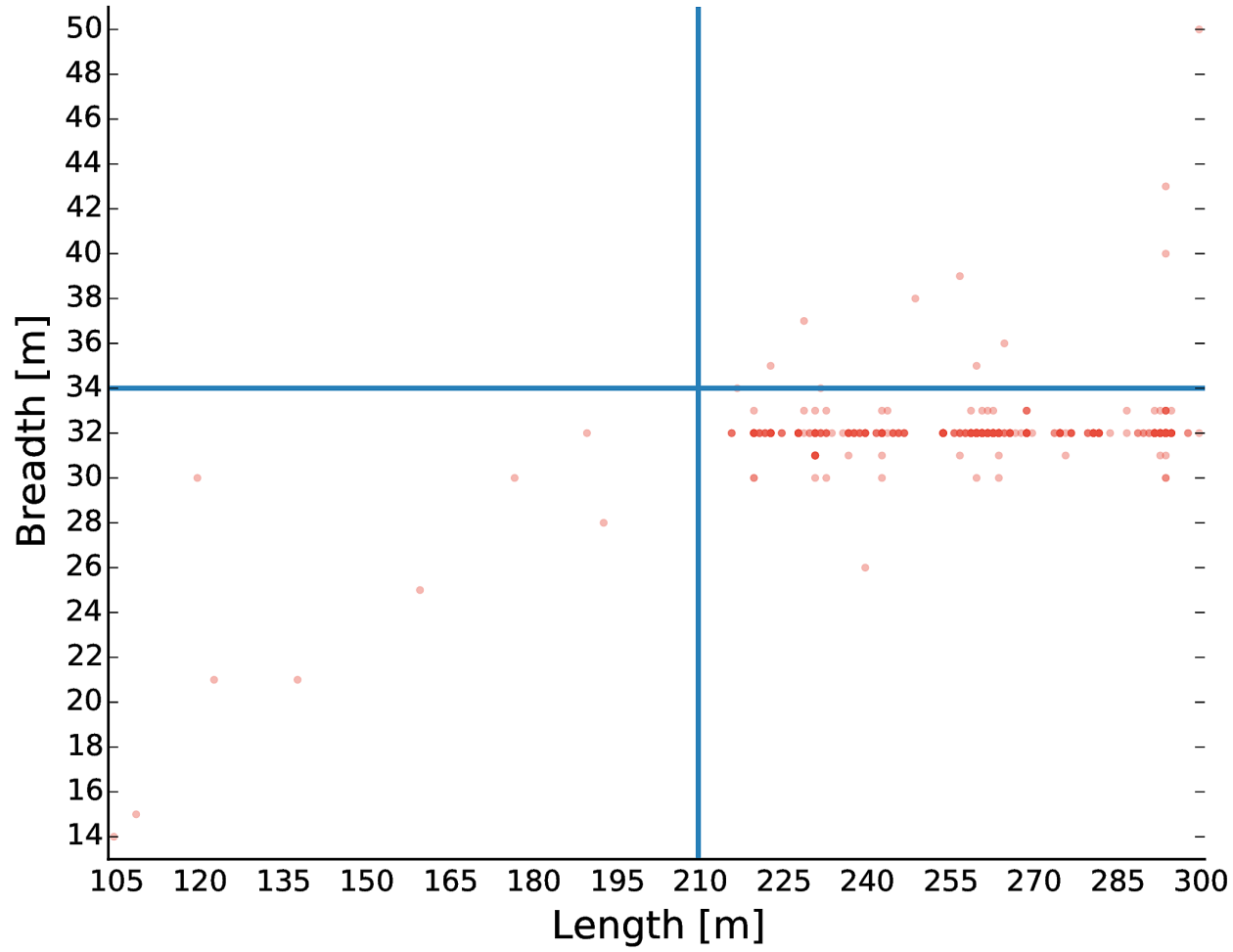
Table G.15: Breadth and length for container vessels in the Panamax category.
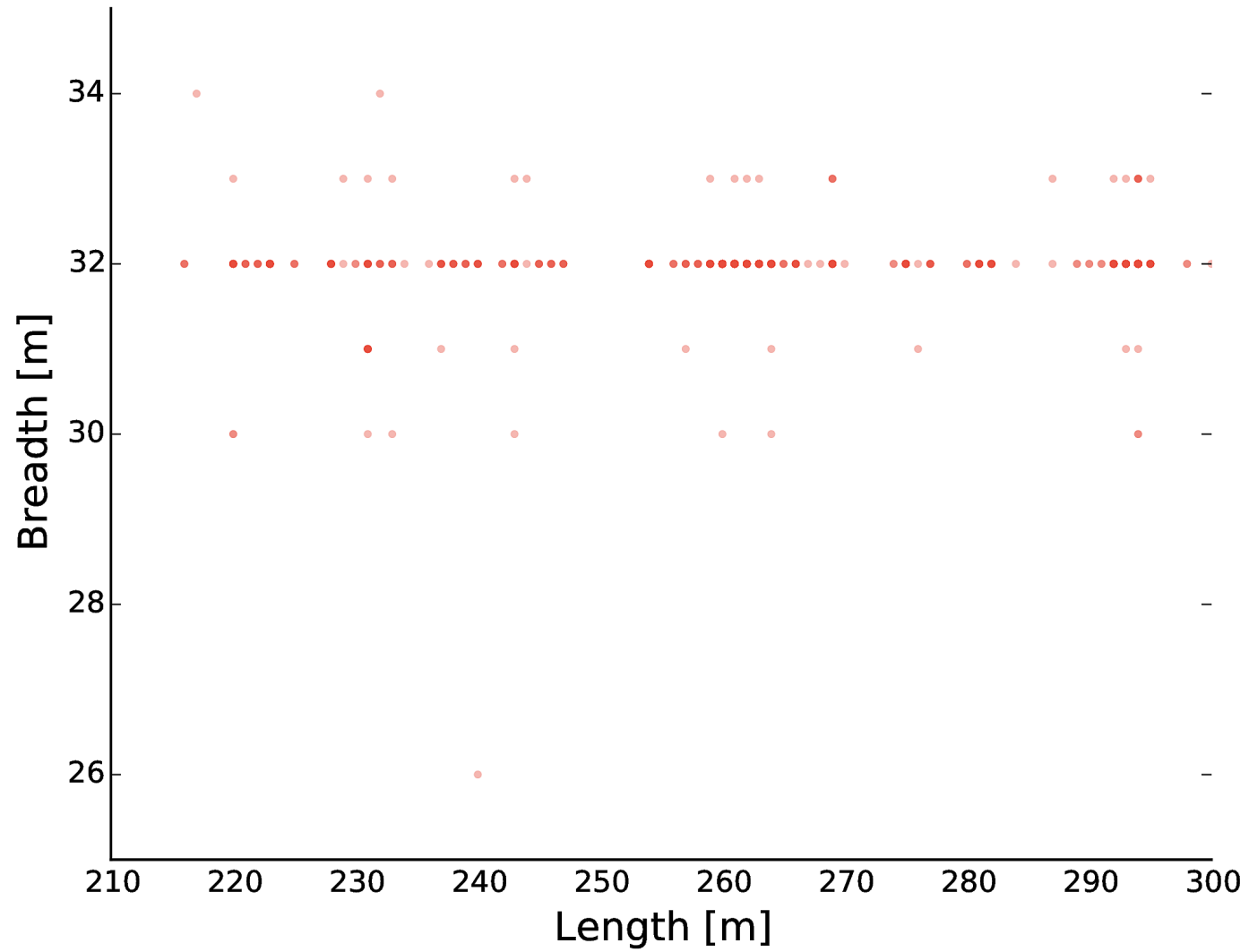
Table G.16: Close up of lower right quadrant of left figure.
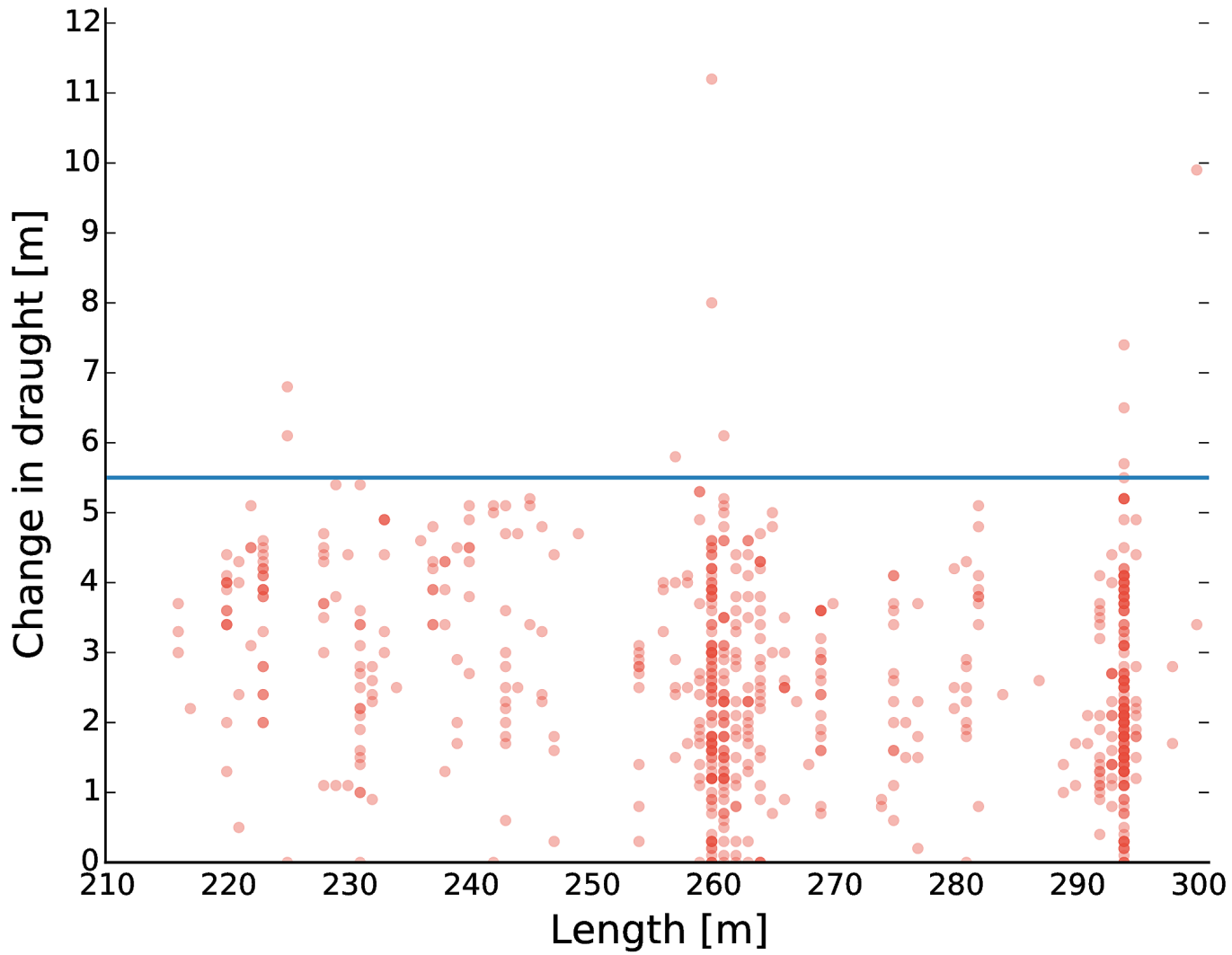
Table G.17: Change in draught for Panamax Container Vessels. 97.4% of the vessels had a change of draught less than 5.5 m, indicated by the blue line.

Table G.18: Maximum recorded speed for Panamax Container Vessels
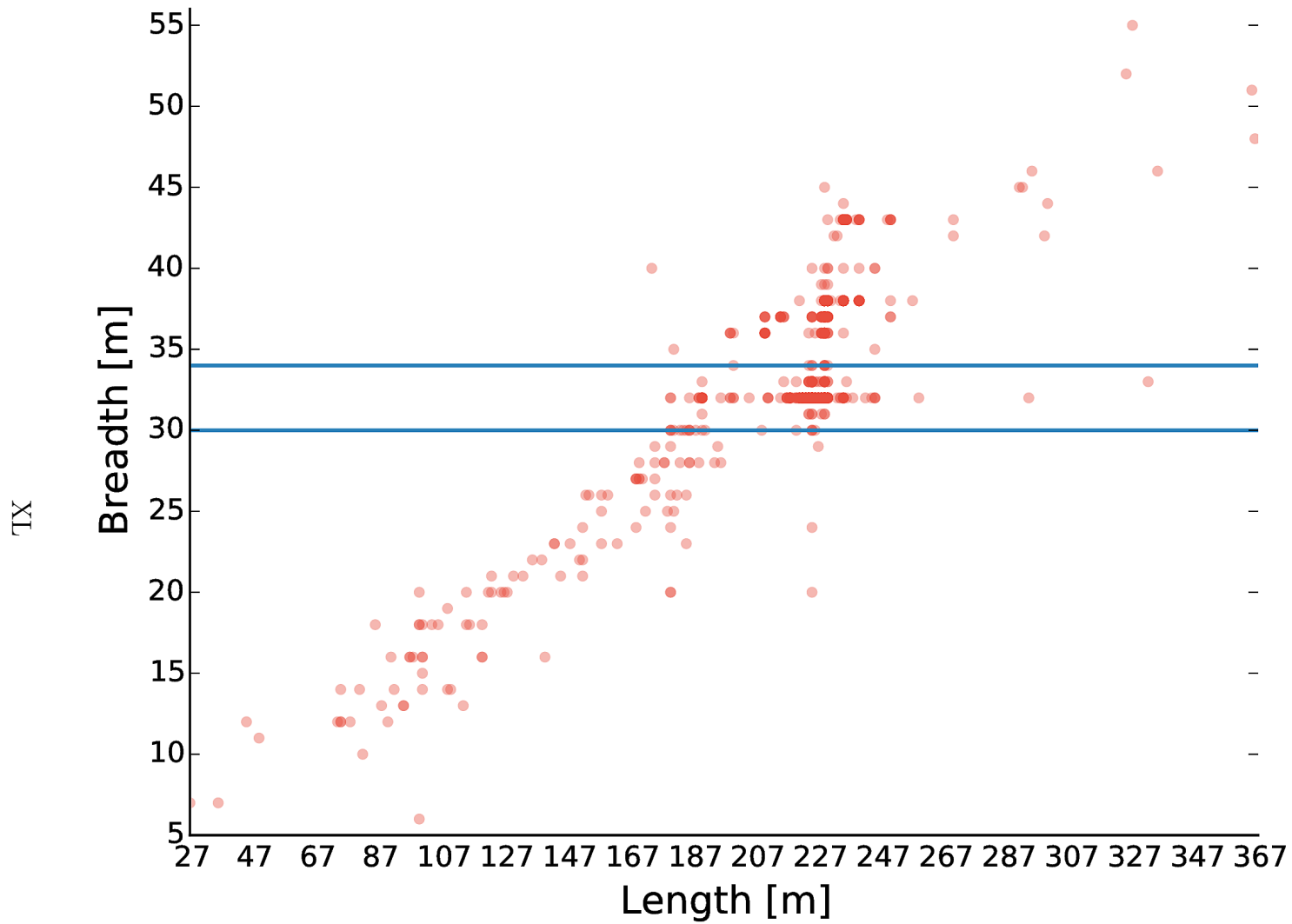
Table G.19: Length and breadth for the ideal group of Panamax sized Bulk Carriers. The horizontal lines indicates the maximum and minimum allowed breadth.
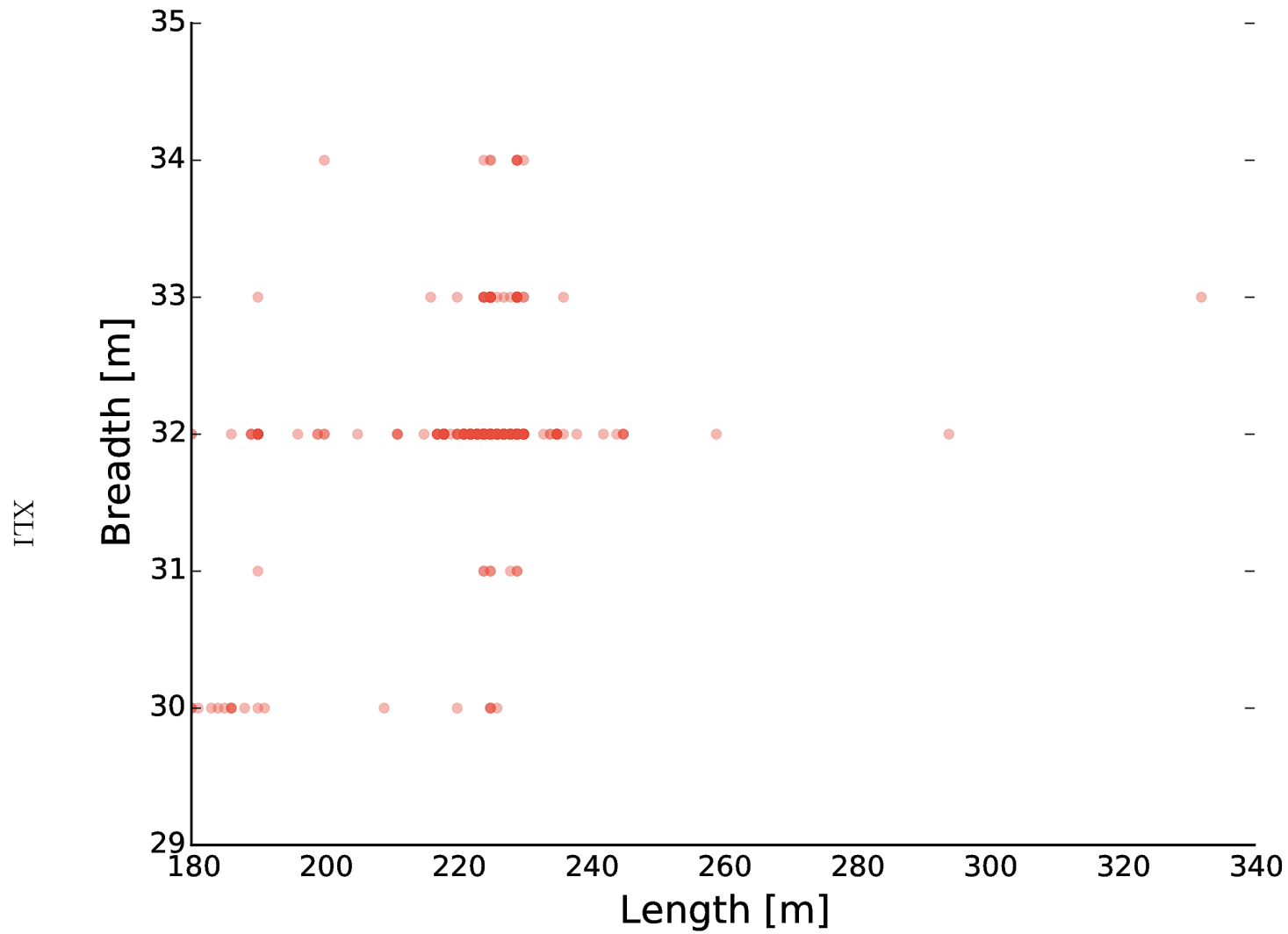
Table G.20: Vessels inside the allowed breadth category.
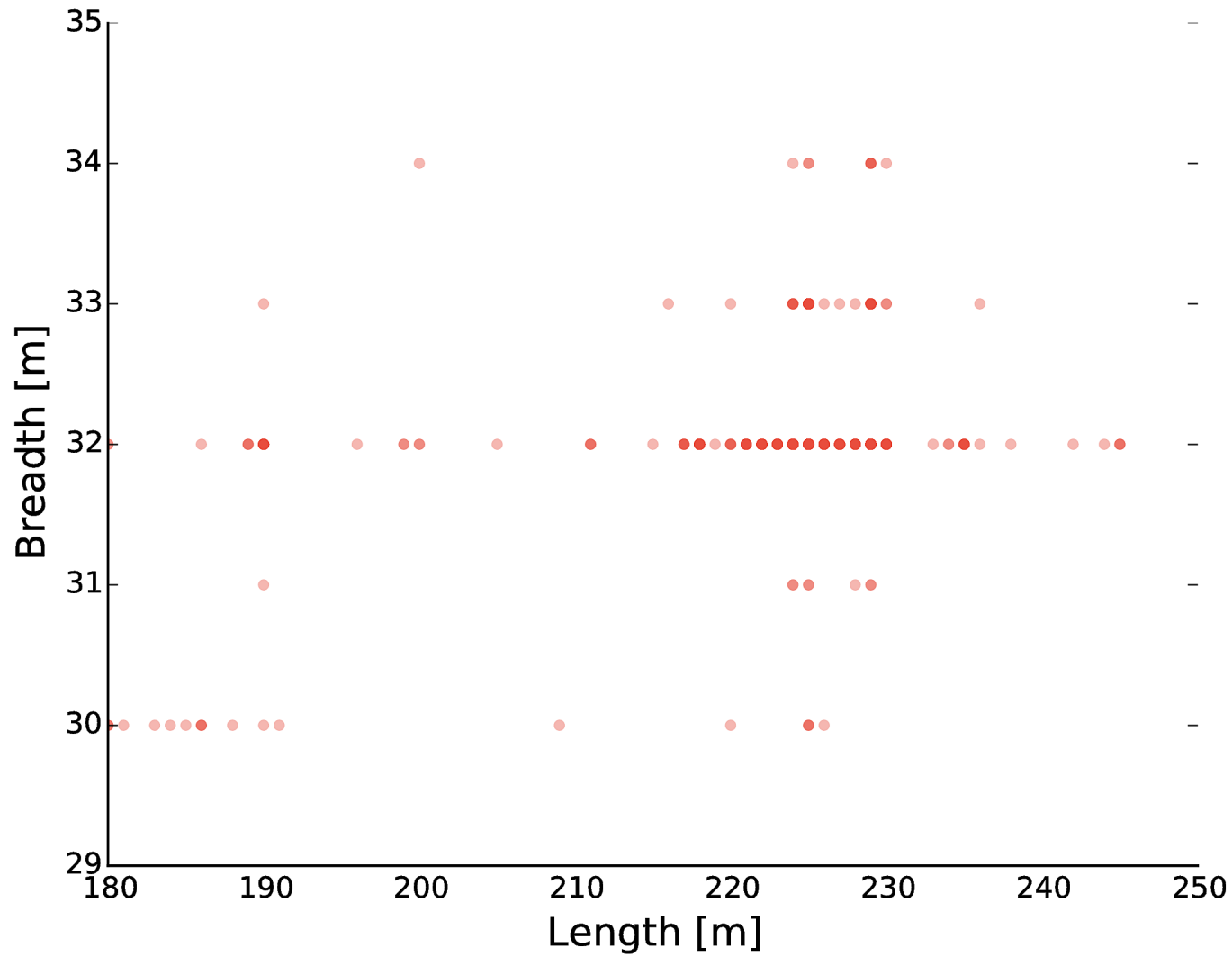
Table G.21: Length and breadth for the final selection of Bulk Carriers, where the breadth is between 30 and 34 m, and the length is under 250m.
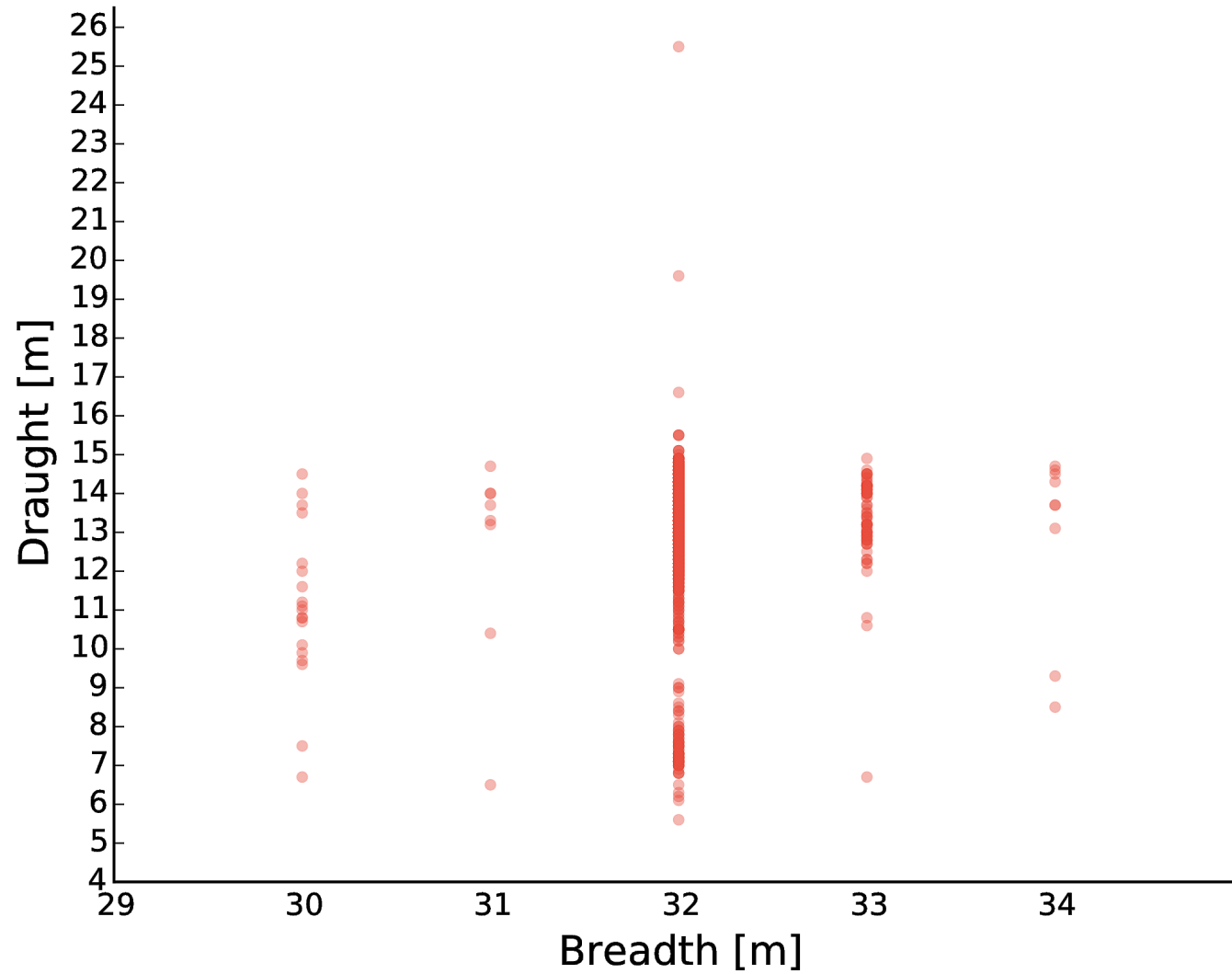
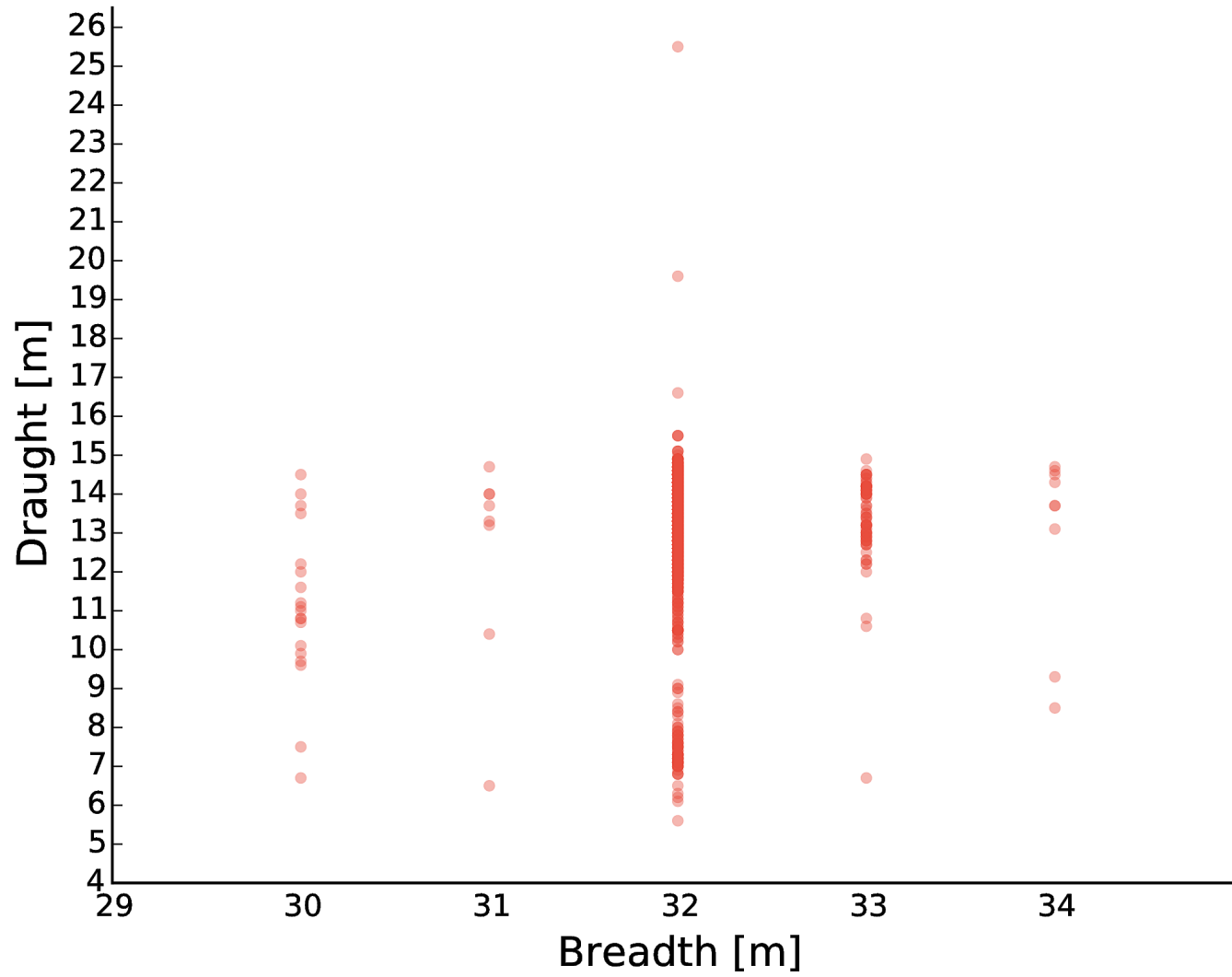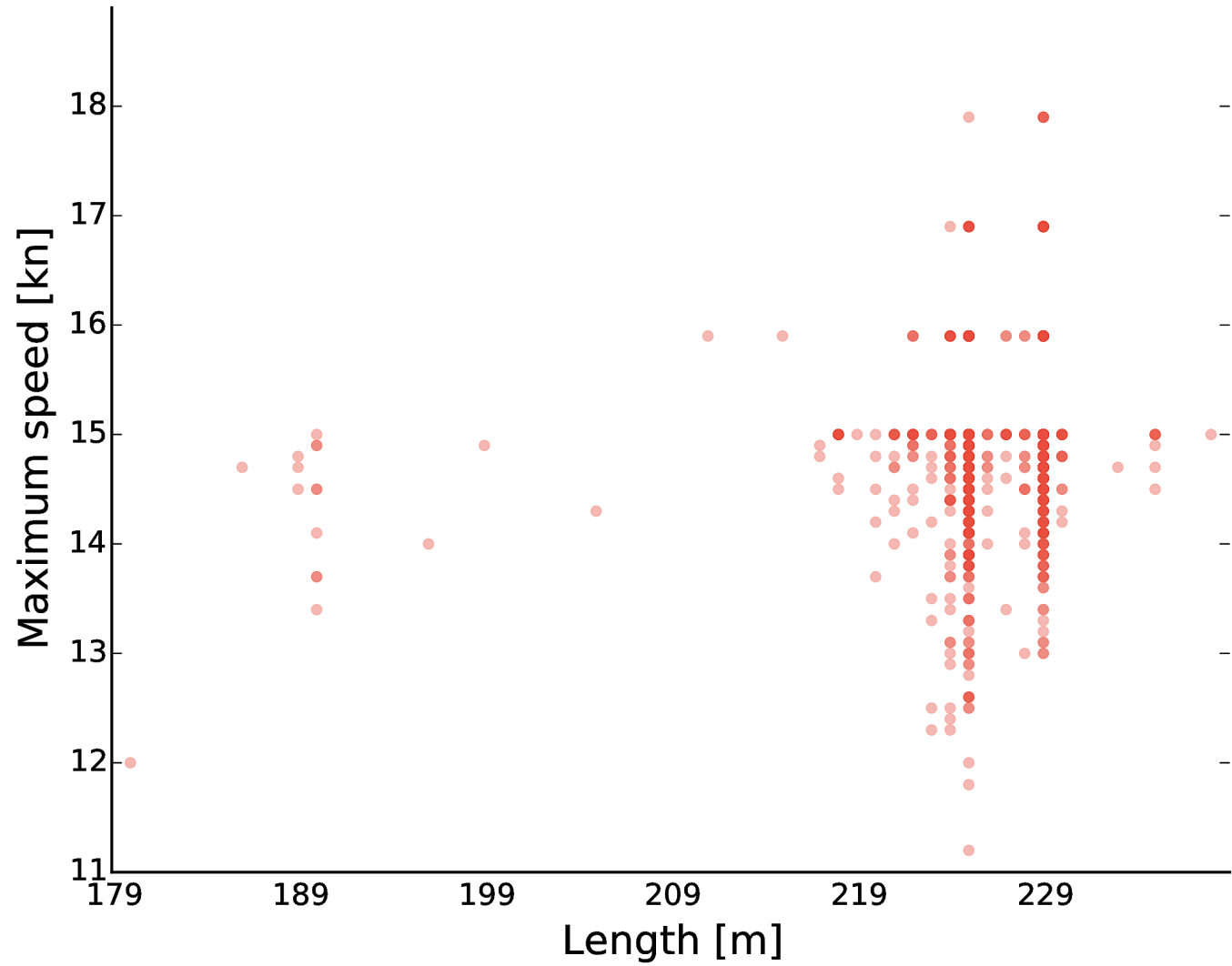Table G.22: Maximum draught and breadth for Panamax Bulk Carriers.

Table G.23: Maximum draught and length for Panamax Bulkc Carriers.

Table G.24: Maximum speed and length for the template Bulk Carrier group.

Total number of voyages: 64112
Median number of messages pr. hour: 0.206982936228
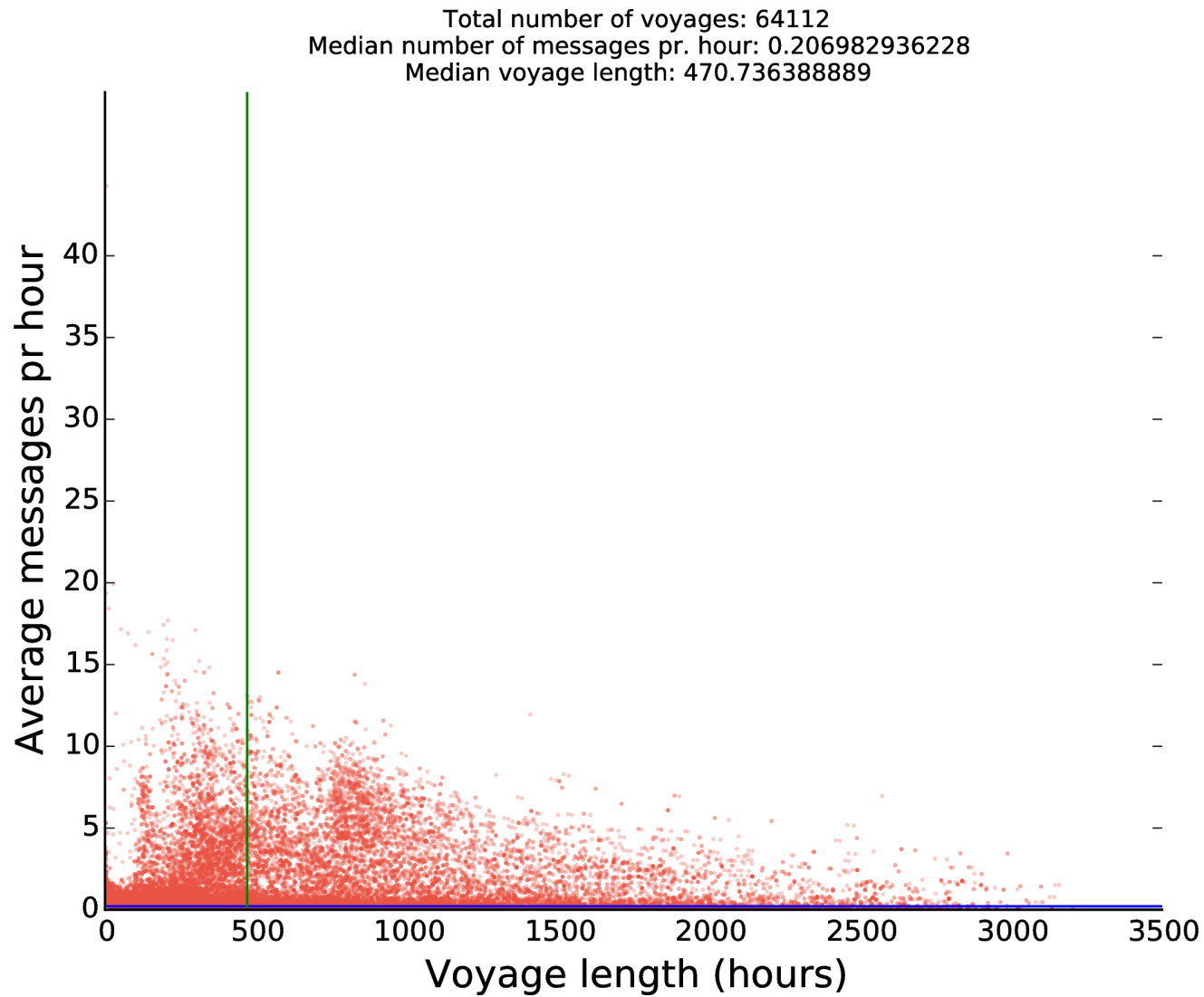Median voyage length: 470.736388889

Table G.25: Scatterplot over voyage length and average number of messages per hour for each voyage. The horizontal line signifies the median number of messages per hour, while the vertical line signifies the median voyage length.
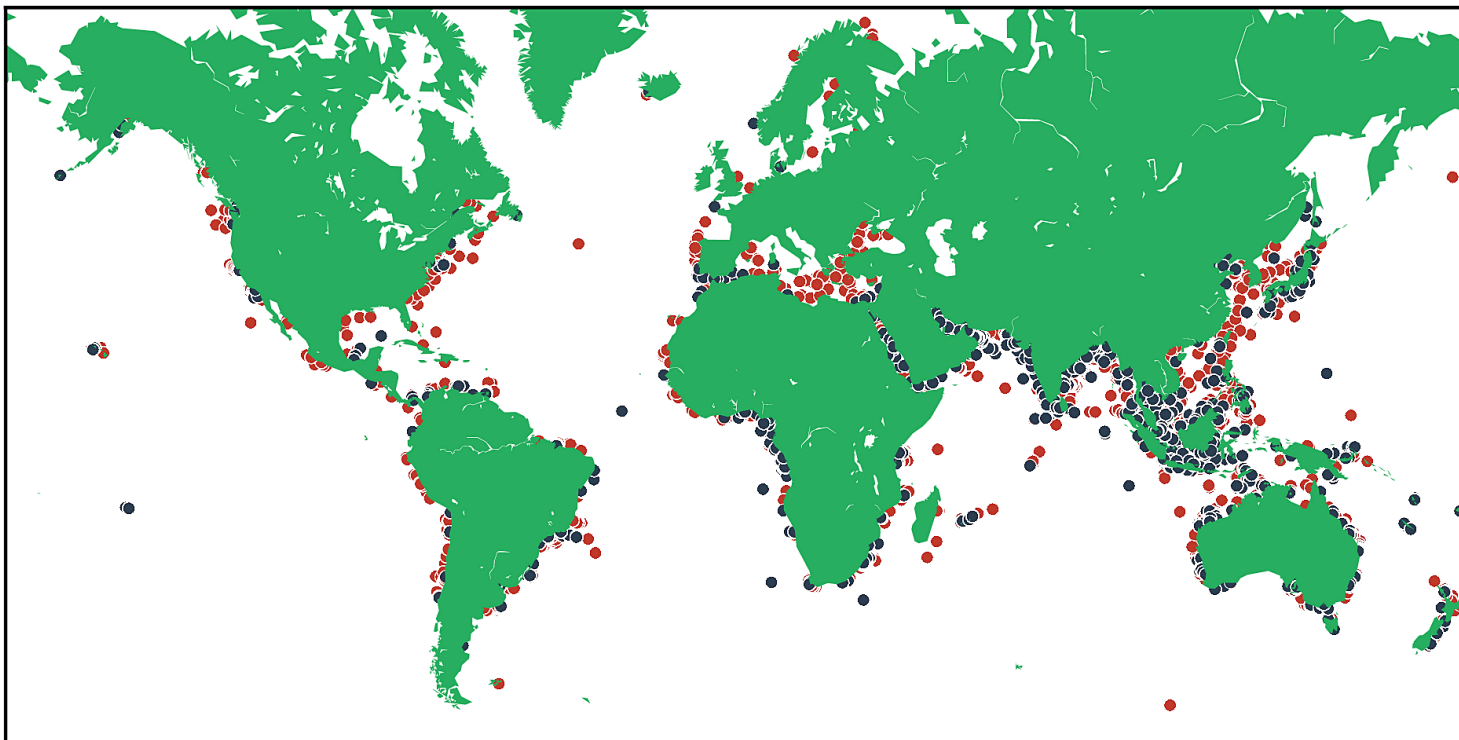
Table G.26: The origin of a selection of the tankers (blue) and cargo vessels (red) that could be backtracked from their entry to Singapore. These also have at least 6 dynamic messages per hour on average, a requirement made to easier visualize their voyage in an animation.