Encounter Probability of Inland Maritime Traffic

MSc Thesis

This thesis is made for the MSc program of Geographical Information Management and Applications (GIMA). A conglomeration between Utrecht University, University of Twente, Wageningen University and Delft University.

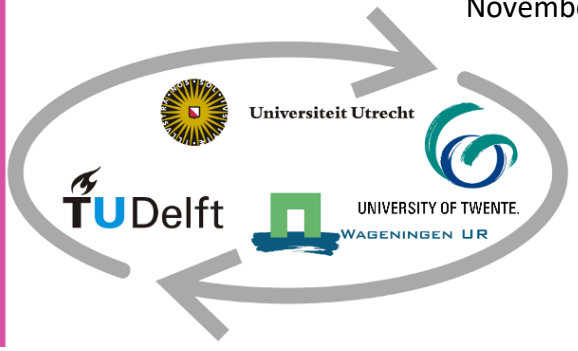
Supervisors:

dr.ir. B.M. Meijers & dr. Y.C. Altan, Delft University of Technology.

By

Thomas Lier

November 2019



13-11-2019
Thesis made by Thomas Lier
Student number: 4021169 (UU)
Email: t.b.lier@students.uu.nl

PREFACE

The bases for this research stemmed from my passions of solving complex spatial problems with seemingly easy solutions, that are reusable for a wide audience. The digital world is developing fast and large amounts of data are generated each moment. As we move further into the digital age, we need ways to access data efficiently. It is my chance to contribute to research how to tackle this challenge. And not only to access data efficiently, but also how to make it usable in a spatio-temporal context. A relatively new method, applied on an inland waterway has been examined and had been made accessible to a wide audience. The work represents a six months full-time effort.

All this work would not have been possible without the supervision of Martijn Meijers, associated with the Faculty of Architecture and the Built Environment, a Postdoc Researcher and Assistant Professor at Delft University of Technology. And the supervision of Yigit Altan, Postdoc Researcher at Eindhoven University of Technology and Assistant Professor at Behcesehir Univesity. I would like to thank Martijn and Yigit for the help they have provided, and for their expertise in the subject which led to good advice.

It was nice to work with relatively new programs and techniques. Looking back, writing a thesis was definitely my most difficult challenge during my Master program, but also a very valuable experience. I hope that this research will help in mapping and therefore mitigating maritime risks, and provide useful insights that can be of use in practice or future research.

Thomas Lier

SUMMARY

The main objective of this thesis is to develop a method to analyse and manage historic AIS data efficiently, and combine this with waterway characteristics to find manoeuvring points that help with the creation of an encounter probability risk map. The main objective is met by answering the following research question: **"In what way can Automatic Identification System (AIS) data be managed efficiently in combination with waterway characteristics in order to find manoeuvring points that can be used for the development of an encounter probability risk map? "** To be able to answer this question the data from historic AIS messages can be used for analysis, therefore the data has to be reconstructed. Because of the possibility of a large amount of historic AIS messages, storage of raw historic AIS data in a DBMS has to happen in an efficient way. The spatio-temporal component of the research, indicates that storage of the data alone is not sufficient. A object-relational database will be used. In this study the choice is made for the PostgreSQL database within pgAdmin with the PostGIS extension. These choices enable the possibilities to perform spatial analysis and typical queries on spatial data.

A case study will help in developing a concept that is applicable in maritime waterways in an international perspective, this includes: straits, gulfs and bays. For purposes of data management, a grid-based analysis had been chosen. In this research, a molecular collision theory is applied on a spatial distribution, on data from long term AIS messages in a limited and congested waterway. At this moment some approaches have tried to apply the theory to real waterways, but those studies remained limited within ship routes. In this research the occurrence of an encounter is taken as follows: whenever two vessel are expected to be within a predefined proximity of the same geographic location at the same time, than an encounter will occur. The probability of such encounters depends on the density of ships, its velocity and the meeting angle of approaching vessels. From the literature review it seems that ship movement remains not predefined and can be random along their route in congested waterways. Therefore the molecular collision theory was not applied route dependent, but upon sectors, that later were subdivided into cells. This grid-based cell design enables a spatial variation of the theory, and the cells have to be at least the size of the collision diameter, to not create artificial encounters.

By applying the theory on this grid-based design, the main factors determining the encounter probability in a given area point out to be: ship density, collision diameter and the relative velocity of the ships, within these cells within a certain time. Using these mathematical, probabilistic concepts, an encounter probability can be calculated, from which a risk map can be developed. The practicability and the applicability of the proposed approach and its integrated management of data was proved in a case study with historic AIS data. The case study has only shown a brief inside in the extensive possibilities of analysing inland maritime traffic. Comparing day and night inland maritime traffic is not the only possibility of comparing different moments. Analysis can be made for various moments, examples are comparisons in different day of the week, different months of even different years. Analysis can also be extended for various circumstances in weather, current, season, ship types or even a risk comparison between different waterways. Doing different spatial analysis on large datasets of historic AIS data invite as starting point for future research.

TABLE OF CONTENTS

List of Figures

List of Tables

Abbreviations

AIS	Automatic Identification System
AIVDO	Abbreviation for own vessel's information
AIVDM	Abbreviation for received data from other vessels
ASCII	American Standard Code for Information Interchange
CoG	Course over Ground
COS	Cosine
DGPS	Differential Global Positioning Systems
DIAMONIS	Dutch Inland AIS Monitoring Infrastructure
ETA	Estimated Time of Arrival
Geo-DBMS	Geographical Database Management System
GIST	Generic Index Structure
Git	Global Information Tracker
GPS	Global Positioning System
IMO	International Maritime Organization
JSON	JavaScript Object Notation
LORAN	Long-Range Navigation
MMSI - number	Maritime Mobile Service Identity - number
NMEA	National Marine Electronics Association
(Post)-GIS	Geographical Information Systems
(Postgre)-SQL	Structured Query Language
RoT	Rate of Turn
RWS	Rijkswaterstaat
SIN	Sinus
SoG	Speed over Ground
UTC	Universal Time Coordinated
Varchar	Variable Character
VTS	Vessel Traffic Service
WGS-84	World Geodetic System — 1984

1. Introduction

World seaborne trade is growing for years and will continue to grow with an average of 3.2 percent between 2017 and 2022 (Review of maritime transport (R.O.M.T.), 2017). Maritime transport is growing accordingly, and roughly 90 percent of world trade is carried by sea (Fiorini, Capata & Bloisi, 2016). The increasing maritime transport has its consequences for oceans and sea. Cargo flows are expected to expand across all segments (R.O.M.T., 2017). Maritime transport is crucial for world-wide economic development, but also represent financial and safety risks (Zhang, Goerlandt, Kujala & Wang, 2016). One of these risks can be the occurrence of a maritime accident. Though maritime accidents are relatively infrequent, the costs of such an accident can be enormous. An accident with a ship involved can have a huge economic and environmental cost (Heij, Bijwaard and Knapp, 2011), but also be of effect on human life or in traffic flows. For these and more reasons it is necessary to monitor the activities happing on waterways. To be able to monitor activities happening on busy waterways, the European Commissioner for maritime affairs and fisheries, wanted to develop activities for adequate and integrated monitoring and surveillance strategy for sea traffic back in 2012 (Fiorini, Capata & Bloisi, 2016).

1.1 Problem statement

One of the activities that help with developing adequate and integrated monitoring of sea traffic, is mapping of maritime risks in busy waterways. It should be acknowledged that almost all human activities involve risks, and the marine environment is a branch of human activities where risk is greater than average (Galor, 2009). Among waterways, there are areas where ship manoeuvring is more limited than in other areas (Galor, 2009). During navigation through a busy waterway, the ship may be affected by various factors which can make navigating difficult. Navigating through the maritime traffic on a properly planned trajectory is not always possible. An important part of this is where a ship decides to manoeuvre in a confined waterway and change its current trajectory. Changing the current trajectory can lead to navigational accidents and in confined waterways these type of accidents happen relatively frequent (Galor, 2009). When a ship drastically will change its course or speed, a manoeuvre occurs. This moment of manoeuvring happens at a certain point that will be defined as a manoeuvring point.

To avoid accidents between two vessels, one has to know if a possible encounter is going to happen on the current trajectory of two or more ships. Therefore, ship movement data is one of the essential parameters that can help in analysing maritime risks. The availability of ship movement data can be found in Automatic Identification System (AIS) data. The capturing of it happens trough Vessel Traffic Services (VTS) ashore. Although AIS was not designed for collision avoidance, using historic AIS data is a valuable source of data that can be used for analysis of waterway movement.

All vessels with a length greater than 20 meter, are obliged to carry an AIS transponder that transmits their position to nearby vessels. The data gathered by an AIS receiver can be used to analyse vessel movement based on real time data. AIS data contains information among others about the vessel identity, position, speed, and course (Perez et al., 2009). In particular, depending on a vessel's speed, AIS transmitters send data every 2 to 10 seconds while

underway and every 3 minutes when a ship is at anchor (Perez et al., 2009). Furthermore, every 6 minutes journey related information is being broadcast by a ship (Meijers, van Oosterom & Quak, 2016). In the Netherlands a network of AIS wall stations is developed by Rijkswaterstaat (RWS) to support vessel traffic management. The network is called Dutch Inland AIS Monitoring Infrastructure (DIAMONIS), and receives real time data, with that it supports a number of specific applications within RWS.

AIS has been proven a valuable source of information for ship traffic. Though it also created its own difficulties. From a previous situation of a scarcity of data, now an overabundance of data has to be worked with (Aarsæther and Moan, 2009). It has created such amounts of data that sifting through it manually is not possible, but should be done in an automated way. The large data amounts ask for automated processes that can help in the development of efficient and user friendly ways to analyse vessel movement. These processes are not widely available and can be seen as lagging behind the technical possibilities.

1.2 Research objectives

The main objective of this thesis is to develop a method to analyse and manage historic AIS data efficiently, and combine this with waterway characteristics to find manoeuvring points that help with the creation of an encounter probability risk map. The main objective is met by answering the research question.

Finding manoeuvring points in limited waterways will help in the creation of an effective risk map for the avoidance of possible accidents. Analyzing the waterway traffic characteristics in combination with historic Automatic Identification System (AIS) data will result in a bases for the risk map. In detail this information will be combined to form the bases of a quantifiable encounter probability map, that helps to understand risky places in waterways. This means that it will be part of the research to analyze historic vessel trajectories. The analysis will be based on historical AIS data.

1.3 Research questions

> In what way can Automatic Identification System (AIS) data be managed efficiently in combination with waterway characteristics in order to find manoeuvring points that can be used for the development of an encounter probability risk map?

This research question can be broken down into different sub questions:

1. What different methods are available for storing AIS data efficiently?
   a. What is AIS data?
   b. How to decode AIS data?
   c. How to store AIS data efficiently?
2. How to use a geodatabase management system (Geo-DBMS) for storing and managing historic AIS data?
   a. What Geo-DBMS is efficient in storing historic AIS data?
   b. How to preform spatial analysis on data stored in a Geo-DBMS, and how does that influence the choice for a certain Geo-DBMS?
3. Where can manoeuvring points in the waterway be observed?

a. What are the waterway characteristics of a certain waterway that will be used as case study?
b. How can the waterway be divided into different sectors where changes are observed in navigational characteristics?
c. How to decide ship traffic distribution along a waterway?

4. How can the historic AIS data be used for analyzing ship traffic?
5. In what way can the data be combined and visualized to create an encounter probability risk map?

1.4 Scope

The scope of this research is to develop a method to analyze and manage historic AIS data efficiently. This will be combined with waterway characteristics to find maneuvering points that help with the creation of an encounter probability risk map. That exits of a method of creating a reliable risk map for inland maritime traffic, in a structured, reproducible and semi-automated way. This involves, two main things. One is the development of queries and functions that can be applied on data in the database. The other is the integration of historic AIS data and waterway characteristics, for the development of a risk map.

1.5 Research limitations

The limitations of this research are on one side the availability of raw AIS messages and on the other side the limitation of AIS data itself. Goerlandt and Kujala (2011) recognized that AIS contains various types of errors. These erroneous can be filtered out before analysing the data. An expected lack of the model was that it does not work for predictions during approaching and departing manoeuvres. And after accessing possible valuable parameters, not all data in the dataset remains useful. Regarding the available amount of test data, the reliabilities of some results produced in the research becomes lower. But still it remains a very valuable indicator of how to implement and reproduce such a method.

1.6 Relevance

From accessing raw AIS messages to the creation of an encounter probability risk map. Enriching historic AIS messages enables different analysis possibilities not only regarding the encounter probability calculations for making risk maps, but various possibilities. The research is very relevant for scientific and more practical purposes of adequate and integrated monitoring of inland maritime traffic. Scientifically the research can be used as a starting point for risk analysis in inland maritime situations. Components of the research can also be applied in a broader perspective of maritime traffic. Practically, it enables the interested audience to map risks in a reliable manner for their own area of interest. Furthermore, the audience can use the explained database management principles for efficient handling of their dataset regarding their questionnaires. At this moment there is a gap in applicability of proven methods applied on large datasets. A manual execution is no longer a possibility for handling these amounts of data. This research not only contributes to this gap by semi-automating a lot of processes, through queries and database functions. It also contributes by explaining a step wise method that can be reused in an international perspective. Especially for research into risk full situations in inland maritime areas. A combination is made between making

historic AIS data easily accessible within a Geo-DBMS and waterway characteristics to indicate a risk in different waterway sector and cells.

1.7 Structure

The thesis will be displayed in the following order. First it starts with a preface, followed by a summary. The table of contents is listed and the structure of the thesis described. The topic will be introduced. Next, the research objectives are explained after which the related work will be discussed. This will contain a descriptive part of AIS data and how to store AIS data. Also, there will be a part on database management systems. Followed by a description of the case study area. These components together will be used to form the theoretical framework, what resulted in a conceptual model and design scheme after which the research methodology will be discussed. After that the results of testing are displayed and described. Next conclusions can be made and a discussion upon the research is given. Finally one can find the references, and an appendix for the GitHub library designed for the thesis research, important queries and database functions and histograms.

2. Related work

Maritime traffic is an international activity and this research will also be applicable to an international situation. The encounter probability risk map will be developed for a specific area, that serves as case study. Parts of the study should be applicable to other international waterways with different contexts and characteristics. Xiao et al. (2011) describes, that local regulation and also behaviour of officers in waterways do differ. Goerlandt and Kujala (2011) argue that their model should be capable of providing detailed information about the moment and conditions in which ships have an encounter. In the output of their model the relevant important parameters are, geographic location, encounter angle, time of day and type, size and speed of striking (Goerlandt & Kujala, 2011). For the encounter probability risk map, it will be important that there are possible similarities and differences of ship behaviour.

2.1 AIS data

To find out what the reasons for the behaviour of ship movement are, aspects like local regulation, behaviour of officers and characteristics of waterways will be looked upon. Also the detailed information about the geographic location, encounter angle, vessel size and its speed will be analysed. The different types of AIS data are described as static, semi-static and dynamic by Aarsæther and Moan (2009).

- Static data: Ship identification number (MMSI number), vessel dimensions.
- Semi-static data: Ship destination, hazard level of cargo and ship draft.
- Dynamic data: Time of broadcast, ship speed, rate of turn, course over ground and position (Aarsæther and Moan, 2009).

From this description it can be derived that AIS data is spatio-temporal object data. AIS data will exist of at least an vessel id, x (latitude), y (longitude) and t (time). It is clear that developing an encounter probability risk map, requires a lot of different data entries. A part of it can be interpreted from AIS data. But factors like waterway characteristics, should be derived from other (geographical) data. Because this work will be based on actual behaviour of ships the first step will be studying historical AIS data to find out what the characteristics of real-life ship movement are.

2.2 Storing historic AIS messages

The historic AIS data is not in a format where ship positions are reported in a sequence, but as a package of different position reports of different ships. To be able to apply the data from historic AIS messages for analysis, reconstruction of data is necessary. An important aspect of the ship position data is that from the AIS data it is given in the non-Euclidian WGS-84 reference system (Goerlandt & Kujala, 2011). For this reason distances and courses should be computed using appropriate formulas for geodetic computations (Goerlandt & Kujala, 2011). This limitation noted, the international telecommunication union defined 27 different top level AIS messages, from which message types 1-3 Position report and 5 Static and voyage related data will be used (International Telecommunication Union, 2014).

The National Marine Electronics Association (2012) (NMEA) defines an ASCII encoding, such that AIS messages can be send over a serial link to other equipment. In NMEA terms that means, an AIS message is a group of sentences (Meijers, van Oosterom & Quak, 2016).

(NMEA) !AIVDM,1,1,,B,13a=3800000CL=4Me?O`eVh20827,0*1A
(Timestamp) 2018-07-04 00:00:03.701389
Figure 1. NMEA sentence with timestamp.

Figure 1 is an example of an NMEA sentence with a timestamp. This example show that such a message can span multiple NMEA sentences, the AIS channel used for broadcasting, the raw AIS message itself, and a checksum. The NMEA messages don't always have the same content, it is depending upon the message type which parameter a raw message contains. Table 1 shows the parameters of the sample message in Figure 2. A NMEA sentence is split in words based on comma's. The raw AIS message is encoded as 6-bit ASCII characters and does not contain a timestamp (however during the gathering of data a base station can add such information). Next to the messages, also meta-information over the signal strength can be recorded with specific NMEA sentences.

Table 1. Parameters of sample AIS message decoded.

Parameter	Value	Description	Meaning of message type 1
Packet Type	AIVDM		
Channel	A		
Message Type	1	Scheduled Position Report	
Repeat Indicator	0		
MMSI	244532000		Maritime Mobile Service Identity - a series of nine digits uniquely identifying ship stations
Navigation Status		Underway Using Engine	"at anchor", "under way using engine(s)", or "not under command"
Rate of Turn (RoT)	0		Right (+) or left (-), 0 to 720 degrees per minute
Speed over Ground (SoG)	0		0 to 102 knots with 0.1 knot resolution
Position Accuracy		False	Longitude and Latitude – to 1/10,000 minute
Longitude	51.916584		
Latitude	4.2468834		
Course over Ground (CoG)	223		Relative to true north to 0.1 degree
True Heading	216		0 to 359 degrees from gyro compass
Time Stamp	0		Coordinated Universal Time (UTC) time accurate to nearest second when this data was generated
Reserved for regional	0		
RAIM flag		False	
Communication State	No value		
Communication Sync State	UTCDirect		
Communication Slot Timeout	2		
Communication Sub Message	No value		
Communication Utc Hour	No value		
Communication Utc Minute	No value		
Communication Time Stamp	No value		
Communication Slot Number	135		
Communication Received Stations	No value		
Communication Slot Offset	No value		

Table 1 only includes message type 1. As discussed, different message types exist. For example, every 6 minutes the AIS transmitter sends additional fields, including: a ship identification number, the vessels name, type of ship, ship dimensions, ships draught, etcetera (Fiorini, Capata & Bloisi, 2016).

The NMEA messages that are 6-bit ACSII encoded can be decoded in different ways. Four different methods of decoding have been tried (Figure 2; 2B JSON encoded, 2C Using a python library, 2D as 6-bit ASCII encoded & 2E as bit vector). Meijers, van Oosterom & Quak (2016) investigated different data types that were identified as promising for storing raw AIS data. Their initial testing results pointed out that the most promising option for storing AIS data for a DBMS is storage as bit vector (Meijers et al., 2016).

2018-07-04 00:00:42.450238
!AIVDM,1,1,,B,33M@vfm0010CLjpMeBK68SuD0DFr,0*4C
Figure 2A. Decoding one AIS messages in four different ways.

2018-07-04 00:00:42.450238
{"class": "AIS", "device": "stdin", "scaled": true, "status": 5, "status_text": "Moored", "heading": 126, "type": 3, "repeat": 0, "mmsi": 232013499, "turn": 0.0, "speed": 0.10000000149011612, "accuracy": false, "lon": 4.2489, "lat": 51.917833333333334, "course": 157.0, "second": 42, "maneuver": 0, "spare": 0, "raim": false, "sync_state": 0, "slot_increment": 5211, "slots_to_allocate": 5, "keep_flag": false}
Figure 2B. JSON encoded AIS message.

2018-07-04 00:00:42.450238
{'id': 3, 'repeat_indicator': 0, 'mmsi': 232013499, 'nav_status': 5, 'rot_over_range': False, 'rot': 0.0, 'sog': 0.10000000149011612, 'position_accuracy': 0, 'x': 4.2489, 'y': 51.917833333333334, 'cog': 157.0, 'true_heading': 126, 'timestamp': 42, 'special_manoeuvre': 0, 'spare': 0, 'raim': False, 'sync_state': 0, 'slot_increment': 5211, 'slots_to_allocate': 5, 'keep_flag': False}
Figure 2C. decoded AIS message using a python library.

2018-07-04 00:00:42.450238
33M@vfm0010CLjpMeBK68SuD0DFr
Figure 2D. Timestamp and raw AIS message (as 6-bit ASCII encoded).

2018-07-04 00:00:42.450238
000011000011011101010000111110101110110101000000000000000000100000001
001101110011001011100001110110110101001001101100011000100010001110010
010101000000000101000010110111010
Figure 2E. Timestamp and raw AIS message (as bit vector).

Storage of raw historic AIS data in a DBMS has to happen in an efficient way because of the large amounts of data that have to be worked with. As seen earlier in table 2, that often null values are present, a data reduction method can be convenient. Each raw AIS record can be interpreted as a point in the world space, multiple points might be grouped based on their proximity in terms of latitude and longitude values (Fiorini, Capata & Bloisi, 2016). Grouping

of multiple points can help in reducing data. Furthermore, because null values are often present, it might not be necessary to have direct access to all attributes of the messages. Other options next to grouping multiple points should be considered in this research.

Inside the scope of the research is looking into ship encountering processes. In this process ships approach each other and if it goes right, they will finally depart from each other (might be with an evasive action). This process has a spatio-temporal component when ships could fall into a dangerous situation with the potential for collision (Huang, Mou, & Chen, 2018). To make this encountering process tangible, certain criteria have to be defined, where the distance and direction between two ships should be less than the threshold, because in that situation a dangerous encounter occurs. The spatio-temporal component of this research, means that raw storage of data alone is not sufficient. Meijers et. al. (2016) state that two parts are needed to answer spatio-temporal queries within reasonable time limits. One part is that access to relevant parameters of AIS messages is needed (among others; MMSI, destination and course over ground), the second part is that, indexing and eventually clustering of the records need to be performed (Meijers et al, 2016).

It should be noted that AIS data does have its limitations. Goerlandt and Kujala (2011) recognized that AIS contains various types of errors, examples of this are data corruption, a possible erroneous MMSI number, a faulty position report or even errors in other parameters. Because of the existence of these errors, to develop a reliable method for data error checking should be incorporated. This further complicates the data handling. An option to check the data can be by means of an algorithm, but Xiao et al. (2015) argue in their research that only a small proportion of signals shows ambiguous positions. Thus, in their research it was not a problem, these ambiguous positions were eliminated in the statistical analysis (Xiao et al., 2015).

2.3 Encounter probability concepts

For deciding on a critical situation where an encounter risk might occur a critical distance can be defined. In literature, 0.5 nautical mile (nm) was suggested (Fowler and Sorgard, 2000). Most collision probability models use a critical encounter distance. This is unrealistic as vessel dynamics in collision avoidance is not taken into account (Goerlandt & Kujala, 2011). Using such an critical distance will only signify the proximity and not indicate whether ships would come in to contact. Though an entering situation of two ships within a critical distance can be a valuable indicator for a risk full event. Although this might be valuable indicator of a risk full situation, most research has been focused on estimating ship collision or grounding (Li et at., 2012) in specific water areas. The models used for this are called ship accident frequency models.

One of the first and most popular approaches to ship collision and grounding models is proposed by Macduff (1974) and Fuiji (1974), where the probability of a vessel being involved in a collision, during passage of a specified area, has been formulated as follows: P=Pa*Pc. In the equation, Pa stand for the geometrical probability for a vessel encountering in an accidental scenario, if no aversive measures are made. Pc stands for the causation probability, what means the conditional probability that a collision occurs in an accidental scenario

(Macduff, 1974). After this first method, many researches have contributed to probability estimations, either causation probability, geometrical probability or both.

Based on this equation the collision probability needs input from two probabilities, the encounter probability and causation probability. Li et al. (2012), states that they are independent probabilities. Though they do seem to have some overlap. Geometrical probability is dependent of at least the geometric parameters of water area, amount of traffic, vessel size, speed over ground (SOG), course over ground (COG) and causation probability is determined by at least captains sailing skills, and the maneuverability of a vessel (Li et al., 2012).

A concept by The International Maritime Organization described risk as a combination of frequency and severity of consequence (IMO, 2002). In this concept, consequence means the outcome of an accident and frequency the number of occurrences per unit time, for example per year (IMO, 2002). Another commonly accepted concept for risk is R = P*C. R represents the risk, P the probability of an unwanted event and C the consequence of that event. This study focusses on the probability, but the consequence is equally important in a regular risk analysis (Goerlandt & Kujala, 2011). The occurrence of an encounter is taken as follows: whenever two vessel are expected to be within a predefined proximity of the same geographic location at the same time, than an encounter will occur.

This corresponds to the collision diameter theory, were the collision will be replaced by encounter. Monocular collision theory is one of several quantitative risk assessment models that can be used for shipping waterways (Li et al., 2012). At this moment some approaches have tried to apply the theory to real waterways, but those studies remained limited within ship routes. Altan and Otay (2017), applied the molecular collision theory on a spatial distribution, on data from long term AIS messages in a congested waterway, where the results are compared with past maritime collision records. For this research, a similar collision theory will be applied to calculate the encounter probability. Applying the collision theory for calculating an encounter probability is not the only method that can be used for finding risk full events in maritime traffic. Ship domain and velocity obstacle are examples of other methods that can be used, in this thesis research the focus will be upon the collision theory.

3. Theoretical framework

When making choices in a suitable database management system for this study several considerations have to be made. First, the choice for a data model has to be made. Second, thought has to be given to the efficiency of the data model and to the usability, for example to integrate the data into different applications.

For this study it seems to be necessary to use a form of object-relational database. Because there is a necessity for being able to perform spatial analysis and typical queries on spatial data. Because of the complexity of the spatial queries that are needed to visualize the data, there will be relied on the PostgreSQL database[1] within pgAdmin4[2] with the PostGIS extension[3]. In the scope of this research there will not be relied upon different types of databases, although they might offer efficient ways of handling data too. PostgreSQL as object-relational database has to possibility to work with geographic objects allowing locational queries to be run in SQL. In addition for more complex spatial queries, PostGIS offers many features rarely found in other competing spatial databases such as Oracle. Also the PostgreSQL database offers possibilities for spatio-temporal data clustering, what is a process of grouping objects. The grouping of objects happens based on their spatial and temporal similarities. The PostGIS spatial index is built using the PostgreSQL GIST (generic index structure) index infrastructure (Paul Ramsey blog, http://blog.cleverelephant.ca/).

Third-party software used at this moment.
PostgreSQL 10.6 was used as database system and the spatial database extender PostGIS was installed to add support for geographic objects. PostgreSQL is released under the "PostgreSQL License", a liberal Open Source license. PostGIS is released under the GNU General Public License.

3.1 Initial test for choosing a data type

Based upon research of Meijers et al. (2016), a data type for storing historic AIS messages in a database was chosen. A load script was used to load to multiline NMEA sentences, their timestamp and payload. The table was created using two columns, one for the timestamp and a column is used for the payload of the AIS messages. The data type of the payload was varchar. Using the data type varchar is not desirable for efficient storage. Using bit varying as data type is better suited. There are two main advantages and one disadvantage of the datatype bit varying. The main advantages exist of; substring functions that are available with the data type and compact storage is achieved (Meijers et al., 2016). A disadvantage of the data type is that parsing cost performance (Meijers et al., 2016).

Interesting is that bit vector and varchar data type did use the same amount of space in the research, it was an unexpected outcome, but still the most promising option seemed to be the bit varying type. The database supplies functions to interpret parts of the bit vector and also allows to make function based indexes (Meijers et al., 2016). Indexes do cost some more storage space, but gives fast access to attributes of the individual AIS messages, what enables access to not frequently accessed attributes possible (Meijers et al., 2016).

For testing with the dataset, with a goal to answer spatio-temporal queries, raw storage of the AIS messages alone is not sufficient. Access is needed to the relevant parameters that AIS

[1] (www.postgresql.org)
[2] (pgadmin.org)
[3] (postgis.net)

messages contain. Examples of the parameters are MMSI number, speed over ground, course over ground, destination, etcetera. To be able to extract these parameters, first functions have to be used to access the relevant parameters. Secondly, indexing and eventually clustering of the records should be performed.

3.2 Functions for access of AIS parameters

Various database functions for accessing different AIS parameters are available for AIS messages stored as bit vector. First the raw NMEA string had to be encoded as bit vector, from a varchar data type. Below is an example of a PL/PGSQL function used to encode NMEA strings as bit vectors.

```
-- Encode nmea string as bit vectors. (AIVDM/AIVDO Payload Armoring)
-- This code should only be used on the data payload

CREATE OR REPLACE FUNCTION ais_bit_vectorize(payload varchar)
RETURNS bit varying AS $$
BEGIN
        RETURN
        replace(replace(replace(replace(replace(replace(replace(replace(replace(replace(replace(replace(replace(repl
        ace(replace(replace(replace(replace(replace(replace(replace(replace(replace(replace(replace(replace(replace(
        replace(replace(replace(replace(replace(replace(replace(replace(replace(replace(replace(replace(replace(repl
        ace(replace(replace(replace(replace(replace(replace(replace(replace(replace(replace(replace(replace(replace(
        replace(replace(replace(replace(replace(replace(replace(replace(replace(payload, '0','000000'), '1'
        ,'000000'), '2', '000010'), '3', '000011'), '4', '000100'), '5', '000101'), '6', '000110'), '7', '000111'), '8', '001000'),
        '9', '001001'), ':', '001010'), ';', '001011'), '<', '001100'), '=', '001101'), '>', '001110'), '?', '001111'), '@',
        '010000'), 'A', '010001'), 'B', '010010'), 'C', '010011'), 'D', '010100'), 'E', '010101'), 'F', '010110'), 'G', '010111'),
        'H', '011000'), 'I', '011001'), 'J', '011010'), 'K', '011011'), 'L', '011100'), 'M', '011101'), 'N', '011110'), 'O',
        '011111'), 'P', '100000'), 'Q', '100001'), 'R', '100010'), 'S', '100011'), 'T', '100100'), 'U', '100101'), 'V', '100110'),
        'W','100111'), '`', '101000'), 'a', '101001'), 'b', '101010'), 'c', '101011'), 'd', '101100'), 'e', '101101'), 'f',
        '101110'), 'g', '101111'), 'h', '110000'), 'i', '110001'), 'j', '110010'), 'k', '110011'), 'l', '110100'), 'm', '110101'), 'n',
        '110110'), 'o', '110111'), 'p', '111000'), 'q', '111001'), 'r', '111010'), 's', '111011'), 't', '111100'), 'u', '111101'), 'v',
        '111110'), 'w', '111111')::bit varying;
END;
$$ LANGUAGE plpgsql
IMMUTABLE
;
```

Figure 3: A PL/PGSQL function for Encoding NMEA string as bit vectors.

An extended set of database function were defined to access the different parameters of AIS messages. These functions are designed for the bit vector datatype. Interpreting the raw AIS data was done through the document of Raymond and detailed explanation of Meijers and Altan (Raymond, 2016; Meijers et al., 2016; Meijers et al., 2017; Altan & Otay, 2017). The following functions for use inside the database where made for decoding the following parameters:

- Bit vectors
- Callsign
- Course over ground
- Destination
- Dimension to bow
- Dimension to port

- Dimension to starboard
- Dimension to stern
- Draught
- Easting
- Northing
- Encode 6 bit
- Maneuver
- MMSI
- Navigation status
- Point (geographical location to a point feature, a function provided by PostGIS)
- Rate of turn
- Ship type
- Speed over ground
- True heading
- (Message) type
- Vessel name

The codes can be viewed through a Git repository that is created for the thesis research. A more extended explanation of the Git principle can be found in Appendix 1: Git.

3.3 Indexing tables

After defining the different possible functions, the following indices on the tables were used directly from Meijers et al (2016):

- Index on MMSI parameter of the AIS message: Index type is B-tree.
- Index on message type parameter of the AIS message: Index type is B-tree.
- Partial index on combined 'latitude' and 'longitude' parameters of the AIS message: Index type is R-tree, and is only created for the messages with message type in (1, 2, 3) and within the geographic domain: ((-90,-180), (90,180)), this excludes AIS messages with (91,181) coordinate.
- Full index on geometry.
- Index on timestamp of the AIS message: index type is B-tree.

3.4 Molecular collision theory on a spatial distribution

When the distance between two ships is equal to the collision diameter, an encounter occurs. The probability of such encounter depends on the density of ships, its velocity and the meeting angle. Corresponding data available from the AIS messages are; ships speed over ground (SOG), course over ground (COG) and ship dimensions. Ship movement is not predefined and can be random along their route in congested waterways. Therefore the molecular collision theory should not be applied route dependent. And the research area is divided into sectors, and will be further subdivided into cells.

As explained in theory, when the distance between two ships is equal to the collision diameter, than there will be a collision. As the collision diameter is influenced by dimensions and velocities of ships, it will change accordingly with different values. Another important

parameter is time. Pedersen (1995), formulated the application of this concept to ship collision as:

$$P_E = \sum_i \sum_j \iint_{z_i z_j} \rho_i(z_i)\, \rho_j(z_j)\, D_{ij}\, V_{ij}\, T\, dz_i dz_j$$

Where,
PE, is the encounter probability.
zi, is the position of the ship on the ith route.
zj, is the position of the ship on the jth route.
ρi (Zi), is the ship density as a function of z in the ith direction.
ρi (Zj), is the ship density as a function of z in the jth direction.
T, is the time period.
Dij, is the collision diameter of the ships on ith and jth route.
Vij, is the relative velocity of the ships on ith and jth route.
(i and j represent the boundaries and the travel direction of the ships)

In the research area the number of ships have to be known in order to find the encounter probability (Altan & Otay, 2017). The ship density represents the number of ships per unit area per time and can be calculated as follows:

$$\rho_i = \frac{Q_i}{v_i \Delta t} \frac{\Delta t}{T} = \frac{Q_i}{v_i T}$$

Where,
Qi, is the number of ships per unit width in the ith direction.
Vi, is te velocity of ships that travel in the ith direction
Δt, is the travel time between the considered area.

Next to considering the number of ith direction of ships in a certain area, the travel speed of these ships is considered as well. With a lower speed, ships will be longer in a considered area and therefore get a higher density compared to ships with faster speeds (Altan & Otay, 2017). Furthermore, time spend inside the given area over the time period of the observation is considered in the formula: Δt/T (Altan & Otay, 2017).

The relative velocity of encountering ships determines the collision diameter line and consequently the encounter probability (Altan & Otay, 2017). It is calculated as:

$$V_{ij} = |V_i - V_j| = \sqrt{V_i^2 + V_j^2 - 2V_i V_j \cos\theta}$$

Where,
Vi, is the velocity of the first ship
Vj, is the velocity of the second ship
Θ, is the angle between the velocity vectors.

The collision diameter depends on meeting angle and size of the ship. And indicates that there will be a collision when two ship centers will be inside of the collision diameter (Altan & Otay,

2017). It should be noted that the collision diameter is developed for the distance between ship centers, therefore the most accurate method here would be using ship centers as collision diameter. The value of the collision diameter is the projection of the two approaching ships on to a line. The calculation of Pedersen (1995) is as follows:

$$D_{ij} = \frac{L_i V_j + L_j V_i}{V_{ij}} \sin\theta + B_j \left\{ 1 - \left(\sin\theta \frac{V_i}{V_{ij}} \right)^2 \right\}^{1/2} + B_i \left\{ 1 - \left(\sin\theta \frac{V_j}{V_{ij}} \right)^2 \right\}^{1/2}$$

where,
L, is the length of the ship.
B, is the width of the ship.

This formula is calculated by collision types, what is a function of meeting angles (Altan & Otay, 2017).

Take-over collision ($\Theta = 0^0 \mp 10^0$)
Crossing collision ($10^0 \leq \Theta \leq 170^0$)
Head-on collision ($\Theta = 180^0 \mp 10^0$)

By combining the different components necessary to make the proposed encounter probability calculation. The encounter probability formula can be applied to each cell individually. As described the cells should at least be larger than the collision diameter, otherwise the collision diameter might exceed the border and create artificial collisions (Altan and Otay, 2017).

$$P_E = \sum_s \sum_i \sum_j \sum_{ik(s)} \sum_{jk(s)} \rho_i(ik(s), s)\, \rho_j(jk(s), s)\, V_{ij}(s)\, D_{ij}(s)\, \Delta l(ik, s) \Delta l(jk, s)\, T$$

The formula that should be applied cell by cell is;

where,

s, represents the sector
i & j, the boundaries (N, S, E, W) of the sector
i = N, S, E, W
j = i, …,W
ik & jk, the cell inside the sector according to the entrance
D_{ij}, collision diameter
Δl, travel distance
T, considered time

The main factors determining the encounter probability in a given area are ship density, collision diameter and the relative velocity of the ships (Altan and Otay, 2017). It will only consider one encounter at a time, since ship encounters are rare events. An expected lack of the model is that it does not work for predictions during approaching and departing maneuvers.

3.5 Conceptual schema & data modelling

By combining the theory of the collision diameter method with a functional data model, the schematic model towards risk mapping of the encounter probability can be designed (Figure 4). The DBMS chosen for this research is a object relational database that has been extended with PostGIS. This choice has to do with the spatio-temporal character of the AIS data and the combination of spatial data created as waterway sectors. These two parameters form the basis of the schematic model. The raw AIS messages on one side of the model and the waterway cells on the other side of the model. The raw AIS messages have to be enriched with data that they contain, from message type 1,2,3 and 5. The SoG, CoG and the timestamp of a ship can be derived from message type 1,2 and 3. The dimensions of a ship from message type 5. The parameter of the waterway cells have to be decided by combining different parameters. Observed manoeuvring points in combination with waterway characteristics will be used to from first sectors, that later can be subdivided in waterway cells.

Having the required input parameters, the schematic model continues with the approach angle and length of all (LOA), and beam (the widest point measured of a ships regular waterline). As described in the previous paragraph of molecular collision theory on a spatial distribution, these values have to be calculated in order to perform the calculations necessary towards an encounter probability calculation. Therefore the next three parameters in the schematic model are the execution of the mathematical concepts. First the calculation of ship density, second the calculation of the relative velocity of ships and third, the calculation of the collision diameter. When these calculations are preformed, the encounter probability calculation can be made. After making that calculation, a risk map can be designed. How to performs these different steps has been visualized in figure 4 and will be explained in further detail in the next chapter.

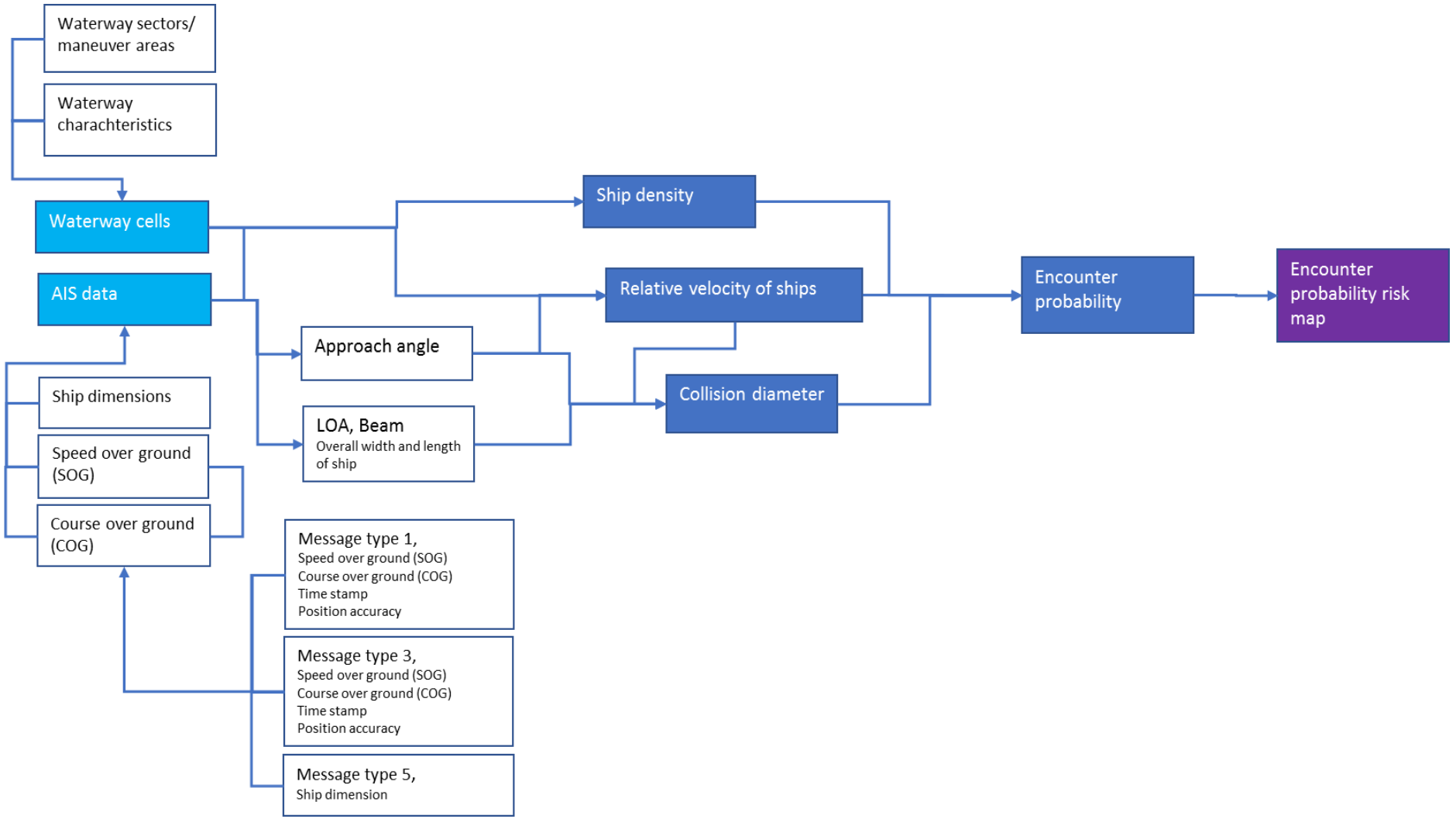


Figure 4. Schematic model of encounter probability risk mapping.

4. Methodology

The scope of the thesis is to develop a method to analyse and manage historic AIS data efficiently. This part will be combination with waterway characteristics to find manoeuvring points that help with the creation of an encounter probability risk map. This chapter will first discuss the organization of the DBMS. After what a case study will be discussed, followed by the spatial application of the collision diameter theory and a data preparation fase. The chapter will be concluded by a paragraph about database implementation.

4.1 Database organization

The basis for a reliable database management system is the data model. The choice for the right database that is suitable for managing, storing and structuring AIS and waterway data is dependent upon requirements, that are based on this case. Klein et al. (2015) developed and described a method for finding a suitable and efficient database organization. This resulted in a methodology that can be implemented for making a database choice (figure 5), the model is based on the research of Klein et al. (2015).

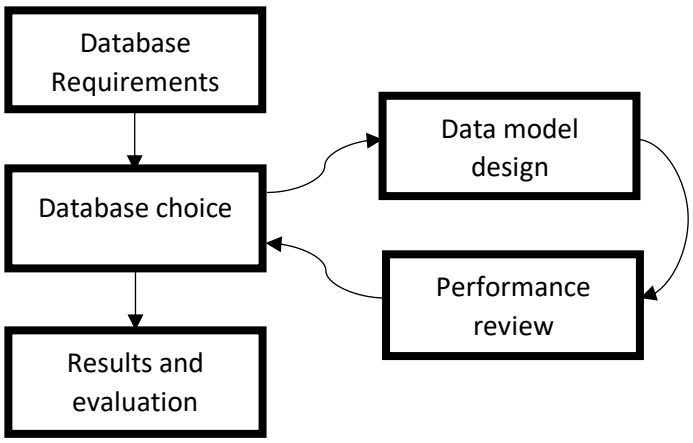


Figure 5. Database organisation method.

The method described is designed in five different parts adapted to the needs of this research. The five parts are: database requirements, the choice for a database, an iterative part of data model design and review of the performance, and results and evaluation. The iterative part helps to verify the right database choice. After which the database can be used to produce results and evaluate on them.

From the related work and the theoretical framework the database requirements have been decided and a database choice has been made. An object-relational database has been chosen because of the necessity for being able to perform spatial analysis and typical queries on spatial data. Furthermore, the main reasons leading to this decision were possibilities in the DBMS like clustering and indexing that might be necessary for large datasets and fast retrieval of data. For the performance review of the database organization, statements about the response time of queries will be made. The storage size will be discussed and finally also the effectiveness of used indexes can be evaluated.

4.2 Data model design part 1

A goal for this research is to use a data model that can be replicated correctly. Therefore implementing a similar method would become less complex. This is relevant for giving a wide audience access to the research. The different queries used for working out de encounter probability will be shown in paragraph 4.6 the implementation of the database. And in this paragraph the initial input to the database will be discussed.

The PostGIS extension makes use of the default 'public' PostgreSQL schema. An UTF8 encoding is used on the create database, with no other specific settings. Based upon two main tables the research can be based. One started with the collection of raw AIS messages. And the other with the waterway cells, stored by their boundaries as polylines. Where an how these boundaries are designed will be discussed in the next part of the case study. Thus by accessing the available parameters of the raw AIS messages by using the different functions and queries designed in chapter 3.2 functions for access of AIS parameters, the following columns have been made and filled with data (Table 2).

Table 2: Database columns of the table containing the raw AIS messages.

Name	Data Type		Name	Data Type
aisencode6bit	**text**		**mmsi**	**integer**
aispoint	**geometry**		navigationstatus	integer
bitvectorcombined	bit varying		nmeastring1	bit varying
callsign	text		nmeastring2	bit varying
channel	character varying		northing	integer
checksum	character varying		numberofsentences	character varying
courseoverground	**integer**		packettype	integer
datetime	**timestamp**		rateofturn	integer
destination	text		sentencenumber	character varying
dimensiontobow	**integer**		sentencetype	character varying
dimensiontoport	**integer**		sequentialmessageid	character varying
dimensiontostarboard	**integer**		shiptype	integer
dimensiontostern	**integer**		**speedoverground**	**integer**
draught	integer		trueheading	integer
easting	integer		vesselname	text
maneuver	integer		trueheading	integer
messagetype	integer			

The 394103 messages each are displayed in an unique row, where for each possible parameter that message contains the rows are filled with data. Therefore, many fields have a 'null' value,

of not containing any data. A good example of a field with many null values is nmeastring2, where only a raw AIS message exists that of two messages will get a value. More specific, taking a close look at the schematic model for data necessary for the data model design. The case study in combination with AIS data, only some attributes shown in table 2 are useful for the next steps in this research. Those are: AIS point (the locational data), date-time, MMSI (identification number of a vessel), SoG, CoG and the different ship dimensions. Datetime is stored as a timestamp. AIS point are stored as geometry. The other values of interest are stored as integers. To be able to retrieve the original data, the raw input should also be stored.

The other main table of interest is the table containing the sector and cell boundaries. The creation of the sectors and cells will be discussed is the next paragraphs.

4.3 Case study; ship behavior inside the waterway

For the case study a part of the Rotterdam harbour is chosen. But the concept applied in this study should be applicable to maritime waterways in an international perspective, that includes: straits, gulfs and bays. A visualization of the research area is made for a part of the Nieuwe Waterweg (Figure 6). The research area is part of the main shipping connection: Nieuwe Maas, Nieuwe Waterweg and Maasmond. Xiao et al. (2012) did describe the Rotterdam waterway in a comparison study, resulting in the following information. The waterway has the characteristics of a straight and narrow waterway with 10 degrees of variance. The navigable channel of the Port of Rotterdam is about 270 meters wide and has traffic in both directions without separation scheme (Xiao et al., 2012). Furthermore, large numbers of ships are expected to pass through every day. The passages are expected to be made by various types of ships with a wide range of dimensions. The waterway is influenced by tide and river discharge.

Figure 6. Research area of a part of The Nieuwe Waterweg, inside the Rotterdam harbor area.

The research area is based on a main transport axle. It is assumed here that the navigable width of the waterway is an addition to a number of different aspects, namely: the number of water strips, usually one for each sailing direction, to be regarded as the envelope of the path

widths of all occurring ships, one or more safety strips between the sailing lanes, the width of which depends on expected ascending or meeting vessels (Koedijk et al., 2017).

According to Xiao et al. (2012) ship courses do not vary much when sailing through the waterway. They explain this by giving three reasons, the location of the case study is a straight, ships normally do not need to change their course during passage. Also, the waterway that is studied is relatively narrow (about 270 m navigable width), and there is no room to change course significantly. Finally, the only reason for changing course would be in an encountering situation (Xiao et al., 2012). In a head-on encounter, the ships normally do not need to shift the lateral position much, as the ships are navigating by the starboard side. That is an interesting statement, so changes in course could mean an encountering was happening. Also this head-on encounter doesn't count for ships that have a different destination than straight through the waterway. Inside the limited width of the waterway there is no room for overtaking (figure 7, Research area with AIS data visualized on MMSI), so ships may change speed to try to make a safe maneuver and avoid an encountering situation (Xiao et al., 2012).

Figure 7. Research area with AIS data visualized on MMSI

It seems that the ship speed is constantly changing in the waterway. The average grades from the research of Xiao et al. (2012) show that the larger ships sail at smaller speed. The different speed choices can be explained with the fact that changing speed for a large ship is not an easy task. Not only does it take more energy for a large ship, it also takes more time to make changes compared to smaller ships (Xiao et al., 2015). A dangerous situation could occur if larger ships would need to reduce or increase speed suddenly.

For purposes of data management, a grid-based analysis has been chosen. The research area is divided into grids, that are called sectors. The entrance or exit of a sector can be analyzed to find different characteristics, that give a representation of the sector. The size of the sector is an important factor. Namely, inside the sector minimum changes in course and speed are expected. To decide where minimum changes in course and speed occur several things are observed. A dynamic map has been made, where the ship movement over time can be observed. Unfortunately this map cannot be displayed due to technical limitations of not being able to display a movie in a paper. Also waterway characteristics are observed, this resulted

in the following sectorization, see figure 9. The research area is divided into 5 different sectors. The schematic representation of these sectors is given in figure 8.

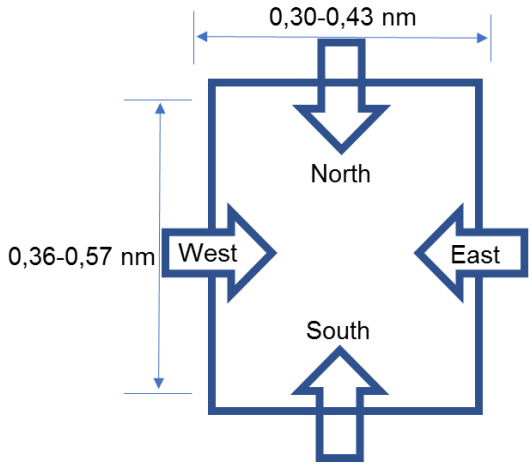


Figure 8. The schematic representation of sector.

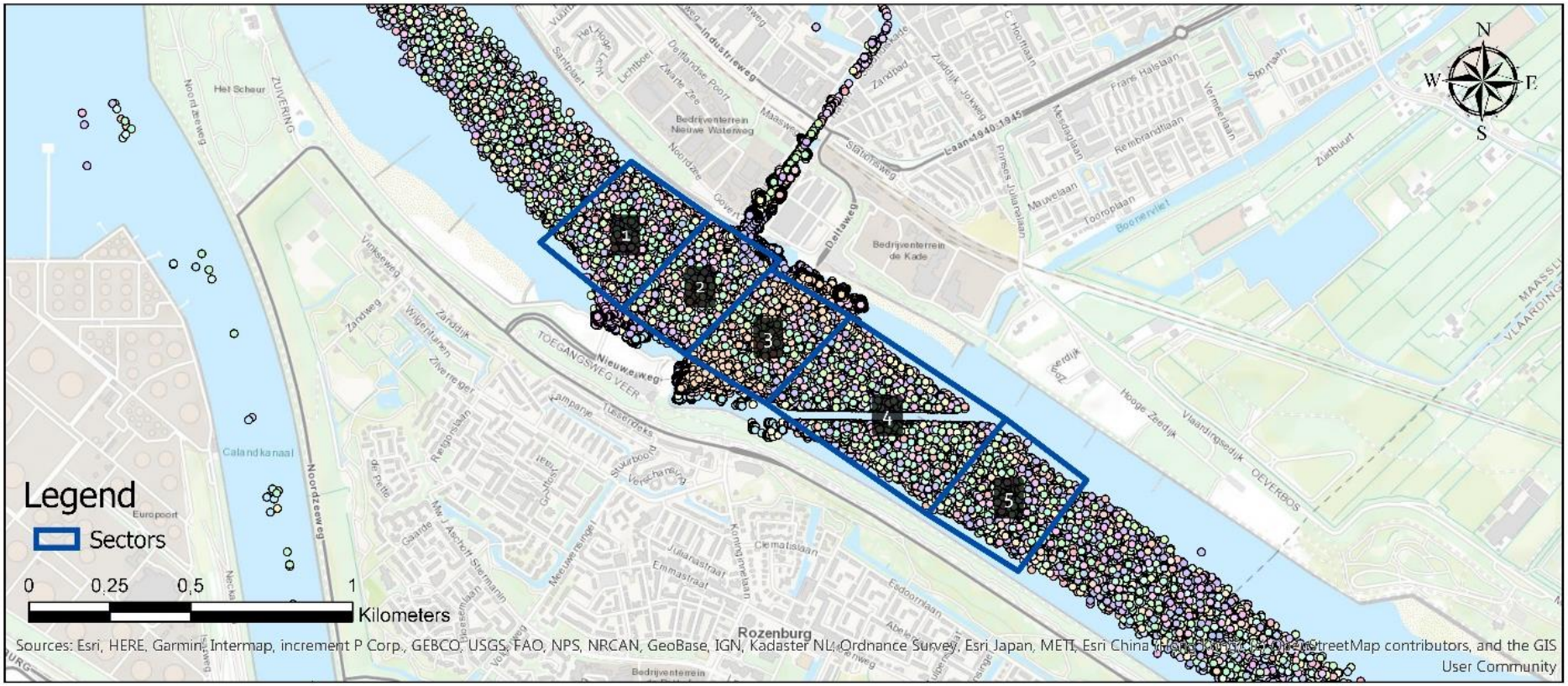


Figure 9. Research area divided in 5 sectors

4.4 Spatial application of the collision diameter theory

Altan and Otay (2017) point out that the main equation by Pederson (1995), gives the number of collisions for given ship routes in the considered area. This means that it is not indicative for collisions where there are no designated routes. Therefore, to achieve a true spatial distribution, the molecular collision theory cannot be applied directly (Altan and Otay, 2017). To be able to achieve spatial variation the theory has to be applied on smaller parts of the research area, which are the grid-based design into sectors (Figure 9) and further subdivision into cells. The cells where the formula will be applied upon should be at least the size of the collision diameter, else artificial encounters will occur.

In the research of Altan and Otay (2017) inside the sectors, change in course over ground and speed over ground are minimum. For the limited research area it is difficult to decide where to create sectors but for creating cells further vessel movement analysis is required. For creating cells, the movement of vessels in the different sectors have been analyzed (figure 10).

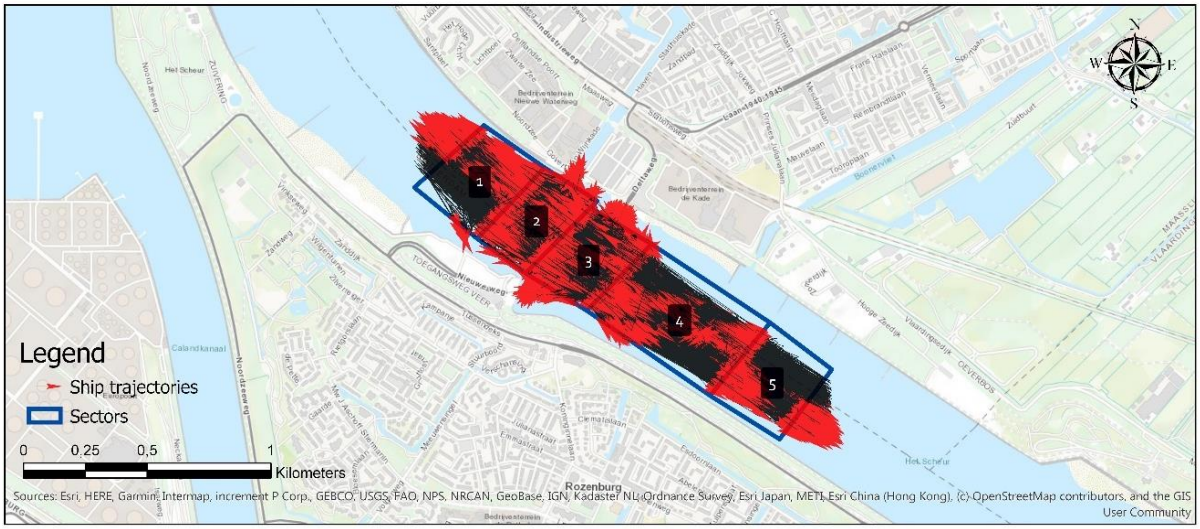


Figure 10. Ship trajectories modelled in the 5 sectors.

To analyze the different variations, the ship directions are displayed with an arrow in the sailing direction. Each arrow represents the trajectory of a ship through a sector, with the arrowhead at the first observed AIS point (within a 5 min time interval) after exiting a sector in the sailing direction. Based on thorough analysis the 5 sectors are subdivided into smaller cells (figure 11). The 5 sectors could be split in half, because most ship movement is east-bound or west-bound. East-bound vessel movement will mainly occur on the south-side of the waterway, and west-bound vessel movement will mainly occur on the north-side of the waterway. Divergent directions vessel most of the vessel move in inside the waterway, are north-bound or south-bound. Either by a passenger ferry, or ship traffic heading to the Maassluis harbor that is on the north-side above the second sector (figure 11, 12).

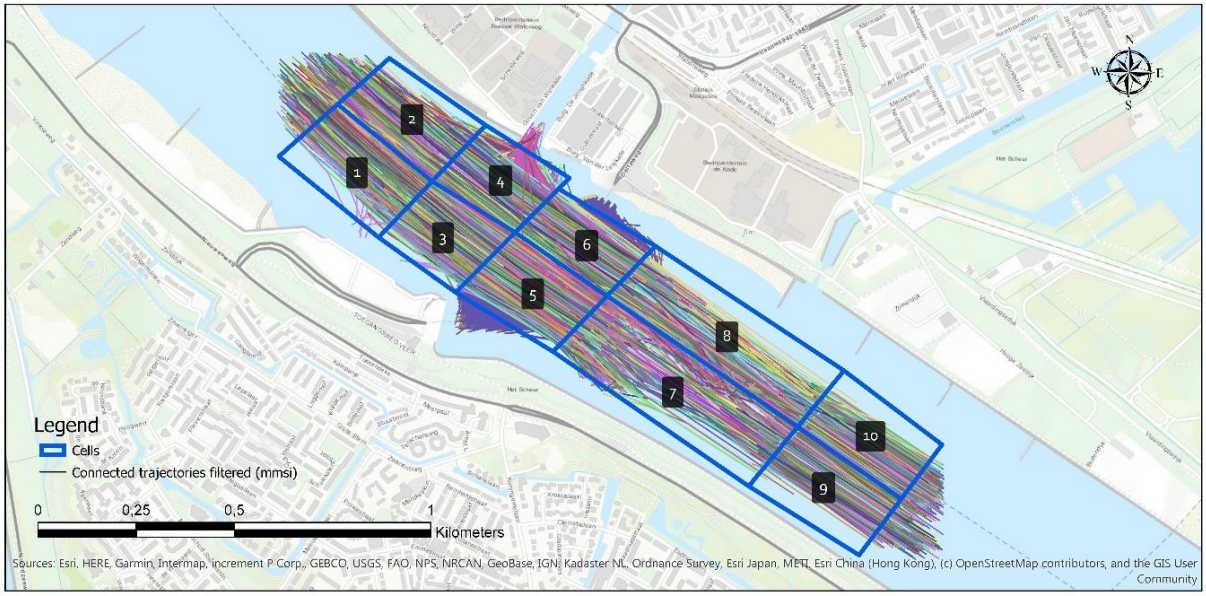


Figure 11. Ship trajectories modelled in 10 different cells.

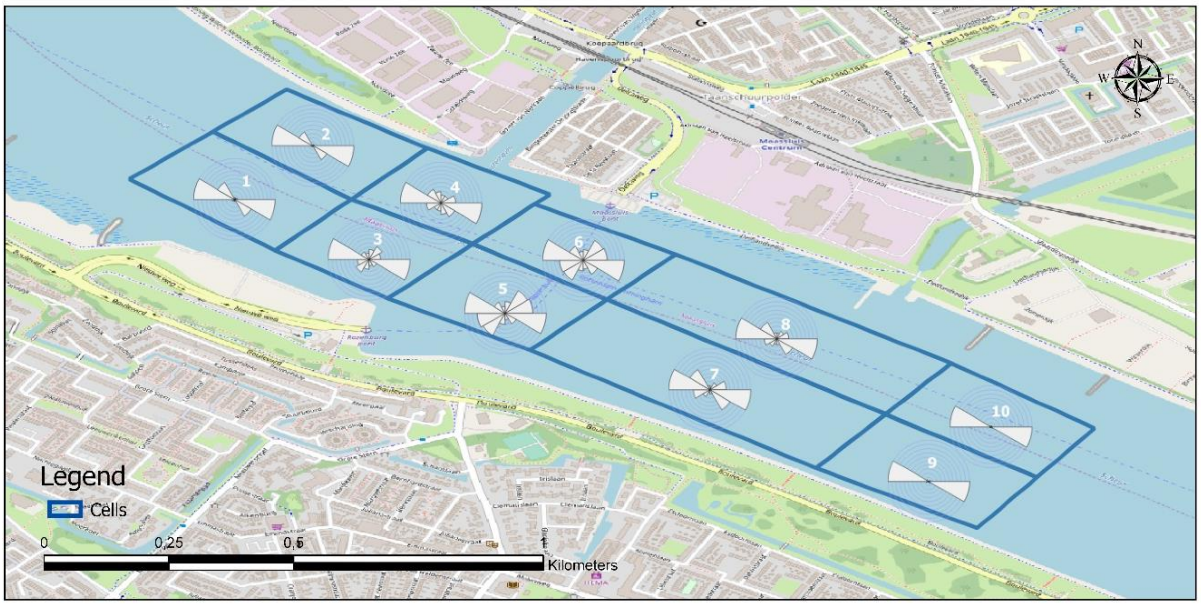


Figure 12. Histograms of line segment directions (logarithmic and area proportional).

4.5 Data model design part 2

Where in paragraph 4.2, the first part of the data model design was discussed, in this paragraph second part will be discussed in regards to the DBMS. The first of two main tables in the database was the table containing AIS data. The second table exist of cell and sector data containing the boundaries. Table 3 contains the data types of the database table. 28 different cell boundaries are stored in the database as geometry features of polylines, this enables spatial analyses with the other geometry attributes of the AIS points. In the second part of table 3 the sectors as polygonal geometry are displayed. These are also needed to join the spatial data to, with that the data can be related to the corresponding cell.

Table 3: Database columns of sectors and sector boundaries.

(Sector/cell boundaries) Name	Data Type
Boundaryid	**Integer**
Boundarygeom	**Geometry**
Shape_length	Double precision
Start_X	Double precision
Start_y	Double precision
Mid_X	Double precision
Mid_Y	Double precision
End_X	Double precision
End_Y	Double precision

(Cells) Name	Data Type
Boundaryid	**Integer**
Boundarygeom	**Geometry**
Shape_length	Double precision
shape_area	Double precision

By generating the two main parts of the research, the typical queries on spatial data can be applied. After which the encounter probability calculations can be performed. Therefore in the remainder of this chapter, the data preparation phase will be discussed after which the different queries will be reviewed.

4.6 Data preparation for the encounter probability

Preparation of the data for calculating the encounter probability exists of various different steps. The spatial application of the explained theoretical and mathematical concepts will be descripted and visualized if possible through either code, maps, charts or graphs. All the different steps of the data preparation phase are made available as different queries via the GitHub page. And all queries that are referenced to in this chapter can be found in appendix 2, also an overview will be given in the next paragraph. First the raw dataset had to be managed and made accessible. The number of AIS records the dataset exists of is 394103 (figure 13).

From the raw AIS messages the different parameters are accessed by applying different database functions and queries as described in 3.2 functions for access of AIS parameters. After accessing possible valuable parameters, not all data in the dataset remains useful. First only message type 1,2 and 3 remain of interest because these messages contain a position report. Were the static and voyage related data from message type 5 is used to enrich the data of message type 1,2 and 3 (Figure 26). After enrichment of the dataset existing of two weeks of AIS messages, the locational data of these messages is used to create points. With the creation of a database table with all the different fields found in table 2 (page 26). The initial set up for data handling is completed.

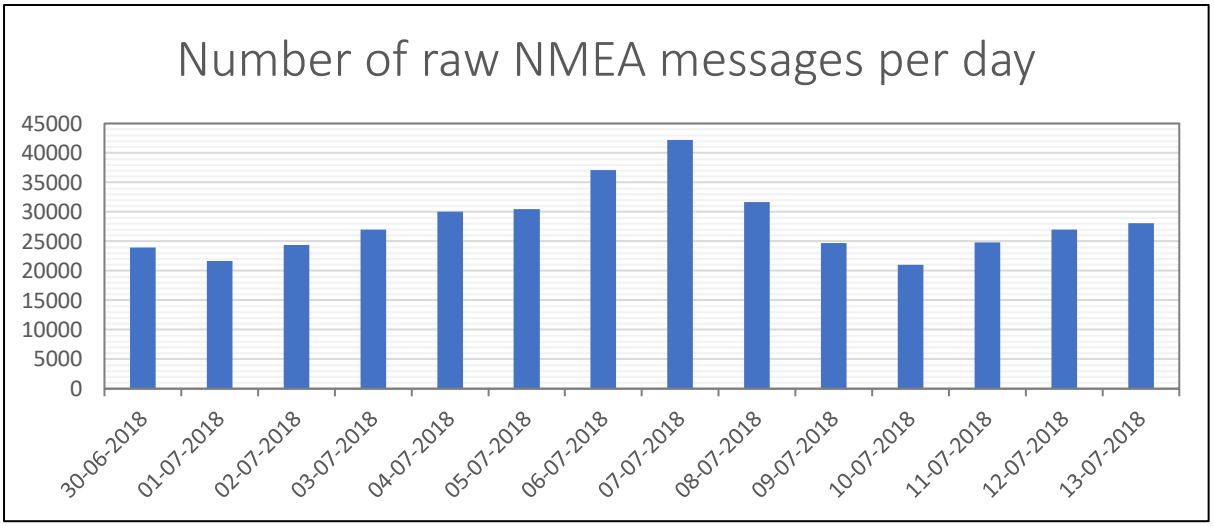


Figure 13. Number of AIS records by day inside the dataset.

4.6.1 AIS points of interest

As described, not all data remains of interest or can be used in further analyses. This means that a selection will to be made. The selection will be made on the points that are close to the cell boundaries. This has several reasons, first a linear crossing of the cell is expected by a vessel, second the points are not constant so the exact point of entering and leaving a cell cannot be selected, but the closest point to the cell boundary can be selected (Figure 27). The selection of the closest points to a cell boundary, is performed for each boundary of the 10 cells. Each cells has 4 boundaries, north, east, south and west boundary. In this research that means this query has been applied on 28 boundaries because many cells have overlapping boundaries and a boundary does not have to be selected multiple times.

4.6.2 Ship trajectories

After selecting points in close proximity of a cell boundary, two points that are within a time interval of 1 min are modelled into a line string representing a track around the cell boundaries (Figure 27). This track represents the vessels trajectory. Furthermore, because the analysis will be done on cell level, only tracks that are crossing the cell boundary remain of interest, these can represent an entry of exit of a cell by a vessel (figure 28). The line passing filter and the trajectory crossings of the cell boundaries are represented in figure 14.

The track crossings can either represent a trajectory that is entering or exiting the cell. In order to derive this data a query is used (figure 29), and a join between the cell boundary and the track is performed. After which the information can be derived whether a vessel is entering or leaving a cell (figure 30). The ship trajectory is not only the line string of exiting or entering a cell, but a combination of both. The two separate line stings of a vessel are connected in order to get the trajectory of the vessel it's movement through a cell (figure 31).

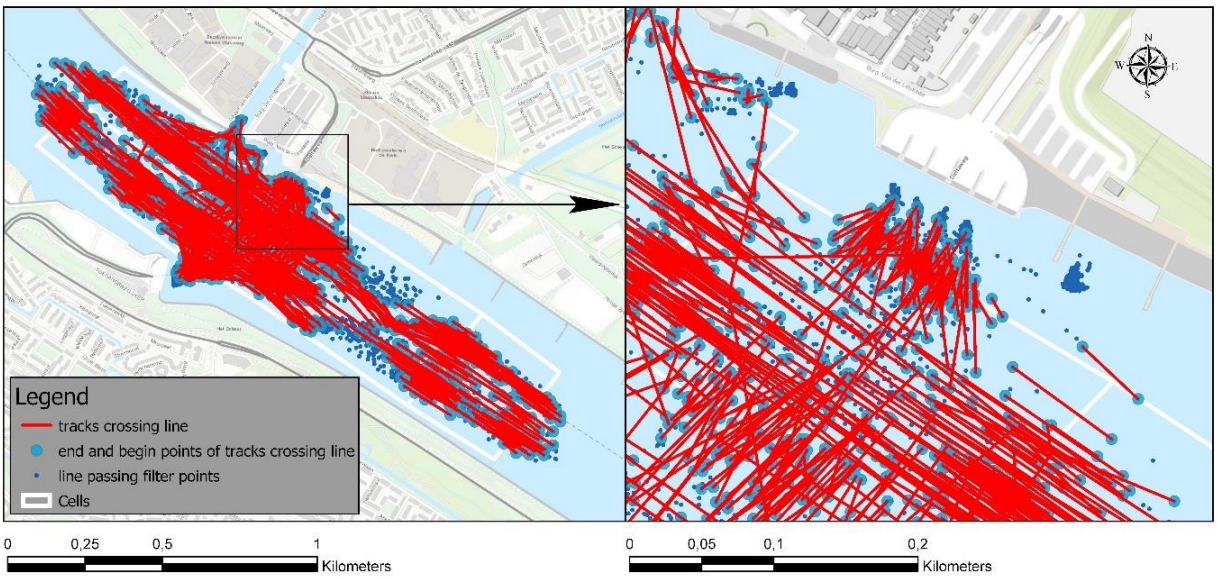


Figure 14. line passing filter and track crossing cell boundaries

4.6.3 Sailing direction

After gathering this information about the trajectories, the direction of sailing can be derived from the trajectories (figure 32). There are 4 sides to each cell: north, east, south and west. And each trajectory consists of two line strings, what makes 12 possible directions, because entering and exiting on the same side are excluded. This would mean a U-turn within the small proximity of a cell, what is very unlikely for a large vessel. Having the vessel trajectories with the underlying data that is enriched. Analysis per cell can now be made. For the purposes of data management and time limitation only 4 directions are considered in this research. These 4 direction are east-bound, west-bound, north-bound and south-bound. That means that the two line strings that are connected into one trajectory are in the same direction. In these 4 directions, related to the cells, the following ship count can be made over the 14 days the dataset exists of (figure 15 and 33).

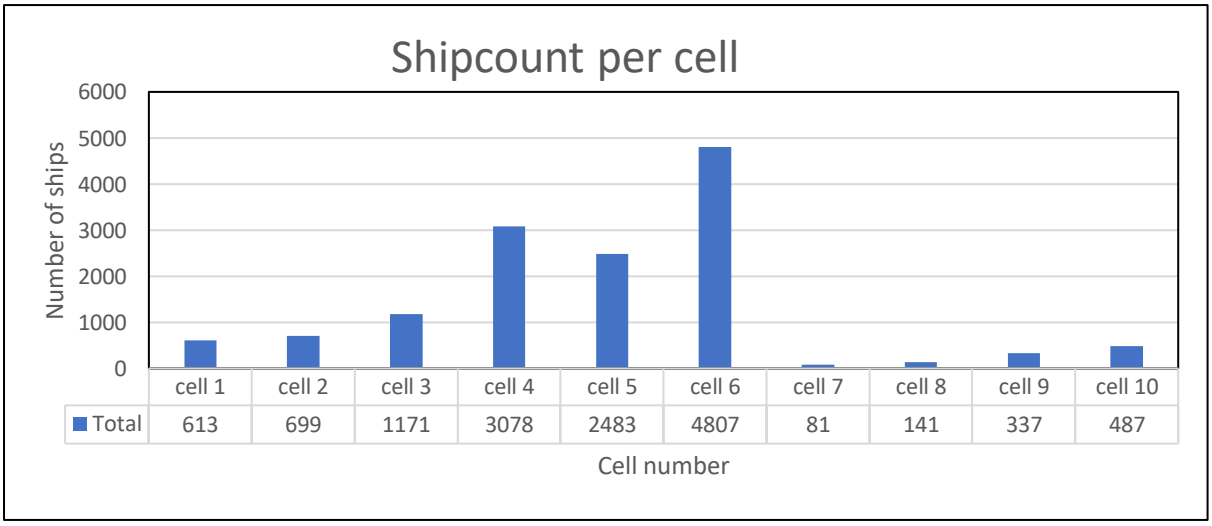


Figure 15. Ship count per cell per direction N,E,S,W.

4.6.4 Spatial analysis

Having enriched data, not only as individual points but also as vessel trajectories, enables various analysis possibilities. The different analysis possibilities are needed to generate the input data for the encounter probability calculations. It can be used to calculate average values of vessels passing through the different cells. Other necessary input values that are needed for the encounter probability calculations are: the average time range spend in a cell (figure 35), the average speed over ground and the average course over ground (figure 36) & the average ship dimensions (figure 37, 38 & 39).

4.7 Database implementation

All different queries that were applied on the database table and referenced to in the previous chapter are given in an overview below, see table 4. The main function of a query is described and the expected return of a query is briefly explained. The queries are listed in the same order as they are implemented, also regarding that many queries need input from the previous ones. For further reference to the queries appendix 1 and 2 can be used.

Table 4: Listing of typical queries

Function name	Main function	Return type
Query for data enrichment of message type 1,2,3.	Select message type	Table with of AIS message type 1,2,3 enriched
Query that defines point that are in close proximity of line strings from cell boundaries.	Buffer and select	Points in close proximity of cell boundaries
Query to connect points within a minute into track around cell boundaries.	Select within time range	Two point connected in a trajectory, based on time range
Query to find trajectories crossing cell boundaries.	Intersect	Trajectories crossing cell boundaries
Join of cell boundaries and track crossing line	Join	Enrich trajectories with the cell of interest
Query for finding whether the start point or the endpoint is inside the cell.	Join	Start and endpoints of trajectories
Query for connecting line strings and applying necessary filters	Join entry and exit trajectories	Connected line strings
Ship direction queries examples	Update and union	Table with directional data
Ship count queries examples	Group	Return the number of ships
Query to calculate time spend - and average speed in a cell	Extract	Time spend & average speed
Query to calculate average SoG in a cell in a certain direction	Calculate average	Averages grouped values of SoG
Query to calculate average CoG in a cell in a certain direction	Calculate average	Averages grouped values of CoG
Queries for calculating LOA and Beam (length and width) averages per cell per direction.	Calculate average	Averages grouped values LOA & Beam

5. Results; spatial application of the encounter probability

The following chapter will be used to explain the application of the equations that are needed to come to a calculation of the encounter probability. The outcome of the calculations will be used as input for the encounter probability risk maps. This chapter is made as a step wise way for calculating an encounter probability. Formulating these calculations as queries that directly can be applied to the database is something that remains of interest, but outside the scope of this thesis.

5.1 Ship density

$$\rho_i = \frac{Q_i}{v_i \Delta t}\frac{\Delta t}{T} = \frac{Q_i}{v_i T}$$

The ship density calculation exists of three things, the ship count per direction per cell, the average velocity per direction per cell and the amount of time from the dataset. The timespan of the dataset is 2 weeks. This has to be recalculated in seconds, because the calculation will be done in seconds, the amount is 1209600. The calculations are performed on sectors that are further subdivided into cells. Subsequently, the calculations are carried out cell-by-cell. Important for the calculation is that the ship density uses the relative ship count per cell, based on the sector. Furthermore in the calculation the average velocity, in meters per seconds was used. This was used in combination with the timespan of the dataset. Therefore the resulting ship density of a cell in a certain direction is a percentage of 1, the default value for a sector (figure 16).

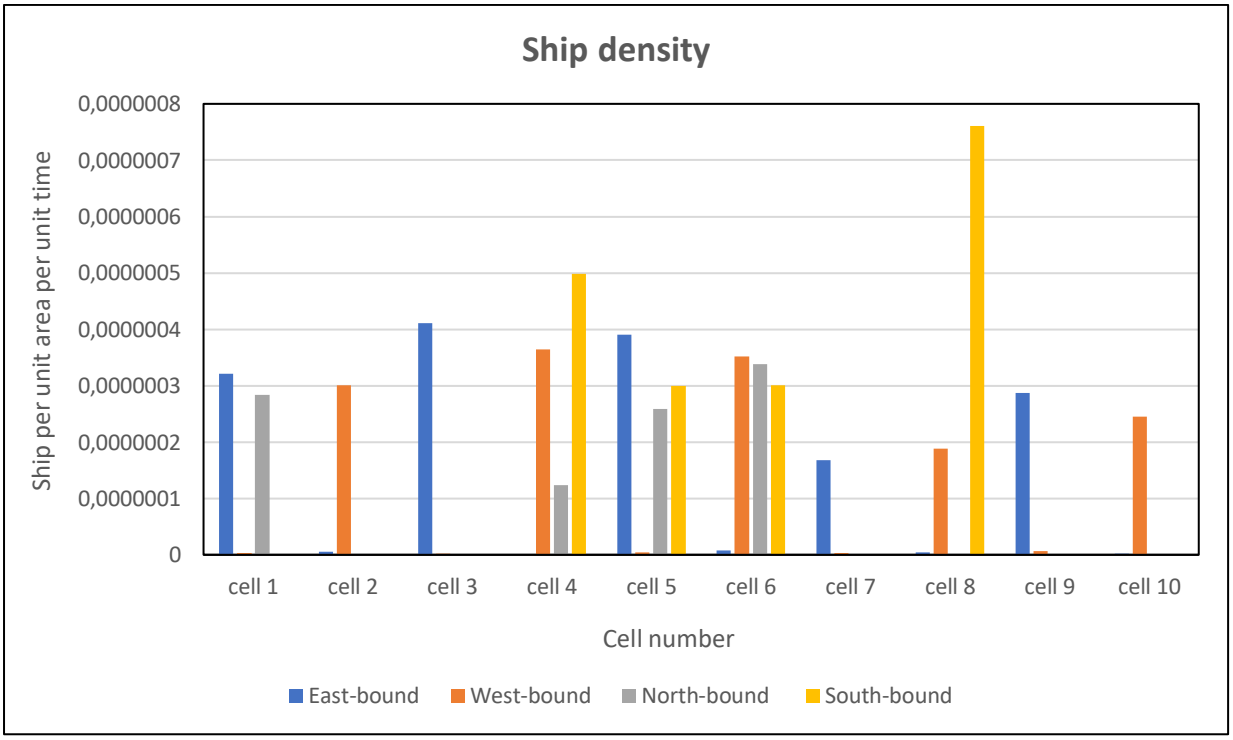


Figure 16. Ship density per cell per direction.

5.2 Relative velocity

$$V_{ij} = |V_i - V_j| = \sqrt{V_i^2 + V_j^2 - 2V_iV_j\cos\theta}$$

The calculation for the relative velocity of ships exists of 7 different steps. Some of those are straight forward and other slightly more complex calculations. The average velocity is needed and also the averages per cell per direction for course over ground.

As discussed earlier, there are 12 different possible directions that are being researched. Because some of the directional combinations give the same values only 6 different options are required for the calculation of the relative velocity. Figure 17 represent the different options where each arrow represents one direction of crossing a cell.

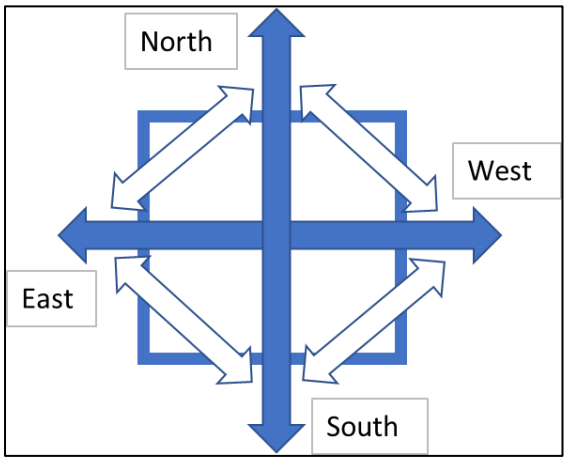
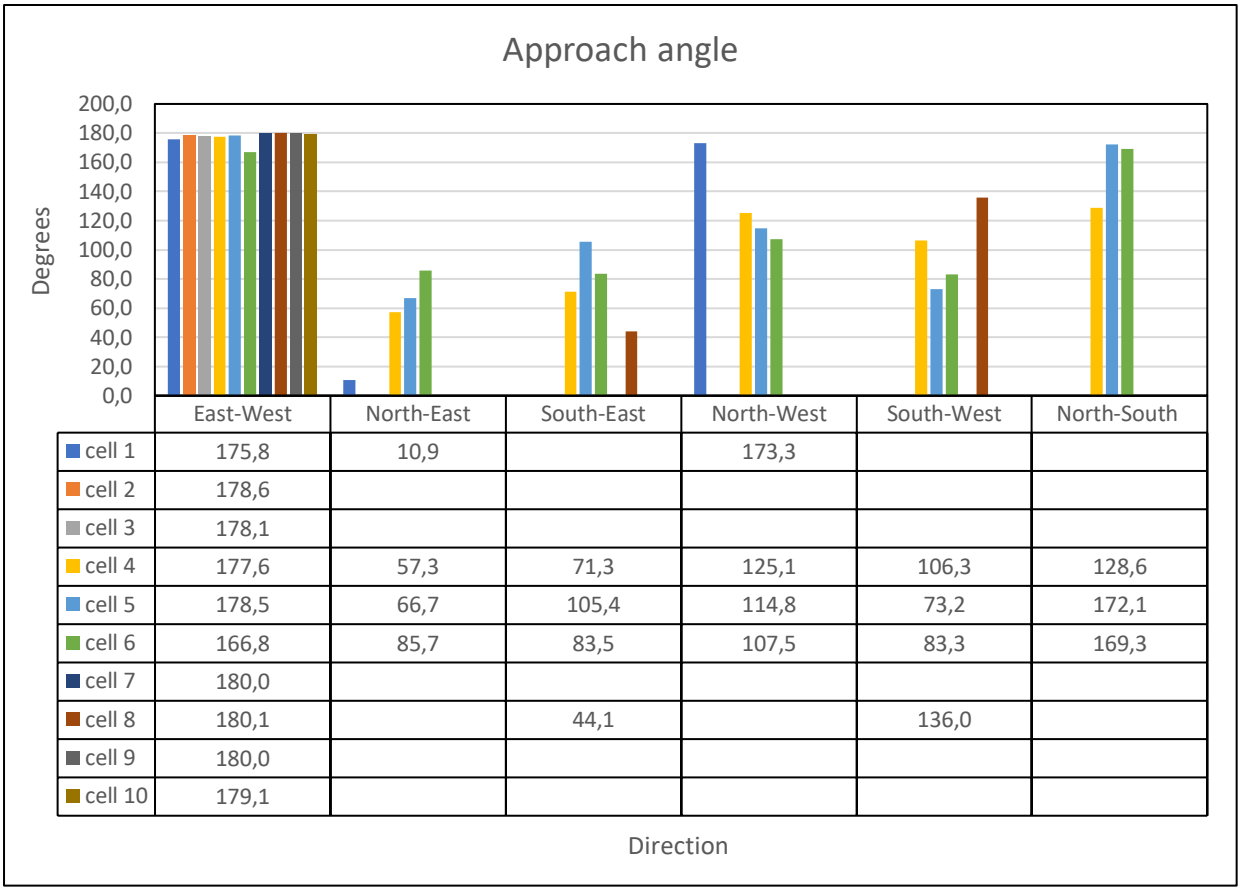


Figure 17. Schematic ship directions crossing a cell.

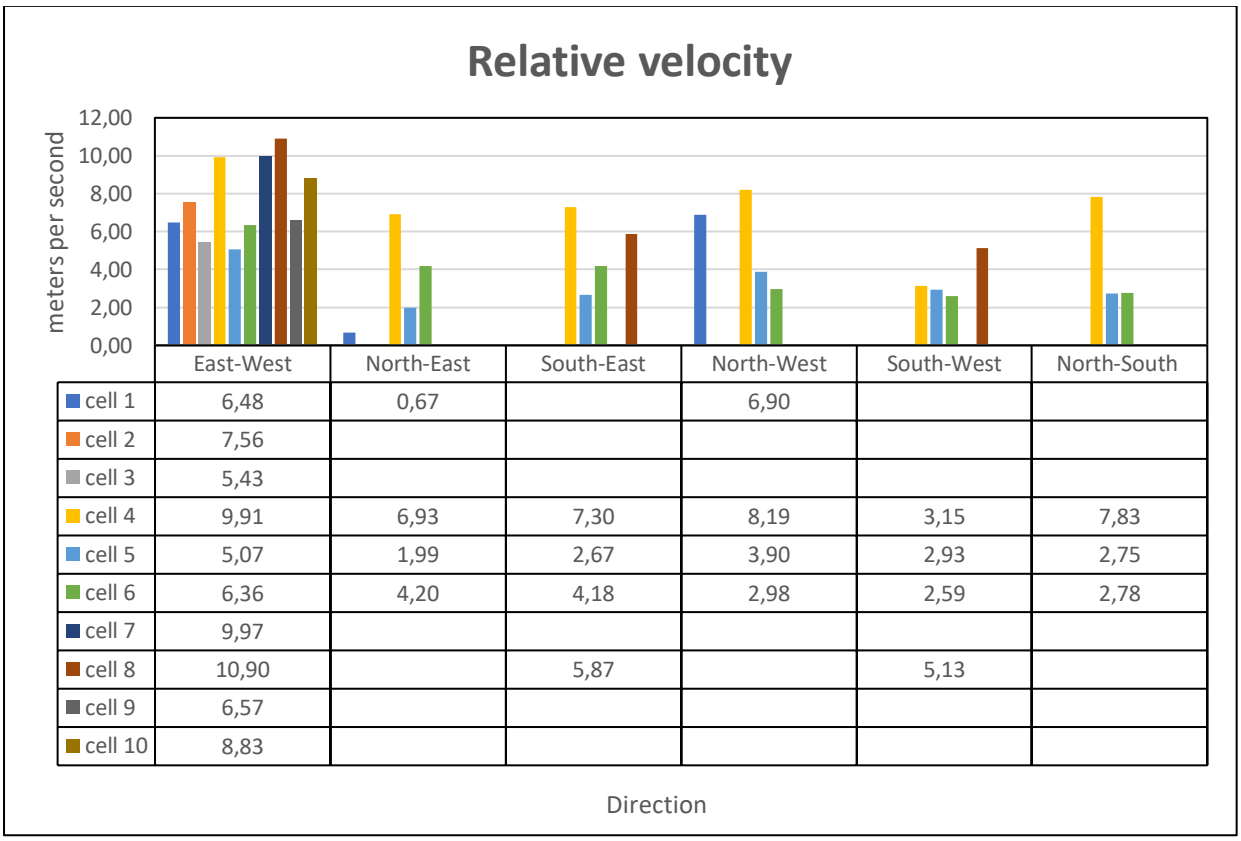
For the equation the meeting angle is required. As explained in the methodology there are 3 types of encounters. A take-over encounter ($\Theta = 0^0 \mp 10^0$), a crossing encounter ($10^0 \leq \Theta \leq 170^0$) and a head-on encounter ($\Theta = 180^0 \mp 10^0$)(figure 18). Executing the formulae gives the relative velocity as result (figure 19).

What becomes clear is that not in each direction ships have been recorded within time period. This means that from such a direction the possibility of an encounter does not exist, and cannot be measured. From the measured directions, some have a limited number of records. At this stage that does not influence the calculation, but for the final calculation of the encounter probability some values have not been considered. This will be explained in more detail in paragraph 5.4.



	East-West	North-East	South-East	North-West	South-West	North-South
cell 1	175,8	10,9		173,3		
cell 2	178,6					
cell 3	178,1					
cell 4	177,6	57,3	71,3	125,1	106,3	128,6
cell 5	178,5	66,7	105,4	114,8	73,2	172,1
cell 6	166,8	85,7	83,5	107,5	83,3	169,3
cell 7	180,0					
cell 8	180,1		44,1		136,0	
cell 9	180,0					
cell 10	179,1					

Figure 18. Approach angle per direction.



	East-West	North-East	South-East	North-West	South-West	North-South
cell 1	6,48	0,67		6,90		
cell 2	7,56					
cell 3	5,43					
cell 4	9,91	6,93	7,30	8,19	3,15	7,83
cell 5	5,07	1,99	2,67	3,90	2,93	2,75
cell 6	6,36	4,20	4,18	2,98	2,59	2,78
cell 7	9,97					
cell 8	10,90		5,87		5,13	
cell 9	6,57					
cell 10	8,83					

Figure 19. Relative velocity per cell per direction.

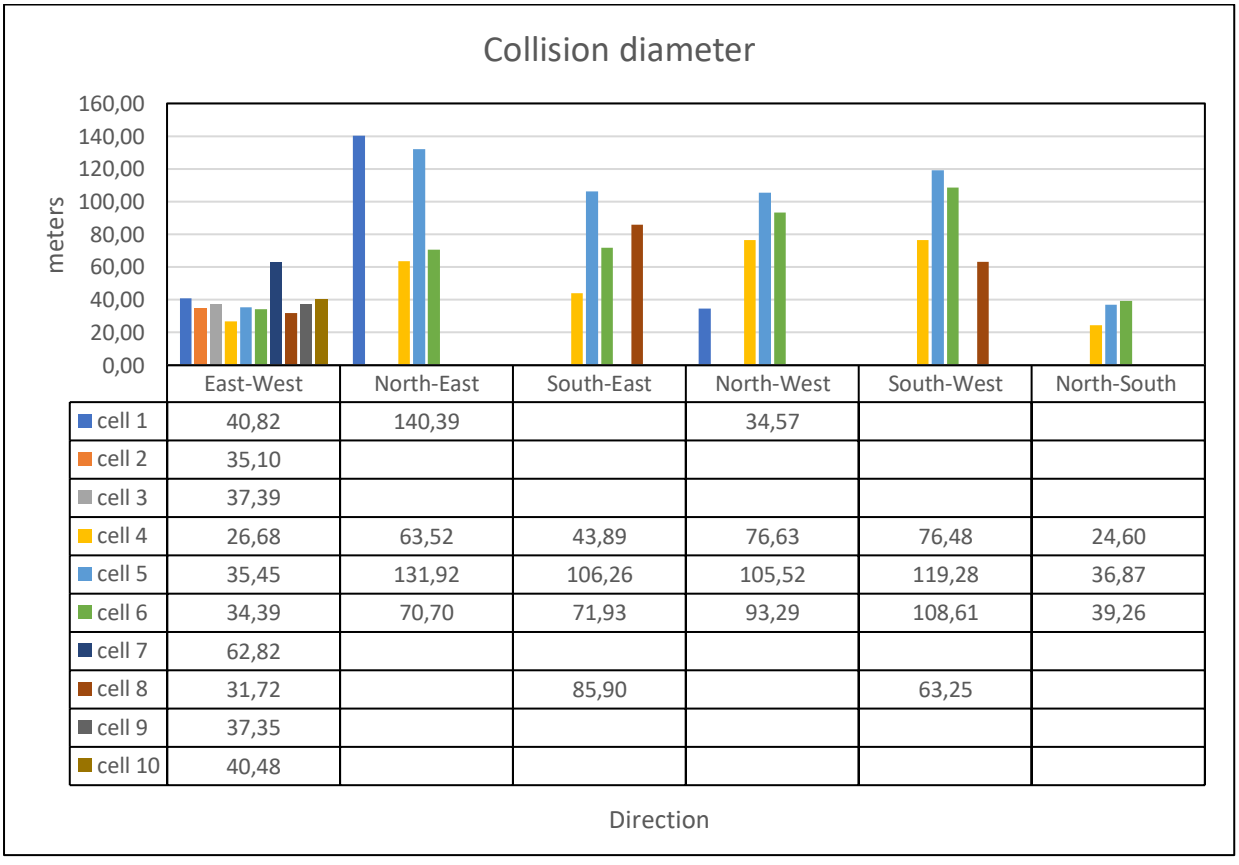
5.3 Collision diameter

$$D_{ij} = \frac{L_i V_j + L_j V_i}{V_{ij}} \sin\theta + B_j \left\{ 1 - \left(\sin\theta \frac{V_i}{V_{ij}} \right)^2 \right\}^{\frac{1}{2}} + B_i \left\{ 1 - \left(\sin\theta \frac{V_j}{V_{ij}} \right)^2 \right\}^{\frac{1}{2}}$$

The collision diameter can be solved by splitting it into 3 different parts.

$$\frac{L_i V_j + L_j V_i}{V_{ij}} \sin\theta \text{ -} \quad \text{part 1 +} \quad + B_j \left\{ 1 - \left(\sin\theta \frac{V_i}{V_{ij}} \right)^2 \right\}^{\frac{1}{2}} + \quad \text{part 2 +} \quad B_i \left\{ 1 - \left(\sin\theta \frac{V_j}{V_{ij}} \right)^2 \right\}^{\frac{1}{2}} \text{ part 3}$$

The average length and width of ships per direction per cell are calculated, and the average velocity per cell per direction will be used. These two values will be combined for part one of the calculation. For part 2 and 3 the same calculation can be used because it is basically the same equation, only in different directions. The 3 different parts of the formulae can be combined and will result in the collision diameter, in meters (figure 20).



	East-West	North-East	South-East	North-West	South-West	North-South
cell 1	40,82	140,39		34,57		
cell 2	35,10					
cell 3	37,39					
cell 4	26,68	63,52	43,89	76,63	76,48	24,60
cell 5	35,45	131,92	106,26	105,52	119,28	36,87
cell 6	34,39	70,70	71,93	93,29	108,61	39,26
cell 7	62,82					
cell 8	31,72		85,90		63,25	
cell 9	37,35					
cell 10	40,48					

Figure 20. Collision diameter per cell per direction in meters.

As described in the previous paragraph some cells have a limited number of ships passing in a certain direction. This caused not well balanced values. For example in cell 1, in North-East direction the collision diameter is 140 meters. This is an unrealistically high value, and would

negatively influence the further calculation of the encounter probability. When using an larger dataset over a longer period of time can prevent such measures.

5.4 Encounter probability

In the encounter probability calculation, the outcomes of the previous calculations will be used as input parameters alongside other parameters.

$$P_E = \sum_s \sum_i \sum_j \sum_{ik(s)} \sum_{jk(s)} \rho_i(ik(s), s)\, \rho_j(jk(s), s)\, V_{ij}(s)\, D_{ij}(s)$$
$$\Delta l(ik, s) \Delta l(jk, s)\, T$$

In order to calculate the encounter probability at a given cell, the joint probability of the observed encounters has to be calculated. Since it is unknown if the encounters are independent events, it is impossible to calculate the joint probability of encounters directly. Therefore, the non-encounter probability for each encounter type can be calculated, by doing this it is possible to calculate the joint probability of non-encounters by just multiplying them.

For example, we have calculated the northbound-eastbound encounter probability as 0.2. If it will be subtracted it from 1, the non-encounter probability for the northbound-eastbound encounter will be 0.8. If we do the same calculation for each encounter type, the non-encounter probability for each encounter type is found. Since non-encounters are independent events, the joint probability is the multiplication of these non-encounter probabilities. The end result of this multiplication gives the non-encounter probability for the given cell. Next the actual encounter probability can be calculated for the creation of a risk map. The encounter probability is the non-encounter probability subtracted from 1.

This results in the encounter probability that can be observed for the given cell. This operation has to be done for each cell to create the risk map. This results in the encounter probability of each cell in the Nieuwe Waterweg (figure 21). This can also be plotted in a map, displaying this encounter probability in a risk map (figure 22).

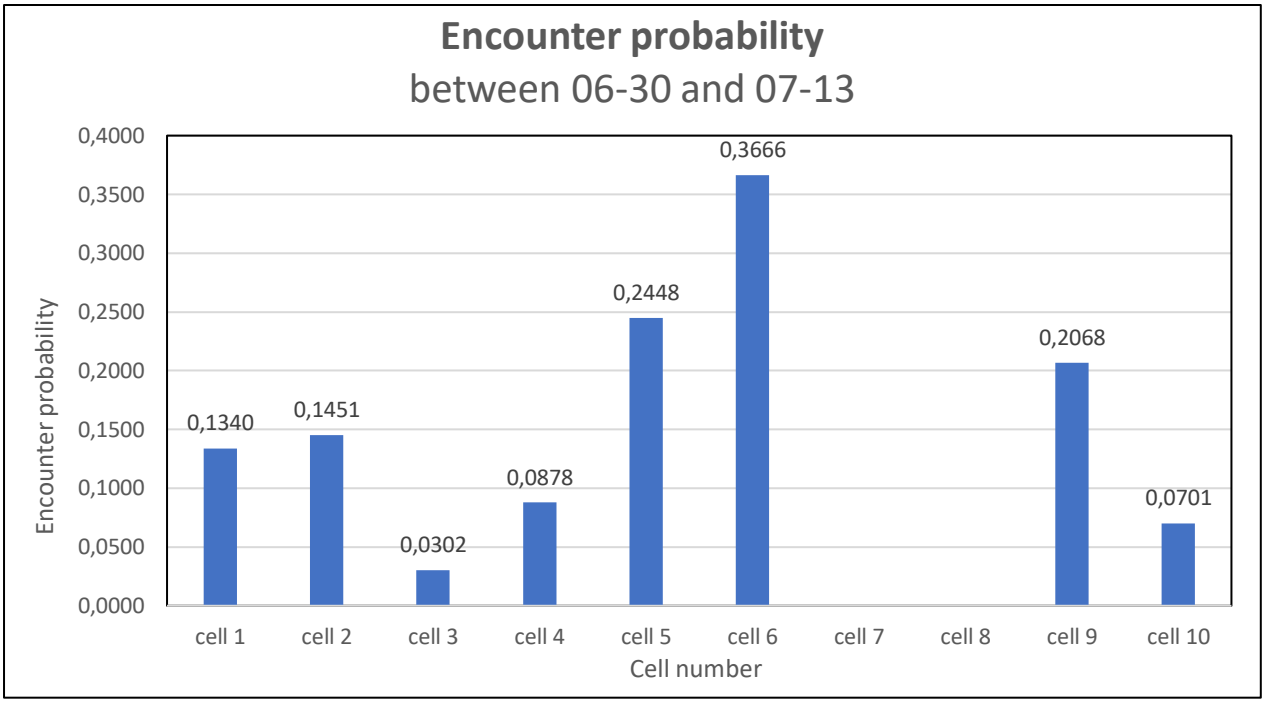


Figure 21. Encounter probability.

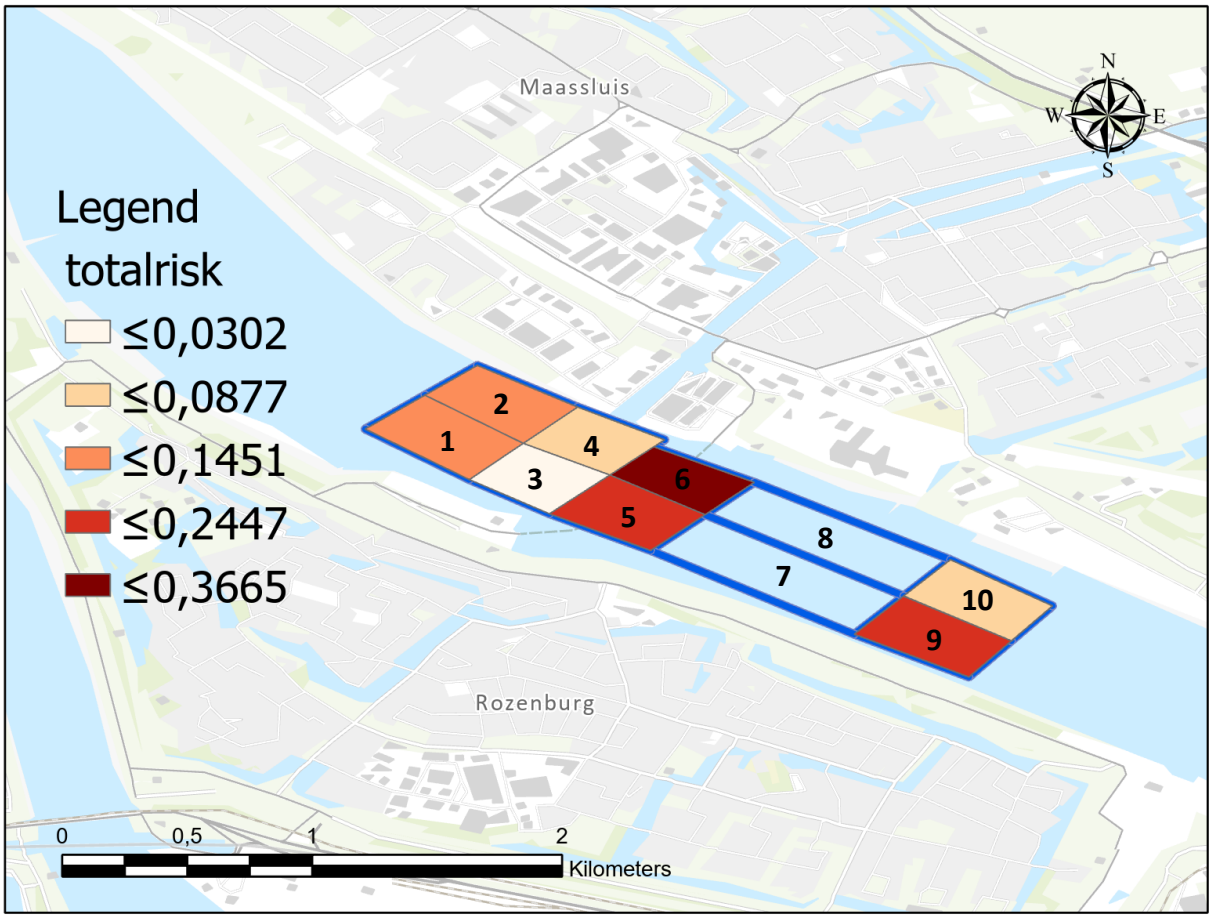


Figure 22. Encounter probability risk map.

The outcome of the encounter probability calculation give an good image of what sectors represents the highest risk of an encounter. Obviously some cells were not shown in the results. This choice was made because of two reasons, the first is that some cells had non-reliable outcomes. In that case they have been given a non-encounter value of 1. This is deemed necessary to generate more reliable results. When working with a larger dataset, with records over a longer period of time this is not necessary. The second reason is that the data gathered around cell 7 and cell 8 did not generate optimal results. Designing the sector and cells happened in an early stadium, and should be an iterative process. In this case the cell length and the location of the sector division caused a lot of trajectories to be generated wrong. By making an iterative process of finding manoeuvring points that help in the generation of a sector, the results can be enhanced. Not only two cells per sector have to be made. But many more are possible as long as the cell remains larger than two times half the collision diameter.

5.5 Database performance evaluation

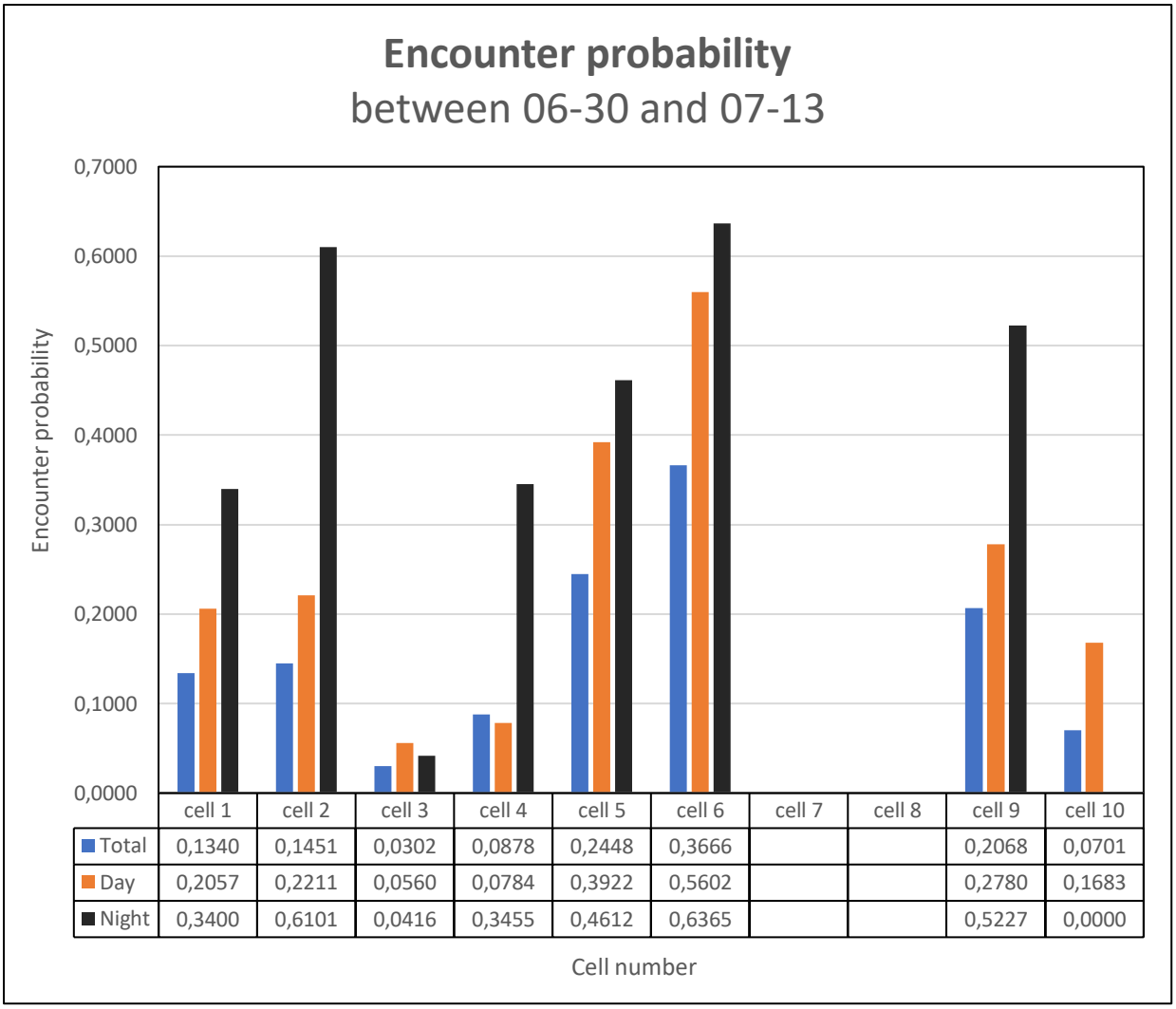
Working with an advanced DBMS with a clear goal has proven that the performance possibilities of the PostgreSQL did not form any limitations in concern with the queries designed. Some larger datasets have been tested outside the scope of calculating the encounter probability. With those data sets containing millions of raw AIS messages, the queries still performed within reasonable time of 15 min maximum. Limitations were met in database storage possibilities. In this research the storage from the database was done locally, and for large datasets of AIS data, than other solutions than local storage should be used. Examples are usage of a specialized data center or cloud storage.

5.6 Risk mapping

In the previous paragraphs the different steps to execute an encounter probability calculation are explained. In this research three different risk maps have been produced. An encounter probability risk map for the whole dataset (figure 21), but also a risk map for all vessel activities during daylight (figure 23) and one for all vessel activities during nighttime (figure 24).

The encounter probability calculations are done for a period of time. The choice has been made to also make a daytime (>05:30)(Figure 24) and a nighttime (>22:00)(Figure 25) calculation. Interesting is that during nighttime the probability is a highest (Figure 23). A possible explanation for this is that the ship density during night time is higher than during the day. Also different sectors carry an higher risk. It is difficult on the two week dataset to make significant conclusions. One of the reasons is the limited amount of data available.

Another remarkable outcome when comparing the day, night and total encounter probability is that in most cells the day and night probability are higher than considering the total period. It would seem logic if the risk calculation of the whole dataset would be the combination of both risks, and get a value in between. Although it seems logic, the outcome does not correspond with the logic. Unfortunately, the reason is yet unknown and invites for further research. To conclude, this research can serve as a good test case how to implement and execute different encounter probability calculation, towards creating one's own risk maps.



	cell 1	cell 2	cell 3	cell 4	cell 5	cell 6	cell 7	cell 8	cell 9	cell 10
Total	0,1340	0,1451	0,0302	0,0878	0,2448	0,3666			0,2068	0,0701
Day	0,2057	0,2211	0,0560	0,0784	0,3922	0,5602			0,2780	0,1683
Night	0,3400	0,6101	0,0416	0,3455	0,4612	0,6365			0,5227	0,0000

Figure 23. Encounter probability histogram.

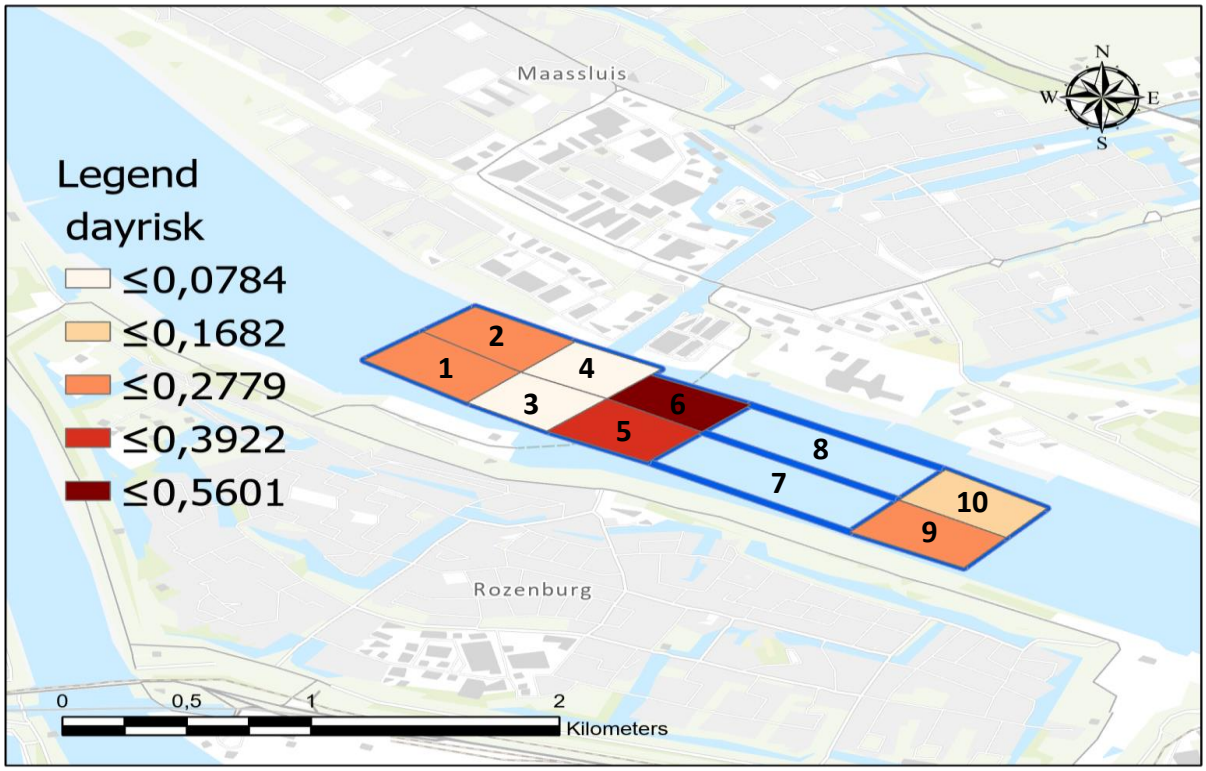


Figure 24. Encounter probability risk map during day time.

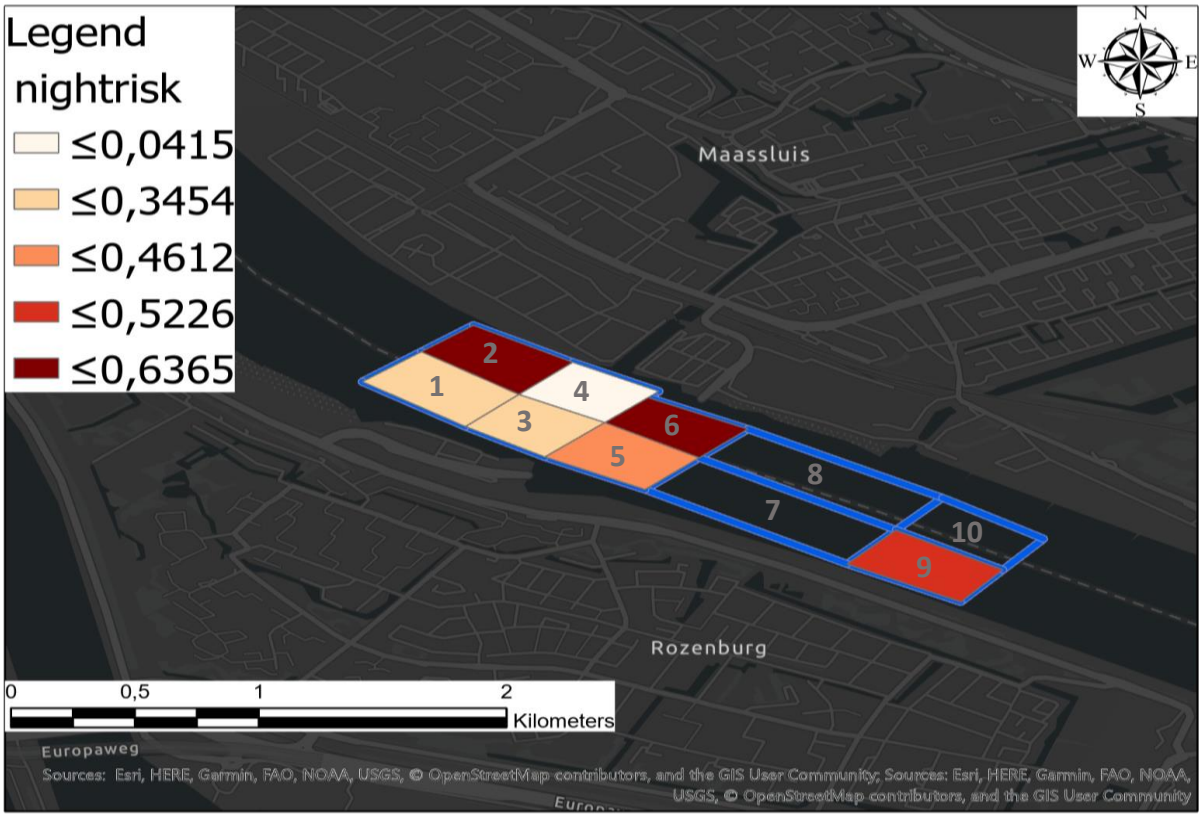


Figure 25. Encounter probability risk map during night time.

6. Conclusion & Discussion

Context aware analysis of encounter probability of inland maritime traffic is of interest in various applications. However complex the implementation of mathematical concepts to spatio-temporal issues is, uniform programmable analysis methods can shed light on such issues. This thesis contributes to fill this gap by presenting and evaluating a method for efficient AIS data management combined with waterway characteristics for the development of an encounter probability risk map.

6.1 Revisiting research questions

The main objective of this thesis is to develop a method to analyse and manage historic AIS data efficiently, and combine this with waterway characteristics to find manoeuvring points that help with the creation of an encounter probability risk map. This was formulated with the following research question: “In what way can Automatic Identification System (AIS) data be managed efficiently in combination with waterway characteristics in order to find manoeuvring points that can be used for the development of an encounter probability risk map?” To answer the research question, first the different sub questions can be answered.

“What different methods are available for storing AIS data efficiently?” this sub question was split into three questions: “What is AIS data? What AIS data is, is clear in this stadium. The abbreviation means Automatic Identification System (AIS) data. A system fitted on ships for identification and navigational marks. The main purpose it was invented for is an aid to navigation. “How to decode AIS data?” The NMEA messages that are 6-bit ACSII encoded can be decoded in different ways. Four different methods of decoding have been tried (2B JSON encoded, 2C Using a python library, 2D as 6-bit ASCII encoded & 2E as bit vector). From these four messages the most promising seemed bit vector. Furthermore, the decoding of AIS data means that functions that have been made for this research can be used to access the different parameters such messages contain.

“And how to store AIS data efficiently?” This sub question can be answered in two parts. Stored as an efficient data type was achieved by choosing bit vector. The storage itself has to be done in an efficient system. For this study it seems to be necessary to use a form of object-relational database. Because of the complexity of the spatial queries that are needed to visualize the data, the solution is a PostgreSQL database, within pgAdmin4, combined with a PostGIS extension. Different types of efficient databases exist but they were not deemed necessary for this research. Examples are MYSQL or MongoDB.

The next research question continues on the previous research question. “How to use a geodatabase management system (Geo-DBMS) for storing and managing historic AIS data?” It is clear that for efficient storage a Geo-DBMS had to be chosen. “What Geo-DBMS is efficient in storing historic AIS data?” For this specific research that is a PostgreSQL database. Though various options exist. Because of the spatio-temporal component of the research, storage of the data alone was not sufficient.

In the database the necessity exist to be able perform spatial analysis and typical queries on spatial data. This has been formulated with the following sub question: “How to preform spatial analysis on data stored in a Geo-DBMS, and how does that influence the choice for a

certain Geo-DBMS?" To be able to perform spatial analyses on data stored in a Geo-DBMS, the different components of the AIS messages had to be accessed. Therefore, different database functions were, and are made available for AIS messages stored as bit vector. This means that the first function that has been made is a function to encode the raw NMEA strings as bit vector, from a varchar data type. After that an extended set of database functions for access to the different AIS parameters has been defined.

For the research not only historic AIS data is relevant. Waterway characteristics are important information, that in combination with the AIS data can result in a desired risk map. This resulted in the following sub question: "Where can manoeuvring points in the waterway be observed?" This means that locations within a waterway were vessels can move skillfully or carefully through that waterway, by making sailing decisions. This resulted in three sub questions of which the first is: "What are the waterway characteristics of a certain waterway that will be used as case study?" To answer this question thorough waterway analysis has been performed. Navigational maps and governmental officials have been consulted.

The next sub question that results from this are: "How can the waterway be divided into different sectors where changes are observed in navigational characteristics?" and "How to decide ship traffic distribution along a waterway?" Interactive maps have been made to see the actual movement of vessels through a waterway. This has been done by plotting historic AIS data on a map dynamically based on its timestamps and identification number. And sectors and cells could therefore by designed in areas of assumed minimal changes in navigational variation. A remark can be made that the design of such sectors can be done in an iterative way or either in multiple designs, for interesting results. This specifically becomes visible after performing encounter probability calculations and creating the first risk maps.

For a more analytical part of the research the next sub question is made: "How can the historic AIS data be used for analyzing ship traffic?" In various different ways, but in this research it has been used for a specific purpose of finding an encounter probability. This could be achieved by applying the molecular collision theory on a spatial distribution. Because whenever two vessel are expected to be within a predefined proximity of the same geographic location at the same time, than an encounter will occur.

The last sub question that is a part of the answer to the research question is: "In what way can the data be combined and visualized to create an encounter probability risk map?" In the simple answer; the data can be combined via typical queries and functions in a Geo-DBMS and visualized within GIS software. A combination of the answer to these different sub questions enables an answer to the main research question in a comprehensive manner.

"In what way can Automatic Identification System (AIS) data be managed efficiently in combination with waterway characteristics in order to find manoeuvring points that can be used for the development of an encounter probability risk map?" The AIS data can be managed efficiently in combination with waterway characteristics through a Geo-DBMS such as PostgreSQL with a PostGIS extension. The different parameters of the AIS data can be accessed through typical queries and database functions. The manoeuvring points can be found by combining the historic AIS data and the waterway characteristics interactively and

through spatial analysis. This can be used for the development of an encounter probability risk map by using the proposed methodology.

6.2 Limitations and reflection

During the research it became clear that AIS data does have its limitations. AIS contains various types of errors. These can be filtered out before analyzing the data. An expected lack of the model was that it does not work for predictions during approaching and departing maneuvers. This has not been proven. After accessing possible valuable parameters, not all data in the dataset remained useful. From accessing raw AIS messages to the creation of an encounter probability risk map. Enriching historic AIS messages enabled different analysis possibilities not only regarding the encounter probability calculations for making risk maps, but various possibilities.

The practicability and the applicability of the proposed approach and its integrated management of data was proved in a case study with historic AIS data. Though remarks can be made for dividing a research area into a grids. This remains an activity where assumptions about main traffic routes have to be made. It might be interesting to develop an method where a spatial distribution can be made that is not dependent upon input from the researcher, but fully based on data or at least a more iterative process. The grid sizes can differ from the once's that have been used in this thesis, as long as they remain larger than the collision diameter, to be sure no artificial collisions occur.

Working with an advanced DBMS did not form any limitations in concern with the queries designed. Some larger datasets also were tested outside the scope of this research. Though it proved the queries still performed within reasonable time limits. Some database were met in storage. In this research the storage from the database was done locally, and for large datasets of AIS data, other solutions should be used.

The amount of data that ensures a certain threshold, for producing reliable efforts, is something that requires some effort. At this moment no guidelines are written. But it could be an enhancement to the research. In this research 4 directions of sailing have been used as input. Whereas more directional possibilities exist, it might have interesting consequences on the encounter probability. Though the amount of work would also increase exponentially with including more directions.

6.3 Recommendations

For this research a linear passage was assumed through a sector, that sectors are designed in square forms. As AIS data is a precise dataset perhaps a different shape of the sectors can be considered, for example hexagonal. That could result in a more precise spatial distribution, and therefore might benefit the outcomes of an encounter probability calculation. The theory only considers one encounter at a time, since ship encounters are rare events. But the encounter of more than two ships at one moment, might result in high risk situations. Therefore, it might be interesting to design a method that also includes more than two ships at one time. Based on the proximity of ships in certain moments of time.

The case study has only shown a brief inside in the extensive possibilities of analysing inland maritime traffic. Furthermore, comparing day and night inland maritime traffic is not the only possibility of comparing different moments. Analysis can be made for various moments, examples are comparisons in different day of the week, different months of even different years. Analysis can also be extended for various circumstances in weather, current, season, ship types or even a risk comparison between different waterways. Doing different spatial analysis on large datasets of historic AIS data invite for future research. Also finetuning and automating more of the detailed and explained processes can be a good starting point for future research.

References

Aarsæther K. G., & Moan T. (2009). Estimating Navigation Patterns from AIS. The journal of navigation, 62, pp. 587-607.

Altan, Y. C., & Otay, E. N. (2017). Maritime Traffic Analysis of the Strait of Istanbul based on AIS data. Journal of Navigation, 70(6), pp. 1367–1382. Doi.org/10.1017/S0373463317000431.

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

Chen, P., Huang, Y., Mou, J. & van Gelder, P. H. A. J. M. (2018). Ship collision candidate detection method: A velocity obstacle approach. Ocean engineering 170 (2018), pp. 186-198.

Debnath, A. K., & Chin, H. C. (2016). Modelling Collision Potentials in Port Anchorages: Application of the Navigational Traffic Conflict Technique (NTCT). Journal of Navigation, 69(1), pp. 183–196. Doi.org/10.1017/S0373463315000521.

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

Fowler, T.G., Sørgård, E., (2000). Modeling ship transportation risk. Risk Analysis. 20, pp. 225–244.

Fuiji, Y,. Yamanouchi, H. and Mizuki N. (1974). Some factors affecting the frequency of accidents in marine traffic. Journal of Navigation 1974, 27, pp. 239-252.

Goerlandt F., & Kujala P. (2014). On the reliability and validity of ship-ship collision risk analysis in light of different perspectives on risk. Safety Science, 62, pp. 348-365.

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

Graser A. (2018). Evaluating Spatio-temporal Data Models for Trajectories in PostGIS Databases. GI_Forum 2018, 1, pp. 16-33.

Hansen, M. G., Jensen, T. K., Lehn-Schoiler, T., Melchild, K., Rasmussen, F. M. & Ennemark, F. (2013). Empirical ship domain based on AIS data. The journal of navigation (2013), 66, pp. 931-940.

Heij, C., Bijwaard, G.E., Knapp, S., 2011. Ship inspection strategies: effects on maritime safety and environmental protection. Transp. Res. Part D 16, pp. 42–48.

Hexeberg, S. (2017). AIS-based Vessel Trajectory Prediction for ASV Collision Avoidance.

IMO (2002). Guidelines for Formal Safety Assessment (FSA) for use in the IMO rule-making process. International Maritime Organization, London.

Klein, J., Gorton, I., Ernst, N., Donohoe, P., Pham, K., and Matser, C. (2015), Performance evaluation of nosql databases: A case study. In Proceedings of the 1st Workshop on Performance Analysis of Big Data Systems, pages 5– 10.

Koedijk, O.C., A. van der Sluijs & M.L.W. Steijn (2017), Richtlijnen Vaarwegen 2017, Kader verkeerskundig vaarwegontwerp Rijkswaterstaat. Rijkswaterstaat. Retrieved at 25-06-2019 from https://staticresources.rijkswaterstaat.nl/binaries/richtlijnen-vaarwegen-2017_tcm21-127359.pdf.

Kuwata, Y., Wold, M. T., Zarzhitsky, D & Huntsberger, T. L (2011). Safe maritime navigation with COLREGs using velocity obstacles. California Institute of Technology. 4800 Oak Grove Drive, Pasadena, CA, USA.

Macduff, T., (1974). Probability of Vessel Collisions. Ocean Industry, 1974. 9(9) pp. 144-148.

Meijers, M., Van Oosterom, P. & Quak, W. (2017). Management of AIS messages in a Geo DBMS, GISt Report No. 71.

Montewka, J., Goerlandt, F., & Kujala, P. (2012). Determination of collision criteria and causation factors appropriate to a model for estimating the probability of maritime accidents. Ocean Engineering, 40, pp. 50–61. Doi.org/10.1016/j.oceaneng.2011.12.006.

Pedersen, P. T., (1995). Collision and grounding mechanics. Proceedings of WEMT 95, pp. 125-157.

Raymond, E. S. (2016). AIVDM/AIVDO protocol decoding. Retrieved at 01-02-2019 from https://gpsd.gitlab.io/gpsd/AIVDM.html.

Review of Maritime Transport (2017). Developments in international seaborne trade. United Nations Conference on Trade and Development. Retrieved at 01-02-2019 from Unctad.org/en/pages/publicationwebflyer.aspx?publicationid=1890.

Shu, Y., Daamen, W., Ligteringen, H., Hoogendoorn, S. (2013), AIS data analysis for vessel behavior during strong currents and during encounters in the Botlek area in the Port of Rotterdam. International Workshop on Next Generation Nautical Traffic Models 2013, Delft, The Netherlands.

Silveira, P., Texeira, A. P. & Guedes Soares, C. (2014). Assesment of ship collision estimation methods using AIS data. Centre for Marine Technology and Engineering (CENTEC), Insituto Superior Tecnico, Universidade de Lisboa, Portugal.

Ventura, M. (2009). COLREGS -International Regulations for Preventing Collisions at Sea COLREGS -International Regulations for Preventing Collisions at Sea COLREGS -International Regulations for Preventing Collisions at Sea -Articles of the Convention on the International Re, pp. 1–74.

Vreede, I. De. (2016). Managing Historic Automatic Identification System data by using a proper Database Management System Structure.

Xiao, F., Ligteringen, H., van Gulijk, C., & Ale, B. (2012). AIS data analysis for realistic ship traffic simulation model. International Workshop on Nautical Traffic Models, (September), pp. 44–49.

Xiao, F., Ligteringen, H., Van Gulijk, C., & Ale, B. (2015). Comparison study on AIS data of ship traffic behavior. Ocean Engineering, 95, pp. 84–93. Doi.org/10.1016/j.oceaneng.2014.11.020.

Zhang, W., Goerlandt, F., Montewka, J. & Kujala P. (2015). A method for detecting possible near miss ship collisions form AIS data. Ocean Engineering 107 (2015), pp. 60-69.

Appendix 1: Git

This repository created for thesis research, is made to keep track of changes made in code over time. Git will be used as a version control system, to keep track of my changes to the code over time (and view past versions), keep track of multiple different versions of the code, and it helps to enable others like my thesis supervisors to stay synced up.

It helps with not having to do manual copying an pasting code, and then emailing snippets of my code. Other version control systems beside Git exist. (Mercurial and Subversion are two other commonly used ones). Keeping track of old versions of code is valuable for debugging, and understanding how to work with a version control system.

The idea behind the functions are that they only need to be written once and can be reused multiple times. This will save me time and effort and support modular programming. The functions improve performance and efficiency of the database. Furthermore, complex programming logic can be decomposed into a number of smaller and simpler functions.

https://github.com/ThomasLier/AIS_Thesis_Research, access can be requested by emailing t.b.lier@students.uu.nl.

Appendix 2: Queries and database functions

The queries and functions displayed below are only a small selection of that are directly referenced to in the thesis, other queries and database functions that were made are found via the GitHub link (Appendix 1).

```
-- update query enrich table type 1,2,3 with 5

SELECT * INTO AIS0630_0713_messagetype123 FROM  "AIS0630_0713"
WHERE messagetype = (0)
OR messagetype = (1)
OR messagetype = (2)
OR messagetype = (3);

SELECT * INTO AIS0630_0713_messagetype5 FROM "AIS0630_0713"
WHERE messagetype = (5);

-- Update table colums with values from other table columns with a where clause (Join function without actual join)

UPDATE "ais0630_0713_messagetype123"
SET dimensiontostarboard = ais0630_0713_messagetype5.dimensiontostarboard,
    dimensiontoport = ais0630_0713_messagetype5.dimensiontoport,
    dimensiontostern = ais0630_0713_messagetype5.dimensiontostern,
    dimensiontobow = ais0630_0713_messagetype5.dimensiontobow,
    shiptype = ais0630_0713_messagetype5.shiptype,
    draught = ais0630_0713_messagetype5.draught
FROM "ais0630_0713_messagetype5"
WHERE ais0630_0713_messagetype123.mmsi = ais0630_0713_messagetype5.mmsi;
```

Figure 26. Query for data enrichment of message type 1,2,3.

```
-- Query that defines point that are in close proximity of line strings from cell boundaries

DROP TABLE IF EXISTS ais0630_0713_line_passing_filter;
CREATE TABLE
        "ais0630_0713_line_passing_filter"
AS
SELECT
        ais0630_0713_messagetype123.date,
        ais0630_0713_messagetype123.mmsi AS mmsi,
        ais0630_0713_messagetype123.aispoint AS geometry,
        ais0630_0713_messagetype123.speedoverground AS speed,
        ais0630_0713_messagetype123.courseoverground AS course,
        ais0630_0713_messagetype123.trueheading AS heading
FROM
        "ais0630_0713_messagetype123"
WHERE
        st_setsrid(
                -- note, coordinates of line are transformed from RD to WGS'84, this is not necessary, they can be RD or WGS.
                st_transform(
                        -- make buffer around the passing line
                        st_buffer(
                                st_setsrid('LINESTRING(4.258744614 51.906376916, 4.261750768
                                51.908930517)'::geometry(LineString), 4326),
                                0.0001
                        )
                , 4326)::box2d,
                4326)::geometry(Polygon, 4326)
        && ais0630_0713_messagetype123.aispoint
AND
        ais0630_0713_messagetype123.messagetype in (0,1,2,3);
```

Figure 27. Query that defines point that are in close proximity of line strings from cell boundaries.

```
--before committing this query, run
-- ALTER TABLE "table" ALTER COLUMN date TYPE timestamptz USING (date::timestamptz); for involved tables.

DROP TABLE IF EXISTS ais0630_0713_tracks_around_line;
CREATE TABLE
	"ais0630_0713_tracks_around_line"
AS
SELECT * FROM
(
	WITH "ais0630_0713_messagetype123" AS
	(
	SELECT
		*
	FROM
		"ais0630_0713_line_passing_filter"
	ORDER BY
		mmsi, date
	)

	SELECT
		mmsi,
		start_date,
		end_date,
		tstzrange(start_date, end_date) AS happened_date,
		duration_secs,
		dist,
		sog1,
		sog2,
		cog1,
		cog2,
		head1,
		head2,
		CASE WHEN duration_secs <> 0 THEN dist / duration_secs
		ELSE null
		END AS speed, geo_segment
	FROM
	(
		SELECT
			mmsi,
			start_date,
			end_date,
			EXTRACT(EPOCH FROM (end_date - start_date)) as duration_secs,
			st_distance(st_transform(geom1, 4326), st_transform(geom2, 4326)) as dist,
			st_makeline(geom1,geom2)::geometry(LineString, 4326) as geo_segment,
			sog1,
			sog2,
			cog1,
			cog2,
			head1,
			head2
		FROM (
			-- make a table with start and end time stamp of segment
			-- where segment is point and next point in sequence (lead)
			SELECT
				mmsi,
				date AS start_date,
				lead(date) OVER w AS end_date,
				-- row_number() OVER w AS num,
				geometry AS geom1,
				lead(geometry) OVER w AS geom2,
				speed as sog1,
				Lead(speed) OVER w AS sog2,
				course as cog1,
				Lead(course) OVER w AS cog2,
				heading as head1,
				Lead(heading) OVER w AS head2
				FROM
				"ais0630_0713_messagetype123"
			WINDOW w AS (PARTITION BY mmsi ORDER BY date)
		) as widetable
		WHERE
			widetable.end_date - widetable.start_date < interval '1 minute'
	) as segmenttable
) AS R
;
```

Figure 28. query to connect points within a minute into track around cell boundaries.

```
-- make a table with segments that cross the line
DROP TABLE IF EXISTS ais0630_0713_tracks_crossing_line;
CREATE TABLE
	"ais0630_0713_tracks_crossing_line"
AS
SELECT
	*
FROM
	"ais0630_0713_tracks_around_line"
WHERE
	st_intersects(
		geo_segment,
		st_transform(
			st_setsrid('LINESTRING(4.258744614 51.906376916, 4.261750768 51.908930517)'::geometry(LineString), 4326),
			4326))
```

Figure 29. query to find trajectories crossing cell boundaries.

```
-- make a join with sectorboundaries for further analysis
##Sectorboundaries, this does create a new record for every sectorboundary that the line intersects with.

DROP TABLE IF EXISTS ais0630_0713_intersects_sectorboundaries;
CREATE TABLE
	"ais0630_0713_intersects_sectorboundaries"
AS
SELECT
			ais0630_0713_tracks_crossing_line.mmsi,
			ais0630_0713_tracks_crossing_line.start_date,
			ais0630_0713_tracks_crossing_line.end_date,
			ais0630_0713_tracks_crossing_line.geo_segment,
			ais0630_0713_tracks_crossing_line.sog1,
			ais0630_0713_tracks_crossing_line.sog2,
			ais0630_0713_tracks_crossing_line.cog1,
			ais0630_0713_tracks_crossing_line.cog2,
			ais0630_0713_tracks_crossing_line.head1,
			ais0630_0713_tracks_crossing_line.head2,
			sectorboundaries.boundaryid,
			sectorboundaries.boundarygeom
FROM "ais0630_0713_tracks_crossing_line"
JOIN "sectorboundaries"
ON ST_intersects((ais0630_0713_tracks_crossing_line.geo_segment), sectorboundaries.boundarygeom);
```

Figure 30. Join of cell boundaries and track crossing line.

```
-- make a table with segments that have their startpoint inside a sectors and the same  but than for endpoint
## Sectors
DROP TABLE IF EXISTS ais0630_0713_startpoint_inside_sector;
CREATE TABLE
	"ais0630_0713_startpoint_inside_sector"
AS
SELECT
			ais0630_0713_intersects_sectorboundaries.mmsi,
			ais0630_0713_intersects_sectorboundaries.start_date,
			ais0630_0713_intersects_sectorboundaries.end_date,
			ais0630_0713_intersects_sectorboundaries.geo_segment,
			ais0630_0713_intersects_sectorboundaries.sog1,
			ais0630_0713_intersects_sectorboundaries.sog2,
			ais0630_0713_intersects_sectorboundaries.cog1,
			ais0630_0713_intersects_sectorboundaries.cog2,
			ais0630_0713_intersects_sectorboundaries.head1,
			ais0630_0713_intersects_sectorboundaries.head2,
			ais0630_0713_intersects_sectorboundaries.boundaryid,
			ais0630_0713_intersects_sectorboundaries.boundarygeom,
			sectors.sectorid,
			sectors.sectorgeom
FROM "ais0630_0713_intersects_sectorboundaries"
JOIN "sectors"
ON ST_within(ST_StartPoint(ais0630_0713_intersects_sectorboundaries.geo_segment), sectors.sectorgeom);
--OR ST_within(ST_EndPoint(ais0630_0713_intersects_sectorboundaries.geo_segment), sectors.sectorgeom);
```

Figure 31. Query for finding whether the start point or the endpoint is inside the cell.

```
-- connect the two linestrings that are the begin and end of a sector.
-- Within a reasonalble time interval, +- 5 min? (enddate and next following startdate with the same sector should be within short time)

DROP TABLE IF EXISTS ais0630_0713_connected_trajectories;
 CREATE TABLE
            "ais0630_0713_connected_trajectories"
            AS
            SELECT
  ais0630_0713_startpoint_inside_sector.mmsi as mmsi,  ais0630_0713_startpoint_inside_sector.start_date as start_date_2,
  ais0630_0713_startpoint_inside_sector.end_date as end_date_2,  ais0630_0713_startpoint_inside_sector.sog1 as sog1_2,
  ais0630_0713_startpoint_inside_sector.sog2 as sog2_2,  ais0630_0713_startpoint_inside_sector.cog1 as cog1_2,
  ais0630_0713_startpoint_inside_sector.cog2 as cog2_2,  ais0630_0713_startpoint_inside_sector.head1 as head1_2,
  ais0630_0713_startpoint_inside_sector.head2 as head2_2,  ais0630_0713_startpoint_inside_sector.direction as direction2,
  ais0630_0713_endpoint_inside_sector.start_date as start_date_1,  ais0630_0713_endpoint_inside_sector.end_date as end_date_1,
  ais0630_0713_endpoint_inside_sector.sog1 as sog1_1,  ais0630_0713_endpoint_inside_sector.sog2 as sog2_1,
  ais0630_0713_endpoint_inside_sector.cog1 as cog1_1,  ais0630_0713_endpoint_inside_sector.cog2 as cog2_1,
  ais0630_0713_endpoint_inside_sector.head1 as head1_1, ais0630_0713_endpoint_inside_sector.head2 as head2_1,
  ais0630_0713_endpoint_inside_sector.direction as directions1, ais0630_0713_startpoint_inside_sector.geo_segment as geometry2,
  ais0630_0713_endpoint_inside_sector.geo_segment as geometry1, ais0630_0713_startpoint_inside_sector.sectorid as sectoridstart,
  ais0630_0713_endpoint_inside_sector.sectorid as sectoridend, st_makeline(ais0630_0713_endpoint_inside_sector.geo_segment
, ais0630_0713_startpoint_inside_sector.geo_segment)::geometry(LineString, 4326) as geo_segment
            FROM
                        "ais0630_0713_startpoint_inside_sector", "ais0630_0713_endpoint_inside_sector"
            WHERE
                        ais0630_0713_startpoint_inside_sector.start_date - ais0630_0713_endpoint_inside_sector.end_date < interval '1 minutes'
            AND
                        ais0630_0713_startpoint_inside_sector.sectorid = ais0630_0713_endpoint_inside_sector.sectorid
            AND
                        ais0630_0713_startpoint_inside_sector.mmsi = ais0630_0713_endpoint_inside_sector.mmsi;

-- again run a time filter of max interval between start_date_1 and end_date_2
DROP TABLE IF EXISTS ais0630_0713_connected_trajectories_filtered;
SELECT * INTO ais0630_0713_connected_trajectories_filtered from "ais0630_0713_connected_trajectories"
WHERE ais0630_0713_connected_trajectories.start_date_1 - ais0630_0713_connected_trajectories.end_date_2 < interval '5 minutes';
```

Figure 32. Query for connecting line strings and applying necessary filters.

```
-- this function is used to decide what direction the vessel is sailing, reference of the first 4 is startpoint inside sector, last 4 is endpoint inside sector.
-- 4 possible directions to go to, north east south or west.
-- create column direction

-- Update a column by setting new values with a where clause
UPDATE "table" SET column = 'valuethatreplaces'
WHERE column = 'valuestobereplaced'

--north (Same for east, south and west)
UPDATE "ais0630_0713_startpoint_inside_sector" SET direction = 'north'
WHERE boundaryid = 21 and sectorid = 5
UPDATE "ais0630_0713_startpoint_inside_sector" SET direction = 'east'
WHERE boundaryid = 3 and sectorid = 6
UPDATE "ais0630_0713_startpoint_inside_sector" SET direction = 'south'
WHERE boundaryid = 21 and sectorid = 6
UPDATE "ais0630_0713_startpoint_inside_sector" SET direction = 'west'
WHERE boundaryid = 1 and sectorid = 5

-- than delete records from table that are not necessary.
delete from "ais0630_0713_startpoint_inside_sector" WHERE direction is null;

##endpoint

--north (Same for east, south and west)
UPDATE "ais0630_0713_endpoint_inside_sector" SET direction = 'north'
WHERE boundaryid = 30 and sectorid = 5
UPDATE "ais0630_0713_endpoint_inside_sector" SET direction = 'east'
WHERE boundaryid = 2 and sectorid = 6
UPDATE "ais0630_0713_endpoint_inside_sector" SET direction = 'south'
WHERE boundaryid = 31 and sectorid = 6
UPDATE "ais0630_0713_endpoint_inside_sector" SET direction = 'west'
WHERE boundaryid = 3 and sectorid = 6

-- than delete records from table that are not necessary.
delete from "ais0630_0713_endpoint_inside_sector" WHERE direction is null;

### union both tables into new table and only keep unique values
DROP TABLE IF EXISTS "ais0630_0713_direction";
SELECT X.* INTO "ais0630_0713_direction"
FROM
        (
        SELECT * FROM ais0630_0713_endpoint_inside_sector
        UNION
        SELECT * FROM ais0630_0713_startpoint_inside_sector
        ) X ;
```

Figure 33. Ship direction queries examples.

```
-- number of ships per sector

DROP TABLE IF EXISTS ais0630_0713_ship_count_sector;
  CREATE TABLE
                "ais0630_0713_ship_count_sector"
  AS
SELECT
 ais0630_0713_connected_trajectories_filtered.sectoridend,
   COUNT(*)
FROM
   ais0630_0713_connected_trajectories_filtered
GROUP BY ais0630_0713_connected_trajectories_filtered.sectoridend;

-- number of ships per sector per direction

DROP TABLE IF EXISTS ais0630_0713_ship_count_sector_1;
  CREATE TABLE
                "ais0630_0713_ship_count_sector_1"
  AS
SELECT
 ais0630_0713_connected_trajectories_filtered.sectoridend,
   COUNT(*)
FROM
   ais0630_0713_connected_trajectories_filtered
WHERE ais0630_0713_connected_trajectories_filtered.directions1 = 'east'
AND ais0630_0713_connected_trajectories_filtered.direction2 = 'east'
GROUP BY ais0630_0713_connected_trajectories_filtered.sectoridend;
```

Figure 34. Ship count queries examples.

```
-- calculating length trough st_length, st_length(geom, false/true) and time calculation.

DROP TABLE IF EXISTS ais0630_0713_average_speed_sector;
  CREATE TABLE
                "ais0630_0713_average_speed_sector"
  AS
  SELECT
    ais0630_0713_connected_trajectories_filtered.start_date_1,
    ais0630_0713_connected_trajectories_filtered.end_date_2,
    EXTRACT(EPOCH FROM (end_date_2 - start_date_1)) as duration_secs,
    ST_Length(ais0630_0713_connected_trajectories_filtered_heading.geo_segment) * 100000 as length_m
  FROM "ais0630_0713_connected_trajectories_filtered"
                ;

-- query for calculating the average speed inside a sector in knots.

SELECT ((ais0630_0713_average_speed_sector.length_m / ais0630_0713_average_speed_sector.duration_secs) * 1.943844) AS avg_speed_kn
FROM "ais0630_0713_average_speed_sector";
```

Figure 35. query to calculate time spend - and average speed in a cell.

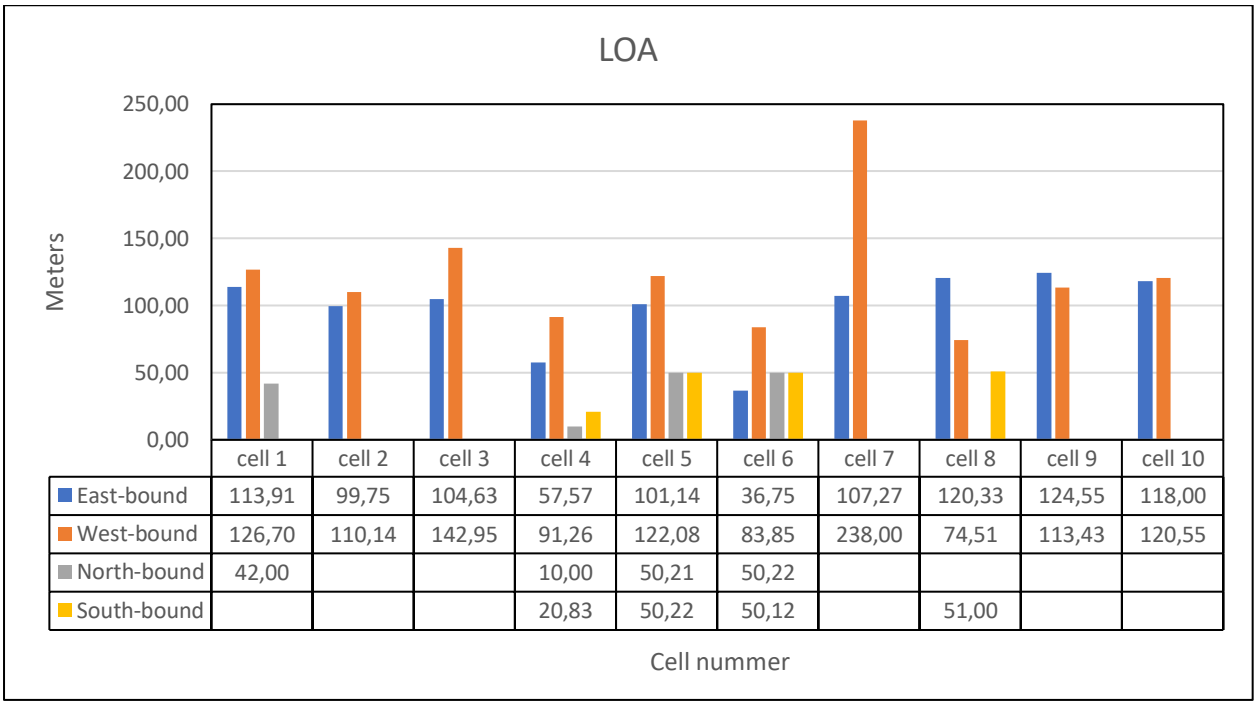
```
-- query for calculating the average speed inside a sector based on sog.(Same query can be used for cog or head).

DROP TABLE IF EXISTS ais0630_0713_average_speed_sector_sog;
 CREATE TABLE
                "ais0630_0713_average_speed_sector_sog"
 AS
 SELECT
     (SELECT AVG(sog)
      FROM   (VALUES(ais0630_0713_connected_trajectories_filtered.sog1_1),
                (ais0630_0713_connected_trajectories_filtered.sog2_1),
                (ais0630_0713_connected_trajectories_filtered.sog1_2)) V(sog)) AS sog_average,
                                                                    ais0630_0713_connected_trajectories_filtered.sectoridend,
          ais0630_0713_connected_trajectories_filtered.directions1,
          ais0630_0713_connected_trajectories_filtered.direction2
FROM   "ais0630_0713_connected_trajectories_filtered";

-- next summarize/ group by sector
DROP TABLE IF EXISTS ais0630_0713_average_speed_sector_sog_1;
 CREATE TABLE
                "ais0630_0713_average_speed_sector_sog_1"
AS
SELECT ais0630_0713_average_speed_sector_sog.sectoridend, AVG(ais0630_0713_average_speed_sector_sog.sog_average) AS sog_average_sector
FROM   "ais0630_0713_average_speed_sector_sog"
WHERE ais0630_0713_average_speed_sector_sog.directions1 = 'east'
AND ais0630_0713_average_speed_sector_sog.direction2 = 'east'
GROUP BY ais0630_0713_average_speed_sector_sog.sectoridend;
```

Figure 36. query to calculate average SoG in a cell in a certain direction.

```
-- ship dimension calculation, length and width calculation.
-- length = dimension to bow and stern, the special value 511 indicates 511 meters or greater;
-- width = dimension to port and starboard,  the special value 63 indicates 63 meters or greater.

DROP TABLE IF EXISTS ais0630_0713_ships_dimension;
 CREATE TABLE
            "ais0630_0713_ships_dimension"
 AS
 SELECT
   ais0630_0713_messagetype123.mmsi,
   (ais0630_0713_messagetype123.dimensiontostern + ais0630_0713_messagetype123.dimensiontobow) as length_dimension,
            (ais0630_0713_messagetype123.dimensiontoport + ais0630_0713_messagetype123.dimensiontostarboard) as width_dimension
FROM "ais0630_0713_messagetype123";

-- enrich table ais0630_0713_connected_trajectories_filtered with dimension of ships

UPDATE "ais0630_0713_connected_trajectories_filtered"
SET length_dimension = ais0630_0713_ships_dimension.length_dimension
FROM "ais0630_0713_ships_dimension"
WHERE ais0630_0713_connected_trajectories_filtered.mmsi = ais0630_0713_ships_dimension.mmsi;

UPDATE "ais0630_0713_connected_trajectories_filtered"
SET width_dimension = ais0630_0713_ships_dimension.width_dimension
FROM "ais0630_0713_ships_dimension"
WHERE ais0630_0713_connected_trajectories_filtered.mmsi = ais0630_0713_ships_dimension.mmsi;

-- calculate average ship dimensions per sector

DROP TABLE IF EXISTS ais0630_0713_average_dimension_sector;
 CREATE TABLE
            "ais0630_0713_average_dimension_sector"
 AS
 SELECT
     (SELECT AVG(length)
     FROM   (VALUES(ais0630_0713_connected_trajectories_filtered.length_dimension)) V(length)) AS length_average,
     (SELECT AVG(width)
      FROM   (VALUES(ais0630_0713_connected_trajectories_filtered.width_dimension)) V(width)) AS width_average,
                                        ais0630_0713_connected_trajectories_filtered.sectoridend,
        ais0630_0713_connected_trajectories_filtered.directions1,
        ais0630_0713_connected_trajectories_filtered.direction2
FROM   "ais0630_0713_connected_trajectories_filtered";

-- calculate average ship dimensions per sector per direction

DROP TABLE IF EXISTS ais0630_0713_average_dimension_sector_1;
 CREATE TABLE
            "ais0630_0713_average_dimension_sector_1"
AS
SELECT ais0630_0713_average_dimension_sector.sectoridend,
     AVG(ais0630_0713_average_dimension_sector.length_average) AS length_average_sector,
     AVG(ais0630_0713_average_dimension_sector.width_average) AS width_average_sector
FROM   "ais0630_0713_average_dimension_sector"
WHERE ais0630_0713_average_dimension_sector.directions1 = 'east'
AND ais0630_0713_average_dimension_sector.direction2 = 'east'
GROUP BY ais0630_0713_average_dimension_sector.sectoridend;
```

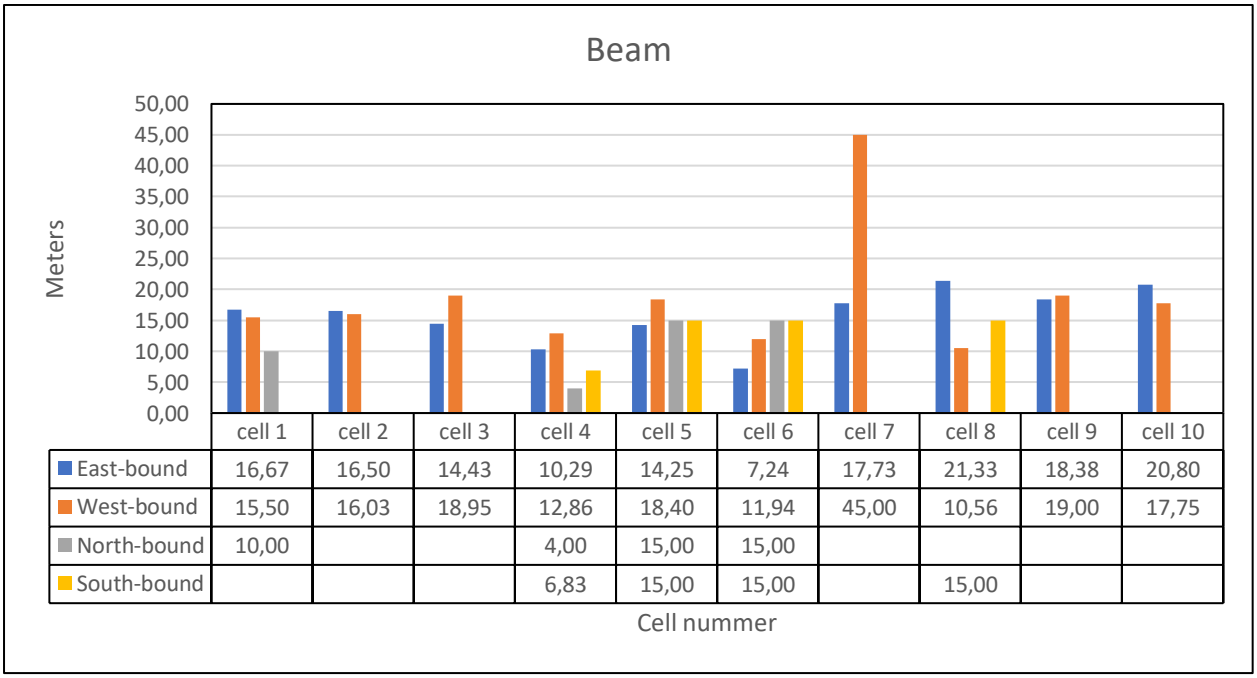
Figure 37. Queries for calculating LOA and Beam (length and width) averages per cell per direction.

Appendix 3: Histograms



	cell 1	cell 2	cell 3	cell 4	cell 5	cell 6	cell 7	cell 8	cell 9	cell 10
East-bound	113,91	99,75	104,63	57,57	101,14	36,75	107,27	120,33	124,55	118,00
West-bound	126,70	110,14	142,95	91,26	122,08	83,85	238,00	74,51	113,43	120,55
North-bound	42,00			10,00	50,21	50,22				
South-bound				20,83	50,22	50,12		51,00		

Figure 38. Average length in meters in N,E,S,W direction



	cell 1	cell 2	cell 3	cell 4	cell 5	cell 6	cell 7	cell 8	cell 9	cell 10
East-bound	16,67	16,50	14,43	10,29	14,25	7,24	17,73	21,33	18,38	20,80
West-bound	15,50	16,03	18,95	12,86	18,40	11,94	45,00	10,56	19,00	17,75
North-bound	10,00			4,00	15,00	15,00				
South-bound				6,83	15,00	15,00		15,00		

Figure 39. Average width in meters in N,E,S,W direction