GIMA
Geographical Information Management and Applications

Mapping Maritime Risk in the Kattegat Using the Automatic Identification System

Luc van der Lecq

1 March 2021

Thesis for the Degree of Master of Science

Universiteit Utrecht
TU Delft
University of Twente.
Wageningen UR

Mapping Maritime Risk in the Kattegat Using the Automatic Identification System

by

Luc van der Lecq

1 March 2021

Thesis for the Degree of Master of Science
Geographical Information Management and Applications (GIMA)
Utrecht University

Mapping Maritime Risk in the Kattegat Using the Automatic Identification System

Thesis for the Degree of Master of Science

Author	Luc van der Lecq
Student number	5719305
Date	1 March 2021
Programme	Geographical Information Management and Applications (GIMA) Faculty of Geosciences Utrecht University (Delft University of Technology, University of Twente, Wageningen University & Research)
Supervision	Martijn Meijers, Delft University of Technology Yigit Can Altan, Delft University of Technology
Responsible professor	Peter van Oosterom, Delft University of Technology

Preface

This thesis on *Mapping Maritime Risk in the Kattegat Using the Automatic Identification System* has been written as part of the graduation process of the Geographical Information Management and Applications (GIMA) Master's programme, a joint programme between the universities of Utrecht, Delft, Twente, and Wageningen.

My work on the topic of this thesis started in early July of 2020, partly inspired by my positive experiences with digital shipping applications at my part time work placement. While initially aiming to create a comparative case study on the effect of traffic management interventions, the focus of my thesis work slowly shifted towards exploring the technical details of modelling vessel movement data and designing a complementary risk analysis system. This balance proved to be a recurring theme in my research process and keeping it has been an enjoyable learning experience. Fortunately, my connection with vessel traffic and the Automatic Identification System does not end with this thesis project, as I will continue my academic endeavours in the maritime traffic domain in the form of a research internship.

I would like to use this opportunity to thank the people that have helped me complete this thesis in times of corona. In particular, I have appreciated the helpful comments of and discussions with my supervisors, Martijn Meijers and Yigit Can Altan. My fellow GIMA students proved to be indispensable in keeping spirits high, despite the unfortunate closure of the GIS facilities at Utrecht University. My friends and family also deserve a big thank you for their valuable feedback and their (online) support on all fronts.

I wish you a pleasant reading experience.

Luc van der Lecq
Utrecht, 1 March 2021

Summary

The Kattegat strait between Denmark and Sweden features high vessel traffic densities and difficult navigational conditions, making it susceptible to ship collisions. This led to the implementation of a new shipping route system as of July 2020. This study has assessed the change in the spatial distribution of maritime risk in the north of the Kattegat after this implementation and its relation to the spatial pattern of vessel movements and recorded maritime incidents.

The maritime risk concept is formed by extracting vessel proximity, encounter type, speed, and evasive manoeuvres from non-accident critical events. An automated maritime risk assessment has been developed using vessel tracking data from the Danish Automatic Identification System (AIS). Several millions of data points have been pre-processed in Python and stored in a PostGIS database using a basic line segment data model. A three-dimensional index has been generated to efficiently intersect segments in space and time. A kernel density estimation has been applied in QGIS to create heatmaps of maritime risk from the detected vessel encounter locations and their risk values.

A 17% decrease in encounter frequency has been observed across the entire research area, which is in line with projections from before the shipping route changes. Maritime risk values increased slightly. Most ship encounters occurred along the central shipping corridor from and towards the Baltic sea and involved cargo vessels. This may be because these areas see the most vessel movements and cargo vessels are the predominant ship type. The maritime risk pattern along this corridor shifted to the east, which is most likely caused by the shipping route changes and the corresponding shift in traffic density. Encounters involving passenger vessels are relatively rare, but have high risk values. This may be due to their crossing of the busy cargo shipping corridor. It is concluded that maritime risk corresponds to traffic intensity in terms of the spatial pattern and vessel type. Due to spatial and temporal scope constraints, the generalisability of this conclusion is limited.

The result do not show a pronounced relationship between the spatial pattern of maritime risk and the locations of recorded accidents. Correcting the estimated maritime risk density for the traffic density bias also did not lead to new insights.

Contents

1 Introduction

1.1 Problem statement

With the demand for global transport and seaborne trade showing an upward trend, vessel traffic intensities across the world's waterways are high and projected to increase (International Transport Forum, 2019; Tran & Haasis, 2015). To fulfil this demand efficiently, vessels are becoming larger and thus require deeper draughts (Tran & Haasis, 2015). Aside from shipping operations, a wide range of other current developments in the maritime domain contribute to increasing pressures on the use of sea space. These developments include the installation of offshore wind farms, the designation of marine protected areas, the exploitation of mineral resources on the sea floor, and maritime recreation.

1.1.1 Shipping activity and navigational conditions

The Kattegat strait between Denmark and Sweden is no exception to these global trends and features heavy international vessel traffic. This strait links the North Sea and the Baltic Sea and saw nearly 75,000 total ship passages in 2019 (approximately 10,000 unique ships) (Grimvall & Larsson, 2014). The established shipping routes are primarily used by cargo ships and tankers (International Maritime Organization [IMO], 2017a). Not only the shipping routes through the Kattegat see heavier use than those in the Baltic Sea; virtually no part of the Kattegat area is free of vessel traffic (HELCOM [Baltic Marine Environment Protection Commission], 2018a). Other relevant groups of vessels are high speed passenger crafts and ferries crossing the main traffic stream, a large fishing fleet, and smaller, seasonal crafts (IMO, 2017a).

Difficult navigational conditions add a dimension of complexity for ships sailing through the Kattegat. Large parts of the water between Denmark and Sweden are less than 30 meters deep, limiting the navigable space for deep-draught shipping (Du, Goerlandt, & Kujala, 2020). This results in a relatively high number of groundings (HELCOM, 2018a; IMO, 2017a). Additionally, the unique environmental configuration of the Kattegat has an impact on navigational safety. The inflow of salt water from the North Sea versus the inflow of fresh water from rivers and rain causes spatial differences in water density (Kerbrat, 2018). This can influence a ship's draught throughout its passage. The Kattegat is also known for its high easterly winds in winter and spring, as well as swell from the North Sea (Kerbrat, 2018). Both of these phenomena result in waves that can impede shipping operations.

1.1.2 Maritime incidents

High vessel traffic densities and difficult navigational conditions make the Kattegat susceptible to ship collisions (Du et al., 2020). Between 2011 and 2018, a total of 474 maritime casualties and incidents were reported in the Kattegat region (European Maritime Safety Agency, 2019). The IMO (2017a) has recorded 20 collisions and groundings between 2000 and 2017. More than half of the maritime casualties and incidents were caused by navigational issues. The mid-water phase of a ship's journey within internal waters or on territorial sea is considered as most unsafe. No significant decreasing trend in terms of maritime accident events has occurred in the last decade (EMSA, 2019; HELCOM, 2018a).

Maritime incidents in the Kattegat have a considerable environmental impact. A large proportion of accidents with pollution (including oil spills) between 2011 and 2016 in the Baltic Sea occurred along Kattegat shipping routes (HELCOM, 2018b). A valid explanation for this spatial pattern is given by Lunde Hermansson and Hassellöv (2020): laws and regulations on emissions are linked to territorial borders and water depth, resulting in a concentration of tank washing operations in a small area. Even though the number and size of oil spills shows a decreasing trend for the Baltic Sea in general, this is not the case for the Kattegat subregion (reference period 2008-2013) (HELCOM, 2018b).

1.1.3 Ship routing changes

In the last ten years, the increase in vessel traffic and number of maritime incidents have prompted the need for traffic management improvements in the Kattegat. This led to the development of a new shipping route system (IMO, 2017a). As of July 2020, these changes have been put in place (Danish Maritime Authority [DMA], 2020). The approved changes include (1) deep water fairway improvements, (2) alterations in the route to the Great Belt, (3) a new, shallower route to the Sound, and (4) a number of new traffic separation schemes. A map of the changes is included in figure 1.1.

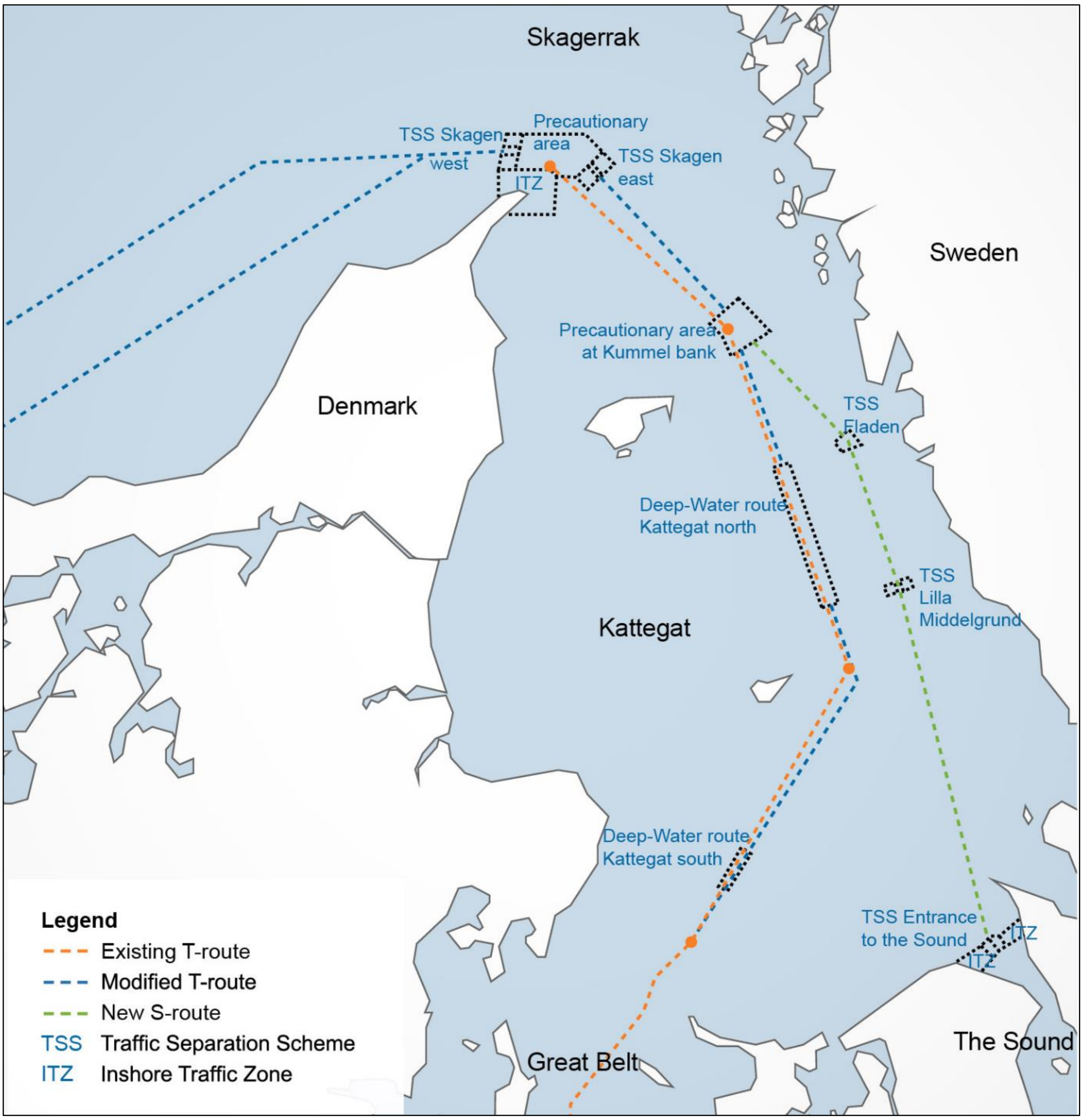

Figure 1.1 Ship routing changes in the Kattegat as of July 2020 (SSPA Sweden, 2018)

The new situation replaces a 40 year old ship routing system deemed unfit for modern ship dimensions and traffic separation requirements (IMO, 2017a; Kerbrat, 2018). The goals of these changes are "to organize the flow of ship traffic in more predictable ways to enhance the safety of navigation and protection of the marine environment" (IMO, 2017a, p. 1). As the Baltic Sea handles up to 15 percent of the cargo traffic in the world, achieving these goals is in the interest of Danish and Swedish authorities and other member states of the IMO (Federal Maritime and Hydrographic Agency [BSH], 2016).

1.1.4 Automated maritime risk assessment

Manually detecting and analysing the potential risk of maritime incidents in a busy maritime region like the Kattegat can be labour intensive, imprecise, or disruptive. Using accident reports proved to be unreliable and the number of reports is too limited for a meaningful statistical analysis (Du et al., 2020).

Advancements in data capture and storage technology have paved the way for the automation of waterway risk assessments. The publication of big digital datasets on vehicular movements is a particularly relevant development. In this matter, *big* data relates to both its volume and velocity (for more detail on the implications of big data for planning transport systems, see the work of Milne and Watling [2019]). The most notable vehicle tracking dataset in the maritime domain originates from the Automatic Identification System (AIS).

The use of vessel encounters extracted from AIS data in lieu of recorded accidents is common practice in maritime research publications (Du et al., 2020). According to the review of Du et al. (2020, p. 14), combining a ship domain violation with other risk indicators (e.g. rate of turn) to detect risky ship encounters is promising. The maritime risk in the Kattegat has not yet been historically evaluated with this approach. Using this combination can lead to new insights relating to the automated maritime risk assessment of waterways using large amounts of vessel movement data.

1.2 Research objectives

The main objective of this research is to identify maritime risk in the Kattegat, with a particular focus on the differences in risk before and after the ship routing changes of July 2020. Gaining insight into the changes of risk may have an added value to the operations of maritime authorities. This main objective requires a definition of risky ship behaviour and its indicators. Spatial patterns of ship traffic must also be mapped to add context to risky ship encounters.

The second objective is of a methodological nature and involves the combination of a ship domain approach with Danish vessel tracking data from the AIS. In order to study maritime risk using this data source, a scalable database management system (DBMS) must be designed to manage the large amounts of AIS data. This system should be able to handle complex queries within reasonable response times. The final DBMS design, as well as the considerations during its development, may be a valuable addition to the body of knowledge on database models for movement data.

1.2.1 Research questions

In order to address the problem statement and research objectives, a main research question is posed:

> To what extent has the spatial distribution of maritime risk changed after the implementation of new shipping routes in the Kattegat on 1 July 2020 and how does this relate to the spatial pattern of vessel movements and to recorded maritime incidents?

Five sub questions serve as building blocks for the main research question. They are:

1. How can maritime risk be defined?
2. How can maritime risk be analysed with spatial movement data?
3. What is the spatial pattern of vessel movements?
4. To what extent has the spatial pattern of maritime risk changed after the implementation of new shipping routes?
5. To what extent does the spatial pattern of maritime risk relate to the spatial pattern of vessel movement and recorded maritime incidents?

The sub questions will hereinafter be referred to by their number.

1.2.2 Spatial and temporal scope

Given the spatial extent of the aforementioned ship routing changes, only vessel movements in a specific part of the Kattegat sea area will be selected to study maritime risk. The northern limit of this research area is roughly between the Skaw in Denmark and Göteborg in Sweden. In the west and east, the port areas of Frederikshavn and Göteborg are excluded. The southern limit of the research area is formed by a line between (1) the Danish town of Sæby, (2) the Danish island of Læsø, and (3) the start of the new S shipping route towards the Sound. The total research area covers more than 2,500 km² and is shown on the map in figure 1.2.

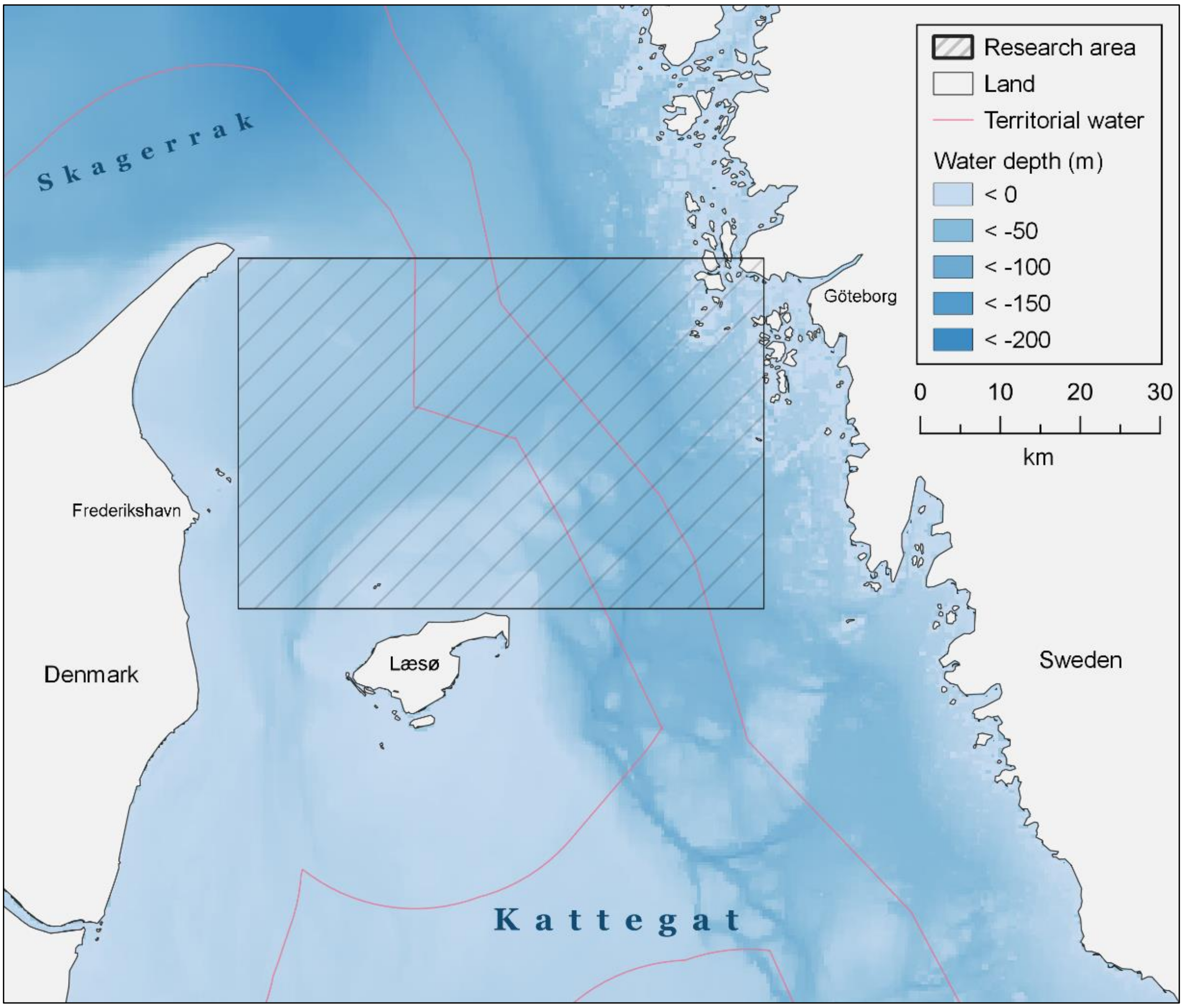

Figure 1.2 Overview of the research area
(Baltic Sea Hydrographic Commission, EuroGeographics, HELCOM)

No recent changes have been made to the ship routing system in the south of the Kattegat. In addition, the maritime situation of the Danish straits is not comparable to the spatial configuration in the north of the Kattegat. For these reasons, the research area definition excludes vessel movements in the southern part of the Kattegat, the Little Belt, the Great Belt, and the Sound from the research scope.

The aforementioned ship routing changes in the Kattegat have been effectuated as of 1 July 2020. To evaluate these changes in terms of maritime risk, vessel movement data from a three month period before 1 July 2020 is selected as a reference period (henceforth referred to as timeframe 1). This data has been compared to data from 1 July 2020 up to and including September 2020 (henceforth referred to as timeframe 2).

1.3 Research structure

The flow of this thesis is summarised in figure 1.3. It starts with a comprehensive review of academic literature. From this theoretical basis, a maritime risk assessment methodology and a suitable DMBS are developed in tandem. Results are generated using this spatial analysis method. Finally, the results of the spatial analysis are reflected upon.

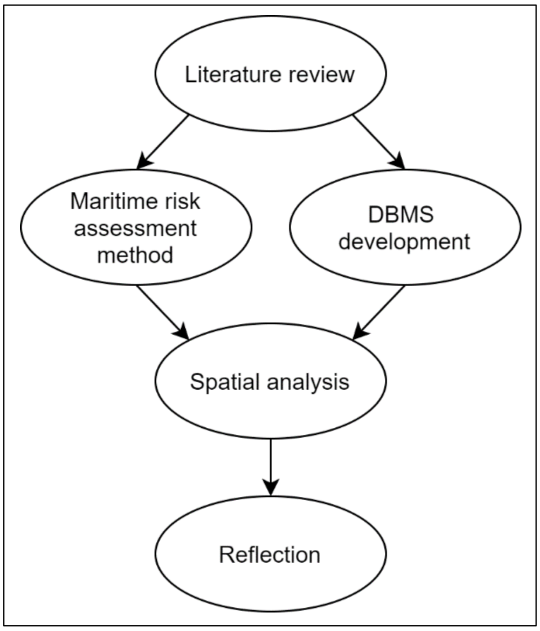

Figure 1.3 The flow of the thesis

This flow of reasoning is incorporated into six chapters. Their respective contents are as follows:

- in chapter 2, existing academic works on maritime risk assessment and database models for movement data are reviewed;
- next, chapter 3 describes the steps that are required to answer the research questions, including a motivation for the choices regarding the methodology of this research;
- the main findings are presented in chapter 4;
- chapter 5 provides an interpretation of the results, their meaning in relation to the theoretical background, as well as the limitations of the research methodology;
- answers to the five sub questions and the main research question are formulated in chapter 6, as well as directions for future research.

2 Theoretical background

The theoretical background of this research starts with a review of risk assessment methods in a maritime context in section 2.1. A definition of the constituents of the maritime risk concept is put forward in section 2.2. This definition contributes towards the answer to sub question 1. Section 2.3 gives an overview of database models that are suited for storing spatio-temporal data. A condensed identification of the knowledge gap is presented in the final section of this chapter.

2.1 Maritime risk assessment methods

The review of Du et al. (2020) identifies a range of methods for studying maritime traffic risk and safety. Examples include the extraction of maritime movement patterns, anomaly detection, ship accident probability and consequences modelling, and the statistical analysis of maritime accidents. A distinction can be made between methods that use non-accident critical events and methods that have recorded accidents as input.

2.1.1 Maritime risk assessment using recorded accidents

The most direct way of studying maritime risk and safety for a given waterway for a given period of time is leveraging maritime incident databases and investigation reports. However, accident reports proved to be unreliable and the number of reports too limited for meaningful statistical analyses (Du et al., 2020). The underreporting of accident data – which is an established issue and not limited to the maritime domain – inhibits the use of accident records for traffic risk assessments, because it may lead to biased risk assessments (Du et al., 2020; Li, Weng, & Fu, 2020). Selective underreporting may also occur, in the sense that only incidents of a certain size, impact or at a certain location from the coast (Li et al., 2020). It may sometimes be in the interest of ship companies to deliberately withhold information during incident investigations or local procedures may be unclear (Li et al., 2020). An added downside of accident-based approaches is the lack of contextual information that incident reports provide (Du et al., 2020).

A statistical analysis of maritime accidents has been performed on the Baltic sea level (BSH, 2016; Helcom, 2018a). Unfortunately, the cause and location of many recorded accidents remains unknown or unreported. Accident reports that did include this information show that collisions are the most common type of accident, especially in the southwestern part of the Baltic Sea with more than 50%. This percentage has shown an increasing trend in the last decade. The leading cause of maritime accidents is unintentional human error relating to navigation. Areas with limited vessel manoeuvrability around ports are main collision hot spots, as well as the mid-water phase of a ship's journey within internal waters or on territorial sea. Construction work at sea and recreational boating in coastal areas are thought to be complicating factor, as this leads to more potential objects to collide with and more traffic.

2.1.2 Maritime risk assessment using non-accident critical events

Given the drawbacks of using recorded accidents as input for maritime risk assessments, it is unsurprising that the number of published articles using non-accident critical events has shown an upward trend between 2006 and 2019 (Du et al., 2020). Non-accident critical events approaches commonly use vessel tracking information such as AIS as input data. The spatial tracking of ships allows researchers to evaluate sailing behaviour and other conditions leading up to an incident. Accident reports may not include this information. Additionally, non-accident critical events have a higher occurrence than ship collisions and can be seen as a safety performance indicator of a maritime traffic system (Du et al., 2020). This has also been found in practice by Hänninen and Kujala (2014), who report on the relationship between incident reporting (including near misses) and accident involvement in the Gulf of Finland.

Using non-accident critical events to study maritime risk is based on relational accident theory. The model structure of this theory can be conceptualised using Hydén's safety pyramid (see figure 2.1). It is assumed that several types of events placed on a continuum of severity occur according to a certain ratio and corresponding frequency (Hydén, 1987). In other words, ship collisions are thought to seldomly occur and have high impacts, whereas potential maritime conflicts happen more frequently but tend to have a lower severity or impact. With this pyramid of proportions in mind, it is only logical that maritime risk assessments rarely focus on collisions.

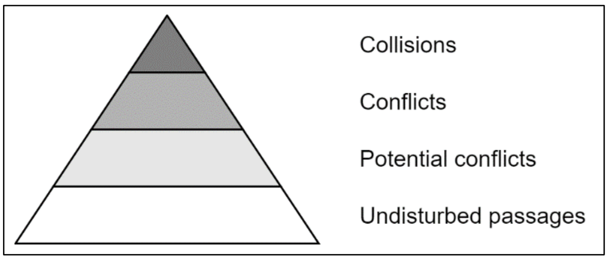

Figure 2.1 Conceptual safety pyramid of traffic conflicts (after Hydén, 1987)

Ship collision risk assessments using non-accident critical events have been performed for many waterways worldwide. Examples include the coast of Portugal, the North Sea, the Singapore Strait, the strait of Istanbul, and the Gulf of Finland (Altan & Otay, 2018; Goerlandt, Montewka, Lammi, & Kujala, 2012; Qu, Meng, & Suyi, 2011; Silveira, Teixeira, & Soares, 2013; Van Westrenen & Ellerbroek, 2017).

As part of the implementation process of the new shipping routes in the Kattegat (see section 1.1.3), a commercial analysis of ship collision risk has been performed using data from the AIS in 2017 (IMO, 2017b). It can be categorised as research on ship accident probability and consequences modelling. This involved a mathematical model that projects historical AIS traffic onto a new shipping route layout to create a scenario for comparison. The new configuration of the shipping routes is expected to lead to a decrease in collisions and groundings of eight percent (IMO, 2017b). However, this analysis has not been performed with the exact new shipping route layout that has been implemented. In addition, it does not consider individual vessel encounters, but uses aggregate data and applies an inflexible probability factor to calculate maritime risk instead.

2.2 Indicators of maritime risk

A key element of the non-accident critical events approach is classifying events according to their degree of risk, thereby effectively placing them on the safety pyramid model of figure 2.1. In a maritime context, this maritime risk concept is formed by (1) modelling ship behaviour variables extracted from movement data and (2) enriching this data with explanatory environmental variables. The underlying idea of this model is that situations with a higher maritime risk have a higher probability of being an accident.

According to Du et al. (2020), the variables speed, course, rate of turn, position, encounter angle, and ship domain belong to the most used variables to model maritime risk using non-accident critical events (n=39). These can be grouped into four key factors that influence maritime risk, which have been conceptualised in figure 2.2. Assessing risky ship encounters (i.e. a non-accident critical event) with a ship domain violation is deemed promising (Du et al., 2020).

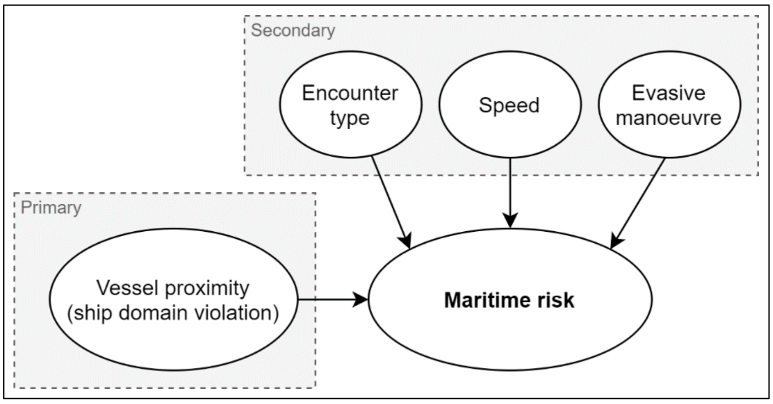

Figure 2.2 Conceptual model of key maritime risk indicators

2.2.1 Ship domain violation

A common denominator in research on maritime risk assessments is the definition of a asymmetrical safety area around a ship that is preferably kept free from other vessels (Du et al., 2020). This ship domain concept has been introduced in a maritime context in the 1970s (Fujii & Tanaka, 1971). Ship domains can be the result of theoretical analyses or experts' knowledge, which tend to be more demanding, heavily parameterised models, whereas more elementary definitions are generally determined empirically (Szlapczynski & Szlapczynska, 2017). Szlapczynski and Szlapczynska (2017) also note that empirically determined ship domains may be less suitable for use in the collision avoidance decision making process due to their relative simplicity. On the other hand, some authors find the complexity of more advanced ship domain models to be impractical (Szlapczynski & Szlapczynska, 2017). Regardless of the type of development, the ship domain is commonly affected by the type of the water region, traffic density, and traffic patterns (Hörteborn, Ringsberg, Svanberg, & Holm, 2019; Szlapczynski & Szlapczynska, 2017).

Even though the ship domain is one of the most used variables to model maritime risk, the geometric definition of this domain has considerable variations in the scientific literature as various authors furthered the work of Fuji and Tanaka between 1970 and 1980 (Du et al., 2020; Szlapczynski & Szlapczynska, 2017). Early mainstream definitions were established by Goodwin, Coldwell, and Davis et al. (see figure 2.3). Various authors approach the safety criterion of the ship domain differently by choosing which domain of the two vessels in an encounter must be violated by the other's presence to constitute risk. A more contemporary approach is that the ship domains of both vessels must overlap for a violation to occur (Szlapczynski & Szlapczynska, 2017).

Hörteborn et al. (2019) have revisited the debate on ship domain definitions by analysing its shape using AIS data in the Baltic Sea, including eight locations in the Kattegat research area of this study. This resulted in a generic ship domain definition of an ellipsoid of approximately 3.3 by 1.7 kilometres (see figure 2.4). Several additional conclusions were drawn:

- The ship domain size must decrease in restricted waters.
- The ship domains must attain a more circular shape in a crossing encounter.
- The ship domain is of little significance in a head-on encounter due to the fairway traffic separation system.
- The ship domain was not dependant on ship size.

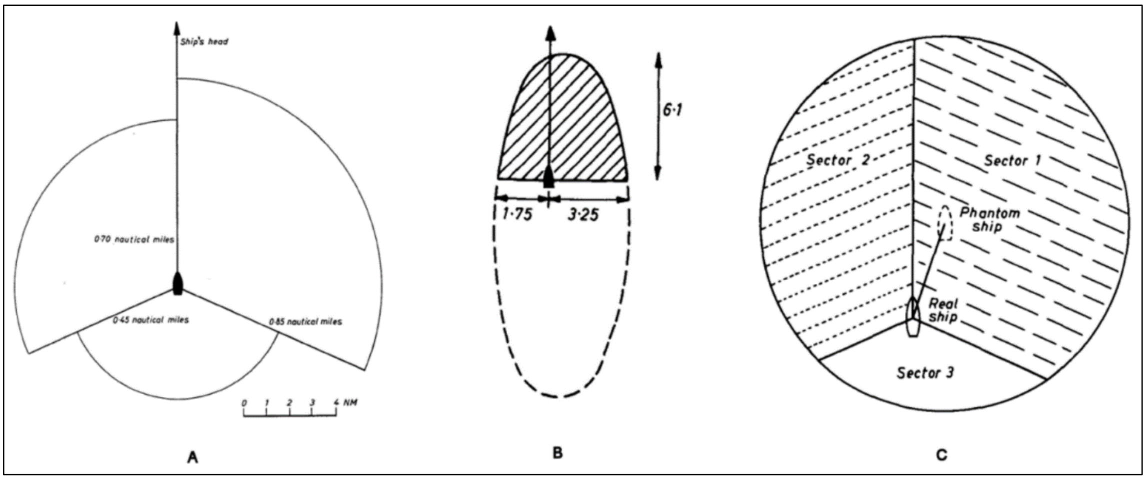

Figure 2.3 Common ship domain definitions (A: Goodwin, B: Coldwell, C: Davis and co.) (adopted from Theunissen & De Groot, 2014)

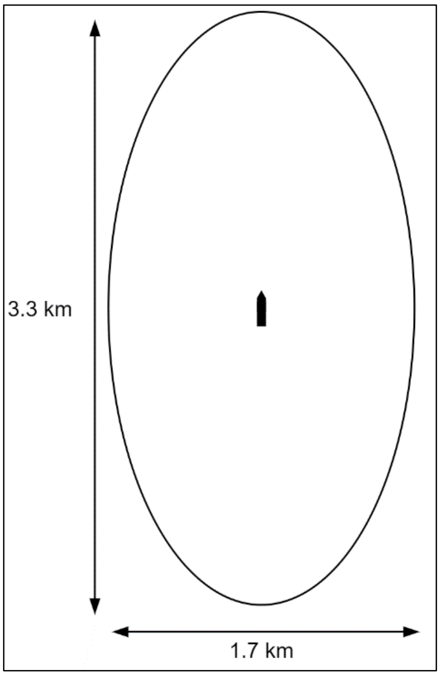

Figure 2.4 Empirically determined generic ship domain definition for the Baltic Sea (after Hörteborn et al., 2019)

2.2.2 Encounter type

The encounter angle is identified as a key indicator for modelling maritime risk (Du et al., 2020). The encounter situation of two vessels can be divided into three subgroups using this variable. When considering the individual trajectories of two vessels during an encounter, the encounter type is deducted from the encounter angle (defined as the difference between the course over ground, ΔCOG) between the two vessel. These three subgroups are generalised as:

- *Head-on encounter*
  $|\Delta COG| = 180°$ with a given offset
- *Overtaking encounter*
  $|\Delta COG| = 0°$ with a given offset
- *Crossing encounter*
  $0° < |\Delta COG| < 180$ with a given offset or $180° < |\Delta COG| < 360$ with a given offset

2.2.3 Speed

The speed difference variable is identified as a key indicator for modelling maritime risk (Du et al., 2020). When considering the individual trajectories of two vessels during an encounter, the speed difference is defined as the difference between the speed over ground (ΔSOG) between the two vessels. The speed difference variable has a positive relation with the maritime risk compound.

2.2.4 Evasive manoeuvre

Evasive manoeuvres during vessel encounters may indicate risky shipping behaviour. Such an evasive manoeuvre can be detected by assessing the rate of turn level of a vessel during its journey, because a high rate of turn level is a rare occurrence during normal shipping operations and may results from last moment evasive action (Mestl, Tallakstad, & Castberg, 2016). This rate of turn (ROT) variable is commonly used as a key indicator for modelling maritime risk, as "it is straight forward and indeed very fast to single out potential critical situations characterized by non-normal manoeuvres" (Du et al., 2020; Mestl et al., 2016, p. 76).

Mestl et al. (2016) note that high rate of turn levels in themselves are not sufficient indicators and propose a method of combining the turning rate with vessel speed to more accurately reflect abnormal shipping behaviour. The product of the ROT and SOG variables allows for the calculation of the centripetal acceleration (CA) of a vessel, which is defined as the change in rotation around the vertical axis to which people on a ship are exposed. The centripetal acceleration variable has a positive relation with the maritime risk compound.

2.2.5 Ancillary risk indicators

A range of additional factors can be identified that lie outside the vessel tracking information domain. The following groups of ancillary risk indicators are not included in the conceptual model of figure 2.2, as they are considered to lie outside of the scope of this research. They are, however, indicators of maritime risk.

Firstly, external causes affect safe shipping operations (BSH, 2016; Helcom, 2018a). Examples includes adverse weather, poor visibility, and strong currents (Shu, Daamen, Ligeringen, & Hoogendoorn, 2013). Second, technical equipment failure is a considerable contributing factor in shipping accidents (BSH, 2016; Helcom, 2018a). An example of this is machinery failure. Thirdly, some psychological and social elements of human error behaviour cannot be deducted from vessel tracking information such as AIS data. A key example is the decreased vigilance of mariners on the bridge.

2.3 Database management systems for moving objects

2.3.1 Storing spatio-temporal data

Given the increasing availability and magnitude of data from vessel tracking systems, finding a suitable data model and implementing it in a database system to store and analyse spatio-temporal data is challenging. Classical spatial database management systems can handle moving objects, but cannot handle time intrinsically. GIS functionality for analysing moving objects is therefore limited, as support for handling temporal information has not entered the mainstream (Graser, 2018).

The current state of the art of open source technology with temporal support has been identified in the work of Graser and Dragaschnig (2020). Two groups of tools and software come to the fore: desktop GIS and spatial databases, interactive notebook environments, and distributed computing solutions. A typology of these two groups is given in table 2.1.

Table 2.1 Typology of current GI technology with support for temporal information (after Graser & Dragaschnig, 2020)

Technology groups	Software and tools	Features and characteristics
Desktop GIS and spatial databases	QGIS, PostGIS, Jupyter, MovingPandas	Easy interactive mapping Built-in trajectory functions Less suited for big datasets Lower hardware requirements
Distributed computing solutions	GeoMesa, M3, SPARK, MobilityDB	Limited trajectory functionality Dedicated spatio-temporal indexing Designed for big datasets Advanced hardware requirements

With growing amounts of data demanding more processing power, recent scientific efforts have been focused on the development of distributed big data computing tools (see the work of Graser, Widhalm, and Dragaschnig [2020] and Bakli, Sakr, and Zimanyi [2020] for novel applications). Even though desktop GIS, spatial databases, and interactive notebook environments are primarily used on a single computer, recent developments have paved the way for parallelisation of these tools, too (Graser, 2018; Graser & Dragaschnig, 2020).

2.3.2 Indexing and clustering spatio-temporal data

The aim of database management systems for moving objects is to introduce the possibility of representing and querying moving objects to existing database technology. Spatial databases allow for efficient querying of stationary two dimensional objects by creating a hierarchical tree using the bounding boxes of the spatial features, which can then be traversed to find objects. This index is stored separately and reduces the duration of a database scan during a query.

Accommodating moving objects is more difficult (Güting et al., 2000). Many researcher have worked on indexing spatio-temporal objects (Graser, 2018). In the context of vessel tracking information systems like the AIS, a given set of spatio-temporal objects can be considered as having a third dimension, namely time. Following from this, indexes aimed at moving vessels may apply a three dimensional bounding box to efficiently access the data of parts of a vessel's trajectory in three dimensional space (Graser, 2018).

The performance of database management systems depends on the data retrieval speed from the physical disk, especially for large databases that transcend the storage capacity of the RAM (Van Oosterom, 2005). Decreasing this speed may be achieved by implementing a clustered index. A clustered index not only stores a hierarchical tree containing the order of objects, but also rearranges the data on the physical disk accordingly. For a spatial database, this means that proximate objects are stored in the same physical data block.

2.4 Knowledge gap

The introduction in the first chapter of this research documentation and the previous sections in this literature have led to the following unexplored scientific opportunities:

- The combination of non-accident critical events approach with key maritime risk indicators including a ship domain violation in the Kattegat is a unique approach in the body of literature on maritime risk assessments.
- Using a relatively accessible data model to represent ship trajectories in a PostGIS DBMS may add to best practices in the movement data science domain.
- The work of Du et al. (2020) reveals that approximately 20 percent of the selected published articles validated their model using a reality check. This leads to a need for scientific evidence that validates already existing research approaches, as opposed to proposing new methods for maritime risk assessments (Du et al., 2020).

3 Methodology

The methodological procedure of this research project is divided into three main sections. The first section contains information on the most important data source of this research project, namely the Automatic Identification System. Section 3.2 describes the workflow of analysing vessel traffic patterns in the research area. The third section on the maritime risk analysis is the largest and first discusses the design of a suitable DBMS for large amounts of spatial movement data. It also elaborates on data pre-processing steps, mapping encounter risk, and validating detected encounters.

3.1 The Automatic Identification System

The Automatic Identification System (AIS) exchanges navigational and voyage related information between ships and shore stations via very high frequency radio (VHF) (Harati-Mokhtari, Wall, Brookes, & Wang, 2007). In its essence, the AIS is a form of remote sensing according to a Lagrangian frame of reference, because it entails the continuous tracking of object (Dodge, Weibel, Ahearn, Buchin, & Miller, 2016). The promise of the AIS is an enhanced level of safety and efficient navigation at sea, because mariners receive continuous and automatic updates on ship movements in their surroundings. This applies especially to areas where more traditional navigational aids such as radar fall short. Examples are detecting ships around bends or behind landmasses. The AIS also poses benefits for maritime environmental protection, situational awareness, and waterway management (DMA, n.d.). These objectives and circumstances of use are in line with the current maritime situation of the Kattegat (see section 1.1). For this reason, vessel information from the AIS is a suitable starting point for a maritime risk analysis of the Kattegat sea area.

3.1.1 Principles of the AIS

In its core, the AIS is made up of ship-born transceivers and land based transceivers (see figure 3.1). The transceivers broadcast AIS messages regularly and automatically. The information that is received can be visualised and used for navigational decision making. This information can contain three types of content (IMO, 2015):

- *Static information*
  This type of information is used to identify a vessel, its type, and to describe its physical dimensions.
- *Dynamic information*
  This type of information covers the vessel location and the corresponding timestamp, as well as other navigational readings from vessel instruments.
- *Voyage information*
  This type of information relates to the cargo of a vessel and its destination, which are to be manually entered for each voyage.

The IMO (2015) distinguishes two shipborne equipment classes, with each class having different technical specifications and intended applications. Class A equipment is required for all vessels of more than 300 gross tonnage on international voyages, cargo vessels of more than 500 gross tonnage not on international voyages, passenger ships of all sizes, and fishing vessels longer than 15 metres. This type of equipment is highly advanced, more expensive, and aimed at commercial use. The reporting interval of class A units ranges from a maximum of 180 seconds when the vessel is at anchor or moored and not moving faster than 3 knots, to a minimum of 2 seconds when moving faster than 14 knots and changing course.

Class B transceivers may be installed on vessels that are not bound by class A regulations. These transceivers are generally less advanced, less costly, and aimed at lighter commercial vessels and smaller, seasonal crafts. In general, class B units have a lower reporting frequency, with the transmitting interval ranging from 180 to 30 seconds.

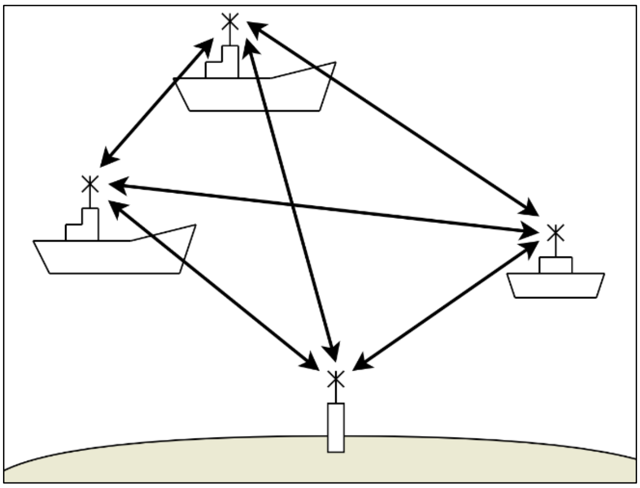

Figure 3.1 Overview of the AIS
(after IMO, 2015)

3.1.2 AIS data from the Danish Maritime Authority

The Danish Maritime Authority (n.d.) captures, processes, and publishes AIS data of vessel traffic in Danish waters and their surroundings. It is made publicly available on the basis of the Directive on the re-use of public sector information (PSI Directive) of the European Union. The historical data is published in a text file format with a delay of approximately three days. Figure 3.2 gives an indication of the magnitude of the data files, with each text file containing millions of data points. A small sample of data points from a single text file is included in appendix A.

The key elements of the Danish AIS data are a timestamp and coordinates, thereby making it spatio-temporal object information. Additional relevant elements of the Danish dataset include vessel characteristics (e.g. identification, vessel type, dimensions, cargo type) and navigational indicators from vessel sensors (e.g. speed, heading, course, rate of turn). The majority of data points in this dataset originate from class A equipment, but it also includes reports from class B units.

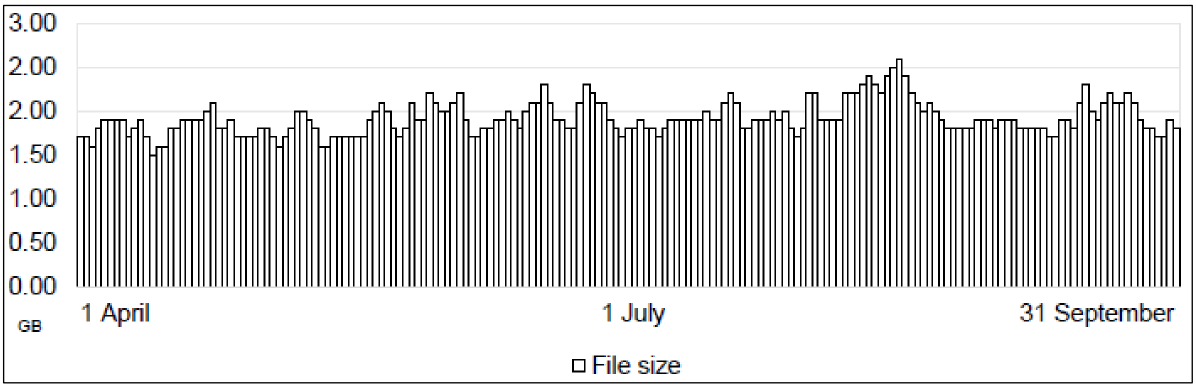

Figure 3.2 File size of AIS text files published by the DMA for timeframe 1 and timeframe 2

3.1.3 AIS data from HELCOM

HELCOM publishes shipping density raster files in the GeoTIFF format for all ships with AIS equipment, as well as separate raster files that only include data from various ship types. These raster files are publicly available without limitations, with attribution being the only requirement for reuse (HELCOM, 2018c). The HELCOM AIS system is fed by national data from Baltic Sea States, including data from Danish sources. Its spatial scope is limited to the Baltic sea. It includes the Kattegat and excludes the Skagerrak.

The shipping density raster files were created by HELCOM according to the following workflow (see HELCOM [2018c] for details). First, a total of 339 port entrance areas were established across the Baltic Sea States, as well as 5 areas representing the Baltic Sea borders. Secondly, AIS point data from the HELCOM AIS system were transformed into lines, ensuring that each line represents a unique ship movement between two ports, from an inbound crossing of the Baltic Sea borders, or towards an outbound crossing of the Baltic Sea borders. Finally, the resulting lines were overlaid on a 1 x 1 km grid where for each raster cell the number of trips crossing it was calculated and stored.

3.2 Vessel traffic analysis

The vessel traffic analysis is aimed at providing context for the subsequent main analysis of maritime risk. Results from this analysis form the basis of the answer to sub question 3.

The workflow of this analysis is shown in figure 3.3. The first step of this procedure was taking the shipping density raster files of nine ship types from HELCOM and clipping them to the Kattegat area. The resulting raster files were locally aggregated according to the grouping in table 3.1, resulting in four raster layers representing a distinct ship type (cargo, passenger, fishing, and other ships). Finally, the ship category with the highest intensity value was selected for each individual cell in order to create a predominant ship type raster map.

This final step of this analysis has been performed using the ArgStatistics algorithm in Esri's ArcGIS Pro, because this implementation of a local maximum operation was not readily available in other common GIS software.

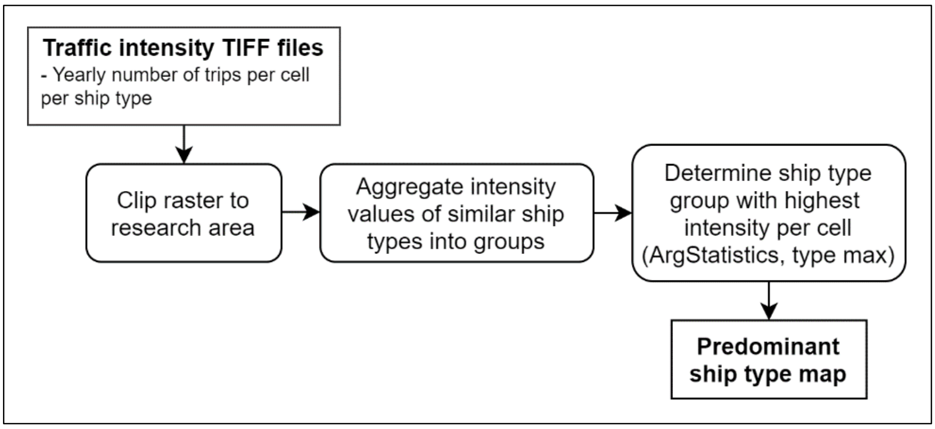

Figure 3.3 Methodological steps of the vessel traffic analysis

Table 3.1 Aggregation of ship types from HELCOM raster files into four distinct groups

Cargo	Passenger	Fishing	Other
cargo container roll-on/roll-off tanker	passenger	fishing	other service unknown

3.3 Maritime risk analysis

The maritime risk analysis is the main methodological process of this research and consists of both exploratory and confirmatory data analyses. Figure 3.4 gives a schematic overview of the maritime risk analysis. The development of this methodology contributes to answering sub question 2, while the results of this analysis contribute towards the answer to sub question 4.

The first step involved importing AIS data into a PostGIS DBMS that suits the functional requirements of subsequent spatial analyses. A process of data preparation in the database was performed to control data quality. Subsequently, the original point locations were transformed into a line segment model. The analysis continued with a main phase of analysis aimed at discovering spatial patterns of maritime risk. The procedure to create maps from AIS text files was identical for both timeframes 1 and 2. The final methodological steps included comparative mapping and the validation of results.

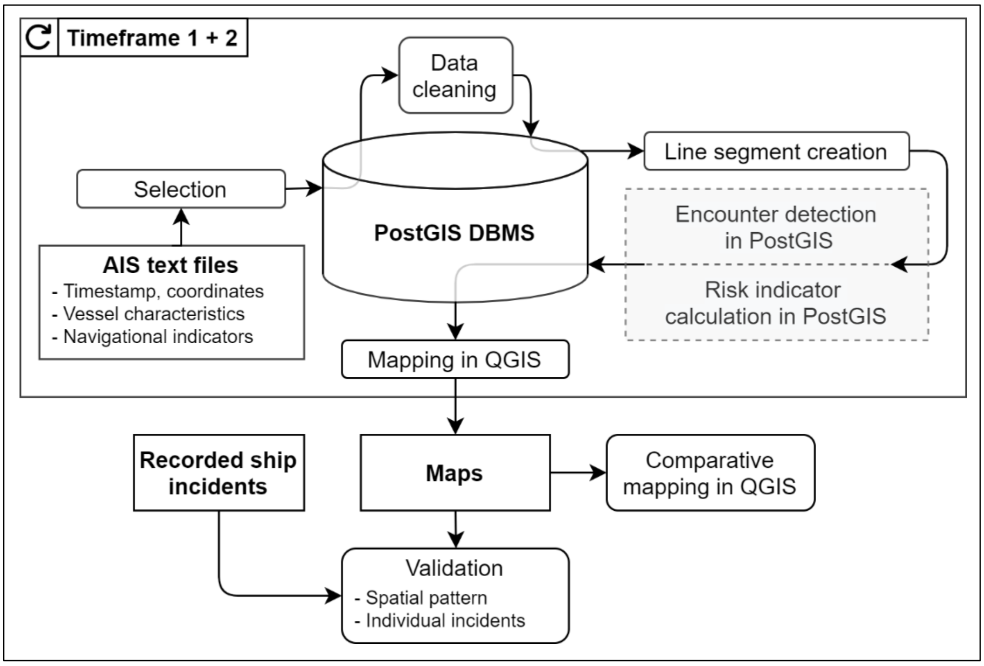

Figure 3.4 Methodological steps of the maritime risk analysis

3.3.1 Data selection and pre-processing

The first part of the maritime risk analysis involved importing AIS data into a PostgreSQL database with the PostGIS extension. Due to the limitations of consumer grade computer equipment in terms of data storage and processing power, a subsample of the available data was analysed. Only the text files of a random sample of 15 days per timeframe were selected (see table 3.2). From this selection of 40 files, only vessel movements within the research area (see figure 1.2) were analysed to reduce complexity. The data pre-processing procedure of reducing the number of data points has been scripted in Python using the pandas library. It includes attribute filters and a bounding box filter (see appendix B for scripts).

Table 3.2 Random sample of 15 days from timeframes 1 and 2 for text file data import

Timeframe 1			Timeframe 2		
2020-04-08	2020-05-05	2020-06-06	2020-07-01	2020-08-07	2020-09-05
2020-04-12	2020-05-16	2020-06-09	2020-07-03	2020-08-16	2020-09-06
2020-04-18	2020-05-24	2020-06-12	2020-07-04	2020-08-21	2020-09-13
2020-04-22	2020-05-29	2020-06-16	2020-07-17		2020-09-18
2020-04-27		2020-06-17	2020-07-18		2020-09-21
2020-04-30			2020-07-30		2020-09-29

A lack of data accuracy and consistency can be expected in AIS datasets (Felski & Jaskólski, 2012; Harati-Mokhtari et al., 2007; Zhao et al., 2018). Controlling this quality was crucial throughout the maritime risk analysis, as the accuracy of navigational indicators had a considerable influence on the detection of vessel encounters. As mentioned in section 3.1.1, AIS specifications generally stipulate that at least one message must be sent every 75 meters or 180 seconds approximately. The system achieves this by making the reporting interval dependant on speed. For this reason, line segments in the DBMS that greatly exceed these limits were deemed invalid during the data cleaning process and have been left out (see appendix B for the corresponding query).

3.3.2 DBMS design

A database design of four tables has been established in the PostGIS DBMS to store the various intermediate products of the maritime risk analysis. Figure 3.5 gives a schematic overview of this design. The points table was filled with the results of data pre-processing procedure. The segments table contained a similar number of data points as the points table, because each segment consists of a start and end point. The encounter_segments table is made up of a subset of this number of rows, as only a relatively small portion of all vessel segments represent encounter situations. This detection process, as well as the grouping of encounter segments into higher level unique encounters is discussed in section 3.3.4.

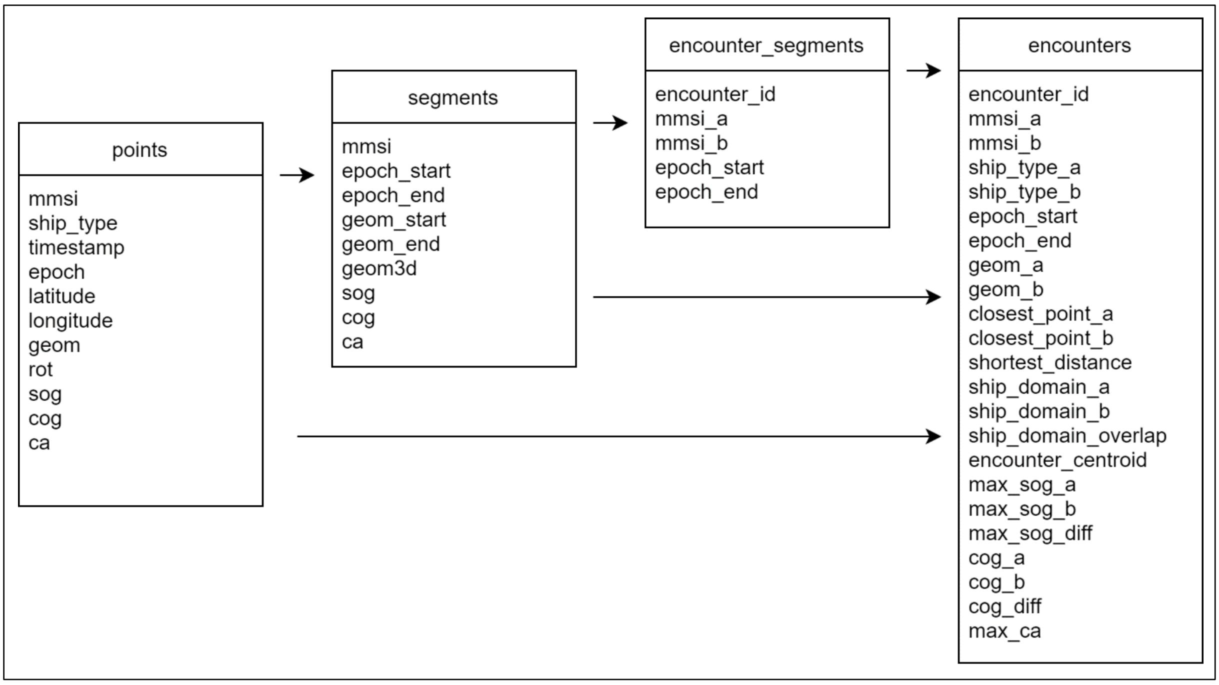

Figure 3.5 Table configuration and general flow of information of the PostGIS DBMS

3.3.3 Vessel trajectory data model

Over the course of this thesis project, four types of data models were considered to approximate a ship's trajectory based on AIS data from the DMA in a PostGIS database. They were:

- *a 2D Point model*
  A two dimensional points representation (x,y) with a timestamp attribute;
- *a basic 2D line segment model*
  A line representation, consisting of a maximum of two two-dimensional points (x,y) with a start timestamp attribute and end timestamp attribute;
- *a basic 3D line segment model*
  A line representation, consisting of a maximum of two three-dimensional points (x,y,t);
- *a full trajectory 3D line model*
  A LinestringM representation, consisting of multiple three dimensional points (x,y,t).

The choice was made to transform the discrete ship position points into a basic 3D line segment data model. The PostGIS query for this step is included in appendix B. Together, these line segments make up the full trajectory of a vessel's journey, as visualised in figure 3.6.

A line segment model has benefits over a point-based model, especially when the historical AIS data has been stored at irregular intervals (Graser, 2018). This is because a series of line segments has a more continuous nature than a series of points. The basic 3D line segment model was also beneficial for the efficiency of the DBMS during analyses, because choosing short lines as opposed to a full trajectory LinestringM data model made it possible to use multidimensional indices (see also section 2.3.2) (Graser, 2018). The implementation of the full trajectory line model has only recently been added to the PostGIS toolbox. It is therefore assumed that using the core functionality associated with the basic line representation benefits from better support in terms of its implementation and its use in other PostGIS functions.

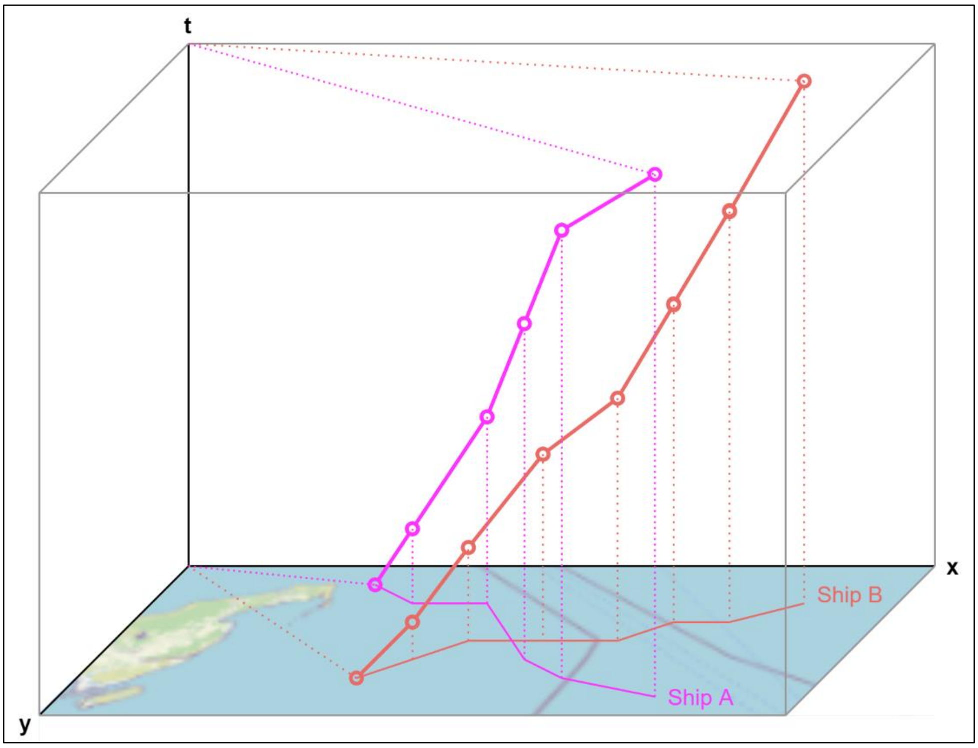

Figure 3.6 Illustrative representation of ship trajectories using the basic 3D line segment model

3.3.4 Encounter detection

The methodological process of this research project continues with the main phase of spatial analysis (light grey in figure 3.4). The steps that are involved with the encounter detection and risk indicator calculation are shown in figure 3.7. The PostGIS queries that correspond to these steps are included in appendix B.

The line segments of the vessel trajectories were the starting point of this phase of analysis. First, a three-dimensional index was created on the geometry of the line segments. This allowed for intersecting large amounts of line segments considering space and time in an efficient way (see section 2.3.2). All ship pairs that were within approximately 1,000 meter of each other and occurred at an overlapping time interval were stored in the encounter_segments table in order to guarantee the ship domains of two vessels to overlap. As an encounter of two vessels can take place over a prolonged period of time, the encounter segments were grouped based on ship identification numbers and time of the encounter. Finally, these groups were dissolved and stored as unique higher level encounters in the encounters table.

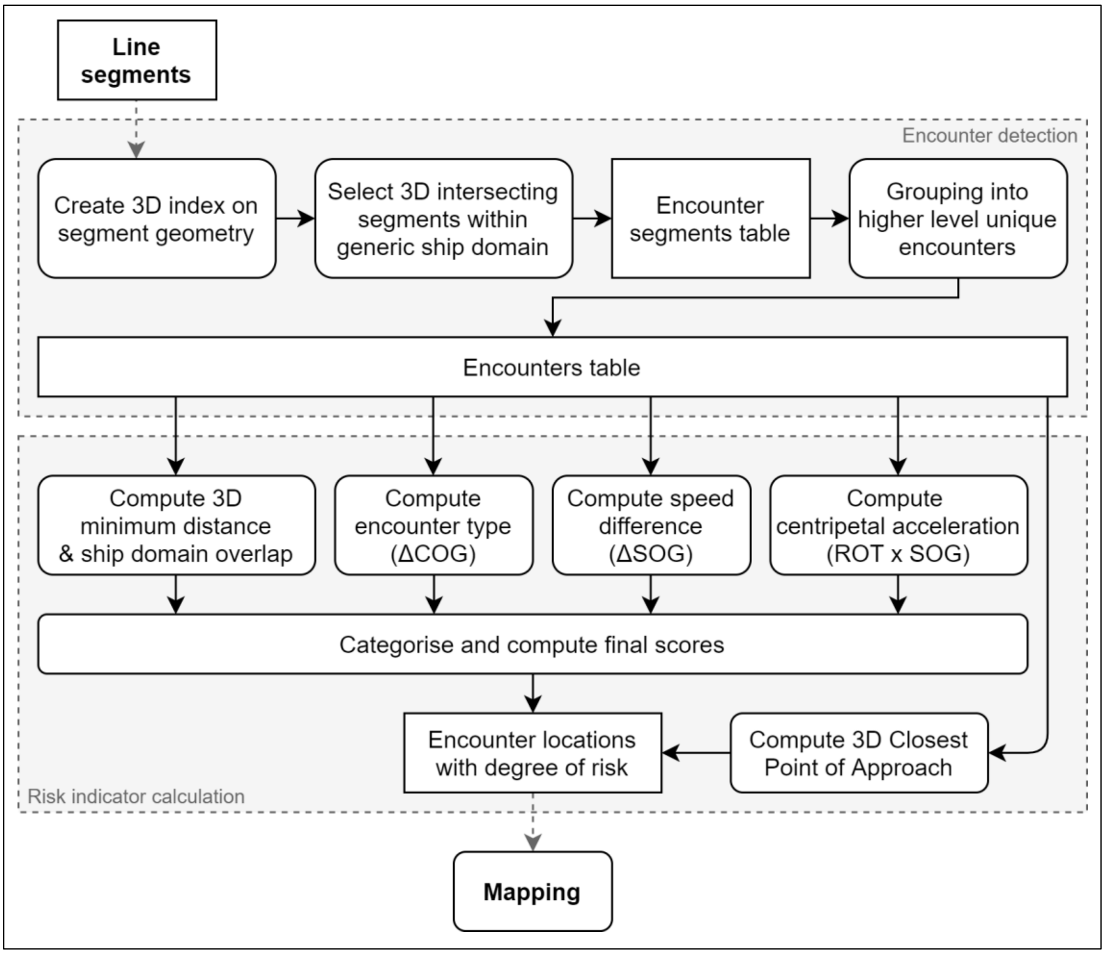

Figure 3.7 Methodological steps of the encounter detection and risk indicator calculation

3.3.5 Risk indicator calculation

A total of five risk indicators were calculated that moderate the maritime risk value of each higher level encounters in the encounters table,. This process is shown in the lower grey box in figure 3.7. The five indicators that were used in this study are based on the maritime risk factors from section 2.2.

The calculated maritime risk indicators for each encounter are presented in table 3.3. See appendix B for the full queries of these calculations. The result of the first three indicators were based on the position in time and space where the two vessels were closest. The calculations of the speed difference and centripetal acceleration also considered the 60 seconds before and after the encounter time interval for additional context.

The final step of the maritime risk analysis is combining the five individual risk indicators. The five values of each encounter in the encounter table were reduced to a final maritime risk score using min-max normalisation. The specifics of these normalisations are provided in table 3.3. As no hierarchical order among the four maritime risk factors has been identified during the literature review, all five risk indicators received equal weights. Note that the 3D minimum distance value of 3,300 meter represents the maximum distance between two ships whose ship domains just overlap. Also note that the ship domain overlap value of 5,610,000 m^2 is the area of the generic ship domain (see figure 2.4). The code of the final score calculation is available in appendix B.

Table 3.3 Normalisation of the five risk indicators

Risk indicator	Input value	Normalised value
3D minimum distance (m)	3,300 0	0 1
Ship domain overlap (m^2)	0 5,610,000	0 1
Encounter type (ΔCOG, °)	0-45 or 315-360 135-225 45-135 or 225-315	0 0.5 1
Speed difference (ΔSOG, kn)	0 30	0 1
Centripetal acceleration (CA, m/s^2)	0 2,500	0 1

3.3.6 Heatmap generation

The next step in the research process is producing heatmaps of the detected encounters and their risk value. This was done by estimating the density of encounter locations for each cell on a raster grid. The fundamental steps of this kernel density estimation are as follows:

- a surface is fitted over each point, which determines the decrease in influence of a point through space based on a kernel function;
- a point attribute is set to adjust the volume of the surface per point;
- the estimated density per grid cell equals the sum of all surface values that cover that cell.

A heatmap was generated from the detected encounters using the Kernel Density Estimation algorithm in QGIS, with a search radius of 5000 meters, a quartic kernel function, the calculated maritime risk value as the weight parameter, and an output raster size of 100 meters. The search radius parameter has been determined using Silverman's rule of thumb, because this parameter has a large impact on the kernel density estimate and can cause rings around spatial outliers if chosen incorrectly (Silverman, 1986).

This heatmap generation process has been carried out for timeframes 1 and 2. The resulting rasters were compared using a local difference operation to assess the change in maritime risk between both timeframes. An average maritime risk heatmap has also been created by combining the individual heatmaps of both timeframes. This average maritime risk heatmap serves as the backdrop for the validation of the maritime risk heatmaps.

The heatmaps are accompanied by descriptive statistics on the detected encounters and their final risk value. These statistical results were partitioned based on timeframe and ship type group. The grouping of ship types is given in table 3.4.

Table 3.4 Aggregation of ship types from DMA AIS data into four distinct groups

Cargo	Passenger	Fishing	Other
cargo	passenger	fishing	diving
tanker	high speed craft		dredging
			law enforcement
			military
			other
			pleasure
			reserved
			sailing
			spare1
			undefined

3.3.7 Validation

The average maritime risk heatmap of both timeframes was validated with the locations of recorded ship accidents in the research area from 2000 and onwards. This is the final step of the maritime risk analysis and contributes to the answer to sub question 5. These accidents records have been published by HELCOM and the IMO. Accidents that involved a collision or other forms of contact between ships have been highlighted.

4 Results

Chapter 4 presents the following results of this research project: a definition of maritime risk suitable for the Kattegat, the spatial pattern of vessel traffic in the research area, and the results of the maritime risk analysis including a validation. These results lay the foundation for answering the research questions.

4.1 Definition and operationalisation of maritime risk

In section 2.2 of this thesis document, the maritime risk concept has been conceptualised using four key factors, namely: vessel proximity (ship domain violation), the encounter type, the speed, an evasive manoeuvre. These factors have been operationalised in the following manner so that they can be deduced from Danish AIS data:

- The vessel proximity factor has been operationalised using two indicators, namely the 3D minimum distance and the ship domain overlap for two vessels at that moment in space and time. A smaller 3D minimum distance and a larger ship domain overlap increase the final maritime risk score. The generic ship domain in this research has been defined as an ellipsoid of approximately 3.3 by 1.7 kilometres (see figure 2.4). It has a decisive influence on the maritime risk compound; it has been assumed that the ship domains of two vessels in an encounter must overlap for there to be risky shipping behaviour. Section 3.3.4 explains the implementation of this condition in the encounter detection process.
- The encounter type of two vessels was deduced from the difference between the encounter angles of two vessels (ΔCOG) at the moment in space and time where the two vessel were closest. It can either be an overtaking encounter, a head-on encounter, or a crossing encounter (see table 3.3). Crossing encounters have the largest impact on the maritime risk compound, followed by head-on encounters and overtaking encounters in that order.
- The speed indicator was calculated as the maximum difference between the speed over ground (ΔSOG) between the two vessels during the encounter time interval and 60 seconds before and after this interval. The speed difference variable has a positive relation with the maritime risk compound.
- The evasive manoeuvre indicator is operationalised as the centripetal acceleration (CA) and derived by taking the product of the ROT and SOG variables. It was calculated as the maximum CA value of either of the two ships during the encounter time interval and 60 seconds before and after this interval. The CA variable entails the change in rotation around the vertical axis to which people on a ship are exposed. The centripetal acceleration variable has a positive relation with the maritime risk compound.

4.2 Spatial pattern of vessel traffic

The vessel traffic analysis using HELCOM statistical AIS data from 2016 serves provides background information about traffic flows in the research area. The results of the vessel traffic analysis are shown in figure 4.1 and figure 4.2. The north of the Kattegat acts as a funnel with large parts of the area seeing thousands of unique ship movements. The main established shipping lanes (see figure 1.1) are clearly visible in figure 4.1. Noteworthy is that the density pattern already shows thousands of yearly ship movements in the south east of the research area using the new S-route towards the Sound. Several secondary flows of traffic can be distinguished that cross the main shipping lanes. These include movements between large ports towns such as Frederikshavn and Göteborg, as well as services to and from the Danish island of Læsø. The highest concentrations of yearly trips (more than 10,000) are found at the final approach to the Göteborg archipelago.

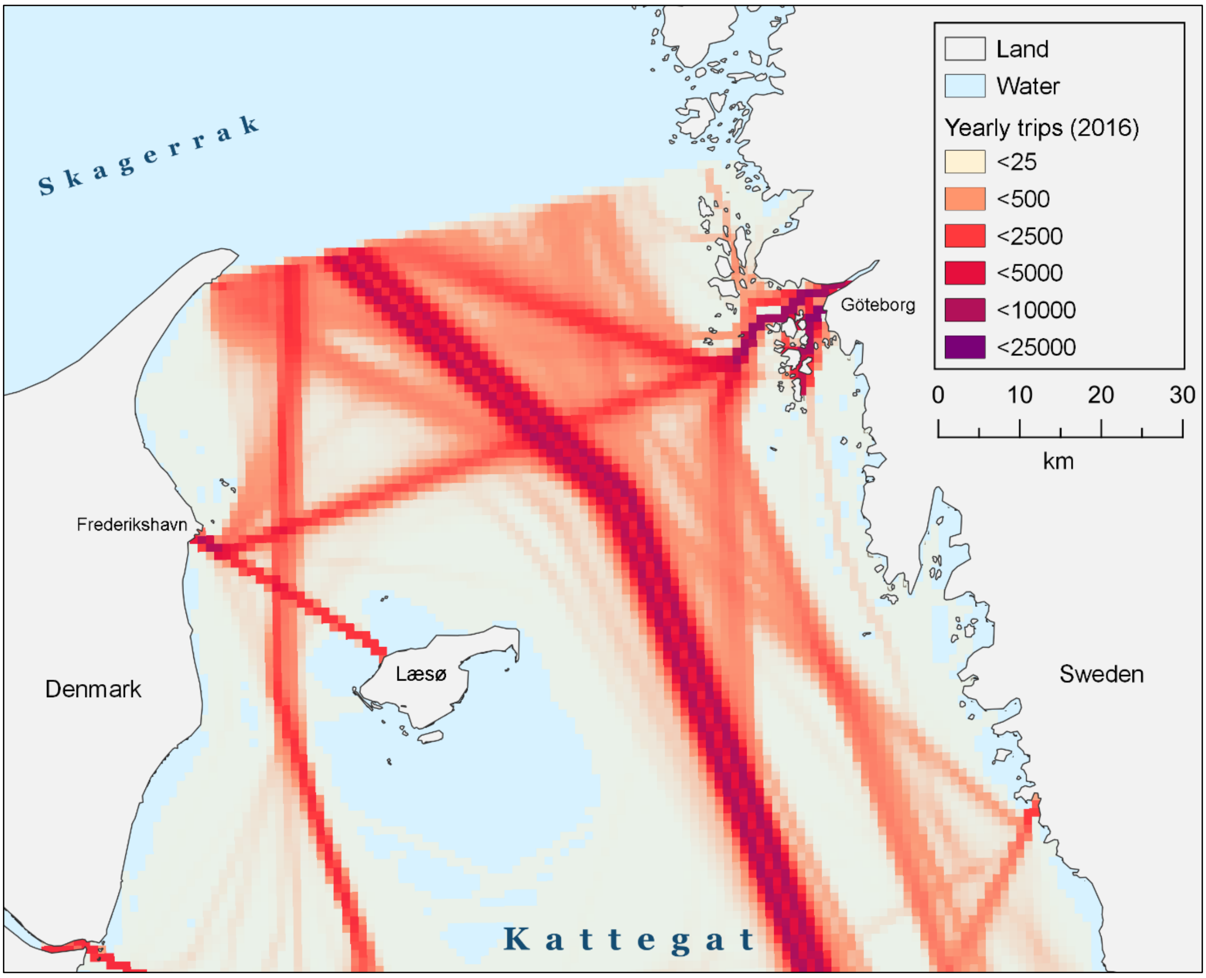

Figure 4.1 Spatial distribution of vessel density for all ship types (2016) (HELCOM)

Figure 4.2 reveals that the majority of the research area is primarily used by cargo vessels. The three large purple branches in the southern part of the Kattegat represent cargo traffic sailing towards the three straits into the Baltic sea. The predominant vessel type in the final approach to the Göteborg archipelago is also cargo. Several other patterns stand out. The flows of traffic perpendicular to the established shipping lanes are mainly made up of passenger ferry services and consist of several thousands of trips throughout the entirety of 2016. Additionally, to the northeast of the research area lies a relatively small region where fishing vessels are the predominant vessel type. Finally, some parts along the fringes of the Kattegat are mostly used by the Other category of vessel. It must be stressed that the vessel density values of these last two classes are relatively low.

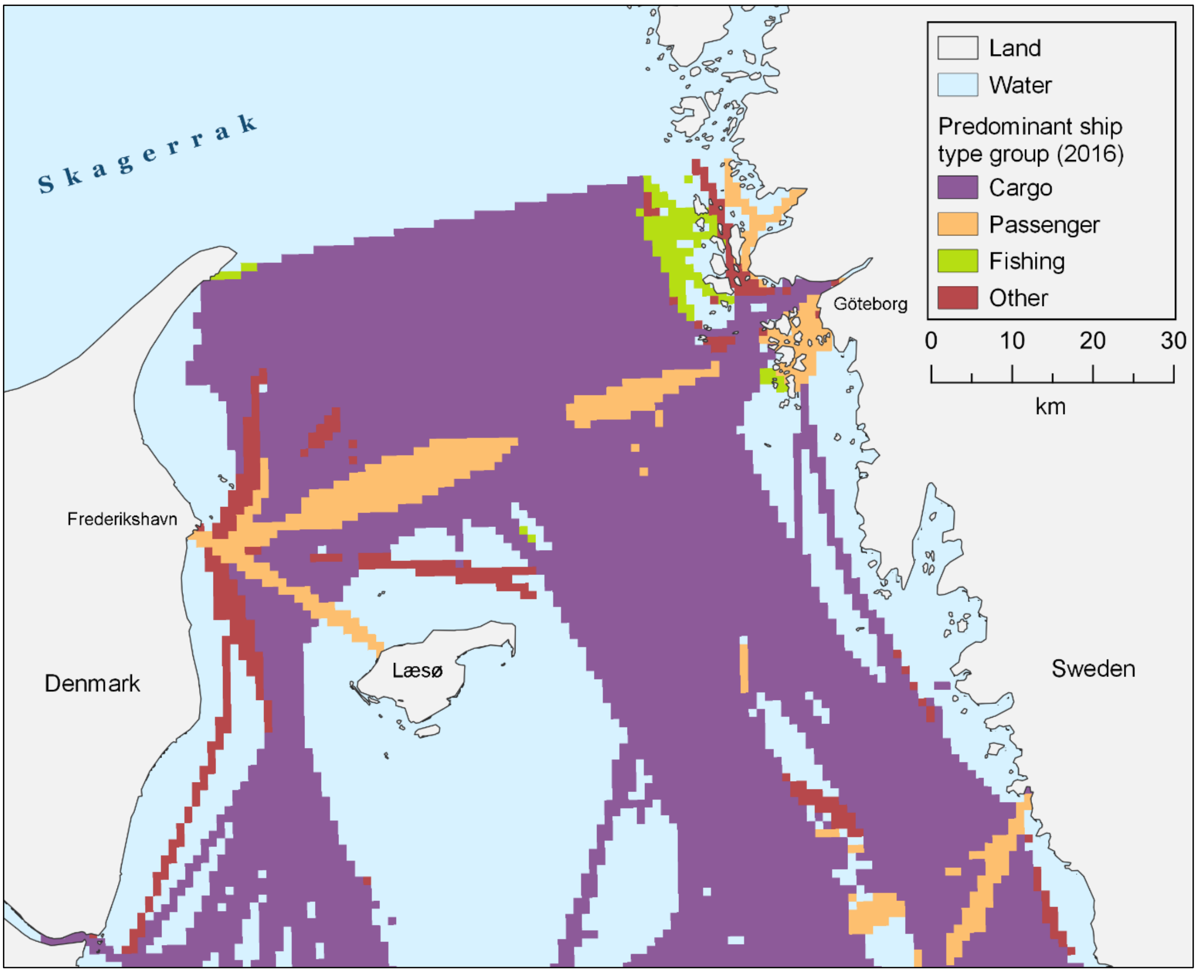

Figure 4.2 Spatial distribution of predominant ship type groups for locations with more than 25 trips per year (2016) (HELCOM)

4.3 Spatial patterns of maritime risk

The vessel movement data from the DMA for all selected days of timeframes 1 and 2 (see table 3.1) and for the research area (see figure 1.2) have been pre-processed in Python and stored in a PostGIS database. This amounts to 58.5 GB of input data. These data have been used to analyse and map maritime risk according to the methodological process that is described in section 3.3.

A total of 3,086 unique encounters were detected across both timeframes. Of this total, 1,685 encounters occurred during timeframe 1 with an average maritime risk score of 0.312. This amounts to an average of 112 encounters per day. The number of detecting timeframe 2 is 1,401 with an average final risk score of 0.320. This corresponds to average of 93 encounters per day. Thus, the number of encounters has decreased by 284 (-17%), while the average maritime risk score has increased slightly by 0.008. This trend is visualised in figure 4.3, which shows the frequency distribution of maritime risk values for timeframes 1 and 2.

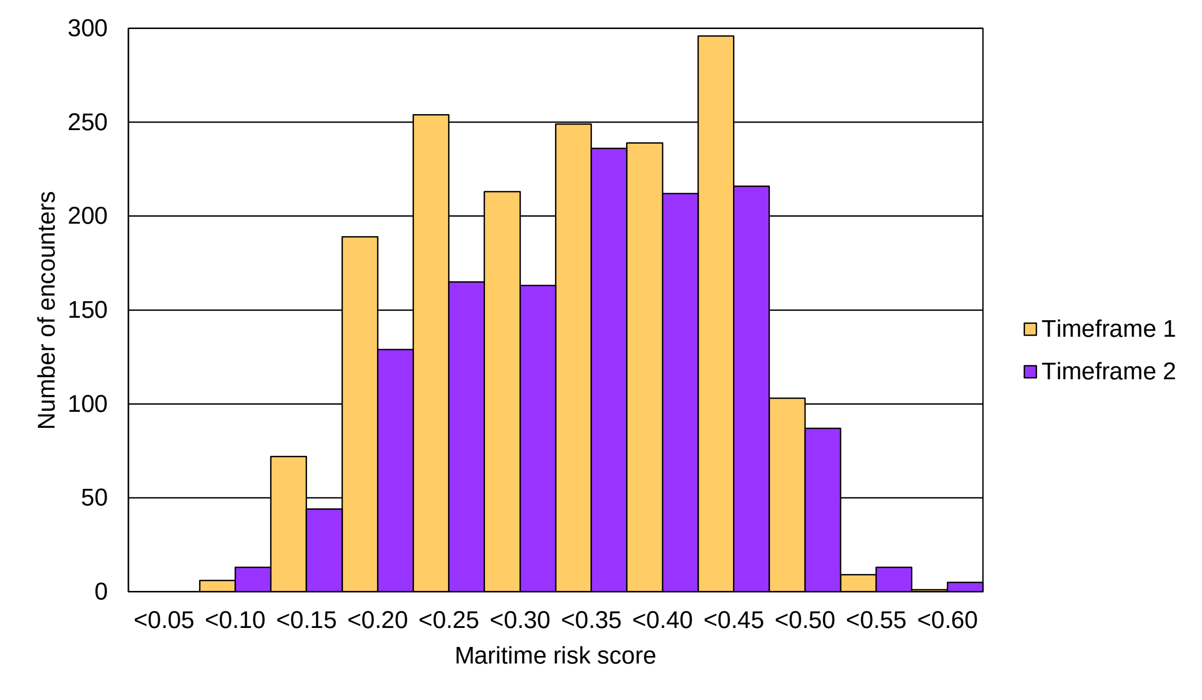

Figure 4.3 Frequency distribution of maritime risk values for timeframes 1 and 2

Table 4.1 compares the encounter and risk figures among four vessel type groups. It shows that cargo ships are most often involved in the detected encounters (75%), followed by fishing vessels (44%). Vessels that have been assigned to the passenger ship category and the Other ship category are involved in the least amount of encounters (9%). Even though relatively few passenger ship encounters are detected, these did receive the highest average risk value across both timeframes (0.358). Furthermore, the frequency of passenger ship encounters increased by 55% between timeframe 1 and 2. Encounters with fishing vessels and Other vessels too received a relatively high average risk value, namely 0.346 and 0.336 respectively. The Other vessels category saw a considerable encounter frequency increase between timeframe 1 (77 encounters) and timeframe 2 (210 encounters).

Table 4.1 Number of encounters and average risk value per vessel type for timeframes 1 and 2

	Timeframe 1		Timeframe 2	
Vessel type involved in encounter	**Number of encounters**	**Average risk value**	**Number of encounters**	**Average risk value**
Cargo	1,250	0.313	1,026	0.320
Passenger	103	0.355	160	0.361
Fishing	845	0.342	500	0.350
Other	77	0.332	210	0.350

The spatial result of the maritime risk analysis are provided as heatmaps in figures 4.4 and 4.5 for timeframes 1 and 2 respectively. The actual encounter points on which the kernel density estimation is based are included as small black dots. These maps show that most ship encounters occurred along the main shipping corridor in the centre of the research area. A distinct maritime risk hotspot lies in the north east of the research area at the approach to Göteborg, which is especially prominent in timeframe 1. Another cluster of relatively high maritime risk density can be found in the north western limits of the research area.

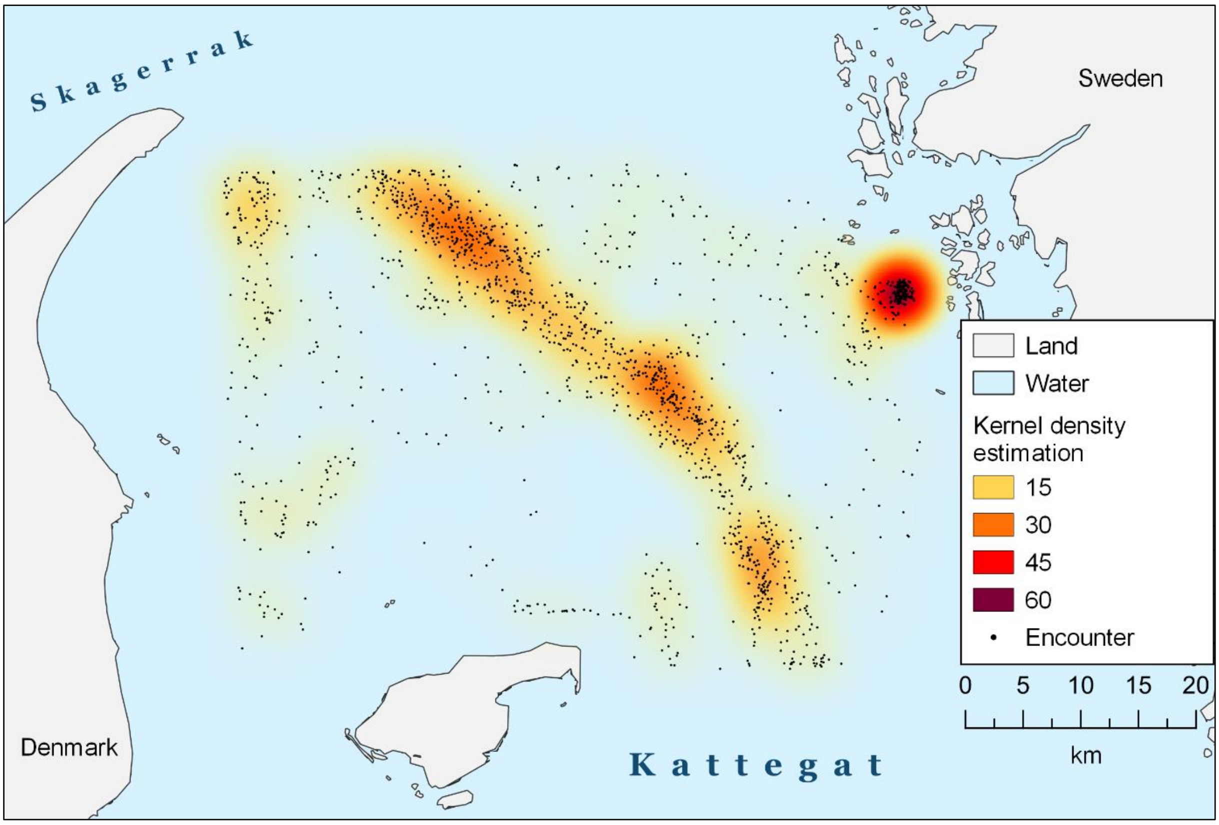

Figure 4.4 Heatmap of vessel encounters weighted by their maritime risk for timeframe 1

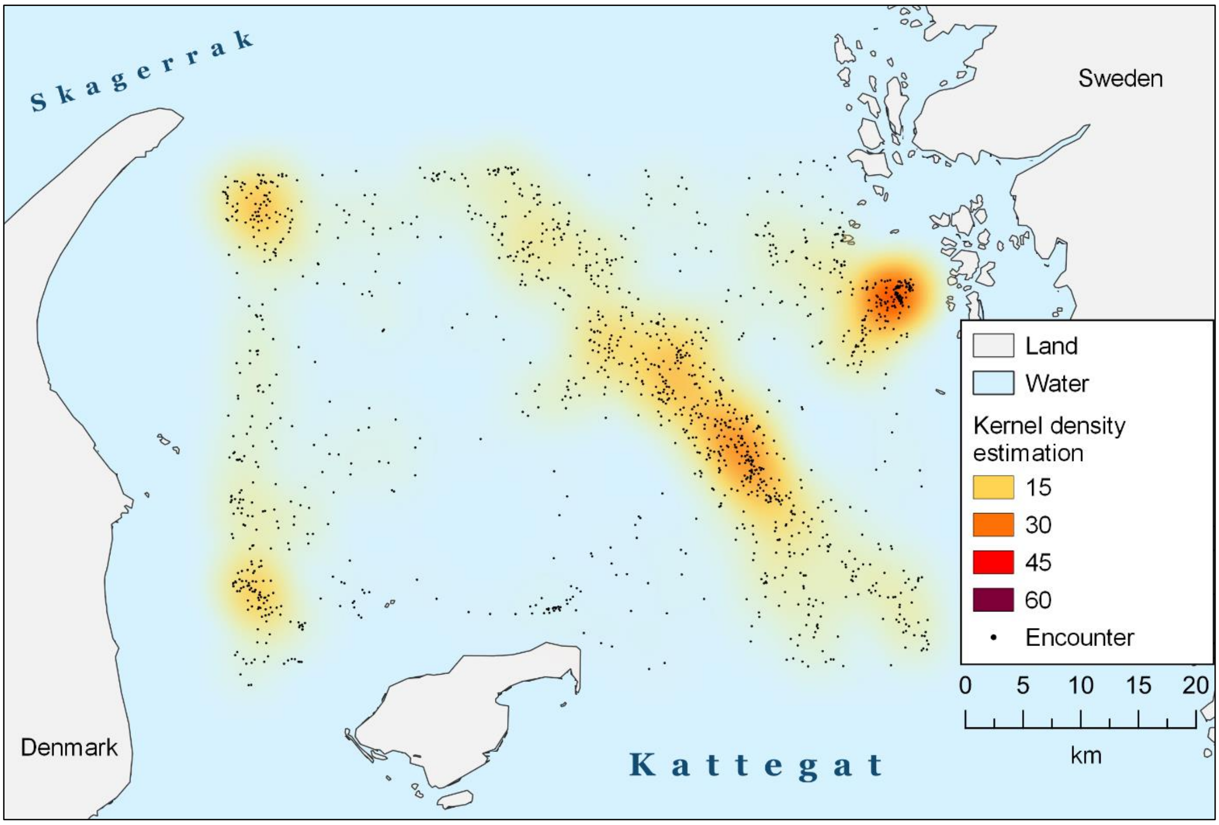

Figure 4.5 Heatmap of vessel encounters weighted by their maritime risk value for timeframe 2

A comparison of the heatmaps of figures 4.4 and 4.5 is given in figure 4.6. The first observation is that this difference map shows the general decrease in average maritime risk. Secondly, it also makes clear that the risk pattern along the north-south shipping corridor has shifted to the east. Thirdly, the green coloured areas in the north east of the research area indicate a considerable decrease in risk values at the entrance to the Göteborg archipelago.

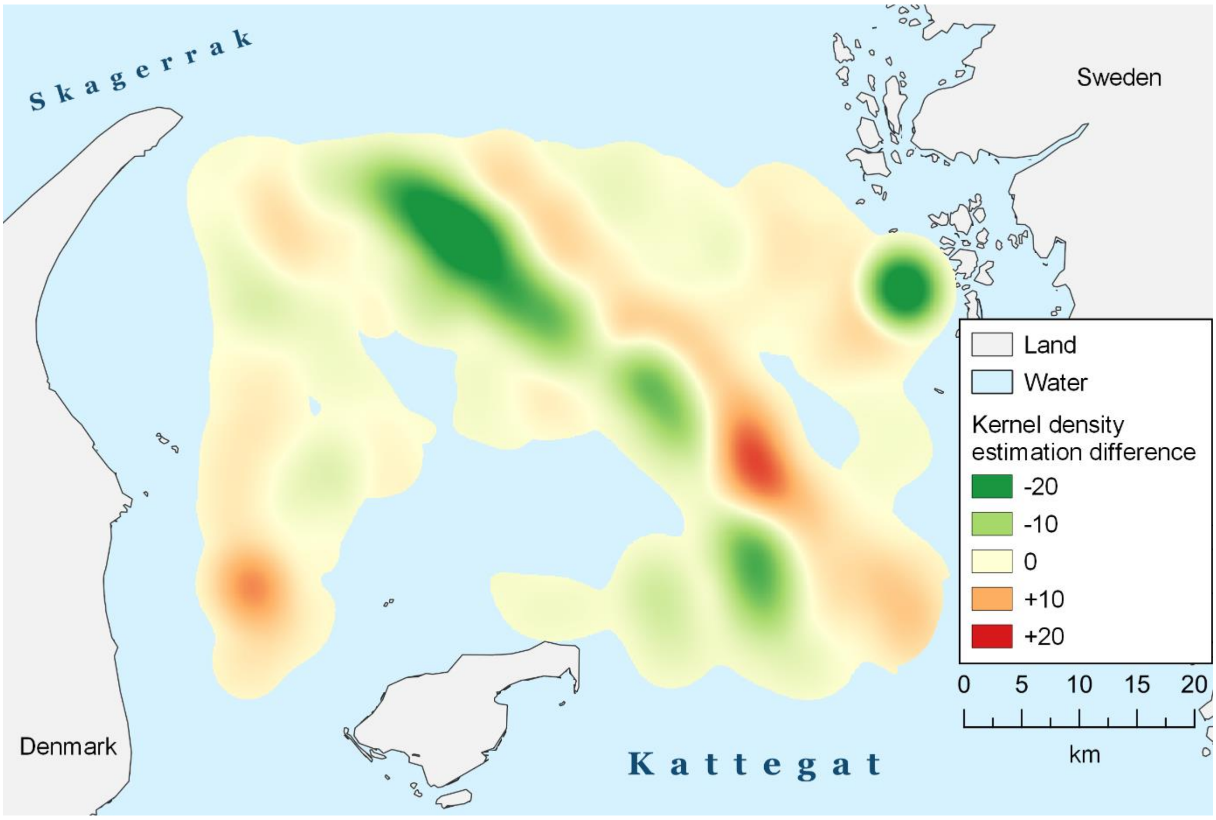

Figure 4.6 Comparison between the maritime risk heatmaps of timeframes 1 and 2

4.3.1 Validation

The average maritime risk of timeframes 1 and 2 has been combined with recorded accidents locations from 2000 and onwards in figure 4.7. It shows that the recorded accidents are spread out across the research area relatively evenly. Ergo, the overlay does not provide a pronounced interrelationship between the spatial pattern of risk and the locations of recorded accidents. However, some area of interest can be identified where a relatively high concentration of recorded accidents corresponds with a relatively high maritime risk density. These are the entrance to the Göteborg archipelago in the north east of the research area and a cluster of accidents in the north western limits of the research area.

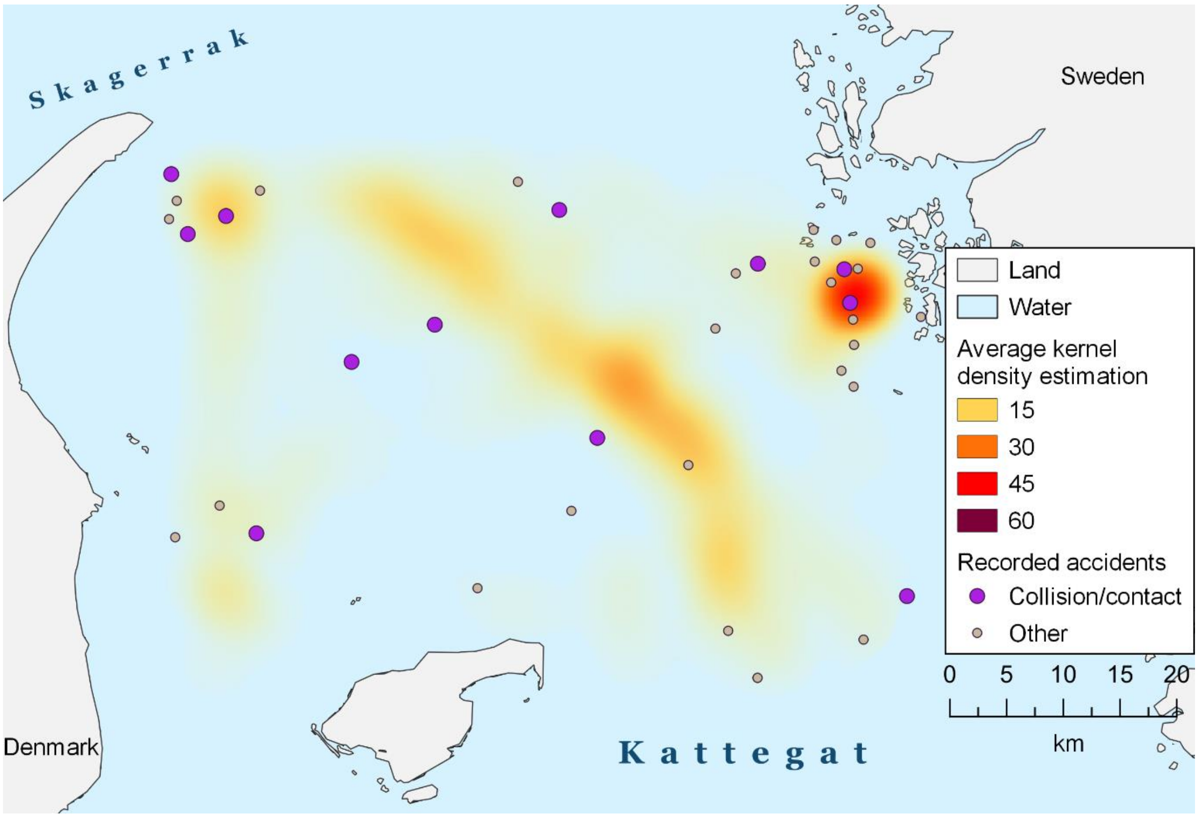

Figure 4.7 Overlay of recorded ship incidents and the average heatmap of timeframes 1 and 2

5 Discussion

The following sections reflect on the results of chapter 4 by outlining the relationship with related academic works and their implications. Influential methodological limitations related to AIS data and the analytical research process are also highlighted.

5.1 Interpretation of the results

The results of chapter 4 indicate that (1) the number of encounters decreased while maritime risk values increased slightly, (2) encounters involving passenger vessels are relatively rare but relatively risky, (3) when it comes to the spatial pattern and the vessel type, maritime risk corresponds to traffic intensity, and (4) no pronounced relation is observed between the average maritime risk pattern and recorded accidents.

From the millions of data points that have been processed, only several thousands of encounters were detected. This means that the number of records in the PostGIS tables decreased throughout the maritime risk analysis process. These observations are in line with the fundamental concepts of relational accident theory, in particular Hydén's safety pyramid (see section 2.1.2) (Hydén, 1987). However, figure 4.3 shows a right-skewed distribution of maritime risk of both timeframes. This indicates the opposite, namely that there are more high risk encounters than encounters with a maritime risk value below the average.

The results of chapter 4 also showed that most ship encounters occurred along the main shipping corridor in the centre of the research area regardless of the timeframe. The most plausible explanation for this pattern is that these areas see the most vessel movements. Correcting the estimated maritime risk density for traffic density did not result in any noticeable changes in spatial pattern. The subsequent maps are therefore not included in chapter 4. The relative risk of encounters involving passenger vessels may also be explained by the traffic pattern in the research area. As follows from the results in section 4.2, passenger vessels tend to cross the busy north-south shipping corridor which may be a source of conflict.

The modification of the T-route (see figure 1.1) is clearly visible in the shifted spatial pattern of figure 4.6. This is an indication that the flow of vessel traffic and therefore the maritime risk pattern has adapted to the changes implemented by the IMO. As part of the implementation process, a decrease in collisions and groundings of eight percent was projected (IMO, 2017b). The results of this research project show a 17% decrease in encounter frequency after the implementation. This gives reason to believe that ship routing changes per July 2020 have indeed achieved their goal, which was to "to organize the flow of ship traffic in more predictable ways to enhance the safety of navigation (IMO, 2017a, p. 1).

5.2 Research limitations

5.2.1 Generalisability of maritime risk results

The results of chapter 4 have been produced within the restraints of a limited spatial and temporal scope due to feasibility restrictions. Consequently, the results from the chosen research area may not be representative of the entire Kattegat area. Similarly, the data points of the maritime risk analysis were derived from 15 days per timeframe of 3 months. Typically, a considerably longer period of time and a larger sample size is required in order to confidently evaluate changes in spatial patterns between two timeframes.

The limited generalizability of the maritime risk results partly stems from feasibility constraints. This resulted in a DBMS with lower hardware requirements and a moderate learning curve, as well as a limited operationalisation of maritime risk. A more advanced distributed computing solution would have allowed for a considerable increase in the spatial and temporal scope of this study (see section 2.3.1). Section 5.2.4 elaborates on possible improvements to the maritime risk compound.

5.2.2 Limitations of AIS data

It must be stressed that not all vessels are fitted with AIS equipment, most notably smaller leisure craft, fishing boats, and warships (IMO, 2015). Data from the AIS may therefore not accurately reflect the marine traffic situation at a given point in time, which may have led to an underestimation of maritime risk.

In addition, the AIS data itself is known to have data quality issues (Felski & Jaskólski, 2012; Harati-Mokhtari et al., 2007; Heymann, Noack, & Banyś, 2013). Examples are missing values, inaccuracies, and incorrect formatting. The AIS does not check the content validity of its information, making it the user's responsibility (IMO, 2015). Even though some steps have been taken to mitigate such issues, there is no guarantee on the level of data quality due to the decentralised nature of the system. In practice, data quality issues that remained uncorrected may have negatively influenced the accuracy of maritime risk calculations.

As the AIS is a self-reporting system, it is also subject to spoofing (Katsilieris, Braca, & Coraluppi, 2013). Especially the content of the AIS message that is dynamic and relates to a vessel's voyage is prone to errors. For example, possible calibration errors in a ship's instruments directly affect the quality of the AIS for other ships in its vicinity, as the dynamic content of AIS messages directly originates from navigational sensor readings. A possible way of improving the consistency of dynamic navigational indicators is to recalculate them for all vessel segments solely based on the spatio-temporal components of AIS messages.

A more fundamental critique of the Automatic Identification System is that it is only one perspective on maritime risk during real vessel encounters. Other sources of navigational and safety-related information are available to mariners that may shift the notion of risk. A possible next step in the development of maritime risk assessment methods is using mixed research methods to combine vehicle tracking datasets with qualitative data on maritime safety, such as logs from vessel traffic service or primary data from mariners on the bridge.

5.2.3 Considerations on modelling and indexing trajectories

The majority of the maritime risk analysis has been performed using a PostGIS DMBS. A distinct limitation of the PostgreSQL database framework and the PostGIS extension is the lack of explicit spatio-temporal indices. A suggestion for improving PostGIS as a suitable DBMS for spatial movement data is given in Graser (2018, p. 31): "explicit spatio-temporal approaches could take advantage of the constraint that time values in trajectory objects are continuously increasing".

As mentioned in section 3.3.3, the vessel trajectories in this study are represented using three-dimensional basic line segments based on periodically reported AIS messages. Despite the benefits over a point based model, such a line segment data model may still not be a true representation of a ship's trajectory. In the example of a given vessel equipped with an AIS device that has made a turn, the ship's path in the database and the real trajectory may show considerable spatial differences, especially when the AIS reporting frequency is relatively low.

5.2.4 Improvements to the maritime risk compound

Section 4.1 presented a suitable method for maritime risk detection using AIS data from the Kattegat, which applied both flexible parameters (e.g. the centripetal acceleration) and non-flexible parameters (e.g. the ship domain). Other academics who studied maritime risk in various locations demonstrated the possibility of adjusting the ship domain based on dynamic navigational indicators (Du et al., 2020; Fiskin, Nasiboglu, & Yardimci, 2020; Goerlandt et al., 2012; Van Westrenen & Ellerbroek, 2017). Likewise, the ship domain can also be divided into sectors, each of which may have a distinctive influence on the maritime risk compound when violated. Furthermore, it may be worthwhile to expand the proposed encounter type detection process by distinguishing encounters where a given vessel crosses in front of another. Implementing one or more of these suggestions may further refine the maritime risk analysis that has been developed in this study.

As discussed in section 1.1.1, the unique environmental configuration of the Kattegat may cause unfavourable navigational conditions for ship traffic. Combining the navigational indicators of ships with additional information on their natural surroundings can add a new dimension to the maritime risk compound (Shu et al., 2013). Possible examples of additional indicators include spatial data on water levels, currents, or winds.

A final improvement is related to the parameter choices of the various function, tools, and algorithms that make up the maritime risk analysis. Efforts have been made to ground parameter settings in theory or infer them from input data. Some parameter values have been set partly on experience. Examples are the maximum distance for the initial encounter detection (see section 3.3.4), the increased time interval of higher level encounters (see section 3.3.5), the course offset during encounter type classification (see section 3.3.5), and the search radius of the kernel density estimation (see section 3.3.6). The best practice would be to subject all influential parameters in the maritime risk analysis to a sensitivity analysis.

6 Conclusion

The goal of this thesis was to study maritime risk in the Kattegat before and after the ship routing changes of July 2020. To this end, the following main research question has been posed:

> To what extent has the spatial distribution of maritime risk changed after the implementation of new shipping routes in the Kattegat on 1 July 2020 and how does this relate to the spatial pattern of vessel movements and to recorded maritime incidents?

To fully answer this main question, a total of five sub questions have been defined. An answer to the five sub questions and the main research question is formulated in section 6.1. Directions for future work on this topic are given in section 6.2.

6.1 Revisiting the research questions

1. *How can maritime risk be defined?*

The maritime risk concept has been conceptualised using four key factors, namely (1) vessel proximity (ship domain violation), (2) the encounter type, (3) speed, and (4) an evasive manoeuvre. A generic ellipsoidal ship domain of 1.7 by 3.3 kilometres has been selected from the literature. The four factors have been operationalised as five indicators, namely (1) 3D minimum distance, (2) ship domain overlap, (3) encounter type (ΔCOG), (4) speed difference (ΔSOG), and (5) centripetal acceleration (CA).

2. *How can maritime risk be analysed with spatial movement data?*

AIS point data from the DMA has been pre-processed in Python and stored in a PostGIS database. The vessel trajectories were represented using a 3D line segment data model (x,y,t). Unrealistic segments were removed during a data cleaning process to control the lack of AIS data quality. A three-dimensional index has been generated to efficiently intersect segments in space and time. The resulting encounter locations and their risk values were mapped in QGIS using kernel density estimation.

3. *What is the spatial pattern of vessel movements?*

Using HELCOM statistical AIS data from 2016, three high density flows of predominantly cargo vessel traffic have been identified. Several secondary traffic flows of ferry passenger services can be distinguished that cross the main shipping lanes. The approach to the Göteborg archipelago has the highest vessel density of the research area.

4. *To what extent has the spatial pattern of maritime risk changed after the implementation of new shipping routes?*

The results of this research project show a 17% decrease in encounter frequency across the entire research area after the implementation of the ship routing changes of July 2020. The average risk value increased slightly. A pattern of relatively high maritime risk densities along the main north-south shipping corridor has shifted to the east.

5. *To what extent does the spatial pattern of maritime risk relate to recorded maritime incidents and the spatial pattern of vessel movement?*

The results do not show a pronounced relationship between the spatial pattern of maritime risk and the locations of recorded accidents. The results do indicate that most ship encounters occurred along the main north-south shipping corridor in the centre of the research area, as these areas see the most vessel movements. Correcting the estimated maritime risk density for traffic density did not lead to a change in spatial pattern.

Main research question
A 17% decrease in encounter frequency has been observed across the entire research area after the implementation of the ship routing changes of July 2020, which is in line with earlier projections. The average risk value increased slightly. Most ship encounters occurred along the main north-south shipping corridor in the centre of the research area and involved cargo vessels, as these areas see the most vessel movements and cargo is the predominant ship type in the research area. This pattern shifted to the east, which is most likely caused by the modification of the T-route and the corresponding shift in traffic density. Encounters involving passenger vessels have relatively high risk values, which may be due to their crossing of the busy north-south shipping corridor.

The result do not show a pronounced relationship between the spatial pattern of maritime risk and the locations of recorded accidents. Correcting the estimated maritime risk density for traffic density also did not lead to new insights.

6.2 Future work

The most prominent way to extended this research project would be to increase its generalizability. Firstly, the generalizability may be improved by including the ship movements for the entire Kattegat area and to increase the size of both intervals. A potential requirement for taking this step is explicit support for storing and querying large amounts of spatio-temporal data in mainstream database systems. Secondly, several improvements to the maritime risk compound can be made. These include the introduction of a dynamic ship domain, the inclusion of risk factors from the natural environmental, and an examination on the sensitivity of influential maritime risk assessment parameters. Exploring these avenues in future research is relevant, because it may develop a more accurate assessment of maritime risk in the Kattegat.

References

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

Bakli, M., Sakr, M., & Zimanyi, E. (2020). Distributed Mobility Data Management in MobilityDB. *Proceedings - IEEE International Conference on Mobile Data Management*, *2020-June*, 238–239. https://doi.org/10.1109/MDM48529.2020.00052

Danish Maritime Authority (DMA) (n.d.). *AIS data management policy*. Retrieved from https://www.dma.dk/SikkerhedTilSoes/Sejladsinformation/AIS/Sider/AIS%20datamanagementpolitik.aspx

Danish Maritime Authority (2020a). *New Shipping Routes in Kattegat and Skagerrak*. Retrieved from https://www.dma.dk/SikkerhedTilSoes/Sejladsinformation/RuterKattegatSkagerrak/Sider/default.aspx

Danish Maritime Authority (2020b). *Passage lines*. Retrieved from https://www.dma.dk/SikkerhedTilSoes/Sejladsinformation/AIS/Sider/Passagelinjer.aspx

Dodge, S., Weibel, R., Ahearn, S. C., Buchin, M., & Miller, J. A. (2016). Analysis of movement data. *International Journal of Geographical Information Science 30*(5), 825–834). https://doi.org/10.1080/13658816.2015.1132424

Du, L., Goerlandt, F., & Kujala, P. (2020). Review and analysis of methods for assessing maritime waterway risk based on non-accident critical events detected from AIS data. *Reliability Engineering and System Safety*, *200*, 1–23. https://doi.org/10.1016/j.ress.2020.106933

European Maritime Safety Agency (EMSA) (2019). *Annual Overview of Marine Casualties and Incidents 2019*. European Maritime Safety Agency. Retrieved from http://www.emsa.europa.eu/we-do/safety/accident-investigation/items.html?cid=141&id=3734

Federal Maritime and Hydrographic Agency (BSH) (2016). *Shipping in the Baltic Sea - Past, present and future developments relevant for Maritime Spatial Planning*. Federal Maritime and Hydrographic Agency. Retrieved from https://vasab.org/project/balticlines/project-outputs/

Felski, A., & Jaskólski, K. (2012). Information unfitness as a factor constraining Automatic Identification System (AIS) application to anti-collision manoeuvring. *Polish Maritime Research*, *19*(3), 60–64. https://doi.org/10.2478/v10012-012-0032-4

Fiskin, R., Nasiboglu, E., & Yardimci, M. O. (2020). A knowledge-based framework for two-dimensional (2D) asymmetrical polygonal ship domain. *Ocean Engineering*, *202*. https://doi.org/10.1016/j.oceaneng.2020.107187

Fujii, Y., & Tanaka, K. (1971). Traffic Capacity. *The Journal of Navigation*, *24*(4), 543–552.

Goerlandt, F., Montewka, J., Lammi, H., & Kujala, P. (2012). Analysis of near collisions in the Gulf of Finland. In Bérenguer, C., Grall, A, & Guedes Soares, C. (Eds.), *Advances in Safety, Reliability and Risk Management* (pp. 2880–2886). London: Taylor & Francis.

Graser, A. (2018). Evaluating Spatio-temporal Data Models for Trajectories in PostGIS Databases. *GI_Forum*, *1*, 16–33. https://doi.org/10.1553/giscience2018_01_s16

Graser, A., & Dragaschnig, M. (2020). Open Geospatial Tools for Movement Data Exploration. *KN - Journal of Cartography and Geographic Information*, *70*(1), 3–10. https://doi.org/10.1007/s42489-020-00039-y

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

Grimvall, A., & Larsson, K. (2014). *Mapping shipping intensity and routes in the Baltic Sea*. Swedish Institute for the Marine Environment. Retrieved from https://havsmiljoinstitutet.se/publikationer/rapportserie/2014/2014-5-mapping-shipping-intensity-and-routes-in-the-baltic-sea

Güting, R. H., Böhlen, M. H., Erwig, M., Jensen, C. S., Lorentzos, N. A., Schneider, M., & Vazirgiannis, M. (2000). A Foundation for Representing and Querying Moving Objects. In *ACM Transactions on Database Systems, 25*(1), 1-42. https://doi.org/10.1145/352958.352963

Hänninen, M., & Kujala, P. (2014). Bayesian network modeling of Port State Control inspection findings and ship accident involvement. *Expert Systems with Applications, 41*(4), 1632–1646. https://doi.org/10.1016/j.eswa.2013.08.060

Harati-Mokhtari, A., Wall, A., Brooks, P., & Wang, J. (2007). Automatic identification system (AIS): Data reliability and human error implications. *Journal of Navigation*, *60*(3), 373–389. https://doi.org/10.1017/S0373463307004298

HELCOM (Baltic Marine Environment Protection Commission) (2018a). *Report on shipping accidents in the Baltic Sea from 2014 to 2017*. HELCOM. Retrieved from https://helcom.fi/media/publications/Report-on-shipping-accidents-in-the-Baltic-Sea-from-2014-to-2017.pdf

HELCOM (2018b). *Operational oil spills from ships - HELCOM core indicator report*. HELCOM. Retrieved from https://helcom.fi/baltic-sea-trends/holistic-assessments/state-of-the-baltic-sea-2018/reports-and-materials/

HELCOM (2018c). Maritime activities in the Baltic Sea. *Baltic Sea Environment Proceedings No.152*. HELCOM. Retrieved from https://www.helcom.fi/wp-content/uploads/2019/08/BSEP152-1.pdf.

Heymann, F., Noack, T., & Banyś, P. (2013). Plausibility analysis of navigation related AIS parameter based on time series. *ENC2013, Wien*. Retrieved from https://elib.dlr.de/84166/

Hörteborn, A., Ringsberg, J. W., Svanberg, M., & Holm, H. (2019). A Revisit of the Definition of the Ship Domain based on AIS Analysis. *Journal of Navigation*, *72*(3), 777–794. https://doi.org/10.1017/S0373463318000978

Hydén, C. (1987). *The development of a method for traffic safety evaluation: The Swedish Traffic Conflicts Technique (Dissertation).*

International Maritime Organization (IMO) (2015). *Revised Guidelines for the Onboard Operational Use of Shipborne Automatic Identification Systems (AIS), Resolution A.1106(29)*. International Maritime Organization.

International Maritime Organization (2017a). Routeing Measures and Mandatory Ship Reporting Systems - General overview for establishment of traffic separation scheme and other routeing measures in the vicinity of Kattegat between Denmark and Sweden, *NCSR 5/3/3*. International Maritime Organization.

International Maritime Organization (2017b). *Routeing Measures and Mandatory Ship Reporting Systems - Report on the sea traffic and consequence analysis and IWRAP Mk2 analysis related to proposals for new routeing measures in the vicinity of Kattegat between Denmark and Sweden, NCSR 5/INF. 3*. International Maritime Organization.

International Transport Forum (2019). *ITF Transport Outlook 2019*. Paris: OECD Publishing. https://doi.org/10.1787/transp_outlook-en-2019-en

Katsilieris, F., Braca, P., & Coraluppi, S. (2013). Detection of malicious AIS position spoofing by exploiting radar information. *Proceedings of the 16th international conference on information fusion*, 1196-1203.

Kerbrat, A. (2018). Traffic Rules and Environmental Conditions in Kattegat and the Sound Regarding Changes Planned for 2020. *Naukowe Akademii Morskiej w Gdyni*, *107*, 59–71. https://doi.org/10.26408/107.04

Li, G., Weng, J., & Fu, S. (2020). Bootstrap-Tobit model for maritime accident economic loss considering underreporting issues. *Transportmetrica A: Transport Science, 1-22* https://doi.org/10.1080/23249935.2020.1829169

Lunde Hermansson, A., & Hassellöv, I.-M. (2020). *Tankrengöring och dess påverkan på havsmiljön*. Swedish Institute for the Marine Environment. Retrieved from https://havsmiljoinstitutet.se/publikationer/rapportserie/2020/2020-6-tankrengoring-och-dess-paverkan-pa-havsmiljon

Mestl, T., Tallakstad, K. T., & Castberg, R. (2016). Identifying and Analyzing Safety Critical Maneuvers from High Resolution AIS Data. *TransNav, the International Journal on*

Marine Navigation and Safety of Sea Transportation, *10*(1), 69–77. https://doi.org/10.12716/1001.10.01.07

Milne, D., & Watling, D. (2019). Big data and understanding change in the context of planning transport systems. *Journal of Transport Geography*, *76*, 235–244. https://doi.org/10.1016/j.jtrangeo.2017.11.004

Qu, X., Meng, Q., & Suyi, L. (2011). Ship collision risk assessment for the Singapore Strait. *Accident Analysis and Prevention*, *43*(6), 2030–2036. https://doi.org/10.1016/j.aap.2011.05.022

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. *IWNTM13: International Workshop on Nautical Traffic Models 2013, Delft, The Netherlands*. Retrieved from https://repository.tudelft.nl/islandora/object/uuid%3A869a23d3-ce82-4f17-8689-79cfff1696f9

Silveira, P. A. M., Teixeira, A. P., & Soares, C. G. (2013). Use of AIS data to characterise marine traffic patterns and ship collision risk off the coast of Portugal. *Journal of Navigation*, *66*(6), 879–898. https://doi.org/10.1017/S0373463313000519

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall.

SSPA Sweden (2018). Maritime spatial planning and ship routeing systems. SSPA Sweden. Retrieved from https://www.sspa.se/safety-security-and-risk/maritime-spatial-planning-ship-routeing-systems

Szlapczynski, R., & Szlapczynska, J. (2017). Review of ship safety domains: Models and applications. *Ocean Engineering*, *145*, 277–289. https://doi.org/10.1016/j.oceaneng.2017.09.020

Theunissen, E., & de Groot, C. A. (2014). Conflict detection and resolution for semi-autonomous vessels. *ENC-GNSS 2014*. https://www.researchgate.net/publication/294871642

Tran, N. K., & Haasis, H. D. (2015). An empirical study of fleet expansion and growth of ship size in container liner shipping. *International Journal of Production Economics*, *159*, 241–253. https://doi.org/10.1016/j.ijpe.2014.09.016

Van Oosterom, P. (2005). Spatial access methods. In Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (Eds.), *Geographic Information Systems: Principles, Techniques, Applications, and Management* (pp. 385–400). Hoboken, NJ: Wiley.

van Westrenen, F., & Ellerbroek, J. (2017). The Effect of Traffic Complexity on the Development of Near Misses on the North Sea. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *47*(3), 432–440. https://doi.org/10.1109/TSMC.2015.2503605

Zhao, L., Shi, G., & Yang, J. (2018). Ship Trajectories Pre-processing Based on AIS Data. *Journal of Navigation*, *71*(5), 1210–1230. https://doi.org/10.1017/S0373463318000188

Appendices

Appendix A Sample of AIS text files published by the DMA

Timestamp	Type of mobile	MMSI	Latitude	Longitude	Navigational status	ROT	SOG	COG	Heading	IMO	Callsign	Name	Ship type	Cargo type	Width	Length	Type of position fixing device	Draught	Destination	ETA	Data source type	Size A	Size B	Size C	Size D
01-07-20 02:46	Base Station	2190071	57.11003	8.64828	Unknown value					…	…		Undefined	…			…		…	…	AIS				
01-07-20 07:46	Class A	305425000	55.70783	10.78516	Under way using engine	0	14.2	187.0	187.0	…	…	BULKNES	Cargo	…	260	1760	…	105	…	…	AIS	1480	280	80	180
01-07-20 20:56	Class A	209681000	54.77408	10.26869	Under way using engine	0	6.8	331.1	328.0	…	…	RIX EXPLORER	Cargo	…	130	890	…	54	…	…	AIS	750	140	20	110
01-07-20 01:10	Class A	740339000	57.47261	8.70183	Under way using engine	0	11.2	57.0	57.0	…	…	JAMES CLARK ROSS	Other	…	180	1000	…	62	…	…	AIS	440	560	50	130
01-07-20 05:45	Class A	220349000	57.71544	10.58469	Engaged in fishing	0	0		47.0	…	…	LADY GAGA S222	Fishing	…	50	180	…		…	…	AIS	140	40	30	20
01-07-20 01:05	Class B	211651370	55.13961	10.78599	Unknown value		0			…	…	KIA KAHA	Pleasure	…	40	120	…		…	…	AIS	40	80	20	20
01-07-20 14:12	Class A	219005901	56.44810	7.83104	Under way using engine		6.0	154.5	213.0	…	…	L70 LIMEA	Fishing	…	40	120	…	20	…	…	AIS	30	90	20	20
01-07-20 08:50	Class A	220138000	56.69792	8.21792	Engaged in fishing	0	0	197.0	162.0	…	…	MICHELLE-JASMINE	Fishing	…	50	260	…	70	…	…	AIS	160	100	30	20
01-07-20 01:34	Class A	219650000	56.70120	8.22251	Engaged in fishing	0	0	165.3	46.0	…	…	L232 TOVE KYNDE	Fishing	…			…	27	…	…	AIS				
01-07-20 08:58	Class A	219001819	55.61079	12.50170	Under way using engine	0	0	193.6	276.0	…	…	LAPIS	Dredging	…	140	660	…	36	…	…	AIS	570	90	40	100

Appendix B Collection of Python scripts and PostGIS queries for data pre-processing, encounter detection and risk indicator calculation (also available on GitHub)

```
# -*- coding: utf-8 -*-

# --- Vessel traffic analysis ---
# determine ship type group with highest value per cell (ArcGIS)
from arcpy.sa import *
from arcpy.management import *

agg_max = ArgStatistics(rasters=[Raster(r"agg_cargo"),Raster(r"passenger"),Raster(r"fishing"),Raster(r"agg_other")],stat_type="max",multiple_occurrence_value=99)
CopyRaster(agg_max, r"agg_max")

# --- Maritime risk analysis ---
# random samples of three months before and three months after 1 July 2020
import pandas as pd
import random
sample_days_before = sorted(random.sample([d.strftime("%Y-%m-%d") for d in pd.date_range('2020-04-01','2020-06-30')], 15))
sample_days_after = sorted(random.sample([d.strftime("%Y-%m-%d") for d in pd.date_range('2020-07-01','2020-09-30')], 15))

# preprocess daily AIS text files from DMA for import in postgis
import os
import pandas as pd

def process_aisdk(file_input_path):
    with open(file_input_path) as file_input:
        df = pd.read_csv(file_input, usecols=['# Timestamp', 'Type of mobile', 'MMSI', 'Latitude', 'Longitude', 'Navigational status', 'ROT', 'SOG', 'COG', 'Ship type'])

        # bounding box filters
        df = df[((df['Latitude'] >= 57.3) & (df['Latitude'] <= 57.7)) & ((df['Longitude'] >= 10.7) & (df['Longitude'] <= 11.7))]
        df = df.loc[~((df['Latitude'] >= 57.6) & (df['Longitude'] >= 11.6))]
        df = df.loc[~((df['Latitude'] <= 57.35) & ((df['Longitude'] >= 10.8) & (df['Longitude'] <= 11.2)))]
        # exclude messages with these 'stationary' statuses
        df = df[~df['Navigational status'].isin(['Moored', 'At anchor', 'Aground'])]
        # only include messages from class A or class B equipment
        df = df[df['Type of mobile'].isin(['Class A', 'Class B'])]
        # exclude messages from vessels that are close to other vessels by definition
        df = df[~df['Ship type'].isin(['Pilot', 'Tug', 'Towing', 'Towing long/wide'])]


    output_path = os.path.splitext(file_input_path)[0] + "_processed" + os.path.splitext(file_input_path)[1]
    with open(output_path, 'w', newline='\n') as file_output:
        df.to_csv(file_output, columns = ['# Timestamp', 'MMSI', 'Latitude', 'Longitude', 'ROT', 'SOG', 'COG', 'Ship type'], index=False, header=False)
```

```sql
--- Maritime risk analysis ---
-- points table creation
CREATE TABLE points
(timestamp TIMESTAMP, mmsi CHAR(9), latitude FLOAT, longitude FLOAT, rot FLOAT, sog FLOAT, cog FLOAT, ship_type CHAR(30));

-- csv import
-- set correct file permission > https://ourtechroom.com/tech/importing-csv-file-in-postgresql-table/
COPY points FROM '<file>' DELIMITER ',' CSV;

-- timestamp conversion
ALTER TABLE points ADD COLUMN epoch INTEGER;
UPDATE points SET epoch = (EXTRACT(EPOCH FROM timestamp));

-- centripetal acceleration calculation
ALTER TABLE points ADD COLUMN ca FLOAT;
UPDATE points SET ca = (rot * sog);

-- point geometry conversion
SELECT AddGeometryColumn ('public','points','geom',4326,'POINT',2);
UPDATE points SET geom = (SELECT ST_SetSRID((ST_MakePoint(longitude, latitude)), 4326));

-- points table index creation
CREATE INDEX idx_points_mmsi ON points(mmsi);
CREATE INDEX idx_points_epoch ON points(epoch);
CREATE INDEX idx_points_geom ON points USING GIST (geom);

-- segments table creation
CREATE TABLE segments (mmsi CHAR(9), epoch_start INTEGER, epoch_end INTEGER, ca FLOAT, sog FLOAT, cog FLOAT);
SELECT AddGeometryColumn ('public','segments','geom_start',4326,'POINT',2);
SELECT AddGeometryColumn ('public','segments','geom_end',4326,'POINT',2);

-- segments creation
WITH cte AS (
    SELECT
        mmsi,
        LAG(epoch, 1, NULL) OVER (PARTITION BY mmsi ORDER BY mmsi, epoch) AS epoch_start,
        epoch AS epoch_end,
        GREATEST(ABS(LAG(ca, 1, NULL) OVER (PARTITION BY mmsi ORDER BY mmsi, epoch)), ABS(ca)) as ca,
        GREATEST(ABS(LAG(sog, 1, NULL) OVER (PARTITION BY mmsi ORDER BY mmsi, epoch)), ABS(sog)) as sog,
        (ABS(LAG(cog, 1, NULL) OVER (PARTITION BY mmsi ORDER BY mmsi, epoch)) + ABS(cog)) / 2 as cog,
        LAG(geom, 1, NULL) OVER (PARTITION BY mmsi ORDER BY mmsi, epoch) AS geom_start,
        geom AS geom_end
    FROM points
    )
INSERT INTO segments SELECT * FROM cte WHERE (epoch_start IS NOT NULL) AND (geom_start IS NOT NULL);

-- transform segments to 3D linestring model
SELECT AddGeometryColumn ('public','segments','geom3d',4326,'LINESTRINGZ',3);
UPDATE segments SET geom3d = ST_SetSRID(ST_MakeLine(
    ST_MakePoint(ST_X(geom_start), ST_Y(geom_start), epoch_start),
    ST_MakePoint(ST_X(geom_end), ST_Y(geom_end), epoch_end)
    ), 4326);

-- remove unrealistic segments
DELETE FROM segments WHERE (ST_Length(ST_Transform(geom3d, 25832)) > 1000) OR (epoch_end - epoch_start > 500);
```

```
-- segments table index creation
CREATE INDEX idx_segments_geom3d ON segments USING GIST (geom3d gist_geometry_ops_nd);
CREATE INDEX idx_segments_mmsi ON segments(mmsi);

-- encounter segments table creation
CREATE TABLE encounter_segments (mmsi_a CHAR(9), mmsi_b CHAR(9), epoch_start INTEGER, epoch_end INTEGER);

-- selecting encounter segments using 3D intersection (lat,lon,time)
INSERT INTO encounter_segments
SELECT
    a.mmsi AS mmsi_a,
    b.mmsi AS mmsi_b,
    LEAST(a.epoch_start, b.epoch_start) AS epoch_start,
    GREATEST(a.epoch_end, b.epoch_end) AS epoch_end
FROM
    segments as a, segments as b
WHERE
    a.geom3d &&& ST_Transform(ST_SetSRID(ST_MakeLine(
        ST_MakePoint(ST_XMin(ST_Transform(b.geom3d, 25832)) - 500, ST_YMin(ST_Transform(b.geom3d, 25832)) - 500, ST_ZMin(b.geom3d)),
        ST_MakePoint(ST_XMax(ST_Transform(b.geom3d, 25832)) + 500, ST_YMax(ST_Transform(b.geom3d, 25832)) + 500, ST_ZMax(b.geom3d))
        ), 25832), 4326)
    AND a.mmsi <> b.mmsi;

-- removing duplicate encounters segments (shipA/shipB shipB/shipA)
DELETE FROM encounter_segments WHERE mmsi_a > mmsi_b
    AND EXISTS (
        SELECT * FROM encounter_segments AS lookup
        WHERE lookup.mmsi_a = encounter_segments.mmsi_b AND lookup.mmsi_b = encounter_segments.mmsi_a
        );

-- grouping encounters segments into higher level unique encounters
CREATE INDEX idx_encounter_segments_mmsi_a_mmsi_b ON encounter_segments(mmsi_a, mmsi_b) INCLUDE (epoch_start, epoch_end);

ALTER TABLE encounter_segments ADD COLUMN encounter_id INTEGER;

WITH cte AS (SELECT mmsi_a, mmsi_b, epoch_start, epoch_end, DENSE_RANK() over (ORDER BY mmsi_a, mmsi_b) AS encounter_id FROM encounter_segments)
UPDATE encounter_segments
SET encounter_id = cte.encounter_id FROM cte WHERE encounter_segments.mmsi_a = cte.mmsi_a AND encounter_segments.mmsi_b = cte.mmsi_b;

-- encounters table creation
CREATE TABLE encounters (encounter_id INTEGER, mmsi_a CHAR(9), mmsi_b CHAR(9), epoch_start INTEGER, epoch_end INTEGER);

-- merging grouped encounter segments to higher level unique encounters
INSERT INTO encounters SELECT DISTINCT
    encounter_id,
    mmsi_a,
    mmsi_b,
    MIN(epoch_start) OVER (PARTITION BY encounter_id) AS epoch_start,
    MAX(epoch_end) OVER (PARTITION BY encounter_id) AS epoch_end
FROM encounter_segments;

-- collecting extended linestring geometry from segments table
SELECT AddGeometryColumn ('public','encounters','geom_a',4326,'LINESTRINGZ',3);
SELECT AddGeometryColumn ('public','encounters','geom_b',4326,'LINESTRINGZ',3);
```

```
WITH cte AS (SELECT mmsi, epoch_start, epoch_end, ST_MakeLine(geom3d ORDER BY epoch_start) AS geoms FROM segments GROUP BY mmsi, epoch_start, epoch_end)
UPDATE encounters
SET geom_a = (SELECT ST_MakeLine(cte.geoms ORDER BY cte.epoch_start) FROM cte
WHERE encounters.mmsi_a = cte.mmsi AND (encounters.epoch_start - 60) <= cte.epoch_start AND (encounters.epoch_end + 60) >= cte.epoch_end);

WITH cte AS (SELECT mmsi, epoch_start, epoch_end, ST_MakeLine(geom3d ORDER BY epoch_start) AS geoms FROM segments GROUP BY mmsi, epoch_start, epoch_end)
UPDATE encounters
SET geom_b = (SELECT ST_MakeLine(cte.geoms ORDER BY cte.epoch_start) FROM cte
WHERE encounters.mmsi_b = cte.mmsi AND (encounters.epoch_start - 60) <= cte.epoch_start AND (encounters.epoch_end + 60) >= cte.epoch_end);

-- calculate closest ship positions in time and space, and distance at that point
SELECT AddGeometryColumn ('public','encounters','closest_point_a',4326,'POINTZ',3);
SELECT AddGeometryColumn ('public','encounters','closest_point_b',4326,'POINTZ',3);
ALTER TABLE encounters ADD COLUMN shortest_distance FLOAT;

UPDATE encounters SET closest_point_a = ST_3DClosestPoint(geom_a, geom_b);
UPDATE encounters SET closest_point_b = ST_3DClosestPoint(geom_b, geom_a);
UPDATE encounters SET shortest_distance = ST_Length(ST_Transform(ST_MakeLine(closest_point_a, closest_point_b), 25832));

-- calculate relative course (encounter type)
ALTER TABLE encounters
ADD COLUMN cog_a FLOAT, ADD COLUMN cog_b FLOAT, ADD COLUMN cog_diff FLOAT;

WITH cte AS (SELECT mmsi, epoch_start, epoch_end, cog FROM segments)
UPDATE encounters
SET cog_a = (SELECT cog) FROM cte
WHERE encounters.mmsi_a = cte.mmsi
AND cte.epoch_start <= ST_Z(encounters.closest_point_a)
AND ST_Z(encounters.closest_point_a) <= cte.epoch_end;

WITH t AS (SELECT mmsi, epoch_start, epoch_end, cog FROM segments)
UPDATE encounters
SET cog_b = (SELECT cog) FROM t
WHERE encounters.mmsi_b = t.mmsi
AND t.epoch_start <= ST_Z(encounters.closest_point_b)
AND ST_Z(encounters.closest_point_b) <= t.epoch_end;

UPDATE encounters SET cog_diff = ABS(cog_a - cog_b);

-- calculate ship domain overlap
SELECT AddGeometryColumn ('public','encounters','ship_domain_a',4326,'POLYGON',2);
SELECT AddGeometryColumn ('public','encounters','ship_domain_b',4326,'POLYGON',2);
ALTER TABLE encounters ADD COLUMN ship_domain_overlap FLOAT;

UPDATE encounters
SET ship_domain_a = (SELECT ST_Transform(
    ST_Rotate(
        ST_Scale(
            ST_Buffer(
                ST_Transform(closest_point_a, 25832)
            , 1000)
        ,ST_Point(0.85, 1.65), ST_Transform(closest_point_a, 25832))
    , RADIANS(180-cog_a), ST_Transform(closest_point_a, 25832))
, 4326));

UPDATE encounters
SET ship_domain_b = (SELECT ST_Transform(
    ST_Rotate(
        ST_Scale(
            ST_Buffer(
                ST_Transform(closest_point_b, 25832)
```

```sql
             , 1000)
         ,ST_Point(0.85, 1.65), ST_Transform(closest_point_b, 25832))
     , RADIANS(180-cog_b), ST_Transform(closest_point_b, 25832))
, 4326));

UPDATE encounters
SET ship_domain_overlap = (SELECT ST_Area(ST_Transform(ST_INTERSECTION(ship_domain_a, ship_domain_b), 25832))
WHERE ST_Overlaps(ship_domain_a, ship_domain_b));

-- calculate speed difference
ALTER TABLE encounters ADD COLUMN max_sog_a FLOAT, ADD COLUMN max_sog_b FLOAT, ADD COLUMN max_sog_diff FLOAT;

WITH cte AS (SELECT mmsi, epoch_start, epoch_end, MAX(sog) AS sog FROM segments GROUP BY mmsi, epoch_start, epoch_end)
UPDATE encounters
SET max_sog_a = (SELECT MAX(sog) FROM cte
WHERE encounters.mmsi_a = cte.mmsi AND (encounters.epoch_start - 60) <= cte.epoch_start AND (encounters.epoch_end + 60) >= cte.epoch_end);

WITH cte AS (SELECT mmsi, epoch_start, epoch_end, MAX(sog) AS sog FROM segments GROUP BY mmsi, epoch_start, epoch_end)
UPDATE encounters
SET max_sog_b = (SELECT MAX(sog) FROM cte
WHERE encounters.mmsi_b = cte.mmsi AND (encounters.epoch_start - 60) <= cte.epoch_start AND (encounters.epoch_end + 60) >= cte.epoch_end);

UPDATE encounters
SET max_sog_diff = ABS(max_sog_a - max_sog_b);

-- calculate centripetal acceleration
ALTER TABLE encounters ADD COLUMN max_ca FLOAT;
WITH cte AS (SELECT mmsi, epoch_start, epoch_end, MAX(ca) AS ca FROM segments GROUP BY mmsi, epoch_start, epoch_end)
UPDATE encounters
SET max_ca = (SELECT MAX(ca) FROM cte
WHERE (encounters.mmsi_a = cte.mmsi OR encounters.mmsi_b = cte.mmsi) AND (encounters.epoch_start - 60) <= cte.epoch_start AND (encounters.epoch_end + 60) >= cte.epoch_end);

-- choose point based on 3D closest point that represents encounter
SELECT AddGeometryColumn ('public','encounters','encounter_centroid',4326,'POINT',2);
UPDATE encounters SET encounter_centroid = ST_Centroid(ST_Collect(closest_point_a, closest_point_b));

-- include ship types for visualisation purposes
ALTER TABLE encounters ADD COLUMN ship_type_a CHAR(30), ADD COLUMN ship_type_b CHAR(30);
UPDATE encounters SET ship_type_a = (SELECT MODE() WITHIN GROUP (ORDER BY ship_type) FROM points WHERE encounters.mmsi_a = points.mmsi);
UPDATE encounters SET ship_type_b = (SELECT MODE() WITHIN GROUP (ORDER BY ship_type) FROM points WHERE encounters.mmsi_b = points.mmsi);
```

```
--- Maritime risk analysis ---
-- Final risk score calculation in QGIS field calculator
final_risk = ((1-( "shortest_distance"/3300))+("ship_domain_overlap"/17624334)+(CASE
WHEN ("cog_diff" <= 45 OR "cog_diff" > 315) THEN 0 WHEN ("cog_diff" > 135 AND
"cog_diff" <= 225) THEN 0.5 WHEN (("cog_diff" > 45 AND "cog_diff" <= 135) OR
("cog_diff" > 225 AND "cog_diff" <= 315)) THEN 1.0 END)+( "max_sog_diff"/30)+(
"max_ca"/2500))/5
```