Rodmar Arntsen

Analysis of AIS-data to identify near encounter situations between ships

Master's thesis in Marine Technology
Supervisor: Stein Haugen
June 2019

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Marine Technology

NTNU
Norwegian University of
Science and Technology

Rodmar Arntsen

Analysis of AIS-data to identify near encounter situations between ships

Master's thesis in Marine Technology
Supervisor: Stein Haugen
June 2019

Norwegian University of Science and Technology
Faculty of Engineering
Department of Marine Technology

NTNU
Norwegian University of
Science and Technology

Preface

This master thesis is the culmination of my two year Master of Science at the Department of Marine Technology, with a specialization in Safety and Asset Management. The thesis has been written in the spring semester of 2019 at the Norwegian University of Science and Technology, and corresponds to 30 ECTS.

The intended reader of this report should have a basic knowledge and interest of maritime risk assessment. It is also seen as beneficial to have some knowledge of marine traffic and navigation for a greater understanding, but are not considered to be an absolute prerequisite.

Trondheim June 5, 2019

Rodmar Arntsen

Acknowledgements

There are several people I would like to thank for their comments, contributions, discussions and valuable assistant during the writing of this thesis.

First I would like to thank my supervisor, Professor Stein Haugen, which has been available, giving comments and guidance throughout my work with the thesis.

The thesis is written in cooperation with Safetec Nordic AS in Trondheim. I would like to thank Safetec for the opportunity of working together with them for this thesis, especially Martin Hassel, Asbjørn L. Aalberg and Peter Ellevseth for help and discussions during the progress of the thesis.

Finally I would like to thank friends and family for supporting and motivating me throughout the two years of studying at NTNU.

Summary

In this thesis, historical AIS data is used to identify near encounters between vessel. To identify near encounters a new ship domain is proposed with the intended application of assessing historical data. Domain violations has been divided into three different risk levels, based on the theory of Heinrich risk triangle. From the risk triangle the term "Near encounter" for the model is defined.

In the model there are also proposed a new way of determining the type of encounter (head-on, overtaking and crossing). All detected encounters are considered toward two parameters - rate of turn and speed change - to identify encounters where evasive maneuverers are performed.

With the new model a case-study is performed at an area in Vestfjorden in North-Norway for a three-year period. The results from the case-study shows that the number of encounters seems unreasonably high considering the traffic picture. The high number of encounters are to a large degree caused by the fact that all consequences are considered. The case-study area has rich fishing grounds for periods of the year, where between 1252-1515 vessels were engaged in the fisheries in the time period assessed. Fishing vessels, at the fishing grounds, are the main contributor to detected encounters.

From manually review of statistics there are found several faults in the AIS-data. The fault is probably caused by human operators operating the AIS transmitter. The faults in the AIS-data is so significant that a manually control is needed. The results also show that the need for manually control is needed to remove encounters classified as near encounters, which are marine operations where vessels are close to each other without it being a near encounter. The challenge of determining a ship domain that fits all vessels in all situations are discussed. Since no standardization is applied to maritime traffic, the domain should be seen as a domain of interest in the risk assessment, and not a domain that applies to all vessels in the dataset.

The conclusion is that the model at current state, with the uncertainties in both dataset and model, should be considered as a risk indicator that can be used to identify areas of interest for further risk assessment.

Sammendrag

I denne oppgaven brukes historiske AIS-data til å identifisere nærhendelser mellom fartøy. For å identifisere nærhendelser er et nytt skipsdomene foreslått med den tiltenkte bruken å vurdere historiske data. Domenebrudd har også blitt delt inn i tre ulike risikonivåer, basert på Heinrichs risikotrekant. Fra risikotrekanten er også begrepet "Nærhendelse" for modellen definert.

I modellen er det også foreslått en ny måte å bestemme typen hendelse på (møtende, innhentende og kryssende). Alle avdekte hendelser vurderes mot to parametere - svinghastighet (ROT) og hastighetsendring - for å identifisere hendelser hvor vike manøvrerer utføres.

Med den nye modellen har det blitt gjennomført en casestudie i Vestfjorden i Nord-Norge for en treårsperiode. Resultatene fra casestudien viser at antall hendelser virker å være urimelig høyt, når man vurderer trafikkbildet. Det høye antallet hendelser er i stor grad forårsaket av at alle konsekvenser vurderes. Case området har rike fiskeplasser i perioder av året, hvor mellom 1252-1515 fartøy var involvert i fisket i studieperioden. Fiskefartøy, på fiskebankene, er hoved bidragsyter til detekterte nærhendelser.

Fra manuell gjennomgang av statistikken er det funnet flere feil i AIS-dataene. Feilene er sannsynligvis forårsaket av menneskelige operatører som opererer AIS-senderen. Det er funnet at feilene i AIS-dataene er så signifikant at manuell kontroll er nødvendig. Resultatene viser også at behovet for manuell kontroll er nødvendig for å luke ut hendelser klassifisert som nærhendelser, som i virkeligheten er maritime operasjoner der fartøyene er nær hverandre uten at det er en nærhendelse. Utfordringen med å bestemme et skipdomene som passer til alle fartøy i alle situasjoner er diskutert. Siden det ikke finnes standardisering av maritim trafikk, bør domenet betraktes som et domene av interesse for risikovurderingen, og ikke et domener som gjelder for alle fartøy i datasettet.

Konklusjonen er at modellen i dagens tilstand, med usikkerhetene i både datasett og modell, bør betraktes som en risikoindikator som kan benyttes for å identifisere områder av interesse for ytterligere risikovurderinger.

Contents

List of Tables

List of Figures

Nomenclature

AIS	Automatic Identification System
ARPA	Automatic Radar Plotting Aid
COG	Course Over Ground
COLREG	Convention on the International Regulations for Preventing Collisions at Sea
CPA	Closest Point of Approach
ECDIS	Electronic Chart Display and Information System
ENC	Electronic Chart System
ETA	Estimated Time of Arrival
FSA	Formal Safety Assessment
GPS	Global Position System
IMO	International Maritime Organisation
Lat	Latitude
Long	Longitude
NCA	The Norwegian Coastal Administration
Nm	Nautical mile
OOW	Officer On Watch
QRA	Quantitative Risk Assessment
ROT	Rate Of Turn
SA	Situational Awareness
SAR	Search And Rescue
SOG	Speed Over Ground
STCW	Standards of Training, Certification and Watchkeeping for Seafarers
TCPA	Time to Closest Point of Approach

VHF	Very High Rrequency
VTS	Vessel Traffic Service
WGS	World Geodetic System

Chapter 1

Introduction

1.1 Background

Automatic Identification System (AIS) was introduced in the early 2000s as an aid to navigation and as a support tool for regulation and monitoring of shipping traffic. The tool was developed for both vessels and on shore facilities like flag states, rescue coordination centers, vessel traffic service and customs to mention some. AIS-messages are transmitted either through VHF radio or satellite and contain both static and dynamic information such as vessel identification, position, course and speed about each vessel, (The Norwegian Coastal Administration, 2019). Today most vessels in the commercial maritime industry are required by regulations to carry AIS. Vessels that are not required to carry it, still often have it installed as a navigational aid. The AIS system generates large amount of data about vessels movements and in later years this data has been stored. This large amount of data is assumed can be utilized in a risk assessment.

Risk assessment has been and is an important part of the maritime industry to be able to reduce the risk as low as reasonably practicable. Traditionally the risk assessments have been performed based on historical accidents data, geometrical considerations, simulation and expert judgements. With large amount of AIS-data, it is considered that these data's can be utilized to make risk assessments, both in a historical perspective and in the moment assessments which will be important for the development of autonomous vessels.

In 2018, Nordkvist proposed a model to utilize historical AIS-data as his master thesis in a collaboration with Safetec and NTNU, (Nordkvist, 2018). The model proposed by Nordkvist is the foundation for this thesis, which has a focus on improving the proposed model based on navigational knowledge and sea experience.

1.2 Objectives

The objectives of this thesis are:

1. Make a definition of the term "Near Encounter".
2. Propose a model that can utilize historical AIS-data, to be used as a risk indicator in coastal areas by identifying near encounters.
3. Compare the results towards findings in Nordkvist thesis (Nordkvist, 2018).
4. Make clear all assumptions and limitations of the model in current state and make recommendations for future improvements of the model.

1.3 Structure

The reminder of this paper is structured as follows:

- Chapter 2 present the theory behind the methods used in the model.
- Chapter 3 gives a presentation of the model with the different methods used to detect near encounters.
- Chapter 4 present the results from a case-study, using the model on an area in Vestfjorden for a three-year period.
- Chapter 5 has a thorough discussion about the model and the assumptions made. A general discussion about the of use of AIS data as risk assessments are performed.
- Chapter 6 recommended for further work to improve the model from current state.
- Chapter 7 conclusion.

Chapter 2

Theory

2.1 AIS

Automatic identification system (AIS) is internationally regulated through regulation 19 of SOLAS Chapter V (IMO, 2018). IMO states "The regulation requires AIS to be fitted aboard all ships of 300 gross tonnage and upwards engaged on international voyages, cargo ships of 500 gross tonnage and upwards not engaged on international voyages and all passenger ships irrespective of size. The requirement became effective for all ships by 31 December 2004." As an amendment to the IMO rules for AIS, the Norwegian Government has also implemented the EU-requirement for fishing vessels, requiring all fishing-vessels above 15m to carry AIS class A (Norwegian Maritime Directorate, 2012). Class B AIS, is for vessels that are not described in the rule and is an optional safety feature for these vessels. Class B only operates at reduced reporting rate or when free time slots are available (IMO, 2015a). Therefore, this will be an uncertain source of information and will not be further considered.

IMO states in their guidelines "AIS should always be in operation when the ship is under way or at anchor. If the master believes that the continual operation might compromise the safety or security of his/her ship or where security incidents are imminent, the AIS may be switched off" (IMO, 2015a) In the guidelines for the onboard operational use of AIS, there are listed some inherent limitations of AIS. These should be known when working with historical AIS-data, since they give important inputs on the limitations of the data used:

- The OOW should always be aware that other ships, in particular leisure craft, fishing boats, warships, and some coastal shore stations including VTS centres, might not be fitted with AIS.
- The OOW should always be aware that other ships fitted with AIS as a mandatory carriage requirement might switch of the AIS under certain circumstances by professional judgement of the master.
- In other words, the information given by the AIS may not be a complete picture of the situation around the ship.
- The OOW should be aware that poorly configured or calibrated ship sensors (position, speed and heading sensor) might lead to incorrect information being transmitted.

- It would not be prudent for the OOW to assume that the information received from other ships is of a comparable quality and accuracy to that which might be available on its own ship.
- The potential of AIS as an assistant for anti-collision device is recognized and AIS may be recommended as such a device in due time.
- Nevertheless, AIS information may merely be used to assist in collision avoidance decision-making. When using AIS in the ship-to-ship mode for anti-collision purposes, the following cautionary points should be borne in mind:
    1. AIS is an additional source of navigational information. It does not replace, but supports, navigational systems such as radar target-tracking and VTS.
    2. The use of AIS does not negate the responsibility of the OOW to comply at all times with the Collision Regulations, particularly rule 7 when determining whether risk of collision exist.

AIS class A gives information about the vessel which can be divided into three categories as shown below. In Norway the AIS system is managed by The Norwegian Coastal Administration (Kleppe, 2015). The information in AIS-messages that is received and stored in Norway is:

- Dynamic information (position, course and speed).
- Static information (Identity (IMO-number, MMSI-number, call-sign), vessel type, dimensions).
- Details about the voyage (Destination, estimated time of arrival (ETA), cargo, draft).

The positional accuracy of AIS is set to approximately 10 meters, but this will depend on the installation of the AIS being correctly and that the sensors used in the AIS are functioning as specified (IMO, 2015a). The static and dynamic information that is considered to be the most important for the validity of AIS data can be seen in Table 2.1

AIS-messages can be transmitted by both VHF (Very High Frequency) radio and satellite. Norway has two satellites going in a polar orbit, at 600km height. The information from the satellite is downloaded through Vardø or Svalbard every 90 minutes to The Norwegian Coastal Administration (Kleppe, 2015). The coastal-administration in Norway has a total of 60 base-stations that collects AIS messages from all vessels equipped with AIS out to 40-60nm from the coast by VHF. Every vessel mandatory to carry AIS class A, shall from the requirements from IMO send the dynamic information as showed in Table 2.2

Table 2.1: Dynamic- and static-information AIS (IMO, 2015a)

Information item	**Information generation, type and quality of information**
Lenght and beam	Set on installation or if changed
Type of ship	Select from pre-installed list
Location of electronic position fixing system (EPFS) antenna	Set on installation or may be changed for bi-directional vessels or those fitted with multiple antennas
Ship's position with accuracy indicators and integrity status	Automatically updated from the position sensor connected to AIS. The accuracy indication is approximately 10m
Position time stamp in UTC	Automatically updated from ship's main position sensor connected to AIS
Course over ground (COG)	Automatically updated from ship's main position sensor connected to AIS, if that sensor calculates COG. This information might not be available
Speed over ground	Automatically updated from the position sensor connected to AIS. This information might not be available
Heading	Automatically updated from the ship's heading sensor connected to AIS
Rate of turn	Automatically updated from the ship's ROT sensor or derived from the gyro. This information might not be available

Table 2.2: Message interval of dynamic information AIS

Type of ship	Reporting interval
Ship at anchor	3min
Ship 0-14 knots	12sec
Ship 0-14 knots and changing course	4sec
Ship 14-23 knots	6sec
Ship 14-23 knots and changing course	2sec
Ship >23 knots	3sec
Ship >23 knots and changing course	2sec

The Norwegian Coastal Administration has also given some general challenges with data, based on AIS on their homepage "Havbase" (Kleppe, 2015). These should be known when using the data source.

- Vessels sailing in area without AIS receivers.
- The signal force from the vessel transponder is not high enough to reach the receiver. This is dependent on type of transponder, height and placement of antenna on board the vessel and the cabling on board the vessel.
- The vessels AIS transponder is not configured to send the right information (MMSI, name and so on).

- The vessels transponder can have an error.
- Error in the GPS system.
- The crew on board the vessel neglect the importance of configuring the AIS transponder to send the right information. This refers to the static information; name of vessel, type of ship, dimensions as well as destination and ETA.

2.2 Maritime risk

The International Maritime Organization defines risk as the combination of the frequency and the severity of the consequence (IMO, 2015b). This can be expressed as shown in equation 2.1, where the risk (R) is found from the probability (P) and consequence (C) of an accident.

$$R = P * C \tag{2.1}$$

For ship collisions, the literature shows that research often has been divided into two parts. One part has investigated the probability or frequency of accidents, and the other part has looked at what the consequences of accidents could be. Even so, there has been performed some quantitative risk assessment (QRA) studies trying to look at both parts (Chai et al., 2017). In the article (Chai et al., 2017), the group concludes that although different factors like ship type, speed and visibility are estimated for the ship collision frequency, the impact of ship type in the accident consequence is not considered. They also noted that due to data limits, this study did not consider the effects of ship size and navigational skill on the causation probability. This shows how complicated it will be to conduct a full risk assessment with both probability of collision and consequences of a collision in a given area. Therefore, the two parts have been divided into two studies and this report will focus on the studies used to determining the probability or frequency of accidents.

When looking into the probabilities studies, the studies can be divided into static models and dynamic models (Goerlandt and Kujala, 2011). Snider (Snider, 1989) defined the difference between static and dynamic risk in his article:

- Static risk is the risk that would exist in an unchanging world. Obviously, the concept of static risk is hypothetical, for in such a world, losses would be continuing to occur, but their total frequency and severity would be constant.
- Dynamic risk are those risks which result from change itself. Dynamic risks may rise from significant changes in the frequency or severity of existing sources of loss or from completely new sources.

2.2.1 Static risk models for ship collisions

For static studies the numbers of collision occurrences over the studied timed period (N_{coll}) is estimated as in equation 2.2

$$N_{coll} = N_A * P_C \tag{2.2}$$

Where N_A is the number of pairwise vessel encounters during the time period resulting in collision, providing no evasive maneuvers is taken and P_C is the probability of failing to avoid a collision (Goerlandt and Kujala, 2011). This is the basic formula developed by Fujii in the early 70's (Fujii and Shiobara, 1971).

In static models there are two different ways of calculating the P_C or causation probability. Causation probability can be estimated by scenario approach and synthesis approach (Kujala et al., 2009). The scenario approach is based on historical accident data, which both is used by Pedersen (Pedersen and Simonsen, 1995), Weng (Weng et al., 2012) and Kristiansen (Kristiansen, 2005). Synthesis approach find the probability of error of vessels leading to collisions, by fault tree or Bayesian Belief network. The fault tree is used both by Pedersen (Pedersen and Simonsen, 1995) and Fowler (Fowler and Sørgård, 2000) where they use both estimations, experts judgments and assumptions to find the top probability of the fault tree. This is then used as the causation probability. For Bayesian Belief network, several studies have been performed to determine the causation probability of collision. The factors human, technology and organization has to different degrees been assessed by for instance Hänninen (Hänninen and Kujala, 2012), Martins (Martins and Maturana, 2013) and Olsen (Olsen, 2017) in her master thesis. These three studies show that the number of causes and the interactions between the causes are very complex. To determine the probability in each node they have used both observations, known probabilities - for instance failure of technical equipment - and expert judgments. The use of BBN is proposed in IMOs guideline for FSA to find subjective probability. IMO defines it as: "The degree of confidence in the occurrence of an event, measured on a scale from zero to one. An event with a probability of zero means that it is believed to be impossible; an event with the probability of 1 means that it is believed it will certainly occur." (IMO, 2015b)

2.2.2 Dynamic risk models for ship collisions

In dynamic risk models the risk or probability of collision varies with time. This is the base of the model made by Merrick in both 2000 and 2001 ((Merrick et al., 2000), (Merrick and van Dorp, 2001)) where they made simulations of different causes with respect to the time domain. Both models can be compared to the steps described as energy flow by Vinnem (Vinnem, 2014). The energy flow is simulated with different factors such as geometrical limitations - similar to those used in static assessments- and human and technical errors similar to the causes found in the Bayesian networks ((Hänninen and Kujala, 2012), (Martins and Maturana, 2013), (Olsen, 2017)). Aarsæther developed a model for simulation of ship traffic in his dissertations (Aarsæther, 2011). In his model he emphasized the importance of the human roll in navigation when simulating marine traffic. He concluded with: "The thesis also highlights deficiencies in simulation studies of marine traffic where the human operator is neglected and that the present development of human operator models for risk analysis does not contain the required model sophistication for inclusion in a time-domain simulation."

2.3 Situational awareness

Situational Awareness (SA) is defined by Endsley as: "Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in near future"(Endsley, 1995). In simplicity it can be said that SA is a person's "understanding" of a situation. SA is often divided into three hierarchical phases (Endsley, 1995)

- *Level 1 SA: Perception of the elements in the environment.*
  The first step in achieving SA is to perceive the status, attributes, and dynamics of relevant elements in the environment.
- *Level 2 SA: Comprehension of the Current Situation.*
  Level 2 SA goes beyond simply being aware of the elements that are present to include an understanding of the significance of those elements in light of pertinent operator goals.
- *Level 3 SA: Projection of Future Status.*
  The ability to project the future actions of the elements in the environment - at least in the very near term - forms the third and highest level of SA.

Endsley points out two elements that are important when discussing SA: "SA consist of an operator's knowledge of the state of the environment at any given point in time, this knowledge includes temporal aspects of that environment, relating both to the past and the future" in other words time is an important element in SA. The other element is space or spatial, she states that in addition to its aspect as a frequent "element" of SA, spatial information is highly useful for determining exactly which aspects of the environment are important for SA.

In maritime domain the cause of accidents is often identified as human error. Hetherington review different sources to determine the human error and found that the human error was responsible in the area between 49% $\sim$ 96% of all accidents (Hetherington et al., 2006). These percentages are in the same region as presented in several other studies (Olsen, 2017), (Chauvin et al., 2013), (Macrae, 2009). IMO defines human error as "A departure from acceptable or desirable practice on the part an individual or a group of individuals that can result in unacceptable or undesirable result" (IMO, 2015b).

In the study performed by Grech of human error in maritime operations, lack of SA was the reason behind 71% of the accidents. Inadequate planning, inadequate observation and poor interpretation stood for 90.4% of the lack of SA (Grech et al., 2002). In the project of enhancing situational awareness on board vessels, the project group found four main reasons for lack of SA (Pico et al., 2015)

- **Fatigue** is the main cause of the SA. Fatigue occurs in two different categories: temporary fatigue and chronic fatigue.
- **Information overload** can cause reduced SA in two ways: The first one is causing fatigue by having information overload for a longer period. The second one is missing the information needed by presenting less useful information on the equipment.
- **Track line attachment syndrome** is a tendency of navigators, commonly younger mates, but sometimes seasoned captains, to develop an unhealthy and obsessive need to place and keep a pictorial representation of their vessel upon a track line or series of lines presented on a computer monitor. This often leads to not using other navigational aids than the ECD/ECDIS such as radar and looking outside. In that way the officer is less aware of vessels in the surrounding area, that possibly are on collision course.

- **Trust and educate**. Another possible cause is the lack of trust in the people below the officers. What happens is that the higher ranked crew member does not trust the lower ranked crew member and spends a lot of extra time to make sure everything goes according to plan.

IMO implemented situational awareness in the Manila amendment to STCW in 2010 (IMO, 2010). In table A-II/1 "Specification of minimum standard of competence for officers in charge of a navigational watch on ships of 500 gross tonnage or more" situational awareness is implemented as a requirement. In table 2.3 excerpts are taken from table A-II/1 with respect to situational awareness.

Table 2.3: Excerpt from Tabel A-II/1, STCW

Column 1	Column 2	Column 3	Column 4
Competence	**Knowledge, understanding and proficiency**	**Methods for demonstrating competence**	**Criteria for evaluating competence**
Maintain a safe navigational watch	Bridge resource management Knowledge of bridge resource management principles, including: .1 ... .2 ... .3 ... .4 obtaining and maintaining situational awareness	....	Team member(s) share accurate understanding of current and predicted vessel state, navigational path, and external environment
Use of ECDIS to maintain the safety of navigation	Proficiency in operation, interpretation, and analysis of information obtained from ECDIS, including: .1... .2... .3... .4... .5... .6 Situational awareness while using ECDIS including safe water and proximity of hazards, set and drift, chart data and scale selection, suitability of route, contact detection and management, and integrity of sensors		Communication is clear, concise and acknowledged at all times in a seamanlike manner

2.4 Detecting near encounters and risk of near encounters

2.4.1 Ship domain

Ship domain has, since it was introduced in the 70s, been used in risk analysis and proposed as a tool for navigators to evaluate the risk. Several definitions for ship domain have been proposed, but a definition that has been used in several studies are Goodwin's "A ship domain may be thought of as the sea around his ship which the navigator would like to keep free, with respect to other ships and fixed objects"(Goodwin, 1975). In his study Pietrzykowski claims that the ship domain shape and size is determined by the officer of the watch who considers several factors such as ship's speed and length, sea area, traffic density, etc. This means the domain boundary varies depending on the current navigational situation(Pietrzykowski, 2008). Several ship domains have over the last 50 years been discussed and proposed.Figure 2.1 shows a comparison done by Wang in 2009 of some of the different domains proposed from different studies for the same vessel (Wang et al., 2009). Figure 2.1 shows that there is quite a large difference in size and shape of different ship domain proposed.

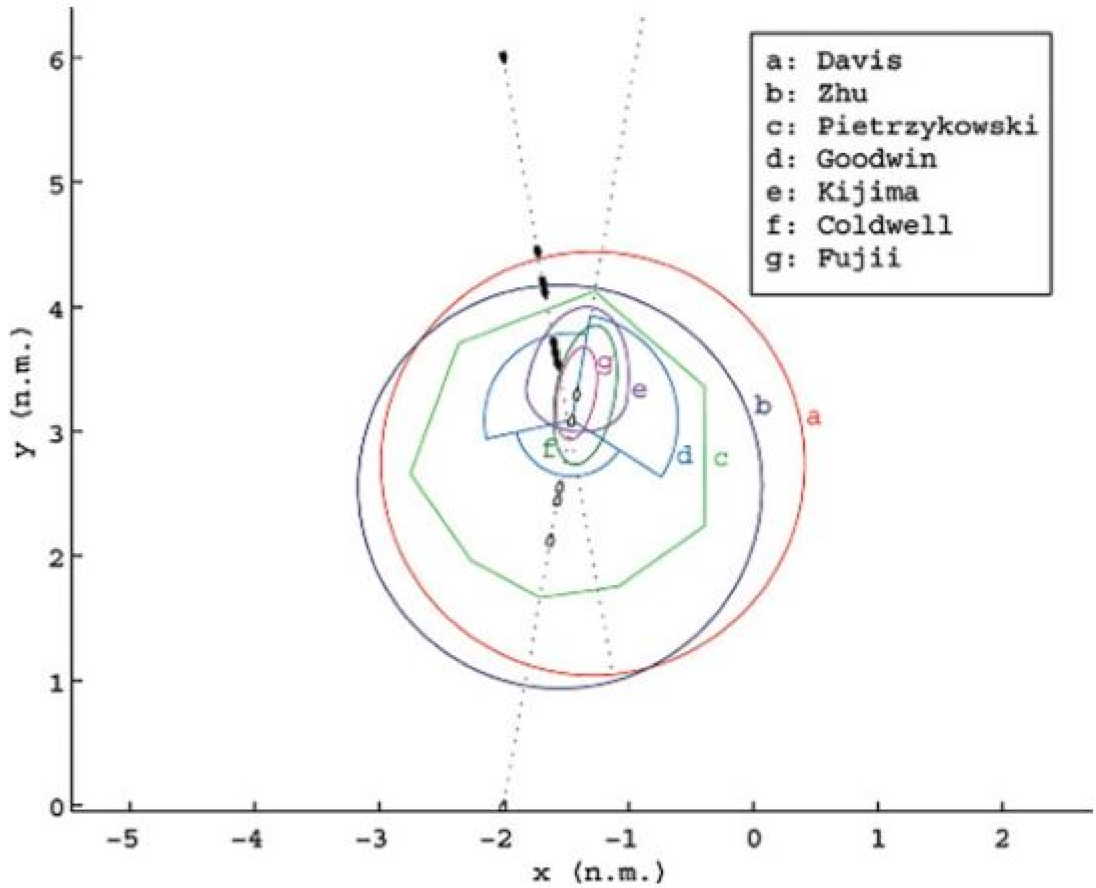


Figure 2.1: Comparison between different ship domains.

The differences of domains can be somewhat explained by how they are developed, if a statistical approach or analytical approach is used. The statistical approach bases itself on either radar data ((Fujii and Tanaka, 1971), (Goodwin, 1973)) or in later year AIS data ((Hansen et al., 2013), (Pietrzykowski and Magaj, 2017), (Zhou et al., 2018)) to determine the size of the ship domain based on actual observations in a specific sea area. Statistical domains still vary depending on where the data is extracted, and which factor are used to define the domain. Some uses only dimensions of the vessel and other both dimensions and speed. A typical shape and size of statistical domain can be seen in Figure 2.2. Here the only parameter determining the size of the domain is the length of the vessel. The domain proposed, called a comfort ellipse by the authors, is defined as the minimum area, where ships in the study area would like to keep clear from obstacles or other ships. The group used historical AIS-data in Danish waters to create the domain (Hansen et al., 2013).

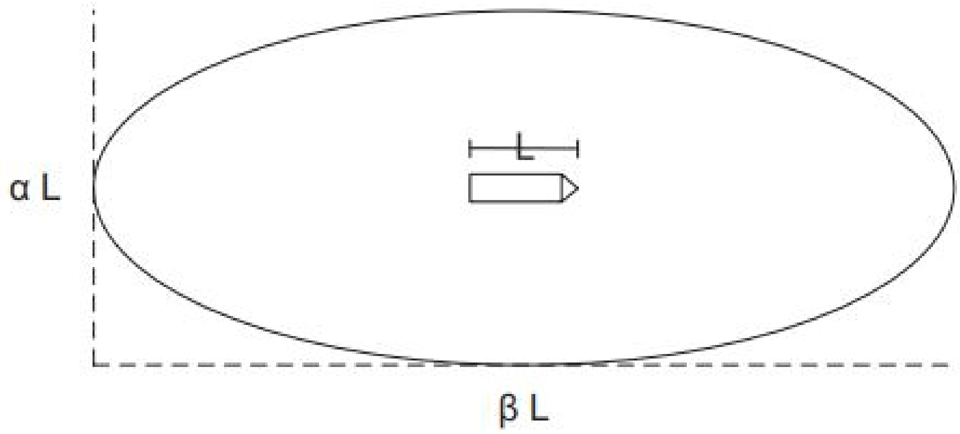


Figure 2.2: Comfort ellipse proposed by Hansen using historical AIS-data (Jensen et al., 2013).
$\alpha = 3.2$, $\beta = 8$ and L is the length of the ship

In the analytical approach the domain is defined using simulations, expert judgements and rules and regulations ((Liu et al., 2016), (Wang, 2010), (Dinh and Im, 2016), (Wang, 2013)). This means that the analytical domains are very dependent on what the intended use of the domain is and for which vessels it applies. Therefore, analytical domains are not looked further into in this thesis.

2.4.2 Detecting collision risk at sea

For navigators at sea the regulation for collision avoidance is COLREG 72. COLREG rule 5 "Look-out" states: "Every vessel shall at all times maintain a proper look-out by sight and hearing as well as by all available means appropriate in the prevailing circumstances and conditions so as to make a full appraisal of the situation and of the risk of collision" (IMO, 1972). Further rule 7 "Risk of collision" and rule 8 "Action to avoid collision" states which situations should be considered to have collision risk and how to maneuver to avoid the situation in any condition of visibility. There are mainly two ways of determining if there is a risk of collision:

- By look-out, use of relative movement of the other vessel and/or compass bearing. If there is no change in relative movement or compass bearing there is a risk of collision.
- Radar and AIS data, when the set limit values for CPA and TCPA are broken there is a risk of collision.

COLREG states that if there is any doubt such risk shall be deemed to exist. Rule 8 also states: "that any actions to avoid collision shall be taken in accordance with the Rules, if the circumstances of the case admit, be positive, made in ample time and with due regard to the observance of good seamanship." It is reason to assume that navigators today base themselves highly on Radar and AIS if these systems are installed and working. These systems give exact passing distance to other vessels and time to closest point of approach at an early state, where look-out and compass bearing can be hard to determine. This means to comply with COLREG the navigators use set limit values for CPA and TCPA in the systems. The limits are either based on the navigator's perception of the water and traffic density or the limit values are set as a default by the master or company. When a situation is defined as a collision situation, there are several rules regulating the encounter and defining a give-away and stand-on vessel for the different encounters (overtaking-, head-on- crossing situations) based on the conditions of visibility. In encounter between vessels in sight of one another, there will be a give-away and a stand-on vessel. Rule 18 states responsibilities between vessels (power driven-, fishing-, sail-vessel and so on) and which vessel has right of way. The only exception is rule 14 head-on situations between motor vessels, where both have a give-away responsibility. The regulation of both give-away and stand-on vessel is given in rules 16 and 17 described below:

- **Rule 16 Actions by give-way vessel.**
  Every vessel which is directed to keep out of the way of another vessel shall, so far as possible, take early and substantial action to keep well clear.

- **Rule 17 Actions by stand-on vessel.**
  (a)
  (I)Where one of two vessels are to keep out of the way the other shall keep her course and speed.
  (II) The latter vessel may however take action to avoid collision by her maneuver alone, as soon as it becomes apparent to her that the vessel required to keep out of the way is not taking appropriate action in compliance with these Rules.

In other words, if a collision situation is developing both vessels have responsibilities to act to avoid the situation in ample time. Even the stand-on vessel must give-way if the give-away vessel does not fulfill the rules. A situation of close quarters between two vessels, rule 8 and 17 gives both vessels the responsibility to avoid the situation to develop into a collision.

Chapter 3

Method

The model's purpose is to identify encounters between vessels, by use of ship domain. The model rank encounters based on the closeness between the vessels and time with use of the principle of Heinrich's accident triangle (Collins, 2011). Figure 3.1 shows the triangle used to define the risk and for this model three different levels has been defined, where level 1 is defined as "Near Encounter". The reason for choosing Heinrich triangle to rank encounters is to differentiate the risk of collision. Risk in maritime industry is defined as the probability and consequences as described in Section 2.2. The triangle is used to represent the higher probability of collision based on the distance between the vessels in an encounter. The model uses historical AIS data and ship domain to define encounters into the three different levels of the risk triangle which is further described in Section 3.1. The model also uses the parameters rate of turn and speed change to indicate evasive maneuverers performed during a time period 3 minutes before CPA till 2 minutes after, by one of the vessels in the encounter. According to COLREG rule 8 and good seamanship, it can be claimed that maneuverers to avoid collision should be performed well before this time and can therefore be an identification of lack of SA.

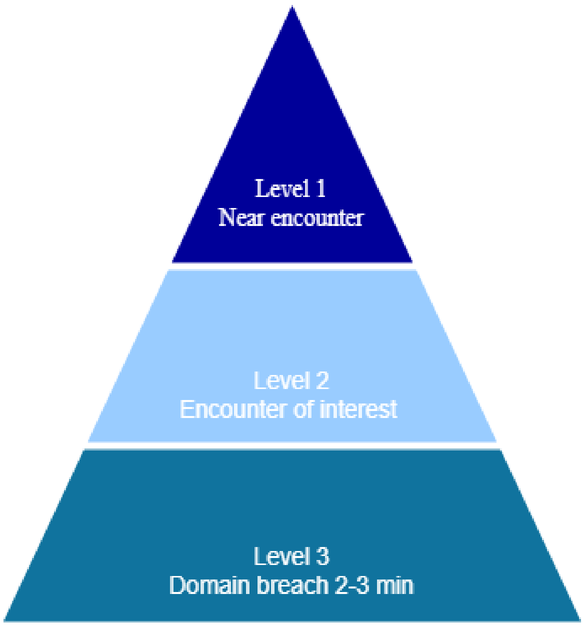


Figure 3.1: Risk triangle for model, divided into three different levels of encounters.

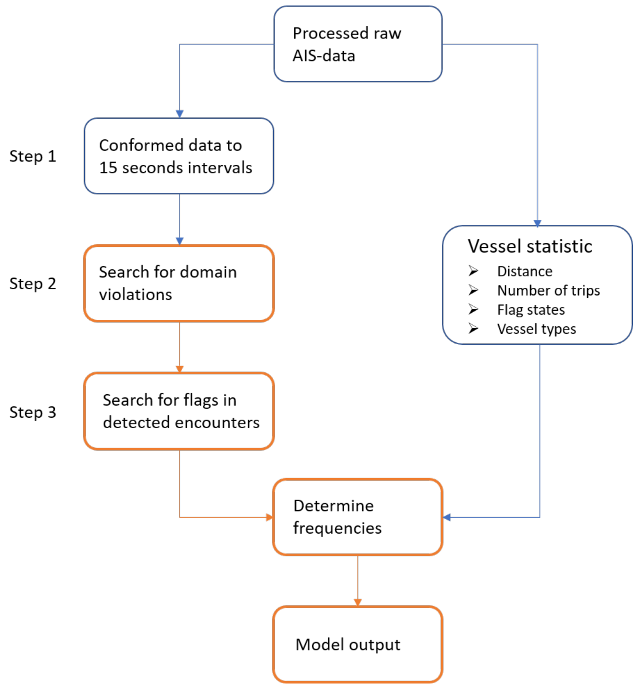


Figure 3.2: Outline of hole model where the main focus in this thesis is the steps marked with orange frame.

The model in this thesis is shown in Figure 3.2 and is based on the model proposed by Nordkvist (Nordkvist, 2018). The focus in this thesis is to improve step 2-3 and to verify the validity of the data in the statistics. Step 1 was performed by Nordkvist and gave datasets from Vågsfjorden for the time period 2013-2016 (Nordkvist, 2018). The datasets are conformed into 15 seconds intervals, except for 2013 which is conformed into 10 seconds intervals using "nearest" method with a tolerance of 4 seconds. For all rows not filled with nearest method, the last known value for static data and the dynamic data -course over ground (COG)- are used for 4-time steps i.e. one minute. COG is filled in with last known value because of problems with linear interpolation when general heading is 000/360°. Position (lat, long) and speed over ground (SOG) are linearly interpolated from last known row towards next row for 4-time steps. After four rows from known values are filled, the remaining empty rows are discarded (Nordkvist, 2018).

3.1 Domain

An encounter is in the model defined as when a vessel enters another's domain. The focus of the domain in this model is in the front of the vessel. It is assumed that a navigator will have a greater focus in the front of the vessel and therefore the future, than to the aft or in other words the past of the journey. The domain used in this model has an oval shape, Figure 3.3, and uses some of the properties defined by Hansen (Hansen et al., 2013). The domain has a minimum circular extension in all directions set to $R = 1.6 * L$, where R is the radius of the circle and L is the total length of the vessel. The aft radius is set to be constant $1.6 * L$ equal to starboard and port radii. Hansen pointed out that a navigator will have a greater focus to the area in front of the vessel compared to the area behind (Hansen et al., 2013), which also is assumed in the model proposed. For the front radii it is in the model dependent on speed and time, compared to Hansen's radii which is given by the length of the vessel. Time is set to conservative 180 seconds or 3 minutes and gives an estimation of where the vessel will be in 3 minutes if course and speed are not altered. The decision of selecting 180s or 3min is based on the property of maritime radars, where most radars can display vectors of 3 or 6 min. It is reasonable to assume most radars used in collision avoidance (ARPA), uses vectors of 3 or 6 minutes in coastal waters. Three rules are selected to determine the length of the front radius:

- Main rule of determining front radius.
  $Y = v * 180s$
- $1.6 * L > Y$
  The front radius is equal to the other radii if Y is smaller than 1.6*L and the domain becomes circular.
- $1nm < v * 180s$
  The front radius is limited to a length of 1nm. Encounters with passing distances above 1nm or 1852m are not considered to be of interest in this model (even if some captains and companies can have a larger passing distance in their orders). This means that up to 20knots the front radius will be speed dependent, but above 20 knots it will be limited to 1nm.

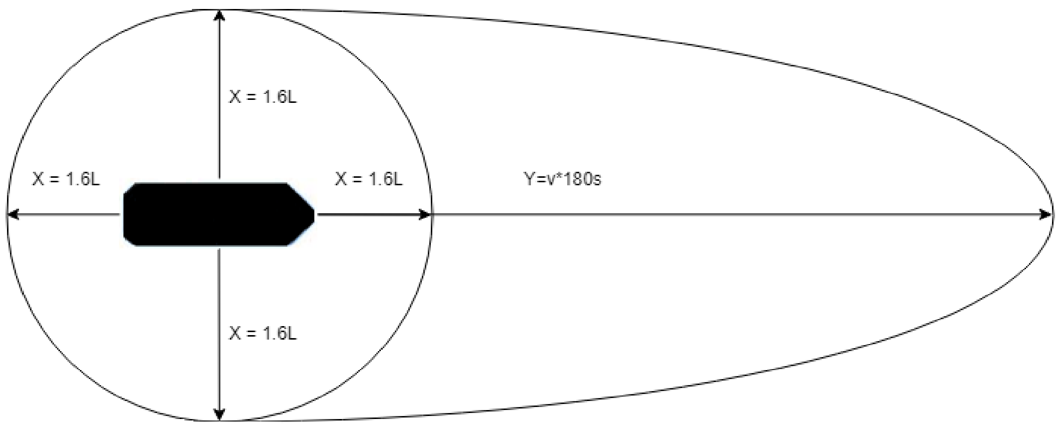


Figure 3.3: Proposed domain used in model. Y is the front radius, v is the vessel speed (SOG), L is the length of vessel and x is the minimum extension in all directions.

With respect to Heinrich's accident triangle, three boundaries are set for the domain to classify the three risk levels, ref Figure 3.1.

- **Level 1:**
  A domain violation of the circular domain or within front radius equal to 1 minute, dependent on which is the greatest, is classified as a level 1 encounter or a near miss. It is assumed that a breach in this area would not be following the term good seamanship, which is an important term used in COLREG. The time to act if something unexpected is to occur is low both for the human operator to react and vessel to maneuver. It is considered to be too low for any vessel type.

- **Level 2:**
  A domain violation between 1-2 minutes in front of own vessel is classified as an encounter of interest. Dependent on the position in the fairway/sea-area, the incident can be of interest. If the encounter is in open water, good seamanship will in many cases not be performed if there is enough water around the encounter to increase the time distance.

- **Level 3:**
  Level three is a domain breach within the area 2-3 minutes in front of the vessel. These encounters can be used to identify trends and define areas of interest, but in most cases, navigators working at sea would under many circumstances be satisfied with a TCPA in the area 2-3min. Most vessels will have time to maneuver if something unforeseen should happen.

Geographiclib for Python is used to solve geodesic problems i.e. distances and angles on the ellipsoid, in the same way as in Nordkvist paper (Nordkvist, 2018). AIS receivers are not normally placed in the center of the vessel and it is therefore necessary to determine the true center of the vessel, as a reference point for the ship domain. The method used to determine the true center is the same as used in Nordkvist model. Using the Geographiclib package, true center is solved using a geodesic problem, equation 3.1. Input variables are coordinates, clockwise azimuth ($azi1\_True$) from 0° North and distance (s12) from AIS position to true ship center.

$$True\_center = Direct.WGS84.Direct(Latitude, Longitude, azi1\_True, s12) \tag{3.1}$$

The true azimuth from AIS position to center of the vessel is found from equation 3.2 & 3.3.

$$azi1\_True = COG + \beta \tag{3.2}$$

$$\beta = \begin{cases} 0 & \text{for} \quad d_{bow} > d_{aft} \wedge d_{starb} = d_{port} \\ 180 & \text{for} \quad d_{bow} < d_{aft} \wedge d_{starb} = d_{port} \\ 90 & \text{for} \quad d_{bow} = d_{aft} \wedge d_{starb} > d_{port} \\ -90 & \text{for} \quad d_{bow} = d_{aft} \wedge d_{starb} < d_{port} \\ arctan\frac{dist\_y}{dist\_x} & \text{for} \quad d_{bow} > d_{aft} \wedge d_{starb} > d_{port} \\ 360 - arctan\frac{dist\_y}{dist\_x} & \text{for} \quad d_{bow} > d_{aft} \wedge d_{starb} < d_{port} \\ 180 + arctan\frac{dist\_y}{dist\_x} & \text{for} \quad d_{bow} < d_{aft} \wedge d_{starb} < d_{port} \\ 180 - arctan\frac{dist\_y}{dist\_x} & \text{for} \quad d_{bow} < d_{aft} \wedge d_{starb} d_{port} \end{cases} \tag{3.3}$$

The distance($s12$) from the AIS position to the true center of the vessel is found by equation 3.4-3.6.

$$s12 = \sqrt{dist\_x^2 + dist\_y^2} \tag{3.4}$$

$$dist\_x = \begin{cases} (d_{bow} + d_{aft})/2 - d_{aft} & \text{for} \quad d_{bow} > d_{aft} \\ (d_{bow} + d_{aft})/2 - d_{bow} & \text{for} \quad d_{bow} < d_{aft} \end{cases} \tag{3.5}$$

$$dist\_y = \begin{cases} (d_{starb} + d_{port})/2 - d_{port} & \text{for} \quad d_{starb} > d_{port} \\ (d_{starb} + d_{port})/2 - d_{starb} & \text{for} \quad d_{starb} < d_{port} \end{cases} \tag{3.6}$$

Figure 3.4 shows the geometry for finding true center of the ship from AIS position.

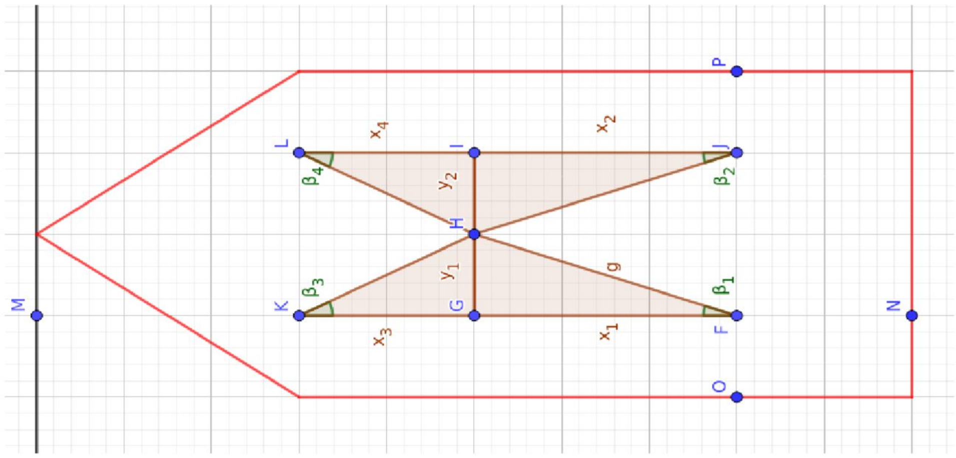


Figure 3.4: Definition of geometries used to find true ship center (Nordkvist, 2018).

From true center of the vessel the new domain, Figure 3.3 is given by the four radii of equation 3.7. In the equation v_{own} is own vessel speed (SOG) and L is equal to total length of own vessel.

$$\begin{cases} R_{fore} = \begin{cases} 1,6 * L & \text{for} \quad v_{own} * 180s \leq 1,6 * L \\ v_{own} * 180s & \text{for} \quad v_{own} * 180s > 1,6 * L \\ 1nm & \text{for} \quad v_{own} * 180s \geq 1nm \end{cases} \\ R_{aft} = R_{starb} = R_{port} = 1,6 * L \end{cases} . \tag{3.7}$$

To determine domain breaches the same method is used as in Nordkvist (Nordkvist, 2018), based on Zhang (Zhang et al., 2016). The angle between own ship COG to target ship center is denoted α and is measured clockwise to the target ship, ref equation 3.8. Illustration of α can be seen in Figure 3.6.

$$\begin{cases} \alpha = 360 - (COG - azi1) & \text{if } azi1 \leq COG \\ \alpha = azi1 - COG & \text{if } azil1 > COG \end{cases} \tag{3.8}$$

$azi1$ is the absolute clockwise angle between own ship center to target ship center and is found using the Geographiclib package from 000° North, equation 3.9. When the target ship is west of own ship the value is negative and therefore calculated anticlockwise.

$$\begin{cases} azi1 = Geodesic.WGS84.Inverse(lat_1, lon_1, lat_2, lon_2) & \text{if } lon_2 \geq lon_1 \\ azi1 = 360 - Geodesic.WGS84.Inverse(lat_1, lon_1, lat_2, lon_2) & \text{if } lon_2 < lon_1 \end{cases} \tag{3.9}$$

With known α the distance to the boundary of own ship with respect to target ship is calculated with equation 3.10.

$$\begin{cases} l_\alpha = \left(\frac{1+\tan^2\alpha}{\frac{1}{R_{fore}{}^2} + \frac{\tan^2\alpha}{R_{starb}{}^2}} \right)^{1/2} & \text{if } \alpha \leq \frac{\pi}{2} \\ l_\alpha = \left(\frac{1+\tan^2\alpha}{\frac{1}{R_{aft}{}^2} + \frac{\tan^2\alpha}{R_{starb}{}^2}} \right)^{1/2} & \text{if } \frac{\pi}{2} < \alpha \leq \pi \\ l_\alpha = \left(\frac{1+\tan^2\alpha}{\frac{1}{R_{aft}{}^2} + \frac{\tan^2\alpha}{R_{port}{}^2}} \right)^{1/2} & \text{if } \pi < \alpha \leq \frac{3}{2}\pi \\ l_\alpha = \left(\frac{1+\tan^2\alpha}{\frac{1}{R_{fore}{}^2} + \frac{\tan^2\alpha}{R_{port}{}^2}} \right)^{1/2} & \text{if } \frac{3}{2}\pi < \alpha \end{cases} \tag{3.10}$$

Domain violations are found when the length between two vessels is shorter than l_α, for one or both of the vessels. Each encounter will have a CPA when the distance between the vessels are smallest. The risk level of an encounter is based on the distance at CPA, in accordance with the levels in Figure 3.1.

3.2 Parameters to indicate evasive maneuvers

A navigator has two possible actions to avoid a near situation, turning or changing speed. The model utilizes these two parameters to indicate evasive maneuvers in encounters, on the bases of set limits. For the rest of this thesis, the term flag and flagged will be used for indicating evasive maneuvers in encounters. COLREG rule 8 states that a maneuver to avoid collision should be positive, made in ample time and with due regard to the observance of good seamanship. In coastal waters an evasive maneuver within three minutes, would in most situations not be considered in ample time and in good seamanship manner Therefore these encounters will be of extra interest.

3.2.1 Turning manoeuvre

Turning is the most used maneuver to avoid situations at sea, because it is easy to detect and normally the most effective way of avoiding situations. When a vessel is turning it results in rate of turn (ROT) degrees per minute. For any vessel ROT can be estimated based on the planned journey of the vessel as equation 3.11. This means that most vessels will have a predefined ROT, they will use in turns in the planned route based on the vessel's maneuverability and speed.

$$ROT = \frac{Speed(knots)}{Selected\ turn\ radius(nm)} \tag{3.11}$$

Table 3.1 shows some ROT values and the context between speed and turning radius. These are ROT that can be considered normal to different degree in planned journeys, depending if the journey is in coastal waters only, or if it is planned for inland waters along the coast as well.

Table 3.1: ROT values and the context between speed and turning radius.

Speed (knots)	Turning Radius (nm)	Rate of turn (°/min)
8	0,2	40
8	0,5	16
8	1,0	8
12	0,2	60
12	0,5	24
12	1,0	12
18	0,2	90
18	0,5	36
18	1,0	18

For determining the limit for evasive maneuverers COLREG rule 8 b states: "Any alteration of course and/or speed to avoid collision shall, if the circumstances of the case admit, be large enough to be readily apparent to another vessel observing visually or by radar; a succession of small alterations of course and/or speed should be avoided." (IMO, 1972). This means that a turn to avoid a situation within 3min, will normally generate quite high ROT. Nordkvist (Nordkvist, 2018) used 70 $^{\circ}/min$ as the limit for all vessels in his model. His results show that most encounters classified as exceptional encounters was between two fishing vessels. If equation 3.11 is considered smaller vessels have much easier to gain a high ROT, because they will have smaller turning radius. Smaller vessels will also be more exposed to weather generating yawing motions, which can generate high ROT by itself. Metsl did a case study on a collision between a ferry and a fishing vessel, where only the ferry was set up to log ROT. The group found that the ferry, travelling in 18 knots, had a ROT < $150^{\circ}/min$ in 99.999% of the time, during half a year study period (Mestl et al., 2016). Once again looking at equation 3.11 speed is the other parameter that can generate a high ROT. Since not all vessels transmit ROT-value in their AIS messages an estimation of the ROT-value is used, equation 3.12. A challenge with using equation 3.11 to determine the ROT is that the dataset is conformed into 15 seconds intervals. In the conforming use of "nearest" method and last known value, some uncertainty is given to the COG value, and the time between two COG values are unknown. The time can be up to 23 seconds by use of nearest method, and even higher with last known value. The time interval of 15 seconds also gives the chance of underestimating the actual ROT value. Often a ROT value will peak in a maneuver and these peaks will not be possible to detect when using the equation and 15 seconds intervals, therefore the risk of underestimating ROT is considered high.

$$\widehat{ROT}_t = \frac{1}{2}\sum_{i=t}^{t+1}\frac{COG_i - COG_{i-1}}{time_i - time_{i-1}} \tag{3.12}$$

When setting the limit for evasive maneuver, size is taken into consideration. Small vessels will normally have the capability of using smaller turning radius than larger vessels and therefore generate higher ROT, ref equation 3.11. Speed will also have the possibility to generate high

ROT, but for the model proposed only size is considered when setting the limits for ROT. The limits for ROT used in the model are:

- **Vessels $\leq$ 50m**: For smaller vessels the ROT limit is set at $150°/min$. Generally higher ROT should be considered normal for evasive maneuvers for small vessels. A vessel sailing at 10knots will need a radius of the turn $\approx$ 0,067nm or 124m to breach the limit, which for a small vessel is considered reasonable. The high limit for ROT, will reduce the number of vessels gaining high ROT because of yawing motion.

- **Vessels > 50m**: For vessels above 50 m, the ROT limit is set at $70°/min$, equal to Nordkvist value. For a vessel sailing at 10 knots, the turn radius becomes $\approx$ 0,14nm or 260m.

If a vessel is breaking the threshold value for ROT during a domain breach the encounter gets flagged.

3.2.2 Change of speed

Speed is the second option for a navigational officer to avoid a situation. Changing speed is often not an effective maneuver in late collision avoidance, especially for larger vessels with high inertia force. Smaller vessels or vessels with high speed potential still can use speed to avoid situations, and if turning is not an option changing speed is the only possibility to avoid collision for a navigator.

For the model speed change will have to be sufficiently large to be defined as a maneuver to avoid a collision. The limit is assumed and set at 3knots change in a 30 second interval within 3 minutes before CPA. The limit value is high, but it is considered reasonable within 3 minutes of CPA to be able to avoid collision. 3 knots speed change equals a sailed distance in 3 minutes of $\pm$ 0,15nm $\approx$ 278m and can be considered large and clear enough to be considered a maneuver to avoid collision. The speed change ΔSOG is calculated in equation 3.13 for a period of 30 seconds. Since the conformed dataset has a interval of 15 seconds this means that 3 values are used in total to determine the ΔSOG. The reasoning for choosing only three values in ΔSOG, is that SOG is one of the values that are quite irregular in the datasets and therefore to be able to determine a ΔSOG a period of 30 seconds is chosen. If $\Delta SOG >$ 3knots during a 30 second interval it will be considered as an evasive maneuver. This limit is as already stated very conservative, especially for large vessels with high inertia force, but for smaller vessels with high power reserves and low inertia it can be considered reasonable. In general, speed change as an evasive maneuver is in practice used by vessels with high power to inertia relationship.

$$\Delta\widehat{SOG}_t = \frac{1}{2}\sum_{i=15s}^{15s} |SOG_t - SOG_{t-i}| + |SOG_t - SOG_{t+i}| \tag{3.13}$$

If $\Delta SOG >$ 3knots the encounter will be flagged.

3.2.3 Determining the type of encounter

The type of encounters is defined in COLREG based on the visibility. Vessels in sight of one another are regulated by COLREG rule 11-18, and vessel in restricted visibility regulated by rule 19. In this model all encounters are considered to be in sight of each other, since visibility is not included in the model. Encounters are also considered to be between two motor-vessels, rule 18 responsibilities between vessels is not considered (IMO, 1972). The reason for neglecting rule 18 is for computer limitations and it is considered that the validity of AIS information such as vessels engaged in fishing, not under command and so on, is in many cases misleading since this information has to be changed by the navigator on the AIS transmitter onboard the vessel. In many cases, fishing vessels will be engaged in fishing in their AIS-message when transiting between locations even though they are then -from COLREG definition rule 3- to be considered a motor vessel (IMO, 1972).

Encounters are therefore in the model based on COLREG rule 13-15, which is encounters between two motor vessels in sight of one another (IMO, 1972). Navigational light sectors can be used to illustrate the different types of encounters, Figure 3.5 shows the navigational light sectors used.

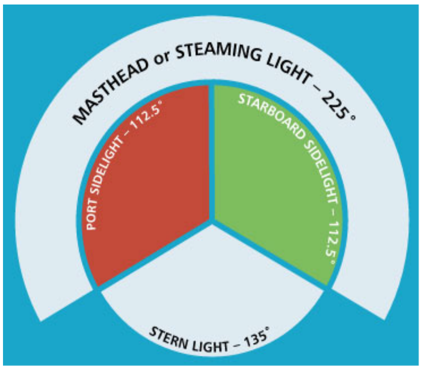


Figure 3.5: Coverage of light-sectors vessels (Roads and Maritime Services, 2019).

Definition of the type of encounter from COLREG.

- Overtaking situation. From rule 13 in COLREG an overtaking situation is defined as "A vessel shall be deemed to be overtaking when coming up with another vessel from a direction more than 22.5° abaft her beam". This will fall under the stern light sector on Figure 3.5.

- Head on situation. Defined in COLREG rule 14 "meeting on reciprocal or nearly reciprocal courses so as to involve risk of collision". This is a definition that can be discussed by the phrase near reciprocal course.

- Crossing situation is defined in COLREG rule 15 as "two power-driven vessels crossing so as to involve risk of collision". In crossing situations, a "give a way" vessel and a "stand on vessel" is determined. A vessel which have the other vessel on its starboard side (green light) should keep out of the way for the other vessel and vessel with vessel on its port side (red light) should stay on course and speed.

To determine what kind of situation encounters are, Nordkvist (Nordkvist, 2018) and Iperen (Iperen, 2015) use relative heading(φ) at CPA. In this model the type of encounter is determined

6 minutes before CPA with a combination of relative heading (φ) between the vessels in the encounter, eq 3.14, and relative bearing (α), eq 3.8. The choice of using 6 minutes is as described earlier because of properties of radars vectors. As shown in Figure3.6 the relative heading will not be enough to determine the type of situation at 6 minutes before CPA, nor will relative bearing. The reasoning behind determining the type of situation at TCPA 6 minutes is that this will be in such a good time before the encounter that the vessels are not affected by each other, since domain violations will occur. Still they will be so close to the encounter that give away responsibility is decided. In the model, table 3.2 specify what is used to classify encounters. A sector of 10 ° to each side is assigned to head-on situations. Table 3.2 shows that two statements are applied to determine if it is a head-on or overtaking situations, any situation not assigned as a head-on or overtaking situation is considered crossing situations. Give-away and stand-on vessel in crossing situations are determine by α at 6 minutes before CPA. Since all situations not found to be head-on nor overtaking is assigned as crossing situations, the sectors for determining give away and stand on vessel are defined around 360°.

$$\varphi = max(COG_1, COG_2) - min(COG_1, COG_2) \tag{3.14}$$

In eq 3.14 COG_1 and COG_2 are the courses of the two vessels in each encounter. Ref Figure 3.6 it can be considered as the courses of vessel A and vessel C in the moment 6minutes before CPA. The reasoning for using max/min values instead of absolute values in the equation is the challenge of courses going to 360°. Therefore, by using max/min value between the two courses, results will always be positive and no consideration has to be taken towards the challenge of the switch between 000 / 360°.

Table 3.2: Defining type of encounters.

Type of encounter	
Encounter type	**Conditions**
Overtaking	$292.5\,° \leq \varphi \leq 67.5°$ $112.5° \leq \alpha \leq 247.5°$
Head-on	$170° \leq \varphi \leq 190°$ $010° \geq \alpha \geq 350°$
Crossing, give away vessel	$000° < \alpha < 180°$
Crossing, stand on vessel	$180° < \alpha < 360°$

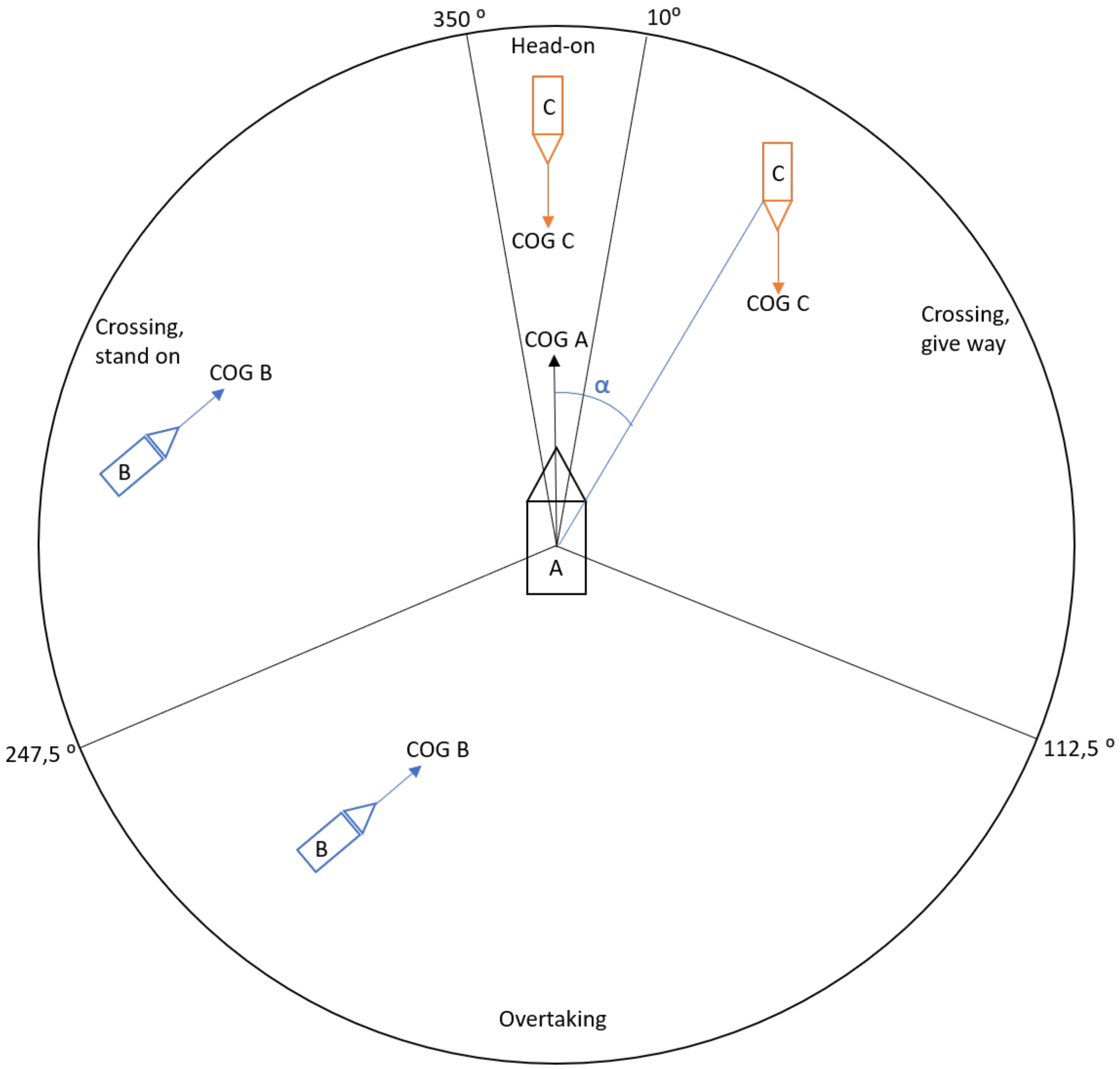


Figure 3.6: Defining of situation of encounters.

Chapter 4

Case study: Vestfjorden

The case-area used in this thesis is the same as used in Nordkvist (Nordkvist, 2018) to have the possibility to compare the results. The area can be seen in Figure 4.1 and are defined by the corner coordinates South-West 67°00'N 011°18'E and North-East 68°00'N 014°00'E and cover the south-west part of Vestfjorden and Lofoten. All plots in charts are, if not other vice specified, in this chapter made by use of QGIS3 with Mercator projection and datum WGS84.

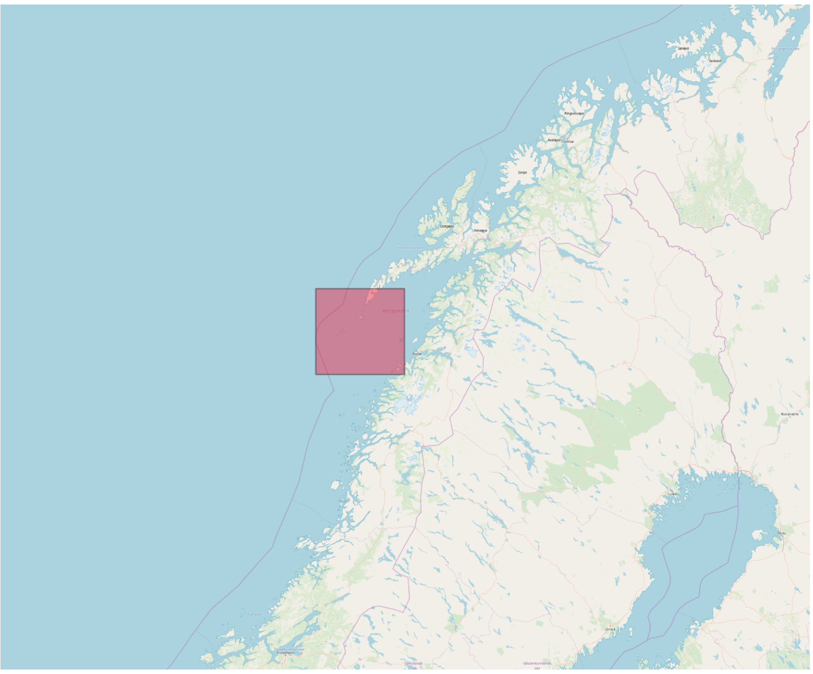


Figure 4.1: Case-study area marked with red rectangle.

Vestfjorden is the natural sailing route for vessels transiting from South to North-Norway, on coastal journeys through Tjeldsundet. In the north-east, Vestfjorden is connected to Ofotfjorden with Narvik harbor located in the fjord. From Narvik harbor the company Luossavaara-Kiirunavaara (LKAB) ships out iron ore products and has a shipping capacity of almost 30 million tonnes per year (LKAB, 2017). Figure 4.2 shows an intensity plot of all ship activity with vessels carrying AIS transmitter, both class A and B, for the time period 2016-2017 (NCA, 2019). From Figure 4.2 it can be seen that traffic mainly goes along the fjord but there are some high intensity routes going across the fjord (south-east to north-west). Figure 4.3 shows the ferry routes in the area, which fits well with the high intensity crossings.

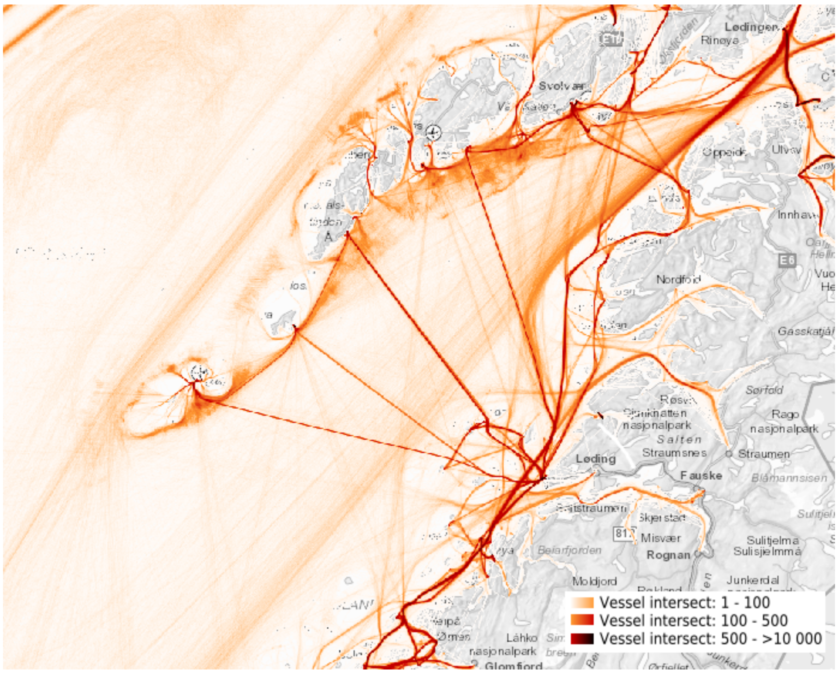


Figure 4.2: Intensity plot of Vestfjorden 2016-2017.

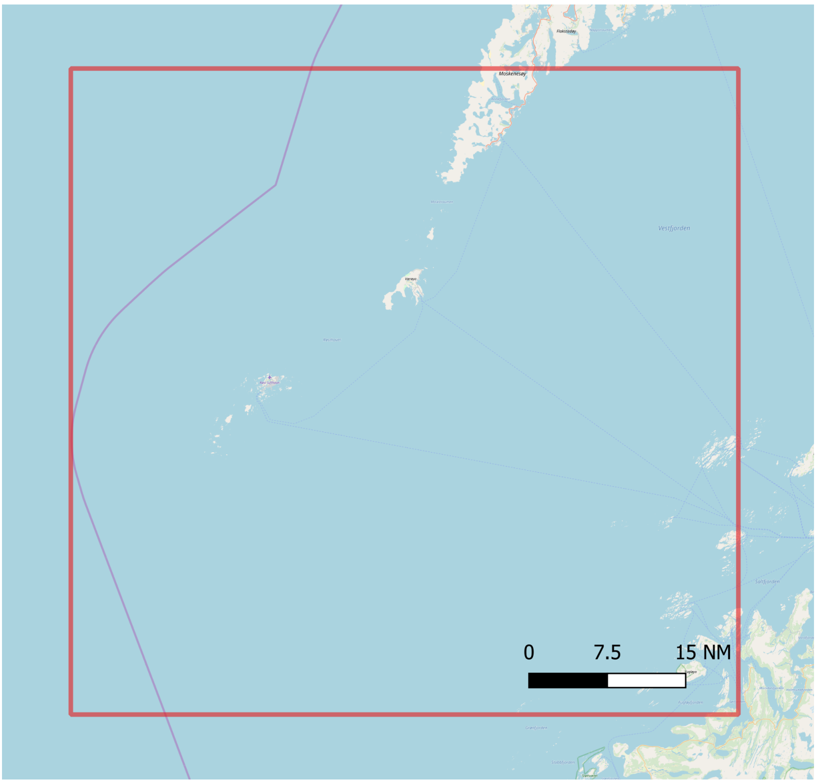
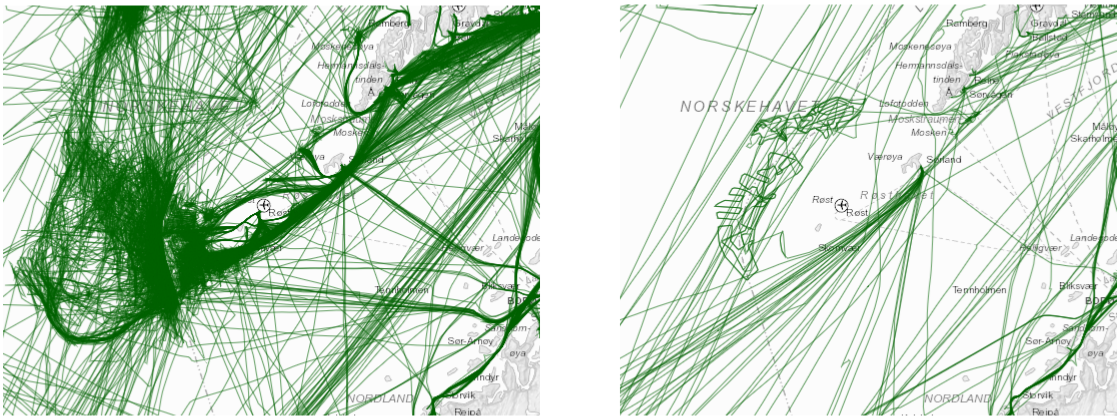


Figure 4.3: Ferry routes in case-study area marked with dashed lines. Case area is marked red frame.

Vestfjorden and Lofoten is also an important area with several fishing grounds. The main fishery in the area are after the north east arctic cod, which each year is coming to the coast to spawn (skreifiske). Skreifiske finds place each year from January to April in the Lofoten area. During skreifiske over a thousand fishing vessels are involved in the fisheries. The statistics for 2013-2015 shows that it was between 1252 and 1515 vessels involved. The majority was from the coastal fishing fleet which goes to harbor each night to unload the catch (Statistics Norway, 2017). Figure 4.4 shows intensity plot of fishing vessel movement in Vestfjorden in two different months in 2015. Figure 4.4a shows March which currently is right in the high season of skreifiske, while Figure 4.4b shows September which is out of season for the fishery. The plots show that during skreifiske it was high activity and the fishing vessels were gathering around the fishing grounds, where the cod were found. This often varies some from year to year but is still around the Lofoten peninsula. A typical picture of the fishing vessels involved in the fishery can be seen sailing out of Røst harbor in the case-study area on Figure 4.5.



(a) March.

(b) September.

Figure 4.4: Fishing vessel movement in March and September 2015. The green lines shows the track from the vessels.

Figure 4.5: Fishing vessels sailing from Røst during skreifiske (NRK, 2014).

4.1 AIS-data

For this thesis the conformed data from Nordkvist (Nordkvist, 2018) is used for the years 2013-2015. 2016 has not been used because of computer limitations and time constrain. The computer used for processing the data is a 4-core processor with a 1.6GHz CPU running Windows 10 with 8GB memory available. The results are even thought 2016 is not used, considered to give good information about the proposed model.

4.1.1 Handling missing data

The conformed datasets are the same as used in Nordkvist thesis. This means that the same problems occur in the dataset, which described by Nordkvist with missing values. For comparison reasons the way of handling missing data (SOG, COG, length) are the same as performed by Nordkvist (Nordkvist, 2018).

Missing SOG
For missing SOG value, it is by default set to 5.001knots. This results in a maximum extension of the domain in the front radius equals $\approx$ 463m. The proposed domain is much more dependent on speed than other domains in the front radius. Therefore, speed is an important parameter for the domain to be able to determine the level of risk with respect to the risk triangle, Figure3.1.

Another challenge with speed, is unreasonably high SOG values in the dataset. This is handled with setting all SOG values > 35knots to 20knots. This is a higher limit than used by Nordkvist, but it is considered that speeds up to 35knots can be assumed reasonable for instance for high speed passenger boats and rescue vessels. The choice of selecting 20knots for SOG values above 35knots, is that with 20knots the front radius will reach its maximum extension of 1 nm.

Missing COG
For missing COG, the heading of the vessel and therefore the front radius is not possible to determine. In these situations, the domain takes a circular shape equal to $1.6 * length$, which is equal to level 1 risk.

Missing length
The same default value of 40.001m is used for missing dimensions to determine length of vessel. Length is an important value, since the length determines the size of the domain direct in three dimensions and indirect in the front dimension. Missing dimensions will also make it impossible to determine true center of the vessel, but this is considered to have less impact on the results that the missing length for domain size. Nordkvist recommended the default length to be further validated. Therefore, results where the default length is used, the vessel will be tried to be identified and check out what the actual length is. The results will be presents later in this chapter.

4.2 Statistics

The statistics used in the thesis are the statistics found by Nordkvist for the years 2013-2015. The statistics is manually controlled and corrected. In the manual control, the vessel type was verified with search on marinetraffic.com or other web pages where information was found.

Several faults were found for the type of vessel which were corrected. In several cases the same vessel has different definitions for type of vessel over the three-year period. Two errors are common:

1. Small vessels (SAR, sailing vessels, tugs, dredger, coast guard vessels) was often identified as fishing vessels for the year 2014.

2. Aquaculture vessels (Fish carriers, work boat etc.) was identified as fishing vessels instead of being categorized as cargo or other, which is more suitable. In correcting these faults fish carriers are categorized as cargo and work boats as other.

A total of 4433 inputs are controlled, and 160 faults is found and changed. This gives a fault of $\approx 3.6\%$ of the total input values. If the input values for fishing vessels are assessed, it is original 1752 inputs defined as fishing vessels, which gives a fault of $\approx 9\%$ since most of the faults was fishing vessels not being fishing vessels. 9% will have an impact of sailed distance, that is not considered to be negligible. The faults found in the statistics is also corrected in the results for encounter, which will be presented later in this chapter. There was a total of 9 encounters where the type of vessel was wrong. 7 of the encounters was between fishing-fishing vessels, but after correction it is between fishing and SAR vessel. The two other vessels that was wrongly classified were a dredger and a fishing vessel, both which are corrected.

The results from the control and cleaning for the years 2013-2015 a total of 2500 unique vessels, determined by the combination of MMSI and name of vessels, was in the case-study area. Figure 4.6 shows the distribution of vessels types and Figure 4.8 shows the distribution of flag-states. The total number of flag-states are 57, but for simplicity the states with less than 10 vessels are collected under other countries.

Figure 4.7 shows the sailed distance for each vessel type in the period 2013-2015. An interesting observation can be made regarding passenger vessels, where there were only a total of 186 vessels in the case-study area, ref Figure 4.6, still they had the second highest sailed distance right below fishing vessels. This gives a good indications that there were several ferries operating quite constant at the ferry routes in the area, ref Figure 4.3. Each passenger vessel sailed much more on average than any other vessel type.

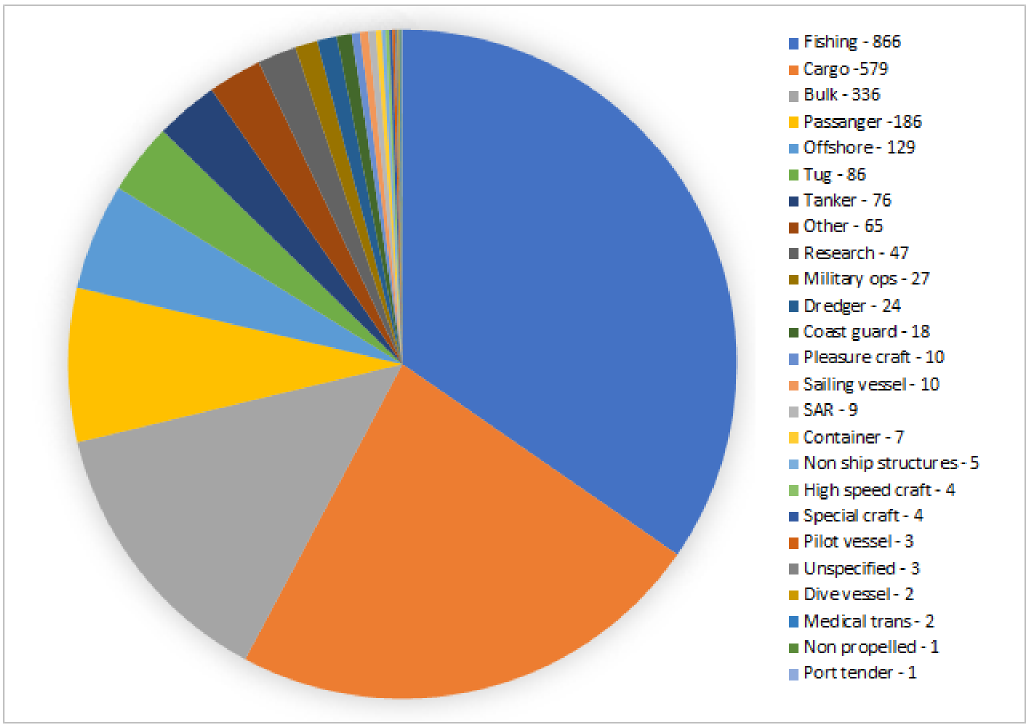


Figure 4.6: Distribution of vessel type in the case-study area in the time period 2013-2015.

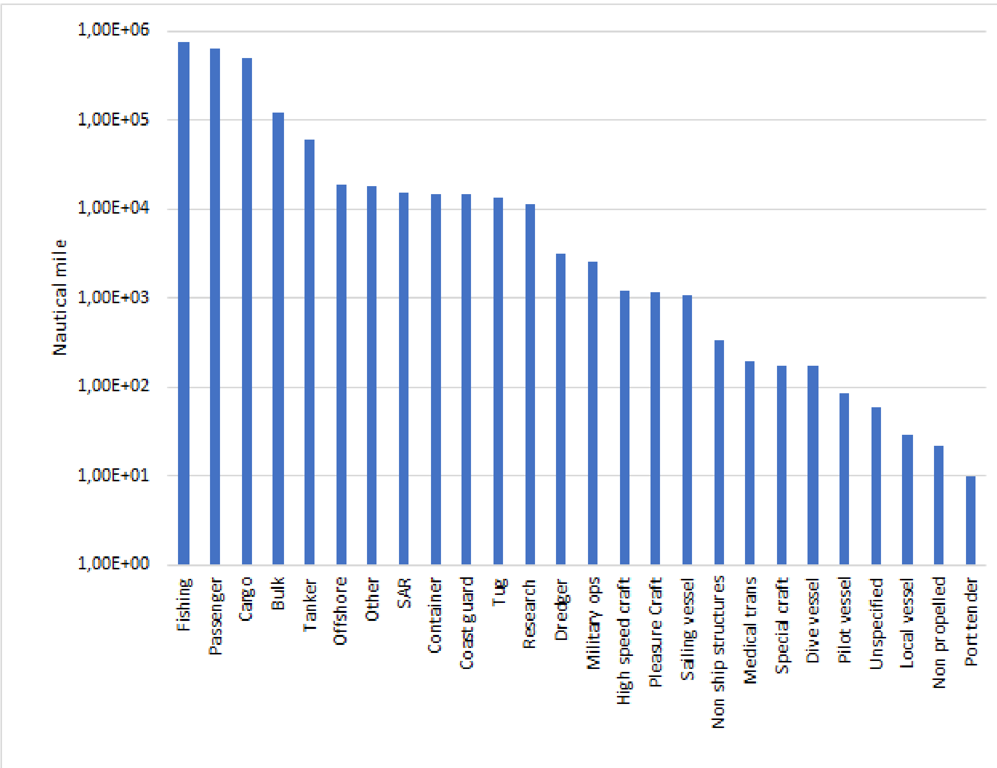


Figure 4.7: Total sailed distance for each vessel type in the case-study area in the time period 2013-2015. The vertical axis is logarithmic.

Figure 4.9 shows the distance for each flag-state. In Figure 4.9 flag-states with less than 1000nm are collected under other countries. Figure 4.9 together with Figure 4.8 shows that there were some correlation between the number of vessel each state had in the area with the sailed distance. This was dependent on the number of trips each vessel took and the nature of the journey, if it was just a sail through or if the vessel was operating in the area. A good example is Cyprus which had only approximately 24% the number of vessels compared to Russia, but still had almost the same sailed distance.

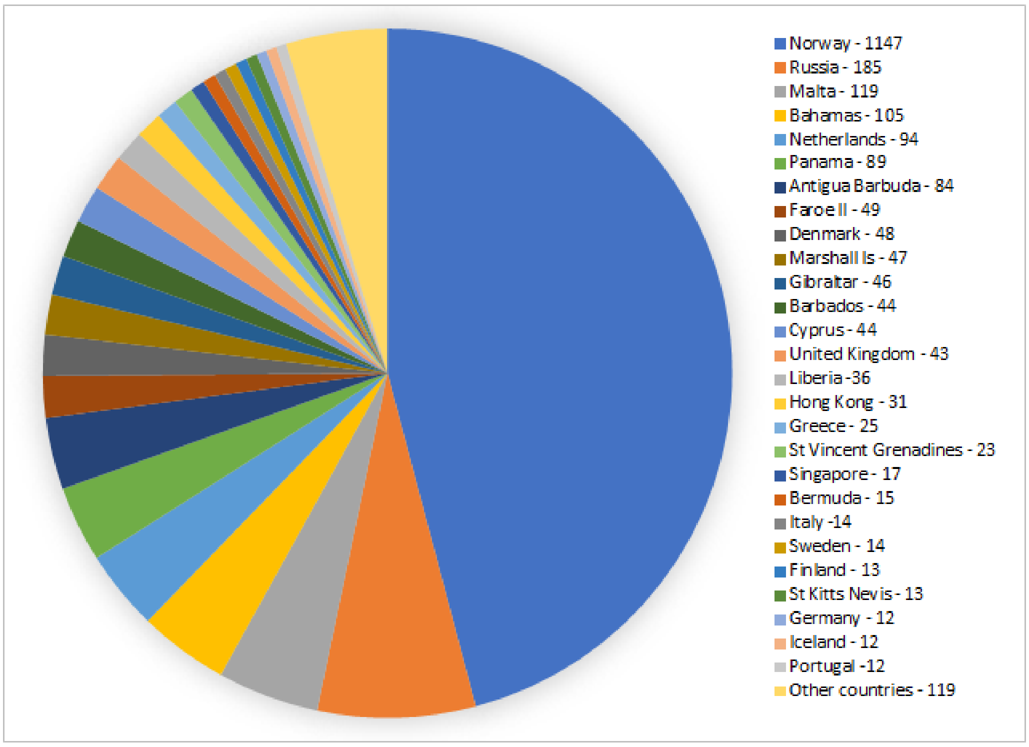


Figure 4.8: Distribution of flag states in the case-study area in the time period 2013-2015.

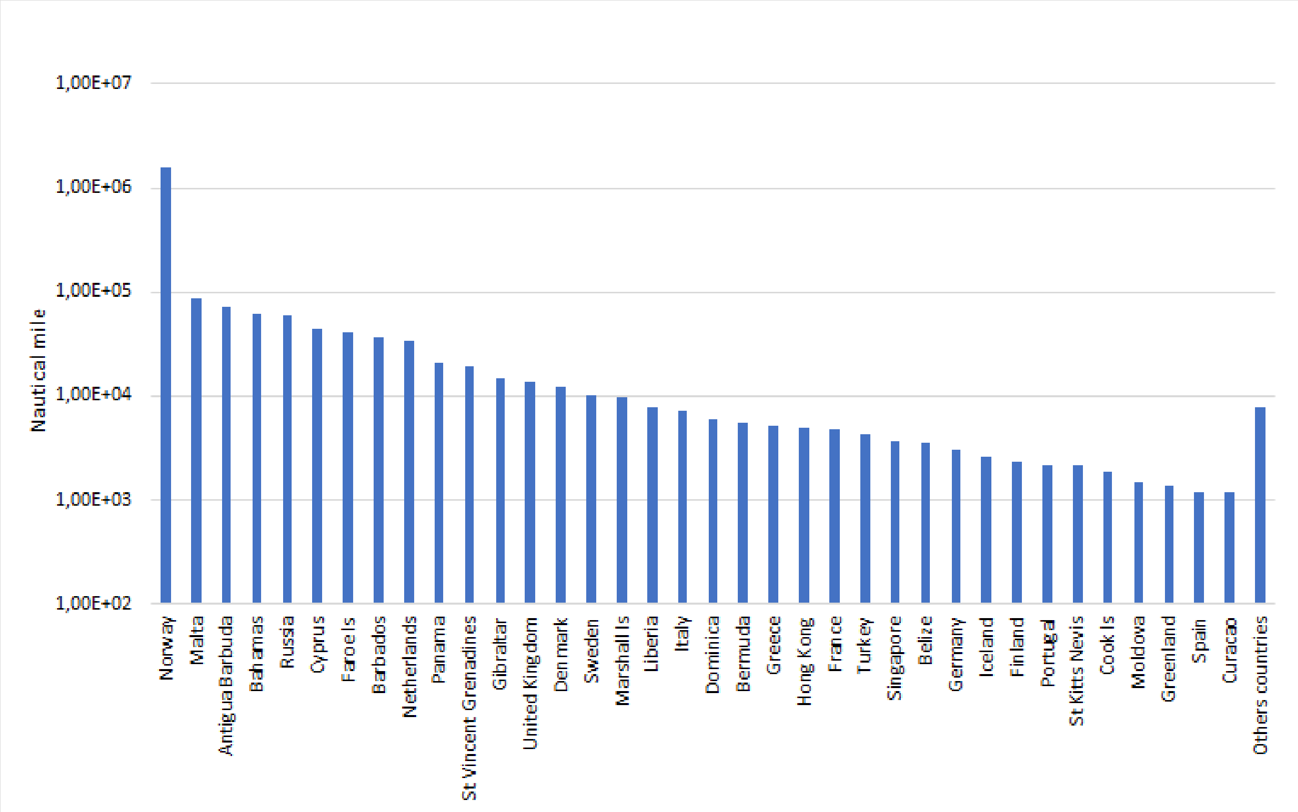


Figure 4.9: Total sailed distance for each flag-state in the case-study area in the time period 2013-2015. The vertical axis is logarithmic.

4.3 Type of encounters

For the search after encounters, Nordkvist blocked some areas close to shore. To be able to compare results the same limits are utilized in this model as well, the limits can be seen in Figure 4.10.

In the model a new way of determining type of encounter is proposed. The results for type of encounters can be seen in table 4.1. The total number of encounters are considered later in this chapter, but there should be made some comments about the number of the different type of encounter.

Table 4.1: Type of encounters for the years 2013-2015

Type of encounter	2013	2014	2015
Head-on	11	13	13
Overtaking	0	0	0
Crossing passing at stern	108	125	122
Crossing passing at bow	202	255	216
Total encounters	321	393	351

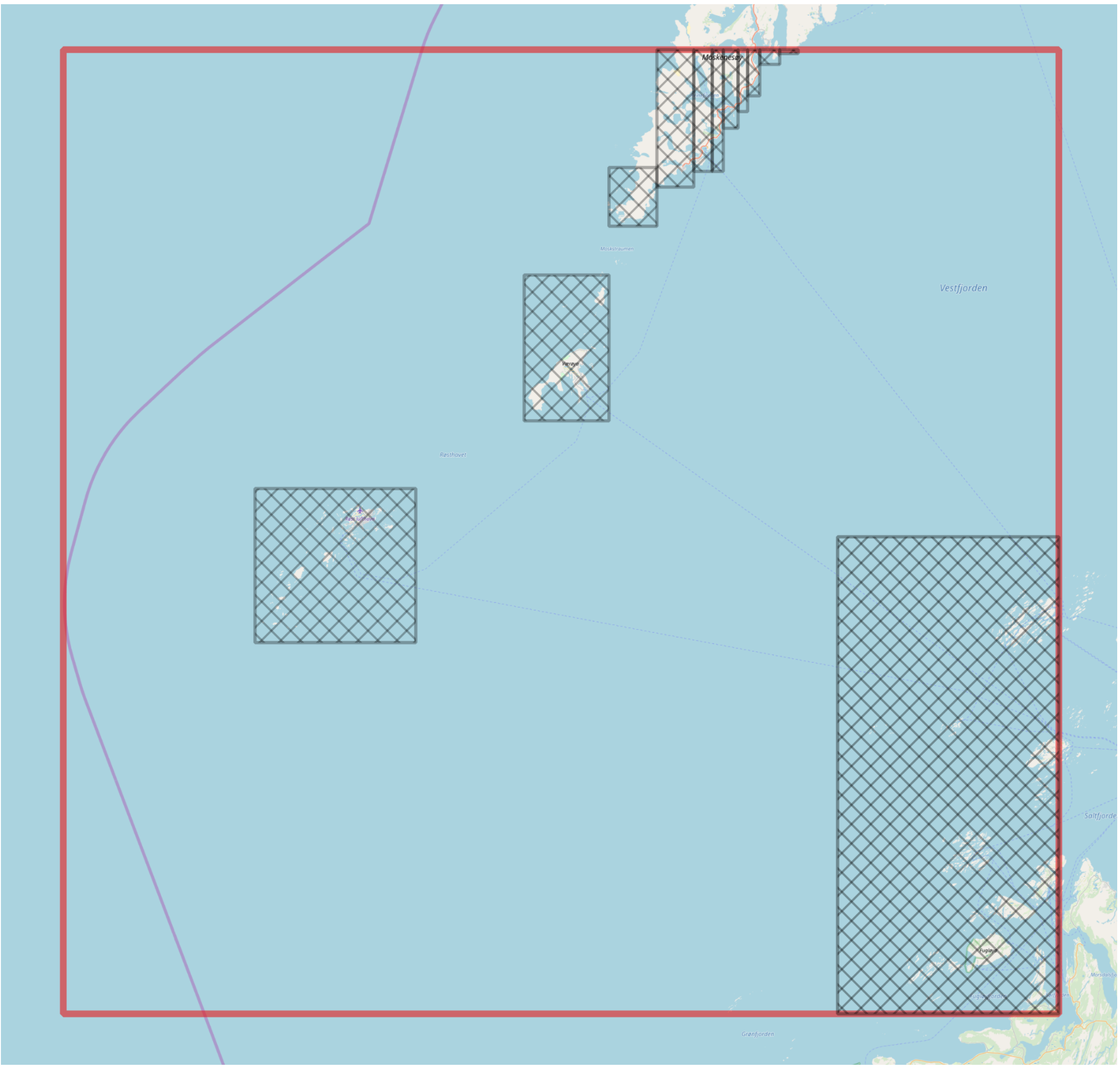


Figure 4.10: Excluded area in case-study area in search after encounters marked with grey hatched rectangular.

In this model the type of encounters are determined 6 minutes before CPA, different from Nordkvist (Nordkvist, 2018) and Iperen (Iperen, 2015) which determined the encounters at CPA. If Figure 3.6 (Defining situation of encounters) are considered, the distribution of situations detected seems reasonable with respect to the different sectors that each type of encounter is assigned in COLREG, and the area under assessment. As a reminder, all encounters are considered in sight of each other and rule 13-15 are used. The number of crossing situation had the majority of encounters each year, which is reasonable for the case-study area if the intensity plot in Figure 4.2 is considered. The intensity plot shows that the traffic in the area to a large extend is spread over the hole area with one exception; ferries which follows the same routes. This implies that most encounters should be expected to be crossing, which also is the output from the proposed method. For head-on and overtaking encounters, there was few and no encounters detected, which is reasonable based on the traffic picture. Head-on encounter will only occur if two vessels has planned the same route with reciprocal courses, which from the intensity plot is unlikely in the area. The where the vessels use the hole area for navigation. A typical head-on encounter is shown in Figure 4.11, where two ferries following the same route is in the encounter. For the study area no overtaking situations has led to domain violation, which is expected in an unrestricted coastal area. Overtaking situations is of nature occurring when one vessel overtakes another because of higher speed. In open waters there is reason to believe that navigators will overtake other vessels with a distance greater that the domain extension in the proposed model. Also, in this model the encounter type is determined 6 minutes before CPA and if the maximum extension of the domain in front radius is considered (1852m), this will be equal to a relative speed between the two vessels of 10knots for the overtaking vessel to catch up and collide with the vessel being overtaken.

The low number of head-on and overtaking situations are considered very reasonable in the area, with the traffic flow from the intensity plot. For crossings situations there is twice as many encounters with passing at bow than stern. This was expected because of the domain and models having a focus to the front of the vessels, instead on stern.

Table 4.2 shows the results from the model used by Nordkvist (Nordkvist, 2018) proposed by Iperen (Iperen, 2015) to determine encounter at CPA. The difference in the results in table 4.1 and table 4.2 are significant. From the argumentation above, it is reason to assume that the number of head-on and overtaking encounters found with the method proposed by Iperen is exaggerated. In real operations the type of encounter between two vessels are determined well ahead of an escalation of the situation to close quarters and maneuvers are performed to avoid collision. Therefore, to determine type of encounter at CPA will often give wrongly classification of the type of encounter, since in close quarters maneuvers are normally performed in the time around CPA. The proposed method in this model, is more realistic and give a more correct classification of encounter by determining it at a set time before CPA.

Table 4.2: Numbers of encounters found by Nordkvist (Nordkvist, 2018)

Type of domain violation				
Year	2013	2014	2015	2016
Number of Head-on encounters	95	52	105	16
Number of Overtaking encounters	57	28	18	27
Number of Crossing passing at bow	55	38	49	33
Number of Crossing passing at stern	36	34	40	26

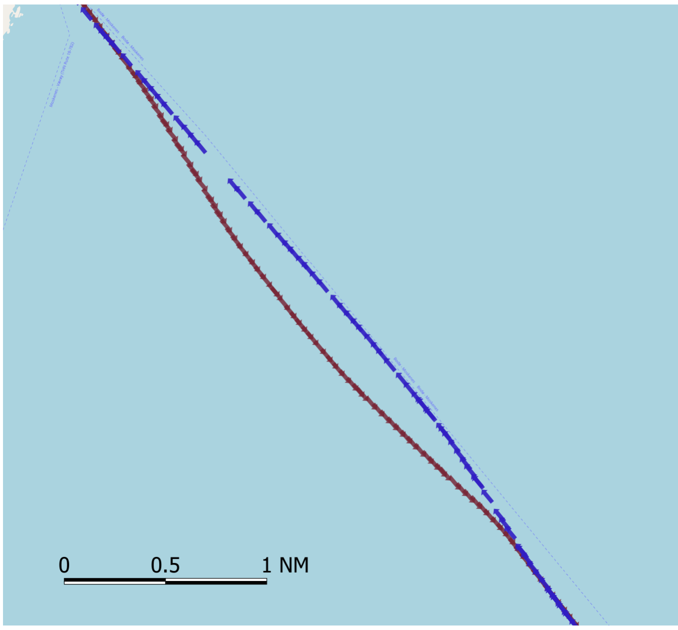


Figure 4.11: Head-on encounter between two passenger vessels 2013. Risk level 1 with a CPA 230m

4.4 Detected encounters

Table 4.3 shows the detected violations for the three year period from the model. The results are divided into two groups, based on the type of vessels that are involved in the encounter:

- "Fishing vessel". Encounters between two fishing vessels.
- "Other encounters". All other encounters that are not between two fishing vessels.

The total number of encounters in the case area seems unreasonably high for the three-year period, especially for fishing vessels. Figure 4.12 shows a plot of all detected encounters. In the plot the excluded areas in Figure 4.10 can be seen from the encounters, especially in the area around Røst and the south-east corner. Some trends can be seen from Figure 4.12, fishing vessel encounters is to a large extend located to the fishing grounds in the western part of the case area and close to shore. Other encounters are mostly towards the ferry routes in the area. For a better overview the two groups will be assessed into different sections below.

Table 4.3: Detected encounters from model year 2013-2015. The results are divided into two groups. Top part of table is "Other encounters" and the lower part is "Fishing vessel".

Risk level	2013			2014			2015		
Number of flags	#0	#1	#2	#0	#1	#2	#0	#1	#2
Other encounters									
Level 1	19	9	6	8	5	1	15	4	1
Level 2	13	1	0	10	0	0	10	2	0
Level 3	16	2	0	19	1	0	11	0	0
Fishing vessel - Fishing vessel									
Level 1	33	37	3	63	25	3	40	21	3
Level 2	64	31	1	91	39	0	81	44	3
Level 3	48	38	0	94	33	1	76	39	1
Total number of flagged encounters	193	118	10	285	103	5	233	110	8
Total number of encounters for year	321			393			351		

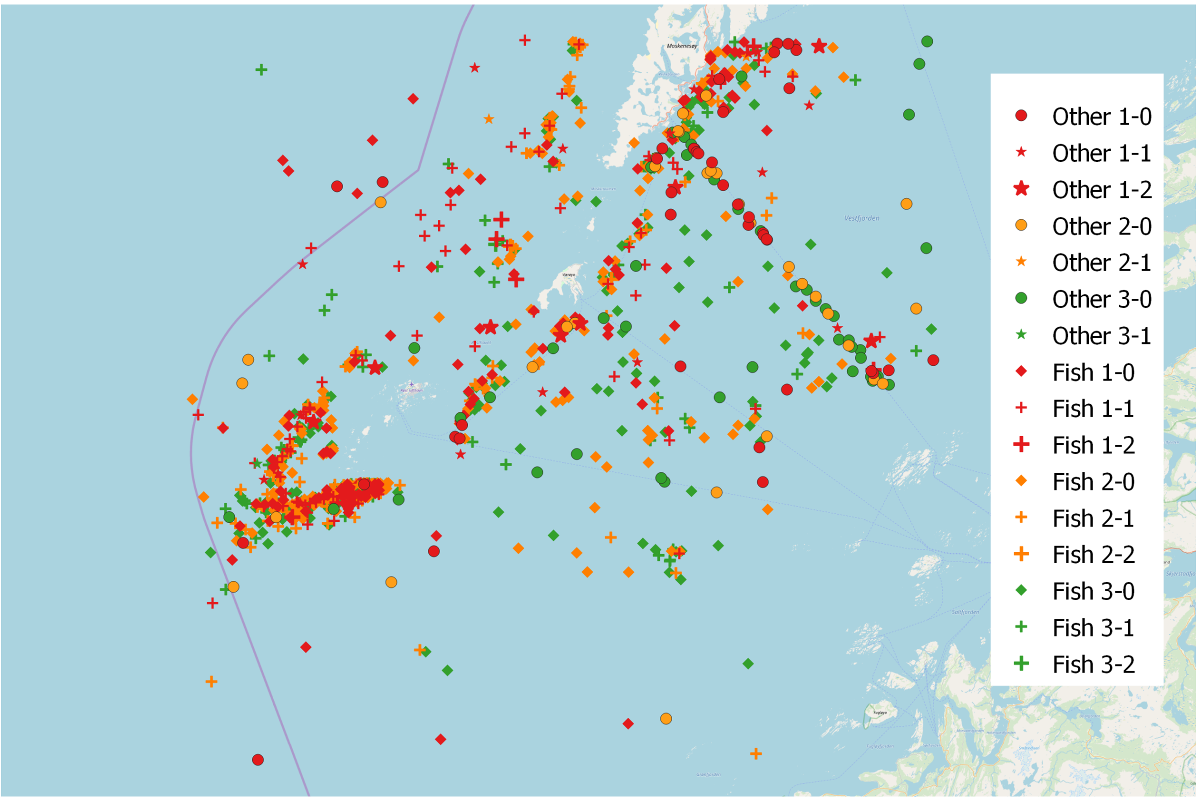


Figure 4.12: Detected encounters in case-area year 2013-2015. Symbols used are defined in the figure. The notation on the symbols are as follows; First, the group the encounters belongs to is defined (other or fish). Second, the first number indicates what risk level the encounters are at (1-3). Finally, the last number is the number of flags the encounters have (0-2). "Other 1-0" is therefore encounters in group "Other encounters" at risk level 1 with no flags, and so on.

Table 4.4 shows actual length of vessels which domain has been breached, and has missing length in the dataset and therefore has used the default length of 40.001m. The total number of encounters where vessels has unknown length is found to be a total of 60, which is 5.5 % of all encounters detected.

Table 4.4: Actual vessel length of vessels with unknown length in dataset that has used the default value in detected encounters.

Vessels with default length 40.001m in model.			
MMSI	**Actual length (m)**	**Type vessel**	**Number of encounters**
982581501	length<15	Work boat Kystverket	1
888888886	102	Cargo	1
331072001	length<10	Man overboard boat	1
273335530	28.2	Tug	1
259377000	23.43	Fishing	3
259350000	21.33	Fishing	1
259122000	40	Fishing	3
259023000	51.3	Fishing	1
258269500	47.33	Barge	1
257775500	13	Fishing	1
257688500	15.0	Fishing	12
257457700	13	Fishing	3
257432500	20	Fishing	2
257409500	27.28	Fishing	12
257258000	22.55	Fishing	1
257116000	52.55	Fishing	1
257071440	11	Fishing	2
257030540	11	Fishing	1
257026440	15	Fishing	10
257003740	10	Fishing	1
219014223	7	Pilot vessel	1

4.4.1 Fishing vessels

Figure 4.13 show all encounters between fishing vessels and Figure 4.14 shows all encounters at risk level 1 for the three year period. In both plots it can be seen that the area around Røst have a high density of encounters, which also fits with the intensity plot in Figure 4.4a showing the fishing vessels movements in March 2015. From Figure 4.14 which also shows the excluded areas, it is reason to assume that in the areas around the islands in Lofoten there would have been several encounters.

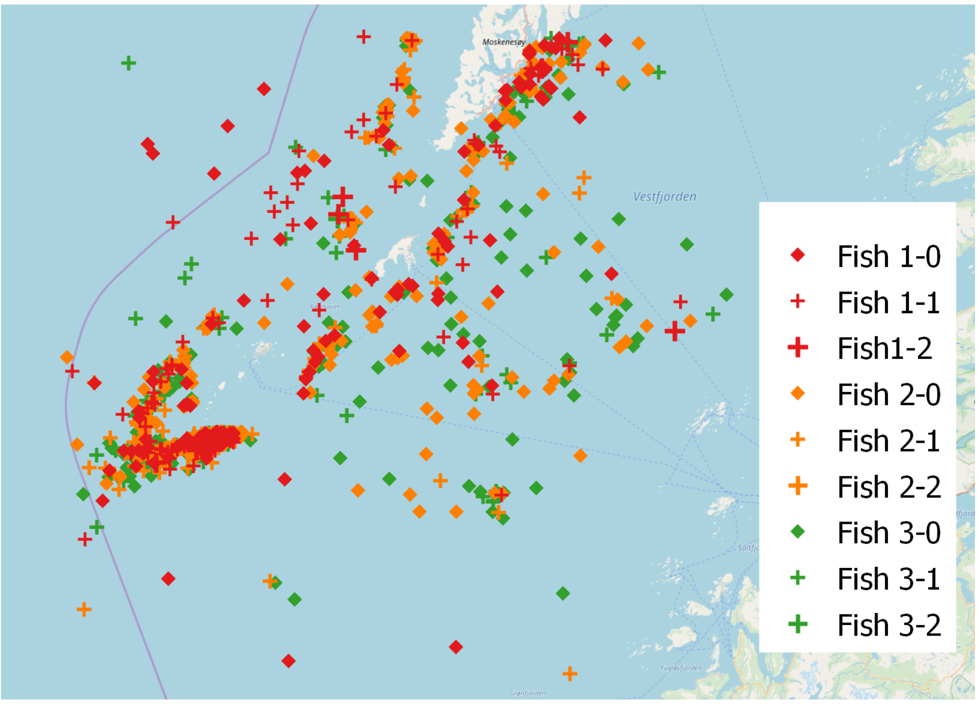


Figure 4.13: Detected encounters between fishing vessels 2013-2015.

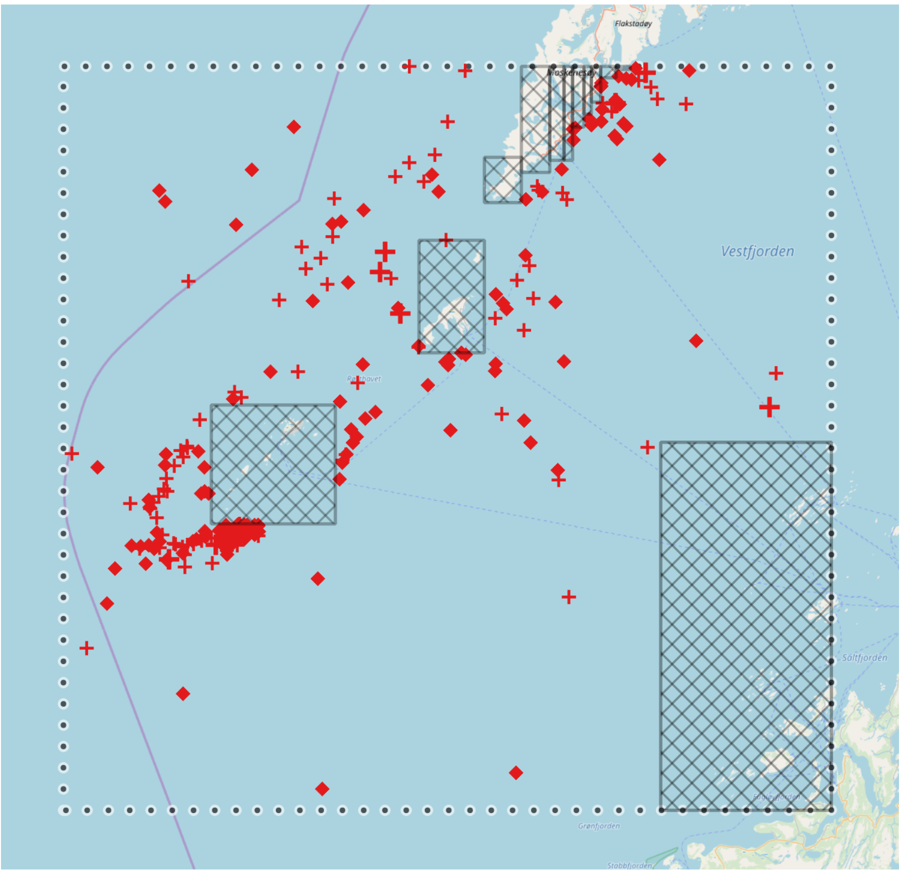


Figure 4.14: Detected encounters between fishing vessels 2013-2015 at risk level 1. Dotted line shows the case-area and grey hatched areas are excluded areas when searching after encounters.

Some examples of fishing vessel encounters follow. Figure 4.15 shows the movement of three different fishing vessel west of Røst, in a 30 minute timespan the 1st of April 2015. The movement leads to three different encounters between the vessels, at risk level 1, 2 and 3. These type of movement at mid-day for fishing vessels engaged in fishing at a fishing ground, is assessed as normal. Also, it is reasonable to believe that the consequences of a collision between two of the vessels would have been low since all vessels have a speed between 0.2-5knots in the 30 minutes assessed in the plot, all the vessels had a length of 11m. With respect to a risk assessment of traffic in the area, these are considered to be not very relevant since the consequences would be very low even if a collision was to occur.

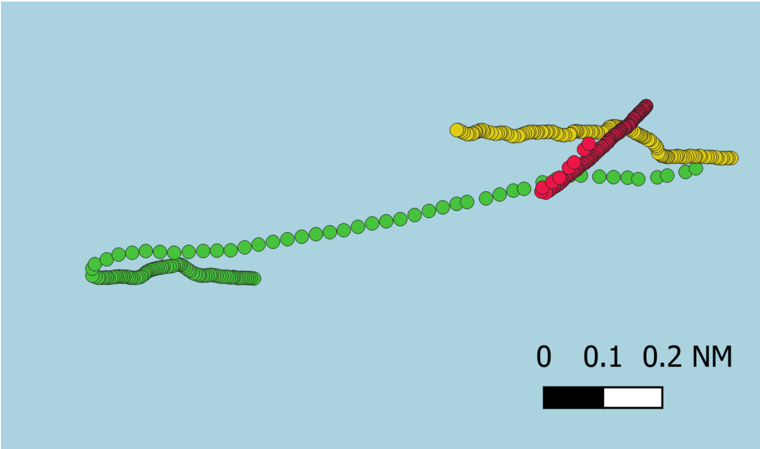


Figure 4.15: Movement of three fishing vessels at the fishing grounds west of Røst, in a time period of 30 minutes at mid-day (01.04.2015).

In Figure 4.16, which is the same area west of Røst, two fishing vessels have an encounter at risk level 1 with 2 flags, CPA in the encounter is 47m. In this encounter the vessels speed is between 4.2-8.5 knots in the 15 minutes assessed. The vessels in the encounter have a length of 20 and 22m and because of the size and speed of the encounter and the close CPA this is typical an encounter that is assessed of interest in a risk assessment point of view.

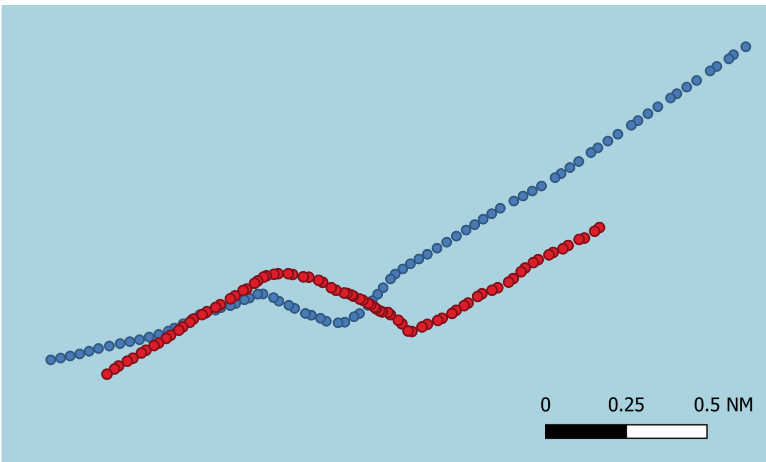


Figure 4.16: Fishing vessel encounter at risk level 1 with two flags. Both vessel sails towards south-west and have a CPA 47m (25.02.2014).

Figure 4.17 shows an encounter at risk level 1 but no flags are triggered. In this encounter some of the challenges by excluding areas can be seen. The vessels involved are sailing with speed 7-8 knots and have a length of 20 and 22m. With respect to size and speed and the very short CPA at 12m this is an encounter that cannot be excluded when identifying near encounters even if they do not break the limits of getting flagged. For this encounter maximum ROT is found to be 25.2 $^\circ/min$ and ΔSOG is 0.2knots.

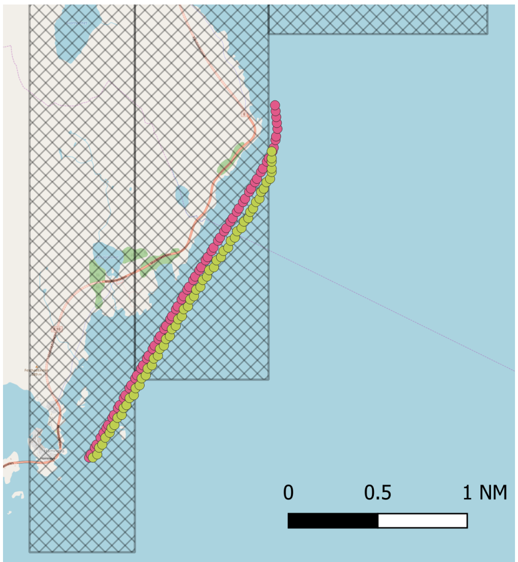


Figure 4.17: Fishing vessel encounter at risk level 1 with no flags. Both vessel sails towards north-east and have a CPA 12m (13.01.2014). Grey hatched area show excluded areas and therefore the vessels are not assessed before they alter their course strait north.

Figure 4.18 shows an encounter at level 3 with no flags. The encounter is in the south-east part of the case-area. The plot shows a classical crossing situation and CPA is 501m. Both vessels are sailing at speeds around 8 knots and have a length of 27 and 28m. In a risk perspective it is considered that this encounter is not of interest for near encounters considerations. Both vessels are keeping steady course and speed over a longer period. To have a CPA of 501m or $\approx$ 0.27 nm is considered that most fishing vessels at this size would consider safe, and that they would have time to act if something should occur.

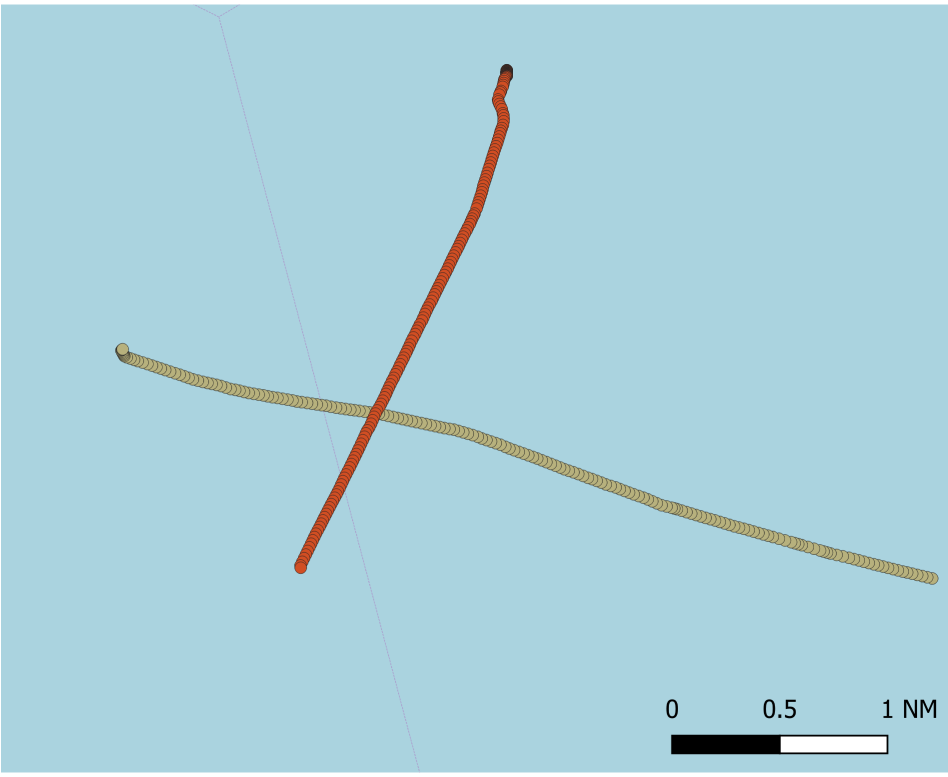


Figure 4.18: Fishing vessel encounter at risk level 3 with no flags. Orange vessel is sailing towards the south-west and the yellow sailing west- north-west. CPA at encounter is found to be 501m (10.06.2013).

It is also found that the same two fishing vessels have several encounters during the three-year period, a place between 1 and 6 encounters for the three-year period is normal. The most extreme pair of vessels are two vessels with a total of 58 encounters during the three-year period. The position of all the encounters at CPA, can be seen in Figure 4.19. In the figure there are gathering of encounters at the fishing grounds west of Røst, and a gathering of encounters in the north-east, near shore. The encounter in Figure 4.17 are one of the encounters in the north-east corner. The gathering of encounters in the north-east corner and the time when encounters occur in this area, fits with this being the area of their home harbor. The encounters are equal distributed for all three years in the first four months of the year, which is the period of the cod fishing season. The CPA of the encounters are between 7 and 809m and the risk level is in all three levels. 22 of the encounters are at level 1. The pair is responsible for 6.4% of all encounters if levels 1-3 is assessed, and 9.9% of all level 1 in the "Fishing vessel" group. The number of encounters is 5.4% off all detected encounters by the hole model, which seems unreasonably high for only two vessels when the total number of vessels are found to be 2500. By owner search of the two vessels, both vessels are found to have owners in Reine in Lofoten. Therefore, it is reasonable to assume that these vessels operate together, sharing information during the cod fisheries and uses the same harbor. Many of the encounters are therefore performed under controlled circumstances and are not the type of encounters the model has been proposed to identify.

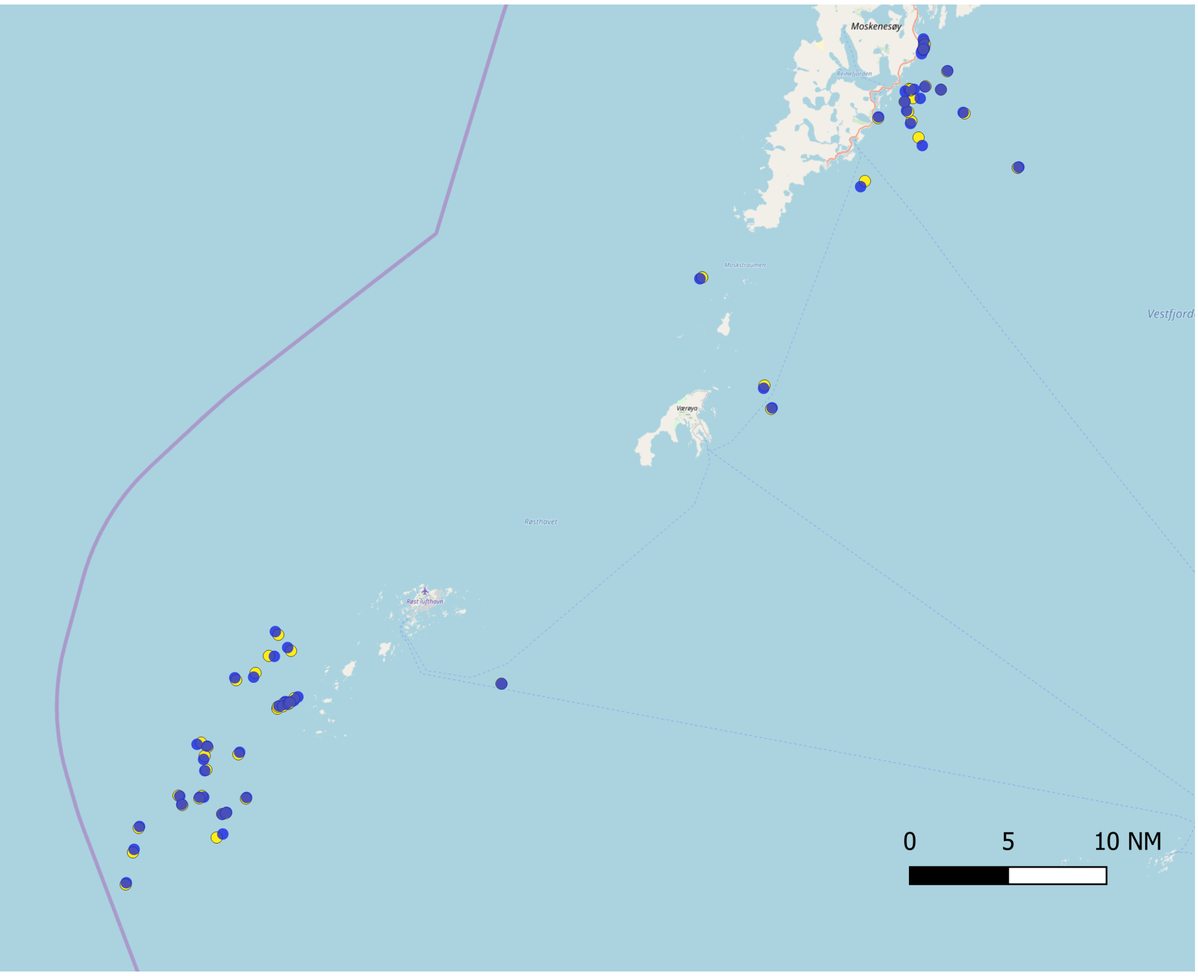


Figure 4.19: Fishing vessel pair having a total of 58 encounters during the three-year period. The two vessels have a length of 20 and 22m, both have owners in Lofoten. Vessel 1 is marked with blue points at CPA and vessel 2 with yellow. In some cases only the blue shows because the position is so close each other.

4.4.2 Other encounters

In this section encounters in group "Other encounters" are assessed. Figure 4.20 shows a plot of all encounters in "Other encounters" during the three year period and Figure 4.21 shows all encounters classified at risk level 1. Figure 4.21 shows that there are several level 1 encounters along the ferry route - Bodø-Moskenes - and especially in the northern part of the route. It should also be commented that if Figure 4.21 and Figure 4.2 (Intensity plot of Vestfjorden) is compared with respect to the excluded area in the south-east, the intensity plot in the area show that it is reason to assume that there would be several encounters in this area. It would be of interest to run the model on the excluded area in the south-east, where it is reasonable that there could be encounters of type overtaking and head-on to verify the proposed method, but because of time constrain this is not performed. The first encounter assessed is one of the encounters on the ferry route Bodø-Moskenes. The head on encounter in Figure 4.11 is between two ferries at level 1 with no flags. The encounter is somewhat typical for the ferries operating along the route, and similar encounters are some of the encounters at risk level 1 along the route. In a risk picture perspective, if consequences of a collision are considered, this is an encounter that will have the possibility of generating high risk since ferries are carrying several people. The data from the encounter show that the CPA is 230m and the ferries are sailing with a speed of 17 and 18knots at CPA. The ferries are sailing at almost reciprocal courses, which gives a relative speed of $\approx$ 35knots. Both vessels have a length of 96m and the forces in a collision between these two ferries would be high. Therefore, it is reasonable to assume that consequences could be high. Figure 4.22 shows 7 other encounters between ferries on the same ferry route between Bodø-Moskenes. 5 of the encounters are in 2013, 2 in 2014 and non is found in 2015, which can be an identification of procedures for ferries passing each other has been changed. The CPA in the encounters are between 119-149m, and speed in the range 16.4-21.2knots, which from a navigational risk perspective is unreasonably high.

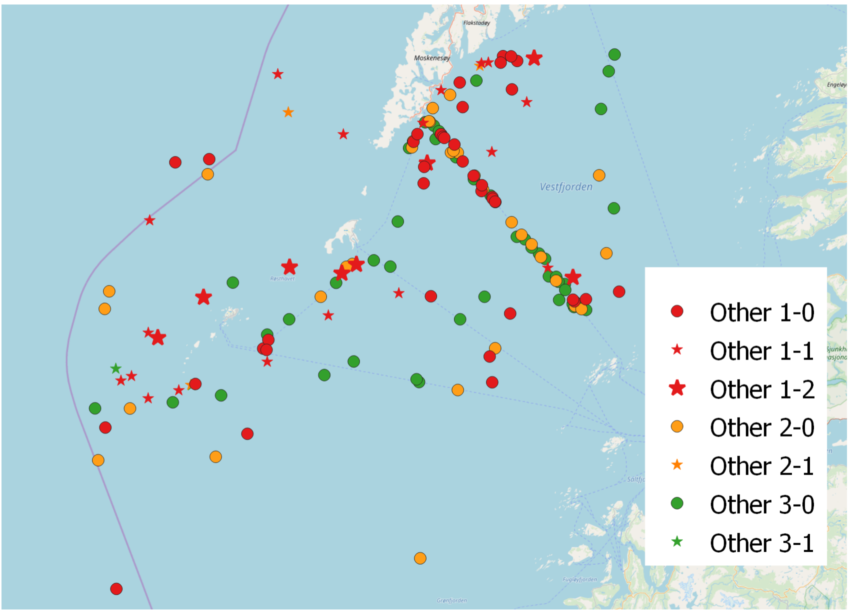


Figure 4.20: Detected encounters between "Other" vessels 2013-2015.

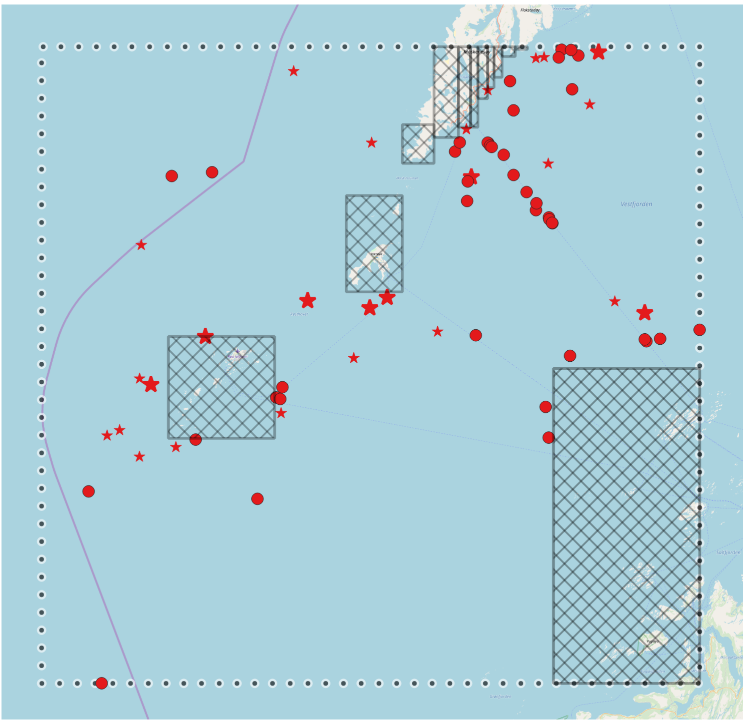


Figure 4.21: Detected encounters between "Other" vessels 2013-2015 at risk level 1.

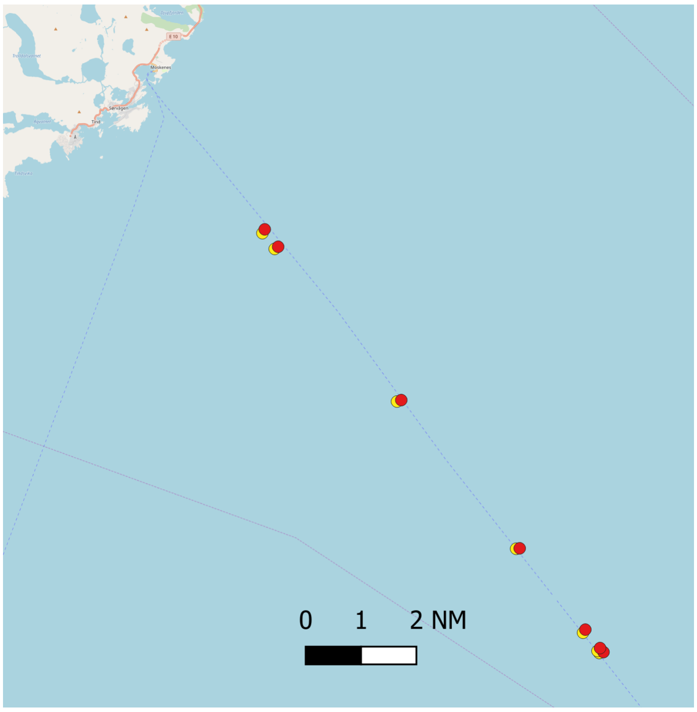


Figure 4.22: Detected encounters between ferries 2013-2014 at risk level 1 on the route Bodø-Moskenes, no encounters found in 2015.

Figure 4.23 shows a known collision between a ferry and a fishing vessel in 2013. This is a known collision used by Mestl (Mestl et al., 2016) in his study to identify if ROT can be used to identify encounters at sea, which is used as argumentation when limits are set in this model. An interesting observation is made that the CPA is found to be 50m in the proposed model, which can be explained with the data utilized uses 10 seconds intervals for 2013 and therefore the time of impact is not in the dataset. The collision was also identified by the model proposed by Nordkvist (Nordkvist, 2018).

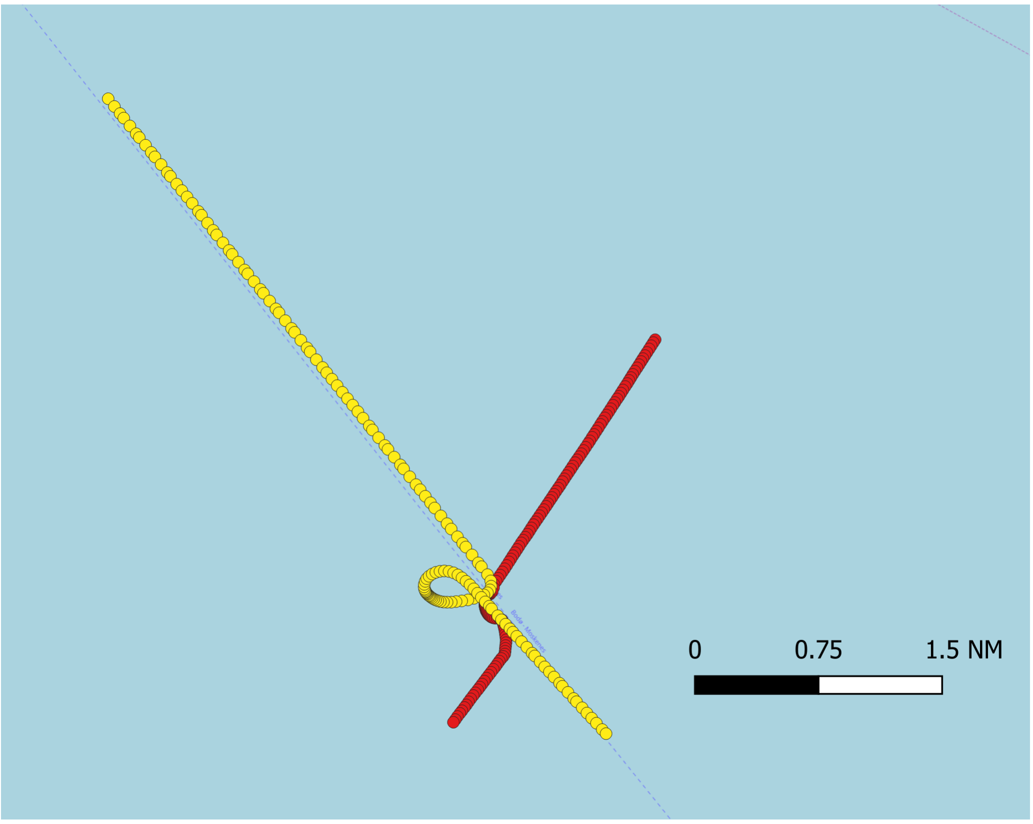


Figure 4.23: Collision between ferry and fishing vessel. The encounter is classified at risk level 1 with 1 flag, CPA 50m (16.06.2013). Ferry is yellow plot and is sailing south-east at speed 18knots until impact while fishing vessel is sailing south-west.

To compare the proposed models behavior toward the model proposed by Nordkvist (Nordkvist, 2018), encounters between ferries and cruise ships are considered since this is encounters plotted by Nordkvist. Figure 4.24 shows two encounters found by Nordkvist. In these cases, the ferries are sailing south-east and are turning away from the cruise ships, which are sailing north-east. In both cases, the domain breached is the cruise vessels domain and the CPA were larger than 600m in both cases. None of these encounters are found in the model proposed in this thesis, mainly because the side sector (radius to port and starboard) is reduced. It is considered that the side sectors are overestimated in the model proposed by Nordkvist. From the data in the plot and the distance at CPA, it is from a navigator's perspective, not reasonable to consider this as a near encounter. It is therefore considered that the proposed model in this thesis, makes the right assessment of not detecting these as close encounters. The speed of the ferries is not given in the plot, and this would be an important parameter for determining the front radius of the ferries in the proposed model. It is reasonable to assume that the speed and the give-away maneuver performed by the ferries, are performed at an early stage with respect to time and therefore the encounters are not detected by the model. In the model in this thesis two encounters between ferries and cruise vessels are detected both at risk level 3. One of the encounters can be seen in Figure 4.25. Based on the courses at CPA and a CPA of 1162m it is considered reasonable that this is a level 3 encounter. If the encounter also is found by Nordkvists model is not possible to determine, since they not are plotted.

The total number of encounters between passenger vessels (ferries and cruise) with the proposed model at level 1 is 12 for the three-year period, distributed equal over the three years. This is mentioned since a collision between two passenger vessels is considered to have the greatest potential with respect to loss of human life's as a consequence.

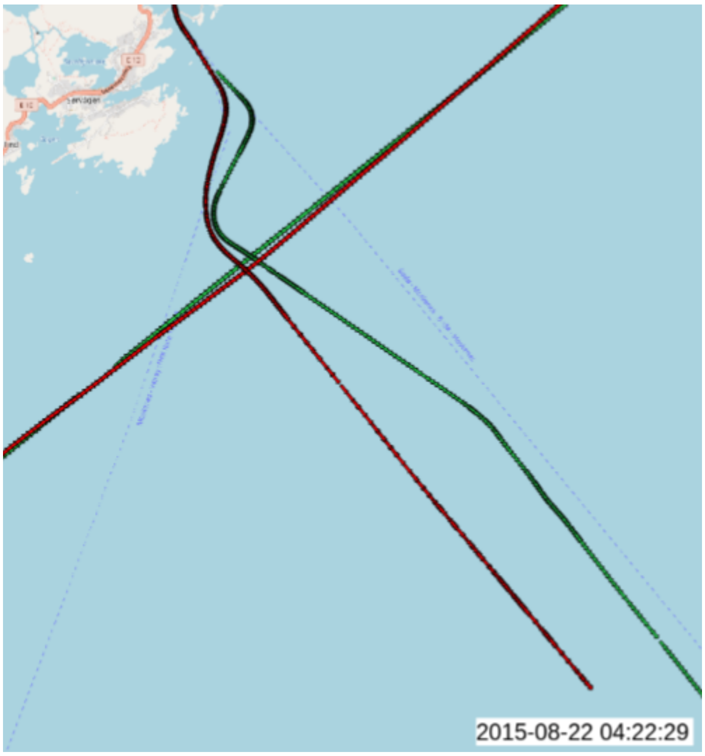


Figure 4.24: Encounter found by Nordkvist (Nordkvist, 2018). The encounters are not detected with model proposed in this thesis.

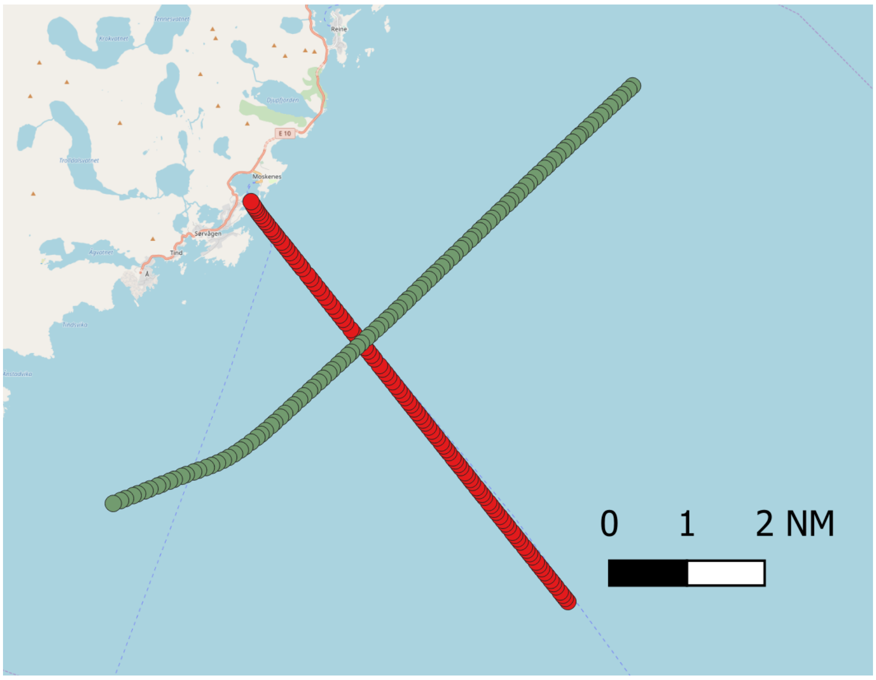
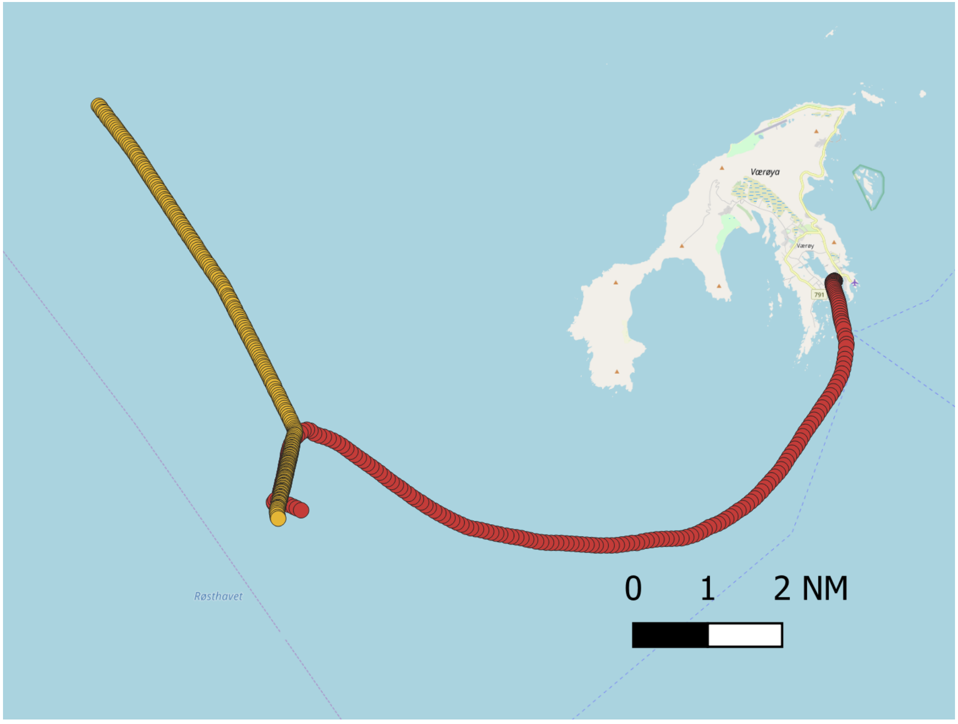
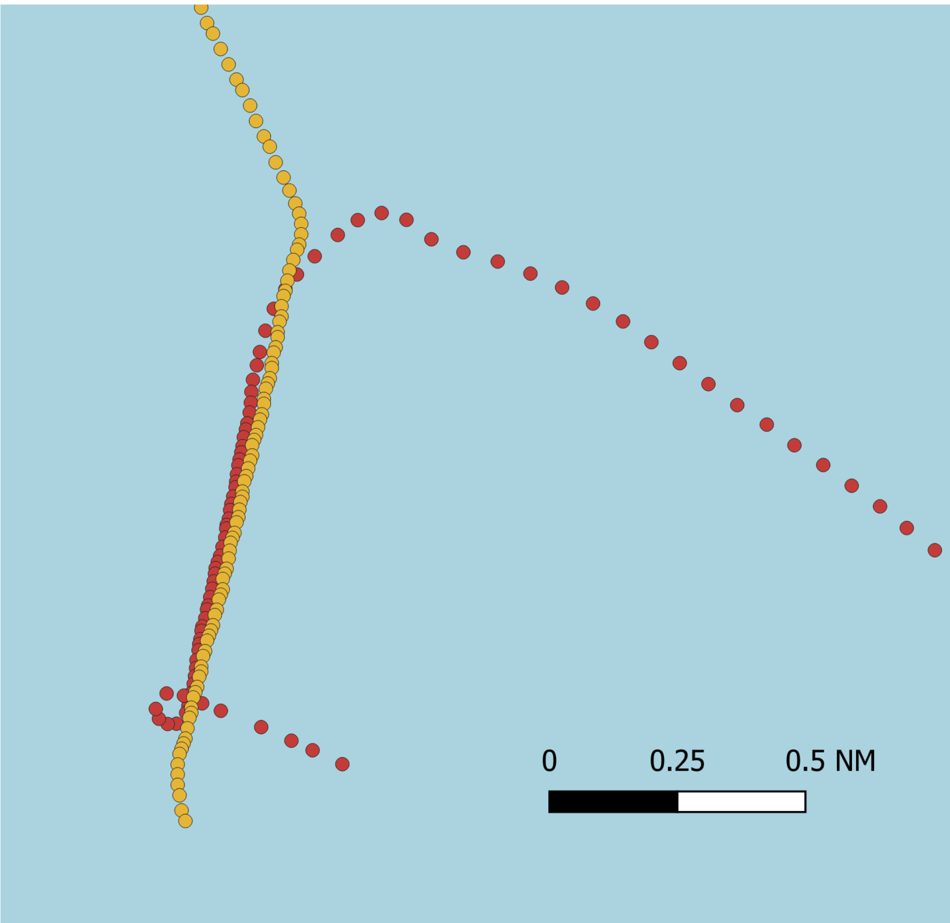


Figure 4.25: Encounter between ferry and cruise at level 3 without flag. Ferry red plot sailing north-west at 16 knots and Cruise green plot sailing north-east at 20knots, CPA 1162m

Further on, the vessel type SAR is involved in 29 encounters at level 1 for "other vessels" during the three-year period. 8 of the encounters with SAR vessels are flagged with two flags and both limit for ΔSOG and ROT are breached. Figure 4.26 show the closes encounter with a CPA of 2.3m found by the model. The encounter is between a SAR vessel and a fishing vessel, classified at risk level 1 with 2 flags. Figure 4.26a gives an overview in a 40 minutes timespan where it is clear that the SAR vessel sets best speed from the harbor and start approaching the fishing vessel at 18:35. Speed increases from 0 to 23knots within 6 minutes, which is normal time they use from letting go at the quay and sailing out of the harbor where its normal to have speed restrictions. The speed of the SAR vessel is steady at 25-26knots all the way until the turn can be seen in Figure 4.26b. After the vessels has sailed with parallel courses for a while, the SAR vessel sets course to harbor again at best speed. When the encounter is plotted and speed/course is evaluated, it is very reasonable to assume that this is not a near encounter in the form that the model is proposed to identify, but a rescue operation where closeness is necessary to perform the operation.



(a) Overview of movement for both vessels, timespan 40minutes.



(b) Close up of near encounter.

Figure 4.26: SAR vessel encounter. SAR vessel is red plot and fishing vessel is orange, CPA 2.3m (20.03.2013).

Figure 4.27 shows another encounter between a SAR vessel and a fishing vessel. In this encounter the level of risk is 1 and from the plot and the course/speed picture this seems to be an encounter of interest in this model. Therefore, if SAR vessels are filtered out of the dataset the risk of losing actual encounters is present. SAR vessels also in general has high speed potential and can therefore generate quite high forces in a collision.

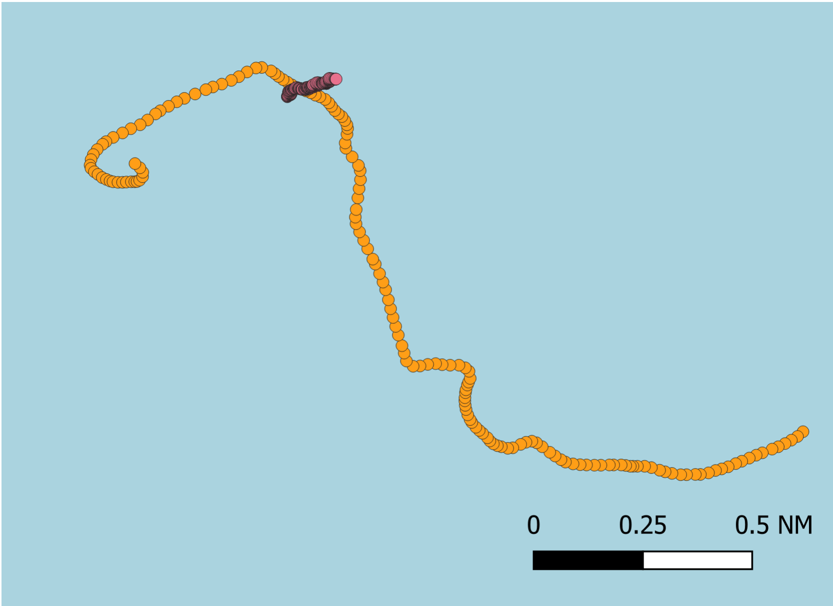
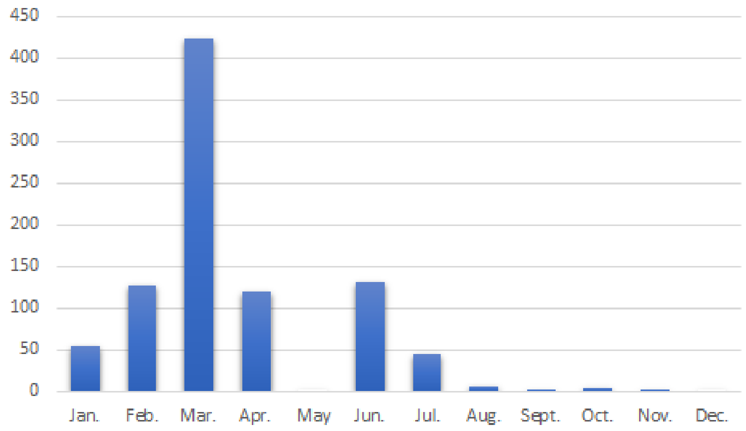


Figure 4.27: SAR vessel encounter where the SAR vessel is the orange plot and fishing vessel is the pink one. Encounter at risk level 1 with no flags, CPA 71m (15.03.2013).

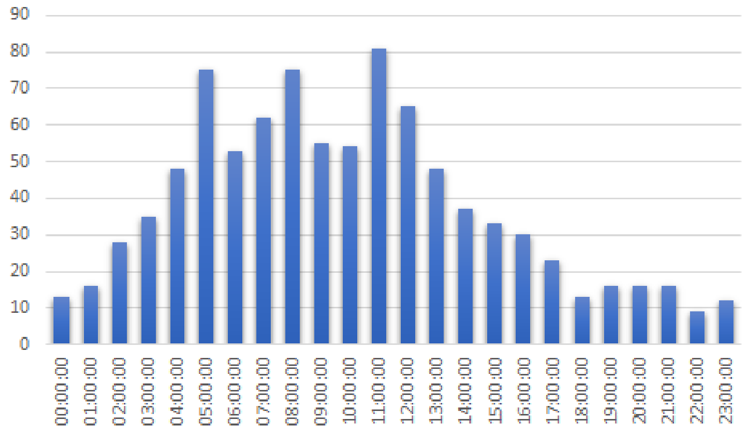
A final observation in the results for other encounters, only 16 encounters involving vessels above 100m during the three years even though several vessels above 100m makes several trips through the area. The low number of encounters involving larger vessels can be explained with quite normal behavior of smaller vessels, keeping a larger distance the bigger the other vessel is. It should also be mention that 6 of the 16 encounters are at risk level 3 with the ferries crossing Vestfjorden, and 4 is with SAR vessels at level 1. Two vessels were also detected with faulty length when assessing vessels above 100m. The first had a dataset length of 440m and from manually control a real length of 33.05m (MMSI 258225000) and the second has a dataset length of 100m and a real length of 21m (MMSI 257548600). The fault of length has a great influence on the side and aft domain size, especially when the fault is so significant.

4.5 Initial frequency assessment

Figure 4.28 and Figure 4.29 show the distribution of encounters during year and day, for both of the groups "Fishing vessel" and "Other vessel". If the group "Fishing vessel" and Figure 4.28 is considered, it is peaking in the beginning of the year with a significant peak in March, Figure 4.28a. The distribution during the year is reasonable because of the high number of vessels participating in skreifiske in the first moths of the year. It is also reasonable that the main peak is in March since this is the period when the cod is closest to shore, therefore the fishing vessels are fishing in the case study area. The distribution during the day is also following the natural behavior of the coastal fishing fleet, Figure 4.28b. The coastal fishing fleet has a normal operation pattern where they sail out in early morning to the fishing grounds and returns in early afternoon to unload. This fits very well to the distribution of encounters during the day.



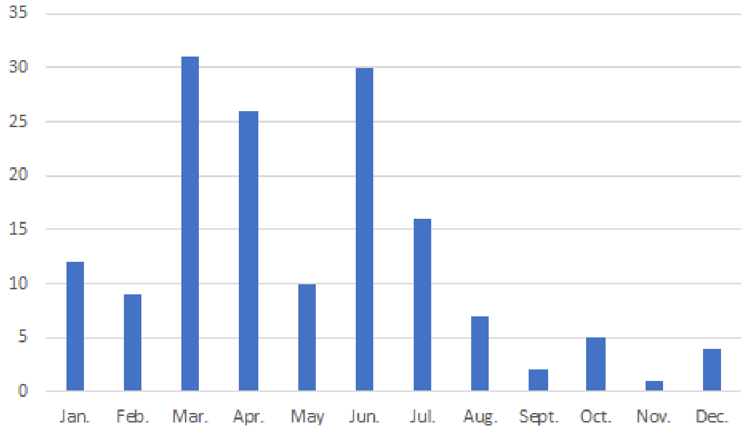
(a) Month.



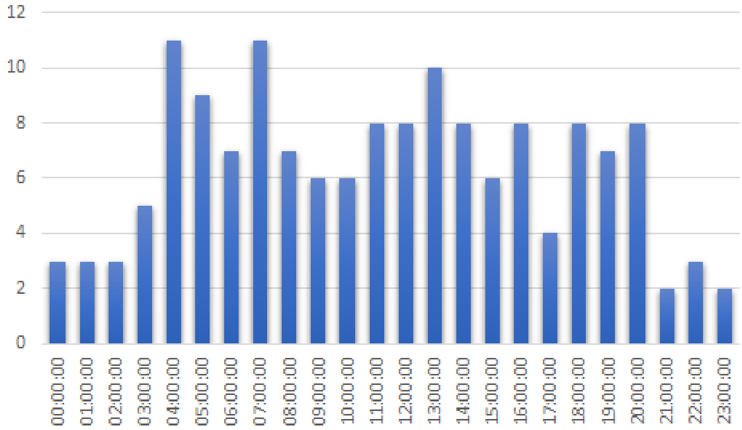
(b) Hour.

Figure 4.28: Distribution of fish encounters for the period 2013-2015 for months and time at day.

For "Other encounter" the distribution during the year, Figure 4.29a is slightly different than for fishing vessel. Most encounters are still in the first half of the year. Since there in other encounters still can be one fishing vessel in the encounter, it is reasonable that there will be encounters between fishing vessels and other vessel types during the skreifiske season. In the section above, the number of SAR vessel detected in encounters at risk level 1 is 29. SAR vessels are involved in 29 encounters with fishing vessel at any risk level, which is mainly during skreifiske. During the month of June and July there are a peak that cannot be seen for fishing vessels. There are still fishing activity in the case area, but the activity is more scattered over the area which can be seen in Figure 4.30. In the summer months there are also more cruise and tourist activity that leads to detected encounters. Figure 4.29b show that the distribution during the day, where several encounters can be seen later in the afternoon and early evening than for the group "Fishing vessel", but at night the same low number of encounters is detected.



(a) Month.



(b) Hour.

Figure 4.29: Distribution of other encounters for the period 2013-2015 for months and time at day

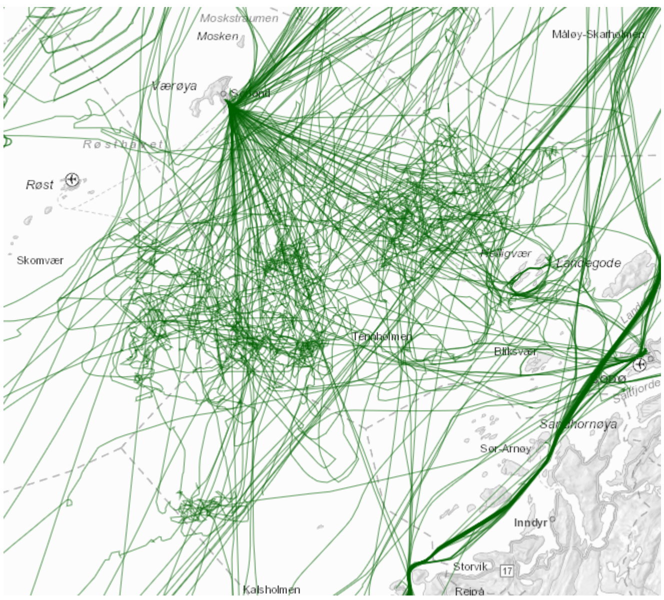


Figure 4.30: Fishing vessel activity June 2015.

From the statistics and the results of encounters a frequency can be calculated. Table 4.5 shows the initial frequency given as the number of expected encounters per 10^5nm. The frequency is given for all encounters detected in the area, and for the groups "Fishing vessel" and "Other encounters". In the table both frequency for all risk levels and for risk level 1 is presented. When calculating the frequency, for the different type of encounters, the statistics used is:

- All encounters, the total distance of all vessels is used.
- Fishing vessel, the total distance for all fishing vessels are used.
- Other encounters, the total distance of all vessels is used since all type of vessels can be involved in the group "Other encounter".

Table 4.5: Number of encounters for each vessel group per 10^5nm

Number of encounters per 10^5nm			
Groupe	**All encounters**	**Fishing vessel**	**Other encounters**
All risk levels	48.3	121.7	7.0
Risk level 1	13.5	30.4	3.1

The results show that the highest frequency of encounters is found amongst fishing vessels, which is reasonable considering the high number of encounters. The fact that the only distance used for calculating the frequency for fishing vessels are the fishing vessels distance, also increases the frequency compared to the other calculation performed. To emphasize this, if encounters where SAR vessels are considered the calculations show that SAR vessels has several near encounters per 10^5nm, at all levels: 189.8 and at level 1: 124.4 based on the number of encounters and the sailed distance for SAR vessels. Another very important aspect when reading the frequencies are the excluded areas when searching for encounters. Figure 4.31 shows the three areas that is considered to have the biggest impact on the frequency as an overlay to the intensity plot for the area. The statistics have been calculated for the hole case-study area by Nordkvist, while for search the encounters has quite large areas that are excluded. The three areas in Figure 4.31 has a total area of 565.7 nm^2, which is significant when considering the sailed distance that is reasonable to believe is generated in these area. A reasonable assumption is also that several encounters would have been detected in the excluded area, based on the intensity of the traffic in these areas. The total case-study area is 3750.7nm^2, meaning that the three excluded areas is covering 15% of the total area. Since there are several uncertainties with the model at current state, and the fact that the statistics does not fit the actual area for encounter search, no further calculations around frequency is performed.

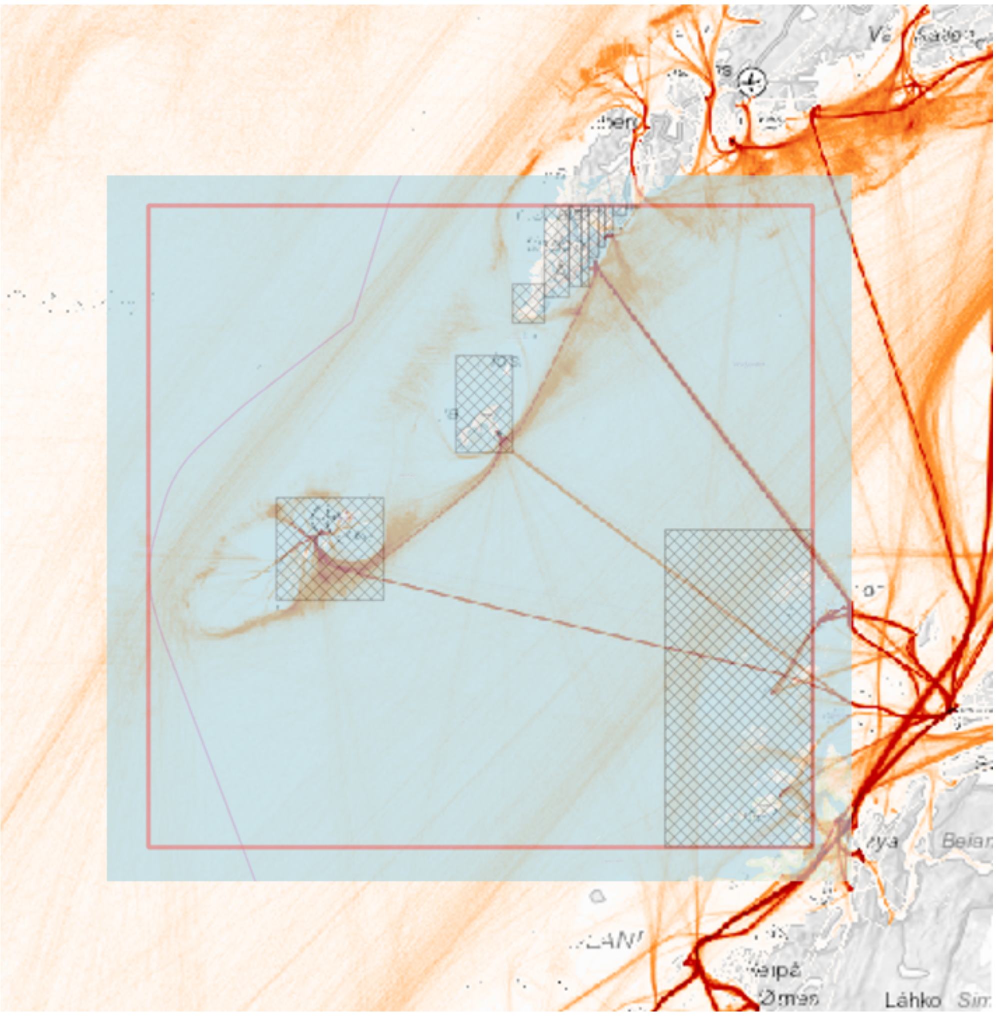


Figure 4.31: Overlay of excluded areas on intensity plot for the case-study area.

Chapter 5

Validity and Discussion

To make the discussion clear it is divided into four separate sections. The first sections discuss the model proposed in this thesis with focus on the method, assumptions and limitations taken. Second section discuss the findings from validation of AIS-data. The third section discuss risk assessment in the current model and in general. The last section is a comparison of results from the model proposed in this thesis compared to the results found by Nordkvist (Nordkvist, 2018).

5.1 Proposed model

The proposed model detects encounters based on the parameters defined in the thesis and rank these encounters with respect to the risk levels introduced in the model. By introducing risk levels, the result is easier to assess. The risk levels defined in this report are based on the distance between the vessels, and the area of TCPA is given by the properties of the three different risk levels. By separating encounters in different risk levels, the model can be utilized to different use. In areas with high activity and many encounters, one can utilize level 1 encounters in the risk assessment and in areas with lower activity all three levels of risk can be utilized.

It is considered that a model for performing risk assessments on historical data and live risk assessments will deviate from each other. In this model a new ship domain is proposed with the purpose of being used as a tool to assess historical AIS-data, which is an important limitation of the model. A domain for a live risk assessment would be designed with more dynamic boundaries, dependent on the area of operation and type of encounter. This is supported by the finding in the study performed by Hörteborn (Hörteborn et al.,) where they concluded with "the geographical characteristics of different waters influence the shape and size of the ship domain; and the type of intersections influences the shape and size of the ship domain.". The proposed domain in the model is a quite simple domain, but still robust and easily understandable as a tool for assessing historical data. The domain has focus to the front of the vessel, which is considered the area that most navigators focus on. The front radius in the domain is determined by the speed of the vessel and a three-minute vector. The choice of selecting a three-minute vector or 180s is from the properties of marine radars where three- or six-minutes vectors are normal. The area behind the vessel is in the domain considered to have the same extension as the sides (port, starboard) of the domain. The area behind a vessel will be covered by the other vessel's domain in a crossing or overtaking situation and is from a navigator's experience not more important than the areas to the sides. The minimum extension of the domain is set based on the findings in Hansens study (Hansen et al., 2013). The minimum extension should be considered as a conservative limit, especially for smaller vessels based on experience from sea. With the input in the current model, being only AIS data, it is considered reasonable that the domain complexity should reflect the data input. Therefore, the quite simple domain in current model is considered advanced enough based on the data available and validity of the data, which will be discussed further below.

Two parameters are chosen to flag encounters in this model, ROT and ΔSOG. It is considered important parameters to identify encounters where something extraordinary happens, which the results support. But, as seen for SAR vessels it is not necessary a near encounters of interests even though the encounter gets flagged. If encounter gets flagged it should be considered as an encounter to be further investigated, not that it necessary is an exceptional encounter. The limits used for both ROT and ΔSOG is in the current model based, to a large degree, on assumptions and few studies has been performed to set reasonable limits for different type of vessels. Further work to determine limits for different types of vessels, to get more correct limits for the different type of vessels is needed. It should also be mentioned that in current model ROT is calculated based on COG, and the option of ROT in the AIS messages are not utilized. In an evasive maneuver it is considered that the ROT would peak at some point during the turn, and it is not reasonable to consider it as constant which is the case in the current model. Using COG and formula, ROT will be flattened which must be reflected when setting the limits for ROT. The ΔSOG in current model is also considered to be very conservative, where the limit is set to $\Delta SOG > 3$knots during a 30 second period. In the results a total of 46 vessels has a $\Delta SOG > 3$knots. The longest vessels breaching the ΔSOG are the ferries going between Bodø-Moskenes, with a length of 96 meters. This shows that ΔSOG has a potential to be an indicator of something extraordinary happening in an encounter, but the limits should be further investigated. It is also worth mentioning that SOG is one of the input values in the data set that are most often missing, which makes it harder to calculate ΔSOG for a time period greater than 30 seconds with the current dataset.

In the model a new way of determining the type of encounter (Head-on, overtaking, crossing) is proposed based on COLREG rule 13-15, motor vessels in sight of one another. For determining which vessel has the right of way, both vessels are in the model considered as motor vessels, since navigational status are not available in current dataset. It is considered that the fault of classifying every vessel as motor vessels, is less that using navigational status in AIS messages, which from experience often is faulty especially for fishing vessels. It is important to emphasize that if rule 18 is to be used in the model, the definition on what is categorized as a fishing vessel, sailing vessel and so on must be considered from rule 3 in COLREG. In current model the type of encounter is determined at the point 6 minutes before CPA, by use of the vessels COG and relative bearing between the vessels. Plots and assessments of encounters in the results, show that the classification of the encounters seems correct. Still it would probably be a better solution to determine the type of encounter over a Δ time before CPA, for example 10 to 6 minutes before CPA, based on the plot of the two vessels in Figure 4.17. Because of computer limitations it has not been performed in current model. It should also be mentioned that the model utilizes GOG in determining the type of encounters, while at sea the actual heading of the vessel will determine the encounter type and these two can deviate some, especially at low speed. The deviation between heading and COG will be influenced by vessel type, speed, weather conditions and current in the area. The results for encounter types in the case-study area seems reasonable, based on navigational knowledge of the area and the nature of the three encounter types.

The current model is designed for vessels from 15m length and above. In the results several vessels of less length than 15m are found, in other words also carrying AIS class A. In further development of the model this must either be considered, or a filter to disregard vessels below 15m must be introduced. Vessels below 15m will have other consequences in a collision and the vessel behavior changes with size. Today a lot of the coastal fishing fleet and work vessels in the aquaculture industry are below 15 m, because of certificate requirements, but still they are carrying AIS class A.

The model does have some challenges regarding how to cope with missing data from the dataset. The different assumptions made in the model for length, SOG and COG will be discussed separately.

Missing length is in the model handled by use of the default length of 40.001m. Length influence the domain size to the sides and aft of the vessel. From the manual control of vessel having used the default value in detected encounters, most of the vessels have a length below 40.001m. This is reasonable since vessels with a length below the default value, will gain a domain larger by using the default value compared to if the actual length had been used. From this it can be considered to reduce the default value for missing length, but without checking manually the hole dataset it is impossible to say anything about vessels missing length which are longer than the default value. For vessels having an actual length greater than the default value, the domain will be smaller by use of the default value compared to if the actual length where used. There is therefore a risk of missing actual encounters by using the default value chosen today. If reality is considered, most larger vessels > 50m in transit in coastal waters, would consider a CPA below 0.1nm as unnecessary close and would not be complying with COLREG rule 8 "with due regard to the observant of good seamanship". With consideration that 0.1nm is an appropriate CPA, it is suggested to increase the default length to 115.751m which give a radius of $\approx$ 185.2m or 0.1nm.

Missing SOG value has a great impact on the proposed domain, since the front radius is determined by the SOG. As mention for ΔSOG, the SOG value is the value that are most often missing in the dataset for all vessels. In the current model the default value is set to 5.001knots when SOG is missing, which gives a front radius of 0.25nm. It is worth mentioning that no vessels in detected encounters has a SOG of 5.001knots, which can be an identification that the default value for the current model is too low. Dependent on what uncertainties is accepted in the results two different improvements are considered. The first suggestion is to select that the model falls back to the circular domain based on length if SOG is missing, then no faults are introduced into the model, but the domain in the front radius is very conservative. The second suggestion, which from a navigational perspective seems most reasonable, is to use a default value of 10.001knots giving the front radius a length of 0.5nm. With respect to SA a distance in front any vessel of 0.5nm seems very reasonable to assess, but the risk of introducing faults encounters in the results is present. At 10knots speed the front radius giving a level 1 encounter is 308m or 0.17nm which in a good seamanship reflection is quite little and would be of interest in coastal waters.

Missing COG values are in the current model being handled with the domain going back to the circular domain, which is considered quite reasonable. The size of the circular domain should be considered with the same argumentation used for missing length. A minimum extension to the front of 0.1nm are reasonable that all vessel in commercial transit would consider to short, and therefore this would be situations of interest. It is therefore suggested that if COG is missing and the length of the vessel is below 115.751m, the length is set to default to 115.751m to make sure that all encounters within 0.1nm are considered at level 1.

5.2 AIS data

From the results and the findings from the control of the statistics, there are challenges with the data input. First, vessels having faulty classification with respect to type of vessel which will have an impact on the frequency if vessel groups are assessed. If only the overall frequency is of interest, this will have no impact on the results. A larger problem with vessels being classified wrongly is if filters are to be implemented for given groups of vessels, such as SAR vessels or fishing vessels to avoid faulty encounters. The challenge with vessel type, is that this input is selected by a predefined list on the AIS transmitter by the navigator onboard the vessel, meaning that it can be caused by human errors if the transmitted type is wrong. That this is chosen onboard the vessels can be shown by assessing coast guard vessels, which was classified as "Fishing vessel, law enforcement, other, military ops, SAR and high-speed craft" where the only wrong classification is fishing vessel. The only correction that is considered in the 160 fault classifications are fishing vessels for coast guard vessels. If filters where to be applied for coast guard vessels, a thorough review of data is needed before the search for domain violations are performed. The size of the conformed CSV file for 2015 is 5.95 GB and consist of a total number of 68.5million rows, in other words the dataset is too large to manually verify. With the data files available for this thesis, if filters are to be introduced it must be for length of vessel not type of vessels.

With respect to the data points needed for making good judgements of encounters, the datasets in this thesis is conformed to 10- and 15-seconds intervals, performed by Nordkvist. If the head-on situations between the two ferries described in the results are assessed, the relative speed between the vessels are approximately 35knots. 35knots will in 15 seconds give a relative movement between the two vessels of 270m, which is larger than the CPA of the encounter. With an interval of for instance 2minutes the encounter could have been missed in total, since the maximum extension of the front radius in the model is set to 1852m and the relative movement of the two vessels would have been 2160m. For the model to work as intended the maximum interval of inputs values should be 15 seconds, ideally it should be shorter but then the size of the data file will increase and therefore the computation time. The time spent to run the algorithm for domain volitions was with the current datasets and the computer described earlier, between 24-34.5 hours. The frequency of input values is also important in current model for estimating ROT. As discussed in the results it is reason to believe that a ROT value will have a peak in an evasive maneuverer and the longer intervals it is between the data points the more flattened the ROT value will become. If ROT is taken directly from the AIS messages and not estimated, this would not have such a large impact. It is important to point out that not all vessels are sending ROT values, typical smaller vessels without other input to the AIS-transmitter than GPS would not send this information.

From the results it can also be concluded that there is a need for current historical AIS data, since the dynamics of sailing activity changes. In the results it can be seen for the ferries sailing between Bodø and Moskenes. If the 2013 data is assessed there would have been reasons to consider that there was a potential risk between the ferries with the 5 encounters with low CPA values. On the other hand, if 2015 is assessed, no risk is found between the ferries. If this change of encounters is caused by change of procedures of the ferries or other circumstances is not possible to determine based on the data but shows the importance of having as current as possible historical data when assessments is performed. Change of pattern in any areas can be caused by several reasons; change of fairways, VTS introducing new procedures in their areas and so on. Pattern over time can of course be of interest, but it is important to have a conscious relationship to the age of data being utilized, when performing assessments of areas.

5.3 Risk assessment

The discussion about risk assessment is separated into two parts, where the first part discuss the initial frequency found in the current model and the second part discuss the model in a wider risk assessment view and therefore some limitations of the proposed model.

5.3.1 Frequency in current model

In the results, frequency of near encounters is presented, and some concerns are raised. The statistics used in the frequency calculations, where developed by Nordkvist and are covering 15% more area than the area of the domain violation search. That the excluded areas were this big, was discovered too late in this thesis and because of time constrain it could not be redone. The excluded areas are reason to assume have a significant impact on the sailed distance of each vessel type, based on the intensity plot of the traffic. It is also reasonable to assume that several encounters would have been identified in these areas. For the frequency of near encounters to reflect the reality of an area, it is important that the area of the statistics and domain violations search matches.

Another challenge found with respect to the statistics, are the faults described in AIS data about the vessel type for determining the frequency for classes of vessels. The results show that a fault of around 9 % is found for fishing vessels, which is considered as significant. Errors in the classification of type of vessel will have an impact of the sailed distance for the type of vessel -either too high or too low- and on the encounter side of the model encounters will be assigned to fault vessel type. If no separation for single type of vessel is of interest this is not a problem, but if single type of vessel is of interest a need for manual cleaning and verification of both statistics and results are needed at current dataset and model.

There is also found other faults in the dataset, such as vessel length having a fault with a factor of 13 in the results. Since not all factors such as length, COG and SOG, which are important parameters to detect encounters, are controlled no assessment can be made on how high the percentage fault is. Since these values are based on manual input and GPS, some degree of faults should be expected. When it comes to length, an interesting observation is worth mentioning. In the current model, length is determined by the dimension-data from each AIS-message in the conformed dataset. This leads to some vessels missing length in some encounters and therefore used the default value, and in other detected encounters has the right length. The model should be altered in such a way that the length of the vessel was determined in the statistics, and this length was utilized in the domain violation search. By this, the use of the default length would be reduced, and it would be possible to determine which vessels was missing length before the domain violation search was started, and improve missing inputs.

5.3.2 General discussion about risk assessment

A general assessment of the use of ship's domain as a risk assessment tool. First, it must be considered that the domain that is assessed safe for one OOW, will vary a lot based on the navigator, type of ship and area of operation. A navigator's background and experience, the bridge team composition and knowledge will often be important factors when the OOW is determining what he/she consider as a safe domain. From experience at sea, it is also reasonable to assume that the OOW often will determine the domain based on the vessel that he/she is in an encounter with. For an OOW on a fishing vessel with a length of 20m in an encounter with another fishing vessel the safe passing distance can be 30m, but in an encounter with a bulk vessel of 250m the same OOW would consider a passing distance of maybe 300-400m as safe. This is also a reasonable assumption of why there are so few encounters with vessels above 100m, OOW change the domain dependent on the other vessel in the encounter. The human factors, that plays an important role in navigation in general (Aarsæther, 2011), is an important factor for the determining of the ship domain. Human operators from different cultures will also have different view of what is safe. In the case-study area, there are 57 flag states so different safety cultures are reasonable to assume is in the area. Several domains that is proposed are trying to utilize a lot of different input such as weather, maneuverability of vessel, area of operations and so on for determining the domain of the vessels, but maybe the most important input value human factor and assessment is not easily to determine. In general, it should be considered that the use of ship domain as a risk assessment tool is trying to standardize something that to a large degree is very unstandardized. If COLREG is studied, there are several formulations that are quite vague, as can be seen in rule 17 where the vessel with right of way, has responsibility of making maneuverers at ample time to avoid a collision. Another statement used in COLREG quite often is the term "good seamanship" which are considered different dependent on the human operator on the vessel. In comparison, flight traffic is to a very large degree standardized with respect do distances between flights both horizontally and vertically (ICAO, 1998). To determine a domain for flights would be quite simple, but no such regulation is found for maritime traffic and therefore to determine a domain that is suitable for all vessels in all situations are considered very difficult. Therefore, the domain should not be overestimated but made quite simple such as the domain in this model is when assessing a large historical dataset. The domain should be considered as the area of interest in the risk assessment, not as the actual domain that applies to all the vessels in the dataset.

In the current model no assessment towards the consequences side of an accident is considered, and therefore all near encounters are considered of interest. In the results, most encounters are between fishing vessels and many of these encounters has very low speed on both vessels. The consequences of an encounter with two fishing vessels of 11m in 2knots speed is reasonable to assume would be very low. On the other hand, if a bulk carrier of 250m and a fishing vessel have an encounter at 2knots, it is reasonable to assume that there is a potential of the fishing vessel sinking because of the inertia force of the bulk vessel. The challenge with determining the consequences is that only the length and speed of vessels are available in the current dataset. In the full AIS messages gross tonnage and deadweight is available, but to introduce a calculation in the model will demand even more computer power. It is considered beneficial to perform a study on consequences for different combination of vessels to be able to set limits that would take out near encounters which are not of interest, because of low consequences. If all near encounters are to be considered as current model detects, there would still be a need to do a review of all encounters. There are in the detected encounters operations such as SAR, tugging and dredging which are not near encounters but operations that are of nature are performed close to each other. These must be removed to get a correct assessment of frequency on near encounters.

With respect to risk, the current model assess pairwise encounters without including the bigger picture of other limitations around the encounter, and especially other vessel. For a navigator to obtain SA the hole traffic picture around him/her must be assessed, and from the SA that they then obtain, the navigators make the best possible decision to perform safe navigation in terms of good seamanship. This means that two fishing vessels can from their SA consider it safer to pass each other within the domain barriers defined in this model, than to get close quarters with an unknown vessel sailing through the area. This would of course be difficult to determine based on only the AIS data but is an important aspect to keep in mind when using the results. There is also one other very important aspect that the model will not be able to detect, and that is communication between vessels before the encounter. After AIS has been introduced as an aid to navigation, communication between vessels on VHF has increased since it is simpler to identify the other vessel by name and calling it up on the radio to clear traffic situations. When a situation is cleared on the radio, OOW will often have a smaller safe domain compared to if no communication between the vessels are performed, because SA is increased by knowing what the other vessel's intentions is.

There is also one other very important factor that is not considered in the risk assessment for the area, and that is vessels without AIS or AIS class B. After the introduction of AIS, several situations have happened with vessels with and without AIS. For vessels operating with AIS, vessels without AIS is seen as ghost ships that do not exist which has led to several near encounters. This would of course be impossible to detect but is an important aspect to keep in mind when the frequencies are read.

5.4 Comparison with Nordkvists results

The objectives in this thesis was to improve the model proposed by Nordkvist by use of navigational knowledge and experience from sea. It was also natural to have an objective to compare the results of the two different models. The comparison has proven difficult since very few encounters were plotted in Nordkvist thesis, since he had a focus on creating the framework of the model. Some comparisons can still be made.

The results from type of encounters show a significant difference between the two different methods used. It is from a navigator's perspective considered that the method proposed in this model will give a more correct classification than the method used by Nordkvist proposed by Iperen (Iperen, 2015). Further verification of the method proposed in this thesis is needed, but the results seems promising. The distribution of type of encounters must be considered toward the sea area under consideration. From a professional opinion, it seems very unlikely that head-on encounters should be the dominating encounter type in the case-study area, which is found by the method used in Nordkvist.

When it comes to detected encounters the results show that more encounters where detected by the new proposed domain in this thesis that by the domain utilized by Nordkvist. A difference between the detected encounters are that in this thesis the encounters are ranked based on the defined risk levels, while in Nordkvist model all detected encounters are considered equal. If the term "Near encounter" is assessed, which in this thesis is defined at risk level 1, the number of detected "Near encounters" in this model is quite equal to the results after Nordkvist sorting algorithm. The sorting algorithm in Nordkvist model was introduced to identify exceptional encounters based on ROT limits. There is a great difference between the two models on how high ROT values are interpret. In Nordkvist he used is a sorting parameter for all domain violations where only encounters that breached the ROT limit was considered as exceptional. In this thesis ROT and speed change has been used to flag encounters and no sorting is performed based on these. From the results in this thesis it can be concluded with that high ROT and speed change not necessarily means that an encounter is exceptional but is an identification that the encounters should be investigated further. In a professional opinion, which is also supported by the statistics presented in Section 2.3, lack of SA and human error is the main cause of collisions at sea. Therefore, a situation of close quarters where no maneuver is performed can be considered as dangerous as a situation where one of the vessels perform an evasive maneuver to avoid close quarters. Therefore, ROT and speed change should be considered as indicators that something happening, but not an algorithm to determine that an encounter is exceptional.

There are few encounters plotted in Nordkvist thesis and therefore the performance of the models at the level of single encounters are hard to determine. The only encounters that can be compared are the once mentions in the results: one is the known collision in the area which both models detect, and the other encounters is between ferries and cruise vessels. The encounters between ferries and cruise vessels are not detected by the model in this thesis and from the argumentation used in the results it is considered that this is not a situation of interest and therefore the proposed model should not detect them, from a navigator's point of view. Based on the few encounters that can be compared, it is not concluded how the two different models perform on single encounter level, for this more encounters must be compared.

Chapter 6

Recommendation for further work

Suggestions for further work to improve the model:

1. Perform a consequence analysis to determine what kind of encounters that are of interest. Should be determine for a set of different vessel combinations, to be able to use right limits and filters to identify encounters of interest with respect to size and speed.

2. Perform a verification on the proposed way of determining type of encounters. Verify the method either by assessing known encounters of the type overtaking and head on encounters or choose an area where these types of encounters are normal, narrow fairway typical a Norwegian fjord or a traffic separation scheme area.

3. A verification of a current set of AIS data, to determine if the error rate is the same or has been changed. For the model the length, SOG and COG are important to know the error rate on and the fault of vessel type is important for frequency calculations. This would be important knowledge for how to set limits, and to know the faults that should be considered in the results of the model.

4. Improve the current model to not only assess pairwise encounters but also consider the bigger picture. To get the bigger picture some suggestions are made to what can be implemented:

   - Include other vessels in and around the encounter, that could have been important factors for the encounter. Suggestion is to verify what is in a 1nm radius around the CPA position.
   - Charts to see if there are some navigational limitations with respect to safe navigations around the encounter. Especially depths and rocks would be of interest, but also aquaculture and oil installation can play an important role in safe navigation. In current model the areas with navigational challenges are to a large degree excluded, which is a possible solution in other areas as well. Then it is important that excluded areas are considered both for statistics and domain violation search.

Chapter 7

Conclusion

The model and method proposed in this thesis shows capability of detecting near encounters by a new ship domain and indicate evasive manouvers, based on ROT and ΔSOG limit. A new method for determining the type of encounter is proposed, and the results seems reasonable for the area. From the results there are several encounters that does not seem to have much potential in form of consequences, if they should have led to a collision. To reduce the number of encounters without actual consequences, a consequence study is needed. From assessing the results there are also found that the most critical situations in form of distance and flagging are SAR operations and in other words not near encounters, as the model is proposed to detect. It has also through manually control of AIS dataset, found several faults that will have an impact on results especially if a special class of vessel is to be assessed.

The uncertainties of the quality in the AIS-data are so significant that to make frequency analysis for areas will generate high uncertainties with respect to what the errors are. There are at current state a need for manual assessments of the results, to remove encounters which are not of interest. Therefore, the model at current state, with the uncertainties in both dataset and model, should be considered as a risk indicator that can be used to identify areas of interest for further risk assessment.

Bibliography

Aarsæther, K. G. (2011). Modeling and analysis of ship traffic by observation and numerical simulation.

Chai, T., Weng, J., and De-Qi, X. (2017). Development of a quantitative risk assessment model for ship collisions in fairways. *Safety Science*, 91:71–83.

Chauvin, C., Lardjane, S., Morel, G., Clostermann, J.-P., and Langard, B. (2013). Human and organisational factors in maritime accidents: Analysis of collisions at sea using the hfacs. *Accident Analysis and Prevention*, 59:26 – 37.

Collins, R. L. (2011). Heinrich and beyond. *Process Safety Progress*, 30(1):2–5.

Dinh, G. H. and Im, N.-K. (2016). The combination of analytical and statistical method to define polygonal ship domain and reflect human experiences in estimating dangerous area. *International Journal of e-Navigation and Maritime Economy*, 4(C):97–108.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1):32–64.

Fowler, T. G. and Sørgård, E. (2000). Modeling ship transportation risk. *Risk Analysis*, 20(2):225–244.

Fujii, Y. and Shiobara, R. (1971). The analysis of traffic accidents. *Journal of Navigation*, 24(4):534–543.

Fujii, Y. and Tanaka, K. (1971). Traffic capacity. *Journal of Navigation*, 24(4):543–552.

Goerlandt, F. and Kujala, P. (2011). Traffic simulation based ship collision probability modeling. *Reliability Engineering & System Safety*, 96(1):91–107.

Goodwin, E. M. (1973). A statistical study of ship domains. *Journal of Navigation*, 26(1):130–130.

Goodwin, E. M. (1975). A statistical study of ship domains. *Journal of Navigation*, 28(3):328–344.

Grech, M. R., Horberry, T., and Smith, A. (2002). Human error in maritime operations: Analyses of accident reports using the leximancer tool. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46(19):1718–1721.

Hansen, M., Jensen, T., Lehn-Schioeler, T., Melchild, K., Rasmussen, F., and Ennemark, F. (2013). Empirical ship domain based on ais data. *Journal of Navigation*, 66(6):931–940.

Hetherington, C., Flin, R., and Mearns, K. (2006). Safety in shipping: The human element. *Journal of Safety Research*, 37(4):401 – 411.

Hänninen, M. and Kujala, P. (2012). Influences of variables on ship collision probability in a bayesian belief network model. *Reliability Engineering and System Safety*, 102:27–40.

Hörteborn, A., Ringsberg, J. W., Svanberg, M., and Holm, H. A revisit of the definition of the ship domain based on ais analysis. *Journal of Navigation*, page 1–18.

ICAO (1998). Manual on airspace planning methodology for the determination of seperation minima.

IMO (1972). Colregs - the international regulations for preventing collisions at sea.

IMO (2010). International convention on standards of training, certification and watchkeeping for seafarers, 1978). Manila Amendments, STCW/CONF.2/32.

IMO (2015a). Revices guidelines for onboard operational use of ship automatic identification system (ais), resolution a.1106(29).

IMO (2015b). Revised guidelines for formal safety assessment (fsa) for use in the imo rule-making process (msc-mepc.2/circ..12/rev.1). `http://research.dnv.com/skj/IMO/MSC-MEPC.2-Circ.12-Rev.1.pdf`.

IMO (2018). Ais transponders. `http://www.imo.org/en/OurWork/Safety/Navigation/Pages/AIS.aspx` Accessed: 2018-11-09.

Iperen, E. (2015). Classifying ship encounters to monitor traffic safety on the north sea from ais data. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, 9(1):51–58.

Jensen, T., Hansen, M., Lehn-Schiøler, T., Melchild, K., Rasmussen, F., and Ennemark, F. (2013). Free flow-efficiency of a one-way traffic lane between two pylons. *The Journal of Navigation*, 66(6):941–951.

Kleppe, B. (2015). Ais norge. `https://www.kystverket.no/Maritime-tjenester/Meldings--og-informasjonstjenester/AIS/AIS-Norge/` Accessed: 2018-11-13.

Kristiansen, S. (2005). *Maritime transportation Safety management and risk analysis*. S.l.].

Kujala, P., Hanninen, M., Arola, T., and Ylitalo, J. (2009). Analysis of the marine traffic safety in the gulf of finland. *Reliability Engineering and System Safety*, 94(8).

Liu, J., Zhou, F., Li, Z., Wang, M., and Liu, R. W. (2016). Dynamic ship domain models for capacity analysis of restricted water channels. 69(3):481–503.

LKAB (2017). Transport. https://www.lkab.com/en/about-lkab/from-mine-to-port/transport/ Accessed: 2019-05-18.

Macrae, C. (2009). Human factors at sea: common patterns of error in groundings and collisions. *Maritime Policy & Management*, 36(1):21–38.

Martins, M. R. and Maturana, M. C. (2013). Application of bayesian belief networks to the human reliability analysis of an oil tanker operation focusing on collision accidents. *Reliability Engineering and System Safety*, 110:89–109.

Merrick, J. and van Dorp, J. (2001). Modeling risk in the dynamic environment of maritime transportation. In *Proceeding of the 2001 Winter Simulation Conference (Cat. No.01CH37304)*, volume 2, pages 1090–1098 vol.2. IEEE.

Merrick, J. R. W., Rene Van Dorp, J., Harrald, T., Mazzuchi, J. E., Spahn, M., Grabowski, M., and Grabowski, J. R. W. (2000). A systems approach to managing oil transportation risk in prince william sound. *Systems Engineering*, 3(3):128–142.

Mestl, T., Tallakstad, K. T., and Castberg, R. (2016). Identifying and analyzing safety critical maneuvers from high resolution ais data. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, 10(1):69–77.

NCA (2019). Havbase. https://havbase.no/ Accessed: 2019-05-18.

Nordkvist, H. A. (2018). An advanced method for detecting exceptional vessel encounters in open waters from high resolution ais data.

Norwegian Maritime Directorate (2012). Forskrift om krav til automatisk identifikasjonssystem (ais) for utenlandske fiskefartøy som lander fangst i norsk havn eller opererer i norsk territorialfarvann. `https://lovdata.no/dokument/SF/forskrift/2012-04-30-375` Accessed: 2018-11-09.

NRK (2014). Lofotfiske på røst. `https://www.nrk.no/video/PS*51109` Accessed: 2019-05-19.

Olsen, T. G. (2017). Modelling of technical, human and organisational factors of ship collision accidents using bbn.

Pedersen, P. and Simonsen, B. (1995). Dynamics of ships running aground. *Journal of Marine Science and Technology*, 1(1):37–45.

Pico, M., Hoogrvegt, D., Bik, R., Wiel, S. v. d., and Batenburg, R. v. B. (2015). Enhancing situational awareness,a research about improvement of situational awareness on board of vessels.

Pietrzykowski, Z. (2008). Ship's fuzzy domain – a criterion for navigational safety in narrow fairways. *Journal of Navigation*, 61(3):499–514.

Pietrzykowski, Z. and Magaj, J. (2017). Ship domain as a safety criterion in a precautionary area of traffic separation scheme. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, 11(1):93–98.

Roads and Maritime Services (2019). Night safety. `https://www.rms.nsw.gov.au/maritime/safety-rules/rules-regulations/night-safety.html` Accessed: 2019-04-29.

Snider, W. H. (1989). The boundaries between dynamic and static risks. *Risk Management*, 36(11):37. Copyright - Copyright Risk Management Society Publishing, Inc. Nov 1989; Last updated - 2013-06-08; CODEN - RMGTDN.

Statistics Norway (2017). Karneval og oljestrid. `https://www.ssb.no/jord-skog-jakt-og-fiskeri/artikler-og-publikasjoner/karneval-i-lofoten--295714` Accessed: 2019-05-19.

The Norwegian Coastal Administration (2019). Ais. https://www.kystverket.no/Maritime-tjenester/Meldings–og-informasjonstjenester/AIS/.

Vinnem, J.-E. (2014). *Offshore Risk Assessment vol 1: Principles, Modelling and Applications of QRA Studies*. Springer Series in Reliability Engineering. Springer London, London, 3rd ed. 2014 edition.

Wang, N. (2010). An intelligent spatial collision risk based on the quaternion ship domain. *Journal of Navigation*, 63(4):733–749.

Wang, N. (2013). A novel analytical framework for dynamic quaternion ship domains. 66(2):265–281.

Wang, N., Meng, X., Xu, Q., and Wang, Z. (2009). A unified analytical framework for ship domains. *Journal of Navigation*, 62(4):643–655.

Weng, J., Meng, Q., and Qu, X. (2012). Vessel collision frequency estimation in the singapore strait. *Journal of Navigation*, 65(2):207–221.

Zhang, W., Goerlandt, F., Kujala, P., and Wang, Y. (2016). An advanced method for detecting possible near miss ship collisions from ais data. *Ocean Engineering*, 124:141 – 156.

Zhou, T., Zhang, Z., Hu, Z., and Pan, J. (2018). Research on ship domain in narrow fairways based on ais data. volume 392. Institute of Physics Publishing.

Appendices

A Domain violation search algorithm

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-

"""
By: Rodmar Arntsen
"""
##### READ THIS###############
# change start year in the bottom to the year the data-set starts
# change end year if the data-set is over several years

from typing import Any, Union

from joblib import Parallel, delayed
import csv
import multiprocessing
import sqlite3 as lite
from datetime import datetime
from datetime import timedelta
import numpy as np
from dateutil.relativedelta import relativedelta
from geographiclib.geodesic import Geodesic


# Angle from heading of own ship to position of other ship
def alpha(lon_0, lat_0, lon_2, lat_2, cog_0):
    azi1 = Geodesic.WGS84.Inverse(lat_0, lon_0, lat_2, lon_2)['azi1']
    if type(cog_0) == str:
        if azi1 < 0:
            return np.radians(360 + azi1)
        else:
            return np.radians(azi1)
    else:
        if azi1 < 0:
            azi1 = 360 + azi1
        if cog_0 >= azi1:
            return np.radians(360 - (cog_0 - azi1))
        else:
            return np.radians(azi1 - cog_0)

# Distance to domain boundaries
def l_a(length_0, lat_0, lon_0, lat_2, lon_2, cog_0, v_0):
    # course unknown circular domain
    if type(cog_0) != float:
        R_fore = R_aft = R_starb = R_port = 1.6 * length_0
        cog_0 = 0
    elif type(v_0) != float:
        R_fore = R_aft = R_starb = R_port = 1.6 * length_0
    else:
```

```
        R_aft = R_starb = R_port = 1.6 * length_0
        R_fore = v_0 * 1852 / 3600 * 180   # From nm/hour to m/s and 180s = 3min
    if 1.6 * length_0 >= R_fore:
        R_fore = 1.6 * length_0
    elif 1.6 * length_0 < R_fore <= 1852:
        R_fore = v_0 * 1852 / 3600 * 180
    elif R_fore > 1852:     # Maximum extension of domain 1nm = 1852m
        R_fore = 1852
    #print(R_fore)
    if alpha(lon_0, lat_0, lon_2, lat_2, cog_0) <= np.pi / 2:
        l = (((1 + np.tan(alpha(lon_0, lat_0, lon_2, lat_2, cog_0)) ** 2) / ((1 / R_fore ** 2)
                + (np.tan(alpha(lon_0, lat_0, lon_2, lat_2, cog_0)) ** 2) / R_starb ** 2)) ** 0.5)

    elif np.pi / 2 < alpha(lon_0, lat_0, lon_2, lat_2, cog_0) and \
            alpha(lon_0, lat_0, lon_2, lat_2, cog_0) <= np.pi:
        l = (((1 + np.tan(alpha(lon_0, lat_0, lon_2, lat_2, cog_0)) ** 2) / ((1 / R_aft ** 2) +
            (np.tan(alpha(lon_0, lat_0, lon_2, lat_2,cog_0)) ** 2) / R_starb ** 2)) ** 0.5)

    elif np.pi < alpha(lon_0, lat_0, lon_2, lat_2, cog_0) and \
            alpha(lon_0, lat_0, lon_2, lat_2, cog_0) <= (3 / 2) * np.pi:
        l = (((1 + np.tan(alpha(lon_0, lat_0, lon_2, lat_2, cog_0)) ** 2) / ((1 / R_aft ** 2) +
            (np.tan(alpha(lon_0, lat_0, lon_2, lat_2,cog_0)) ** 2) / R_port ** 2)) ** 0.5)

    elif (3 / 2) * np.pi < alpha(lon_0, lat_0, lon_2, lat_2, cog_0):
        l = (((1 + np.tan(alpha(lon_0, lat_0, lon_2, lat_2, cog_0)) ** 2) / ((1 / R_fore ** 2) +
            (np.tan(alpha(lon_0, lat_0, lon_2, lat_2,cog_0)) ** 2) / R_port ** 2)) ** 0.5)
    #print(l)
    return l

# dir_x = 1 towards bow and 2 towards aft, dir_y = 1 towards starboard and 2 towards port
def True_center(lat_0, lon_0, d_bow, d_aft, d_starb, d_port, cog_0, startyear):
    if type(cog_0) == str:
        return lat_0, lon_0
    elif type(d_bow) == str:
        return lat_0, lon_0
    elif type(d_aft) == str:
        return lat_0, lon_0
    elif type(d_starb) == str:
        d_starb = d_port = 0
    elif type(d_port) == str:
        d_port = d_starb = 0
    else:
        ##hack for 2013/2014 missing d_starb
        if startyear == (2013 or 2014):
            d_starb = d_port
        try:
            if d_bow > d_aft:
                dist_x = (d_bow + d_aft) / 2 - d_aft
```

```
        dir_x = 1
        if d_starb > d_port:
            dist_y = (d_starb + d_port) / 2 - d_port
            dir_y = 1
            s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
        elif d_starb < d_port:
            dist_y = (d_starb + d_port) / 2 - d_port
            dir_y = 2
            s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
        elif d_starb == d_port:
            dir_y = 0
            dist_y = 0
            s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
    elif d_bow < d_aft:
        dist_x = (d_bow + d_aft) / 2 - d_bow
        dir_x = 2
        if d_starb > d_port:
            dist_y = (d_starb + d_port) / 2 - d_port
            dir_y = 1
            s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
        elif d_starb < d_port:
            dist_y = (d_starb + d_port) / 2 - d_port
            dir_y = 2
            s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
        elif d_starb == d_port:
            dir_y = 0
            dist_y = 0
            s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
    elif d_bow == d_aft:
        dist_x = (d_bow + d_aft) / 2
        dir_x = 0
        if d_starb > d_port:
            dist_y = (d_starb + d_port) / 2 - d_port
            dir_y = 1
            s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
        elif d_starb < d_port:
            dist_y = (d_starb + d_port) / 2 - d_port
            dir_y = 2
            s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
        elif d_starb == d_port:
            dir_y = 0
            dist_y = 0
            s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
    if dir_x == 0 and dir_y == 0:
        t_lon_0 = lon_0
        t_lat_0 = lat_0
    elif dir_x == 1 and dir_y == 0:
        vector = Geodesic.WGS84.Direct(lat_0, lon_0, cog_0, s_12, outmask=1929)
```

```
            t_lon_0 = vector['lon2']
            t_lat_0 = vector['lat2']
        elif dir_x == 2 and dir_y == 0:
            vector = Geodesic.WGS84.Direct(lat_0, lon_0, (cog_0 + 180), s_12, outmask=1929)
            t_lon_0 = vector['lon2']
            t_lat_0 = vector['lat2']
        elif dir_x == 0 and dir_y == 1:
            vector = Geodesic.WGS84.Direct(lat_0, lon_0, (cog_0 + 90), s_12, outmask=1929)
            t_lon_0 = vector['lon2']
            t_lat_0 = vector['lat2']
        elif dir_x == 0 and dir_y == 2:
            vector = Geodesic.WGS84.Direct(lat_0, lon_0, (cog_0 - 90), s_12, outmask=1929)
            t_lon_0 = vector['lon2']
            t_lat_0 = vector['lat2']
        elif dir_x == 1 and dir_y == 1:
            vector = Geodesic.WGS84.Direct(lat_0, lon_0, (cog_0 + np.degrees(np.arctan(dist_y /
dist_x))), s_12,
                                           outmask=1929)
            t_lon_0 = vector['lon2']
            t_lat_0 = vector['lat2']
        elif dir_x == 1 and dir_y == 2:
            vector = Geodesic.WGS84.Direct(lat_0, lon_0, (cog_0 + (360 - np.degrees(np.arctan(dist_y /
dist_x)))),
                                           s_12, outmask=1929)
            t_lon_0 = vector['lon2']
            t_lat_0 = vector['lat2']
        elif dir_x == 2 and dir_y == 2:
            vector = Geodesic.WGS84.Direct(lat_0, lon_0, (cog_0 + (180 + np.degrees(np.arctan(dist_y /
dist_x)))),
                                           s_12, outmask=1929)
            t_lon_0 = vector['lon2']
            t_lat_0 = vector['lat2']
        elif dir_x == 2 and dir_y == 1:
            vector = Geodesic.WGS84.Direct(lat_0, lon_0, (cog_0 + (180 - np.degrees(np.arctan(dist_y /
dist_x)))),
                                           s_12, outmask=1929)
            t_lon_0 = vector['lon2']
            t_lat_0 = vector['lat2']
        return (t_lat_0, t_lon_0)
    # if dimensions are missing
    except:
        return lat_0, lon_0

def Risk_level (dom_len_0_2, domain_overlap_0_2, v_0, length_0):

    if domain_overlap_0_2 > 0:
        return 0
    elif domain_overlap_0_2 <= 0:
```

```
        risk_level = dom_len_0_2 + domain_overlap_0_2
    if risk_level <= 1.6 * length_0 or risk_level <= v_0*1852/3600*60:
        return 1
    elif risk_level > v_0 * 1852 / 3600 * 60 and risk_level <= v_0 * 1852 / 3600 * 120:
        return 2
    elif risk_level > v_0 * 1852 / 3600 * 120 and risk_level <= v_0*1852/3600*180:
        return 3
    else:
        return 0

def violations(i):
    print(i)
    if i.date().day == 1:
        d = i + timedelta(days=3)
    elif i.date().day == 4:
        d = i + timedelta(days=3)
    elif i.date().day == 7:
        d = i + timedelta(days=3)
    elif i.date().day == 10:
        d = i + timedelta(days=3)
    elif i.date().day == 13:
        d = i + timedelta(days=3)
    elif i.date().day == 16:
        d = i + timedelta(days=3)
    elif i.date().day == 19:
        d = i + timedelta(days=3)
    elif i.date().day == 22:
        d = i + timedelta(days=3)
    elif i.date().day == 25:
        d = i + timedelta(days=3)
    elif i.date().day == 28:
        # first of next month
        d = i + relativedelta(months=+1, day=1)
    multiprocessing.Lock().acquire()
    conn = lite.connect(f'C:\Test\Vestfjorden_{startyear}.db')   #endre plassering på disk
    c = conn.cursor()
    # select all entries in one month starting with date i
    c.execute(f"select * from reindexed_vest_{startyear} WHERE Time >= '{i}' AND Time < '{d}' order
by Time;")
    entries = c.fetchall()
    # joblib is releasing the lock, otherwise multiprocessing.Lock().release()
    for entry_0 in range(len(entries)):
        timestamp_0 = datetime.strptime(entries[entry_0][0], "%Y-%m-%d %H:%M:%S")
        mmsi_0 = entries[entry_0][1]
        lon_0 = entries[entry_0][3]
        lat_0 = entries[entry_0][4]
        cog_0 = entries[entry_0][5]
        try:
```

```
            imo_0 = entries[entry_0][6]
        except:
            imo_0 = 9
        # hack for nonetype
        try:
            try:
                v_0 = entries[entry_0][2]
                if v_0 > 35:
                    v_0 = 20
            except:
                v_0 = 5.001
            try:
                d_bow_0 = entries[entry_0][8]
                d_aft_0 = entries[entry_0][9]
                length_0 = d_bow_0 + d_aft_0
            except:
                length_0 = 40.001
            d_port_0 = entries[entry_0][10]
            if type(d_port_0) == str:
                # To be able to recognize vessels where dimensions are missing they are sett with 3 decimal
"accuracy"
                d_port_0 = 4.501
            d_starb_0 = entries[entry_0][11]
            if type(d_starb_0) == str:
                d_starb_0 = 4.501
            try:
                width_0 = d_port_0 + d_starb_0
            except:
                # dimensions is sett to
                width_0 = 9.002
        except:
            cog_0 = ""
            length_0 = 40.001
            width_0 = 8.001
            v_0 = 5.001
            d_bow_0 = 0
            d_aft_0 = 0
            d_port_0 = 0
            d_starb_0 = 0
        for entry_2 in range((entry_0 + 1), len(entries)):
            timestamp_2 = datetime.strptime(entries[entry_2][0], "%Y-%m-%d %H:%M:%S")
            # only cheak ais transmissions at the same time instance
            if timestamp_2 > timestamp_0:
                break
            # do not cheak for close encounters close to land
            if (13.4 <= lon_0) and (67.0 <= lat_0 <= 67.52):
                break
            if (11.82 <= lon_0 <= 12.25) and (67.39 <= lat_0 <= 67.55):
```

```
            break
        if (12.55 <= lon_0 <= 12.78) and (67.62 <= lat_0 <= 67.77):
            break
        if (12.78 < lon_0 <= 12.91) and (67.82 <= lat_0 <= 67.88):
            break
        if (12.91 < lon_0 <= 13.01) and (67.86 <= lat_0):
            break
        if (13.01 < lon_0 <= 13.06) and (67.876 <= lat_0):
            break
        if (13.06 < lon_0 <= 13.09) and (67.876 <= lat_0):
            break
        if (13.09 < lon_0 <= 13.13) and (67.92 <= lat_0):
            break
        if (13.13 < lon_0 <= 13.156) and (67.937 <= lat_0):
            break
        if (13.156 < lon_0 <= 13.189) and (67.953 <= lat_0):
            break
        if (13.189 < lon_0 <= 13.243) and (67.985 <= lat_0):
            break
        if (13.243 < lon_0 <= 13.293) and (67.996 <= lat_0):
            break
        t_lat_0 = True_center(lat_0, lon_0, d_bow_0, d_aft_0, d_starb_0, d_port_0, cog_0,
startyear)[0]
        t_lon_0 = True_center(lat_0, lon_0, d_bow_0, d_aft_0, d_starb_0, d_port_0, cog_0,
startyear)[1]
        mmsi_2 = entries[entry_2][1]
        lon_2 = entries[entry_2][3]
        lat_2 = entries[entry_2][4]
        cog_2 = entries[entry_2][5]
        try:
            imo_2 = entries[entry_2][6]
        except:
            imo_2 = 9
        try:
            try:
                v_2 = entries[entry_2][2]
                if v_2 > 35:
                    v_2 = 20
            except:
                v_2 = 5.001
            try:
                d_bow_2 = entries[entry_2][8]
                d_aft_2 = entries[entry_2][9]
                length_2 = d_bow_2 + d_aft_2
            except:
                length_2 = 40.001
            d_port_2 = entries[entry_2][10]
            if type(d_port_2) == str:
```

```
            d_port_2 = 4.501
          d_starb_2 = entries[entry_2][11]
          if type(d_port_2) == str:
            d_port_2 = 4.501
          try:
            width_2 = d_port_2 + d_starb_2
          except:
            width_2 = 9.001
        except:
          cog_2 = ""
          length_2 = 40.001
          width_2 = 8.001
          v_2 = 5.001
          d_bow_2 = 0
          d_aft_2 = 0
          d_port_2 = 0
          d_starb_2 = 0
        t_lat_2 = True_center(lat_2, lon_2, d_bow_2, d_aft_2, d_starb_2, d_port_2, cog_2,
startyear)[0]
        t_lon_2 = True_center(lat_2, lon_2, d_bow_2, d_aft_2, d_starb_2, d_port_2, cog_2,
startyear)[1]
        # domain length in direction from ship 0 to ship 2 [meter]
        dom_len_0_2 = l_a(length_0, t_lat_0, t_lon_0, t_lat_2, t_lon_2, cog_0, v_0)
        # domain length in direction from ship 2 to ship 0 [meter]
        dom_len_2_0 = l_a(length_2, t_lat_2, t_lon_2, t_lat_0, t_lon_0, cog_2, v_2)
        # distance center to center [meter]
        distance_center = Geodesic.WGS84.Inverse(t_lat_0, t_lon_0, t_lat_2, t_lon_2)['s12']
        # if distance between true ship centers are less than safety domain of ship_0
        if dom_len_0_2 > distance_center and mmsi_0 != mmsi_2:
          if mmsi_0 < mmsi_2:
            if dom_len_2_0 < distance_center:
              Violated_domain = mmsi_0
            else:
              Violated_domain = 2
            ID = int(str(mmsi_0) + str(mmsi_2))
            domain_overlap_0_2 = distance_center - dom_len_0_2
            domain_overlap_2_0 = distance_center - dom_len_2_0
            Vessel_name_0 = entries[entry_0][7]
            Vessel_name_2 = entries[entry_2][7]
            risk_level_0_2 = Risk_level(dom_len_0_2, domain_overlap_0_2, v_0, length_0)
            risk_level_2_0 = Risk_level(dom_len_2_0, domain_overlap_2_0, v_2, length_2)
            fields = [ID, timestamp_0, mmsi_0, mmsi_2, imo_0, imo_2, Violated_domain,
distance_center,
                 domain_overlap_0_2, domain_overlap_2_0, v_0, v_2, cog_0, cog_2, t_lat_0,
t_lon_0, t_lat_2,
                 t_lon_2, length_0, length_2, width_0, width_2, Vessel_name_0, Vessel_name_2,
                 risk_level_0_2, risk_level_2_0]
            multiprocessing.Lock().acquire()
```

```
                with open(csvfile, 'a') as f:
                    writer = csv.writer(f)
                    writer.writerow(fields)

            else:
                if dom_len_2_0 < distance_center:
                    Violated_domain = mmsi_0
                else:
                    Violated_domain = 2
                ID = int(str(mmsi_2) + str(mmsi_0))
                domain_overlap_0_2 = distance_center - dom_len_0_2
                domain_overlap_2_0 = distance_center - dom_len_2_0
                Vessel_name_0 = entries[entry_0][7]
                Vessel_name_2 = entries[entry_2][7]
                risk_level_0_2 = Risk_level(dom_len_0_2, domain_overlap_0_2, v_0, length_0)
                risk_level_2_0 = Risk_level(dom_len_2_0, domain_overlap_2_0, v_2, length_2)
                fields = [ID, timestamp_0, mmsi_2, mmsi_0, imo_2, imo_0, Violated_domain,
distance_center,
                      domain_overlap_2_0, domain_overlap_0_2, v_2, v_0, cog_2, cog_0, t_lat_2,
t_lon_2, t_lat_0,
                      t_lon_0, length_2, length_0, width_2, width_0, Vessel_name_2, Vessel_name_0,
                      risk_level_0_2, risk_level_2_0]
                multiprocessing.Lock().acquire()
                with open(csvfile, 'a') as f:
                    writer = csv.writer(f)
                    writer.writerow(fields)
            # if distance between true ship centers are less than safety domain of ship_2
        elif dom_len_2_0 > distance_center and mmsi_0 != mmsi_2:
            if mmsi_0 < mmsi_2:
                if dom_len_0_2 < distance_center:
                    Violated_domain = mmsi_2
                else:
                    Violated_domain = 2
                ID = int(str(mmsi_0) + str(mmsi_2))
                domain_overlap_0_2 = distance_center - dom_len_0_2
                domain_overlap_2_0 = distance_center - dom_len_2_0
                Vessel_name_0 = entries[entry_0][7]
                Vessel_name_2 = entries[entry_2][7]
                risk_level_0_2 = Risk_level(dom_len_0_2, domain_overlap_0_2, v_0, length_0)
                risk_level_2_0 = Risk_level(dom_len_2_0, domain_overlap_2_0, v_2, length_2)
                fields = [ID, timestamp_0, mmsi_0, mmsi_2, imo_0, imo_2, Violated_domain,
distance_center,
                      domain_overlap_0_2, domain_overlap_2_0, v_0, v_2, cog_0, cog_2, t_lat_0,
t_lon_0, t_lat_2,
                      t_lon_2, length_0, length_2, width_0, width_2, Vessel_name_0, Vessel_name_2,
                      risk_level_0_2, risk_level_2_0]
                multiprocessing.Lock().acquire()
                with open(csvfile, 'a') as f:
```

```
                writer = csv.writer(f)
                writer.writerow(fields)
        else:
            if dom_len_0_2 < distance_center:
                Violated_domain = mmsi_2
            else:
                Violated_domain = 2
            ID = int(str(mmsi_2) + str(mmsi_0))
            domain_overlap_0_2 = distance_center - dom_len_0_2
            domain_overlap_2_0 = distance_center - dom_len_2_0
            Vessel_name_0 = entries[entry_0][7]
            Vessel_name_2 = entries[entry_2][7]
            risk_level_0_2 = Risk_level(dom_len_0_2, domain_overlap_0_2, v_0, length_0)
            risk_level_2_0 = Risk_level(dom_len_2_0, domain_overlap_2_0, v_2, length_2)
            fields = [ID, timestamp_0, mmsi_2, mmsi_0, imo_2, imo_0, Violated_domain,
distance_center,
                      domain_overlap_2_0, domain_overlap_0_2, v_2, v_0, cog_2, cog_0, t_lat_2,
t_lon_2, t_lat_0,
                      t_lon_0, length_2, length_0, width_2, width_0, Vessel_name_2, Vessel_name_0,
                      risk_level_0_2, risk_level_2_0]
            multiprocessing.Lock().acquire()
            with open(csvfile, 'a') as f:
                writer = csv.writer(f)
                writer.writerow(fields)


startyear = 2016
startmonth = 1
endyear = startyear
endmonth = 12
csvfile = f"dom_vio{startyear}_{startmonth} - {endyear}_{endmonth}.csv"

d_1 = [datetime(m // 12, m % 12 + 1, 1) for m in range(startyear * 12 + startmonth - 1, endyear * 12
+ endmonth)]
d_4 = [datetime(m // 12, m % 12 + 1, 4) for m in range(startyear * 12 + startmonth - 1, endyear * 12
+ endmonth)]
d_7 = [datetime(m // 12, m % 12 + 1, 7) for m in range(startyear * 12 + startmonth - 1, endyear * 12
+ endmonth)]
d_10 = [datetime(m // 12, m % 12 + 1, 10) for m in range(startyear * 12 + startmonth - 1, endyear *
12 + endmonth)]
d_13 = [datetime(m // 12, m % 12 + 1, 13) for m in range(startyear * 12 + startmonth - 1, endyear *
12 + endmonth)]
d_16 = [datetime(m // 12, m % 12 + 1, 16) for m in range(startyear * 12 + startmonth - 1, endyear *
12 + endmonth)]
d_19 = [datetime(m // 12, m % 12 + 1, 19) for m in range(startyear * 12 + startmonth - 1, endyear *
12 + endmonth)]
d_22 = [datetime(m // 12, m % 12 + 1, 22) for m in range(startyear * 12 + startmonth - 1, endyear *
12 + endmonth)]
```

```
d_25 = [datetime(m // 12, m % 12 + 1, 25) for m in range(startyear * 12 + startmonth - 1, endyear *
12 + endmonth)]
d_28 = [datetime(m // 12, m % 12 + 1, 28) for m in range(startyear * 12 + startmonth - 1, endyear *
12 + endmonth)]

# creates list of dates
d_1.extend(d_4)
d_1.extend(d_7)
d_1.extend(d_10)
d_1.extend(d_13)
d_1.extend(d_16)
d_1.extend(d_19)
d_1.extend(d_22)
d_1.extend(d_25)
d_1.extend(d_28)
d_1.sort()

conn = lite.connect(f'C:\DB\Vestfjorden_{startyear}.db') #endre plassering på disk
c = conn.cursor()
c.execute("""CREATE TABLE IF NOT EXISTS Domain_violation_2 (ID INT,Time DATETIME, mmsi_0
        INT,mmsi_2 INT,imo_0 INT
,imo_2 INT, Violated_domain INTEGER, distance_center float, domain_overlap_0_2     float,
domain_overlap_2_0   float,
v_0 float,v_2 float,cog_0        float, cog_2 float,t_lat_0 float,t_lon_0 float,t_lat_2
        float,t_lon_2   float,
length_0 float,length_2 float width_0   float,width_2   float,Vessel_name_0
        TEXT,Vessel_name_2 TEXT)""")
index1 = (f"CREATE INDEX IF NOT EXISTS mmsi_index_reindex ON
reindexed_vest_{startyear}(MMSI);")
c.execute(index1)
index2 = (f"CREATE INDEX IF NOT EXISTS time_index_reindex ON
reindexed_vest_{startyear}(Time);")
c.execute(index2)
write_column_names = ['ID', 'timestamp_0', 'mmsi_0', 'mmsi_2', 'imo_0', 'imo_2',
            'Violated_domain', 'distance_center', 'domain_overlap_0_2', 'domain_overlap_2_0',
            'v_0', 'v_2', 'cog_0', 'cog_2', 't_lat_0', 't_lon_0', 't_lat_2', 't_lon_2', 'length_0',
            'length_2', 'width_0', 'width_2', 'Vessel_name_0', 'Vessel_name_2',
            'risk_level_0', 'risk_level_2']
with open(csvfile, 'a') as f:
  writer = csv.writer(f)
  writer.writerow(write_column_names)
par = Parallel(n_jobs=16, verbose=10)
do_something = delayed(violations)
par(do_something(i) for i in d_1)
```

B Flagging algorithm

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-

"""
@author: Rodmar Arntsen
"""

import sqlite3 as lite  # sql
import numpy as np
import csv
import pandas as pd
from datetime import datetime
from datetime import timedelta
from geographiclib.geodesic import Geodesic
from Formulas_RA import l_a, True_center, cog_delta, alpha, Risk_level


def rate_of_turn(entries, threshold):   #Finds ROT to flag encounters
   if len(entries) < 4:
      return 0
   else:
      rot = []
      header = ('time', 'mmsi', 'lat', 'lon', 'sog', 'cog')
      entries_df = pd.DataFrame(entries, columns=header)
      entries_df = entries_df.drop_duplicates(subset='time', keep='first')
      entries_df = entries_df.reset_index(drop=True)
      for i in range(1, len(entries) - 1):
         try:
            cog_next = entries_df.loc[i + 1][5]
            cog = entries_df.loc[i][5]
            cog_prev = entries_df.loc[i - 1][5]
            time_next = datetime.strptime(entries_df.loc[i + 1][0], "%Y-%m-%d %H:%M:%S")
            time = datetime.strptime(entries_df.loc[i][0], "%Y-%m-%d %H:%M:%S")
            time_prev = datetime.strptime(entries_df.loc[i - 1][0], "%Y-%m-%d %H:%M:%S")
            time_delta_2 = (time - time_prev).seconds / 60
            time_delta_3 = (time_next - time).seconds / 60
            rot_MA = ((cog_delta(cog, cog_prev) / time_delta_2) + (cog_delta(cog_next, cog) /
time_delta_3)) / 2
            rot.append(rot_MA)
         except:
            pass
   try:
      if max(rot) > threshold:
         return (1, max(rot))
      else:
         return (0, max(rot))
   except:
      return (-2, 101)
```

```
def delta_speed(entries, threshold):    #Finds speed changes to flag encounters
    if len(entries) < 4:
        return 0
    else:
        speedchange = []
        header = ('time', 'mmsi', 'lat', 'lon', 'sog', 'cog')
        entries_df = pd.DataFrame(entries, columns=header)
        entries_df = entries_df.drop_duplicates(subset='time', keep='first')
        entries_df = entries_df.reset_index(drop=True)
        for i in range(1, len(entries)-1):
            try:
                sog_next = entries_df.loc[i + 1][4]
                sog = entries_df.loc[i][4]
                sog_prev = entries_df.loc[i - 1][4]
                sog_delta = ((abs(sog - sog_prev) + abs(sog - sog_next)) / 2)
                speedchange.append(sog_delta)
            except:
                pass
    try:
        if max(speedchange) > threshold:
            return (1, max(speedchange))
        else:
            return (0, max(speedchange))
    except:
        return (0, 0)


def speed_mean(entries):
    sogs = []
    for i in entries:
        try:
            sogs.append(float(i[4]))
        except:
            pass
    return np.nanmean(sogs).astype(np.float)


def format_timedelta(td):
    if td < timedelta(0):
        return '-' + format_timedelta(-td)
    else:
        return str(td)


def give_way(lat_1_6, lon_1_6, lat_2_6, lon_2_6, cog_1_6, mmsi_1, mmsi_2):
    if 0 <= alpha(lon_1_6, lat_1_6, lon_2_6, lat_2_6, cog_1_6) <= np.pi:
        return (mmsi_1)
```

```
    else:
        return (mmsi_2)

def delta_cog(cog_1_6, cog_2_6):
    if cog_1_6 is None or cog_2_6 is None: #If one of the vessels COG has no value encounter can't be
determined
        d_cog = 90
        return d_cog
    else:
        if cog_1_6 - cog_2_6 > 0:
            d_cog = cog_1_6 - cog_2_6
            return d_cog
        elif cog_2_6 - cog_1_6 >= 0:
            d_cog = cog_2_6 - cog_1_6
            return d_cog

def interaction(lon_2_C, lat_2_C, lon_1_C, lat_1_C, cog_1_C, cog_2_C, lat_1_6, lon_1_6, lat_2_6,
lon_2_6, cog_1_6,
                cog_2_6, mmsi_1, mmsi_2):  # C indicates at CPA and 10 is 6 minutes before CPA.
    try:
        if 292.5 < delta_cog(cog_1_6, cog_2_6) < 67.5 and \
                (112.5 / 180) * np.pi <= alpha(lon_1_6, lat_1_6, lon_2_6, lat_2_6, cog_1_6) <= (
                247.5 / 180) * np.pi or \
                292.5 < delta_cog(cog_1_6, cog_2_6) < 67.5 and \
                (112.5 / 180) * np.pi <= alpha(lon_2_6, lat_2_6, lon_1_6, lat_1_6, cog_2_6) <= (
                247.5 / 180) * np.pi: # conversion from degree to rad (degree/180)*pi,
            # alpha must be checked from both vessels therefore or statement
            interaction = 'Overtaking'
            vessel_give = 0
        elif 170 < delta_cog(cog_1_6, cog_2_6) < 190 and \
                (10 / 180) * np.pi <= alpha(lon_1_6, lat_1_6, lon_2_6, lat_2_6, cog_1_6) <= (
                350 / 180) * np.pi:  # conversion from degree to rad (degree/180)*pi
            interaction = 'Head_on'
            vessel_give = 2
        else:
            if give_way(lat_1_6, lon_1_6, lat_2_6, lon_2_6, cog_1_6, mmsi_1, mmsi_2) == mmsi_1:
                vessel_give = mmsi_1
                # angle form stand on vessel to give way vessel if mmsi_1 gives way
                if np.pi / 2 < alpha(lon_2_C, lat_2_C, lon_1_C, lat_1_C, cog_2_C) < (3 / 2) * np.pi:
                    interaction = 'Crossing passing at stern'
                else:
                    interaction = 'Crossing passing at bow'
            else:
                vessel_give = mmsi_2
                if np.pi / 2 < alpha(lon_1_C, lat_1_C, lon_2_C, lat_2_C, cog_1_C) < (3 / 2) * np.pi:
                    interaction = 'Crossing passing at stern'
                else:
                    interaction = 'Crossing passing at bow'
```

```
    except:
        interaction = 'unknown'
        vessel_give = 0
    return (interaction, vessel_give)


def dom(dom_id, dom_start, dom_end, j):
    c.execute(
        f"SELECT *, min(distance_center) FROM Domain_violation_2 WHERE ID = {dom_id} and\
         Time > '{dom_start}' and Time < '{dom_end}';")
    entry = c.fetchone()
    try:
        time_CPA = datetime.strptime(entry[1], "%Y-%m-%d %H:%M:%S")
    except:
        print(dom_id, '  ', entry[1])
    start_init = str(time_CPA - timedelta(minutes=3))
    end_init = str(time_CPA + timedelta(minutes=2))
    start_inter = str(time_CPA - timedelta(minutes=6))  #Type of interaction is determined 6 minutes
before CPA
    mmsi_1 = entry[2]
    mmsi_2 = entry[3]
    dom_vio = entry[6]
    length_1 = entry [18]
    length_2 = entry [19]
    c.execute(f"SELECT * FROM Statistics_{year} WHERE MMSI = {mmsi_1};")
    ship_1_info = c.fetchone()
    c.execute(f"SELECT * FROM Statistics_{year} WHERE MMSI = {mmsi_2};")
    ship_2_info = c.fetchone()
    c.execute(f"select Time, MMSI, Latitude, Longitude,SOG,COG  from reindexed_vest_{year}
WHERE"
            f" MMSI = '{mmsi_1}' and Time >= '{start_inter}'and Time < '{end_init}' order by Time");
    entries_inter_1 = c.fetchall()                    #Data to determine type of interaction
    lat_1_6 = entries_inter_1[0][2]
    lon_1_6 = entries_inter_1[0][3]
    cog_1_6 = entries_inter_1[0][5]
    c.execute(f"select Time, MMSI, Latitude, Longitude,SOG,COG  from reindexed_vest_{year}
WHERE"
            f" MMSI = '{mmsi_2}' and Time >= '{start_inter}' and Time< '{end_init}' order by Time");
    entries_inter_2 = c.fetchall()                    #Data to determine type of interaction
    lat_2_6 = entries_inter_2[0][2]
    lon_2_6 = entries_inter_2[0][3]
    cog_2_6 = entries_inter_2[0][5]
    c.execute(f"select Time, MMSI, Latitude, Longitude,SOG,COG  from reindexed_vest_{year}
WHERE"
            f" MMSI = '{mmsi_1}' and Time > '{start_init}' and Time< '{end_init}' order by Time");
    entries_init_1 = c.fetchall()
    N_entries_1 = len(entries_init_1)
    if N_entries_1 < 10:
```

```
        r_1 = (0, 0)
        ds_1 = (0, 0)
        s_1 = 0
    else:
        if length_1 <= 50:
            r_1 = rate_of_turn(entries_init_1, 150)
            ds_1 = delta_speed(entries_init_1, 3)
            s_1 = speed_mean(entries_init_1)
        elif length_1 > 50:
            r_1 = rate_of_turn(entries_init_1, 70)
            ds_1 = delta_speed(entries_init_1, 3)
            s_1 = speed_mean(entries_init_1)
    c.execute(f"select Time, MMSI, Latitude, Longitude,SOG,COG  from reindexed_vest_{year}
WHERE\
     MMSI = '{mmsi_2}' and Time > '{start_init}' and Time< '{end_init}' order by Time");
    entries_init_2 = c.fetchall()
    N_entries_2 = len(entries_init_2)
    if N_entries_2 < 10:
        r_2 = (0, 0)
        ds_2 = (0, 0)
        s_2 = 0
    else:
        if length_1 <= 50:
            r_2 = rate_of_turn(entries_init_2, 150)
            ds_2 = delta_speed(entries_init_2, 3)
            s_2 = speed_mean(entries_init_2)
        elif length_1 > 50:
            r_2 = rate_of_turn(entries_init_2, 70)
            ds_2 = delta_speed(entries_init_2, 3)
            s_2 = speed_mean(entries_init_2)
    Flagg = 0

    if ((r_1[0] == 1 and ds_1[0] == 1 and dom_vio == mmsi_1) or (r_2[0] == 1 and ds_2[0] == 1 and
dom_vio == mmsi_2)
            or ((r_1[0] == 1 and ds_1[0] == 1 or r_2[0] == 1 and ds_2[0] == 1) and dom_vio == 2)):
        Flagg = 2
    elif((r_1[0] == 1 and dom_vio == mmsi_1) or (r_2[0] == 1 and dom_vio == mmsi_2) or (
            (r_1[0] == 1 or r_2[0] == 1) and dom_vio == 2)):
        Flagg = 1
    elif((ds_1[0] == 1 and dom_vio == mmsi_1) or (ds_2[0] == 1 and dom_vio == mmsi_2) or (
            (ds_1[0] == 1 or ds_2[0] == 1) and dom_vio == 2)):
        Flagg = 1
        if ship_1_info[4] == ship_2_info[4] == 'Fishing':
            with open(f'ff_{dom_id}_{j}.csv', 'w', newline='', encoding='utf-8') as out:
                csv_out = csv.writer(out, delimiter=',')
                csv_out.writerow(top_row)
        else:
            with open(f'{dom_id}_{j}.csv', 'w', newline='', encoding='utf-8') as out:
```

```
            csv_out = csv.writer(out, delimiter=',')
            csv_out.writerow(top_row)
    start = str(time_CPA - timedelta(minutes=30))
    end = str(time_CPA + timedelta(minutes=30))
    c.execute(
        f"select * from reindexed_vest_{year} WHERE (MMSI = '{mmsi_1}' OR MMSI = '{mmsi_2}') and\
         Time > '{start}' and Time< '{end}' order by Time,MMSI");
    #This parts does a review of the data that is found in first step (ship_domain) only difference is
that in this
    #a more turrow look after COG that in ship domain. Therefore some encounters does not show ut
as encounters in
    #this model. Cleaning of the encounters found in ship_domain
    entries = c.fetchall()
    a = []
    b = []
    if len(entries) > 5:
        for idx in range(len(entries) - 1):
            if entries[idx][0] == entries[idx + 1][0]:
                cog_1 = entries[idx][5]
                if type(cog_1) == str:
                    try:
                        int(entries[idx - 2][5])
                        cog_1 = entries[idx - 2][5]
                    except:
                        cog_1 = 0.1
                sog_1 = entries[idx][2]
                if type(sog_1) == str:
                    try:
                        int(entries[idx - 2][2])
                        sog_1 = entries[idx - 2][2]
                    except:
                        sog_1 = 0.1
                time_0 = entries[idx][0]
                cog_2 = entries[idx + 1][5]
                if type(cog_2) == str:
                    try:
                        int(entries[idx - 1][5])
                        cog_2 = entries[idx - 1][5]
                    except:
                        cog_2 = 0.1
                sog_2 = entries[idx + 1][2]
                if type(sog_2) == str:
                    try:
                        int(entries[idx - 2][2])
                        sog_2 = entries[idx - 2][2]
                    except:
                        sog_2 = 0.1
                t_lat_1 = \
```

```
                True_center(entries[idx][4], entries[idx][3], entries[idx][8], entries[idx][9],
entries[idx][11],
                        entries[idx][10], cog_1, year)[0]
            t_lon_1 = \
                True_center(entries[idx][4], entries[idx][3], entries[idx][8], entries[idx][9],
entries[idx][11],
                        entries[idx][10], cog_1, year)[1]
            t_lat_2 = \
                True_center(entries[idx + 1][4], entries[idx + 1][3], entries[idx + 1][8], entries[idx + 1][9],
                        entries[idx + 1][11], entries[idx + 1][10], cog_2, year)[0]
            t_lon_2 = \
                True_center(entries[idx + 1][4], entries[idx + 1][3], entries[idx + 1][8], entries[idx + 1][9],
                        entries[idx + 1][11], entries[idx + 1][10], cog_2, year)[1]
            try:
                length_1 = (entries[idx][8] + entries[idx][9])
            except:
                length_1 = 40.001
            try:
                length_2 = (entries[idx + 1][8] + entries[idx + 1][9])
            except:
                length_2 = 40.001
            # domain length in direction from ship 1 to ship 2 [meter]
            dom_len_1_2 = l_a(length_1, t_lat_1, t_lon_1, t_lat_2, t_lon_2, cog_1, sog_1)
            # domain length in direction from ship 2 to ship 0 [meter]
            dom_len_2_1 = l_a(length_2, t_lat_2, t_lon_2, t_lat_1, t_lon_1, cog_2, sog_2)
            # distance center to center [meter]
            distance = Geodesic.WGS84.Inverse(t_lat_1, t_lon_1, t_lat_2, t_lon_2)['s12']
            # negative values indicate domain violation
            overlap_ship2_into_shipdomain_1 = distance - dom_len_1_2
            overlap_ship1_into_shipdomain_2 = distance - dom_len_2_1
            if str(time_CPA) == time_0:
                CPA = 1
                lat_1_C = t_lat_1
                lon_1_C = t_lon_1
                lat_2_C = t_lat_2
                lon_2_C = t_lon_2
                cog_1_C = cog_1
                cog_2_C = cog_2
                sog_1_C = sog_1
                sog_2_C = sog_2
                overlap_c_2_1 = overlap_ship2_into_shipdomain_1
                overlap_c_1_2 = overlap_ship1_into_shipdomain_2
                distance_c = distance
                dom_len_1_c = l_a(length_1, t_lat_1, t_lon_1, t_lat_2, t_lon_2, cog_1, sog_1)
                dom_len_2_c = l_a(length_2, t_lat_2, t_lon_2, t_lat_1, t_lon_1, cog_2, sog_2)
                Risk_level_1 = Risk_level(dom_len_1_c, overlap_c_2_1, sog_1_C, length_1)
                Risk_level_2 = Risk_level(dom_len_2_c, overlap_c_1_2, sog_2_C, length_2)
                try:
```

```
                max_rot_1 = r_1[1]
            except:
                max_rot_1 = -1
            try:
                max_rot_2 = r_2[1]
            except:
                max_rot_2 = -1
            try:
                delta_speed_1 = ds_1[1]
            except:
                delta_speed_1 = -1
            try:
                delta_speed_2 = ds_2[1]
            except:
                delta_speed_2 = -1

            encounter = interaction(lon_2_C, lat_2_C, lon_1_C, lat_1_C, cog_1_C, cog_2_C, lat_1_6,
lon_1_6,\
                                    lat_2_6, lon_2_6, cog_1_6, cog_2_6, mmsi_1, mmsi_2)[0]
            give_way_vessel = interaction(lon_2_C, lat_2_C, lon_1_C, lat_1_C, cog_1_C, cog_2_C,
lat_1_6,\
                                          lon_1_6, lat_2_6, lon_2_6, cog_1_6, cog_2_6, mmsi_1, mmsi_2)[1]
            if encounter == 'Overtaking':
                if s_1 > s_2:
                    give_way_vessel = mmsi_1
                else:
                    give_way_vessel = mmsi_2
        else:
            CPA = 0
        TCPA = format_timedelta(datetime.strptime(time_0, "%Y-%m-%d %H:%M:%S") - time_CPA)
        fields_1 = (dom_id, time_0, mmsi_1, distance, dom_len_1_2,
overlap_ship2_into_shipdomain_1, CPA, TCPA,
                    t_lon_1, t_lat_1, cog_1, sog_1, ship_1_info[4])
        fields_2 = (dom_id, time_0, mmsi_2, distance, dom_len_2_1,
overlap_ship1_into_shipdomain_2, CPA, TCPA,
                    t_lon_2, t_lat_2, cog_2, sog_2, ship_2_info[4])
        a.append(fields_1)
        b.append(fields_2)
    a.extend(b)

    c.execute(f"INSERT INTO Flagging_{year}
VALUES(?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,\
    ?,?,?,?,?,?,?,?,?,?)", \
      (dom_id, time_CPA, distance_c, mmsi_1, mmsi_2, dom_len_1_c, dom_len_2_c,
overlap_c_2_1, overlap_c_1_2,
       encounter, give_way_vessel, max_rot_1, delta_speed_1, max_rot_2, delta_speed_2, lat_1_C,
lon_1_C, lat_2_C,\
```

```
            lon_2_C, sog_1_C, sog_2_C, cog_1_C, cog_2_C, ship_1_info[1], ship_2_info[1],
ship_1_info[3], \
            ship_2_info[3], ship_1_info[4], ship_2_info[4], ship_1_info[5], ship_2_info[5],
ship_1_info[6],\
            ship_2_info[6], length_1, length_2, Risk_level_1, Risk_level_2, Flagg, 0))
        conn.commit()
        if Flagg == 1:
            if ship_1_info[4] == ship_2_info[4] == 'Fishing':
                with open(f'ff_{dom_id}_{j}.csv', 'a', newline='', encoding='utf-8') as out:
                    csv_out = csv.writer(out, delimiter=',')
                    csv_out.writerows(a)
            else:
                with open(f'{dom_id}_{j}.csv', 'a', newline='', encoding='utf-8') as f:
                    writer = csv.writer(f)
                    writer.writerows(a)


year = input("input year form 2013-2016: ")
conn = lite.connect(f'C:\DB\Vestfjorden_{year}.db')
c = conn.cursor()
c.execute(f"CREATE TABLE IF NOT EXISTS Flagging_{year} (dom_id int,time_CPA datetime,distance_c
float,mmsi_1 int,"
          f"mmsi_2 int,dom_len_1_c float,dom_len_2_c float,overlap_c_2_1 float,overlap_c_1_2 float,\
          encounter text,give_way_vessel int,max_rot_1 float,delta_speed_1 float,max_rot_2 float,
delta_speed_2 float,\
          lat_1_C float,lon_1_C float,lat_2_C float, lon_2_C float,sog_1_C float,sog_2_C float,cog_1_C
float, \
          cog_2_C float, IMO_1 int,IMO_2 int,Flag_1 text, Flag_2 text,Vessel_type_1 text,Vessel_type_2
text,\
          Sailed_distance_1 float, Sailed_distance_2 float, N_trips_1 int,N_trips_2 int,Length_1
int,Length_2 int,\
          Risk_level_1 int, Risk_level_2 int,Flagg_algorithmn int,Visual_inspection int)")
top_row = ('Identifyier', 'Time', 'mmsi', 'Distance', 'Domain_length', 'domain_violation', 'CPA', 'TCPA',
'lon', 'lat',
           'cog', 'sog', 'vessel_type', 'length', 'length_domain',)

c.execute("SELECT distinct ID FROM Domain_violation_2 ORDER BY ID")
violation_ID = [int(i[0]) for i in c.fetchall()]
for x in violation_ID:
    c.execute(f"SELECT * FROM Domain_violation_2 WHERE ID = {x};")
    entries = c.fetchall()
    dates = [datetime.strptime(entries[0][1], "%Y-%m-%d %H:%M:%S")]
    for i in range(1, len(entries)):
        g = datetime.strptime(entries[i][1], "%Y-%m-%d %H:%M:%S")
        f = datetime.strptime(entries[i - 1][1], "%Y-%m-%d %H:%M:%S")
        hours = (g - f).seconds / 3600
        days = (g - f).days
        if days > 1 or hours > 4:
```

```
        dates.append(g)
j = 0
for i in dates:
    j += 1
    dom_start = str(i - timedelta(minutes=60))
    dom_end = str(i + timedelta(minutes=60))
    dom(x, dom_start, dom_end, j)
    print(i, dates, dom_start, j)
```

C Formulas

```python
#!/usr/bin/env python3
#-*- coding: utf-8 -*-

import numpy as np
from geographiclib.geodesic import Geodesic

def cog_delta(cog_1,cog_2):
   if abs(cog_1-cog_2)<=180:
      return abs(cog_1-cog_2)
   else:
      return (min(cog_1,cog_2)+(360-max(cog_1,cog_2)))

def sog_delta(sog_1, sog_2):
   abs(sog_1-sog_2)

def alpha(lon_1, lat_1, lon_2, lat_2, cog_1):
   azi1 = Geodesic.WGS84.Inverse(lat_1, lon_1, lat_2, lon_2)['azi1']
   if type(cog_1) == str:
      if azi1 < 0:
         return np.radians(360 + azi1)
      else:
         return np.radians(azi1)
   else:
      if azi1 < 0:
         azi1 = 360 + azi1
      if cog_1 >= azi1:
         return np.radians(360 - (cog_1 - azi1))
      else:
         return np.radians(azi1 - cog_1)

def l_a(length_0, lat_1, lon_1, lat_2, lon_2, cog_1, v_0):
   # course unknown circular domain
   if type(cog_1) != float:
      R_fore = R_aft = R_starb = R_port = 1.6 * length_0
      cog_1 = 0
   elif type(v_0) != float:
      R_fore = R_aft = R_starb = R_port = 1.6 * length_0
   else:
      R_aft = R_starb = R_port = 1.6 * length_0
      R_fore = v_0 * 1852 / 3600 * 180   # From nm/hour to m/s and 180s = 3min
   if 1.6 * length_0 >= R_fore:
      R_fore = 1.6 * length_0
   elif 1.6 * length_0 < R_fore <= 1852:
      R_fore = v_0 * 1852 / 3600 * 180
   elif R_fore > 1852:     # Maximum extension of domain 1nm = 1852m
      R_fore = 1852
   #print(R_fore)
   if alpha(lon_1, lat_1, lon_2, lat_2, cog_1) <= np.pi / 2:
```

```
        l = (((1 + np.tan(alpha(lon_1, lat_1, lon_2, lat_2, cog_1)) ** 2) / ((1 / R_fore ** 2)
             + (np.tan(alpha(lon_1, lat_1, lon_2, lat_2, cog_1)) ** 2) / R_starb ** 2)) ** 0.5)

    elif np.pi / 2 < alpha(lon_1, lat_1, lon_2, lat_2, cog_1) and \
            alpha(lon_1, lat_1, lon_2, lat_2, cog_1) <= np.pi:
        l = (((1 + np.tan(alpha(lon_1, lat_1, lon_2, lat_2, cog_1)) ** 2) / ((1 / R_aft ** 2) +
            (np.tan(alpha(lon_1, lat_1, lon_2, lat_2,cog_1)) ** 2) / R_starb ** 2)) ** 0.5)

    elif np.pi < alpha(lon_1, lat_1, lon_2, lat_2, cog_1) and \
            alpha(lon_1, lat_1, lon_2, lat_2, cog_1) <= (3 / 2) * np.pi:
        l = (((1 + np.tan(alpha(lon_1, lat_1, lon_2, lat_2, cog_1)) ** 2) / ((1 / R_aft ** 2) +
            (np.tan(alpha(lon_1, lat_1, lon_2, lat_2,cog_1)) ** 2) / R_port ** 2)) ** 0.5)

    elif (3 / 2) * np.pi < alpha(lon_1, lat_1, lon_2, lat_2, cog_1):
        l = (((1 + np.tan(alpha(lon_1, lat_1, lon_2, lat_2, cog_1)) ** 2) / ((1 / R_fore ** 2) +
            (np.tan(alpha(lon_1, lat_1, lon_2, lat_2,cog_1)) ** 2) / R_port ** 2)) ** 0.5)
    return l

def True_center(lat_1, lon_1, d_bow, d_aft, d_starb, d_port, cog_1, year):
    if type(cog_1) == str:
        return lat_1, lon_1
    elif type(d_bow) == str:
        return lat_1, lon_1
    elif type(d_aft) == str:
        return lat_1, lon_1
    elif type(d_starb) == str:
        d_starb = d_port = 0
    elif type(d_port) == str:
        d_port = d_starb = 0
    else:
        ##hack for 2013/2014 missing d_starb
        if year == (2013 or 2014):
            d_starb = d_port
        try:
            if d_bow > d_aft:
                dist_x = (d_bow + d_aft) / 2 - d_aft
                dir_x = 1
                if d_starb > d_port:
                    dist_y = (d_starb + d_port) / 2 - d_port
                    dir_y = 1
                    s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
                elif d_starb < d_port:
                    dist_y = (d_starb + d_port) / 2 - d_port
                    dir_y = 2
                    s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
                elif d_starb == d_port:
                    dir_y = 0
                    dist_y = 0
```

```
        s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
elif d_bow < d_aft:
    dist_x = (d_bow + d_aft) / 2 - d_bow
    dir_x = 2
    if d_starb > d_port:
        dist_y = (d_starb + d_port) / 2 - d_port
        dir_y = 1
        s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
    elif d_starb < d_port:
        dist_y = (d_starb + d_port) / 2 - d_port
        dir_y = 2
        s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
    elif d_starb == d_port:
        dir_y = 0
        dist_y = 0
        s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
elif d_bow == d_aft:
    dist_x = (d_bow + d_aft) / 2
    dir_x = 0
    if d_starb > d_port:
        dist_y = (d_starb + d_port) / 2 - d_port
        dir_y = 1
        s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
    elif d_starb < d_port:
        dist_y = (d_starb + d_port) / 2 - d_port
        dir_y = 2
        s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
    elif d_starb == d_port:
        dir_y = 0
        dist_y = 0
        s_12 = np.sqrt(dist_x ** 2 + dist_y ** 2)
if dir_x == 0 and dir_y == 0:
    t_lon_1 = lon_1
    t_lat_1 = lat_1
elif dir_x == 1 and dir_y == 0:
    vector = Geodesic.WGS84.Direct(lat_1, lon_1, cog_1, s_12, outmask=1929)
    t_lon_1 = vector['lon2']
    t_lat_1 = vector['lat2']
elif dir_x == 2 and dir_y == 0:
    vector = Geodesic.WGS84.Direct(lat_1, lon_1, (cog_1 + 180), s_12, outmask=1929)
    t_lon_1 = vector['lon2']
    t_lat_1 = vector['lat2']
elif dir_x == 0 and dir_y == 1:
    vector = Geodesic.WGS84.Direct(lat_1, lon_1, (cog_1 + 90), s_12, outmask=1929)
    t_lon_1 = vector['lon2']
    t_lat_1 = vector['lat2']
elif dir_x == 0 and dir_y == 2:
    vector = Geodesic.WGS84.Direct(lat_1, lon_1, (cog_1 - 90), s_12, outmask=1929)
```

```
            t_lon_1 = vector['lon2']
            t_lat_1 = vector['lat2']
        elif dir_x == 1 and dir_y == 1:
            vector = Geodesic.WGS84.Direct(lat_1, lon_1, (cog_1 + np.degrees(np.arctan(dist_y /
dist_x))), s_12,
                                    outmask=1929)
            t_lon_1 = vector['lon2']
            t_lat_1 = vector['lat2']
        elif dir_x == 1 and dir_y == 2:
            vector = Geodesic.WGS84.Direct(lat_1, lon_1, (cog_1 + (360 - np.degrees(np.arctan(dist_y /
dist_x)))),
                                    s_12, outmask=1929)
            t_lon_1 = vector['lon2']
            t_lat_1 = vector['lat2']
        elif dir_x == 2 and dir_y == 2:
            vector = Geodesic.WGS84.Direct(lat_1, lon_1, (cog_1 + (180 + np.degrees(np.arctan(dist_y /
dist_x)))),
                                    s_12, outmask=1929)
            t_lon_1 = vector['lon2']
            t_lat_1 = vector['lat2']
        elif dir_x == 2 and dir_y == 1:
            vector = Geodesic.WGS84.Direct(lat_1, lon_1, (cog_1 + (180 - np.degrees(np.arctan(dist_y /
dist_x)))),
                                    s_12, outmask=1929)
            t_lon_1 = vector['lon2']
            t_lat_1 = vector['lat2']
        return (t_lat_1, t_lon_1)
    # if dimensions are missing
    except:
        return lat_1, lon_1

def Risk_level (dom_len_1_c, overlap_c_2_1, sog_1_C, length_1):

    if overlap_c_2_1 > 0:
        return 0
    elif overlap_c_2_1 <= 0:
        risk_level = dom_len_1_c + overlap_c_2_1
    if risk_level <= 1.6 * length_1 or risk_level <= sog_1_C*1852/3600*60:
        return 1
    elif risk_level > sog_1_C * 1852 / 3600 * 60 and risk_level <= sog_1_C * 1852 / 3600 * 120:
        return 2
    elif risk_level > sog_1_C * 1852 / 3600 * 120 and risk_level <= sog_1_C*1852/3600*180:
        return 3
    else:
        return 0
```

Rodmar Arntsen

Analysis of AIS-data to identify near encounter situations between ships

NTNU
Norwegian University of Science and Technology