Ville Hakola

PREDICTING MARINE TRAFFIC IN THE ICE-COVERED BALTIC SEA

Faculty of Information Technology and Communication Sciences
M. Sc. Thesis
March 2020

ABSTRACT

Ville Hakola: Predicting marine traffic in the ice-covered Baltic Sea
M. Sc. Thesis, 62 pages, 0 appendices pages
Tampere University
Master's Degree Programme in Software Development
February 2020

Icebreaking activity and seasonal ice propose challenges for marine traffic prediction in the Baltic Sea. Traffic prediction is a vital part in the planning of icebreaking activities, but it remains largely as a manual task. The aim of this thesis is to examine factors influencing marine traffic modelling in ice-covered waters and propose a novel A*-based method for modelling traffic in ice. The current state of the marine traffic modelling and factors affecting vessel movement are concluded by examining the literature and historical vessel tracks.

The field of traffic modelling research is growing rapidly. Currently the biggest challenges are evaluation of results and the lack of publicly available datasets. Moreover, the current approaches to model vessel movement in ice are promising but fail to capture how icebreaking activity influences vessel routes.

The proposed model consists of sea, maneuverability, route and speed modelling. The model uses historical AIS data, topography of the sea, vessel type and dirways as main data inputs. The model is trained with summer tracks and dirways are used for modelling the ice channels kept open by icebreakers. The accuracy of the model is evaluated by examining route, speed, traffic and ETA (estimated time of arrival) prediction results separately. Moreover, the area between the actual and predicted route is introduced as an accuracy measure for route prediction.

The model shows that winter route prediction can be improved by incorporating dirways to the modelling. However, the use of dirways did not affect the speed, traffic or ETA prediction accuracy. Finally, the datasets and source code used in this thesis are published online.

Keywords: Marine traffic, AIS, Route Prediction, Trajectory prediction, ETA Prediction, Traffic Prediction, Winter Navigation

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

PREFACE

The research presented in this thesis has been conducted to aid the development of a distributed icebreaker information system called IBNet. The Finnish Transport Infrastructure Agency is the owner of the system and the technical development is done by Solita Oy.

I would like to thank Solita Oy and The Finnish Transport Infrastructure Agency for providing such an interesting and challenging topic to research. Especially, I would like to thank three following people from Solita: Ari Koivula for overseeing my work, Mika Alatalo for helping to choose the research topic and Kati Oravainen for enabling the work for Solita.

I would like to thank my supervisor Professor Martti Juhola from Tampere University for the constructive feedback during my thesis work. Likewise, I would like to thank University Lecturer Timo Poranen from Tampere University for helping with the final polish of the thesis. Finally, I wish to express sincere gratitude for my family and friends for their continuing support throughout my studies.

In Tampere, Finland, on 5 February 2020

Ville Hakola

Contents

LIST OF ABBREVIATIONS AND TERMS

AIS	Automatic Identification System, used for collecting vessel tracking data.
CONT	Container ship.
Dirway	A set of directed waypoints through which vessels should travel in order to get icebreaking assistance.
ESRI	Environmental Systems Research Institute, provider of geographic information systems.
Fairway	Navigable channel near port, river or shallow waters
GC	General cargo vessel.
GDAL	Geospatial Data Abstraction Library.
IBNet	A distributed winter navigation system, used for coordinating icebreaking activities in the Baltic Sea.
IMO	International Maritime Organization.
MMSI	Maritime Mobile Service Identity number unique to a vessel. Used to identify vessels in AIS-messages.
NM	Nautical mile.
PAS	Passenger ship.
RORO	Roll-on/roll-off vessel used for vehicle transport.
Safety zone	Area which is not safe for vessels to travel.
SPA	Shortest Path Algorithm.
T	Tanker ship.
Voyage	A journey of a vessel from one port to another.
VTS	Vessel Traffic Service.

1 INTRODUCTION

Maritime transport is the most cost-effective method of transport and accounts around 80% of the volume of all global trade (UNCTAD 2018). Countries around the Baltic Sea are dependent on the sea for trade. In Finland, maritime transport accounts for over 85% of the gross national income (Toivola 2016).

The Baltic Sea is one of the most travelled waterways where ice conditions in winter affect maritime transport. Seasonal ice means that icebreaking is needed to combat the seasonal variation in trade (Lépy 2013). In Northern Baltic Sea, all ports freeze and become unreachable during typical winter without icebreaking activities. The background of marine navigation and specialities of winter navigation are explored in Chapter 2.

Operating the icebreaking fleet efficiently can lead to safer and faster transportation to ports and reduce fuel costs for both icebreakers and merchant vessels. One of the tools in the operational planning of icebreakers is prediction of traffic situation. However, the number of vessels travelling to different ports in an area size of the Baltic Sea means there is a need for an automatized way to predict vessel positions. The previous research about vessel movement prediction is described in Chapter 3.

This thesis leverages data collected during years 2017 – 2019. The data consists of historical vessel positions, vessel metadata, port locations, dirways and map data. The collection and processing of data is described in detail in Chapter 4**.** The data is collected from IBNet which is developed to enhance icebreaking cooperation between Finland and Sweden. IBNet is in use on all Finnish and Swedish icebreakers (Berglund *et al.* 2014).

The prior literature does cover vessel movement prediction in ice only partially. Thus, a novel approach for traffic prediction in ice is proposed. The proposed method combines modelling of sea, traffic, speed and open ice channels using grid-based methods and A* algorithm. The proposed method is described in Chapter 5. The results and conclusions are discussed in Chapters 6 and 7.

The goal of this thesis is to examine the current state and requirements of winter marine traffic modelling and to create a novel method for year-round modelling of marine traffic in the Baltic Sea. The method is intended to be used as a starting point for further research and to evaluate the viability of marine traffic modelling in ice. The method is evaluated separately in summer and winter conditions.

The thesis aims to respond to the following questions: what factors affect vessel movement in ice and what are the key components of modelling of marine traffic.

2 MARINE NAVIGATION IN THE BALTIC SEA

This chapter explores the aspects of marine navigation and seasonal ice relevant to traffic prediction in the Baltic Sea. The chapter aims to give a frame of reference for the remaining of the thesis and highlight the complex nature of modelling winter navigation.

2.1 Background of Marine Navigation

This section describes marine navigation in general and identifies the specialities of marine traffic in the Baltic Sea. The information has been gathered by exploring the current literature and the data gathered during icebreaking season 2017 – 2018.

Marine navigation has unique characteristics when compared to vehicle movement on different substances such as land or air (Guinness et al. 2014, Tu et al. 2018, Ueno et al. 2009). Three major differences can be identified. First, ship cannot suddenly stop, turn or reverse like land vehicle. Second, ship movement happens on a plane whereas an aircraft or a submarine moves in a three-dimensional space. Third, turning a vessel is costly in terms of time and fuel usage which leads to vessels minimizing course changes (Montewka et al. 2017).

Although vessels typically move quite freely on the ocean, vessel movement can be restricted by the following factors: shallow waters, safety zones, fairways, ice and capacity of ports (Guinness *et al.* 2014, Lai and Shih 1992, Löptinen and Axell 2014, Seong et al. 2011). All the factors except ice and port capacity are largely related to the depth of the sea. Port capacity restriction can mean that a vessel must wait near the port before entering. The effect of ice to vessel routes is described in Sections 2.2 and 2.3.

The Baltic Sea is shallow compared to other seas (Finnish Meteorological Institute 2020a). The average depth of the Baltic Sea is 55 meters whereas the average depth of the Mediterranean Sea is 1500 meters (Schroeder 2019). In addition, the Baltic Sea contains many archipelagos where vessel movement is highly restricted as depicted in Figure

1. In shallow waters such as near port or in an archipelago, vessels typically follow fairways to avoid running ashore.

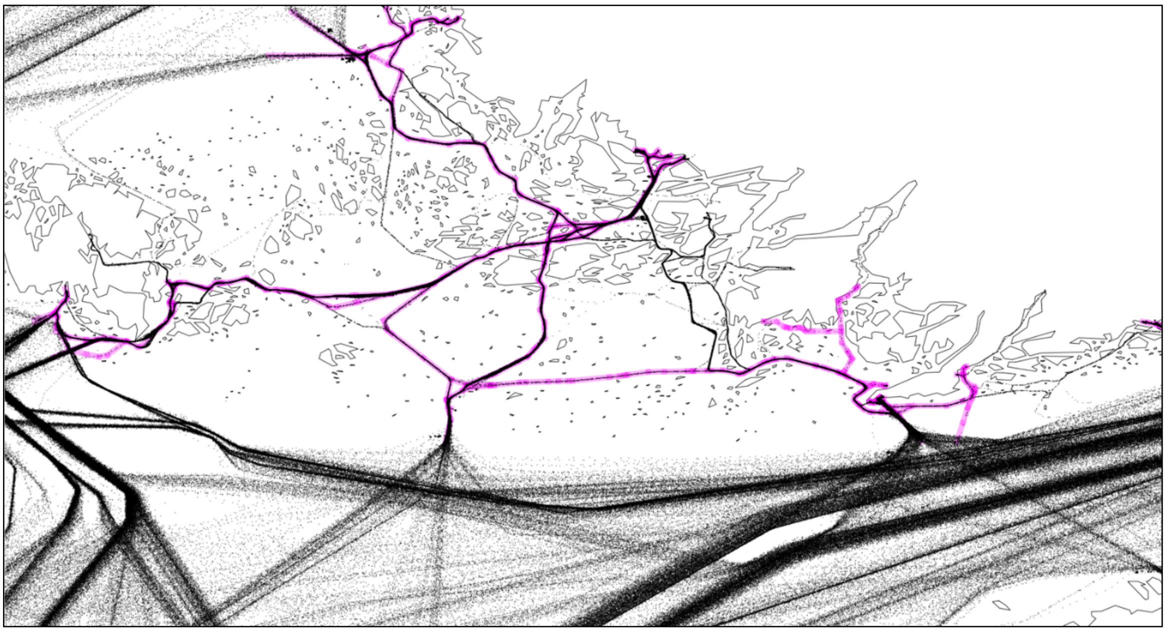


Figure 1. Historical vessel movements (black) combined with fairways (magenta) around the Turku archipelago.

2.2 No Two Winters Are the Same

In this section, the variance in ice coverage and formulation between winters is examined. This section aims to highlight how differences in ice coverage and formulation can affect vessel routes.

Typically, seasonal ice starts to extend from the Northern Bay of Bothnia towards the south of the Baltic Sea around mid-November and lasts until early May. In the last 10 years, the ice coverage in the Baltic Sea has ranged from 14% to 82% of the total area of the Baltic Sea (Finnish Meteorological Institute 2020b). The highest and the lowest ice coverages in the last 10 years are visualized in Figure 2.

As ports start to freeze, the maritime authorities begin to issue port-specific restrictions called *traffic restrictions*. Traffic restrictions describe the minimum requirements for a vessel to fulfil in order to safely and efficiently travel to a specific port. Finnish and Swedish traffic restrictions are updated through the IBNet based on the prevalent ice-conditions (Berglund *et al.* 2014). The active restrictions in the Baltic Sea are published daily on the internet (Guinness *et al.* 2014).

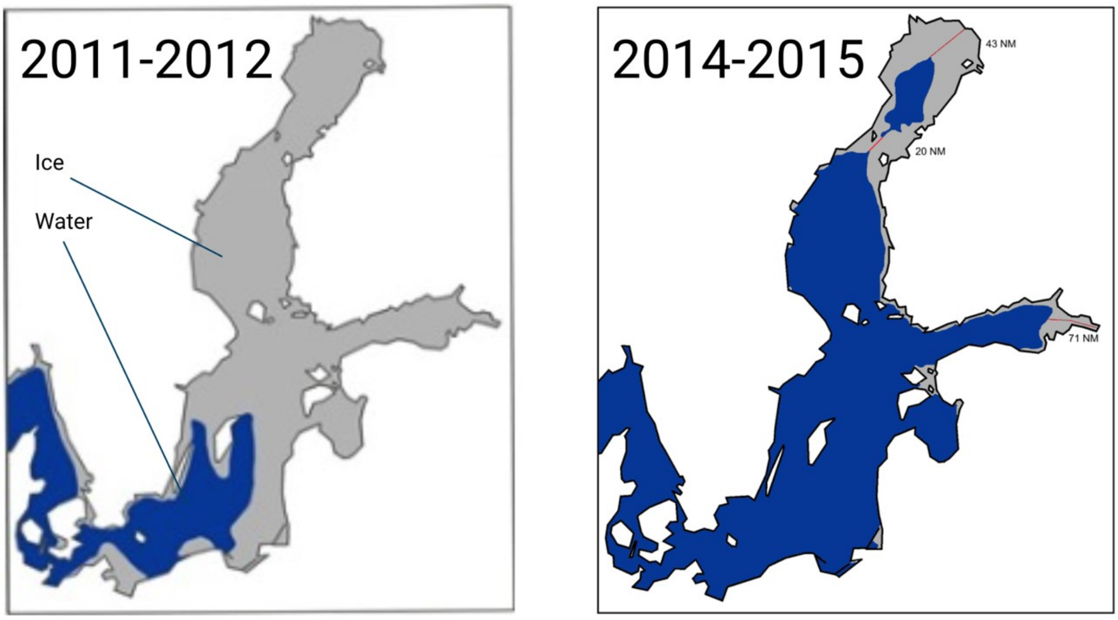


Figure 2. Highest (left) and lowest (right) ice-coverage in the Baltic Sea in the last 10 years (Finnish Meteorological Institute 2020b)

Traffic restrictions are based on the ice class rules that classify a vessel's capability to navigate through ice (Riska and Kämäräinen 2011, RMRS 2019). In the Baltic Sea, two different ice class rules are used: the Finnish-Swedish Ice Class Rules (FSICR) and the Russian Maritime Register of Shipping (RMRS) Ice Class Rules (Non-Arctic Sea Area Requirements). The FSICR rules are more prevalent as majority of the restricted ports belong to Finland and Sweden (BIM 2018). The ice classes could be used in the modelling of vessel's speed in ice.

The ever-changing ice-conditions dictate that vessels can't travel the same routes on winter as they do in the summer since icebreakers can only keep open a limited number of routes (Guinness *et al.* 2014, Löptinen and Axell 2014). Furthermore, an open ice channel can change multiple times during a winter based on the movements of ice and the changes in ice ridges, compression and concentration.

The differences in vessel routes between a summer and a winter can be seen in Figure 3. The division into the summer and the winter routes is based on the official start and end dates of the icebreaking season 2017 - 2018 (Arctia 2018). Icebreakers are filtered out of the data. The winter routes in the Northern Bay of Bothnia differ the most from the summer routes as the area was most affected by the seasonal ice.

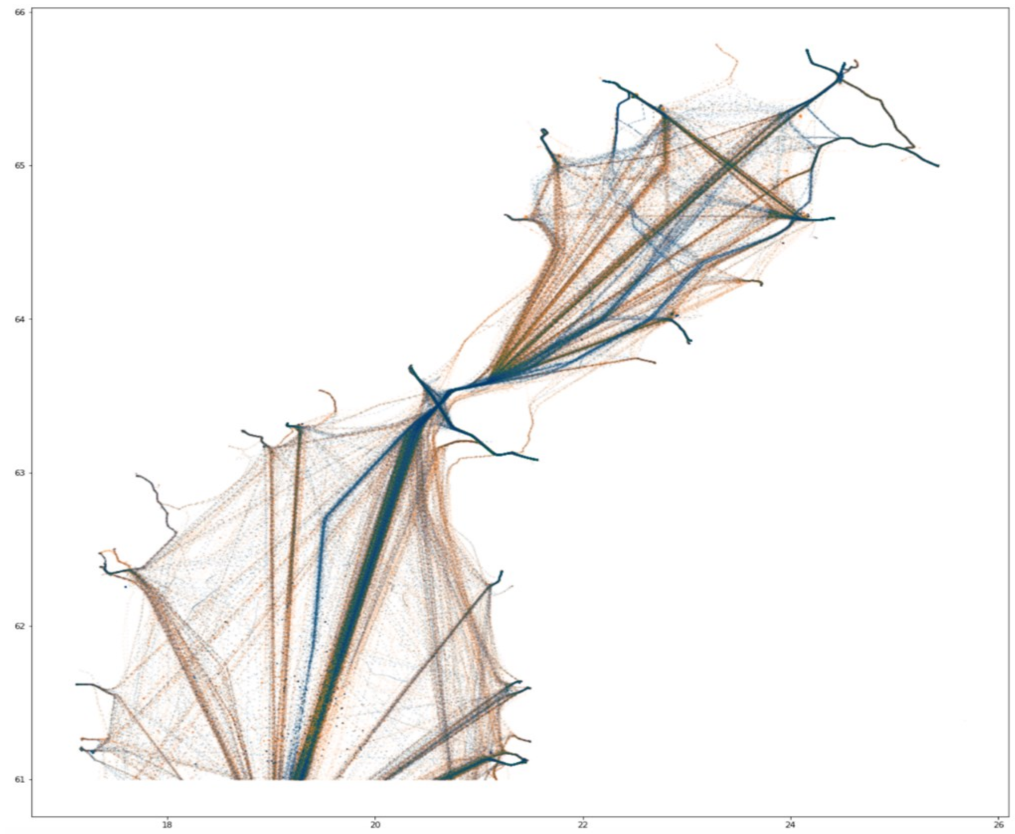


Figure 3. Visualization of the differences in summer and winter routes during years 2017 - 2018. Summer routes are coloured in orange and winter routes in blue.

In summary, seasonal ice is the primary factor that affects vessel routes in winter. The model for the winter routes cannot be extracted from a single winter alone since all winters are different and even winters with similar ice coverage can vary vastly depending on how the ice field develops during the winter.

2.3 Operational Challenges

In this section, the operational challenges that affect marine traffic are examined. The operational challenges in the Baltic Sea can be divided into two types: general and winter related operational challenges (Montewka *et al.* 2017, Sormunen *et al.* 2018). The operational challenges affect both the speed of a vessel and the route of a vessel. The operational challenges are depicted in Figure 4.

Figure 4. The operational challenges affecting speed and route of a vessel (Guinness *et al.* 2014, Montewka *et al.* 2017, Sormunen *et al.* 2018).

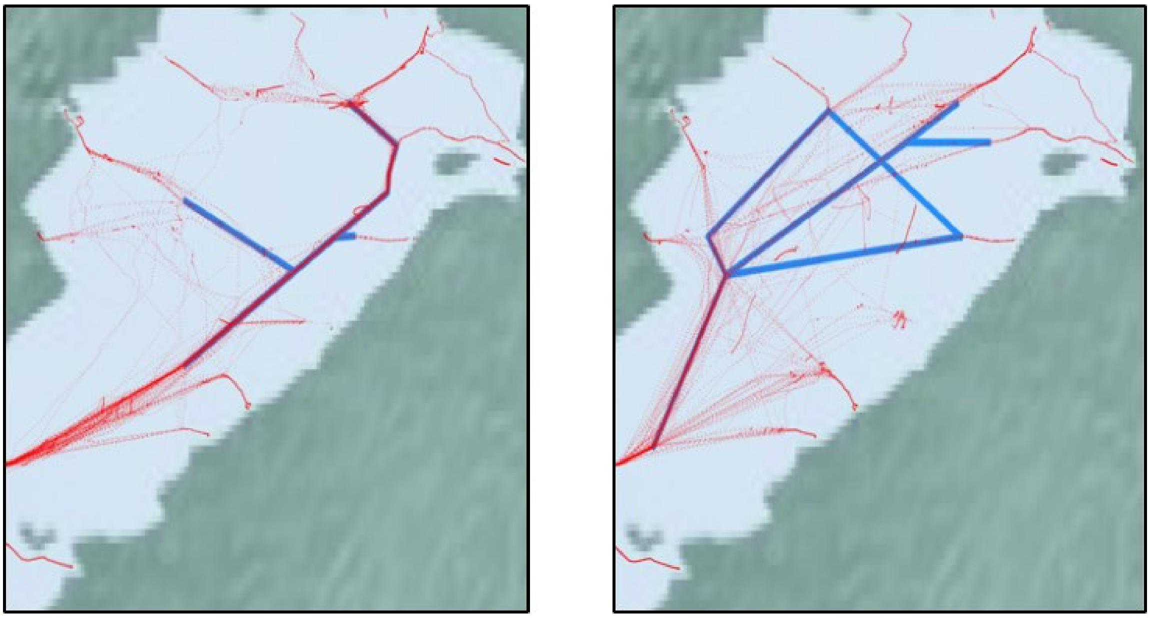
Many of the operational challenges are not well researched and their effect to vessel movement remains unclear. However, the presence of ice, icebreakers and dirways have been shown to affect vessel movement in winter (Guinness *et al.* 2014, Montewka *et al.* 2017, Riska and Kämäräinen 2011). These factors are examined in detail in the following sections.

2.3.1 Vessel Routes in Ice

In winter, when a ship is travelling to a port with a traffic restriction, icebreakers, maritime and VTS authorities direct it through a dirway (Guinness *et al.* 2014, Montewka *et al.* 2017). The current traffic situation and the ice conditions inform the dirway creation process. The main inputs from the traffic situation are the number of vessels that are predicted to require assistance. The prediction is done using the locations and destinations of vessels heading towards or out of frozen ports. The main inputs from the ice conditions are the current and the predicted ice thickness, pressure and drift. Furthermore, the depth of the sea influences waypoint selection, as icebreakers avoid navigating in shallow waters where the risk of accidents increases.

Dirways are created and maintained by icebreaker captains since they have the expertise in travelling through ice, the latest information about the ice field and the viability of different routes. The decision to direct a ship is based on the vessel's ice class and the crew's experience in navigating through harsh ice conditions (Guinness *et al.* 2014). The vessels that are not directed often also benefit from using dirways as the dirways are travelled repeatedly and thus they are more likely to be traversable.

The distance between two waypoints in a dirway can be tens of nautical miles and the navigation though waypoints is left to the vessel crew (Guinness *et al.* 2014). Dirways can be valid for weeks but in a typical winter are often updated in 2-3-day intervals. The change in dirways during winter 2019 is depicted in Figure 5.



10.02.2019 **25.02.2019**

Figure 5. Change in dirways (blue) and vessel routes (red) inside two weeks during icebreaking season 2018 - 2019.

Lehtola *et al.* 2019 compared actual routes of vessels through ice in the Baltic Sea against the planned routes by a seasoned ice navigation specialist. The specialist had access to the latest ice and weather forecasts and observations. They noted that the presence of dirways had the biggest effect to the vessel path in ice. The same outcome can be seen in the Figure 5 where most of the vessel routes towards the north of the Bay of Bothnia follow the dirways.

Kotovirta *et al.* (2009) reported that merchant vessel captains are not eager to use optimal ice routing software in the Baltic Sea because ships must follow dirways in order to receive assistance as fast as possible. If a ship deviates from the given waypoints, it will be placed on the bottom of the icebreakers' assistance queue (Guinness *et al.* 2014). Thus, the vessels that might require assistance tend to follow the given dirways.

To conclude, this section has described how the presence of dirways, properties of the icefield, the topography of the sea, the ice class of a vessel and the crew's experience in ice affect vessel routes. Dirways appear to model the vessel routes in the Baltic Sea most accurately as vessels are often required to follow them. In addition, they indirectly encompass many of the other factors affecting vessel movement in ice.

2.3.2 Vessel Speed in Ice

The prior research of vessel performance and speed in ice is more complete than the literature of vessel routes. The theoretical ice speed has been modelled and the effect of traffic in ice has been explored thoroughly (Kotovirta *et al.* 2009, Guinness *et al.* 2014, Löptinen and Axell 2014, Montewka *et al.* 2017, Sormunen *et al.* 2018).

Ice concentration, compression and level have been shown to affect the speed of a vessel (Kotovirta *et al.* 2009). Ice concentration C can be used to approximate vessel transit speed in ice using the following formula:

$$v_{tr} = \begin{cases} v_{ow}, \ C \leq C_0 \\ \frac{(C_1-C)\, v_{ow}+(C-C_0)v_{i,eq}}{(C_1-C_0)}, \ C_0 < C < C_1 \\ v_{i,eq}, \ C \leq C_1 \end{cases} \quad (1),$$

where C_0=70%, C_1=95%, v_{ow} is open water speed and $v_{i,\,eq}$ is speed affected by ice level and ridges.

Although the formula has been shown to model the theoretical ice speed well, the formula might not be able to capture ice speed accurately when vessel is travelling through a dirway or is assisted by an icebreaker. Sormunen *et al.* (2018) compared theoretical ice speeds of vessels to the ice speeds collected from AIS data in the Baltic Sea. They discovered that the available resolution of the ice forecasts is not good enough for accurate ice speed calculation. This is most notable when vessel is travelling through a dirway as the ice properties of the dirway can differ vastly compared to the surrounding ice field.

When a vessel is stuck on ice or has high probability of besetting, icebreaker will assist the vessel through the ice field by leading or towing (Goerlandt 2017). The waiting time for a vessel to get assistance depends on the length of the icebreaker's assistance queue. Icebreakers often gather several ships in a convoy for efficiency (Montewka *et al.* 2017, Goerlandt 2017). However, this can lead to vessels needing to wait until a critical mass of vessels is gathered around the same area and are travelling to the same destination.

Convoy can also be formed if an icebreaker is heading to assist a vessel that is stuck in ice and at the same time leads another vessel to that direction. Screenshot from the IBNet visualizing forming of a convoy is presented in Figure 6. When travelling in a convoy, vessels tend to travel slower in order to minimize the risk of collisions and to save fuel.

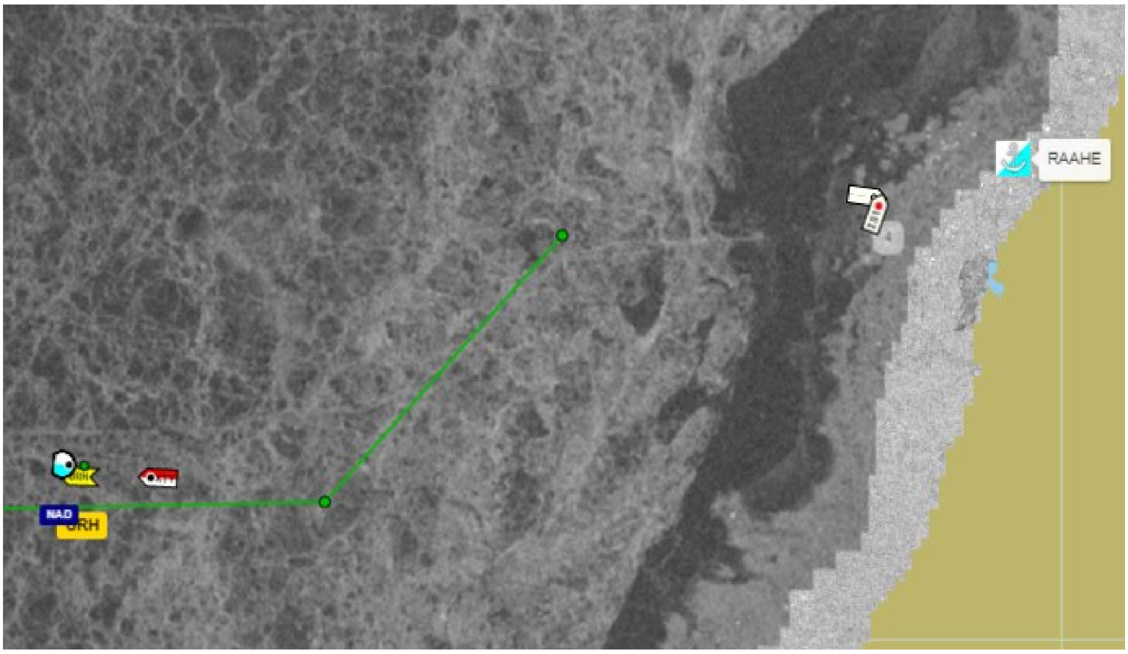


Figure 6. Icebreaker Urho (yellow) leading one vessel (red-white) while going to help another vessel (white-blue) that is stuck on ice. The vessels are not following the dirway (green) closely.

Although the factors influencing the waiting time of a vessel are known, the approximation of it is a complex task. The accurate approximation would require information that is not easily available such as the assisting icebreaker, length of the assistance queue, is the icebreaker going to use convoys or assist vessels one by one, how far the icebreaker will assist the vessels in the queue and the ice speeds of the icebreaker and the vessels in the queue.

In summary, it is evident that the presence of ice, dirways and icebreakers affect the vessel speed significantly. However, the prior literature doesn't reliably describe how vessel ice speed can be modelled in operational situations. Still, the effect of ice should be considered when modelling ice speed if the properties of the ice field are known. Lastly, the accuracy of the ice speed modelling should increase as more accurate ice forecasts become available.

3 TRAFFIC MODELLING

This chapter is structured as follows: first, an overview of the literature related to traffic modelling is given. After that, the remaining chapter is divided into two sections based on the underlying method used in the research. The sections are artificial neural networks and shortest path algorithms. The covered methods were chosen based on the amount and merit of the prior literature. Nonetheless, there are other approaches such as clustering based methods (Pallotta *et al.* 2013, Vries and Someren 2009) and ant colony algorithm (Choi *et al.* 2015) that have been used to model vessel routes.

3.1 Overview

Marine traffic modelling can be divided into four subcomponents: sea modelling, manoeuvrability modelling, route estimation and speed prediction (Tu *et al.* 2018). Although only few studies about the prediction of marine traffic exists, the subcomponents are researched more thoroughly. Topics covered in this chapter relate to traffic modelling or one of its subcomponents. The covered topics are traffic prediction, route estimation, path planning, arrival time estimation, anomaly detection and speed prediction.

Path planning or route estimation relates to predicting the optimal path for a vessel to travel (Guinness *et al.* 2013, Topaj *et al.* 2019, Tu *et al.* 2018). The optimal path is often modelled in terms of time, safeness or fuel usage. The route estimation and the path planning methods have many similarities with traffic modelling and often the methods could be used to predict traffic situation with minor changes. The notable difference to traffic modelling is that in route estimation one is only interested in modelling the route of the vessel, not the speed along the route.

Estimated time of arrival prediction can be done either by modelling the whole route of a vessel to a port and calculating the travel time for the vessel (Alessandrini *et al.* 2018) or by predicting the arrival time based on the current position of the vessel without route estimation (Bodunov *et al.* 2018).

Anomaly detection refers to constructing a model of the normal traffic flow and building a method for detecting objects that deviate from the model (Osekowska *et al.* 2014). Anomalies can be related to position, speed or time of a vessel (Tu *et al.* 2018).

The frequency of AIS based research has increased rapidly in the recent years as shown in Figure 7. The AIS data is often linked with additional data sources as the AIS data is not usually enough for accurate analysis of complex topics like ship performance analysis or maritime safety and risk analysis (Lensu and Goerlandt 2019).

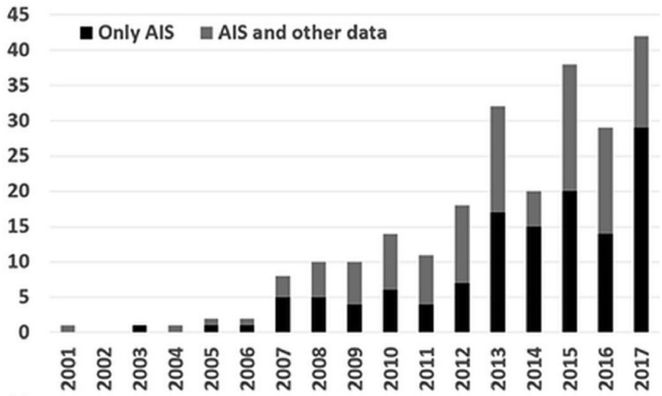


Figure 7. History of Scopus-index articles using AIS data. Only articles with a focus on operational use, e.g. marine traffic monitoring or collision avoidance, are selected (Lensu and Goerlandt 2019).

3.2 Artificial Neural Networks

In this section, artificial neural networks (ANNs) are described in general and the approaches using ANNs are presented and evaluated. ANNs can be used in modelling of non-linear systems, regression and time series forecasting (Kaastra and Boyd 1996, Zissis et al. 2015, Tu *et al.* 2018). ANNs can be used at modelling complex datasets like enriched history AIS data without any additional mathematical modelling or prior information. Artificial neural networks can be used to model traffic without modelling any of the subcomponents (Zissis *et al.* 2016).

3.2.1 Description

A multilayer feed forward ANN has three layer types: an input layer, a hidden layer and an output layer as depicted in Figure 6 (Kaastra and Boyd 1996). Multilayer feed forward neural network refers to an ANN where information moves only one direction (Svozil *et al.* 1997). Network can have multiple hidden layers, but only one input and output layer.

The input layer takes several input signals and transmits them to the neurons of the hidden layer. The number of neurons in the hidden layer can vary although several rule-of-thumb methods exists for selecting the number of neurons in the hidden layer (Berry and Linoff 2004, Boger and Guterman 1997, Gougoulidis 2008, Huang and Babri 1998, Sheela and Deepa 2013). In addition, several algorithmic ways to select number of neurons have been proposed (Sheela and Deepa 2013).

The hidden layer uses an activation function to compute and map the results to the output layer. Finally, the neurons in the output layer sum up all the inputs from the hidden layer and give the output of the network. When the network is trained the weights of the neural networks are adjusted until the network produces desirable outputs. A basic structure of an artificial feed-forward neural network is depicted in Figure 8.

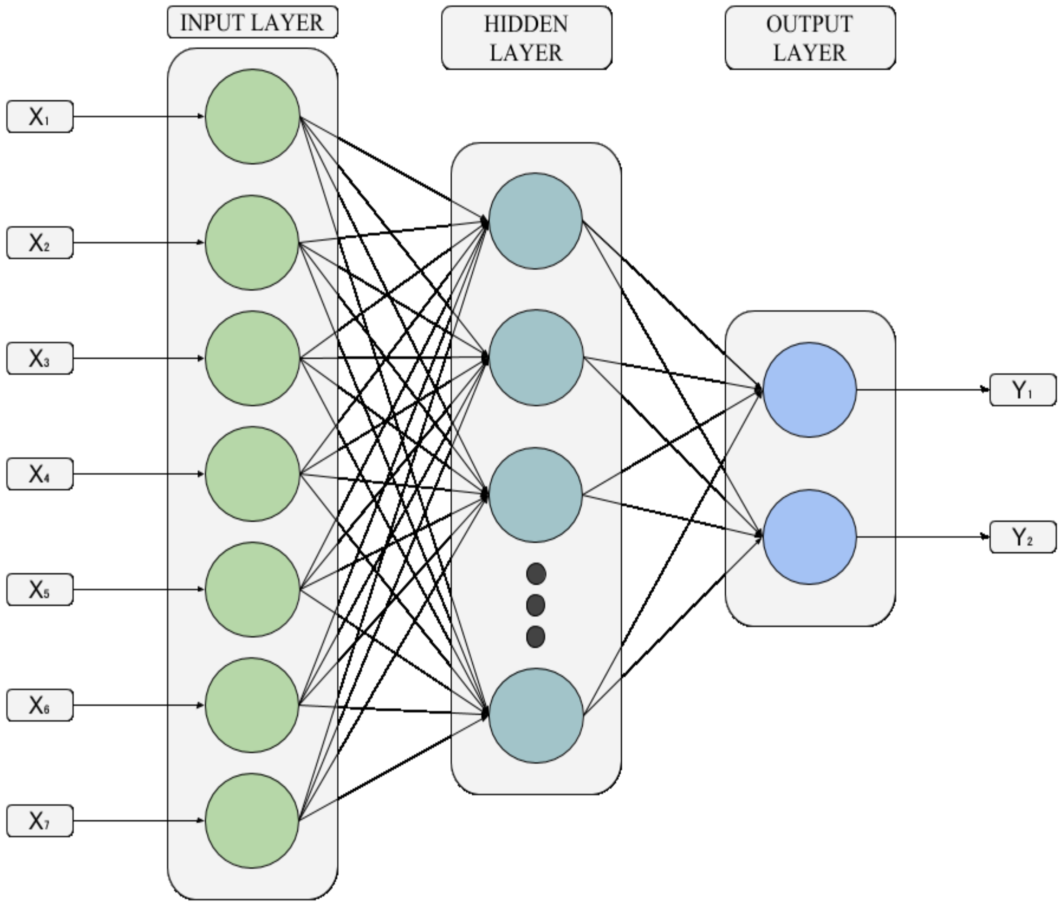


Figure 8 Basic structure of an artificial feed-forward neural network. Modified from Svozil *et al.* (1997).

In general, the advantages of ANN approach are that they are studied thoroughly, performance is generally good, and they can be applied to variety of different problems. ANNs have a strong fitting ability and therefore don't require any prior information (Tu *et al.* 2018). Nonetheless, disadvantages of ANN are that the training process requires a large dataset, can be time consuming and in most cases the network architecture must be determined empirically based on the prediction task.

3.2.2 ANNs in Intelligent Maritime Systems

ANNs have been employed to solve variety of different problems in maritime systems. Kim and Lee (2018) present a deep learning method for predicting marine traffic in a small area. Goal is to give VTS operators a tool for monitoring future traffic in a caution area where ship collision risk is high. Method uses AIS data and can predict traffic for the caution area 20 to 50 min intervals. Training data consisted of 8 million observations. The method shows promise as it improves the accuracy compared to the benchmark methods. However, the method is only able to predict how many vessels are inside the caution area, not the actual positions of the vessels.

Mao *et al.* (2018) used extreme learning machine (ELM) to predict vessel routes for 20-minute and 40-minute time periods. Error for the predictions was 0 - 2.5 nautical miles for the 20 min interval and 0 – 6 nautical miles for the 40 min interval. Error was measured as surface distance between the predicted position and the real position. Error was calculated using Haversine formula. The primary goal of the research wasn't to improve existing approaches but to establish a standardized AIS database for maritime modelling and big data research.

Daranda (2016) proposes a traffic prediction approach which uses combination of AIS data, clustering and ANN for prediction. The AIS data was enriched with turning points of the vessels which are calculated by detecting when vessel course changes. After that, the turning points are filtered by clustering. Finally, an ANN is used to predict vessel routes by predicting the next turning points. Although the results seem promising, the accuracy was measured by visualizing three predicted routes on top of the actual route. No quantitative accuracy measurements were given.

Zissis *et al.* (2016) developed an approach utilizing ANN for real-time vessel route prediction. In contrast to the traditional way of training an ANN where the ANN is trained asynchronously, and the model is used to do vast number of predictions, the ANN was trained per vessel in real time on user request. The data used to train the ANN consists of the last 72 hours of AIS data in 15-minute steps. The approach demonstrates the usability of ANNs for traffic prediction even with a limited dataset. The approach can predict vessel positions up to 124 hours into the future. Although the research included both qualitative and quantitative accuracy measures, the size of the validation set was 3-5 vessels. The size of the validation set is estimated from the figures and tables used in the research as the exact number was not given.

In summary, the prior literature about ANNs in marine applications is few and far between even though the results are promising. The main reason for the lack of research has been the absence of standardized and publicly available AIS datasets (Zissis *et al.* 2015, Mao *et al.* 2018). Furthermore, there is no standardized way of measuring accuracy and the prediction horizons are often short.

3.3 Shortest Path Algorithms

Shortest path algorithms (SPAs) solve the shortest path problem by finding the optimal path from a network of connected nodes using a cost function or a static weight between points (Ahuja *et al.* 1990). Nodes are represented as a graph.

In intelligent maritime system research, SPAs are commonly used in route estimation and path planning. Furthermore, SPAs are prevalent in other domains such as car and pedestrian navigation (Lauther 2004, Van Toll *et al.* 2012, Zhan and Noon 1998).

3.3.1 Description

There are multiple different SPAs that can be applied to marine traffic prediction. The most common algorithms are Dijkstra's (Zhu *et al* 2016, Topaj *et al.* 2019), A* and variations of the A* (Chabini and Lan 2002, Guinness *et al.* 2014, Kotovirta *et al.* 2009, Lehtola *et al.* 2019, Montewka *et al.* 2017, Topaj *et al.* 2019).

SPAs are based on the dynamic programming theory (Fu *et al.* 2006). The dynamic programming theory refers to a heuristic used in computer science to solve complex problems by dividing them into subproblems (Bellman 1954). The subproblems are solved once and their solutions stored. If the algorithm encounters already solved subproblem again, it looks up the solution instead of solving it again and thus saving computation time.

The shortest path between two nodes in a graph can be found by going through the graph from the origin node to the destination node using a recursive cost function. The cost function refers to a function which selects the cheapest route by minimizing the distance or weight between nodes (Fu *et al.* 2006). The cost function can take many different inputs. An example result of a SPA is pictured in Figure 9.

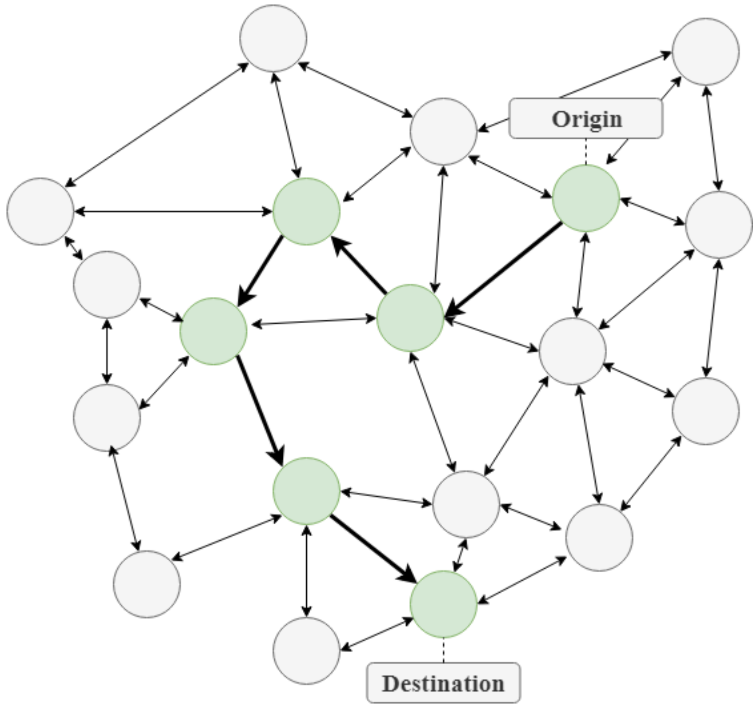


Figure 9. Result of a shortest path search. Modified from Fu *et al.* 2006.

In the maritime domain, the graph for a SPA is usually constructed by dividing a sea area into a grid and connecting the cells inside the grid (Chabini and Lan 2002, Guinness *et al.* 2014, Kotovirta *et al.* 2009, Lehtola *et al.* 2019, Montewka *et al.* 2017, Topaj *et al.* 2019). This approach allows the use of additional location-based data in the cost function if it can be represented as points or cells.

All of SPA implementations are not suitable for real-time predictions in traffic networks (Fu *et al.* 2006). For example, the Dijkstra's algorithm is simple outward search technique that doesn't use any prior knowledge of the position of the origin and destination nodes which makes the algorithm inefficient when the graph is large. To that end, most of the SPAs try to solve this issue by using one or more of the following strategies: limiting the search area, decomposing the search problem or reducing the links (Fu *et al* 2006).

In general, the main benefits of SPAs are that they are well researched, have been employed to traffic and route prediction and the cost function can be easily adjusted to take into account additional information sources such as map and weather data (Fu *et al.* 2006, Topaj *et al.* 2019, Tu *et al.* 2018). Furthermore, the use of SPAs in winter navigation applications is well researched (Guinness *et al.* 2014, Kotovirta *et al.* 2009, Lehtola *et al*. 2019, Montewka *et al.* 2017, Tarovik *et al*. 2017, Topaj *et al.* 2019).

3.3.2 SPAs in Intelligent Maritime Systems

SPAs in winter navigation research are used mainly for path planning (Guinness *et al.* 2014, Kotovirta *et al.* 2009, Lehtola *et al.* 2019, Tarovik *et al*. 2017, Topaj *et al.* 2019). Few of the studies even consider how icebreaking activities affect the vessel movement in ice. The geographical areas of interest are the Baltic Sea and the Arctic Ocean. Besides winter navigation, other points of interests are route estimation in general, prediction of estimated time of arrival (Alessandrini *et al.* 2018) and anomaly detection (Osekowska *et al.* 2014).

Guinness *et al.* (2014) propose a modified A* algorithm to predict optimal ice routes in the Baltic Sea. The algorithm optimizes travel time by modelling the Baltic Sea, vessel manoeuvrability, vessel ice speed and icebreaker assistance. The dataset comprised of AIS data enriched with ice forecasts from the Finnish Meteorological Institution. Although the dataset contained AIS data, the method did not model historical vessel routes in any way. The vessel manoeuvrability modelling allowed more realistic representation of vessel movement inside the graph compared to the traditional adjacent neighbour modelling. The optimal routes were not tested in simulation or by having a vessel to travel the routes.

Tarovik *et al.* (2017) used a similar approach as Guinness *et al.* (2014) to predict optimal ice routes in the Arctic Ocean. The method used A* algorithm, icebreaker assistance and modelling of the sea. In addition, their implementation used pre-defined routes such as fairways and internal port routes that vessels are obligated to follow. The dataset comprised of historical AIS data enriched with weather data. In contrast to Guinness *et al.* (2014), the predicted optimal routes were tested in simulations. The study highlights the need for multidisciplinary approach when developing a winter navigation system.

Topaj *et al.* (2019) continue the work of Tarovik *et al.* (2017) by adding icebreaking assistance as a key component to the route optimization. The approach considers the amount of icebreaking assistance that is needed for a single vessel and the time for an icebreaker to reach the point where the vessel would need assistance. The approach was tested in simulations and was shown to be able to estimate the amount of icebreaking assistance needed and the optimal route for a vessel. However, the authors note many uncertainties regarding the results.

In a similar research, Montewka *et al*. (2017) propose a method for modelling winter maritime traffic in the Baltic Sea. To model the traffic, the authors use a hybrid model which combines A* algorithm, AIS data, ice forecasts and ship manoeuvrability modelling. The method can predict vessel speed in ice accurately but fails to predict the vessel route in a meaningful way. The vessel route prediction fails to consider how the icebreaking activities affect the vessel routes. Furthermore, the authors note that the ice forecast data differed from the real conditions and failed to capture the effect of wind closing ice channels. The testing was qualitative and comprised of examination of one voyage in ice. Still, the study highlights the variety of factors influencing the winter vessel routes in the Baltic Sea.

Finally, Lehtola *et al.* 2019 model optimal ice routes in the Baltic Sea by using A* algorithm, ice speed modelling, AIS data, ice forecasts and modelling of artificial ice channels by using historic AIS data. The optimal routes were validated against the AIS data and two routes planned by an ice navigation specialist. The optimal routes and the specialist routes differed significantly from the real AIS routes. The authors noted that the approach fails to consider how dirways and icebreaking activity affect optimal routes in ice.

3.4 Discussion

The intelligent maritime systems are relatively new development which is apparent in the current literature. The principles governing vessel movement are well known but their affect to marine traffic and optimal routes are not well studied. The future research could benefit from multidisciplinary approach where domain experts and marine scientists are included in the modelling process.

The validation of path planning research is a difficult task as it requires either real world test sailings of routes or accurate simulations which do not currently exist. Thus, often the validation is done only by visually comparing the planned routes to the historical AIS tracks. This evaluation approach can give meaningful results if the aim is to produce more efficient routes. However, if the goal is to improve safety of navigation or minimize besetting probability, the approach often fails to validate results. Same observation has been done by Guinness *et al.* (2014). In addition, the visual accuracy examination does not scale for evaluating large validation sets.

The topic of traffic modelling has not been covered extensively by the prior literature. Still, the path planning research provides a good starting point for traffic modelling as the underlying methods could be used to model traffic as well (Montewka *et al.* 2017). The accurate modelling of winter traffic in the Baltic Sea requires considering the icebreaking activity as integral part of the modelling.

The prior literature of ANNs shows promise but they have not been used to predict vessel movement in ice. SPAs have been more rigorously studied and the benefits and limitations are better understood.

Finally, the lack of standardized datasets can hinder the replicability results. In addition, the collection of AIS and additional data is a time-consuming task (Lensu and Goerlandt 2019). This can make starting a completely new research more difficult as AIS data is often enriched with additional data sources such as weather, ice and map data. Although Lensu and Goerlandt (2019) have created a database containing AIS and ice forecast data from the last 9 years, it has not been made publicly available. A comparable database would provide a good base for further winter navigation research.

4 DATA PROCESSING

Data processing can be regarded as one of the most important and time-consuming steps in any machine learning application (Zanin et al. 2016). Thus, data processing is explored thoroughly in this chapter. Methods of data visualization and description are used as they are crucial when trying to understand complex and big datasets (Porter and Heppelmann 2014, Vassakis et al. 2018). This chapter examines separately the collection and the preparation of the data.

The use of open data sources and publishing of datasets is often overlooked in the big data research (Boyd and Crawford 2011, Bruns 2013, Mao *et al.* 2018). This is evident in the prior literature covered in the previous section also. Thus, the validity and proves of AIS analytics research can be difficult to assess. To improve the availability of datasets, the raw AIS data including the training and validation sets used in this thesis are published online (Hakola 2020a).

4.1 Automatic Identification System (AIS)

This section describes the implementation of Automatic Identification System (AIS) and the properties and quality of the data it produces. AIS is an automatic vessel tracking system that was introduced to promote safe and efficient maritime navigation (International Maritime Organization 2001, Harati-Mokhtari *et al*. 2007). One of the goals was to improve the quality of vessel traffic surveillance. Furthermore, AIS provides crucial data for understanding the maritime traffic domain and the data can be used to analyse movement patterns of vessels (Lensu and Goerlandt 2019, Mao *et al.* 2018, Tu *et al.* 2018).

The International Maritime Organization (IMO) requires AIS to be active on all passenger ships and ships with gross tonnage (GT) above 300 tons. AIS gives information about vessel's identity, location and the current voyage. This information is transmitted through VHF radio channels between AIS systems such as base stations, buoys, vessels and VTS centres (International Maritime Organization 2003).

Contents of AIS transmissions can be divided into static, dynamic and voyage information (International Maritime Organization 2003). Contents of the different information types are presented in Figure 10**.** AIS data is not sent in one transmission but is divided into different message types. The most common AIS message types are 1, 2, 3 and 5. Messages 1, 2, 3 contains the dynamic information and the vessel's MMSI. The message 5 contains both the static and voyage related information. The division helps to reduce the amount of data sent as the static data is not sent alongside the frequently updated dynamic data.

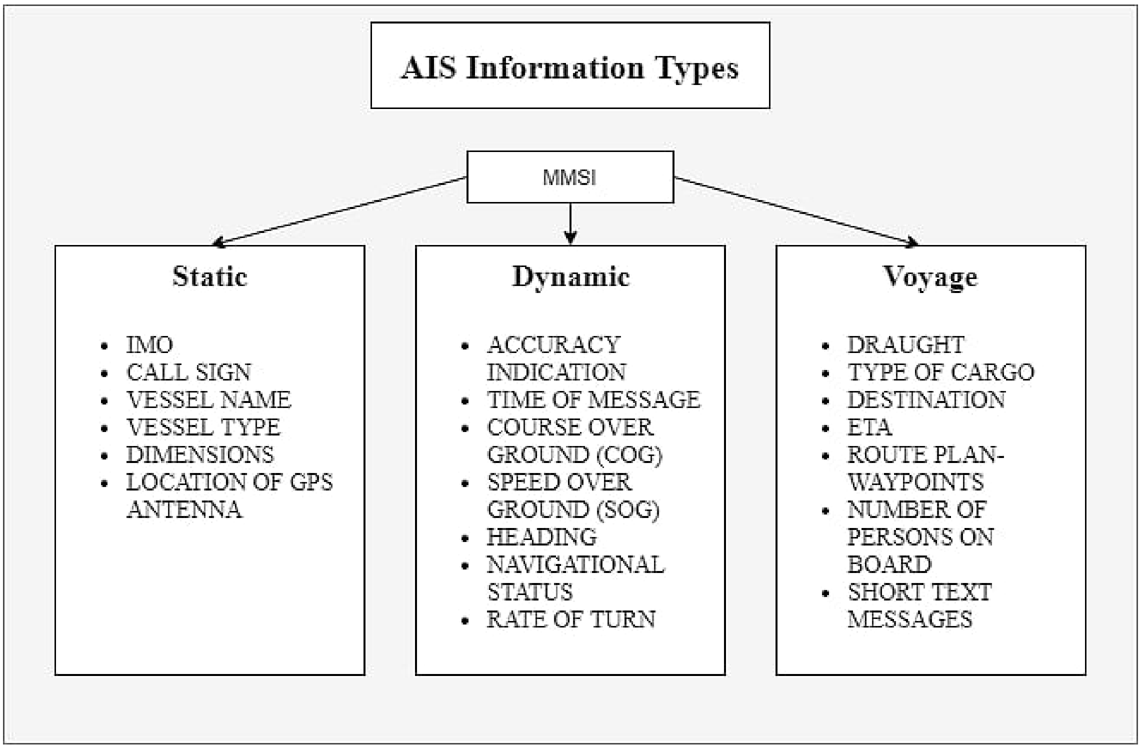


Figure 10. Different AIS information types and their contents (International Maritime Organization 2003).

Although the sending of AIS data is automated, only the dynamic information is read directly from ship's navigational sensors (Harati-Mokhtari *et al.* 2007). The static and the voyage data is manually entered by the ship crew which can affect the validity of the data. The dynamic data can also have errors due to problems with integrations to ship's sensors (Banyś *et al.* 2012). Still, the dynamic data is significantly more reliable than the static and voyage data.

Harati-Mokhtari *et al.* (2007) report that the static and voyage data is often erroneous. They reported problems such as 56 percent of vessel types being wrong and dimensions being incorrect in 47 percent of messages. In addition, the destination and ETA fields were not updated frequently or contained wrong information due to input errors.

4.2 Data Collection

The AIS, port location, dirway and map data used in this thesis are collected from IBNet. Nearly all parts of the dataset can be gathered from open data sources. Similar AIS data is available through multiple sources (Digitraffic 2020, Marinetraffic 2020, Aishub 2020). In addition, similar port location, dirway and vessel metadata is also publicly available (Digitraffic 2020).

The collection area for the AIS data is a rectangle that covers most of the Baltic Sea. The area's most southwest point is 53.5 latitude and 9.4 longitude and most northeast point is 66.1 latitude and 36.1 longitude. The collection rate of the AIS data in different sea areas can vary as IBNet receives AIS data from multiple sources with different AIS coverages.

Map data is taken from the official S-57 map materials provided by Finnish and Swedish governments for the development of IBNet. S-57 refers to a map format used in maritime charts. S-57 material is composed of vectors representing the S-57 object model and has been developed by the International Hydrographic Organisation (GDAL 2019). Traficom provides Finnish S-57 maritime charts as open data (Avoindata 2019). The map material is unofficial, and the data might differ from the data used in this thesis. Swedish map material is not publicly available.

The S-57 map data is converted to ESRI shapefiles as an intermediate step. The ESRI Shapefile refers to a vector data format containing geospatial and geographic information. (ESRI 1998) The specification is developed by ESRI to improve data compatibility among geographic information systems. The conversion is done using an open-source python library (Schylberg 2019) which uses GDAL and OGR libraries to convert the map-data. The shapefile extraction process is visualized in Figure 11.

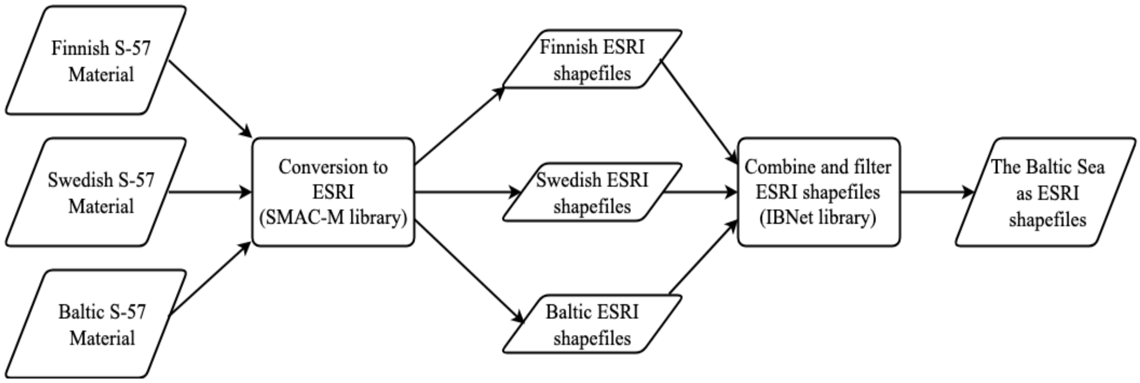


Figure 11. Process of creating map of the Baltic Sea as shapefiles from multiple S-57 materials.

The shapefiles used in this thesis are the land area of the Baltic Sea and the shallow water area. The shallow water area is a subset of the Baltic Sea where the depth of the sea is less than 10 meters. The shapefiles are used in the modelling of the sea described in Chapter 5.

4.3 Data Preparation

The AIS data is merged with the vessel metadata using the MMSI number. The metadata fields are appended to each observation. Then, the AIS dataset is divided into training and validation sets. The preparation of the training and validation sets are described individually in the following sections.

4.3.1 Training Set

In this section, the selection and the creation of the training set is described. The AIS observation time periods for the training set are 01.11.2017 – 20.12.2017 and 25.05.2018 – 21.10.2018. Since vessel movement in winter does not represent the typical vessel movement behaviour as discussed in Chapter 2, the data from winter is left out. Winter data could distort the normal vessel routes and affect the accuracy of the model.

Each observation in the AIS data includes time, coordinate, course, heading and speed information. To avoid skewing the model towards a speed of zero, the dataset is filtered down to observations where speed is greater than 0.1 knots. Average speed of the vessels is not affected by removing observations where the vessel is not moving.

Some vessel types have observations that differ vastly from the rest of the training set. For example, the difference between the type “OTHER” and the rest of the training set is visualized in Figure 12. To reduce the observation amount and simplify the modelling, the five most common vessel types are selected. The remaining vessel types are:

- GC (General Cargo),
- T (Tanker),
- PAS (Passenger ship),
- CONT (Container) and
- RORO (Roll-on, Roll-off ship).

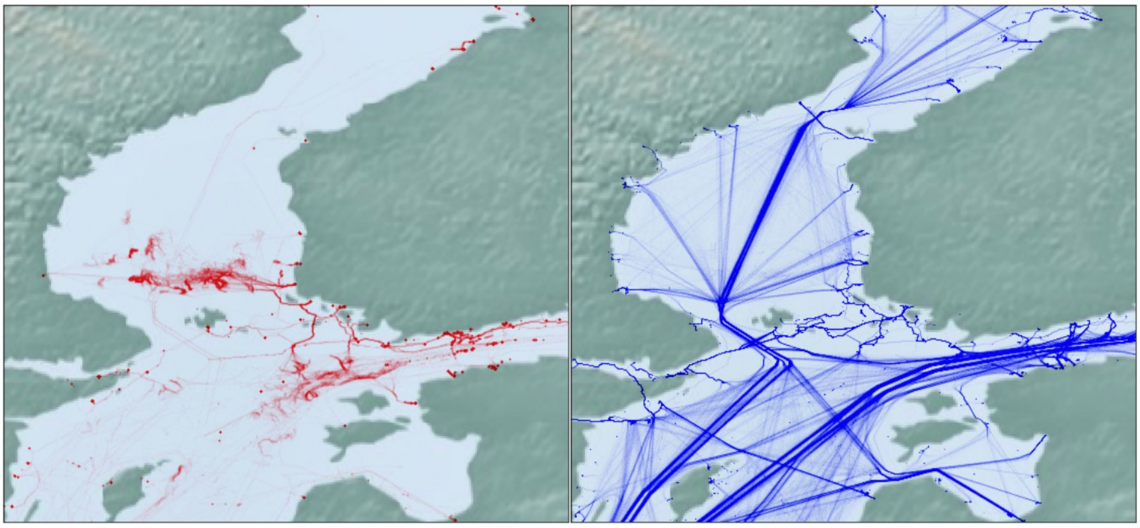


Figure 12. The difference in vessel movements between vessel type OTHER (left) and rest of vessel types (right).

Finally, the resulting training set contains nine million observations from 18 thousand different vessels. There are notable differences between the mean speeds of the selected vessel types. The mean speeds are visualized in Figure 13.

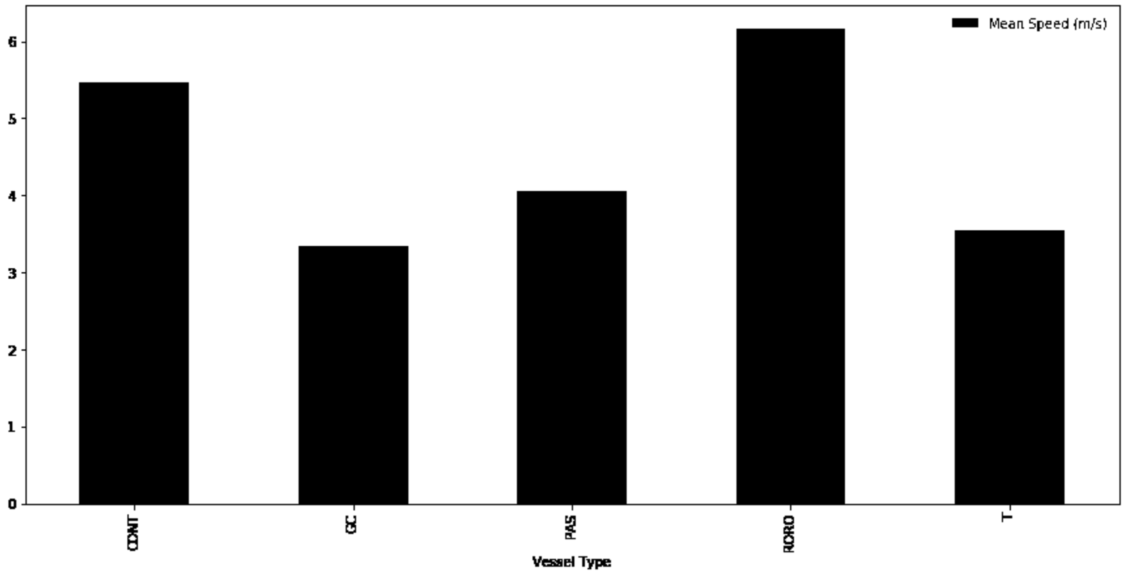


Figure 13. Mean speeds by vessel type in the training set.

4.3.2 Validation Sets

The time periods for the validation sets are 01.02.2019 – 28.02.2019 and 11.05.2019 – 09.06.2019. The first validation set is referred as the winter set and the latter as summer set. As with the training set, only the five most common vessel types are considered.

As the voyage and port visit data in AIS messages has been deemed to be unreliable in Section 4.1, the voyage data is derived from the historical vessel positions. To do this, a simple algorithm is developed that uses the AIS data, haversine distance and port locations to generate voyages (see Algorithm 1). The calculated voyages are used for measuring the accuracy of the model. In short, the algorithm iterates through the AIS observations vessel by vessel and detects voyages by examining if a moving period is between two consecutive stop periods and the stop periods are inside different port areas.

The distance of a vessel to a port is calculated by using the haversine distance which considers the curvature of the earth (Sinnot 1984). The formula for the calculation is

$$d((x_1, y_1), (x_2, y_2)) = 2r_0\sqrt{\sin^2\left(\frac{y_2-y_1}{2}\right) + \cos(y_1)\cos(y_2)\sin^2\left(\frac{x_2-x_1}{2}\right)} \qquad (2),$$

where r_0 refers to the radius of the earth in meters, x_i refers to the latitude of the first and the second point and y_i refers to the longitude values. Haversine distance is used to approximate the geographical distance.

Algorithm 1 Simple voyage calculation

Input: *ais* (AIS observations), *ports* (list of ports)

```
port_radius ← 10 km
max_time_between_messages ← 60 minutes
voyages ← list()
voyage ← dict (id: 0)
for vessel, observations in ais do
  prev ← observations[0]
  for obs in observations do
    obs.in_port, obs.port ← min_distance_to_port(obs, ports) < port_radius_km
    if obs.time - prev.time > max_time_between_messages do
      voyage ← dict(id: voyage.id)
      prev ← obs
      continue
    end if
    if voyage.atd is not empty:
      voyage.observations.append(obs)
    end if
    if prev.in_port and not obs.in_port and is_long_move(obs, observations):
      voyage.atd = obs.timestamp
      voyage.start_port = prev.port
    end if
    if obs.in_port and is_moving(prev) and is_long_stop(obs, observations):
      voyage.end_port ← obs.port
      voyage.ata ← obs.timestamp
      voyages.append(voyage)
      voyage ← dict(id: voyage.id+1)
    end if
  end for
end for
return voyages
```

The algorithm gives a rough approximation of the voyages. The start and end points of the resulting voyages are not precise. However, when testing the model, any subset of the voyage is good enough to test how the model predicts the journey between two points.

The accuracy of the voyages is most dependent on how accurate and dense the AIS data is and how accurately the ports have been modelled. In the dataset, the ports have been modelled as points which does not accurately represent the port area. Ports usually have multiple berths for vessels to stop. Preferably, the berths inside a port would be modelled as polygons so that the voyage could be more accurately generated.

The voyage calculation mapped the voyages correctly in most of the cases. However, in some instances the algorithm failed to detect the stopping of a vessel to a port which resulted in a very long voyage. An example of an erroneous voyage mapping is visualized in Figure 14.

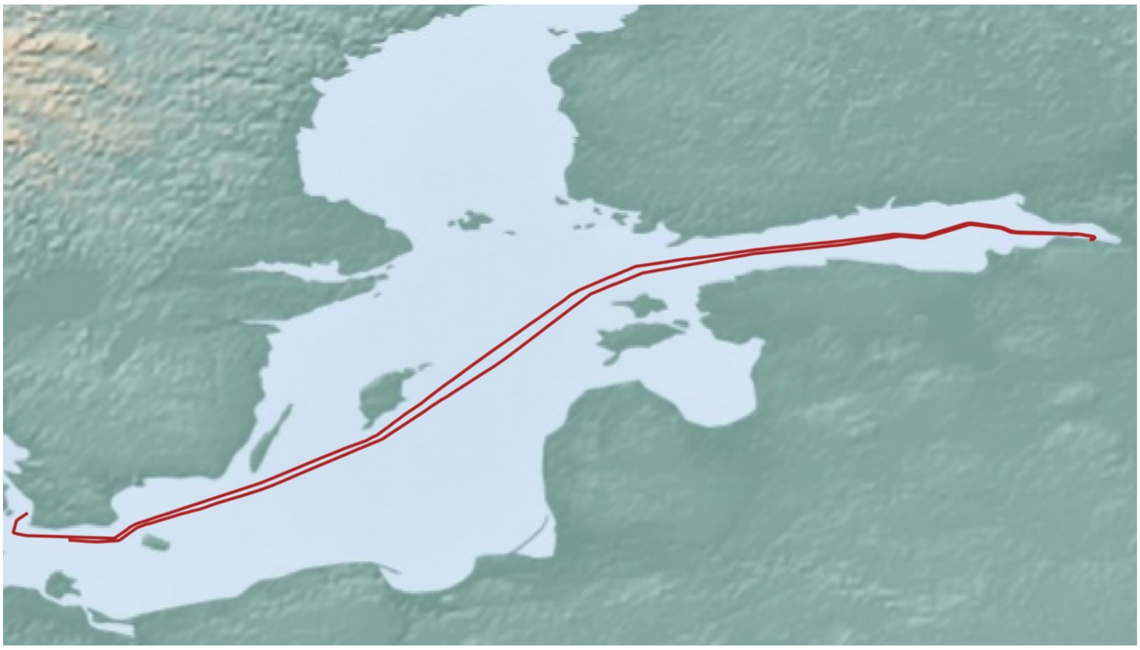


Figure 14. An incorrectly calculated voyage. The calculation failed to detect the port visit of the vessel in the east which resulted in the addition of the return trip to the voyage.

The errors in voyage calculation were detected by comparing the calculated voyage distance to the distance between the start and end point of the voyage. The faulty voyages were filtered out by removing voyages with the length of the voyage being over 100 % longer than the haversine distance between the start and the end point. The filtering resulted to removal of 15.2 % of the original voyages.

Finally, the sea is divided into ten sea areas (see Figure 15) and the calculated voyages from summer are sampled without replacement by selecting 30 voyages ending to each sea area. The winter voyages are sampled without replacement by selecting 30 voyages ending or starting from the Bothnian Bay where ice was present during the collection period of the winter validation set. The resulting validation sets contain 121 thousand observations from 380 voyages.

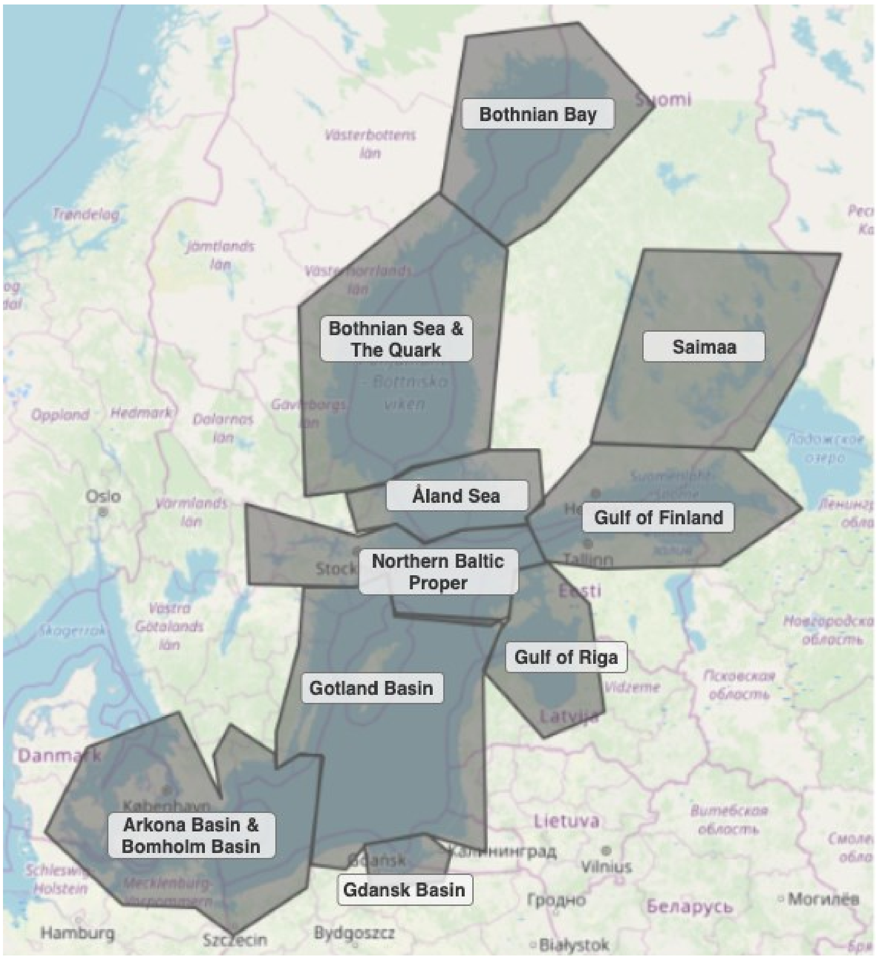


Figure 15. The sea areas used for creating the validation set. Modified from Helcom (2020).

5 METHODOLOGY

In this chapter, a grid-based approach for winter marine traffic modelling is described. The model consists of the following submodels: sea, vessel manoeuvrability, route, speed and icebreaking activity model. The icebreaking activity is modelled through dirways as they have been shown, in Chapters 2 and 3, to affect vessel movement in ice.

5.1 Model Structure

In this section, the submodels of the proposed method are described. As in the previous chapter, different ways of describing and visualizing the method are used. The GitHub repository containing the source code of the approach also contains 17 *Jupyter Notebooks* that go through the whole modelling process in detail (Hakola 2020b). Jupyter Notebooks is an application for creating and sharing documents with live code and visualisations (Jupyter 2020).

5.1.1 Sea Modelling

A sea can be modelled by using *the discretization of the sea* (Guinness *et al.* 2014). The discretisation of the sea refers to a process where the sea is modelled by a set of points s and the number of points in the set is represented by $m=|S|$. The discretisation is a meaningful way to model the sea if the goal is to link additional location-based data to the sea. The linked data in this case are vessel observations, dirways and depth of the sea.

To construct the graph, first the area of the Baltic Sea is discretized into a grid of $L \times H$ square tiles. Second, the land areas are filtered out to avoid predicting impossible routes. Land areas are filtered by going through every tile in the grid and removing the tile if the tile is within a land area. As a result, a uniform grid containing only tiles representing an area of the sea is formed.

After the sea has been modelled, a shallow area graph is created by going through every node in the sea graph and adding the node into the shallow graph if the node is inside the shallow water polygon created in the previous chapter. The resulting two graphs are visualized as points in Figure 16.

The number of tiles in the grid directly relates to the resolution available to model traffic patterns. However, as the resolution grows the computational requirements increase also. To balance the resolution, accuracy and computational requirements, two tile sizes are tested. The tile sizes are 5 km x 5km and 2.5 km x 2.5km.

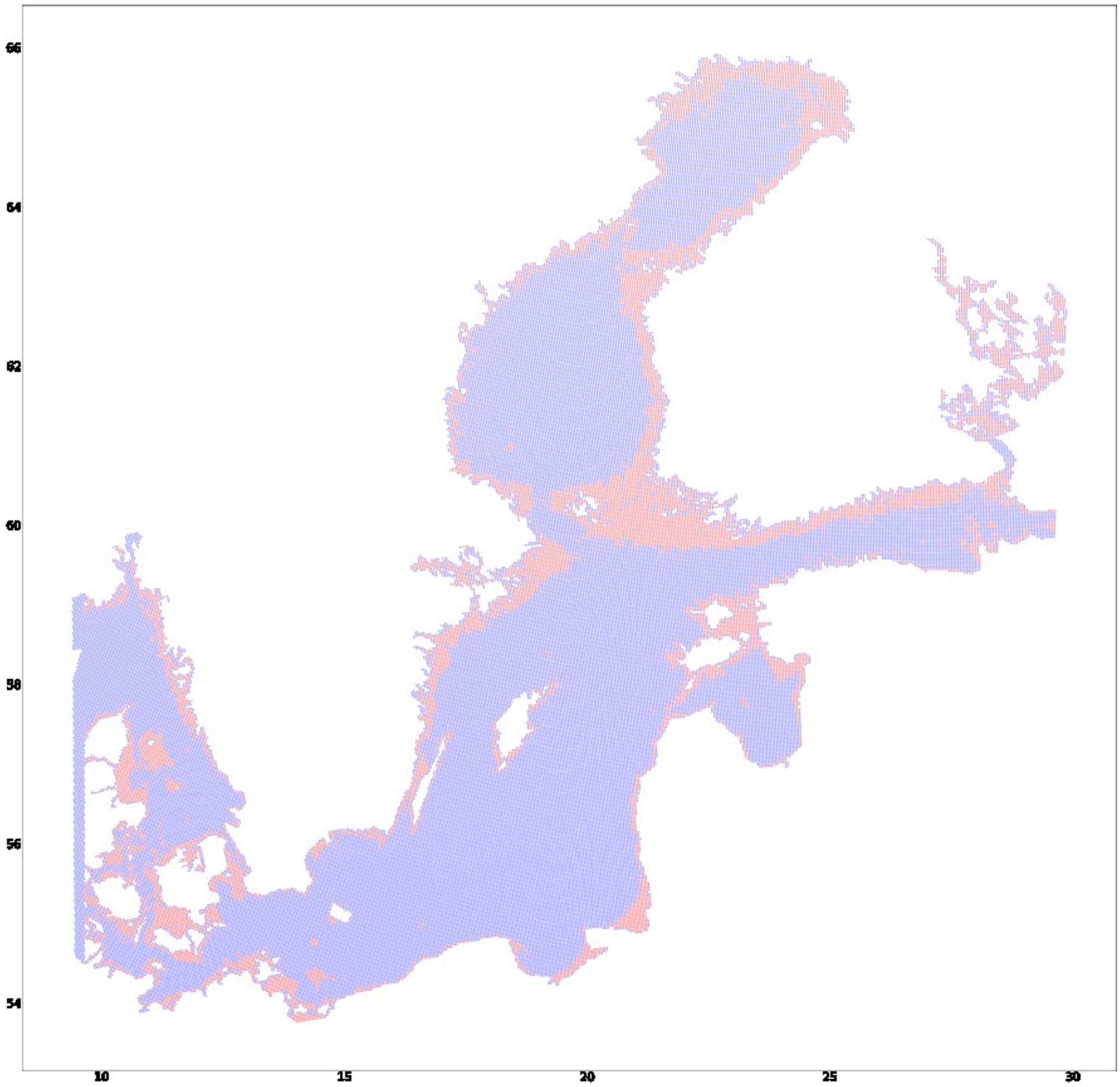


Figure 16. The shallow water nodes (red) drawn on top of the sea graph (blue).

5.1.2 Manoeuvrability Modelling

The grid constructed in the previous section does not yet contain any information about how a ship can move through it. To connect the tiles in the grid, a set of neighbours is defined for each of the tiles in the grid. The neighbours can be constructed by using the adjacent tiles, but this does not accurately model the manoeuvrability of a ship (Guinness *et al.* 2014). Thus, ship manoeuvrability modelling modified from Guinness *et al.* (2014) is used to model ship's movement options in the grid. The adjacent tile modelling is used as a benchmark for the manoeuvrability model. The different ways of modelling movements options are depicted in Figure 17.

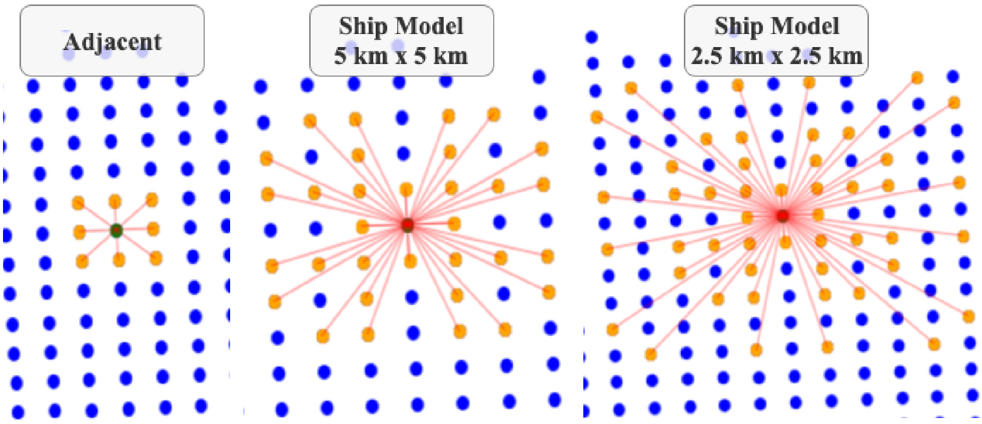


Figure 17. The maximum neighbourhoods of a single node with different ways of modelling.

Guinness *et al.* (2014) used a maximum of 56 neighbours for each tile if all the neighbours are traversable. Moreover, the model describes manoeuvrability when the tile size is 1 km x 1km and the maximum distance between two neighbours is approximately 10 kilometres. Thus, the neighbourhood size is dependent on the dimensions of the tiles if the maximum distance is constant.

In this thesis, larger tile sizes are used which decreases the size of the neighbourhood if the same maximum distance between neighbours is applied. To increase the neighbourhood size, a longer maximum distance is used with the 5 km x 5 km tile size. The maximum distance with 5 km x 5 km tiles is 18 kilometres and 10 kilometres with 2.5 km x 2.5 km. The neighbourhood sizes are 36 for the 5 km x 5 km tiles and 52 for the 2.5 km x 2.5 km tiles.

5.1.3 Route Modelling

In modelling of traffic, we are interested in modelling the typical vessel movement patterns. The traffic patterns are modelled by calculating a transition probability from one tile to another. In short, the probability describes how often a vessel has travelled between two grid tiles. Thus, the weight represents transition probability from one node to another. A weight from node *i* to node *j* is described as

$$W(i \rightarrow j), \qquad (3)$$

The weights are calculated by going through vessel's observations and detecting instances where the vessel moves from one tile to the next. The weight is a sum of all the transitions found in the dataset. The route modelling is visualized in Figure 18.

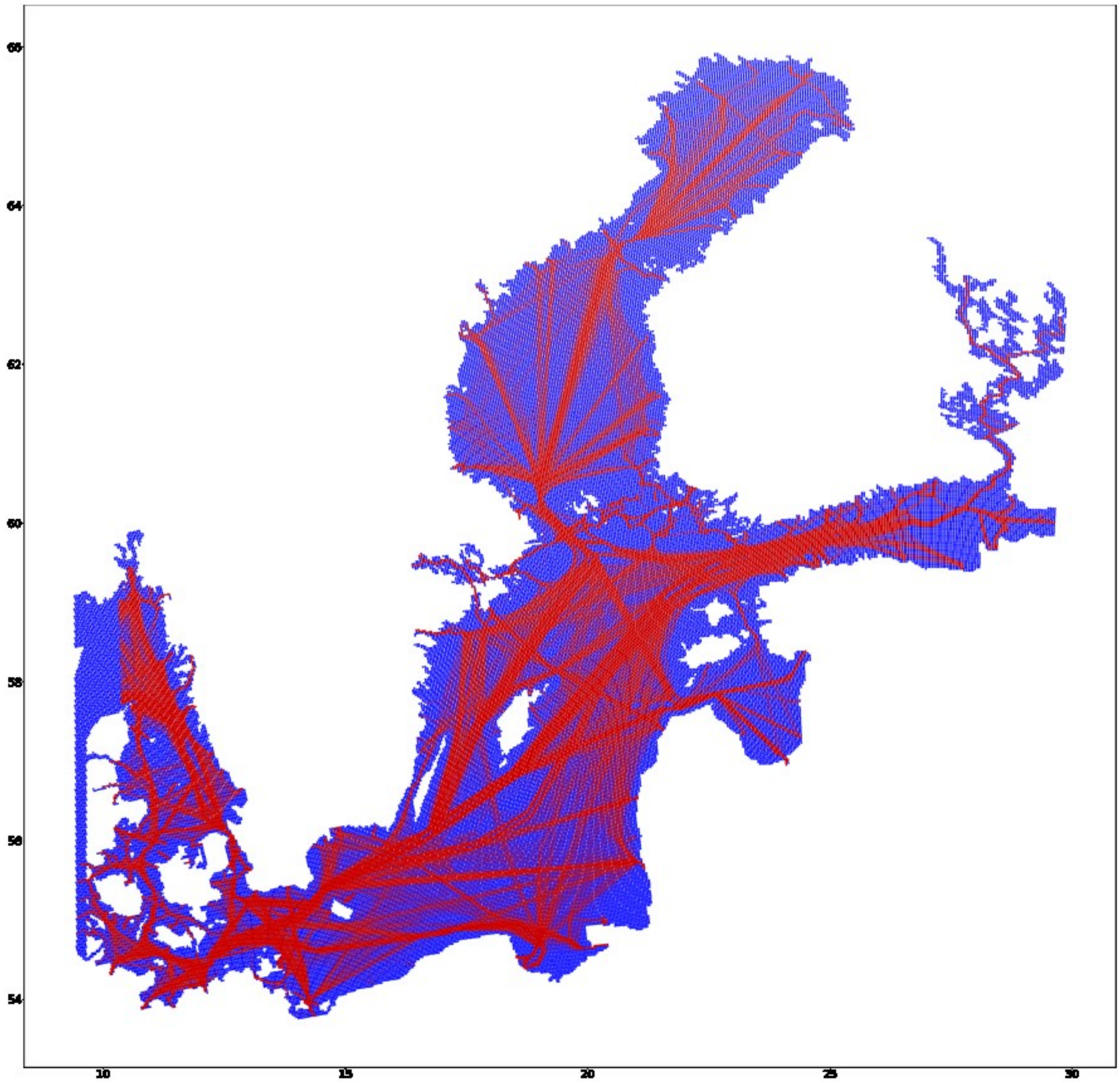


Figure 18. Result of route modelling drawn on top of the sea model

The transitions are detected in two different ways depending on the manoeuvrability model used. When adjacent tiles are used, the transition is simply detected by checking if the next position is on a neighbouring tile. If the AIS data has big enough gap to cause the next observation to be in a tile that is not in the neighbourhood of the current one, the transition is discarded. One way to solve this issue would be to interpolate observations between two distant points. However, the dataset is dense enough and the tile sizes are big enough that the effect of discarded transitions to the weights is insignificant.

When the ship movement model is used, the transitions are mapped using the Algorithm 2. In short, the transition is mapped if the vessel has travelled from the current tile to the neighbour in a 45-minute time window.

Algorithm 2 Transition detection using ship movement model

```
Input: ais (AIS observations)
time_window ← 45 minutes
transitions ←list()
for vessel, observations in ais do
  node ← empty
  for obs in observations do
    if node is obs.node do
      continue
    end if
    node ← obs.node
    neighbours ← get_neighbours(observations)
    future_nodes ← observations.filter(obs.time < x <= obs.time + time_window)
    future_nodes ← remove_duplicates(future_nodes)
      for neighbour in neighbours do
        if neighbour in future_nodes do
          transitions.append([node, neighbour])
        end if
      end for
    end for
end for
return transitions
```

5.1.4 Speed Modelling

Once the traffic has been modelled as a graph, speed modelling can be performed. Modelling of speed is used to predict vessel's location along a predicted route.

Speed is modelled by calculating the mean speed by vessel type in every node in the graph. If no vessel has passed through a node, the speed in the previous node is used. The winter observations and observations with a speed less than 0.2 m/s (~0.4 knots) are filtered out. This is done so that vessels moored at waiting areas, fuel stations and port areas do not distort the speed modelling.

As the nature of the ship acceleration is slow, a friction is introduced to the speed model at the start of the predicted route. The real vessel speed at the start of the predicted route is more accurate than any velocity predicted by the model. Thus, the predicted speed will gradually fall towards the modelled speed from the starting speed. The speed function is

$$v_{ij} = \frac{\left((T_{max} - T_{count})v_{prev} + (T_{count} - T_{max})v_{\text{type mean } ij}\right)}{T_{max}} \quad (4),$$

where T_{max} is the maximum number of transitions for which the friction is used, T_{count} is the number of transitions the model has predicted if the transition count is the equal or less than T_{max} , otherwise it is T_{max}. Mean speed in the node for a vessel type is $v_{\text{type mean } ij}$ and v_{prev} is the speed in the previous node. The previous speeds are not stored in the graph but are kept in memory when doing the path prediction as described in Section 5.3.

To conclude, the proposed approach for speed modelling is quite naïve and gives only a rough estimation of a speed in each node. The use of friction is introduced to give more weight to the actual speed. Additionally, the speed modelling could be improved by incorporating variables such as safe speeds, vessel's maximum speed and ice forecasts to the modelling.

5.1.5 Cost Function

The cost function is the combination of a transition probability to the next node and the travel time from the next node to the destination. Distance to the destination is used to prevent selection of highly traversed routes that do not make sense in terms of travel time. For example, selecting a long path that is commonly used when a significantly shorter but less traversed route is available.

Formula for the cost function is the following:

$$F_{ij} = G_{ij} + H_{ij} \quad (5),$$

where F_{ij} is the cost to travel to the next node with coordinates *(i, j)*, G_{ij} is the transition cost from the current node to node *(i, j)* and H_{ij} is the cost from node *(i, j)* to the destination.

The transition cost G_{ij} is a sum of transition probabilities from the start node to the next node *(i, j)*. The function is

$$\sum_{k=\text{start}}^{(i,j)} (D \rightarrow 0.05) \wedge \left(\neg D \rightarrow P_{(i,j)} + S\right) \qquad (6),$$

where D represents the presence of a dirway in the next node, $P_{(i,j)}$ is the transition probability from the current node to the node *(i, j)* and *S* is a shallow water penalty. The constant dirway cost is introduced to model the winter lanes kept open by icebreaker. The value of the dirway cost has been selected by trial and error and might not suit all situations.

The dirway cost cannot be zero as the dataset lacks information about which points of the dirway the vessel has been directed through. If the cost is zero, then the shortest path algorithm produces highly unusual routes. In addition, the chosen approach enables the use of dirway cost for every ship travelling through an area where dirways are present whether they are directed through the dirways or not.

The shallow water penalty is introduced to avoid selecting routes near archipelagos or shores. As established in Chapter 2, vessels avoid travelling near a shore or an archipelago where the risk of getting stuck is high.

The heuristic for modelling the time to the destination is modified from Guinness *et al.* (2014). They used the heuristic to predict optimal vessel routes in ice. The modified heuristic H_{ij} is

$$\mathrm{H}_{ij} = \frac{\sqrt{(x_i - \mathrm{x}_{\mathrm{end}})^2 + \left(\mathrm{y_j} - \mathrm{y}_{\mathrm{end}}\right)^2}}{v_{ij}} \qquad (7),$$

where x_i is the x-coordinate and y_j is the y-coordinate of the next node *(i, j)*, x_{end} is the x-coordinate and y_{end} is the y-coordinate of the destination node and v_{ij} is the predicted speed in the next node *(i, j)*.

5.2 Route Prediction

In this section, a short description of the route prediction method is given. In addition, the effect of different grid sizes and manoeuvrability models to the prediction result are examined visually. The route prediction method is not elaborated extensively as the python implementation of it is available online (Hakola 2020b).

5.2.1 Description

The route prediction is done using the A* algorithm which calculates the cheapest route from the origin node to the destination node according to the cost function described in the previous section (Hart *et al.* 1968).

The A* algorithm was chosen because of the computational complexity is low compared to other shortest path algorithms such as Dijkstra's. In addition, it has been used extensively in the previous research. However, any shortest path algorithm could be used. The Algorithm 3 describes the pseudo code for the route prediction.

Algorithm 3 Route Prediction

Input: *model (traffic model*), *start_coords*, *end_coords*, *v* (speed*), start_time*

```
open ← set() ← s
closed ← set()
start ← model.get_node(start_coords)
end ← model.get_node(end_coords)
while open is not empty do
  current ← n in open with min(n.g)
  if current is end then
    path.add([end_coords, current.speed])
    while current.parent is not None do
      path.add([current.coords, current.speed])
      current ← current.parent
    end while
    path.add([start_coords, v])
    return reverse(path)
  neighbors ← model.edges[current]
  for n in neighbors do
    n.transitions ← current.transitions + 1
    speed ← model.get_speed(current, type, n.transitions)
    if n in closed then
      continue
    end if
    n.speed ← speed
    n.g ← current.g + model.transition_cost(current, n, speed)
    n.h ← model.cost_to_end(n, end)
    n.f ← n.g + n.h
    if n not in open then
      open.add(n)
      continue
    end if
    for o in open do
      if o.pos is n.pos and o.g > n.g then
        o ← n
        if n in closed then
          closed.remove(n)
          open.add(n)
        end if
      end if
    end for
  end for
  open.remove(current)
  closed.add(current)
end while
```

5.2.2 Discussion

In this section, the accuracy and performance of the route prediction method are discussed and examined visually. Focus is on how different tile sizes and manoeuvrability models affect the prediction results and how the use of dirways impact the prediction in winter. In addition, this section aims to highlight the shortcomings and strengths of the prediction method.

To examine the summer route prediction, two example voyages have been selected from the validation set. The first test voyage is from Halmstad to Luleå and second from Kaliningrad to Kotka. The voyages have been selected to cover most of the sea areas depicted in Figure 14 in Section 4.3.2.

The adjacent model and ship model produce good and similar results with the first voyage using the 5 km x 5 km tile size as can be seen in Figure 19. The adjacent model performs slightly better in the middle of the voyage whereas the ship model predicts the start and end parts of the voyage more accurately.

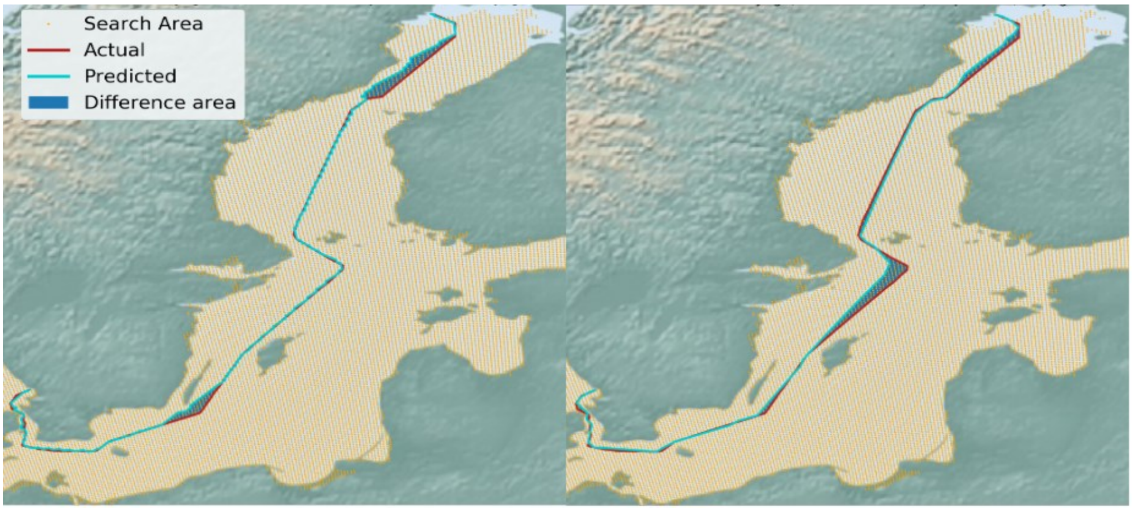


Figure 19. Summer route prediction results visualized using the adjacent model (left) and the ship model (right). Tile size 5 km x 5 km. Voyage id in the validation set 802.

The predictions of the first voyage differ slightly when the 2.5 km x 2.5 km tile size is used as evident in Figure 20. Even though the adjacent model gives almost the same result as with the bigger tile size, the start of the voyage in the Southern Sweden is less accurate and has many sharp turns. In this case, the ship model yields highly inaccurate results in the middle and end part of the voyage.

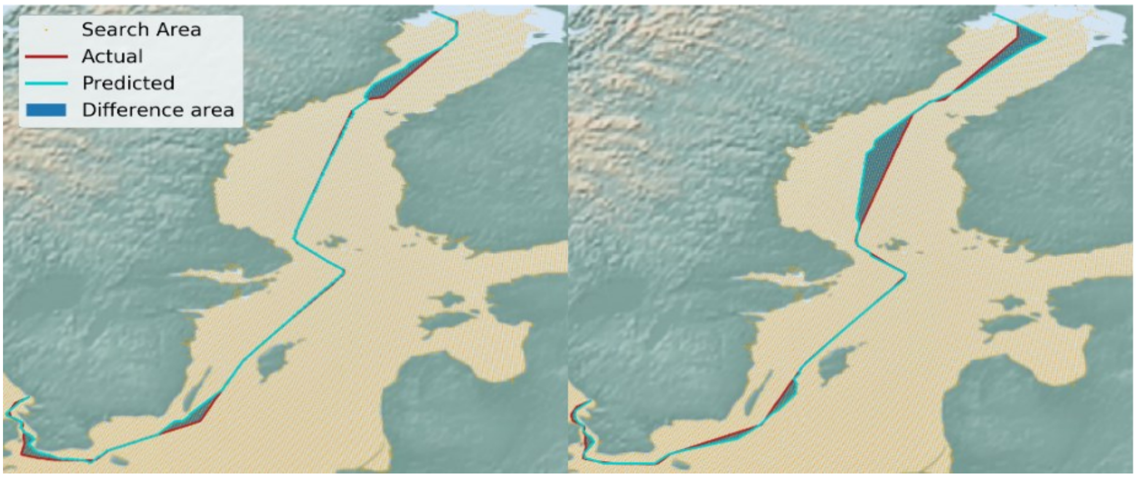


Figure 20. Summer route prediction results visualized using the adjacent model (left) and the ship model (right). Tile size is 2.5 km x 2.5 km. Voyage id in the validation set is 802.

The second voyage predictions with the 5 km x 5km tile size are depicted in Figure 21. The figure shows that the adjacent model fails to predict the route in a meaningful way. Most likely, one of the high-volume traffic lanes from the Southern Baltic Sea to the Northern Baltic Sea, that are visible in Figure 18, distorts the prediction. However, the ship model is not affected by the high-volume traffic lanes and produces a significantly better prediction.

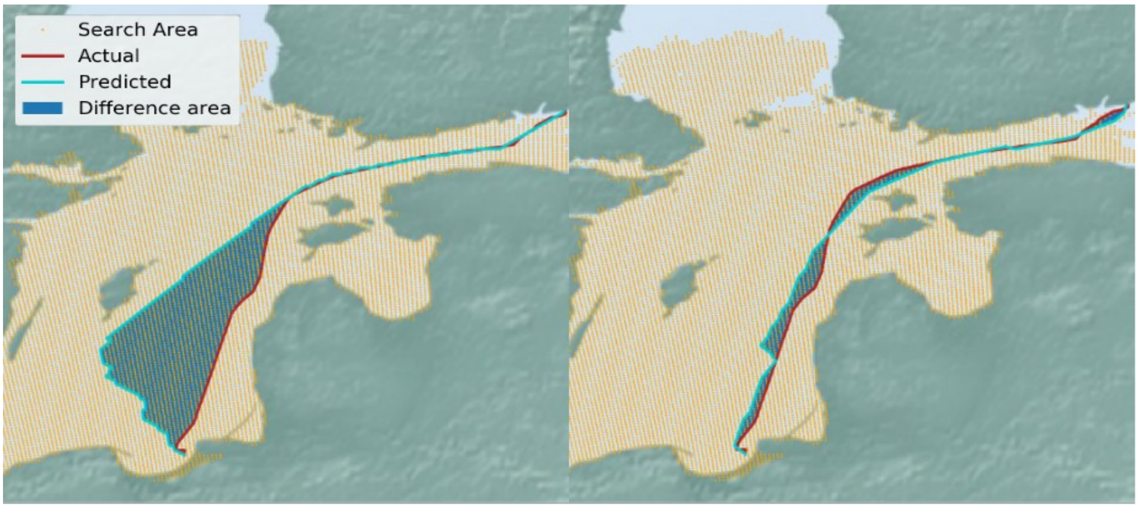


Figure 21. Summer route prediction results visualized using the adjacent model (left) and the ship model (right). Tile size is 5 km x 5 km. Voyage id in the validation set is 5672.

The adjacent model produces slightly improved prediction when using the 2.5 km x 2.5 km tile size which is clear from Figure 22. The ship model prediction has fewer sharp turns but the predicted route deviates more from the actual route compared to Figure 21.

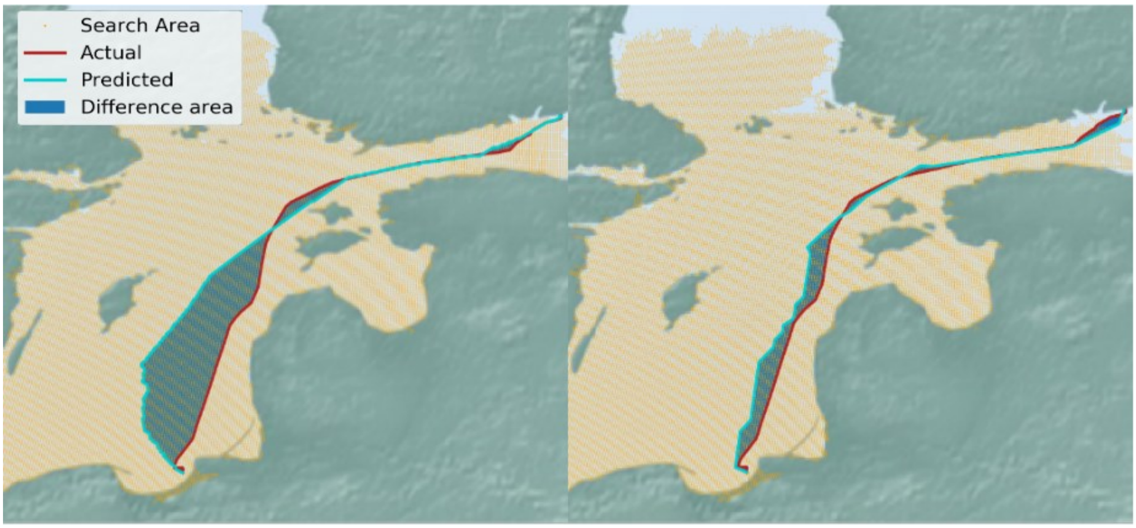


Figure 22. Summer route prediction results visualized using the adjacent model (left) and the ship model (right). Tile size is 2.5 km x 2.5 km. Voyage id in the validation set is 5672.

Next, the impact of dirways to winter route prediction is examined. Two voyages from the validation set have been selected to highlight instances where the use of dirways yields significantly better results. First, Figure 23 shows that the use of dirways improves the prediction greatly in the Northern Bay of Bothnia where the ice field is the thickest.

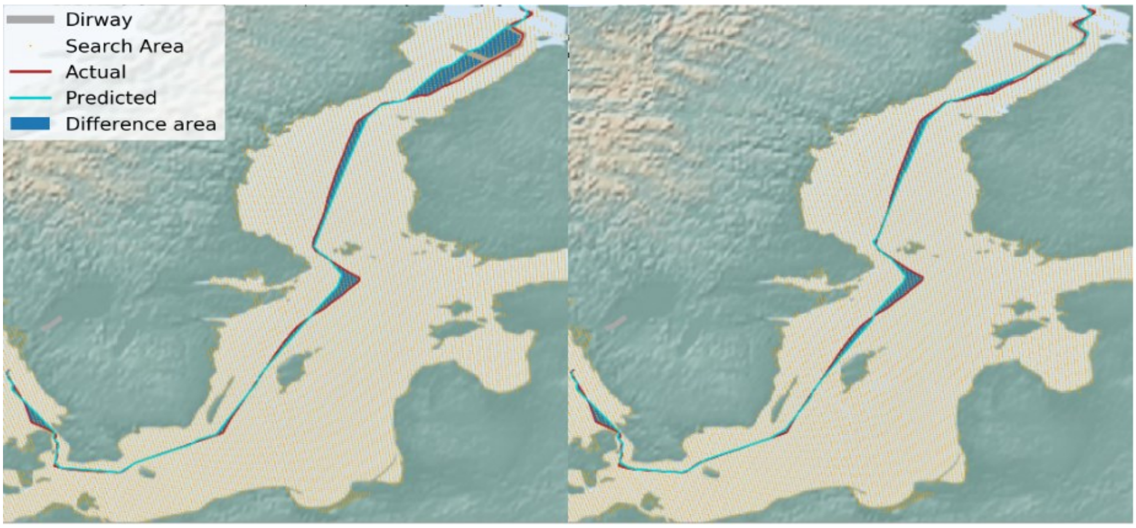


Figure 23. Winter route prediction result comparison with dirways off (left) and on (right). Both predictions have been done using the tile size 5 km x 5 km and the ship model. Voyage id is 97.

The improvements from using dirways in the route prediction are present in Figure 24 also. In this case, the vessel track differs significantly from a typical route to Kemi as it follows the dirways closely. The model using dirways predicts the vessel route accurately in ice.

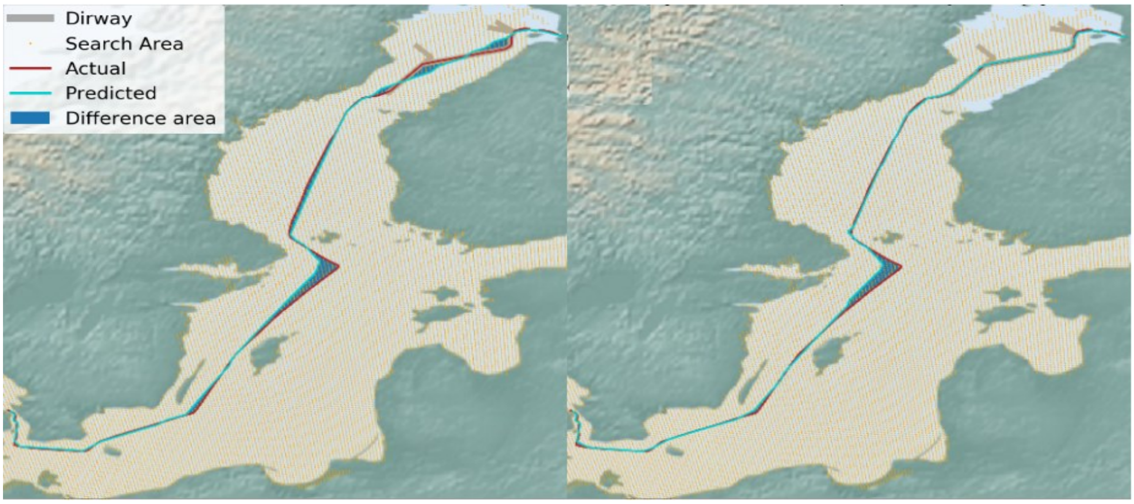


Figure 24. Winter route prediction result comparison with dirways off (left) and on (right). Both predictions have been done using the tile size 5 km x 5 km and the ship model. Voyage id is 587.

In conclusion, the visual examination highlights areas where the model suffers and excels. There are voyages and sea areas that are harder for the model to predict. When predicting winter routes, the use of dirways can improve the accuracy when the route goes through an icefield where dirways are present.

5.3 Traffic Prediction

The traffic prediction method is described in Algorithm 4. The implementation has the advantage of simulating the whole voyage which enables measuring prediction accuracy in different time instants quickly. The traffic prediction method does not consider a case where vessel has multiple port visits in the future. It is developed for the purposes of testing the voyages in the validation set.

Algorithm 4 Traffic prediction

Input: *model, vessel_positions, dirways*

voyages ← *list*()
for *obs* **in** *vessel_positions* **do:**
  voyage ← Route Prediction(*graph*, *obs*.pos, *obs*.end_pos, dirways, *obs*.speed)
  voyages.extend(voyage)
end for
return calculate_timestamps(voyages)

In a real-world implementation of the traffic prediction algorithm, the timestamps for each node should be calculated in the route prediction as it is trivial to do and reduces the time-complexity of the algorithm. The detaching of timestamp calculation gives the advantage of changing the speed model without needing to predict every voyage in the validation set again.

The traffic prediction method could be altered to work in a real-time environment. One possible solution for real-time traffic prediction is described in Algorithm 5. Although the algorithm predicts a single position for each vessel, the use of a SPA means that the algorithm needs to calculate the whole route for the vessel before calculating the single position along the route.

Algorithm 5 Traffic prediction real-time

```
Input: model, observations, dirways, mins_to_future
predicted_positions ← list()
for obs in vessel_positions do:
  portvisits ← find_portvisits[obs.mmsi]
  start_pos ← obs.pos
  start_time ← obs.timestamp
  speed ← obs.speed
  course ← obs.course
  for p in portvisits do:
    if start_time > time_now + mins_to_future do
      break
    endif
   voyage ← Route & Timestamp Prediction (graph, start_pos, end_pos, start_time,
dirways, speed, course)
    start_pos ← p.pos
    start_time ← p.etd (estimated time of departure)
    speed ← 0
    course ← 0
  end for
  if voyage is present do
    predicted_positions[obs.mmsi] ← interpolate_position(voyage, mins_to_future)
  endif
end for

return predicted_positions
```

6 RESULTS

This chapter is divided into sections based on the evaluated component of the model. These components are route, speed, traffic and ETA modelling. Results of each area are discussed in the sections. Lastly, conclusions are drawn from the results.

6.1 Route Prediction Results

Route prediction accuracy is measured by calculating the area between the actual route and the predicted route. The area between routes as prediction error is introduced as a numerical way to measure the route prediction accuracy to improve the deficiencies of prior literature described in Section 3.4. The area between routes has been visualised in the figures in Section 5.2.2.

The ship model performs significantly better than the adjacent model in summer conditions as can be seen from Figure 25. In winter, the ship model performs best without dirways but is the worst performer when dirways are used. Moreover, the use of dirways improves the winter results remarkably with every manoeuvrability model. This is in line with the visualisation of predicted routes in the previous chapter.

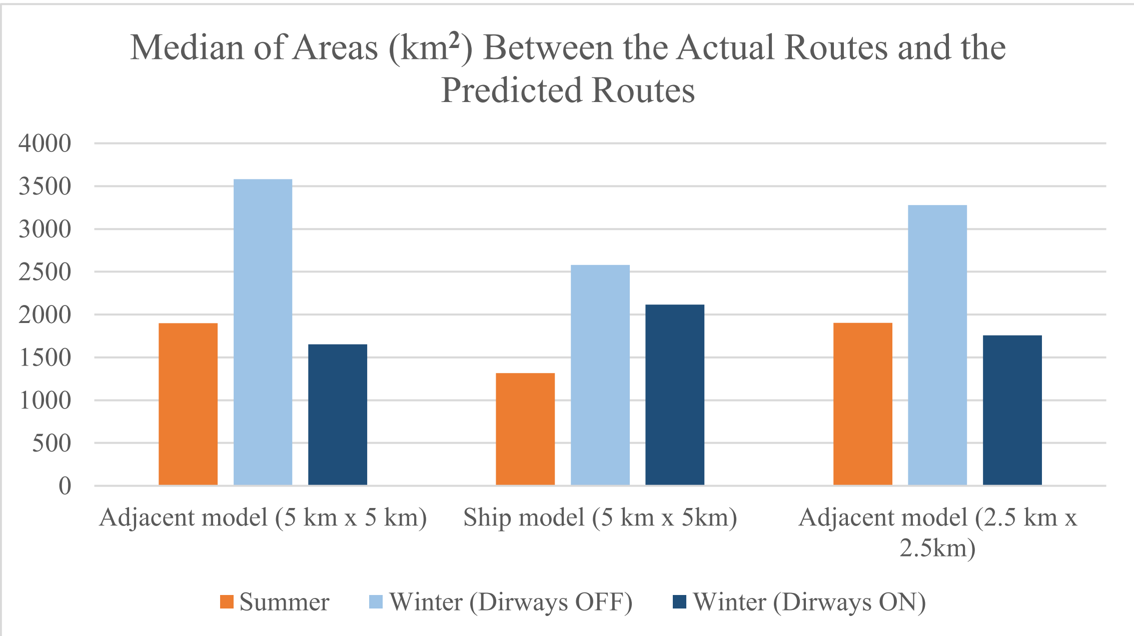


Figure 25. Route prediction accuracy.

The summer route prediction results by sea area are shown in Table 1. The routes are divided into subroutes based on which sea area a part of a route belongs. The area error references to the route prediction accuracy measure described above. Standard deviation is abbreviated as SD. The count indicates the number of routes inside a sea area.

Sea Area	Median Area Error	Mean Area Error	SD of Area Error	Count
Saimaa	13 km^2	31 km^2	43 km^2	21
Gdansk Basin	82 km^2	104 km^2	73 km^2	61
Åland Sea	180 km^2	1127 km^2	2796 km^2	121
Bothnian Bay	213 km^2	413 km^2	460 km^2	36
Gulf of Riga	227 km^2	1318 km^2	2199 km^2	33
Gulf of Finland	292 km^2	620 km^2	787 km^2	84
Arkona Basin & Bomholm Basin	380 km^2	3376 km^2	7865 km^2	113
Gotland Basin	606 km^2	2129 km^2	3916 km^2	182
Bothnian Sea & The Quark	718 km^2	1638 km^2	2658 km^2	79
Northern Baltic Proper	981 km^2	1584 km^2	1774 km^2	126

Table 1. Summer route prediction error to and from different sea areas using 5 km x 5 km tile size and the ship model.

The predicted routes inside Saimaa are the most accurate by a vast margin. This is explained most likely by the smaller variance in vessel routes as Saimaa is significantly shallower than the Baltic Sea.

The big variance in standard deviation illustrates areas where the route prediction has issues. The areas where the route prediction is the most stable are Saimaa, the Gdansk Basin and the Bothnian Bay. Winter results of the Bothnian Bay are shown in Table 2.

Model	Dirways	Median Area Error	Mean Area Error	SD of Area Error
Adjacent (2.5 km x 2.5 km)	ON	269 km^2	1094 km^2	1909 km^2
Adjacent (5 km x 5 km)	ON	367 km^2	1376 km^2	2275 km^2
Ship (5 km x 5 km)	ON	367 km^2	1303 km^2	2186 km^2
Adjacent (2.5 km x 2.5 km)	OFF	1589 km^2	2819 km^2	2960 km^2
Ship (5 km x 5 km)	OFF	1717 km^2	2856 km^2	2861 km^2
Adjacent (5 km x 5 km)	OFF	2122 km^2	3068 km^2	2884 km^2

Table 2. Winter route prediction error in the Bothnian Bay using 5 km x 5 km tile size and the ship model.

After removing the routes with area error over 5000 km^2, the Table 3 shows that the mean and median error drop drastically in every sea area. The sea areas that are least effected by the big errors are the Åland Sea, the Bornholm Basin and the Bothnian Bay. The sea area results indicate that the prediction error is not related to the voyage length as the average voyage length increased as the results improved.

The route prediction error increases steadily up to voyage length of 800 km length as seen in Figure 26. After 800 km, the rise stagnates and starts to fall.

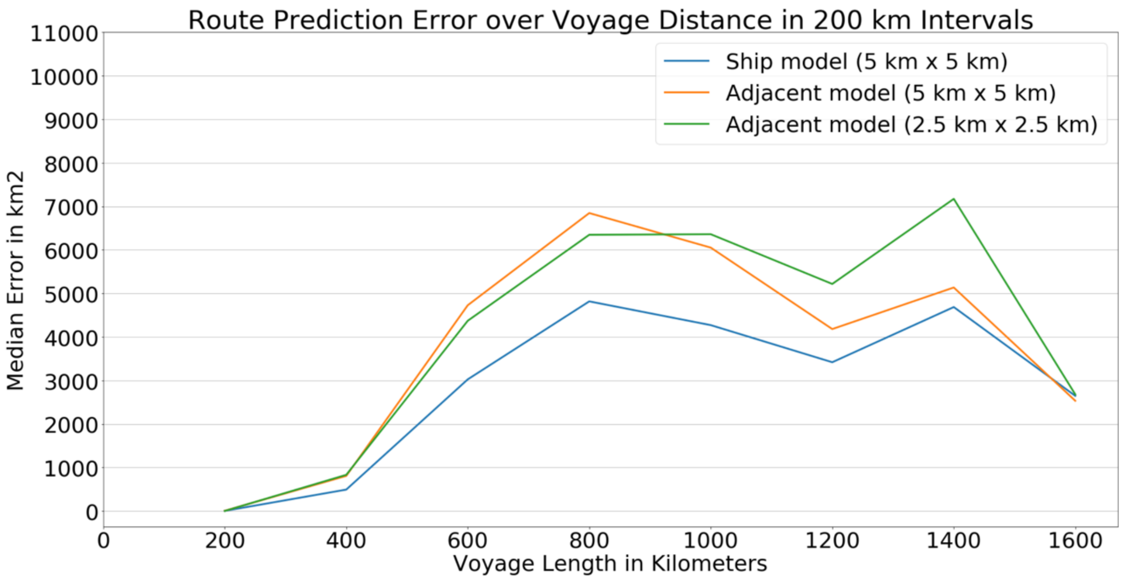


Figure 26. Median route prediction over voyage length in 200 km intervals different grid sizes and manoeuvrability models.

Route prediction accuracy was the most accurate for the vessel type 'PAS' as the Table 3 shows. It also had the least big errors. The passenger ships travel the same routes repeatedly and thus the routes could be easier to predict.

Type	Mean Area Error (km^2)	SD of Area Error (km^2)	Mean Voyage Length (km)	SD of Voyage Length (km)	Big Errors (> 5000 km^2)	% of Big Errors	Count
PAS	1992	4834	339	111	8	8,3 %	96
CONT	2903	2461	471	126	3	27,3 %	11
GC	4176	7562	524	366	35	22 %	159
T	6499	8199	684	400	16	37,2 %	43
RORO	9808	8637	946	281	10	47,6 %	21

Table 3. Prediction error by vessel types using 5 km x 5 km tile size and the ship model.

6.2 Speed Prediction Results

Speed prediction accuracy is measured by calculating the difference between the actual speed at every observation point of the actual voyage and the predicted speed on the predicted route. The mean speed prediction error is slightly smaller in summer compared to winter as seen in Figure 27.

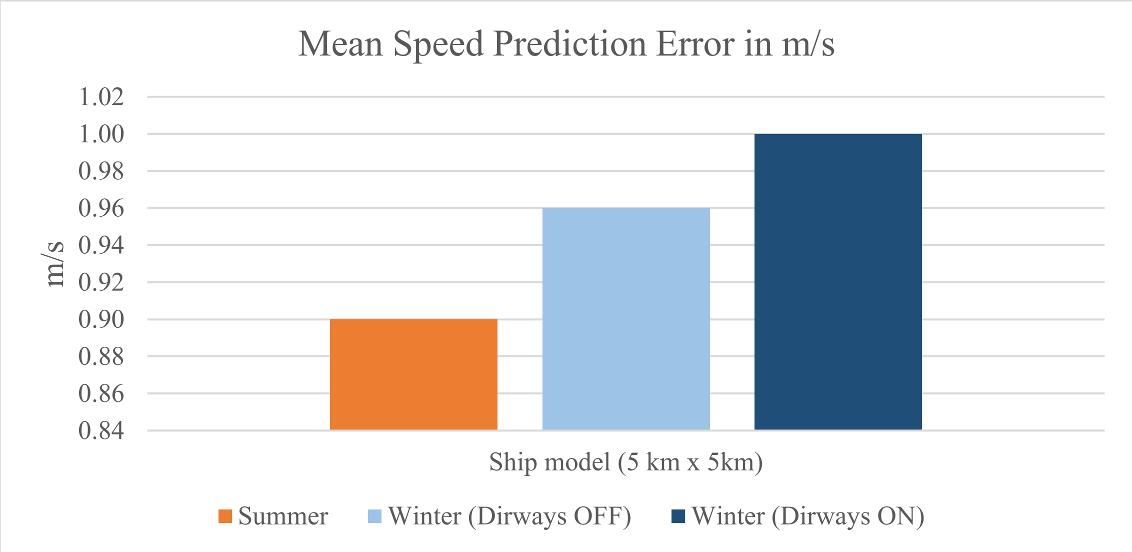


Figure 27. Mean speed prediction error during summer and winter using the 5 km x 5 km tile size and ship model.

The speed model tends to over predict the vessel speed in summer as shown by Figure 28. Even small errors in speed prediction can lead to big errors in traffic prediction with long voyages.

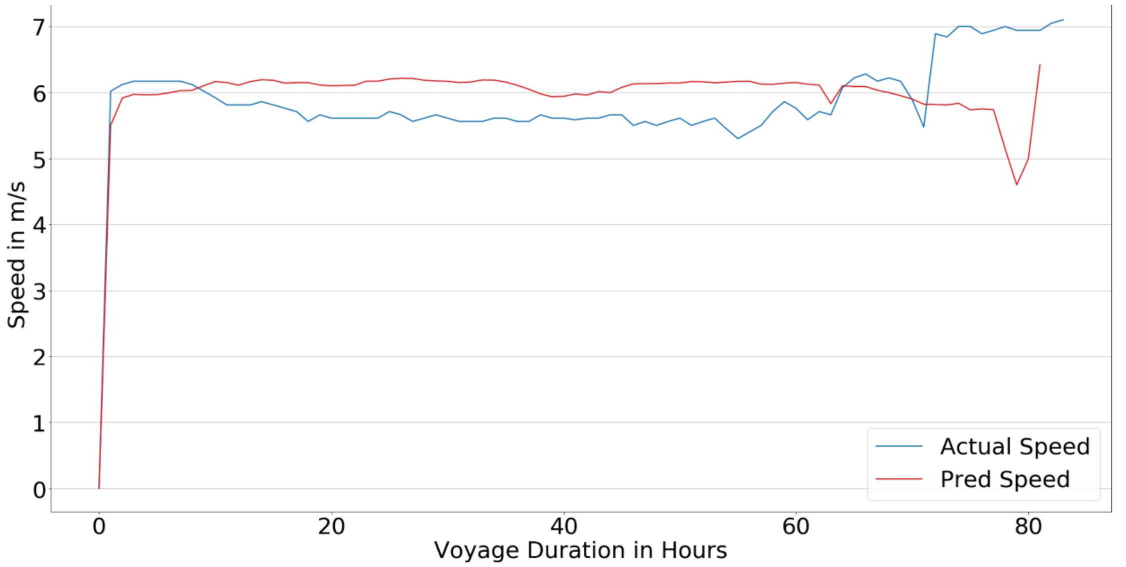


Figure 28. The mean actual speed versus the mean predicted speed over voyage duration in summer using 5 km x 5 km tile size and the ship model.

Overall, the speed prediction is more accurate in summer than in winter as seen in Figure 29. Moreover, the use of dirways does not make meaningful difference to the accuracy of speed prediction.

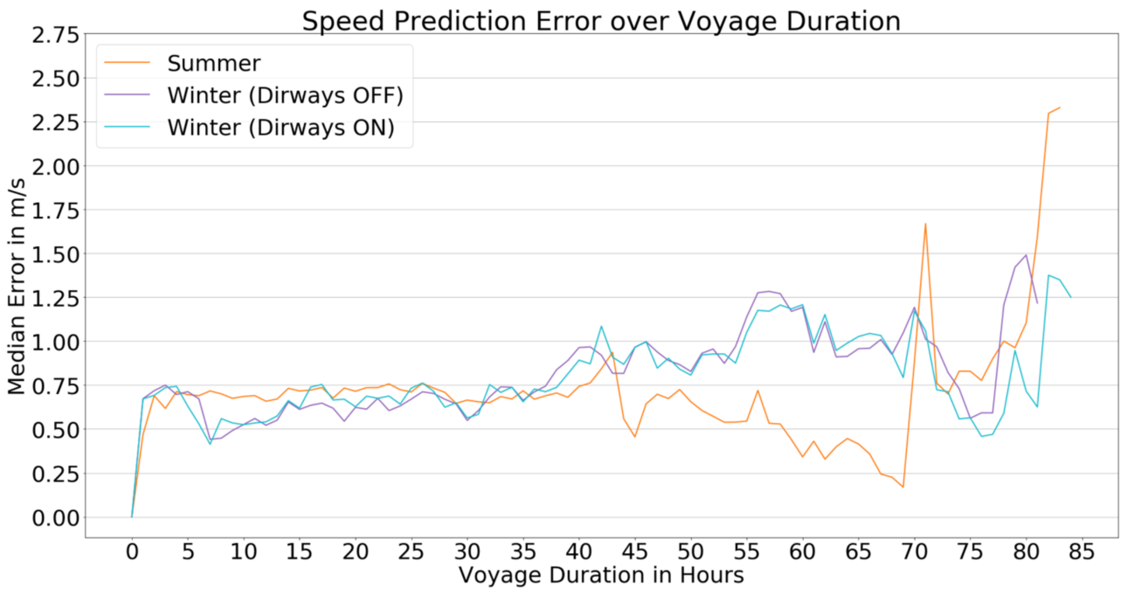


Figure 29. The median speed prediction error with summer and winter validation sets over voyage duration using 5 km x 5 km tile size and the ship model.

Speed prediction performs better with tanker and general cargo vessel types than other types as seen in Figure 30. Interestingly, passenger vessels are the most difficult for the speed model to predict even though one could assume that they tend to travel the same speed from voyage to voyage as they travel the same routes repeatedly.

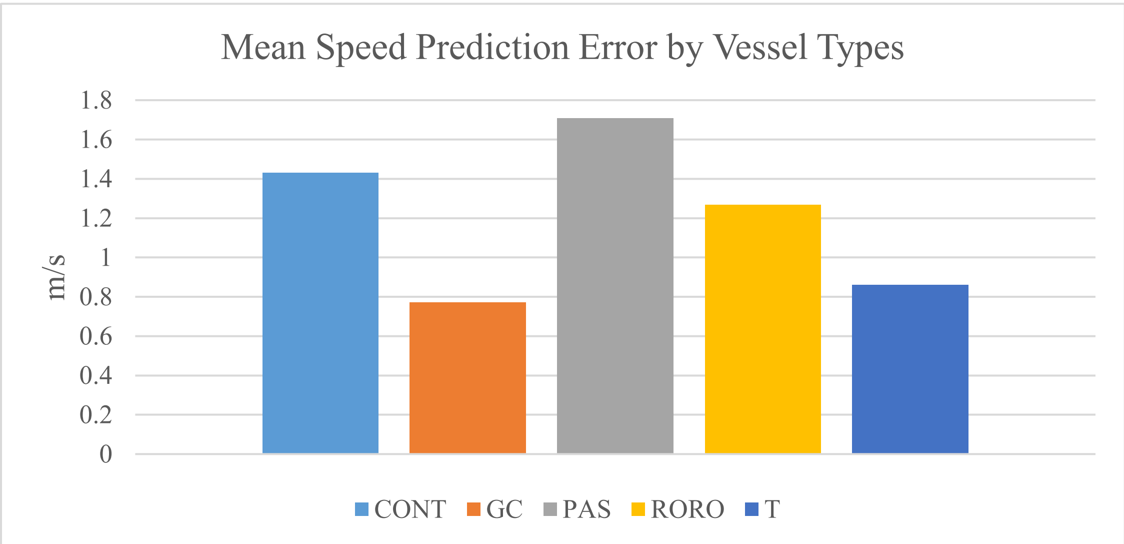


Figure 30. Speed prediction accuracy as the mean difference between the predicted speed and the actual speed. Summer, adjacent model, 5 km x 5 km tile size.

6.3 Traffic Prediction Results

Traffic prediction accuracy is measured by calculating the difference between the actual and predicted position in a point in time in nautical miles. Traffic prediction accuracy is calculated for every observation in the validation set.

The traffic model performs significantly better in summer than in winter as seen in Figure 31. The use of dirways does not significantly affect the traffic prediction accuracy, despite the improvements to the route prediction described in Section 6.1.

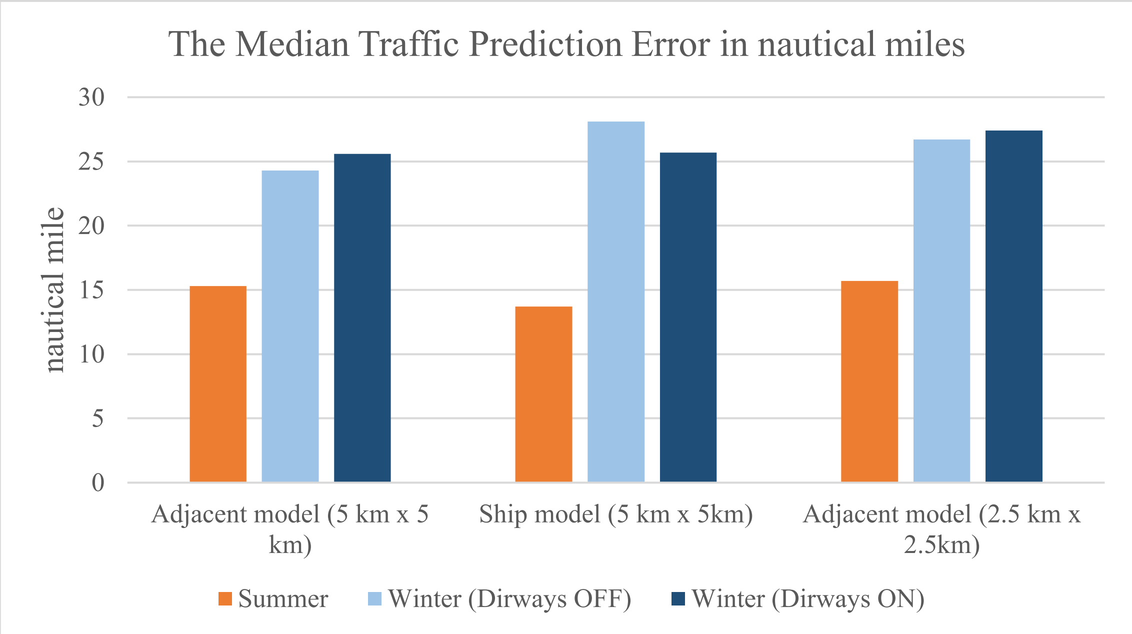


Figure 31. Median error in traffic prediction accuracy with different manoeuvrability models and validation sets.

Traffic prediction accuracy varies between sea areas. Table 4 shows the traffic prediction error by sea areas with the summer validation set. In this case, a voyage is mapped to a sea area if it starts or ends from the sea area. The high standard deviation with many of the sea areas indicate that there are voyages where the route and speed predictions are not accurate.

Sea Area	Median Error	Mean Error	SD of Error	Median Voyage Duration	Mean Voyage Duration	SD of Voyage Duration	Count
Saimaa	3 nm	4 nm	4 nm	305 min	374 min	308 min	60
Åland Sea	11 nm	18 nm	20 nm	530 min	630 min	466 min	50
Gulf of Fin-land	11 nm	19 nm	22 nm	666 min	822 min	649 min	86
Gulf of Riga	12 nm	18 nm	20 nm	798 min	876 min	595 min	33
Bothnian Sea & The Quark	12 nm	23 nm	28 nm	900 min	1036 min	758 min	56
Gotland Basin	13 nm	21 nm	22 nm	585 min	762 min	621 min	77
Northern Bal-tic Proper	15 nm	20 nm	18 nm	409 min	585 min	533 min	47
Arkona Basin & Bomholm Basin	16 nm	26 nm	29 nm	830 min	1092 min	925 min	96
Gdansk Basin	18 nm	25 nm	28 nm	621 min	815 min	711 min	66
Bothnian Bay	28 nm	38 nm	34 nm	1312 min	1534 min	1106 min	38

Table 4. Traffic prediction error by sea areas with the 5 km x 5 km tile size, ship model and summer validation set.

The erroneous voyages are evident in histogram of voyage mean error in Figure 32. The figure shows the amount of cases where traffic prediction fails because of faulty route or speed prediction.

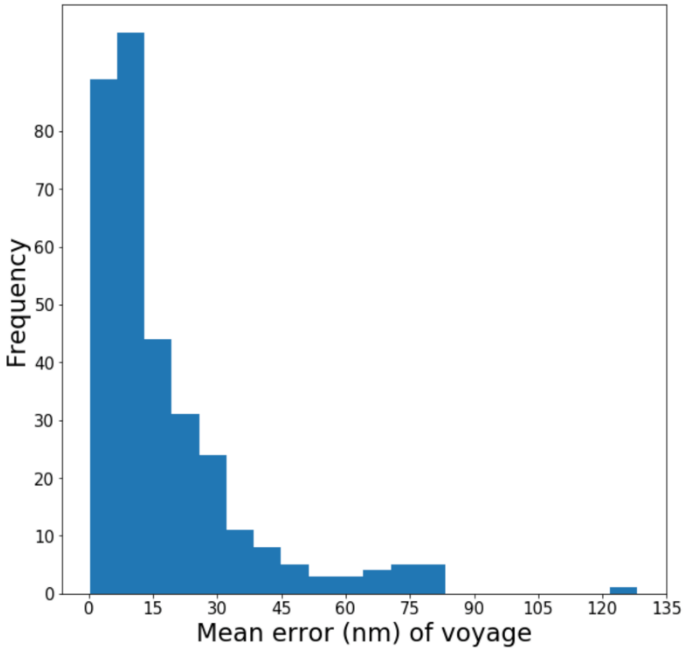


Figure 32. Mean error of voyage histogram.

The mean prediction error over prediction time is visualized in Figure 33. There is no notable difference between the summer and winter validation sets with or without dirways. As the Table 4 and Figure 32 suggested, the prediction accuracy increases significantly when voyages with route prediction mean error over 5000 km[2] are removed from the summer validation set.

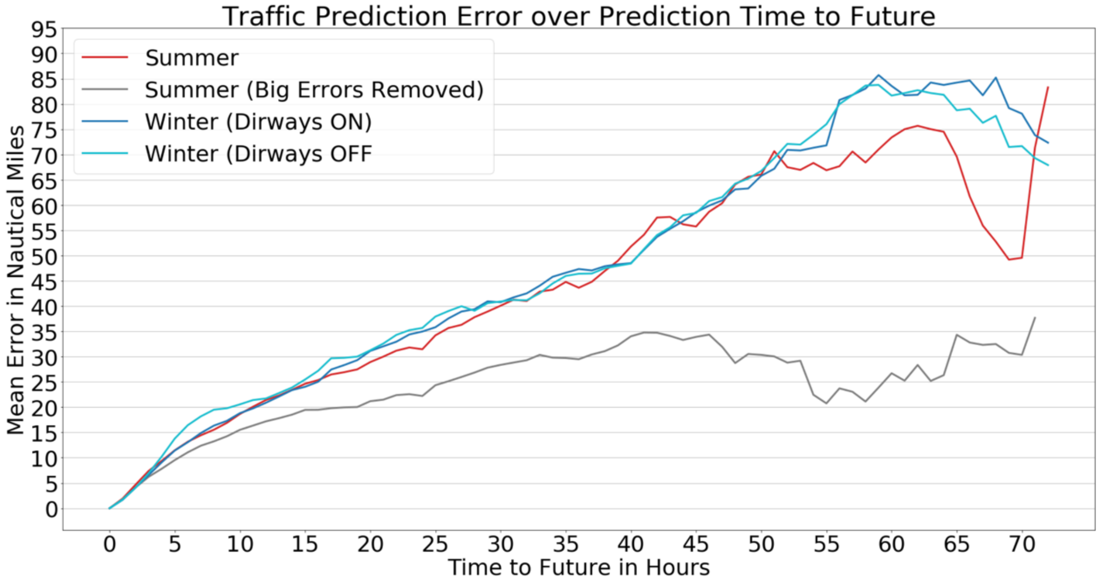


Figure 33. Traffic prediction error comparison over prediction time with 5 km x 5 km tile size. Prediction done using the ship model and 5 km x 5 km grid size.

6.4 Estimated Time of Arrival Prediction Results

Estimated time of arrival accuracy is measured by calculating the difference between the actual and predicted arrival time in hours. Overall, the model over predicted the estimated arrival time in summer with ship model and 5 km x 5 km tile size in 74 percent of voyages. The model predicted ETAs in summer more accurately than in winter as evident in Table 5. However, the summer voyages were significantly shorter than the winter voyages.

Model	Median Diff	Mean Diff	SD of Diff	Median Voyage Duration	Mean Voyage Duration	SD of Voyage Duration	Count
Summer (ship)	1,7 h	2,9 h	3,7 h	18,6 h	23,7 h	16,9 h	330
Winter (Dirways OFF, ship)	6,4 h	4,6 h	5,6 h	56,7 h	48,9 h	24,1 h	80
Winter (Dirways ON, ship)	6,3 h	4,3 h	5,4 h	56,7 h	48,9 h	24,1 h	80

Table 5. ETA prediction error results with summer and winter validation sets.

The median error rises steadily with the duration of a voyage up to duration of 40 hours as seen in Figure 34. After 40 hours, the variance in error increases considerably.

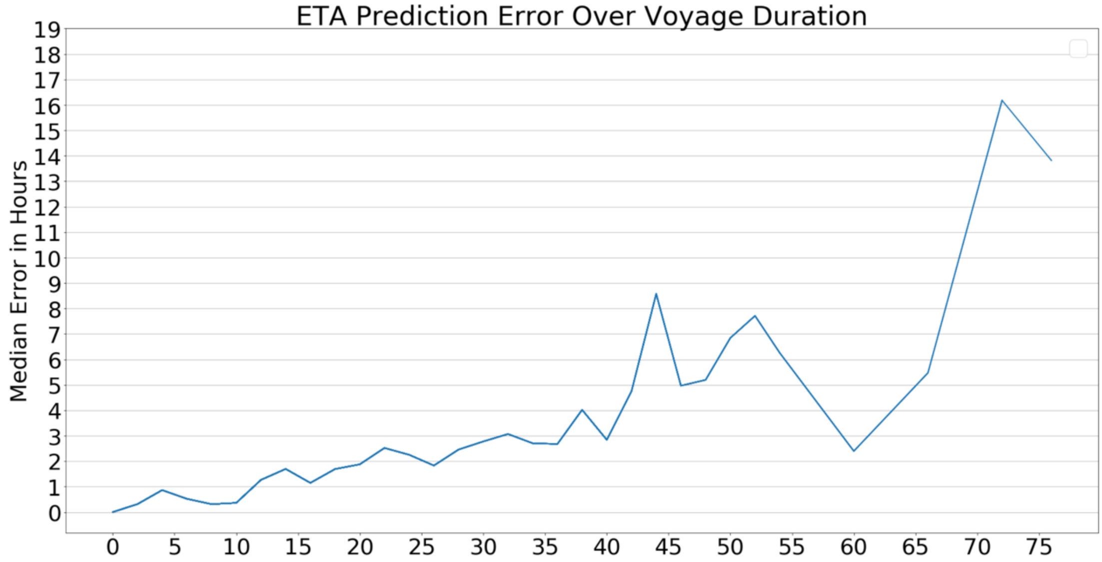


Figure 34. Median ETA prediction error over time averaged for all the voyages in the summer validation set. Prediction done using the ship model and 5 km x 5 km grid size.

The performance of the proposed model is considerably worse than the benchmark study as can be seen in Figure 35. However, it is important to note that the vessel tracks in the Alessandrini *et al.* (2018) results are heading to only one port whereas the vessel tracks in the summer validation set are to 72 different ports and in the winter validation set to 25 ports. In addition, the vessel tracks are from different seas and thus the vessel movement patters between the two seas can vary.

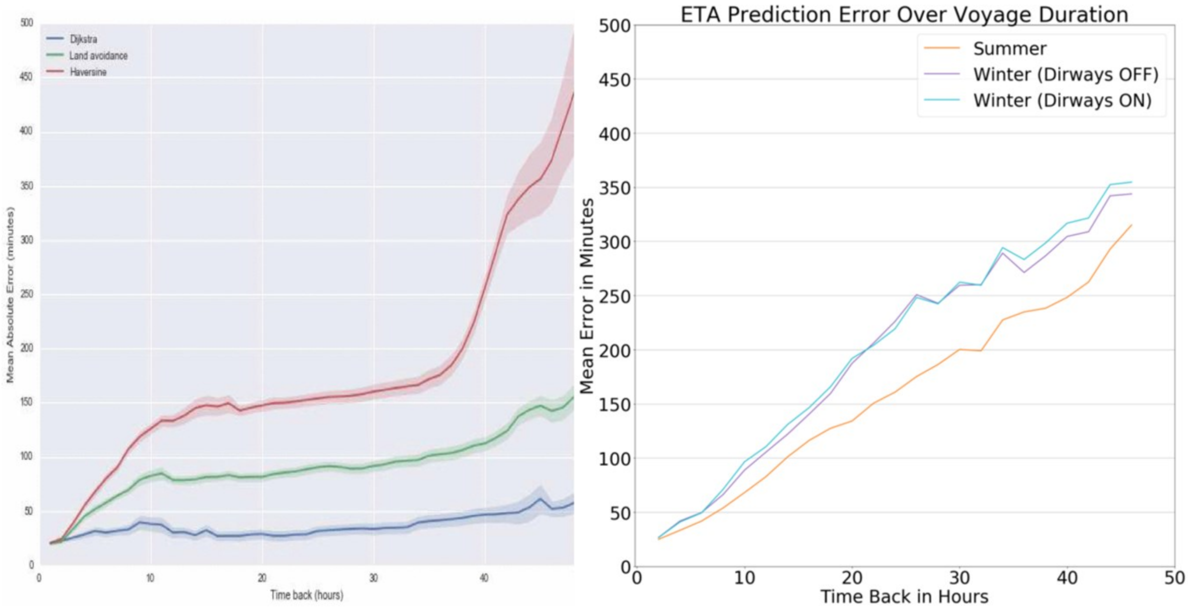


Figure 35. Comparison of ETA prediction error averaged over all the vessel tracks entering the port of Trieste (Alessandrini et al. 2018) and over all the vessel tracks in the validation set using the method described in this thesis.

The prediction error varies between voyages ending to different sea areas as shown in Table 6. These results indicate again that the performance is not only tied to voyage length or duration but also how complex the traffic is in the sea areas which the voyage passes.

Sea Area	Median Diff	Mean Diff	SD of Diff	Median Voyage Duration	Mean Voyage Duration	SD of Voyage Duration	Count
Saimaa	0,54 h	0,9 h	0,97 h	4,33 h	6,77 h	6,22 h	30
Gulf of Finland	0,91 h	2,02 h	2,66 h	19,91 h	22,04 h	11,43 h	56
Arkona Basin & Bomholm Basin	1,14 h	3,03 h	5,04 h	14,78 h	21,44 h	17,63 h	66
Northern Baltic Proper	1,46 h	1,45 h	1,27 h	13,09 h	15,39 h	10,42 h	17
Gotland Basin	1,52 h	3,1 h	4,24 h	17,99 h	20,13 h	10,91 h	47
Åland Sea	1,73 h	2,52 h	2,56 h	17,84 h	20,53 h	9,65 h	20
Gdansk Basin	1,93 h	2,98 h	3,98 h	17,54 h	20,16 h	11,2 h	36
Gulf of Riga	2,69 h	3,4 h	2,58 h	28,9 h	29,08 h	9,7 h	3
Bothnian Sea & The Quark	3,04 h	3,86 h	4,36 h	30,05 h	33,38 h	16,15 h	26
Bothnian Bay	3,45 h	5,61 h	5,68 h	50,98 h	48,56 h	23,57 h	8

Table 6. ETA prediction error by sea areas with the 5 km x 5 km tile size, ship model and summer validation set.

6.5 Discussion

The winter results are only indicative of the model's performance in ice as the ice-coverage during the selected winter was relatively small. There were no dirways present south of the Bay of Bothnia. A potential problem in harsh winters is that as the dirways do not contain information to which ports they are assigned, the model could overfit the route to follow dirways leading to different ports along the route. This problem could be resolved by using only the dirways that the vessel is directed through. However, this information is not publicly available although it is present in IBNet.

The accuracy improvements from using the ship model are most likely result of two things. First, the ship model simulates the movement patterns of a ship more accurately than the adjacent model. Second, the ship model produces routes with less turns than the adjacent model. The adjacent model routes could be improved by applying a smoothing algorithm to the routes. It would be interesting to see how much of the accuracy difference comes from the jaggedness of the adjacent model routes.

The estimated time of arrival prediction could already be useful in the planning of ice-breaking activities. One potential use-case for the ETA prediction would be to use it to predict arrival times to the edge of an icefield. This could give icebreaker captains and marine authorities valuable information about how many vessels are expected to be near an edge of an icefield at any time point in time.

The effect of ice-class to the modelling accuracy was not examined due to the validation set being skewed in terms of ice-class. The validation set contained 71 vessels with ice-class 1A and 9 vessels with ice-class 1A Super. As the 1A Super vessels can operate in ice without icebreaking assistance most of the time, it would be interesting to see if the use of dirways would improve the route prediction accuracy.

The time complexity optimization of the proposed method was not in focus in this thesis. This is evident in the CPU run times shown in Table 7. The time complexity could be vastly reduced by using for example Pruned Landmark Labeling instead of A* (Akiba *et al.* 2013).

Manoeuvrability Model	**Tile Size**	**Run time**	**Count**
Adjacent	5 km x 5 km	11 min	300
Ship	5 km x 5 km	1h 38 min	300
Adjacent	2.5 km x 2.5 km	1h 23 min	300
Ship	2.5 km x 2.5 km	3h 49 min	300

Table 7. The CPU run times of route prediction algorithm with summer validation set and different manoeuvrability models and tile sizes.

7 CONCLUSIONS

The domain of marine traffic modelling is complex, and the prior research is still in its infancy in many ways. Vast number of factors affect the route and speed of a vessel and all are not yet clearly established. Ice coverage in winter, numerous archipelagos and shallow water areas complicate the traffic modelling in the Baltic Sea compared to other seas.

The domain suffers from a lack of standardised public datasets and the models are often not evaluated thoroughly. This thesis addressed these issues by publishing the dataset and source code used in this thesis (Hakola 2020a, Hakola 2020b). In addition, a novel measure for route prediction accuracy was introduced and the model was evaluated thoroughly.

The main target of this thesis was to assess the viability of year-round modelling of marine traffic in the Baltic Sea and the results are promising. The results indicate that the use of dirways improves the route prediction accuracy in ice-covered waters. Moreover, the ship model improved both the route and the traffic prediction.

The performance of the model varied between different sea areas significantly which indicates underlying problems in the route and speed modelling. These issues should be addressed in future research for example by using a more comprehensive training set, dynamically sized graph or improving the cost function. In addition, the time complexity of the method is an issue that should be resolved for real-time predictions.

As the long-term marine traffic prediction is not a well-studied issue, these results act as reference point for future research. Several directions for future research are available: validation of dirway model with AIS data from several winters, integration of ice forecast data into the model for more accurate vessel movement modelling and validation of the results against different shortest path and neural network approaches.

Besides the Baltic Sea, the results could be utilized in the traffic modelling of the Arctic Ocean as it has substantial amount of icebreaking activity and numerous archipelagos and shallow water areas. Moreover, if the rise in Earth’s surface temperatures continues, the Arctic Ocean could soon host multiple major shipping routes where ice coverage varies considerably between summer and winter. For maritime modelling research, data availability will be a crucial issue in the Arctic Ocean as the operating icebreakers are from multiple countries.

Beyond the ice-covered seas, year-round traffic modelling is within reach. The grid-based methods have been shown to capture traffic flows accurately when ice is not present and the resolution requirements are low. Furthermore, the flexibility of grid-based methods might prove beneficial if specialities of different seas surface in the modelling process.

Finally, marine traffic modelling holds tremendous potential for optimising maritime supply chains if integrated with other parts of the logistics chain. To accomplishing this, vast amount of new research and big advances in data availability are required. To make traffic modelling useful for practice, collaboration over academic, country and organisation borders is needed.

REFERENCES

Ravindra K. Ahuja, Kurt Mehlhorn, James Orlin and Robert E. Tarjan. 1990. Faster algorithms for the shortest path problem. *Journal of the ACM (JACM)* 37, 2, 213-223.

Aishub. 2020. A free AIS data sharing service. https://www.aishub.net/. Read 12.09.2019.

Takuya Akiba, Yoichi Iwata and Yuichi Yoshida. 2013. Fast exact shortest-path distance queries on large networks by pruned landmark labeling. In: *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, 349-360.

Alfredo Alessandrini, Fabio Mazzarella and Michele Vespe. 2018. Estimated time of arrival using historical vessel tracking data. *IEEE Transactions on Intelligent Transportation Systems*.

Arctia. 2018. Jäänmurtokausi 2017-2018 päättyi. http://arctia.fi/2018/05/21/jaanmurtokausi-2017-2018-paattyi/. Read 20.2.2019.

Avoindata. 2019. S-57 aineisto. https://www.avoindata.fi/data/fi/dataset/s-57-aineisto. Read 10.4.2019.

Paweł Banyś, Thoralf Noack and Stefan Gewies. 2012. Assessment of AIS vessel position report under the aspect of data reliability. *Annual of Navigation* 19, 1, 5-16.

Richard Bellman. 1954. The theory of dynamic programming. *Bulletin of the American Mathematical Society* 60, 6, 503-515.

Robin Berglund, Renne Tergujeff and Teppo Veijonen. 2014. IBNext: Future needs and development of the icebreaker information system: A prestudy. *Winter Navigation Research Reports*, 85.

Michael J. Berry and Gordon S. Linoff. 2004. *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.

BIM. 2018. *Baltic Sea Icebreaking Report 2017-2018*. Baltic Icebreaking Management, http://baltice.org/app/static/pdf/BIM%20Report%2017-18.pdf. Read 12.8.2019.

Oleh Bodunov, Florian Schmidt, André Martin, Andrey Brito and Christof Fetzer. 2018. Real-time destination and eta prediction for maritime traffic. In: *Proceedings of the 12th ACM International Conference on Distributed and Event-based Systems*, 198-201.

Zvi Boger and Hugo Guterman. 1997. Knowledge extraction from artificial neural network models. In: *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, 4, 3030-3035.

Danah Boyd and Kate Crawford. 2011. Six provocations for big data. In: *A decade in internet time: Symposium on the dynamics of the internet and society*, 21.

Axel Bruns. 2013. Faster than the speed of print: Reconciling ‘big data’ social media analysis and academic scholarship. *First Monday* 18, 10.

Ismail Chabini and Shan Lan. 2002. Adaptations of the A* algorithm for the computation of fastest paths in deterministic discrete-time dynamic networks. *IEEE Transactions on intelligent transportation systems.* 3, 1, 60-74.

Minjoo Choi, Hyun Chung, Hajime Yamaguchi and Keisuke Nagakawa. 2015. Arctic sea route path planning based on an uncertain ice prediction model. *Cold Regions Science and Technology* 109, 61-69.

Andrius Daranda. 2016. Neural network approach to predict marine traffic. *Transaction in Baltic Journal of Modern Computing* 4, 3, 483.

Digitraffic. 2020. Open data from Finnish waterways. https://www.digitraffic.fi/en/marine-traffic/. Read 10.11.2019.

ESRI. 1998. ESRI Shapefile Technical Description. https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf. Read 10.4.2019.

Finnish Meteorological Institute. 2020a. Seas. https://en.ilmatieteenlaitos.fi/seas. Read 09.09.2019.

Finnish Meteorological Institute. 2020b. Itämeren jäätalvet. https://ilmatieteenlaitos.fi/jaatalvet. Read 20.2.2019.

Liping Fu, D. Sun and Laurence R. Rilett. 2006. Heuristic shortest path algorithms for transportation applications: State of the art. *Comput.Oper.Res.* 33, 11, 3324-3343.

GDAL. 2019. IHO S-57 (ENC). https://www.gdal.org/drv_s57.html. Read 10.4.2019.

Floris Goerlandt, Habtamnesh Goite, Osiris A. V. Banda, Anders Höglund, Paula Ahonen-Rainio and Mikko Lensu. 2017. An analysis of wintertime navigational accidents in the northern Baltic Sea. *Safety Science* 92, 66-84.

George Gougoulidis. 2008. The utilization of artificial neural networks in marine applications: *An overview. Naval Engineers Journal* 120, 3, 19-26.

Robert E. Guinness, Jarno Saarimaki, Laura Ruotsalainen, Heidi Kuusniemi, Floris Goerlandt, Jakub Montewka, Robin Berglund and Ville Kotovirta. 2014. A method for ice-aware maritime route optimization. *IEEE/ION PLANS* 2014, 1371-1378.

Ville Hakola. 2020a. Vessel tracking (AIS), vessel metadata and dirway datasets. *IEEE Dataport*. http://dx.doi.org/10.21227/j3b5-es69.

Ville Hakola. 2020b. Modelling marine traffic in the ice-covered Baltic Sea. https://github.com/hakola/marine-traffic-modelling. Read 26.02.2020.

Abbas Harati-Mokhtari, Alan Wall, Philip Brooks and Jin Wang. 2007. Automatic identification system (AIS): Data reliability and human error implications. *The Journal of Navigation* 60, 3, 373-389.

Peter E. Hart, Nils J. Nilsson and Bertram Raphael. 1968. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics* 4, 2, 100-107.

Simon Haykin. 2004. Neural networks: A comprehensive foundation. *Neural Networks 2*, 2004, 41.

Helcom. 2020. Physical description of the Baltic Sea. http://stateofthebalticsea.helcom.fi/in-brief/our-baltic-sea/. Read 20.12.2019.

Guang-Bin Huang and Haroon A. Babri. 1998. Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions. *IEEE Transactions on Neural Networks* 9, 1, 224-229.

International Maritime Organization. 2001. *Guidelines for the Onboard Operational Use of Shipborne Automatic Identification System (AIS).* http://www.imo.org/en/KnowledgeCentre/IndexofIMOResolutions/Assembly/Documents/A.917(22).pdf. Read 26.03.2020.

International Maritime Organization. 2003. *Guidelines for the Installation of a Shipborne Automatic Identification System (AIS)*. http://www.imo.org/en/OurWork/Safety/Navigation/Documents/227.pdf. Read 13.05.2019.

Jupyter. 2020. Open-source project to support interactive data science and scientific computing. https://jupyter.org/. Read 24.2.2020.

Iebeling Kaastra and Milton Boyd. 1996. Designing a neural network for forecasting financial and economic time series. *Neurocomputing* 10, 3, 215-236.

Kwang-Il Kim and Keon Lee. 2018. Deep learning-based caution area traffic prediction with automatic identification system sensor data. *Sensors* 18, 9, 3172.

Ville Kotovirta, Risto Jalonen, Lars Axell, Kaj Riska and Robin Berglund. 2009. A system for route optimization in ice-covered waters. *Cold Regions Science and Technology* 55, 1, 52-62.

K. K. Lai and Katharine Shih. 1992. A study of container berth allocation. *Journal of Advanced Transportation* 26, 1, 45-60.

Ulrich Lauther. 2004. An extremely fast, exact algorithm for finding shortest paths in static networks with geographical background. *Geoinformation und Mobilität-von der Forschung zur praktischen Anwendung* 22, 219-230.

Ville Lehtola, Jakub Montewka, Floris Goerlandt, Robert Guinness and Mikko Lensu. 2019. Finding safe and efficient shipping routes in ice-covered waters: A framework and a model. *Cold Regions Science and Technology* 165, 102795.

Mikko Lensu and Floris Goerlandt. 2019. Big maritime data for the Baltic Sea with a focus on the winter navigation system. *Marine Policy* 104, 53-65.

Élise Lépy. 2013. The recent history of Finnish winter navigation in the Baltic Sea. *Polar Record* 49, 1, 33-41.

Ulrike Löptien and Lars Axell. 2014. Ice and AIS: Ship speed data and sea ice forecasts in the Baltic Sea. *The Cryosphere* 8, 6, 2409-2418.

Katrin Schroeder. 2019. Current systems in the Mediterranean Sea. *Encyclopedia of Ocean Sciences (Third Edition)* 219-227.

Shangbo Mao, Enmei Tu, Guanghao Zhang, Lily Rachmawati, Eshan Rajabally and Guang-Bin Huang. 2018. An automatic identification system (AIS) database for maritime trajectory prediction and data mining. In: *Proceedings of ELM-2016.* Springer, 241-257.

Jakub Montewka, Robert Guinness, Lauri Kuuliala, Floris Goerlandt, P. Kujala and Mikko Lensu. 2017. Challenges in modelling characteristics of maritime traffic in winter conditions and new solution proposal. In: *Developments in Maritime Transportation and Exploitation of Sea Resources-Proceedings of IMAM*, 247-256.

Ewa Osekowska, Henric Johnson and Bengt Carlsson. 2014. Grid size optimization for potential field based maritime anomaly detection. *Transportation Research Procedia* 3, 720-729.

Giuliana Pallotta, Michele Vespe and Karna Bryan. 2013. Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. *Entropy* 15, 6, 2218-2245.

Lokukaluge P. Perera, Paulo Oliveira and C. G. Soares. 2012. Maritime traffic monitoring based on vessel detection, tracking, state estimation, and trajectory prediction. *IEEE Transactions on Intelligent Transportation Systems* 13, 3, 1188-1200.

Michael E. Porter and James E. Heppelmann. 2014. How smart, connected products are transforming competition. *Harvard Business Review* 92, 11, 64–88.

Maria Riveiro, Fredrik Johansson, Göran Falkman and Tom Ziemke. 2008. Supporting maritime situation awareness using self organizing maps and gaussian mixture models. *Frontiers in Artificial Intelligence and Applications* 173, 84.

Marinetraffic. 2020. API for vessel positions in a predefined area. https://www.marinetraffic.com/fi/ais-api-services/detail/ps05/vessel-positions-in-a-predefined-area. Read 04.05.2019.

RMRS. 2019. Rules for the Classification and Construction of Sea-Going Ships. *Russian Maritime Register of Shipping, Saint Petersburg*.

Lars Schylberg 2019. SMAC-M - Scripts for Map and Chart Manager. https://github.com/LarsSchy/SMAC-M. Read 10.05.2019.

Yu Chang Seong, Jung Sik Jeong and Gyei-Kark Park. 2011. The Relation with Width of Fairway and Marine Traffic Flow. *Transport Systems and Processes: Marine Navigation and Safety of Sea Transportation.*

K. G. Sheela and Subramaniam N. Deepa. 2013. Review on methods to fix number of hidden neurons in neural networks. *Mathematical Problems in Engineering* 2013.

Roger W. Sinnott. 1984. Virtues of the haversine. *Sky and Telescope* 68, 159.

O. V. Sormunen, R. Berglund, M. Lensu, L. Kuuliala, F. Li, M. Bergström and P. Kujala. 2018. Comparison of vessel theoretical ice speeds against AIS data in the Baltic Sea. *Marine Design XIII* 2, 841-849.

Daniel Svozil, Vladimír Kvasnicka and Jiȋí Pospichal. 1997. Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems* 39, 1, 43-62.

Oleg V. Tarovik, Alex Topaj, Andrey A. Bakharev, Andrey V. Kosorotov, Andrey B. Krestyantsev and Aleksander A. Kondratenko. 2017. Multidisciplinary approach to design and analysis of arctic marine transport systems. *Volume 8: Polar and Arctic Sciences and Technology; Petroleum Technology*.

Jarkko Toivola. 2016. Winter navigation and Icebreaking services, Baltic countries cooperation Urban Node 2016. https://vayla.fi/documents/20473/205877/Jarkko+Toivola.pdf/a1944554-cfba-4127-ac2f-17fb4051907b. Read 05.03.2019

A. G. Topaj, O. V. Tarovik, A. A. Bakharev and A. A. Kondratenko. 2019. Optimal ice routing of a ship with icebreaker assistance. *Applied Ocean Research* 86, 177-187.

Enmei Tu, Guanghao Zhang, Lily Rachmawati, Eshan Rajabally and Guang-Bin Huang. 2018. Exploiting AIS data for intelligent maritime navigation: A comprehensive survey from data to methodology. *IEEE Transactions on Intelligent Transportation Systems* 19, 5, 1559-1582.

Michio Ueno, Yasuo Yoshimura, Yoshiaki Tsukada and Hideki Miyazaki. 2009. Circular motion tests and uncertainty analysis for ship maneuverability. *Journal of Marine Science Technology* 14, 4, 469.

UNCTAD. 2018. Review of maritime transport. United Nations Conference on Trade and Development. https://unctad.org/en/PublicationsLibrary/rmt2018_en.pdf. Read 08.08.2019.

Wouter G. Van Toll, Atlas F. Cook IV and Roland Geraerts. 2012. Real-time density-based crowd simulation. *Computer Animation and Virtual Worlds* 23, 1, 59-69.

Konstantinos Vassakis, Emmanuel Petrakis and Ioannis Kopanakis. 2018. Big data analytics: Applications, prospects and challenges. In: *Mobile Big Data*. Springer, 3-20.

Gerben de Vries and Maarten van Someren. 2009. Unsupervised ship trajectory modeling and prediction using compression and clustering. In: *18th Annual Belgian-Dutch Conference on Machine Learning,* 7-12.

Massimiliano Zanin, David Papo, Pedro A. Sousa, Ernestina Menasalvas, Andrea Nicchi, Elaine Kubik and Stefano Boccaletti. 2016. Combining complex networks and data mining: Why and how. *Physics Reports* 635, 1-44.

F. B. Zhan and Charles E. Noon. 1998. Shortest path algorithms: An evaluation using real road networks. *Transportation Science* 32, 1, 65-73.

Xianming Zhu, Hongbo Wang, Zihao Shen and Hongjun Lv. 2016. Ship weather routing based on modified dijkstra algorithm. In: *2016 6th International Conference on Machinery, Materials, Environment, Biotechnology and Computer*.

Dimitrios Zissis, Elias K. Xidias and Dimitrios Lekkas. 2015. A cloud based architecture capable of perceiving and predicting multiple vessel behaviour. *Applied Soft Computing* 35, 652-661.

Dimitrios Zissis, Elias K. Xidias and Dimitrios Lekkas. 2016. Real-time vessel behavior prediction. *Evolving Systems* 7, 1, 29-40.