2019

Toward Using High-frequency Coastal Radars for Calibration of S-AIS Based Ocean Vessel Tracking Models

Ben Freidrich
frie0450@mylaurier.ca

Toward Using High-frequency Coastal Radars for Calibration of S-AIS Based Ocean Vessel Tracking Models

by

Benjamin K. Friedrich

Master of Science, Wilfrid Laurier University, 2019

Thesis

Submitted to the Department of Geography and Environmental Studies

in partial fulfilment of the requirements

for the Masters of Science in Geography

Contents

Acknowledgements

I want to give thanks to the ones who helped me get through this thesis till the end. First and most off I want to thank Dr. Steven Roberts for all of his help from start to finish. He spent countless hours reading over revisions, helping debug codes, providing resources of information, and showed tremendous patience with the writing process where deadlines kept extending. Thanks to my co-adviser Dr. Colin Robertson for reading over and providing feedback on the thesis and helping prepare for the defence. Thank you to Dr. Alfaro and Dr. Venkiteswaran for being readers for my defence. Special thanks to the Laurier Spatial Lab who were of crucial help in overall well being when struggling at different parts within the thesis. Finally thanks to my friends and family who were supportive over the course of this entire endeavour.

Abstract

Most of the world relies on ships for transportation, shipping, and tourism. Automatic Identification System messages are transmitted from ships and provide a wealth of positional data on these open ocean vessels. This data is being utilized to determine the optimal path for ships, as well as predicting where a ship may be going in the near future. It has only been in the past decade that Automatic Identification Systems (AIS) signals have been easily received with satellites (S-AIS) so there have been few studies that look at using available information and pairing it with the new abundance of ship positional data. This study attempts to use High Frequency (HF) radar data that measures the velocity of surface ocean currents off the West Coast of North America and incorporates North Pacific Automatic Identification Systems data to create a basic prediction model that uses the radar data to refine the positional accuracy of the prediction. Determining the effects of ocean currents on a ship using these data sets allows for later calibration of currently available position prediction models using high frequency radar data. While the study was unable to obtain consistent prediction correlation results the work systematically analyzes inconstancy and limitations of existing S-AIS and HF radar data that is a valuable contribution to the field.

Chapter 1

Introduction

Ocean ships have been one of the main modes of transportation for thousands of years. Today seaborne shipping accounts for 90% of global trade and with an increasing demand for goods the amount of required shipping continually is on the rise. “There are over 50,000 merchant ships trading internationally, transporting every kind of cargo” (International Chamber of Shipping, 2015). More ships on the ocean may lead to increasing potential of problems and risks. More ship based traffic increases probability of ship collisions occurring. It also increases the potential risk of a ship drifting off course due to engine failure or other catastrophic failures and require assistance from a Search and Rescue team. There is a growing issue with how much pollution is created from large ships, a 2009 Daily Mail article stated that “16 of the world’s largest ships can produce as much lung-clogging sulphur pollution as all the world's cars" (Pearce, 2009). With a continuing increase in ship numbers it becomes paramount to track them for a variety of reasons: ship collision avoidance (Silveira et al., 2013), search and rescue missions (Balduzzi et al., 2014), analyzing more fuel efficient paths to lower pollution output (Carson-Jackson, 2012), tracking of piracy and illegal behaviour (Vespe et al., 2015), and illegal fishing activities (Natale et al., 2015) . Technologies available today are capable of tracking ocean vessels using satellite technology.

The advancement and cost effectiveness of Low Earth Orbiting (LEO) satellites has led to an abundance of maritime shipping data (Ristic et al., 2008). LEO satellites are now able to collect data transmitted from Automatic Identification Systems (AIS) which are placed on all: passenger ships that have been constructed since 2002; ships with weight greater than 3,000 tonnes since 2002; tanker ships with international voyages; and ships with gross tonnage over 300 on international voyages (International Maritime Organization, 2016). AIS is a device on a ship that continuously reports information about the ship such as identity and velocity. AIS were implemented on ships starting in 2002 for safety on busy waterways. AIS allows ships to communicate information on their size and velocity to other ships in their vicinity. AIS broadcast at different rates depending on a ship's movements, if a ship is travelling at high speeds or is changing course the AIS transmits a message every 2 seconds. If a ship is at anchor the AIS transmits every 2 min. Static information is transmitted from AIS every 5 min, which provides information on the destination of the ship, Identification Number of the ship, and the dimensions of the ship (International Maritime Organization, 2016). LEO satellites have only just recently taken advantage of AIS and can almost continuously track ships on international voyages and store the information on the locations over time. The AIS messages that are received by a LEO satellite is considered in this thesis to be a satellite based automatic identification system (S-AIS) messages. LEO satellites tend to be smaller and "may be included as additional payload on previously planned launches [that allow the] dual benefit of reducing the launch planning time considerably and, as they can be included as a part of a larger launch, reducing launch costs" (Carson-Jackson, 2012) .This allows more satellites to be launched for AIS detection, which increases the volume of available data. This information has been used to not only track ships but to predict patterns of movement to determine a ship's future position. Since S-AIS data has only recently become available there have been a lack of studies that attempt to enhance the capabilities of S-AIS based systems.

An issue that is present with AIS and S-AIS messages is data reliability. An AIS data mining study (Harati-Mokhtari et al., 2007b) found that 8% of the AIS messages contained an error in the reported information. Issues arise in system design since manual input is required for some AIS,“’30% of ships were detected as displaying incorrect [navigational] status information”(Harati-Mokhtari et al., 2007a) that can possibly be attributed to human error. Tracking ships using S-AIS messages has potential issues with data gaps that occur when a message is no longer being sent by an AIS, poor satellite coverage causing no data to be received, or when an AIS is turned off. These gaps in S-AIS messages cause problems in generating accurate routing data, creating collision avoidance tests, and search and rescue applications. One way to address these issues is to create prediction models that estimate a ship’s position at a given time or fill in gaps of information where a ship may have been. In creating prediction models it is important to have more access to different types of relevant and accurate data (Pallotta et al., 2013). Environmental conditions, such as wind and waves contribute to the movement of a ship. Waves are a major force on a ship during travel and can significantly alter its path (Szelangiewicz, 2014).

A technology that has been available for a longer period of time than AIS and LEO satellites are High Frequency (HF) radars which are set up along coasts all around the world including the west coast of North America (Gurgel and Schlick, 1996). HF radars have been around for over three decades and measure surface ocean currents for distances up to 200 km off the coastline. The information collected from HF radars are in 1 hour intervals and come in resolution of 500m, 1km, 2km, and 6km, which provides broad coverage on the West Coast (National Oceanic and Atmospheric Administration, 2013). The tracking and measurement of waves is an important piece of information when considering ocean travel due to waves creating additional stress on ships, affecting their movements. (Szelangiewicz, 2014). Gurgel et al., 2010 writes about the accessibility, range, and relevance HF radar data which for this study makes it an excellent option to use to provide wave information to enhance S-AIS message information. AIS messages are gathered in near real time depending on the mechanism receiving the data since the AIS on ships should be sending positional data at a relatively constant pace (Harati-Mokhtari et al., 2007a). The AIS also provides information on the time when the message was sent. HF radar data is collected at rates between 18 minutes for lower-frequency systems, to 4 minutes for higher-frequency systems, and then the velocities are averaged to the mean velocity of waves observed in an hour (Harlan et al., 2010). Since both data sources have short time updates it makes them worth comparing to establish a possible model relationship that can be used in a near real time database of S-AIS and HF radar information. S-AIS messages are currently being used in tracking ocean vessels and predicting a vessel's future location (Pallotta et al., 2013).

This study matches S-AIS messages with HF radar data in space and time to determine the top ocean current velocities that were present when each AIS message was received.

This process generates the following two research questions,

Question 1: Can top ocean wave currents account for message to message discrepancies in predicted position and actual recorded positions?

If a correlation exists, between the measured movement of ships through S-AIS messages and the wave velocities present from HF radars then ,

Question 2: Can HF radars be used to calibrate preliminary S-AIS based location prediction models to enhance prediction accuracy?

Chapter 2 of this paper provides an overview of the related literature. First the literature review addresses AIS messages and how the data was received and the information to date that was generated from it. Next, research that has been done in analyzing AIS messages for ship mapping and tracking of shipping routes is reviewed. This is followed by looking at the research conducted with HF radars and how the information provided by them have been used for different uses in shipping and how they function in different systems. Next is a look at how the effects from the environment can impact a ships movement when travelling or when stationed at port. Then how AIS messages have been used for prediction methods in determining the location of a ship over time and where the ship may be headed is discussed. This leads into how working with big data can be troublesome, S-AIS messages data in particular are discussed.

Chapter 3 reviews research methodology used in thesis. The first topic is data collection of the messages and their migration to a SQL database. Next is the description of how S-AIS data is analyzed to determine gaps in the data over space and time, then the study area is described which in this thesis is the port of Long Beach, California, USA. In this chapter the collection of HF radar data is explained along with the process by which it is correlated with the AIS messages. Next is a description of how S-AIS messages were grouped into trip segments that can be analyzed. Finally, chapter 3 reviews the development of a prediction model meant to estimate the position of a ship which can be compared against the actual location. This information can now be compared against HF radar data in a regression analysis.

Chapter 4 presents the results of the research. First go over the study area and data that has been subsetted for the study. Next task is select messages that are physically possible for a ship to undergo based on position over time. The results of the gap analysis are created to portray the information gaps within the dataset. The trip identification is then reviewed to show how a ship trip was identified. Next the prediction model and error vectors are created. These error vectors are then compared against the wave magnitudes and correlation values are calculated between them. Finally the results of the calibration of the prediction model using the wave magnitudes.

Chapter 5 reviews results and details how this information can be useful in other research. The results were interpreted by first breaking down the types of error that are present in the S-AIS data. Next the results found from the prediction model are analyzed. The correlation tests done between the error vectors and the wave magnitudes helped in answering if a correlation is present. The thesis then goes over the results of the calibration to the prediction model to see if it helped increase accuracy in the prediction by using HF radar data. Next, look at changes that can be made to the study by methods such as enhancing the physical model, expanding the datasets, and using near real time S-AIS data. Finally it look at future uses of this research for the purpose of dynamic ship routing, search and rescue, and calibration of existing vessel prediction models.

Chapter 2

Literature Review

Seaborne shipping accounts for 90% of total world merchandise traffic and the amount of goods traded in 2015 is forecasted to increase by 60% by the year 2030. From 1968 to 2008 seaborne trade has increased by 4 times from 12.9 trillions tonne-kilometres to 51.5 trillion tonne-kilometres and in 2015 is at an estimated 88.5 trillion tonne-kilometres (International Chamber of Shipping, 2015). With an increase in global trade by seaborne ships the advances in maritime technology becomes more valuable for analysis to aid in safety and efficiency. With the advancement and increased cost effectiveness of satellite technologies the information available reporting ship positions has become viable for sophisticated analysis and decision support applications utilize Automatic Identification Systems (AIS) data. Data fusion using existing technologies with ship position data opens new possibilities in research involving maritime tracking. A factor that affects seaborne shipping is environmental conditions which can create high winds and waves. These additional forces place strain on ships which affect their rate of speed and position (Szelangiewicz, 2014). High Frequency (HF) Radars have been installed on coasts around the world for over 30 years providing data on near-shore wave velocity. This technology can provide accurate wave information to help better analyze the relationship between open ocean waves and the trajectory of seaborne ships.

In the following chapter the literature revolves around the ideas of S-AIS based prediction models and environmental covariate effects on a ship. The topics are: AIS and pathing, finding effects on a ship, predicting ship's location, managing the data, and recently related research. AIS and pathing views the uses of AIS messages and how AIS is used in pathing a ship's route. Finding effects on a ship section discusses the use of HF radars to find factors on a ship and different environmental effects that influence a ship's movement. Predicting ship's location discusses research that has developed models to predict a ships path and how a Kalman filter is used in location prediction. Managing the data examines the difficulties working with big data, data reliability with AIS and HF radar, and the limitations of the data used. Finally the chapter goes over recently related research which looks at a study that uses ship movements to predict wave magnitudes in the local area.

2.1 AIS and Pathing

2.1.1 Automatic Identification Systems

The availability and cost effectiveness of advanced Low Earth Orbiting (LEO) satellites has led to an abundance of data for maritime purposes. LEO satellites are now able to collect data emitted from AIS which are placed on all: passenger ships that have been constructed since 2002, ships with weight greater than 3,000 tonnes since 2002, tanker ships with international voyages, and ships with gross tonnage over 300 on international voyages (Henningsen et al., 2000). An AIS is a device on a ship that periodically transmits information about the ship such as identity, position, and velocity. AIS were implemented on ships starting in 2002 for safety concerns on busy waterways. AIS allow ships to communicate information on the ship's size, velocity, and position to ships close by, all in the attempt to avoid a collision (International Telecommuniations Union, 2010). The AIS broadcasts at different rates depending on a ship's movements. If a ship is travelling at high speeds or is changing course the AIS transmits a message every 2 seconds, if a ship is at anchor the AIS transmits every 2 min, static information is transmitted from AIS every 5 min. This provides information on the destination of the ship, the identification number of the ship, and the dimensions of the ship (Henningsen et al., 2000). Further detailed information on AIS messages specific to this study is outlined in sections 3.1 & 3.2

AIS messages have led to "an overabundance of data [when there was a] previous situation of data scarcity" (Gunnar Aarsætherr and Moan, 2009) in maritime data. This technology "can also improve the quality of vessel traffic surveillance (VTS)" (Harati-Mokhtari et al., 2007a) which has been studied by different researchers over the past couple decades. "Maritime transportation represents approximately 90% of global trade by volume, placing safety and security challenges as a high priority"(Pallotta et al., 2013), this makes the need for vessel tracking more important. With current capabilities of storing large amounts of data and faster processing power in computers it is possible to analyze vast quantities of AIS data.

Researchers have made use of AIS data for analyzing aspects of ship behaviour. Xiao et al., 2015 states "AIS data provides valuable input parameters in ship traffic simulation models for maritime risk analysis and the prevention of shipping accidents". Xiao et al., 2015 analyzed ship traffic patterns in Chinese and Dutch waterways using AIS data. Using AIS the researchers were able to appropriately determine traffic characteristics in each region to draw upon similarities. The researchers compared position, speed, course, time intervals and were able to use these values to determine the types of manoeuvre a ship is undergoing such as head on head, overtaking, or collision avoidance. Carson-Jackson, 2012 talks about the development of AIS technology and its benefits for ship identification and vessel tracking. AIS has a "proven effectiveness to assist in search and rescue as well as vessel monitoring in offshore areas" (Carson-Jackson, 2012). The Australian Emergency Response Centre use S-AIS as an important source of information for search and rescue. Further capabilities are continually emerging over time with the increase in satellite coverage, accessibility of data, and efficient computing methods for faster analysis.

2.1.2 Ship Paths

A major part of vessel tracking is identifying navigation patterns from ships. Modelling a ship's navigation is necessary to determine if a ship is on the most economically appropriate path (Lane et al., 2010). Recent studies have used AIS messages to determine high density shipping routes that are taken by ocean vessels for the tracking and flagging of ships that are deviating from these routes. Pallotta et al., 2013 created a detection system that automatically generated paths of ships in vector form, a list of the generated paths, and organizes them in to a list of regular paths. Then the researchers developed a method of detection that highlights ships whose paths are outliers. This was classified as a "fully unsupervised learning" (Pallotta et al., 2013) system which they called the Traffic Route Extraction and Anomaly Detection (TREAD) system. Lane et al., 2010 determined common routes of open ocean vessels for anomaly detection and compared the common routes against a Bayesian network of different potential anomalies, to flag ships that had the highest potential of being an anomaly. Lane et al., 2010 determined there are five different types of anomalies for a ship deviation from standard route, unexpected AIS activity, unexpected port arrival, close approach, and zone entry. Deviation from a standard ship route is an anomaly consisting of a ship is travelling on one of the determined routes created through the TREAD system that then manoeuvre away from the route into a new unfamiliar route. Unexpected AIS activity would be a scenario where a ship stops sending messages and later received messages occur at a substantial distance away. Unexpected port arrival is an anomaly where a ship deviates from the originally destined port and travels to a different one. Close approach is when two ships come within close contact with each other which is against many ocean safety standards. The final anomaly is zone entry which refers to scenarios where a ship travels into restricted areas such as protected habitat zones.

Gunnar Aarsætherr and Moan, 2009 looked into grouping patterns of routes that ships take in navigating around the Norwegian coastline by mapping AIS message coordinates. They were able to simplify a vector shape for each ship sequence by factoring out small deviations within the ships path to make a straighter sequence of lines. These researchers designed to group ship paths that were similar making the information useful for looking at detection of shipping patterns and to lay a foundation down for an “automatic estimation of manoeuvre patterns from AIS” (Gunnar Aarsætherr and Moan, 2009). This study determined that a “combination of automated grouping and the availability of AIS data opens up a range of possibilities for analysis of ship traffic and manoeuvring” (Gunnar Aarsætherr and Moan, 2009).

Chu et al., 2013 looked at meteorological and oceanographic (METOC) forecast systems for aiding in optimal ship routing for minimal fuel consumption. The researchers worked with United States Navy in developing an accurate Smart Voyage Planning (SVP) decision aid which assists the ship route planners to optimize a ship’s fuel consumption. “A ship’s speed and route can be optimized based on the wind, waves, and currents, taking into account the ship’s performance criteria such as hull shape, horsepower, load, trim, ballast, pitch and roll limits, and other factors” (Chu et al., 2013). A real world test of this system took place on a navy vessel called the USS Princeton and noticed fuel savings of nearly 20% on parts of the route. The researchers mention that since the test took place during good weather with little weather uncertainty there is slight bias and they mentioned next trying it in different weather types. The SVP model was found to be very sensitive to a number of different factors such as “location, direction, seasonal synoptic/mesoscale weather, hull/propulsion type and condition, route length, specific model improvements, and ensemble methods”(Chu et al., 2013). These factors added to the statistical uncertainty when determining fuel optimization in the SVP decision aid.

Vandecasteele and Napoli, 2012 created ship paths for analyzing anomaly detection based on changeable rules set by the user. The use of ship position to generate paths allowed the researchers to determine anomalous ships and also to be able to change what conditions may be considered an anomaly using a Semantic Web Rule Language (SWRL) (Vandecasteele and Napoli, 2012). Vandecasteele and Napoli, 2012 determined further development was needed to allow for a more user friendly interface with the current system. These researchers were all able to create common vessel paths using only AIS data. Pallotta et al., 2013 and Gunnar Aarsætherr and Moan, 2009 extended their implementations by using their common paths for the use of future position estimation. This becomes more challenging because for accurate estimations since “AIS are not reliable in many cases and therefore [we] cannot wholly trust the equipment” (Harati-Mokhtari et al., 2007a) making the estimations unreliable if the underlying data is not accurate, see section 2.4.2. This requires calibration from other accessible technologies like High Frequency Radars (HF radars).

2.2 Finding Effects on a Ship

2.2.1 HF Radar

HF radars were developed in the early 1980s and are used to detect coastal waves up to 250km off the coast using Bragg scattering. “Bragg scattering explains the effects of the reflection of electromagnetic waves on periodic structures whose distances are in the range of wavelength” (Wolff, 2017) which is portrayed in figure 2.1. Bragg scattering is a x-rays reflection that occurs off regular periodic structures which causes constructive interference. This constructive interference can be used to solve for distance between the reflective surface. This method is used with HF radars on the periodic structure of waves which allows for sensors to measure the constructive interference and determine the distance between the waves. By sending multiple signals the velocity of the wave can be calculated using this method.

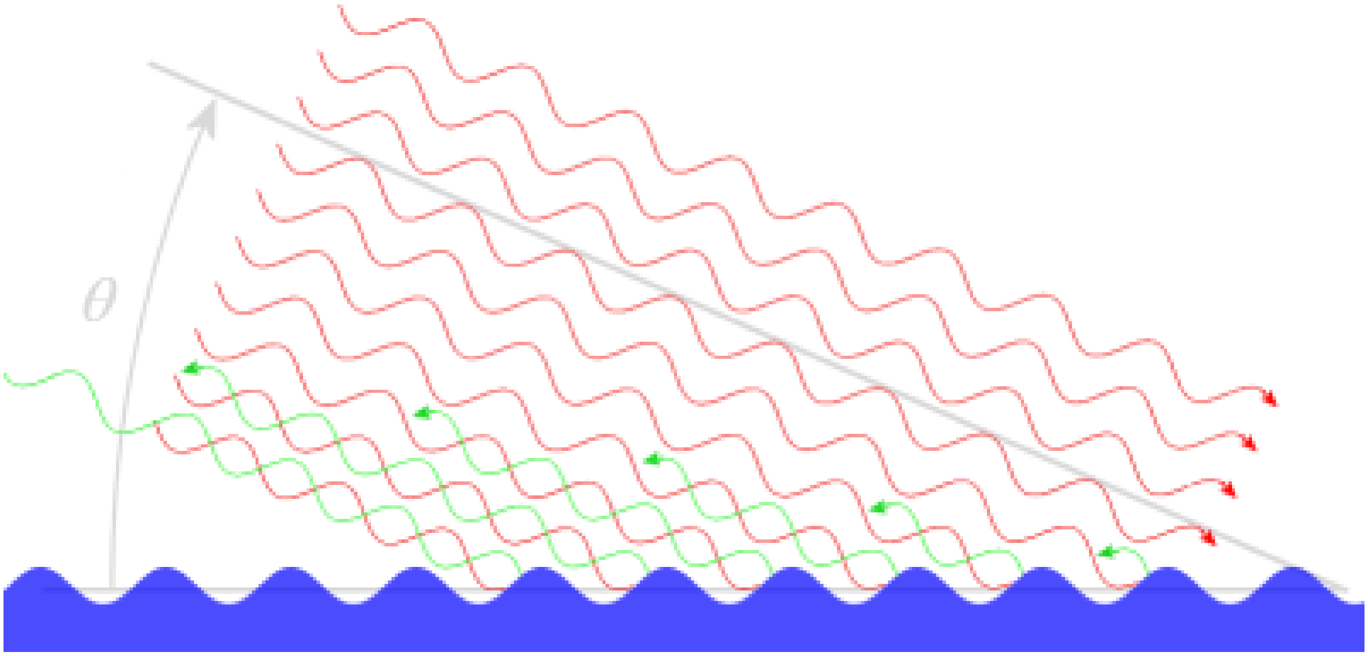

Figure 2.1: Bragg Scattering by Christian Wolff

$$d = \frac{\lambda_t}{2\cos\theta}$$

$d =$ distance of reflective sub-surface

$\lambda_t =$ transmitted wavelength

$\theta =$ transmitted angle of incidence

HF radars provide information on wave velocity which is updated in near real time to allow for top ocean current updates (Gurgel et al., 1999b). This technology has improved in accuracy over the years with the development of the Wave Radar (WERA) HF radar system which is “ advantageous for studying near-shore ocean dynamics”(Gurgel et al., 1999b) because of its better spatial and temporal coverage. Recent studies have explored using HF radars as a new method for detecting ocean vessels near shore based on back scattering that occurs and pairing the data with AIS messages to calibrate the accuracy of position (Dzvonkovskaya and Rohling, 2010).

Dzvonkovskaya and Rohling, 2010 investigated the use of HF Radar to detect ships in its range and tried to match the detection with what was found through AIS during the same time span. This was done to determine the accuracy of high frequency radar by comparing the results to the location of ships through AIS as a means to calibrate the high frequency radars. They found “the maximum detection range for very large and large cargo vessels can be up to 200 km, for the medium-sized vessels up to 160 km, for the small-sized vessels up to 140 km, and for the vessels of very small size up to 120 km”(Dzvonkovskaya and Rohling, 2010) . This paper showed the limitations in distance that the HR radars had for ship detection as well as the accuracy of the radars. This study used AIS messages to measure the accuracy HF radars have detecting ships while our study attempts to use HF radars to enhance the accuracy of AIS based location prediction models.

Vesecky et al., 2010 looked at continuous vessel monitoring using HF radar and AIS within a country's exclusive economic zone. The researchers used an estimate of signal to noise ratio to analyze the probability of detection of a ship off the coast using AIS and HF radar. From this study the researchers found that for ships up to 125 km off shore the probability of a ship's detection is 90% or greater. The accuracy was also affected by the radar coverage in the area which is lower in accuracy when there exists a large spacing of radars. They concluded that "both HF radar and AIS ship monitoring systems provide very significant marine domain awareness with respect to coastal ship traffic" (Vesecky et al., 2010). This aspect of analyzing ship detection only assists in detecting the location of ships, a method that can be used for path prediction involves using the wave data collected from HF radars as a means to better predict the future location of a ship which is what this study explores.

2.2.2 Environmental Effects

Wave effects on a ship add stress on the current path of a ship and causes deviation in the path compared to a ship travelling in calm water. "Wave action is responsible for ship motions, which reduce propeller thrust and cause increased drag from steering corrections"(Vlachos, 2004). Incorporating wave induced motion on a ship can better predict the future path of a ship and can be beneficial when calibrating prediction models that are based on AIS geographic data. Studies have been done looking at the effects that waves have on an object and how it affects the object's drift.

Abascal et al., (2012) looked at testing if wave velocity data provided by High Frequency radar can allow backtracking to the original position of an object. This research is focused on search and rescue operations. The researchers used 2 different types of buoys, with drogue and without drogue (types of stabilizing anchor). The study took place at the Bay of Vigo, Spain. They used a modified Lagrangian trajectory model

$$\frac{d\overrightarrow{x}}{dt} = (C_C\overrightarrow{U}_c + C_D\overrightarrow{U}_w)(\overrightarrow{x}_i, t) + \sqrt{\frac{6D}{\Delta t}}(\overrightarrow{x}_i, t)$$

$\overrightarrow{x}_i =$ particle i position

$t =$ time

$C_C =$ radar uncertainty coefficient

$C_D =$ wind drag coefficient

$\overrightarrow{U}_c =$ surface current velocity

$\overrightarrow{U}_w =$ wind velocity

$D =$ diffusion coefficient

to predict the path of the buoys and found that when a buoy had a drouge the effects of wind are nearly negligible. The study demonstrates the ability to detect the original location within a search area of 3.2km^2 for buoys with drouges and 2.11km^2 for buoys without drouges. "Although the validation has been performed using drifting buoys, the results could be also applied for other floating objects" (Abascal et al., 2012) which in theory can be upscaled to ships. Analyzing the effects that waves have on a ship involves the added resistance that acts on a ship and how it affects its planned movements.

Szelangiewicz's 2014 study developed a model that determines the optimal speed for a ship given the current surrounding weather conditions; it uses the ship's physical attributes and the vectors of wind and wave. The researcher found how much the ship should slow down based on different wind and wave scenarios. The maximum scenario of high winds on a ship would require it to slow its speed by 20%. For waves the ship may need to slow by over 80% to avoid the risk of tipping or rollovers. This study was done on bulk carriers which are major means of freight transport in open ocean (Szelangiewicz, 2014).

Another study that looked at added resistance on a ship was Pérez Arribas who in 2007 found the "added resistance in waves is an important part of ship dynamic[s] due to its economical [*sic*] effect on ship exploitation" (Pérez Arribas, 2007). Pérez Arribas looked at different models for calculating added resistance to a ship and tried to determine which provided the most accurate depiction of resistance on a ship. The author found that "Radiated energy method is perhaps the better method to obtain the added resistance" (Pérez Arribas, 2007). The research in analyzing a ocean vessel's path using both AIS and wave velocity data is still fairly recent and not much has been done to create a better prediction methodology using both resources.

2.3 Predicting Ship's Location

2.3.1 Prediction Models and Methods

There have been several different types of models created for forecasting a marine vessel's future position. Certain techniques have been measured at events such as the Makridakis, or M-Competitions which was designed to compare the accuracy of different forecasting methods. The most recent event, the M-3 which took place in 2000, concluded that Automated Artificial Neural Networks (AANN) head the rankings for forecasting positional location (Makridakis and Hibon, 2000). This was the precedent used by researchers Zissis, Xidias, and Lekkas in 2015 who looked at the development of an Artificial Neural Network (ANN) for the use of prediction of ship location. The researchers acquired their data from marinetraffic.com which provides free marine position data.

The ANN is a program that learns a desired output by training the program. The creator trains the ANN by providing the program with verified input and output pairs. The is input data is connected to a hidden layer of nodes which then interconnected to the final output. These connections are represented by the arrows in Figure 2.2. The interconnections are weighted values that change over time when the training process takes place. Once the training is completed a test set is then given to see if the ANN can identify the desired outputs within a given error threshold. This is shown in Figure 2.2.

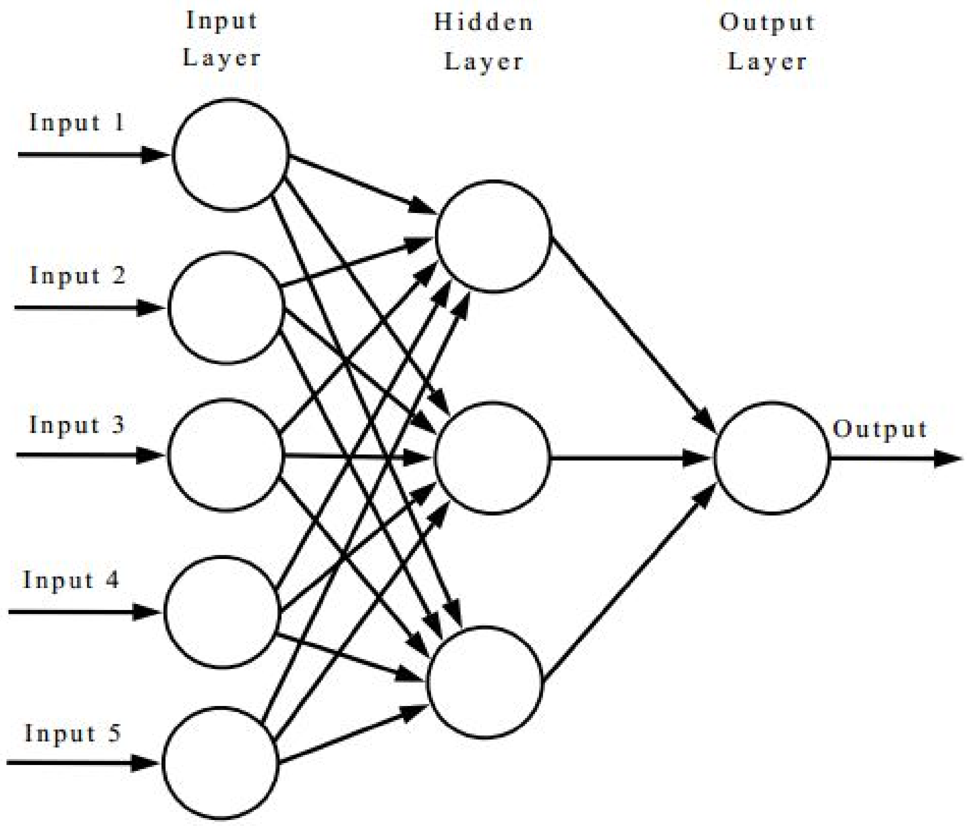

Figure 2.2: Artificial Neural Network

The ANN these researchers developed imported the S-AIS data and identified patterns that certain ships take to accurately determine the path the ship takes within 0.01 mean squared error (MSE) of predicted longitude and latitude. The researchers set up the system to predict in 15 minutes increments as well as 4 hour increments in a study area of Aegeon, Greece with passenger ships. The researchers developed a web based interface for this application and programmed it in C#. According to the authors the "proposal can potentially be used as the predictive foundation for various intelligent systems, including vessel collision prevention, vessel route planning, operation efficiency estimation and even anomaly detection." (Zissis et al., 2015).

A study conducted by Dolinskaya, 2012 presents a navigation model that uses the surrounding environment of a marine ship to determine optimal path planning over short distances. The researchers use a dynamic programming model to analyze fastest paths between two points based on a ship's current trajectory and quickest feasible turn rate. Dynamic programming in its simplest form is the method of breaking down a large complex problem into smaller manageable problems. Dynamic programming was used in order to include constraints in the model to improve its accuracy. The researchers state that "computational demand and run-time of the optimal path finding algorithm is of particular significance." (Dolinskaya, 2012) so they recommend parallel computing to decrease run-time. Dolinskaya created the model in MATLAB for its user friendly environment but mentions the use of C++ for decreasing run-times. The model was tested using a random generated dynamic wave field to simulate waves around a marine vessel to distances up 18km away because this study was only considering the use of short range trips. The reason for analysis on short range trips was that ship radars have a relatively small distance of response. The results found that there was an observed improvement up to 9.7% with an average of 4-6% improvement in optimal pathing time when compared to a model that neglects wave-field data.

Pallotta et al., 2014 researched marine vessel prediction method using Ornstein-Uhlenbeck stochastic processes,

$$dx_t = \theta(\mu - x_t)dt + \sigma dW_t$$

where $\theta > 0, \mu$ and $\sigma > 0$ are parameters and W_t denotes the Weiner process

which forecasts patterns based on historic data. The historic routes were collected through AIS and historic patterns determined using their previously mentioned TREAD system. The study area chosen was a 400 x 600 nautical mile area in the Mediterranean Sea and covers May 1 to September 10 of 2012. The new method took a partially observed track and attempts to match that path based on positional data with another previously recognized route. “These recurrent routes allow prediction of the position of a vessel that is following one of these routes, surprisingly, by several hours ’ ’(Pallotta et al., 2014) with an uncertainty of only a few kilometres. This method of vessel forecasting requires large amounts of historic positional data which is now readily available through LEO satellites able to collect AIS data.

2.3.2 Kalman Filtering

An algorithm used by researchers for estimating a ships position is a Kalman filter. A Kalman filter is used in some forms of ship navigation to estimate a ships position by using previous estimations to over time create an estimate closer to the mean. This algorithm is inclined to have greater accuracy in estimation then a static estimation. Figure 2.3 by Gershenfeld in 1999 shows the flow of a Kalman filter and that it over time alters its Kalman gain factor to increase prediction accuracy.

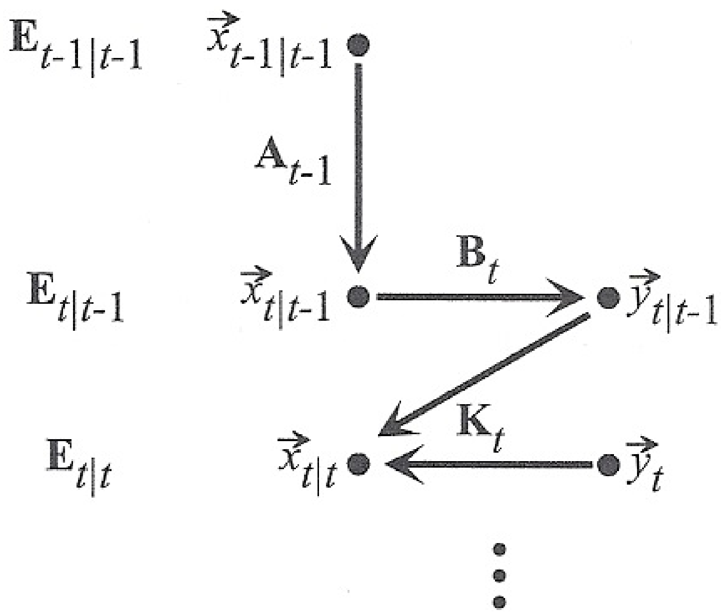

Figure 2.3: Gernshenfeld (1999) steps of a Kalman Filter equation

In Figure 2.3 it starts by the initial estimate ($x_{t-1|t-1}$) and error ($\mathbf{E}_{t-1|t-1}$). The estimate is applied to a prediction ($\mathbf{A}_{t-1}$) to find the new estimate that is then applied to the observable ($y_{t|t-1}$) with a control model ($\mathbf{B}_t$). This can be used to compute the Kalman gain ($\mathbf{K}_t$ with the error ($\mathbf{E}_{t|t-1}$) between the observable ($y_{t|t-1}$) and estimate ($x_{t|t-1}$). The Kalman gain is then applied to determine the new estimate.

Borkowski, 2017 discusses the creation of the NAVDEC, a decision support system. The system used a data fusion of different sensors and a multi-sensor Kalman filter to estimate the ship's position. Borkowski, 2017 states most ships use a " prediction based on the extrapolation of current traffic parameters (speed, acceleration), assuming that these parameters will not change significantly in the future" (Borkowski, 2017) so the proposed algorithm will have higher accuracy. The system used gyrocompass, automatic radar plotting, GNSS (global navigational satellite system), AIS, and electronic navigation charts as information inputted into the algorithm. The multi-sensor Kalman filter weigh the averages of estimation from the different sensors and then calculates the Kalman gain, the relative weight that changes after each iteration, which is based on these new values compared against the previous estimation and the changing Kalman gains current value. This means the Kalman gain changes over time until it reaches equilibrium and the prediction become more uniform. Borkowski, 2017 concluded that one of the main causes in ship collision is human error and the decision support system will help assist in decision making.

Sadowski and Czapiewska, 2015 analyzed different ways to determine the minimum amount of stored S-AIS messages required to accurately replicate a ships path. One of the methods used in this study was a Kalman filter in trying to estimate a ship's position. Sadowski and Czapiewska, 2015 looked at different types of ship path situations (turning into port, straight away, narrow channels) and determined what percent of the messages were required to be stored to accurately generate the same path within error thresholds. The thresholds were looked at 10 metres, 50 metres, and 100 metres of error accuracy for the different cases. In the study the Kalman filter was calculated to be slightly more accurate over the linear algorithm but "linear algorithm is much less complex than Kalman filtering authors suggest using the linear one" (Sadowski and Czapiewska, 2015). The Kalman filter was able to replicate the selected ship paths using only 2-4% of the available messages when a 100 metres threshold on error was allowed.

A study by Laws et al., 2011 applied a Klaman filter to their existing model of combining S-AIS ship detection and HF radar ship detection. The researchers used raw HF radar data to detect the presence of ships off the California coastline and using S-AIS messages to confirm positions. From HF radars there is signal noise caused from Doppler shifts from top ocean currents. The researchers decided to use Kalman filtering to track ship paths and help detect false alerts in the HF radar ship detection. Laws et al., 2011 found an issue when applying the filter, that is when to terminate a track of path and revert back to the previous information. This information could be determined by supplementing previous

S-AIS messages to count parts of the ships trip, which is discussed in this thesis. Overall the researchers saw potential benefits of the Kalman filter algorithm but it caused biases in there study that need to be resolved. Laws et al., 2011 found that careful monitoring of the measurement data is required for the application of Kalman filtering which can be resolved with extensive cleaning of S-AIS messages which is examined in sections 3.2, 3.3, 4.3.

Kalman filtering provides plenty of benefits when creating a prediction model and estimating the location of a ship over time with increased accuracy. A Kalman filter was considered for this study because it provides a better location estimate but the way it generates this is by applying a correction factor for each time. This correction factor may cause a problem when trying to find a correlation with wave magnitudes because the correction factor would normalize the error variance that is used to correlate with the wave magnitudes.

2.4 Managing the Data

2.4.1 Working with Big Data

When working with large data sets the way the data is integrated and analyzed becomes an ever increasing issue. Certain computational techniques become impractical due to processing time limitations when the process is applied to larger data sets. An example of this problem is generalized in a fundamental problem of computing called combinatorial explosion.This problem occurs when trying to determine a solution to a problem and with increasing inputted information exponentially increases in possible combinations for a system to solve the problem directly so the solution is to create a selective sequence for the computer to analyze (Tsang, 2015). Different methods have been used in figuring out the best methods for analyzing large amounts of data and cutting down the time to process it. In the current problem domain, Sadowski and Czapiewska, 2015 decided to look at taking large AIS datasets and compressing the data by only retaining a fraction of the entire data set based on critical points in the data set. The researchers look at replicating and comparing three different types of data compression techniques for storing ship data; Linear Algorithm, Circle Algorithm, and Kalman Filter. The first method uses the idea that ships move in fairly linear travel paths, the second method takes into account that ships cannot take sharp turns and works in arcs, and Kalman filtering continuously predicts errors to calibrate future forecasting. Using these methods the researchers analyzed different ships by taking positional data in intervals of 10 metres, 50 metres and 100 metres. They found that they were able to compress the amount of data needed to maintain an acceptable amount of error to accurately track the vessel. The amount of data needed to be stored changes drastically based on certain shipping situations such as open ocean where only 0.5% of the data was needed at 50m threshold but when in a port area 21% of the data needed to be stored. This means we can "reduce stored data to only few percent of the origin[al] number and recall later the routes of the vessels with acceptable error" (Sadowski and Czapiewska, 2015) allowing faster computing to be done on these smaller subset of data.

Maulidi et al., 2014 did a quick study that integrates AIS data into a shipping database server. The study was created to show how to make an integrated database that could be used as a "backbone of marine safety and environment monitoring system[s]" (Maulidi et al., 2014). The data was related using the Maritime Mobile Service Identity (MMSI), unique identifier for a ship, and the gross tonnage of a ship to match the AIS messages and the information present in the shipping database. This processing took place in a MySQL database was then linked to a web browser interface to allow a user to conveniently "[click] each ship [to display the] amount of emission" (Maulidi et al., 2014) each ship produces over time. "By using the web-based traffic control, it is more efficient to to monitor the marine traffic, safety and environment in term of air pollution in the specific area." (Maulidi et al., 2014). This method of integrating AIS data to a shipping database allows for some simple data analysis.

2.4.2 Data Reliability

AIS messages are important for maritime safety and vessel tracking. The reality though is the that AIS messages are susceptible to, signal error, human error and malicious attacks. This causes issues in analyzing AIS message data because it can produce messages that less accurately portray the real world movements of a vessel or blatantly change the position and information of a ship. A study conducted by Harati-Mokhtari et al., 2007a looked over previous studies and wanted to determine the reliability and human error that are present in AIS messages. "Navigational status is dynamic information that has to be manually entered by the officer of the watch (OOW) and changed or updated as necessary by the navigation officer ... in the VTS-based AIS study 30% of ships were detected as displaying incorrect status information" (Harati-Mokhtari et al., 2007a). This information in the AIS message is important because based on the navigational status the AIS changes its rates of transmission. For the positional information of AIS messages the study found that "1% [of messages] had shown latitude of more than 90°and longitude of more than 180°or the position 0°N/S, 0°E/W" (Harati-Mokhtari et al., 2007a) which could be attributed to the position system not working. The human errors listed in the study include installation, program design, training, equipment regulation, and violations of AIS accuracy. These errors cause issues with reliability of AIS messages. The study makes suggestions on what steps must be taken for human training to minimize the uncertainty in AIS messages and improve overall accuracy of the systems (Harati-Mokhtari et al., 2007a).

Balduzzi et al., 2014 conducted a study to evaluate how secure AIS were against malicious attacks using software and hardware. In this study using software they were able to change information of major AIS message providers (Marine Traffic, AisHUB, Vessel Finder) and were able to change a ship's position, create imaginary ships/buoys, and replace information on AIS messages for a specified ship (Balduzzi et al., 2014). With the use of their own altered AIS the researchers were able to accomplish the same results as the software attacks as well as more. Using an altered AIS they were able to send distress beacons which if received by a ship which by "law is required to join a rescue operation upon receiving a search and rescue message" (Balduzzi et al., 2014). Balduzzi et al., 2014 were also able to override AIS systems messages and make a ship's AIS seem silent by methods of slot starvation, frequency hopping, or timing attacks. The information from this study shows how susceptible AIS messages are to attacks and can be used for changing historic information or real time incoming information as well. The researchers reached out to message providers and AIS developers in hoping to fix these potential methods of attack and are actively working to stop any malicious attacks.

2.4.3 Limitations of the Data

With any source of data there are likely to be errors created by the physical limitation of the technology or mis-reporting information and so measurements must be taken to either find these errors or understand where accuracy may be hindered. This is apparent in both AIS messages and HF radar datasets. Heymann et al., 2011 collected AIS messages in the month of September of 2011 and analyzed “the contained information ... against the plausibility of their values according to other parameters in the time series” (Heymann et al., 2011). The fields of the AIS messages looked at were Course Over Ground (COG), Speed Over Ground (SOG), Position, Heading, Navigational Status, and MMSI. For each of these fields there were critical values assigned based on either AIS message parameters set by International Maritime Organization or by physical limitations on a ships manoeuvrability set by the researchers. An example of a critical value is SOG where they calculated the distance and time between messages to calculate their own average speed. This average speed was seen as the critical value for reported SOG. If the critical values were exceeded then it was seen as not being plausible. The study areas were Rostock harbour of Germany and the Baltic sea. The study showed that there were less overall message errors in the Baltic sea which may have been explained by the larger Baltic sea region that has more dynamic routes and less congestion when compared to the tighter compacted harbour which has a higher density of ship traffic (Heymann et al., 2011). This is an important note since the Baltic sea is a closer resemblance to the study area used in this thesis mentioned in section 3.5. In the AIS message data 10% of the messages in harbour were unlikely with 44% having an error in the heading field, 34% for COG, and 5% for SOG. Heymann et al., 2011 concluded that on average the AIS messages contained critical values less than 10% of the time with most errors occurring from the heading and COG information.

Studying HF radars Gurgel et al., 1999a analyzed the physical limitations of HF radars for remotely sensing surface currents. The study explains factors that affect the radars such as “material parameters of both the sea and atmosphere [such as] salinity and temperature” (Gurgel et al., 1999a). The spatial resolution has physical limitations as well since the “scattering range has to extend over a large number of HF wavelengths in order to allow resonant backscatter” (Gurgel et al., 1999a), this means that data further away from the radar will need to be at a lower resolution to receive the necessary information. At further distances radio interference must be minimized by changing the radar signals which lowers data resolution. The researchers also determined that “disturbances can only arise from nearby radio stations” (Gurgel et al., 1999a). The authors concluded that HF radars are dependent on high salinity of the water which is why it cannot be used in freshwater lakes. The working conditions for HF radars also set limitations. At 100 kilometres away resolution is 8 kilometres and the working range for finer resolution of 0.3 kilometres is within the first 50 kilometres (Gurgel et al., 1999a). This information is important when combining HF radar data with other sources since information closer to shore will have greater accuracy.

2.5 Recent Related Research

A study that took place concurrently with this thesis is that by Guichoux et al., 2016 whose goal is to estimate fine scaled sea surface currents by calculating error between predicted location of a ship and given position, based on AIS messages, which follows similar methods presented in this paper. Their study took place south of Sicily and was "was chosen mainly because of the great variety of [sea current] circulation features" (Guichoux et al., 2016). The researchers state that the methods available for measuring near-surface currents fail to provide information on the small scale features in ocean currents. In the study they sampled AIS messages in minute intervals and only looked at cases where a ship was travelling at a reported speed greater than 6 knots to exclude cases where a ship is actively manoeuvring which would cause a strong bias in their analysis because changing course changes how the external forces act on a ship. Guichoux et al., 2016 determine an error in ship position, that is an expected association of ship drift produced by near-surface currents, and calculate their calculated near-surface ocean currents for that given time. From this study locations of 5 eddies were located and compared against a bathymetry map to compare the likelihood of these eddies existing. The sampling method of AIS messages used in the Guichoux et al., 2016 study will be partly replicated for this study to see if the condition of constant straight line motion produces a stronger relationship between drift and HF radar near-surface current values.

2.6 Summary

Based on a review of previous research using AIS, "the viability of AIS as a data source appears good"(Gunnar Aarsætherr and Moan, 2009) for analyzing navigation patterns of open ocean vessels. AIS allows for nearly continual updates on ship positioning to better track and manage ship routing. HF radar has also proved useful in detecting ships (that may not have an AIS aboard for reporting data) that are within 200 kilometres of a coast that has HF radars on the shore. The studies that use HF radar to detect ships have acknowledged that "ship detection and tracking results based on the real HF radar data processing show a good performance"(Gurgel and Schlick, 1996). Finally, incorporation of wave data produced by HF radars can be seen as beneficial in determining the projected paths of ocean vessels based on the resistances that waves produce. This continuous update of real time data can assist in ship location predictions when AIS signals are not available. These studies inform an understanding of AIS and the effects produced from environmental covariates which for this thesis will focus on waves detected by the coastal HF radars. Using these previous studies the thesis will expand on the available knowledge by looking deeper into S-AIS and HF radar data and what relations may exist between them. The thesis will also use the previous research to guide in some of the tests by comparing methods of selecting data that may produce suitable conditions of ship and wave interactions.

Chapter 3

Methodology

This chapter reviews the collection process of satellite based S-AIS messages and HF radar data of top ocean currents. The process the data was cleaned for data processing and data fusion is explained in detail as are and the methods used to address the issues around S-AIS message gaps. The creation of a preliminary prediction model that determines a ship's position over time and the error in that prediction will be discussed. The error of these predictions are then compared against HF radar data to answer,

Question 1: Can top ocean wave currents account for message to message discrepancies in predicted position and actual recorded positions?

This was done by creating a spatial temporal join between the S-AIS and the HF radar data, then comparing the errors in the preliminary prediction model to the waves present during that time. S-AIS message based trips that met an inclusion criteria where the waves would be the greatest contributing factor were subset for further analysis. Another subset was analyzed based on the Guichoux et al., 2016 study's condition of constant straight line of S-AIS messages which according to the researchers is the strongest case of measurable wave effect. Finally a method is designed to answer,

Question 2: Can we use HF radars to calibrate preliminary AIS-based location prediction models to enhance prediction accuracy?

This was looked at by calibrating the preliminary prediction model using the results of the prediction error and HF radar relation.

The following Figure 3.1 represents the work flow diagram used in answering these questions.

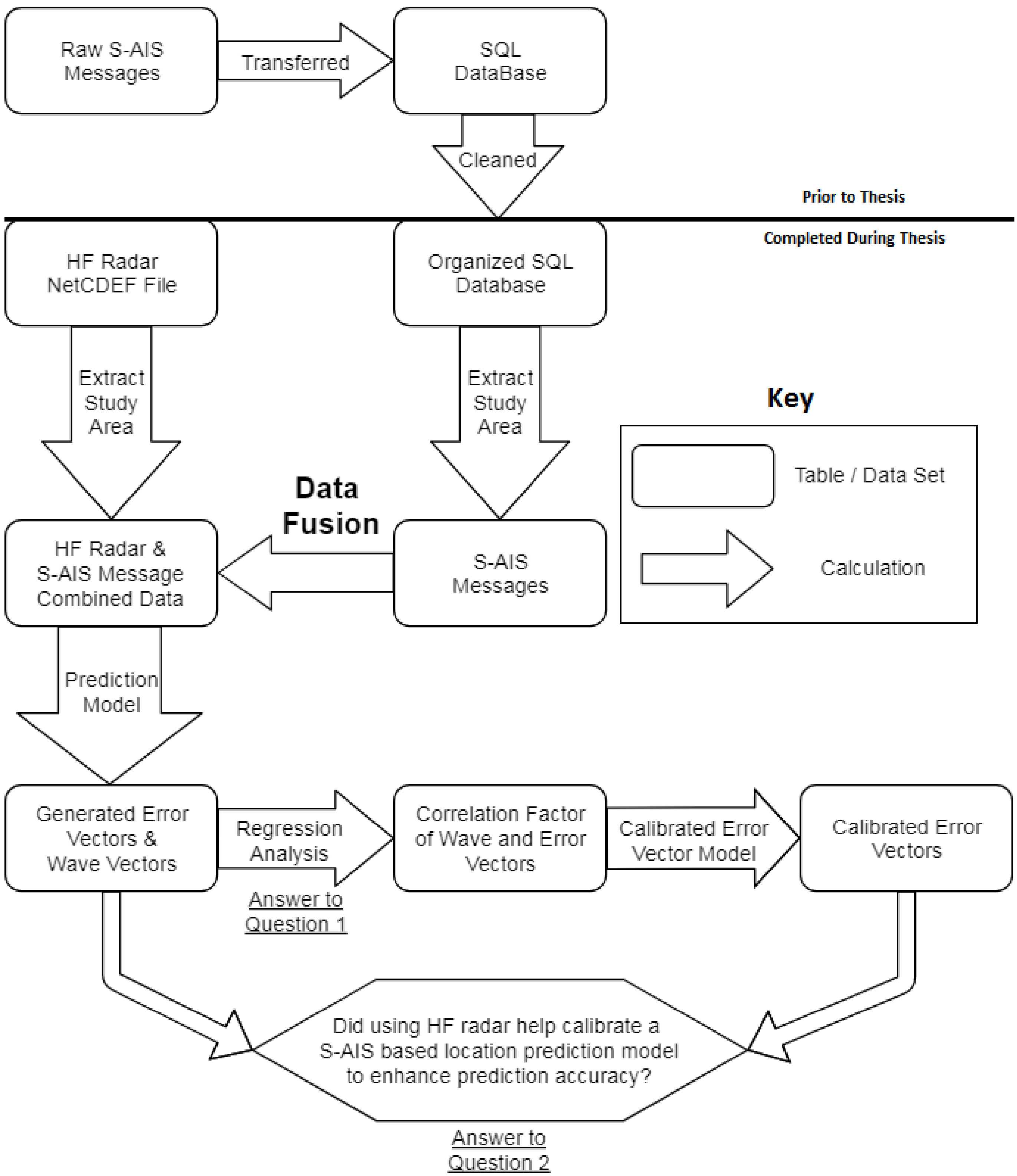

Figure 3.1: Work Flow Diagram

3.1 Satellite Automatic Identification System Ship Data Importation to SQL Database

S-AIS data is a large dataset with billions of messages received by LEO satellites each year. The S-AIS data for this study was provided by original study funders who allowed the information to be used for our analysis. The S-AIS message data was sent in two parts, the first half of the data was sent by physical hard drives delivered to the University, and the second half of the data was downloaded through private web access. The files were all text/comma separated value (csv) files that contained S-AIS messages signals, received by LEO satellites, that intercepted signals transmitted from ships in the North Pacific Ocean in the time span of Nov 1 2012 - Oct 31 2013. Once all the data was imported it was inputted into a PostgreSQL database on a server where it was prepped for cleaning and reorganization.

3.2 Cleaning of Ship Data in SQL Database

After the S-AIS data was transferred into the PostgresSQL database it was cleaned and prepared to both minimize its size and extract only needed information (see appendix B). Not all the S-AIS data set required for this study and therefore could be removed. There were a total of 138 columns provided within the S-AIS dataset and from this it was determined that 34 of them provided enough of the necessary information. The columns that were not used were either inconsistent enough to be unusable (beige), entirely blank (blue), or not appropriate for the study(red). Table 3.1 below is a listing of all the columns dropped and kept.

Columns Kept		Columns Dropped	
MMSI	Millisecond	Name_extension_padding	Offset_4
Message_ID	Region	Message_ID_1_1	Number_slots_4
Repeat_Indicator	Online_data	Offset_1_1	Timeout_4
Time	Group_code	Message_ID_1_2	Increment_4
Country	Sequence_ID	Offset_1_2	ATON_type
Base_station	Channel	Message_ID_2_1	ATON_name
Vessel_name	Data_length	Offset_2_1	off_position
Call_sign	AIS_version	Destination_ID_A	ATON_status
IMO	Regional	Offset_A	Virtual_ATON
Ship_Type	RAIM_flag	Increment_A	Channel_A
Dimension_to_bow	Communication_flag	Destination_ID_B	Channel_B
Dimension_to_stern	Communication_state	Offset_B	Tx_Rx_mode
Dimension_to_port	Fixing_devise	Increment_B	Power
Dimension_to_starboard	Transmission_control	data_msg_type	Message_indicator
Draught	Sequence	Station_ID	Channel_A_bandwidth
Destination	Destination_ID	Z_count	Channel_B_bandwidth
Navigational_status	Retransmit_flag	num_data_words	Transzone_size
ROT	Country_code	health	Longitude_1
SOG	Functional_ID	unit_flag	Latitude_1
Accuracy	Data	display	Longitude_2
Longitude	Destination_ID_1	DSC	Latitude_2
Latitude	Sequence_1	band	Station_Type
COG	Destination_ID_2	msg22	Report_interval
Heading	Sequence_2	offset1	Quiet_Time
Maneuver	Destination_ID_3	num_slots1	Part_number
UTC_year	Sequence_3	timeout1	Vendor_ID
UTC_month	Destination_ID_4	Increment_1	Mother_Ship_MMSI
UTC_day	Sequence_4	Offset_2	Destination_indicator
UTC_hour	Altitude	Number_slots_2	Binary_flag
UTC_minute	Altitude_sensor	Timeout_2	GNSS_status
UTC_second	Data_terminal	Increment_2	spare
ETA_month	Mode	Offset_3	spare4
ETA_day	Safety_text	spare2	
ETA_hour	Nonstandard_bits	Name_ extension	
ETA_minute		spare3	

Table 3.1: Columns of input AIS data: Kept and Dropped

After extracting only the necessary columns the next task was to remove S-AIS messages that were anomalous or were not from ocean vessel ships, as some signals came from ground stations and airplanes. AIS messages have a maritime mobile service identity (MMSI) and this is unique for each ship's radio equipment. MMSI is a nine digit number that uniquely identifies a single vessel but not every nine digit number can be used. The data was further cleaned by removing MMSI values that did not relate to the vessels we were interested in. First we removed any message that had an MMSI value lower than 200000000. This is because a MMSI that starts with a 0 is a radio signal registered to a coast station and a MMSI that starts with a 1 is set aside for the use of search and rescue aircraft. The next set of MMSIs that were required to be removed were those equal or greater than 800000000. A MMSI that begins with an 8 is specific to a hand held radio and a MMSI that begins with a 9 are devices using a free-form number identity. Afterwards all that remained were messages with a MMSI starting with a number from 2-7. Some remaining messages had positional information that was null. Thus we removed all messages that did not contain a longitude or a latitude attributed to it (International Telecommuniations Union, 2010).

The next part of the data cleaning was the removal of repeated messages within the dataset. A message's Message ID specifies the type of message being sent. If a Message ID is number 5 it is a static message that provides additional information on a ship that is otherwise not sent in a standard Message ID of 1, which provides positional information on a ship. Thus, only one Message ID of value 5 is needed for a ship. Redundant messages were removed if a ship's MMSI, Message ID of 5, and Time values were all the same and only one message was kept. If a Message ID was lower than 5 it is a positional report of a ship. With messages being sent at a rates as high as one message every two seconds there can be signal disruption or delay from the information being received by the satellite which can cause repeats in data being received. To remove replicated messages within the data, the messages were sorted based on MMSI and Time, if a message's MMSI was the same as the previous one, the Time stamp was exactly the same as the one previous, and the Message ID was not 5 then the message was removed. With the extraction of only necessary columns, ship messages, and unique messages the data was ready for the next phase of the study. A specific concern with S-AIS data is involving the time stamp of the message and whether it is associated with the time the AIS message was sent or the time the AIS was received by the satellite. The data set had gone under alterations with labelling and un-available metadata so it could not be confirmed what the time stamp was associated with. In this study it was initially assumed that time in the S-AIS messages is the time the AIS message was sent. This is used as a working assumption in the analysis which is later proved may be an issue when reviewing the reliability of the time stamp in S-AIS messages in Section 5.1.1.

3.3 Selecting Physically Realizable Values

Using suggestions from the paper by Heymann et al. (2011), messages were filtered out if they exceed certain physical limits. The first physical limitation that was easy to filter out were ships that exceeded a given Speed Over Ground (SOG) value. The limit chosen for this study was 50 knots (92.6 km/h), this value was chosen because the maximum speed of tanker ships is roughly 30 knots and with different types of ships present in the data set this seemed like a reasonable threshold. This also eliminates the SOG default value of 1023 which would create errors later on.

The next scenario measured was the plausibility of a ship's location over a given time span. To calculate the possibility of a ship being at a given location we assumed the ships speed was 50 knots and calculated the maximum distance that can be travelled in a set period of time. The speed of 50 knots was selected after searching online resources for fastest ship speeds and determining that 50 knots was greater than top speeds of ship's that are not specifically designed for only speed. To determine this for each message the difference in time between 2 sequential messages were taken as well as the distance between their positional information. From this a maximum distance was calculated by multiplying difference in time (seconds) by 25 m/s (50 knots) and if this maximum distance is less then the distance between the given positions then the following message was removed. An example of a error that this removed were scenarios where a ship would suddenly jump 30 kilometres off course in a message and the following message would return to the previous position. The number of cases where this situation occurred was substantial throughout the data set. Figures 3.2 & 3.3 depict the difference in time and difference in distance between messages. The line increments at 25 metres for every second so any point above the line is a values that exceeds the ships maximum speed. The time was limited to 200,000 seconds because the max distance that could be between two messages in this study area is 500 kilometres.

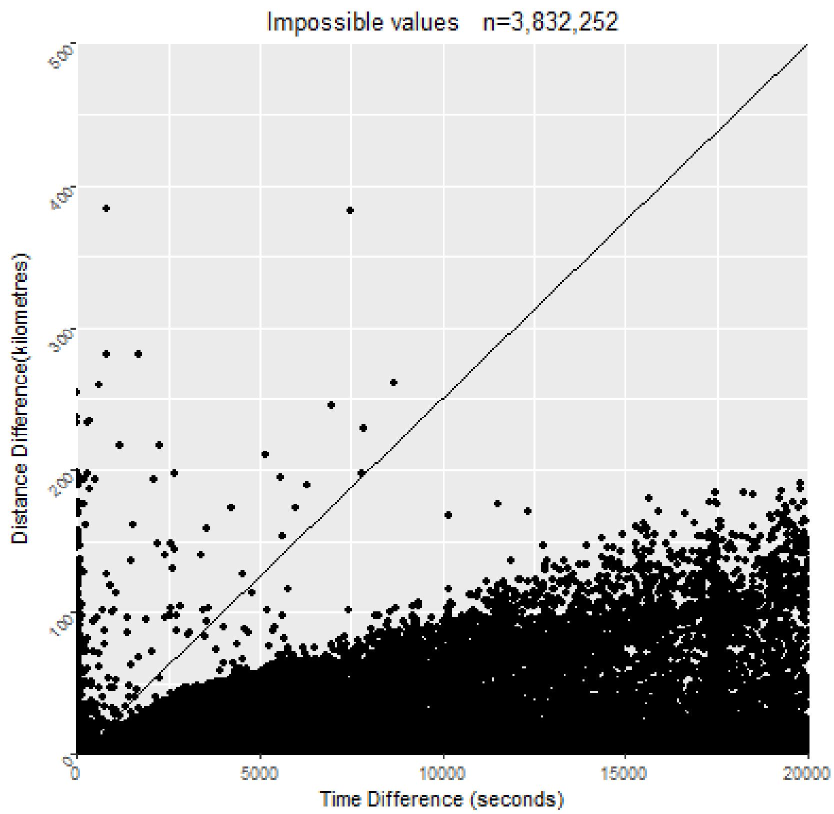

Figure 3.2: Difference in time and distance between each message

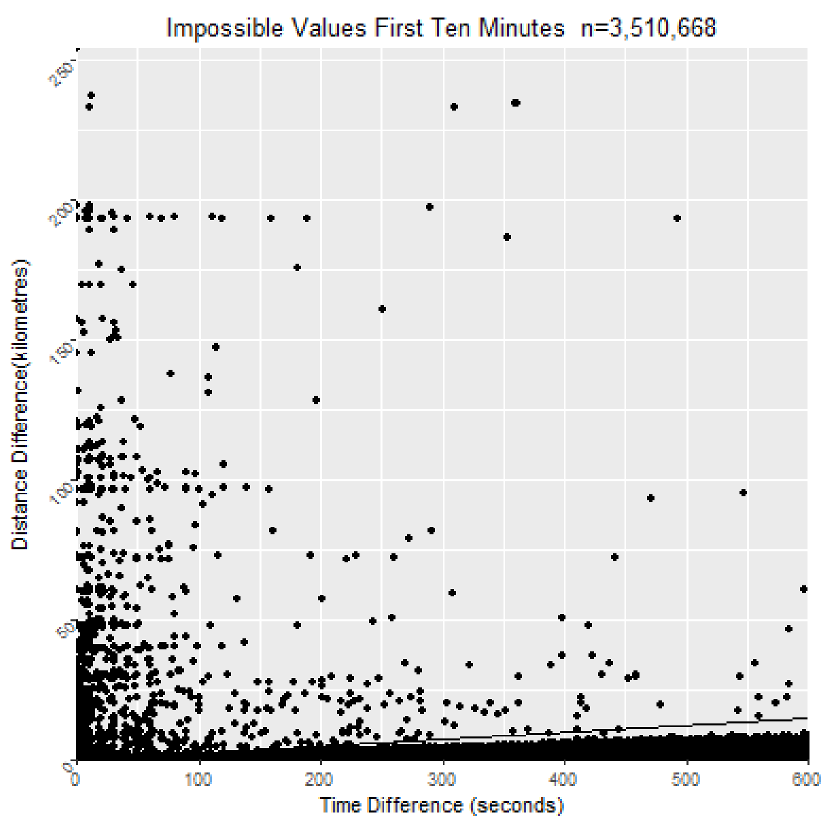

Figure 3.3: Difference in time and distance between each message

3.4 Gap Analysis of Ship Data

A gap analysis was conducted to express the significant need for a prediction model to determine a ship's path during periods where no messages are available. Message gaps in this study can occur for a number of different reasons such as, a satellite not being present in the area to receive the AIS message, a form of interference occurring during the time the AIS message is sent to when the message is received by the satellite, or if a AIS is turned off and is not sending out messages at all. In this study a gap will be defined as a break in a ship's path when no message is received within a 6 hour time span from the previous message and the ship has moved further then 5 kilometres after the next message is received. The reasons for these parameters are because the satellites that provided information for this study comprised a constellation where the maximum time a given area would not have satellite coverage in the given year was 6 hours. The requirement that a ship needs travel a distance greater than 5 kilometres is because if a ship is in port it may turn off its AIS because there is no movement taking place over a large period of time. Ships in port may also have gaps generated by radio interference because of the density of AIS messages that take place in port from multiple ships. It is only ships that are moving greater distances than within port that are of interest in finding the gaps for in this study.

The code to find ship gaps starts with a reorganization of the data set. First we sort the data based on the ship's MMSI, followed by the time. This groups the S-AIS messages based on the proper time sequence that a ship travels over the course of a year. Next a time difference calculation is done that takes the S-AIS message time value and calculates the difference from the previous S-AIS message time value and outputs the difference in those times.

The next step is to calculate the difference in position between messages. The difference is calculated between an S-AIS message and the previous S-AIS message's longitude and latitude fields. Pythagorean Theorem can be used to calculate the distance from an AIS message and the previous AIS message by $\sqrt{\Delta Longitude^2 + \Delta Latitude^2}$. The distances are calculated assuming locally flat space as the magnitudes are what are important and not precision so the error in this assumption is acceptable.

The last step before finding the gaps is to create a subset to not include the first message of each S-AIS message sequence based on MMSI. This is done because in the sorted list when a new MMSI starts all the difference calculation that are performed will be against a different ship making the messages not applicable. To determine ship's gaps now all that is done is a selection criteria from the S-AIS messages (with the first message of a new MMSI removed) where a S-AIS messages time difference is greater than 6 hrs and has a positional change greater than 5 kilometres. This information is used to determine the gaps present in the data to illustrate the need for prediction models to fill the missing information in satellite based AIS data sets. (see Appendix A for code used for these calculations)

3.5 Selection of the Study Area

The study area chosen for analysis was the ocean area surrounding the Port of Long Beach, California. This area was chosen based on recommendations from the original funders of the study who were interested in this port due to the amount of traffic at the port and by how many international ships travel across the Pacific to deliver there. The reason this is also an appropriate study area other then its heavy traffic are the physical attributes of the study area. The Port of Long Beach is a directly open ocean port with minimal barriers coming into the port. This means a lot of ships remain anchored in large portions of open ocean which seemed ideal for the study because information on anchored ships was a required input for parts of the study. The spatial extent used for the study area is Longitude: (-118.2, -120.2) Latitude: (32.8, 34.8). These dimensions were chosen to encompass the entire study area and for convenience of collecting HF radar data. If the spatial extent were any larger the HF radar file from Coastal Observing Research & Development Center website (http://cordc.ucsd.edu/) would be too large for download requiring multiple files to be downloaded and merged to create a larger area. After the study area was determined, AIS message information was subset to the dimensions of the study area so that it may be used in the initial analysis. The AIS messages were subset into a group based on their longitude and latitude being within the study area boundaries, Longitude: (-118.2, -120.2) Latitude: (32.8, 34.8) (see figure 3.4).

3.6 Trip Identification Creation

Next we describe the way we organized the message data into groups based on a sequence of messages for a given ship's trips. In this study, a ship trip was defined as a sequence of messages with in-motion navigational status between two end points of messages with at-rest navigational status. Trips were broken down by using a trip code for each message formed from the MMSI number, followed by trip and year of the message. The columns used to determine the trips were, MMSI, Date, and Navigation Status. The table of message data were first sorted by MMSI number then by the Date and Time the message was received. An index number was given to each message for sorting of the data. A TripID column was added which serves as the trip identifier for each message. The last column added was a SequenceID which is an identification of what part of the trip that each message was received.

SQL code was written to achieve the next step (see Appendix B). A SearchCursor iterates through the rows and compares the previous row to determine what the TripID and SequenceID are. The code uses two different counters to determine the trip number and

Figure 3.4: Study Area

the sequence number. The code first looks at MMSI, if the MMSI is not the same as the MMSI of the previous row then both counters are reset. If the MMSI remains the same as the row prior then the navigation status is viewed. The start of a trip is defined as a ship moving from a previously anchored location, and an end of a trip is defined as a ship that is either moored or anchored. When looking at the navigational status, after MMSI is checked, if the navigational status of the previous row is not a 1 or 5 then the trip counter remains the same but the sequence counter increases by one. If the navigational status of the prior row is a 1 or 5 and the navigational status of the current row is a 1 or 5 then the trip counter remains the same and the sequence counter goes up by 1. If the previous row has a navigational status of 1 or 5 and the current row has a navigational status of anything other than a 1 or 5 then the trip counter increases by one and the sequence counter resets to 1. For each message the TripID field is filled as “The MMSI” + “T” + “TripID”. The SequenceID is equal to the sequence counter. Each message can now be determined to be part of a specific trip that the ship took during the given year and how many messages occurred during in the trip. The former is used to easily extract single trips for analysis.

3.7 Prediction Model & Generating Error Vectors

When a vessel does not change its rate of turn and maintains constant speed in a controlled environment the distance traveled and new location can be estimated using the ship's velocity and time passed. If a ship at position A has coordinates (x_1,y_1), speed of (V), and heading of (θ) then position B (x_2,y_2) after change in time (Δ t) is then calculated.

$$\begin{aligned} x_2 &= x_1 + \cos(\theta) \cdot V_1 \cdot \Delta t \\ y_2 &= y_1 + \sin(\theta) \cdot V_1 \cdot \Delta t \end{aligned} \tag{3.1}$$

The S-AIS messages position were provided in latitude and longitude coordinates so they were transformed into Universal Transverse Mercator UTM zone 11, which is specific for this region, so the calculations could use metres as a distance measure. This was completed using R's spTransform function. The *x* and *y* coordinates are in metres north and east allowing for simple trigonometric calculations. This equation is the basis of creating the preliminary location prediction model.

A problem with the equation above is that in the real world there are external factors that affect these results. The largest contributing factor for error in positional estimation of a ship is environmental covariates, especially waves (Szelangiewicz, 2014). By calculating the expected position of a ship over time and comparing it against the real world position the vector difference between the two points can be compared with wave velocities. To calculate the error vector we need ship position A (x_1,x_2), position B (x_2,y_2) heading (θ) at position A , speed (V) at position A , time (t) at position A message and time at position B message. First the expected position (*Ex, Ey*) must be calculated.

$$\begin{aligned} E_{x_2} &= x_1 + \cos(\theta) \cdot V_1 \cdot \Delta t \\ E_{y_2} &= y_1 + \sin(\theta) \cdot V_1 \cdot \Delta t \end{aligned} \tag{3.2}$$

The error vector is calculated by comparing the differences between position B(x_2,y_2) and expected position (E_x, E_y). Error Vector North (E_y - y_2) , Error Vector East (E_x - x_2). This is completed over the entire dataset with the following equation.

$$\begin{aligned} \mathrm{Err}_{x_i} &= x_i - E_{xi} \\ \mathrm{Err}_{y_i} &= y_i - E_{yi} \\ \text{where } i &= 1...n-1 \;\text{ for n points} \end{aligned} \tag{3.3}$$

The Northward and Eastward error vectors (see Figure 3.5 & 3.6) can now be used to run a regression analysis against the Northward and Eastward wave velocities in the cell. In any case where an error vector crosses into a new cell boundary the cell which the starting point is referenced is used.

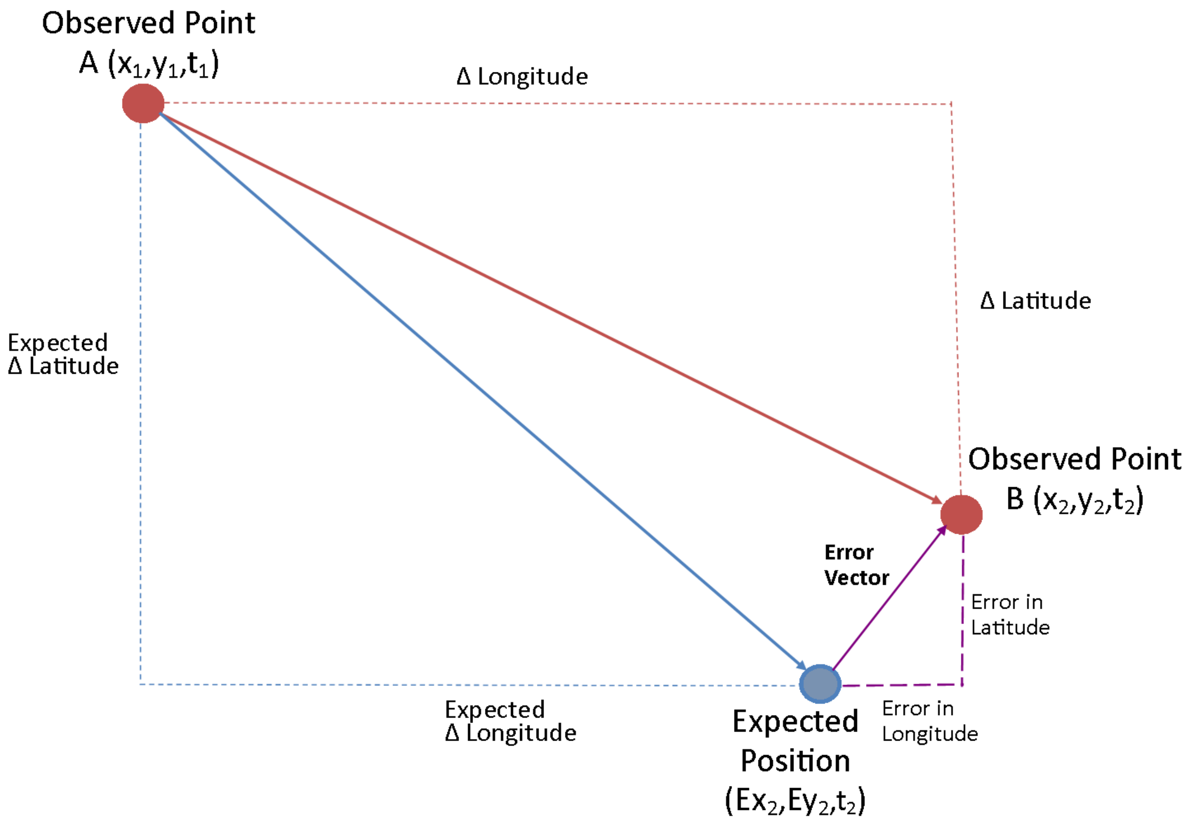

Figure 3.5: Error Vectors

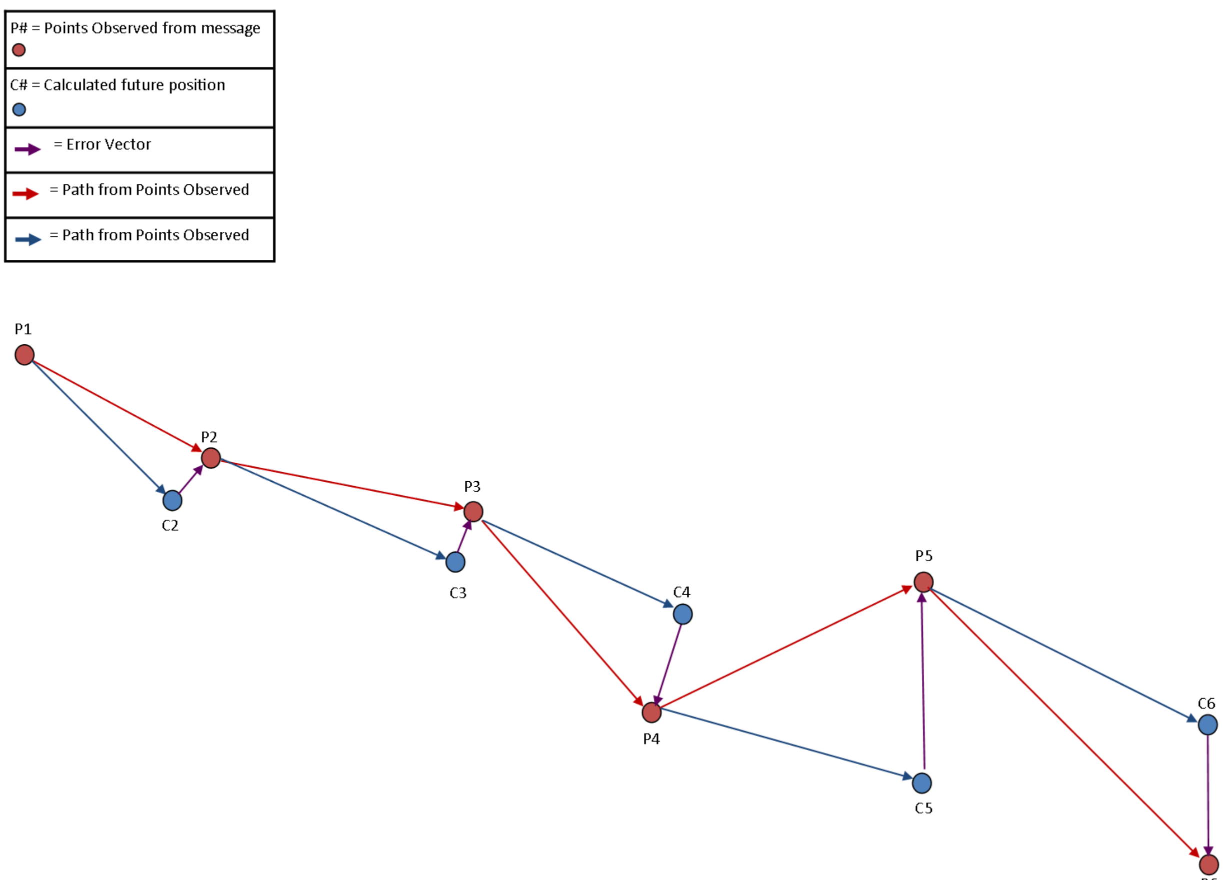

Figure 3.6: Error Vectors

3.8 Collection of High Frequency Radar Data

HF radar data is openly available through the Coastal Observing Research & Development Center (CORDC) website (http://cordc.ucsd.edu/). The data provided by CORDC have a temporal range from October 1 2011 until the present date and can be provided in 500 metres, 1 kilometre, 2 kilometres, 6 kilometres spatial resolution grid cells provided in a NetCDF format. A NetCDF file is formatted so larger amounts of data take up less space. The way a NetCDF is designed is shown in Figure 3.7, it uses an x,y,z to index values where x,y are space coordinates and z is the time coordinate. The HF radar data provides information on the longitude and latitude of each grid cell, a time value, hourly average eastward wave velocity, and hourly average northward velocity. The time span chosen for the HF radar was the range that matches the S-AIS data which ranges from November 1 2012 00:00:00 until November 1 2013 00:00:00. The resolutions downloaded for use of the case study were 1 kilometres (Figure 3.8), 2 kilometres (Figure 3.9), 6 kilometres (Figure 3.10). These three were downloaded because although the higher resolution data would be preferable the 500m resolution data was only sparsely populated while the lower resolution, though not as precise, was more complete within the study area.

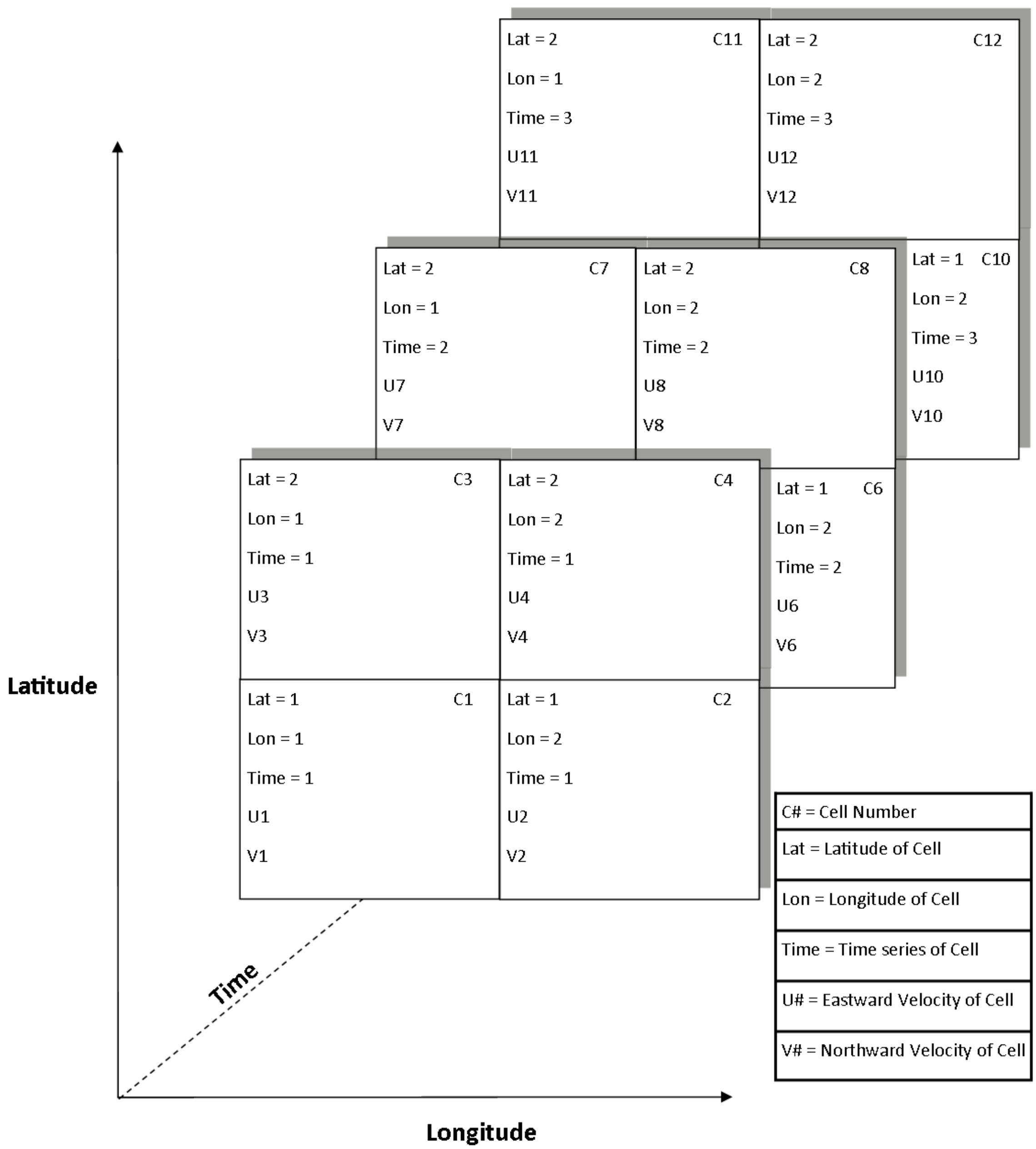

Figure 3.7: Structure of NetCDF

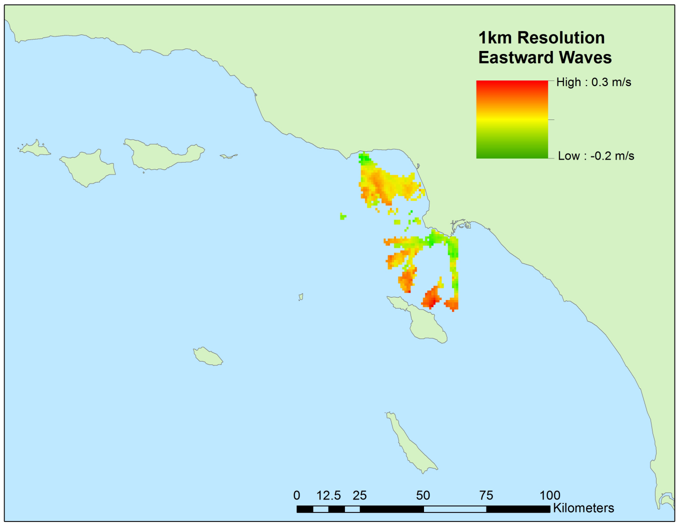

Figure 3.8: 1 km resolution HF radar data

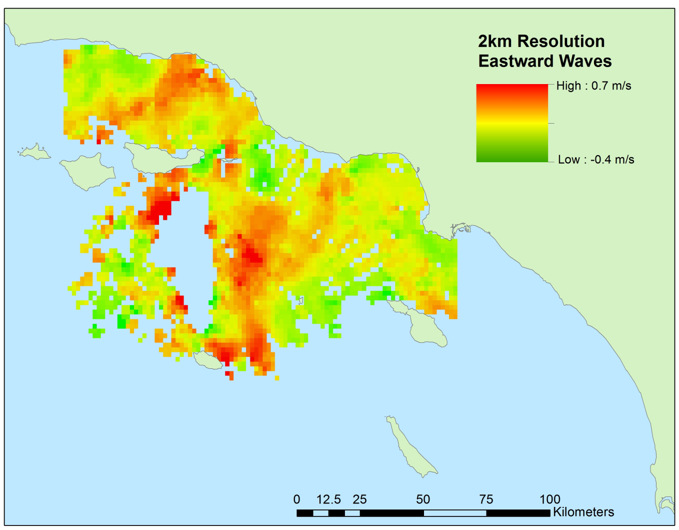

Figure 3.9: 2 km resolution HF radar data

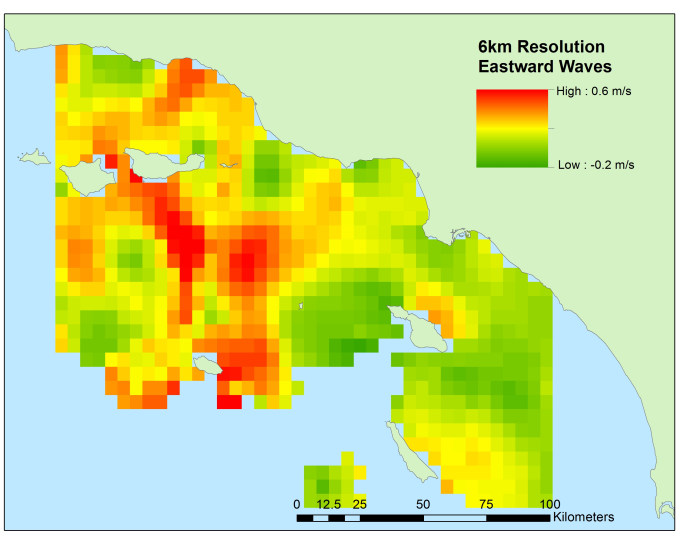

Figure 3.10: 6 km resolution HF radar data

3.9 Spatial and Temporal Join of the Ship and HF Radar Data

To spatially and temporally join the S-AIS messages and the HF radar data a few common attributes needed to be be properly matched. The first was a time relation between the two datasets because each had their own unique style of time reference. The HF radar data's time scale is single increments of hours since November 1 2011, to create a common relation with the S-AIS message dataset the time field of the S-AIS messages has to be converted into epoch time, seconds since January 1 1970, and then converted to seconds since November 1 2011 by removing the difference in time. Then the time was converted to hours and rounded down so that it is represented in the same time scale as the HF radar data. This is completed using the R code below.

```
###Converts date to seconds since Jan/1/1970###
ships$time_conv<- as.POSIXct(ships$time, "%Y-%m-%d %H:%M:%S")

###Converts to hours since 2011-10-01 (365952 is hours from 1/1/1970 - 10/1/2011)###
ships$hour_Conv<-round(((as.numeric(ships$time_conv)/3600)-365952), digits=0)
```

where 3600 represents seconds in an hour and 365952 is the number of hours between January 1 1970 and October 1 2011,

When the time relation is set for a temporal join the next consideration is the relation of the position of the S-AIS messages and to the HF radar grid cells for a spatial join. The join will provide the northward and eastward velocities that were present for each S-AIS message. The HF radar data used for this study consisted of 1 kilometre, 2 kilometres, and 6 kilometres resolutions of an eastward and northward hourly average velocity for each grid cell. To spatially join the HF radar data with each S-AIS message the latitude and longitude must be matched to the respective HF radar grid cell. The grid cells in the HF radar file are in ascending order of longitude and latitude starting point values for each cell. Using R's findInterval function two lists were generated, one for longitude and one for latitude, which provides the interval in which each S-AIS message is associated with spatially.

```
longCode<-findInterval(shipping$longitude,radar$dim$lon$vals, all.inside = T)
latCode<-findInterval(shipping$latitude,radar$dim$lat$vals, all.inside = T)
```

The wave velocities were then spatially and temporally matched using the latitude, longitude, and hour interval values.

This process is repeated for each resolution of data, 1 kilometres , 2 kilometres, and 6 kilometres for each message. The reason for this is that the coverage of each data set contains large spatial gaps so by merging the data resolutions it creates the largest possible coverage. The differences in values for northward and eastward velocities are negligible at the different resolutions, therefore by merging these resolutions there was no measurable impact on data consistency. The highest spatial resolution was taken if possible, if the data was not available then next possible value of the next highest resolution was selected.

This allows for the largest possible coverage of wave velocities to be applied to the S-AIS message data.

3.10 Correlation of the HF Radar and Error Vectors

To answer Question 1: Can top ocean wave currents account for message to message discrepancies in predicted position and actual recorded positions? A Pearson correlation coefficient was calculated between the generated error vectors and the wave magnitudes. This is done by using a Pearson correlation with the error vector and wave magnitude for the northward and eastward units. The next step was to try and isolate scenarios that were expected to show the strongest correlation values. The first subsets analyzed were ships that reported being at anchor which is represented by a S-AIS messages where the navigation status was equal to 1. For ships at anchor it is expected that the only movement in position must be from wave impacts. Then the ships that were moving were isolated into a subset. Next the data was split by the ship's rate of turn. It is expected that when a ship has a rate of turn it is actively moving against the motion induced by the waves or otherwise actively navigating so finding scenarios where no rate of turn is present will produce strong correlations. So a subset was created of the data based on whether a ship's message had a rate of turn equal to 0 or not.

The next type of scenario generated was based the length between messages. For this the thought was that when gaps in time occur between messages the overall magnitude of the errors are increased since the waves have longer to act on the ship over time. This would better represent the overall movement of a ship and would show a stronger correlation than smaller intervals in time would. This is done by selecting messages where the difference in time was greater than one minute and less than 10 minutes.

The next correlation test was done based on S-AIS ship's heading and the direction of the wave magnitude. To determine the wave's heading we used the function $\operatorname{atan2}(EV, EU)$ (where EV is northward wave velocity and EU is Eastward wave velocity). The wave magnitude was calculated using $\sqrt{EV^2 + EU^2}$. Cases were subset based on the wave direction being parallel to the ships heading ($\pm$ 5%) or if the wave direction was perpendicular ($\pm$ 5%). The correlation was then calculated between the error magnitude and the wave magnitude.

The final scenario considered was ship trips that contained the largest wave magnitudes. For this scenario trips were first subset based on the number of continuous messages received in sequence with no time gaps greater than 10 minutes and no errors in the messages that were mentioned in section 3.3. This used to select full trips where the number of messages in a sequence were greater than 50 and then out of those trips ones where the wave velocity was greater than 0.3m/s in either the northward or eastward was selected. The reason for the 0.3m/s selection is because this in the upper range of wave velocity in the selected dataset.

3.11 Comparison with Related Research

Following the study by Guichoux et al., 2016 we wanted to replicate the scenario generated in their study to see if their conditions of collected ship messages assisted in generating a greater correlation between the error vector and HF radar wave velocities. In the study the authors cite that the AIS messages used in the study “were gathered at a one minute sampling rate. For the study, we selected AIS messages of merchant ships with a speed over ground superior to 6 knots in order to exclude the impact of voluntary ship manoeuvres on our results” (Guichoux et al., 2016). These are the parameters we used to try and reproduce a similar data set.

To replicate a similar scenario in our data set the first step is to remove all message that have a speed over ground (SOG) less than 6 knots. The next task is creating time intervals of around 1 minute between ship messages. For this a time difference was calculated between the messages where Δt is t(i+1) - t(i). The messages that have a time difference greater then 120 seconds will be removed to minimize overall processing time. R’s difference function was used on t(i) and t(i+1) to calculate the difference (Δt) between the values. If the difference is less then 60 seconds then message (i+1) was removed and the change in time from t(i) was compared to t(i+2) until a difference in time is greater then 60 seconds. When that difference is met the new time message point starts where that greater than 60 second gap ended and repeat the loop from there. This helped isolate messages to make a scenario as if we only collected messages every minute and not have messages at much lower intervals of 2-10 seconds. From this point we rerun the generating of error vectors with this new subset to see if a correlation exists.

A final ideal case scenario was the expected ideal conditions where we expected a correlation to exist. This ideal case was a ship continuously moving at speeds greater then 6 knots so it can be assumed no drastic manoeuvres are being taken. The rate of turn (ROT) of the messages must be 0 so that we can assume the ship is not actively steering against the current to minimize the effect of drift. The time difference between messages was set no greater then 120 seconds so that no large gaps in time occurred in our continuous ideal scenario. The number of messages in sequence for this ideal trip must be greater then 50, and finally from these requirements the trip that consists of the strongest current velocity was chosen. It is under these conditions of a trip where there is no active manual manoeuvring against the current with a strong force that it is expected to have the greatest impact and therefore the strongest correlation to exist between the error vector and wave velocity.

3.12 Prediction Model Calibration

The final question to be answered was Question 2: Can we use HF radars to calibrate preliminary AIS-based location prediction models to enhance prediction accuracy? This was done by taking the preliminary prediction model that was created in this study (see figure 3.5) and incorporating the wave magnitudes into the prediction. This is done using 2 different models, the first model that was used to calibrate the error in prediction was,

$$\begin{aligned} Cx_i &= x_{i-1} + (\cos(\theta) \cdot V_{i-1} \cdot \Delta t) + (EU_{i-1} \cdot \Delta t) \\ Cy_i &= y_{i-1} + (\sin(\theta) \cdot V_{i-1} \cdot \Delta t) + (EV_{i-1} \cdot \Delta t) \end{aligned} \tag{3.4}$$

where EU is eastward velocity and EV in northward velocity. This was a simple change by adding the wave magnitudes directly to the model.

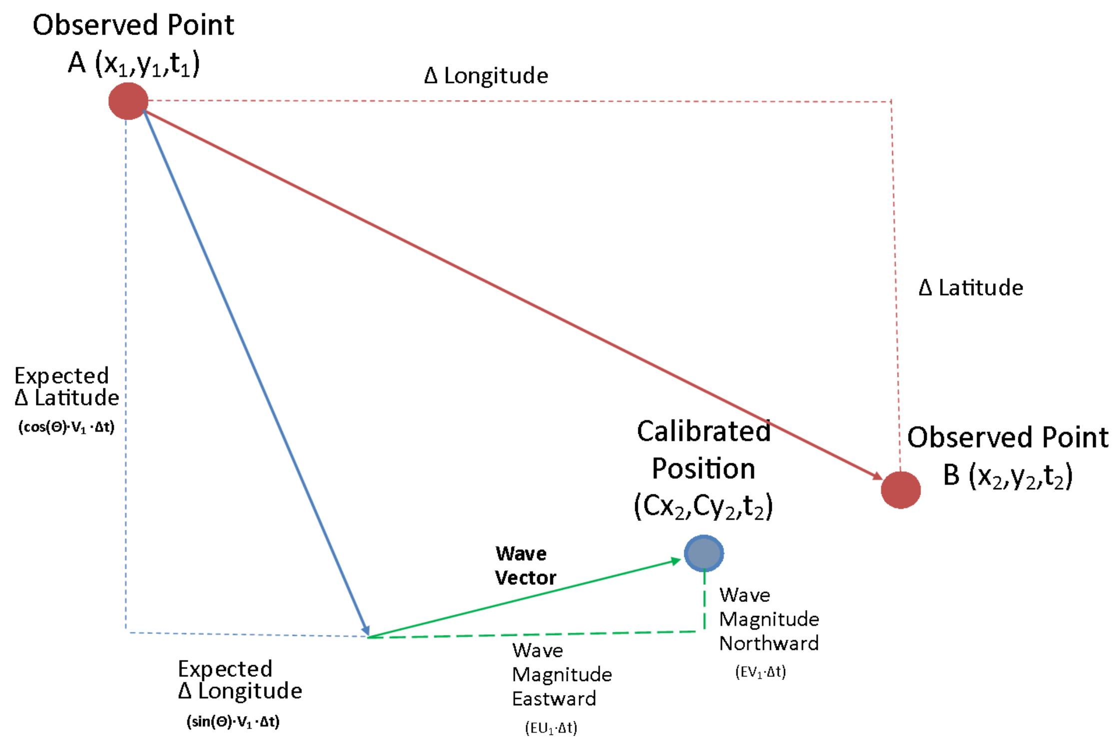

Figure 3.11: Calibration using wave magnitudes

Figure 3.11 visualizes how the wave vector is applied to the model in creating a new prediction location.

A further model was created where the wave's angle of interaction takes effect. The wave interaction was calculated separately for the wave acting perpendicular on the ship and the wave acting parallel on a ship. The maximum ship sizes that can use the the

Panama Canal, New Panama Canal, and Malaca Canal were taken to calculate a ratio of 7:1 from ship's length(7) and width/beam(1). This was done to incorporate the notion that waves would act in proportion based to the expressed surface area of a ship. The model that was created to represent this was,

$$\begin{aligned} RCx_i &= x_{i-1} + (\cos(\theta_{i-1}) \cdot V_{i-1} \cdot \Delta t) + (EU_{i-1} \cdot \Delta t \cdot (\cos(\theta_{i-1}) + \sin(\theta_i) \cdot \frac{1}{7})) \\ RCy_i &= y_{i-1} + (\sin(\theta_{i-1}) * V_{i-1} \cdot \Delta t) + (EV_{i-1} \cdot \Delta t \cdot (\sin(\theta_{i-1}) + \cos(\theta_i) \cdot \frac{1}{7})) \end{aligned} \tag{3.5}$$

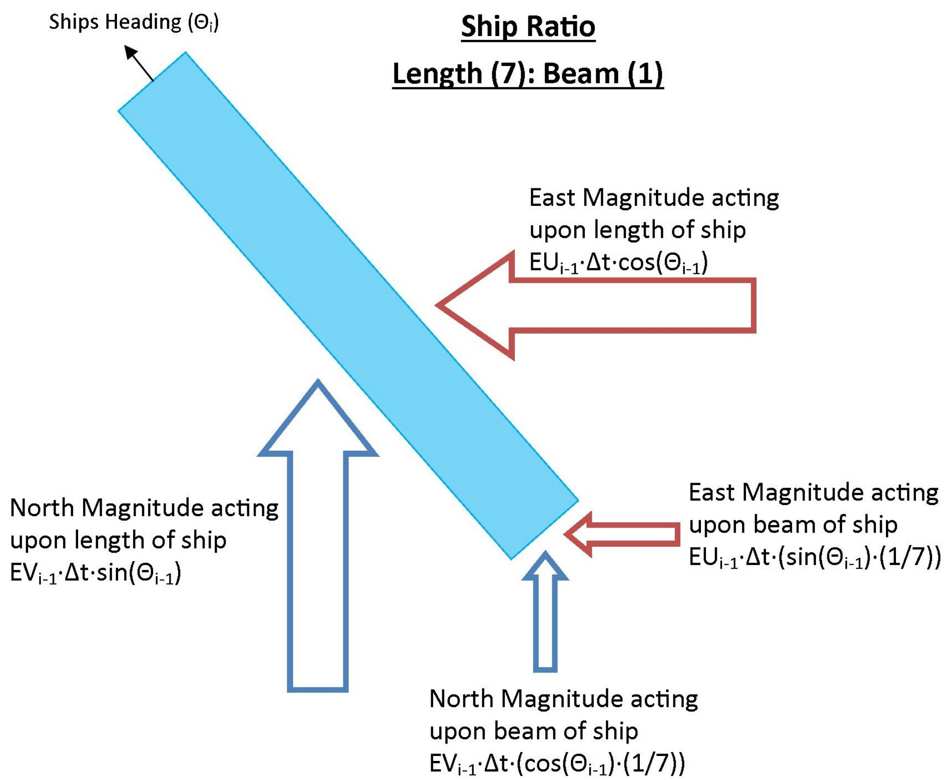

Figure 3.12: Realistic calibration of wave effects

Figure 3.12 gives a indication on how the wave is applied in the realistic calibrated model. The northward and eastward wave magnitudes are applied to both the length and beam of the ship separately so a wave has more contribution dependent of the surface area it comes in contact with on a ship. This is meant to simulate a more realistic approach on how a wave would act upon a ship where more surface area in direct contact with a wave would produce more foreceful displacement.

These models are run to generate new positional predictions and the overall error will be compared to the overall error that was produced in the preliminary model. This was done by comparing the following equations.

$$\begin{aligned}\sum_{i=1}^{n} Errx_i &= \sum_{i=1}^{n} \sqrt{(x_i - E_{xi})^2} \\ \sum_{i=1}^{n} Erry_i &= \sum_{i=1}^{n} \sqrt{(y_i - E_{yi})^2}\end{aligned} \tag{3.6}$$

$$\begin{aligned}\sum_{i=1}^{n} CErrx_i &= \sum_{i=1}^{n} \sqrt{(x_i - Cx_i)^2} \\ \sum_{i=1}^{n} CErry_i &= \sum_{i=1}^{n} \sqrt{(y_i - Cy_i)^2}\end{aligned} \tag{3.7}$$

$$\begin{aligned}\sum_{i=1}^{n} RCErrx_i &= \sum_{i=1}^{n} \sqrt{(x_i - RCx_i)^2} \\ \sum_{i=1}^{n} RCErry_i &= \sum_{i=1}^{n} \sqrt{(y_i - RCy_i)^2}\end{aligned} \tag{3.8}$$

Chapter 4

Results

The results of this study where designed to help determine,

Question 1: Can top ocean wave currents account for message to message discrepancies in predicted position and actual recorded positions?

and if a correlation exists, between the measured movement of ships through S-AIS messages and the wave velocities present from HF radars then determine if we

Question 2: Can we use HF radars to calibrate preliminary S-AIS-based location prediction models to enhance prediction accuracy?

The chapter will present the results. Next we discuss the steps used to remove messages that were not physically possible and exceeded speed limitations. The results from the gap analysis are then explained, detailing the number of gaps present in the dataset. Next the results from the spatial temporal join show how much of the S-AIS and HF radar data match. The results of the prediction model and the errors when compared to the actual positions are provided. The correlation between the wave magnitudes and the error in position is given for all the different scenarios. Finally the calibrated prediction models are used to generate new error vectors which were then compared to the errors in the preliminary model.

4.1 Study Area & High Frequency Radar Data

The S-AIS messages extracted from the SQL database were based on messages with longitude: (-118.2, -120.2) and latitude: (32.8, 34.8). This subset contained a total of 3,965,723 S-AIS messages that were transmitted from November 1st 2012 to November 1st 2013. Figure 4.1 is a kernel density estimation map that gives a representation of the location of S-AIS messages that were collected.

The HF radar data was downloaded in 1 kilometres, 2 kilometres, and 6 kilometres resolution from November 1st 2012 to November 1st 2013 in a study region of longitude: (-118.2, -120.2) and latitude: (32.8, 34.8). HF radar data of 6 kilometres resolution from

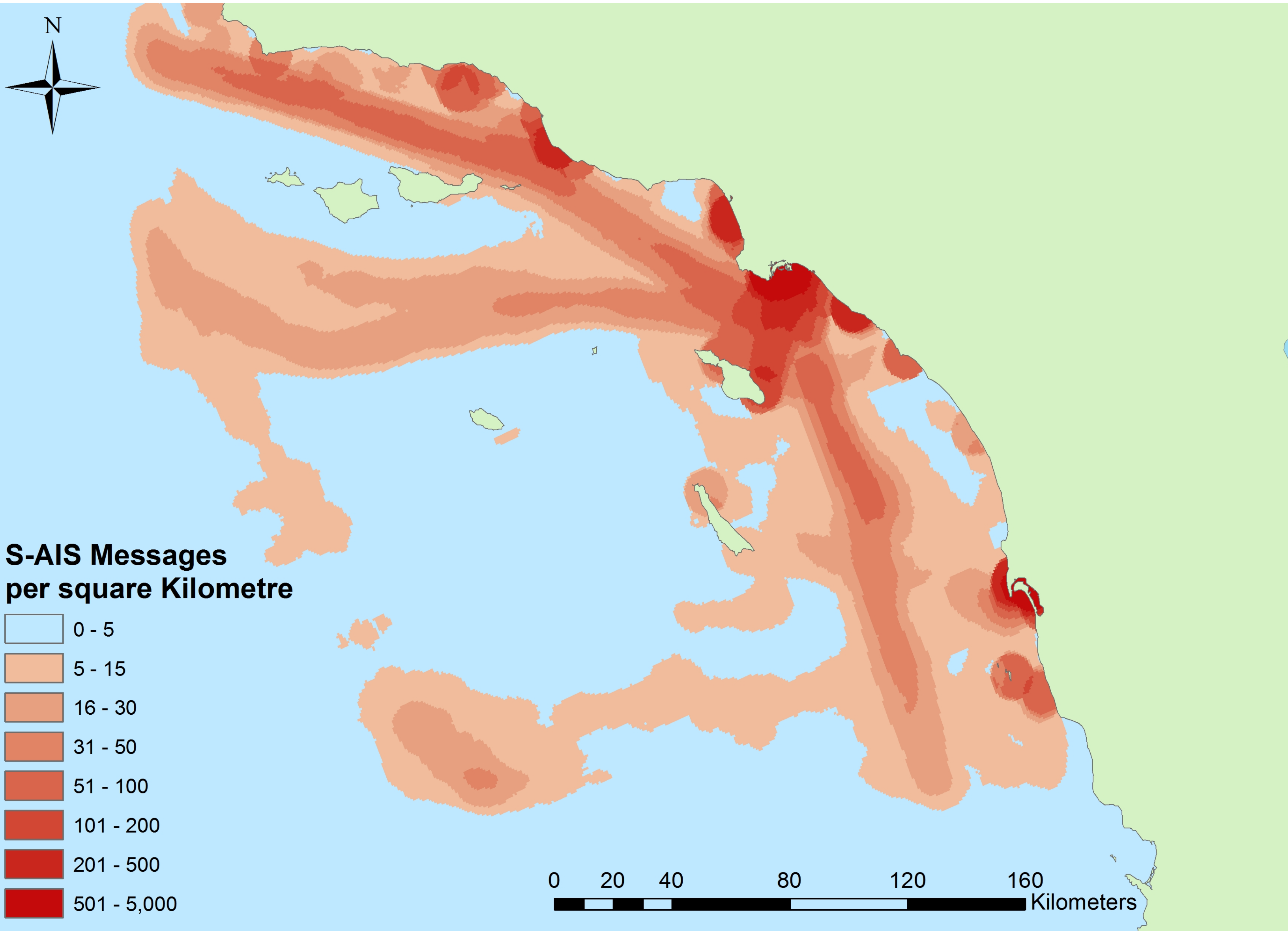

Figure 4.1: S-AIS Message Density form November 1st 2012 to November 1st 2013

the date March 3rd 20:00 onward was unavailable so the study's S-AIS messages were subset to follow this new time span of November 1st 2012 to March 3rd 2013 at 20:00 coordinated universal time. This subset had a total of 1,354,999 messages to be used for the study when combining the HF radar and S-AIS messages. These S-AIS messages were cleaned to remove messages where the MMSI was not in between 200000000 and 800000000 which removes 2.98 % of the messages. Then messages where the MMSI and time value were the exact same which consisted of 6.62 % of the messages were also removed leaving 1,224,886 messages for further analysis. This may have been a mistake to do and should remove messages on complete exact matches in all fields of the data. There may be important information being removed that could possibly help correct errors in the dataset that is detailed in section 5.1.1.

4.2 Physically Realizable Values

The physically realizable values were subset by finding values that were not possible for the ships in our study to accomplish. The first selection was finding S-AIS messages with a reported SOG greater than 50 knots. Of the 1,224,886 S-AIS messages analyzed there were 207 messages with a reported SOG of 50 knots. Through marinetraffic.com, an online database of ship AIS messages, the MMSI of these ships were selected and they all come from tug boats. All the tug boats are registered on MarineTraffic as having a max speed over ground of 7.2 knots making these messages of 50+ knots an obvious error.

The next scenario was the plausibility of the S-AIS messages position in accordance to its previous position and time. The distance between two continuous messages in a trip was calculated and if that distance exceeded the time between the two messages (Δ t) multiplied in seconds by 25 metres per second (50 knots). There were 4,535 messages that exceeded this limit meaning there is at least 2,267 separate cases where a S-AIS message was misreported and provided impossible values. An example of this is the sequence of messages in Table 4.1 which shows a sequence of three messages that span 1 minute and 30 seconds. In this sequence the S-AIS message says the ship travelled 198 kilometres in 11 seconds and returned to the near same position 1 minute later.

MMSI	Time	SOG	Longitude	Latitude	COG
367395280	2/10/2013 7:08:23	0	-118.2566	33.76407	333.4
367395280	2/10/2013 7:08:34	6.8	-118.6933	32.01549	353.8
367395280	2/10/2013 7:09:53	0	-118.2566	33.76406	12.2

Table 4.1: S-AIS Message Error

4.3 Gap Analysis

A gap analysis was completed by using the S-AIS messages in the study over the entire year which consisted of 3,965,723 messages. This was to allow for a larger set of data compared to using only the messages that were between November1 2012 and March 2 2013. First we looked at time gaps greater than 6 hours in a sequence of messages for a single ship. There were 126,894 messages that appear over 6 hours since the last reported S-AIS message which is 3.2% of the message dataset. The next issue was those messages in which the gap between messages were a distance greater than 5 kilometres. There were 33,930 messages in the data set where a ship's reported S-AIS message was collected over 6 hours after the previous message and the distance reported between the 2 messages was equal or greater than 5 kilometres. This occurred in 0.86% of all messages in the data set. An example of a gap is represented in Table 4.2. Between these two messages there was a gap in time of 9 hours and 27 minute and the ship had travelled more than 22 kilometres. The first message reported a navigational status of 0 meaning the it was using its engine while the second message report 5 meaning it is moored (tied to port). This meant there was obvious movement of the ship between the time of the two messages where the AIS message was unreported or not received by prevailing satellites.

MMSI	Time	SOG	Longitude	Latitude	Navigational Status
354357000	4/8/2013 10:23	10.8	-118.1463	33.55324	0
354357000	4/8/2013 19:50	0.2	-118.2556	33.73239	5

Table 4.2: S-AIS Message Error

4.4 Spatial Temporal Join

The spatial temporal join between the S-AIS and HF radar data provided the information on the wave velocities that were present at the time the AIS messages was received by the satellite. This spatial temporal join was performed on 1,200,390 S-AIS messages, these messages represented all messages from November 1st 2012 to March 20 2013 that had all anomalous messages extracted and had a reported navigational status. After the completion of the spatial temporal join between the S-AIS messages and the HF radar data there were 875,165 S-AIS messages that had a measured wave velocity based on the HF radar data. This equates to 72.91% messages being associated with a corresponding wave velocity. From this join between HF radar and S-AIS messages there were 77,391 messages that used 1 kilometre resolution HF radar data, 157,863 messages that used 2 kilometres resolution data, and 639,911 messages that used 6 kilometres resolution data.

4.5 Trip Identification

The creation of trip identifiers groups the S-AIS messages into trip segments to allow for comparison between trips. Since the only requirements to determine which trip each S-AIS message was associated with was the MMSI, navigational status, and time all S-AIS messages were used with a total of 875,165 messages. As defined in Section 3.6 a trip is a sequence of messages with in-motion navigational status between two end points of messages with at-rest navigational status. Table 4.3 is a transition sequence from a sequence of S-AIS messages changing from one trip into the start of another with the start of the second trip highlighted in blue. Figure 4.2 is a map of these S-AIS messages transitional sequence between the trips where a ship is travelling, goes to port then continues on a new trip. The purpose of this is to show how when a trip is transitioning there are gaps in space and time that occur which can cause issues when analyzing, especially when transitioning to a new trip. There are also gaps in time that occur during a a ship's trip, the first message of these gap types are highlighted in beige in table 4.3.

MMSI	Time	Navigational Status	Longitude	Latitude	TripID	SOG	Heading	Sequence
354942000	11/2/2012 8:16	0	-118.6431	33.58934	354942000T1	10.1	91	1
354942000	11/2/2012 8:16	0	-118.6425	33.58931	354942000T1	10.1	91	2
354942000	11/2/2012 8:16	0	-118.6421	33.58929	354942000T1	10.1	91	3
354942000	11/2/2012 9:01	0	-118.4895	33.58573	354942000T1	9.7	89	4
354942000	11/2/2012 10:32	0	-118.2347	33.6169	354942000T1	9.6	57	5
354942000	11/2/2012 19:11	5	-118.2021	33.74356	354942000T1	0	90	6
354942000	11/2/2012 19:17	5	-118.2021	33.74356	354942000T1	0	90	7
354942000	11/2/2012 19:35	5	-118.2021	33.74356	354942000T1	0	90	8
354942000	11/2/2012 21:17	5	-118.2021	33.74357	354942000T1	0	90	9
354942000	11/3/2012 6:23	5	-118.2021	33.74353	354942000T1	0	90	10
354942000	11/3/2012 7:42	5	-118.2021	33.74355	354942000T1	0	90	11
354942000	11/3/2012 7:59	5	-118.2021	33.74355	354942000T1	0	90	12
354942000	11/3/2012 9:29	5	-118.2021	33.74348	354942000T1	0.1	90	13
354942000	11/3/2012 20:05	5	-118.2021	33.74354	354942000T1	0	90	14
354942000	11/3/2012 20:35	5	-118.2021	33.74355	354942000T1	0	90	15
354942000	11/3/2012 20:38	5	-118.2021	33.74358	354942000T1	0	90	16
354942000	11/4/2012 1:41	5	-118.2021	33.74359	354942000T1	0	90	17
354942000	11/4/2012 6:11	5	-118.2025	33.74345	354942000T1	0	90	18
354942000	11/4/2012 20:18	5	-118.2021	33.74356	354942000T1	0	90	19
354942000	11/5/2012 5:44	5	-118.2021	33.74357	354942000T1	0	90	20
354942000	11/5/2012 5:53	5	-118.2021	33.74353	354942000T1	0	90	21
354942000	11/5/2012 7:20	5	-118.2021	33.74356	354942000T1	0	90	22
354942000	11/5/2012 8:44	5	-118.2021	33.74359	354942000T1	0	90	23
354942000	11/5/2012 15:47	0	-118.7549	33.65145	354942000T2	11.7	274	24
354942000	11/5/2012 15:49	0	-118.7627	33.65186	354942000T2	11.7	273	25
354942000	11/5/2012 15:49	0	-118.764	33.65193	354942000T2	11.7	273	26
354942000	11/5/2012 15:49	0	-118.7653	33.65201	354942000T2	11.7	273	27
354942000	11/5/2012 15:50	0	-118.7666	33.65208	354942000T2	11.7	273	28
354942000	11/5/2012 19:02	0	-119.7649	33.67559	354942000T2	19.7	304	29
354942000	11/5/2012 19:03	0	-119.7662	33.67614	354942000T2	19.7	308	30
354942000	11/5/2012 19:25	0	-119.8875	33.74339	354942000T2	20.7	279	31
354942000	11/5/2012 19:26	0	-119.8958	33.74451	354942000T2	20.9	280	32
354942000	11/5/2012 19:27	0	-119.9019	33.74536	354942000T2	20.8	280	33
354942000	11/5/2012 19:28	0	-119.9061	33.74599	354942000T2	20.8	280	34

Table 4.3: S-AIS Message Error

Figure 4.2: Transition of a trip in S-AIS messages

This example identifies inconsistencies with the data when attempting to analyze. There are gaps in time between in motion messages as high as 3.2 hours and even longer gaps occurring between in motion and anchored messages. These gaps in time show the need for understanding the predicted position of a ship when it is unclear how long until the next message is received using S-AIS. In this example we illustrate that when a trip is labelled for this study and by starting a trip with the first reported in motion navigational status, which for this case is message 24, we avoid the complications with the initial transition from anchored to moving. “At anchor” messages are included in at the end of the trip and remain associated to the trip but are subset into separate trip segments in further analysis.

4.6 Prediction Model and Error Vectors

From the S-AIS messages there were 875,165 that have a spatial and temporal pairing with a HF radar wave value. In creating the prediction model from the messages one of the required fields in the S-AIS messages is the ship’s heading. All messages that contained a heading value that was not null were subset leaving 574,907 messages that can be used for the prediction model and generate error vectors.

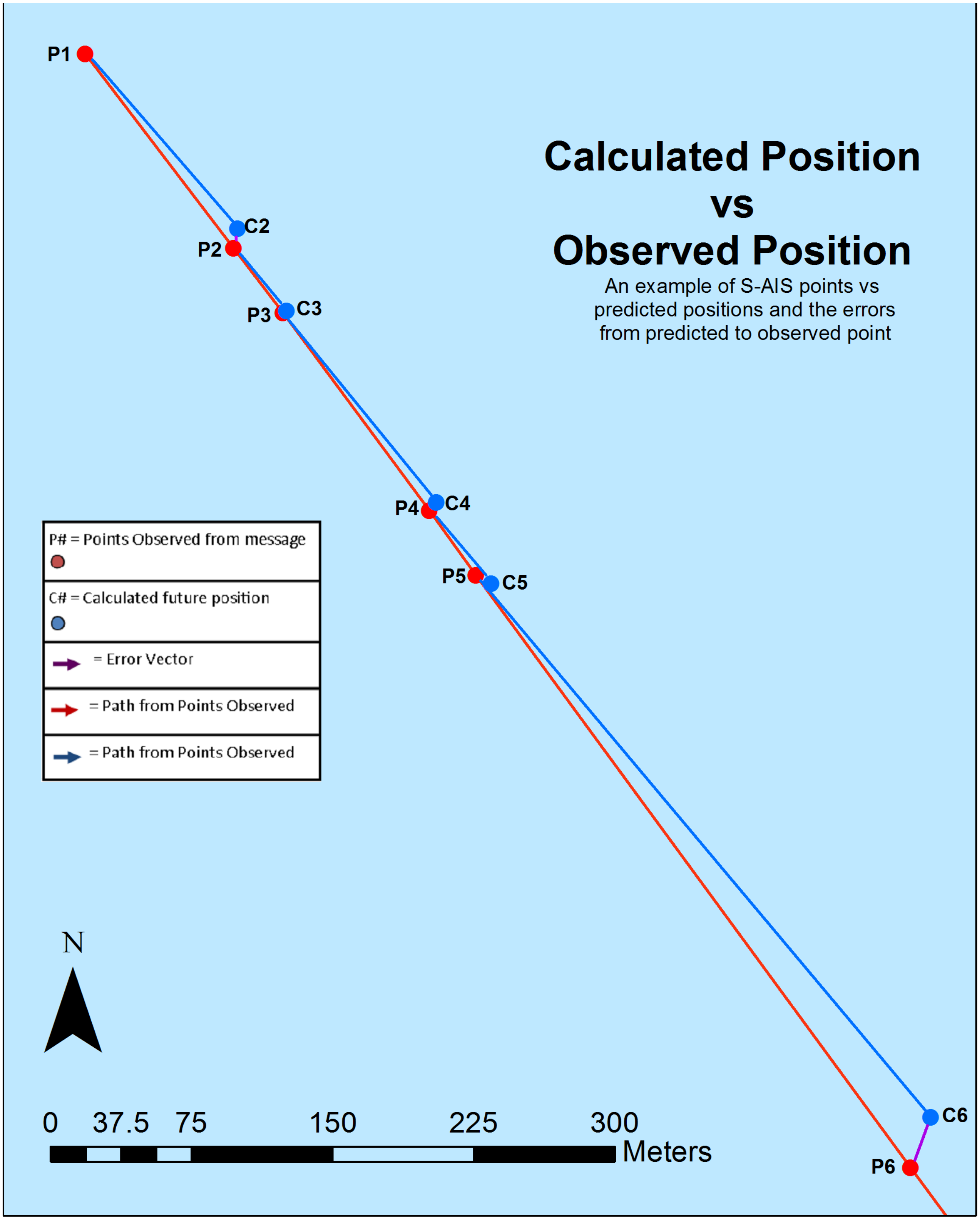

Figure 4.3: S-AIS prediction model and generated error vector map

Figure 4.3 is an example of the predction model and the generated error vectors. In the prediction model the average error between S-AIS messages and the predicted location were 582 metres on the north axis and 660 metres on the east axis with the average time between messages being 1422 seconds and speed of 3 metres per second. These values may be skewed by outliers so when looking at only moving ships the average error between S-AIS messages were 576 metres on the north axis and 654 metres on the east axis with an average time difference of 1043 and average speed of 4.7 metres per second. This shows there was not a significant change when isolating moving ships in the average error values for the initial prediction model.

4.7 Correlation Values

A regression analysis was completed to calculate the correlation between the prediction error vector and the wave vector acting on a ship over time. This was done using the Pearson correlation coefficient. These correlations were broken down based on northward and eastward magnitudes.

This initial test compared all S-AIS messages that had wave magnitude data paired with them as well as removing large time gaps of 24 hours (see Table 4.4). From this regression analysis there was no correlation present.

Correlation test on all messages	**n=872,440**
Northward Correlation	0.010899
Eastward Correlation	-0.000367

Table 4.4: Correlation for all prediction errors and with matching wave magnitudes, time gaps greater than 24 hours removed

Table 4.5 split the S-AIS messages based on the navigational status to determine if a difference is present between ships at anchor and those with an active engine. The expectation would be that there would be a stronger correlation with ships at anchor because the major force applied to its drift is waves. There was no correlation in either scenario and this may be because of the gaps in S-AIS messages present with ships at anchor which an example can be viewed in Table 4.3.

Correlation test split based on navigational status	
At Anchor n=18,855	
Northward Correlation	-0.0143
Eastward Correlation	-0.07036
Under way using engine n=655,188	
Northward Correlation	0.05095
Eastward Correlation	0.004013

Table 4.5: Correlation for all prediction errors and with matching wave magnitudes split based on the ships navigational status

Table 4.6 split the datasets further based on if the S-AIS had a reported rate of turn (ROT). The reason for this split was because the expectation was that ships that were actively moving direction may be doing so to actively negate the effects that may be causing drift on a ship's course. In the data there are multiple cases where in a continuous string of S-AIS messages have alternating reported ROT of 10 and -10 which lead us to this assumption. From the test we see slightly stronger correlations with S-AIS messages that have no rate of turn but the values are not significant.

Correlation test splitting based on navigational status and rate of turn	
At Anchor	
No Rate of Turn n=6,623	
Northward Correlation	-0.06788
Eastward Correlation	-0.3111
With rate of turn n=12,232	
Northward Correlation	-0.009644
Eastward Correlation	-0.05457
Under way using engine	
No Rate of Turn n=113,365	
Northward Correlation	0.1538
Eastward Correlation	-0.00512
With rate of turn n=541,823	
Northward Correlation	0.026283
Eastward Correlation	0.006289

Table 4.6: Correlation for all prediction errors and with matching wave magnitudes split based on the ships navigational status and split again based on rate of turn

Since there was a slight increase in correlation of S-AIS messages with no reported ROT this was split further in Table 4.7 that viewed only S-AIS messages that had a time difference from the previous message between 1 - 10 minutes. This was to determine if time gaps allowed for greater displacement in prediction which would generate more significant correlation values. There was no significant correlation from this subset.

Correlation test split based on navigational status and selecting messages with time gaps of 1-10 min	
At Anchor n=1,756	
Northward Correlation	-0.05177
Eastward Correlation	-0.1554
Under way using engine n=12,275	
Northward Correlation	-0.07499
Eastward Correlation	-0.004408

Table 4.7: Correlation for all prediction errors and with matching wave magnitudes split based on the ships navigational status and selecting cases where time gap is 1-10 minutes

Table 4.8 looked at partial ship trips where there were at least 50 continuous messages with time gaps less than 10 min and all reported wave velocities of 0.3 m/s which is strong in this area. This was an attempt to try and isolate a case where wave correlation would have the greatest opportunity to exist. This subset reported a small negative correlation but is not substantial.

Correlation test on ship trips that had high waves present and long continuous trip sequence	**n=6,478**
Northward Correlation	-0.20078
Eastward Correlation	0.02774

Table 4.8: Correlation for all prediction errors and with matching wave magnitudes, messages for this test must contain wave velocity greater than 0.3m/s, time gap less that 10 min, and a continuous ship trip sequence of 50.

The next subset was to look at S-AIS messages with a reported moving navigational status and split based on the ship's heading and the wave direction in Table 4.9. The split was based on if the the wave impacted the ship along the path of movement or perpendicular to its movement. This was to see if there is a difference in correlation if the wave was acting on the ship's width or length. The criteria for angle of interaction was given a $\pm$ 5°. From this subset there was again no correlation present.

Correlation test split based on the angle between ships heading and wave direction	
Parallel angle (0° or 180° $\pm$ 5°) between heading and wave n=32,597	
Northward Correlation	0.03855
Eastward Correlation	0.00844
Perpendicular angle (90° or 270° $\pm$ 5°) between heading and wave n=32,995	
Northward Correlation	-0.0616
Eastward Correlation	0.017

Table 4.9: Correlation test based on the ships heading orientation and wave angle, split based on if angle is parallel to ship movement or perpendicular

Table 4.10 repeats a test to try and replicate the scenario presented by researcher Guichoux et al., 2016 for their study. Time gaps of 1-2 minutes were simulated as intervals between messages. Then only S-AIS messages with a reported SOG of 6 knots or greater and 0 ROT were selected. From this criteria of data the regression analysis showed no consistent correlation.

Replication of E-Odyn message conditions	n=21,330
Northward Correlation	-0.14825
Eastward Correlation	-0.02367

Table 4.10: Correlation test based on S-AIS messages with a moving navigational status, speed over ground greater than 6 knots, simulated time intervals between 1-2 minutes, 0 rate of turn, and a continuous message sequence of 5 messages

The final case that was set up to try and find a significant correlation was in Table 4.11. This ideal case selected the parameters where there would be an expected strong correlation with prediction error between S-AIS messages and the prevalent wave magnitudes. The parameters for this test were ships with a SOG of 6 knots, time intervals less than 2 minutes, and have a continuous messages sequence of at least 30. This ideal case was meant to isolate partial ship trips of linear moving ships that are not actively moving and have a consistent string of S-AIS messages to reference. Even with these ideal parameters there was no significant correlation present.

Ideal case scenario of expected correlation	n=53,273
Northward Correlation	-0.02914
Eastward Correlation	-0.01888

Table 4.11: Correlation test based on the ideal scenarios of ship trips that would be expected to have a strong correlation between prediction error and wave magnitudes. The ideal cases were segments where when the S-AIS are moving, speed over ground greater than 6 knots, a ROT of 0, time intervals less than 2 minutes, and have a continual sequence of 30 messages.

These Pearson correlation coefficients measure the linear correlation between the error vector of the prediction between S-AIS messages and the wave magnitudes that took place between the S-AIS messages.

4.8 Calibrated Prediction Models

Using the prediction model to determine ship position from an S-AIS message, error vectors were generated based on the positional difference of the predicted location to the reported S-AIS position. These error vectors were calculated in all three model calibrations to be used to determine the overall accuracy in the predictions.

From Table 4.12 we see a comparison of the overall errors produced based on each calibration type. The prediction error increased overall with the calibrated model compared to the initial prediction model by nearly 1%. The overall error of the directional calibrated model minimized the overall error when compared to the initial model by 4%. Based on these values, even though the overall error did decrease when using the directionally calibrated model, the change in error is small and not reliable based since no correlations were present between the error and wave vectors.

Model Type	Sum Absolute North South Error (m)	Mean Error North South Error (m)	Sum Absolute East West Error (m)	Mean Error East West Error (m)
Prediction Error Model	333,097,373	582	377,701,393	660
Calibrated Prediction Error Model	336,092,772	591	377,599,883	664
Directional Calibrated Prediction Error Model	320,636,744	564	361,477,102	636

Table 4.12: A comparitive look at the difference in overall error in the prediction model, calibrated prediction model using wave magnitudes, and directionally calibrated prediction model using wave magnitudes and directional impact on ships orientation

Chapter 5

Discussion

Through the course of this thesis the aim was to determine,

Question 1: Can top ocean wave currents account for message to message discrepancies in predicted position and actual recorded positions?

and if a correlation exists, between the measured movement of ships through S-AIS messages and the wave velocities present from HF radars then

Question 2: Can HF radars be used to calibrate preliminary S-AIS based location prediction models to enhance prediction accuracy?

In this chapter the results from the previous chapter are interpreted to look at the finding of the thesis. Discussed are the errors that were found in S-AIS messages, the problems they may cause, and how this issue can be addressed. Following this the results from the prediction model and how accurate the initial prediction model are looked at. Then the thesis will look at answering if top ocean wave currents account for message to message discrepancies in predicted position and actual recorded positions. These results are used to then determine if HF radars can be used to calibrate preliminary S-AIS-based location prediction models to enhance prediction accuracy.

Next will discuss changes that can be made to the study for further research. Changes can be made to the prediction model to enhance the accuracy in determine the ship's position using developing ship models. Then discussed are the ways to expand the wave dataset to gain a greater field of view of the areas around the AIS messages. Then the thesis goes over ways to implement near real time data in the analysis of ship prediction models so that it may be implemented for decision support systems.

The last part of the discussion is future uses of the research found in this study. This will look into the process of dynamic ship pathing and how this research can be beneficial. The next part looks at the benefits this thesis may be in the field of search and rescue. Finally, discuss the possibility of enhancing current prediction models using the information gathered in this research.

5.1 Interpretation of Results

5.1.1 Errors in S-AIS

With any dataset there comes issues with data quality and reliability, this is no different with S-AIS messages. In this study the first thing to be done was cleaning of the data, described in section 3.2, which reviewed the removal of S-AIS messages that had invalid MMSI value or were a repeated message that had the same MMSI with the same time stamp. It turns out that in reviewing in detail the process of S-AIS system sampling the AIS data it may have been incorrect to remove this data and they in fact should have been cleaned differently because it may have removed messages that reported a different position that could correct for misreported messages. The reliability of the data came into question when analyzing the real world possibility of the messages occurring. By calculating the distance in position and difference in time between 2 messages in sequence S-AIS messages from a ship we were able to calculate the speed of that ship according to the S-AIS message. A speed max of 50 knots (92.6km/h) was set as our threshold and from this we were able to find 2,267 S-AIS message cases that went past this threshold. This is important to find in the dataset so that it may be removed from the model since these values can greatly skew results by having ships travel much farther then possible over time meaning the prediction model error will far exceed expectation.

A specific abnormal case detected using this method was from a ship with the MMSI 235414000. In this case we had normal sequence of S-AIS messages that did not exceed any limits in space over time until one moment there is a gap in time of 7 minutes 39 seconds (highlighted in blue). Over this time the S-AIS messages reported a positional change of 30.18 kilometres which makes the detected average speed of 236km/h between the S-AIS messages. This change in position is detected in other cases but as seen in table 4.1 they normally return to the same position and the single message can be seen as an error. What makes MMSI:235414000 unique is the following sequence of S-AIS message continue from new position. This is shown in figure 5.1

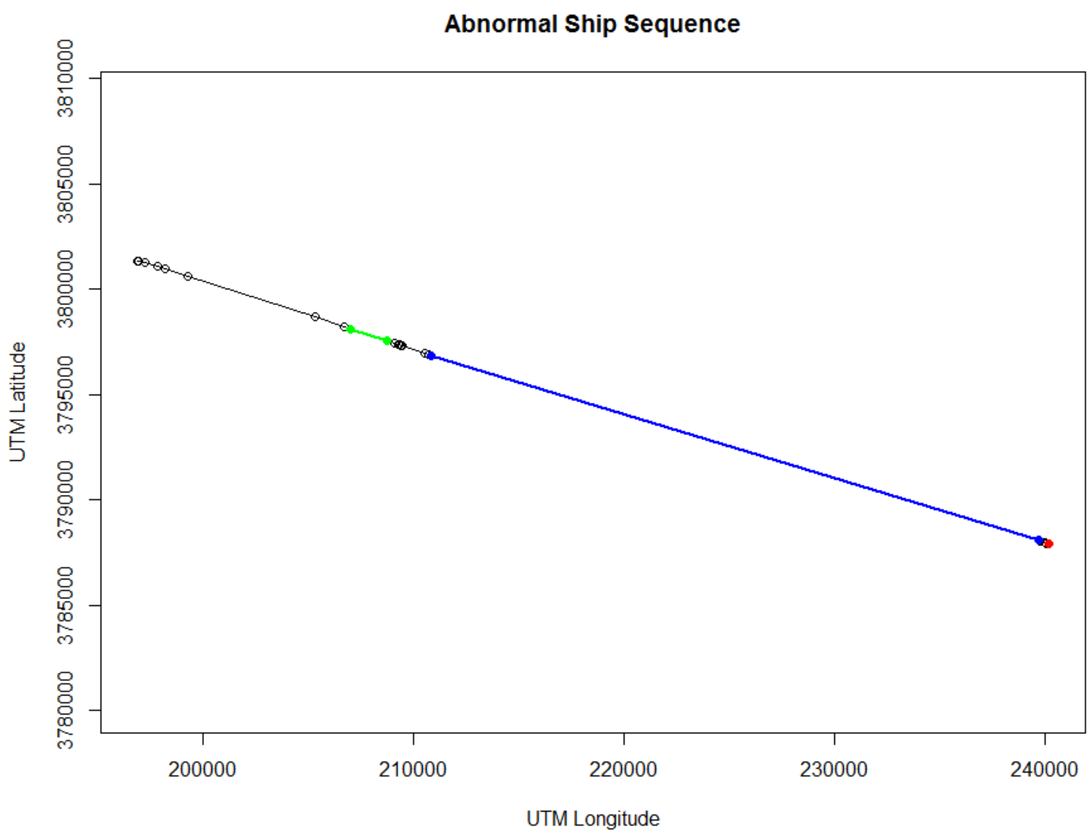

Figure 5.1: Trip sequence where messages continue after a gap occurs

The uniqueness of this case goes further, after a sequence of 10 messages following where the gap in space, highlighted in green, occurs there is a delay in time of 1 hour where the ship only travels 1.8 kilometres even though the reported speed over ground is roughly 32 km/hr. This is type of error differs from the typical types selected from the gap analysis which are commonly, one time, misreported positions spikes. This error may have been produced by miscommunication between the AIS and satellite making this a specific S-AIS error type. The issue might be with the time stamps reported by the AIS or from time stamps generated by the satellite and then self corrected after the next reliable message was received. The way the gap analysis is designed for this study it still detected the gap in space and identified the second S-AIS message after the gap in space as the start of a new trip. This would allow us to still use the gap analysis accurately for the prediction model with the exception of the time gap delay correction being undetected. The methods in determining gap analysis and identifying a ship trip is a beneficial tool when analyzing and working with S-AIS as it can help identify errors like that shown in figure 5.1.

The gap analysis results showed there are gaps in S-AIS message data sets so there is a need to predict a ships movement to fill in these missing message gaps that can help identify a ships position in the past when no S-AIS messages were received. This is evident with seeing that 3.2% of the messages in the dataset occurred after a time greater than 6 hours in the message sequence. This is the maximum amount of time that the satellite constellation would be absent from an area. Of all the S-AIS messages 0.86% have the following message reported after more then 6 hours and travelling a distance greater than 5 kilometres. This shows that the ship was moving over this gap of time and the satellite was unable to directly receive a message by the ship either because the ship's AIS was not sending signals or the satellite was unable to collect the message. This gap analysis helped identify these gaps and where prediction is required to fill in the gaps of positional information. From the results reported in sections 4.2 and 4.3 we see that there is still misreported messages and gaps in the data that need to be either identified or filled in.

5.1.2 Prediction Model

The prediction model that was created for this study was meant to create basic positional predictions to generate error vectors to be compare against the wave velocities. The error vectors that are produced overall are large at an average error of 582 metres northward and 660 eastward with an average time between messages of 1422 seconds. There are still large errors generated that may be skewing these values because of missing messages that were either unreported or not received. If we look at only messages with maximum time difference of 1 hour we see how it may have been skewed. With a time difference less than 1 hour the average error is 38 metres northward and 37 metres eastward with and average time difference of 127 seconds and average reported speed of 3.2 metres per second. The average distance between messages was calculated to be 257 metres away which makes the prediction average accuracy off by 20%. The prediction model created for this study needs to be redesigned in further testing to maybe use existing models like that created by Guichoux et al., (2016) as the basis to enhance initial accuracy and produce greater reliability in the prediction errors.

5.1.3 Does a correlation exist between prediction error and waves?

The first question of this thesis was to determine **Can top ocean wave currents account for message to message discrepancies in predicted position and actual recorded positions?** The way this was tested was by comparing the relation between the errors in the prediction model and the wave magnitudes that took place over the prediction. The relation was completed using the Pearson correlation coefficient and the data was broken into subsets based on the movements of a ship as well the different ways a wave may interact with it the S-AIS messages were broken into categories to find where the waves may have the strongest relation. These correlations were also broken down based on Northward or Eastward relation. Ideally through these tests the expected outcome would for there to be a strong negative correlation when running the regression analysis between the error vector and the wave magnitude. This is based on the knowledge that if a ship was predicted to travel further north then what was later reported it is thought that waves travelling south assisted in this new outcome causing a negative correlation.

The initial regression analysis showed no consistent or significant correlation in either eastward or northward directions so the first subsets were split based on the navigational status of the ship. The trips were split based on if the S-AIS message navigational status reported as ***under way using engine*** versus ***at anchor***. This was to see if a ship at anchor may have slight drift caused by only external forces but this showed no correlation. Next these subsets were split further based on the reported rate of turn. The assumption was that a ship actively turning may be negating the effects caused by waves, this may have been the case for ships at anchor since the messages with no rate of turn showed a small negative correlation shown in Table 4.5. The next test looked only at messages that had a time gap of 1-10 minutes. This was to see if maybe a significant correlation is present when a greater amount of time passes and a greater error vector is produced. This showed no correlation from the 14,031 S-AIS messages that were tested.

To try and create ideal condition where we expect a correlation to exist we subset S-AIS messages that were reported in areas that had wave velocities greater than 0.3 metres per second. The messages also must have had a time gap less than 10 minutes and be in a ship trip sequence of messages greater than 50. From these conditions there was was a slight negative correlation northward but was showed no correlation eastward. The subset may be having some larger values skewing the data and not be representative of the common message that occur in the dataset. This would need to be analyzed in future studies to perhaps better represent common messages in the data. The next correlation test was to see if there is a relation based on how the wave impacts a ship. The messages were split based on if the waves angle of impact was at the bow or stern of the ship versus the waves angle of impact being on the port or starboard of the ship. Neither of these situations produced a significant correlation.

Guichoux et al., (2016) state that they generate wave predictions based on similar error vectors. This would mean that error vectors should have a direct correlation with the wave velocities so the S-AIS collection method used in that research was simulated from the dataset used in this study. For this the S-AIS messages must report a navigational status of ***under way using engine***, speed over ground >6 knots, time intervals of 1-2 minutes were simulated, reported rate of turn of 0, and be part of a sequence of at least 5 messages. From all of these conditions there were 21,330 messages and there was still no correlation in either northward or eastward. This lead to the final correlation test where we generated what are the ideal conditions for a correlation to exist between the error vectors and wave magnitudes. It met the same conditions as the previous case except there was no simulated 1-2 min intervals but instead no time gaps greater than 2 minutes, and the sequence of messages must be greater than 30. After all these conditions were met there is still no significant correlation present overall. This was a surprising result since it would be expected that the error from the predicted location, that is generated by speed over ground and heading, to the next reported message would somehow be correlated with the waves present in the area.

There were 1,195 unique partial trips that met all the set conditions. Below is a frequency distribution of the correlations in these partial trips. Figure 5.2 shows correlation northward of the partial trips and has a bit of a uniform distribution but with more negative correlation cases present. Figure 5.3 which presents the correlation eastward of the partial trips has a normal distribution around 0 but with a tail in the negative correlation.

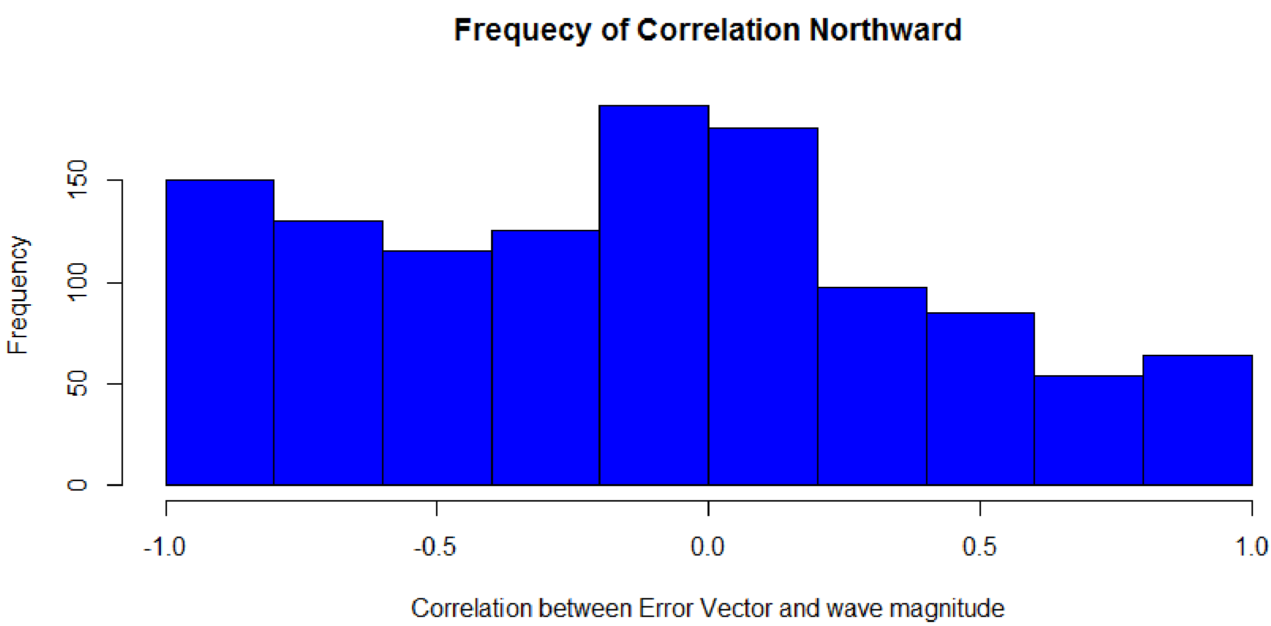

Figure 5.2: Frequency of correlation values for ideal partial trips

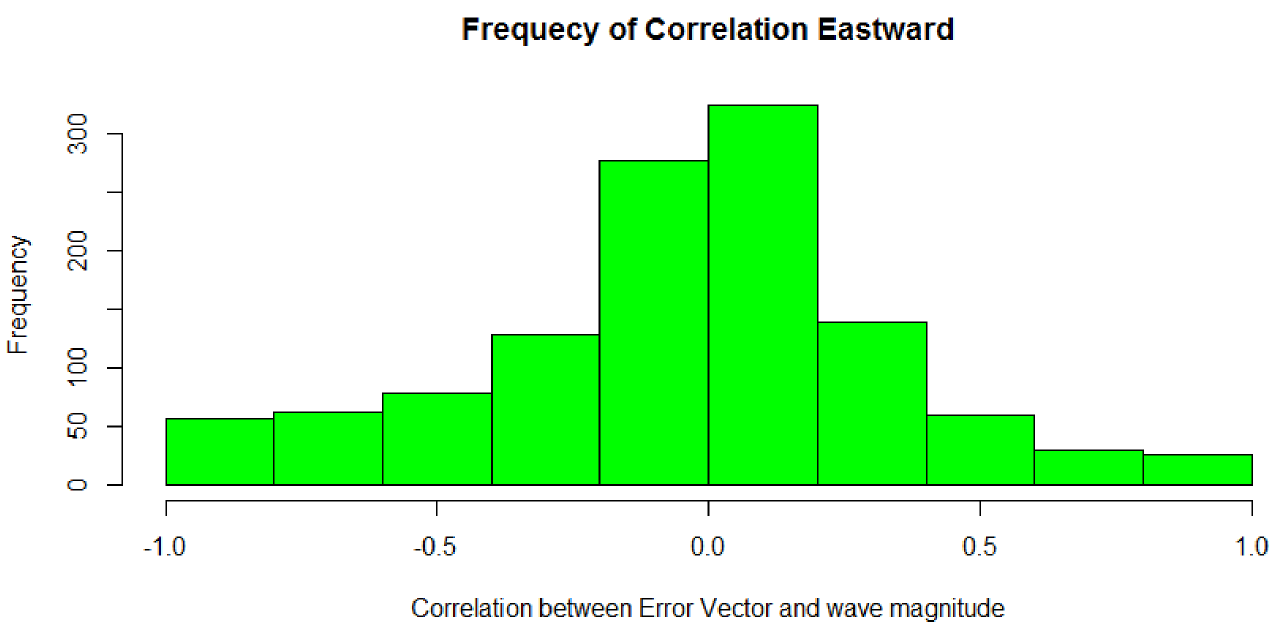

Figure 5.3: Frequency of correlation values for ideal trips

From the ideal cases that were generated we looked at 3 different correlation scenarios. From these cases it was expected for them all to report strong negative correlation since they met the ideal conditions for waves to have the greatest impact on the ships movement.

Figure 5.4 is a partial trip where the ship is travelling northwest at and average speed of 4.2 metres per second. In this case the waves were travelling northeast at 0.143 metres per second. When compared to the errors in the prediction between messages and wave magnitudes there was a strong negative correlation northward at -0.923 and eastward correlation at -0.740. This is the ideal scenario that was expected when first conducting the study but Figure 5.5 & 5.6 that is not the case.

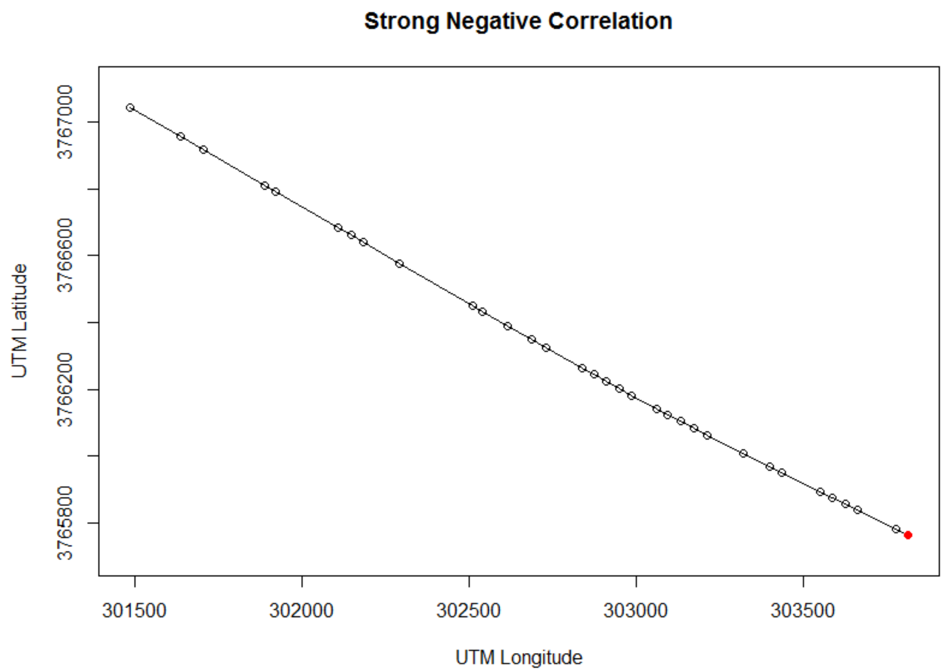

Figure 5.4: Partial trip of ideal case where strong negative correlation exists

Figure 5.5 is similar to 5.4 with the initial conditions but had a greater average speed of 7.8 metres per second. The average wave velocities during this partial trip were south at -0.35 metres per second. The correlation between message to message error and wave magnitude were a strong positive correlation northward at 0.98 and small positive eastward at 0.26. These results show cases were the predictions in ship position were further south then the reported message even though the waves were actively pushing south.

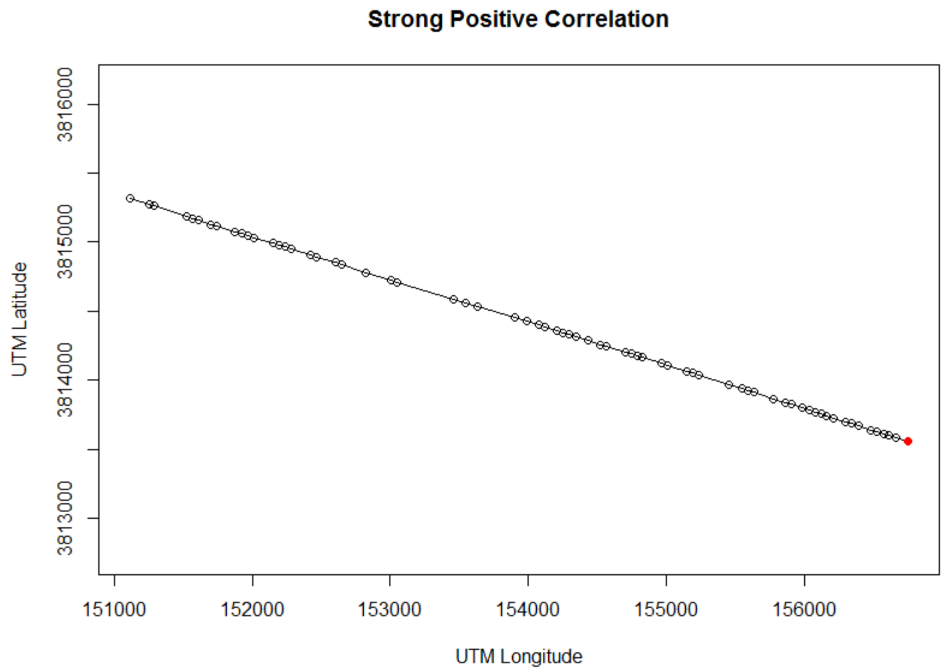

Figure 5.5: Partial trip of ideal case where strong negative correlation exists

This last figure is of a case where the waves were very strong with a south west waves at a velocity of 0.63 metres per second which is one of the highest reported in the dataset. From this case the correlation was measured to be 0.03 northward and 0.2 eastward which represents no overall correlation between prediction error and wave magnitudes. This ideal case had the strongest waves where we expected that to reflect in the drift from prediction but no correlation was measurably present.

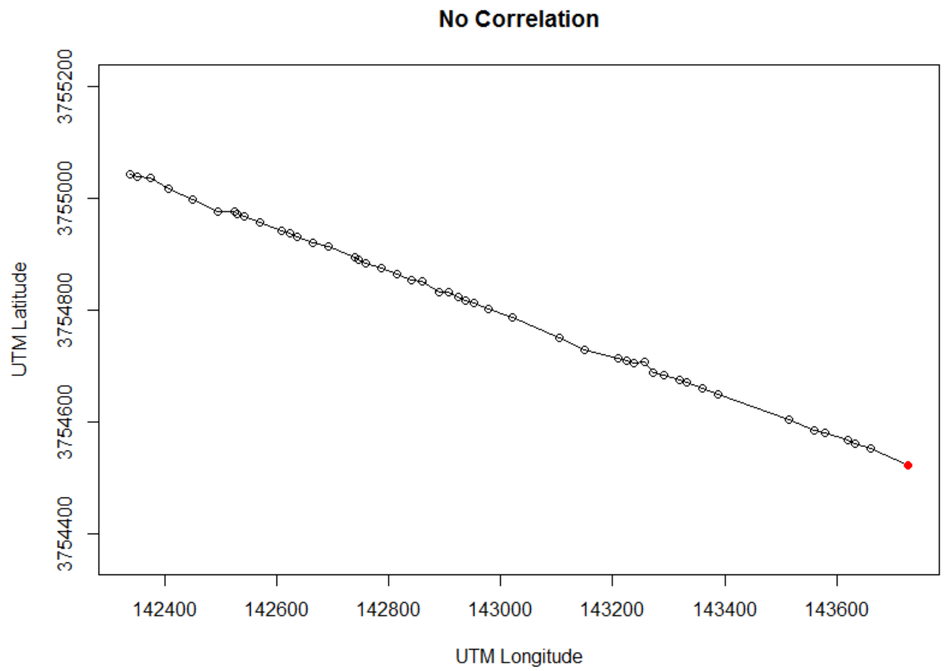

Figure 5.6: Partial trip of ideal case where strong negative correlation exists

Even though all the presented cases showed similar conditions they all presented different correlation values between error in prediction and wave magnitude.

5.1.4 Can we use HF radar data to calibrate the prediction model to minimize error?

From the results of the regression analysis between error vectors and wave magnitudes it was determined that no consistent correlation was present. This makes using HF radar wave velocities not as reliable as we initially thought when attempting to calibrate the prediction model but it was still completed to show if it helped minimize error in prediction to any degree. The first model calibration directly adds the wave magnitudes to the prediction of the ships movement. This direct addition was to be later applied with a weighted factor based upon the results of the regression analysis but since no correlation was found this step was not used. Another calibration was applied that took into effect the angle of impact on a ship with the speculation that a wave that hits the broad side of a ship, therefore more surface area, would have a greater effect then waves that hit the ship straight on. When looking at Table 4.12 we see that the initial calibration did not change the overall error to any significant degree and even increased the overall error on the north axis. The directional calibrated prediction error model did however lower the overall error by 3.74% on the north axis and by 4.3% on the east axis. From the dataset used in the prediction model the average time difference between S-AIS messages was 1422 seconds and an average speed of 3 metres per second. This means the average distance travelled was 4266 meters and with the initial prediction being 880 metres off from the reported messages position. With the use of the directional calibrated prediction model that average error was minimized to about 840 metres. Due to no correlation being detected from our tests we currently can not state significant findings in the calibration of the prediction model.

5.2 Changes to Study

There are changes to the study that could supplement the results and benefit the research. By looking at improvements to the prediction model, wave dataset, and using a near real time updated database this study could better analyze the effects that waves may have on a ship and improve the prediction models.

5.2.1 Enhancing the Physical Model of Prediction

One part of this study was to create a preliminary model to predict the ships location after a given amount of time. This is an important step in the analysis because the deviation of these predictions compared to the ship's next reported position that is considered to be the error that is to be attributed to external forces like waves. The predictions that were made used only the ship's reported speed over ground and heading making the prediction susceptible to errors. This study was set to enhance the model using the information gathered with the HF radar correlations but it may require a more comprehensive base model. Researcher Borkowski, 2017 explains the trajectory prediction algorithm used on board ships for decision support for navigators. The techniques used in this study involve data fusion of multiple sensors on board a ship, implementing this information and S-AIS data you could develop a more complex prediction algorithm with the expectation of enhancing positional accuracy prediction. An example is the creation of an ANN which is trained by "analyzing longer data strings ... of a steady movement (non-manoeuvring vessels)" (Borkowski, 2017). The trained ANN continues to calibrate its weighted factors to enhance the accuracy in its prediction explained in section 2.3.1.

Another way to enhance the model would be further in-depth corrections of the physical model of a ship's movement. Perez, 2007 studies sea-keeping theory, the study of ship motion where waves are present, and used this information to alter kinematic models used in sea-keeping. These models were described as able to be a "basis for models that can be used in guidance and motion control systems" (Perez, 2007). Sea-keeping theory is based on a ship keeping constant speed and course so may only be applied to certain sections of a trip in this study but may be applicable for an enhanced prediction model. By selecting S-AIS messages that report a constant speed and course and using sea-keeping theory we can ideally create a model with greater accuracy then the one developed in this study. This means that the error in prediction would be more influential in the correlation tests that take place and potentially showing a significant correlation.

5.2.2 Expanding the Wave dataset

For this study the wave data was obtained from the Coastal Observing Research and Development Center (CORDC) with the collection done by Integrated Ocean Observing System Program (IOOS) which uses WERA (Wellen Radar) HF radars to collect the data. The accessibility of the data and range it provided allowed for a large spatial coverage of wave velocities off of the US coastline. The issue with using this data set is the resolution with which the data is accessible. The resolutions used for this study were 0.5 kilometres, 1 kilometre, 2 kilometre, and 6 kilometres but problems occurred due to the varying spatial extent of available data for each of the resolutions. The 0.5 kilometre resolution was not available in the study region that was selected and was only provided in one small region in the Bay of San Fransisco. The 1 kilometre resolution data also only covers small and selective areas, for this study it did cover areas up to 30 kilometres off the coastline within the study area. The 2 kilometre and 6 kilometre resolution was capable of covering nearly the entire spatial extent of the study area with issues only occurring due to the HF radar signals being obstructed by geological barriers such as Santa Catalina Island located south west of Long Beach where no messages south west of that point were able to be collected for this study.

To address some of these issues in future studies the use of data from NOAA's National Data Buoy Center (NDBC) may be of use which uses a system of ocean buoys to collect local data on wave and wind. The wave data that is collected from these buoys are significant wave height, dominant wave period, average wave period, and mean wave direction (National Data Buoy Center, 2017). The information from the NDBC may supplement the available HF radar data to give a more accurate representation of a wave's impact on a ships movement. Heij and Knapp, 2015 and Szelangiewicz, 2014 both look at risk assessments associated with wave height and how it impacts the way a ship has to travel to avoid any incidents of potential risk. An additional benefit specific to the data provided by the buoys is the local accuracy of the information and the fact that is not being interpolated from surrounding data. The need for localized data for waves is one of the main reasons for the research by Guichoux et al., 2016 who state that there is not much data provided at a finer scale so they are unable to track "mesoscale structures [such] as eddies or filaments that can escape the altimetry techniques." (Guichoux et al., 2016). Using this information may prove to be beneficial in analyzing small segments of an S-AIS ship trip that are relatively near the buoy. Calculating the error in prediction for these select cases using accurate localized wave information, instead of the broader spanning averages used in this study, may provide a higher correlation with the waves and the error then was determined in this study.

5.2.3 Near Real Time Implementation

The data for this study were selected based on their rate of update which is near real time. AIS's are designed to send a message from every 2 seconds to 2 minutes depending on the ship's manoeuvres and navigational status. Due to cost effectiveness of LEO satellite launches a broader constellation of satellites for AIS message detection is developing. This leads to fewer S-AIS message losses and moves towards real time collection of all AIS messages. HF radar data is collected between 4 to 18 minutes and aggregated to an average over an hour so the shortest period to update the HF radar data is 1 hour. The change that this study could benefit from is a near real time update of S-AIS message and HF radar data to a SQL database. Incorporating a near real time update feed of both S-AIS and HF radar data would allow the data set to continuously update and allow a near real time prediction of where a ship may be based on the most recently received S-AIS message.

5.3 Future Uses

5.3.1 Dynamic Ship Pathing

A potential use of this study is dynamic ship pathing, in which a ship's path could adapt to present wave conditions to lower fuel or time consumption. An early paper by Vlachos, 2004 developed the idea of determining optimal ship routes based on wind and wave forecasts for either saving time of the trip or prioritizing the safety of the trip. This idea was advanced by Dolinskaya, 2012 with the use of ship mounted radars for close proximity wave detection for optimal path finding. This study ran into an issue of needing to use a "rolling horizon" (Dolinskaya, 2012) method of pathing caused by the range limitation of the on board radar. The limitation was determined to be 10 minutes worth of travel time which could be assisted with some aspects of this study. By incorporating some of the techniques of spatial and temporal join with nearby HF radar the data they provide can assist a further optimal path calculation. Incorporating the wave data available to further expand a ship's wave field could allow for an optimized path to be generated. The path optimization can assist in either lowering the travel time of a trip or reduce the fuel consumption of a ship based on the specified priority. Fuel saving ship routing is determined by Chu et al., 2013 who used meteorological and oceanic forecasts to achieve this. Introducing an updated near real time database feed to a ship in this situation can assists in dynamic routing based on changing weather conditions instead of a static forecast determined prior to ship departure. Fuel efficient pathing can assists in minimizing green house gas as well as save ship operators money since the “maritime sector is likely to experience high fuel prices in part due to increased oil scarcity and demand for oil from developing countries but also due to the introduction of sulphur regulations and increasingly stringent emission constraints” (Smith et al., 2013).

5.3.2 Search and Rescue

Vessel tracking is a crucial tool for search and rescue (SAR) missions as they try to pin point a vessel’s exact location for first responders. “Risks at sea are once again on the rise, thus demanding an evolution in vessel traffic management systems (VTMS)”(Zissis et al., 2015) to help assist human operator’s decision making capabilities. The addition of near real time data of related waves and the ship’s forecasted trajectory can assist in minimizing risk of a ship in vessel collision prevention. When a situation occurs where a vessel requires SAR, time is an important factor requiring quick response. ”Knowing which vessels are located in a specific area, their state vector properties (e.g., position, velocity and direction) ... are essential to many governance responsibilities in the open sea, in particular to Search and Rescue (SAR) and security operations” (Vespe et al., 2015). Incorporating near real time HF radar and S-AIS message data of the ship will assist in more accurate position estimates of a ship allowing first responders to more quickly reach the destination.

5.3.3 Calibration of Currently available vessel prediction

There are currently several different vessel tracking systems being used to determine a ship's current location and future position. By creating a spatial temporal join with the ship's S-AIS and recent HF radar data the newly available data may be able to enhance current models to increase prediction accuracy with further development. Moreira et al., 2007 developed a guidance control system for autonomous ships based on a ship's line of site distance. Autonomous guidance is important for marine vehicles for reliability and lower transportation costs (Moreira et al., 2007). By providing the shipping model with additional information on wave conditions it provides useful information in generating a path in a spatial decision support role. Last et al., 2014 stated that "AIS data needs to be integrated with further data to allow for a reliable state estimation of vessels in matters of collision avoidance [models]" (Last et al., 2014) this expresses the need for spatial and temporal matching of S-AIS with other data to assist in prediction modelling.

5.4 Summary

From the results the thesis was unable to find a significant correlation and could not significantly lower the overall error in out prediction model. What was found though are new techniques to process and analyze S-AIS messages to detect errors and further express the gaps in information we have in S-AIS data. Finding the different types of errors in S-AIS is important and we helped show errors that are caused by the AIS, e.g. mis-reporting a message or mis-classifying ship information, or the LEO satellite, .e.g. incorrect time stamps or mixing up ship positions, which is important to identify when analyzing. Steps forward were provided to help in furthering this research, by enhancing the base model we may find a significant correlation with ship drift and waves which would allow for improvements to the study. We may also expand the wave datasets by including additional sources outside HF radars. Buoy's have been used to track wave velocities and may provide more accurate fine scale information that could assist in the analysis. This also leads into the further implementation of a near real time prediction model that could use a data feed of S-AIS and HF radar data to track the predicted movements of ships. The discussion also went over future uses this study may assist in with regards to dynamic ship pathing which could help determine fuel efficient paths that use wave velocities. Search and rescue rely on recent positional data for finding a ship that is in need of assistance so by enhancing the prediction in position in any way is beneficial. The methods used in the spatial temporal join of S-AIS and HF radar may be used in currently available vessel prediction models to provide more information vessel prediction that can help if we are able to later correlate HF radar and S-AIS.

Chapter 6

Conclusion

This thesis was designed to look at a way to create a ship trajectory prediction model using S-AIS messages and further enhancing the accuracy of these predictions using HF radars. This merging of datasets led to the first question that was **Question 1: Can top ocean wave currents account for message to message discrepancies in predicted position and actual recorded positions?** This information was necessary to further the research and answering **Question 2: Can we use HF radars to calibrate preliminary S-AIS-based location prediction models to enhance prediction accuracy?** This study was unable to find any correlation with the error vectors in the prediction model and the prevailing waves but was able to show techniques and information for further research to potentially address these issues.

Past literature was first reviewed to understand the scope of AIS, S-AIS, HF radar, and prediction models. Then limitations of the data and data reliability was looked into, this information became helpful when trying to identify certain types of errors especially with mis-reported messages that caused message spikes as shown in Figure 3.2. Previous recent studies were also useful in trying to replicate conditions that had favourable results in their research but these did not produce the same results when attempted in this paper. This can be attributed to not having a similar prediction model which can significantly change results or not having direct access to AIS data (only S-AIS being utilized).

Next the thesis laid out the methods and results to answer the questions proposed. The steps used in cleaning and identifying errors became of interest because there were more errors found then initially expected.The thesis was able to identify errors that were generated by the AIS and the receiving satellite making S-AIS not completely accurate when compared to real world position of a ship. This was expressed in Section 5.1.1 which identified errors present in S-AIS messages. It further shows a need for more S-AIS research to find solutions so errors like this can be corrected. The next step was creating a spatial temporal join between the S-AIS messages and HF radar NetCDF files. The methods used in this study can be helpful for researchers who look to merge these datasets in the future and comparatively use the information in a efficient manner. The steps used to process the data became important because of the size of the dataset so it is important to reference appendix A when joining the data sets or identifying ship trips in other datasets. Next the prediction model was generated which predicted a ship's future location using the last reported heading and speed. From these predictions error vectors were calculated based on the distance from the next reported position to the predicted position. These error vectors were then compared with the waves present during the time of the message in a regression analysis to see if a significant correlation exists and answering **Question 1** of this paper. Multiple conditions were set where a significant correlation was expected to exist but were unable to identify a situation were this occurred. Even though no significant correlation was found we still went forward in applying the model calibration to show how it may be done if a correlation were found in later studies. The calibration incorporated the HF radar data in two ways. The first calibration directly added the wave magnitudes that took place and the second calibration used the angle of impact that the wave interacted with a ship in the idea that interacting with the broadside of a ship has more surface area therefore more force acting upon the ship. From the calibrations the second approach was able to decrease overall error in prediction by 3.74% on the north axis and by 4.3% on the east axis. Since there was no measurable correlation between error vectors and wave magnitudes we do not have significant confidence in these calibration values.

Lastly the thesis reviewed changes that can be made to the study and future uses. For advancement of this study an enhanced prediction model must be designed to improve the initial accuracy of the prediction. This is important because the error vectors that are generated from the model are key to the regression analysis with wave magnitudes. These could assist in determining the relation between wave magnitude and prediction error. The wave dataset would also benefit from incorporating more information from additional sources. In addition to HF radar data, buoy wave velocity could be added to the data set to give precise wave velocity data at distinct points to enhance accuracy at a finer scale for some areas. The final change to the study would be the implementation of near real time S-AIS and HF radar data. This updated feed could assist in providing more information to the study and allow the design of a near real time updated prediction system that incorporates HF radar data.

This leads to the future uses this study may be of assistance to with the design of dynamic ship pathing. In receiving updated wave velocities a ship may use wave velocity to assist in movement to a destination and save fuel over the course of the trip. This study could also be beneficial in field of search and rescue. Joining new information to S-AIS message data that could assists in prediction accuracy would allow for aid to a ship in need of assistance at a faster rate with knowing a more precise position. Finally the methods of message error detection and joining of HF radar data may assist in the calibration of currently available vessel prediction models. Incorporating additional information would allow users using the vessel prediction model to view an additional factor that may be altering the accuracy of the positions outcome.

In this study we were unable to find out if top ocean currents account for message to message discrepancy or be able to calibrate a location prediction model using HF radar data with confidence. What was determined though was ways to detect certain errors and cleaning of S-AIS data. This type of detection is useful before analysis is undergone so errors do not skew results for future researchers. It also identifies a problem in S-AIS ,misreported messages and impossible data values, that must be looked into to help solve the issues causing these errors. This study also helped create tools that can be used to in the spatial temporal join between S-AIS and HF radar data so other researchers can pair the data for their research. Even though in this paper we were unable to calibrate a S-AIS based prediction model reliably with HF radar data it is able to assist in future research based on the tools and methods produced for S-AIS and HF radar analysis.

Appendix A

R Code

```
### Revisioned Source COde for cleaning dataset, correlating HF and AIS, creating error vectors,
# April/20/2017 BKF
# July/27/2017 BKF
# October/19/2017 BKF
# January/15/2018 BKF
###############################
setwd("E:/Masters/LongBeachTest/")

#### Installing all packages ####
install.packages(c("ncdf4","sp","rgdal","readr"))
library(ncdf4)
library(sp)
library(rgdal)
library(readr)

#### Reading in Radar and Ship data ####

#ships <- read.csv("Masters/LongBeachTest/longbeach.csv",header = TRUE, sep = ",")
ships <- read.csv("longbeach.csv",header = TRUE, sep = ",", stringsAsFactors=FALSE)

radar <- nc_open("RadarData/RTV_HFRADAR,_US_West_Coast,_1km_Resolution,_Hourly_RTV_best.nc")
radar2 <- nc_open("RadarData/RTV_HFRADAR,_US_West_Coast,_2km_Resolution,_Hourly_RTV_best.nc")
radar3 <- nc_open("RadarData/RTV_HFRADAR,_US_West_Coast,_6km_Resolution,_Hourly_RTV_best.nc")


#### Subsetting the ship data to only contain relevant coulumns ####
Keep <- c("mmsi", "time", "navigation_status", "rot", "sog",
          "longitude", "latitude", "cog", "heading")
ships<- ships[Keep]

rm(Keep)

#### Converting Ships time to Correlate with Radar Data which is Nov/1/2012/00:00:00 - Nov/1/2013/00:00:00 ####
ships$time_conv<- as.numeric(as.POSIXct(ships$time, "%Y-%m-%d %H:%M:%S"))
ships$hour_Conv<-round(((as.numeric(ships$time_conv)/3600)-365952), digits=0) #Converts to hours since 2011-10-01 (365952 is hours from 1/1/1970 - 10/1/2011)


ships<-ships[complete.cases(ships$mmsi),]

### Limitations on data variables ###

# Proper latitude and longitude boundaries (ships2)
ships<-ships[ships$latitude >32 & ships$latitude <35 & ships$longitude > -121 & ships$longitude < -117,]

# March 3rd 20:00 cut off where the HF radar corrupts
march3<-ships[ships$hour_Conv<12477,]

# Sort the dataset based on MMSI and Time
march3<- march3[order(march3$mmsi,march3$time),]

#Remove scenarios where MMSI is invalid (<20000000 or >80000000)
march3<- march3[march3$mmsi>200000000 & march3$mmsi<800000000,]

###Remove duplicate rows based on mmsi and time
march3<-march3[!duplicated(paste(as.character(march3$mmsi),as.character(march3$time))),]

#make arbitrary Number Key for organizing
march3$Key<- 1:nrow(march3)

#Remove where ships exceed expected liitation of 50 knots
march3<-march3[which(as.numeric(march3$sog)<50),]

#### Setting Physical Limitations based on distance gaps ###
#1 Knots -> 0.514444 m/s
march3$timeDiff <- c(0, diff(march3$time_conv))

#Setting UTM coordinates for the data
cord.dec = SpatialPoints(cbind(march3$longitude, march3$latitude), proj4string = CRS("+proj=longlat"))  #Collecting the Long Lat points of the data
cord.UTM <- spTransform(cord.dec, CRS("+init=epsg:26711")) #Transforming points to UTM by Code for UTM zone 11 Nad 83
march3$UTMlog<-cord.UTM$coords.x1
march3$UTMlat<-cord.UTM$coords.x2

rm(cord.dec)
rm(cord.UTM)
```

```
# Distance between message cant exceed physical capabilities
march3$distLat <- c(0, diff(march3$UTMlat))
march3$distLon<- c(0,diff(march3$UTMlog))
march3$distDiff<- sqrt((march3$distLon^2)+(march3$distLat^2))

#Remove Values where mmsi is the same but distDiff exceeds time and speed possibilities
march3$mmsidiff<-c(0,diff(march3$mmsi))
march3$removable<- march3$mmsidiff==0 & (march3$timeDiff*25) < march3$distDiff

march3<-march3[march3$removable==FALSE,]

march3<-march3[,-(19:20)]



#### Trip Identification Number ####

#Changing ship data to a list and setting up values for loop
march3<-march3[!is.na(march3$navigation_status),]
SortTest<- as.list(march3)
NavStat <- c(1,5)
j <- 1
SortTest$TripID <- 0
SortTest$TripID[1]<- paste(as.character(SortTest$mmsi[1]),'T',as.character(1),sep="")
oldNav <-SortTest$navigation_status[1]
oldMMSI <-SortTest$mmsi[1]

#Loop to assign a Trip ID to each message
ptm<-proc.time()

for (i in 2:nrow(march3)) {
  if (SortTest$mmsi[i] == oldMMSI){
    if (SortTest$navigation_status[i] != NavStat[1] & SortTest$navigation_status[i] != NavStat[2]){
      if ((oldNav == NavStat[1] | oldNav == NavStat[2])){
        j = j+1
        print(i)
      }
    }
  }
  else {
    j = 1
    oldMMSI <- SortTest$mmsi[i]
  }
  SortTest$TripID[i]<- paste(as.character(SortTest$mmsi[i]),'T',as.character(j),sep="")
  oldNav <- SortTest$navigation_status[i]
}

#Convert list back into dataframe

shipping<- as.data.frame(SortTest, stringsAsFactors=FALSE)
shipping<- shipping[order(shipping$mmsi,shipping$time),]
rm(SortTest)





#### Wave Correlation of the Ship messages and the radar data ####

## Pre Defining ncvar_get start locations
# Make midpoits of each cell
# to source which waves are associated with each
# ship message
#### Recoded Oct 18th
#####
ptm<-proc.time()
# Radar 1
longCode<-findInterval(shipping$longitude,radar$dim$lon$vals, all.inside = T)
latCode<-findInterval(shipping$latitude,radar$dim$lat$vals, all.inside = T)
hourCode<-sapply(shipping$hour_Conv[1:nrow(shipping)],function(x)(which(radar$dim$time$vals == x)))
proc.time() - ptm

# Radar 2
longCode2<-findInterval(shipping$longitude,radar2$dim$lon$vals, all.inside = T)
latCode2<-findInterval(shipping$latitude,radar2$dim$lat$vals, all.inside = T)
hourCode2<-sapply(shipping$hour_Conv[1:nrow(shipping)],function(x)(which(radar2$dim$time$vals == x)))
proc.time() - ptm

# Radar 3
longCode3<-findInterval(shipping$longitude,radar3$dim$lon$vals, all.inside = T)
latCode3<-findInterval(shipping$latitude,radar3$dim$lat$vals, all.inside = T)
hourCode3<-sapply(shipping$hour_Conv[1:nrow(shipping)],function(x)(which(radar3$dim$time$vals == x)))
proc.time() - ptm

## Selection Function for wave data

# Selection for Northward Velocity(V)
waveV2<-function(r,s,t,u,v,w,x,y,z){
  res<- ncvar_get(radar,"v",start=c(r,s,t),count=c(1,1,1))
  if (is.nan(res)){
    res<- ncvar_get(radar2,"v",start=c(u,v,w),count=c(1,1,1))
  }
  if (is.nan(res)){
    res<- ncvar_get(radar3,"v",start=c(x,y,z),count=c(1,1,1))
  }
  return(res)
}

# Selection for Eastward Velovity(U)
waveU2<-function(r,s,t,u,v,w,x,y,z){
  res<- ncvar_get(radar,"u",start=c(r,s,t),count=c(1,1,1))
  if (is.nan(res)){
    res<- ncvar_get(radar2,"u",start=c(u,v,w),count=c(1,1,1))
  }
  if (is.nan(res)){
    res<- ncvar_get(radar3,"u",start=c(x,y,z),count=c(1,1,1))
  }
```

```
  return(res)
}
proc.time() - ptm
## Running the functions into to create 2 lists that contain Northward and Eastward velocities which match currently sorted Shipping dataset

ptm<-proc.time()

testWaveV2<-mapply(function(r,s,t,u,v,w,x,y,z) waveV2(r,s,t,u,v,w,x,y,z)
                   ,longCode[1:nrow(shipping)],latCode[1:nrow(shipping)],hourCode[1:nrow(shipping)]
                   ,longCode2[1:nrow(shipping)],latCode2[1:nrow(shipping)],hourCode2[1:nrow(shipping)]
                   ,longCode3[1:nrow(shipping)],latCode3[1:nrow(shipping)],hourCode3[1:nrow(shipping)])
proc.time() - ptm

testWaveU2<-mapply(function(r,s,t,u,v,w,x,y,z) waveU2(r,s,t,u,v,w,x,y,z)
                   ,longCode[1:nrow(shipping)],latCode[1:nrow(shipping)],hourCode[1:nrow(shipping)]
                   ,longCode2[1:nrow(shipping)],latCode2[1:nrow(shipping)],hourCode2[1:nrow(shipping)]
                   ,longCode3[1:nrow(shipping)],latCode3[1:nrow(shipping)],hourCode3[1:nrow(shipping)])
proc.time() - ptm

#Merging the data

shipping$EV<- testWaveV2
shipping$EU<- testWaveU2



#### Creating the Error Vectors Northward and Eastward for Expected position -> actual ####

# Setting up data to run through loop
shipping$heading<-as.numeric(as.character(shipping$heading)) # Changing Heading from Factor -> Numeric
shipping$sog<-as.numeric(as.character(shipping$sog)) #Changing Speed over Ground from Factor -> Numeric
shipping$predN<-"na"
shipping$predE<-"na"
shipping$errE<-"na"
shipping$errN<-"na"
shipping$waveMagN<-"na"
shipping$waveMagE<-"na"
shipping$speed<- shipping$sog* 0.514444 # Knots -> m/s

# Flattened loop and corrected degree to radian argument in trig. functions
ptm<-proc.time()

imax <- nrow(shipping)

shipping$predN[2:imax] <-round((cos(shipping$heading[1:(imax-1)]*pi/180)*(shipping$speed[1:(imax-1)]*(shipping$timeDiff[2:imax])))+(shipping$UTMlat[1:imax-1]),digi
shipping$predE[2:imax] <-round((sin(shipping$heading[1:(imax-1)]*pi/180)*(shipping$speed[1:(imax-1)]*(shipping$timeDiff[2:imax])))+(shipping$UTMlog[1:imax-1]),digi


shipping$errE[2:imax] <- round((sin(shipping$heading[1:(imax-1)]*pi/180)*(shipping$speed[1:(imax-1)]*(shipping$time_conv[2:imax] - shipping$time_conv[1:(imax-1)]))
proc.time() - ptm
shipping$errN[2:imax] <- round((cos(shipping$heading[1:(imax-1)]*pi/180)*(shipping$speed[1:(imax-1)]*(shipping$time_conv[2:imax] - shipping$time_conv[1:(imax-1)]))
proc.time() - ptm
shipping$waveMagE[2:imax]<- round(shipping$EU[1:(imax-1)]*(shipping$time_conv[2:imax] - shipping$time_conv[1:(imax-1)]),digits=4)
proc.time() - ptm
shipping$waveMagN[2:imax]<- round(shipping$EV[1:(imax-1)]*(shipping$time_conv[2:imax] - shipping$time_conv[1:(imax-1)]),digits=4)
proc.time() - ptm


shipping<-shipping[complete.cases(shipping$mmsi),]
savepoint<-shipping

### Adding sequence counter within each trip
shipping$sequence<-ave(shipping$TripID,shipping$TripID, FUN = seq_along)


#### Isolation of Ideal Case scenarios###

waveOnly<-subset(shipping, !is.na(shipping$EV)) # wave related only

rot0<-c('None',0,'0.0') # Possible 0 values for ROT
noROT<-waveOnly[waveOnly$rot %in% rot0,] # extracting only 0 rot values

withHead<- subset(noROT, !is.na(noROT$heading))

#
#
#
######## Division must be made here bewteen anchred and non anchored ships for analysis####
#
#
#
## Break up the data in anchored ships in a row
anchored<-withHead[withHead$navigation_status==1,]

#Extracting the sequences of a certain length in a row
anchored$caseClass<-cumsum(c(1, diff(anchored$Key) -1))#Givethe difference in Key values but makes repeating values remain the same
temp2<-rle(anchored$caseClass)
temp3<-which(temp2$lengths>10)
imax<- length(temp3)

# Gives the starting rows of the sequences greater then 10 in withHead table
temp4<-NA
for (i in 1:imax) {
  temp4[i]<- sum(temp2$lengths[1:temp3[i]])-temp2$lengths[temp3[i]]+1
}

caseKeys<-anchored$caseClass[temp4]
anchoredCase<-anchored[anchored$caseClass %in% caseKeys,]

#Make sure classes are set properly
anchoredCase$errE<-as.numeric(anchoredCase$errE)
anchoredCase$errN<-as.numeric(anchoredCase$errN)
anchoredCase$waveMagE<- as.numeric(anchoredCase$waveMagE)
anchoredCase$waveMagN<- as.numeric(anchoredCase$waveMagN)


##########
#
```

```
#
#Seeting up the paremerters for the E-Odyn style dataset
#
#
twoMin<-withHead[withHead$timeDiff<=120 & withHead$timeDiff>0,] # only selecting when messages are less then 2 min apart (622715) changed from 120
speedLim<- twoMin[twoMin$sog>=6,]# Speeds over 6 knots

speedLim$caseClass<-cumsum(c(1, diff(speedLim$Key) -1))#Givethe difference in Key values but makes repeating values remain the same
temp2<-rle(speedLim$caseClass)
temp3<-which(temp2$lengths>50)
imax<- length(temp3)

# Gives the starting rows of the sequences greater then 10 in withHead table
temp4<-NA
for (i in 1:imax) {
  temp4[i]<- sum(temp2$lengths[1:temp3[i]])-temp2$lengths[temp3[i]]+1
}

caseKeys<-speedLim$caseClass[temp4]
speedLimCase<-speedLim[speedLim$caseClass %in% caseKeys,]

#Make sure classes are set properly
speedLimCase$errE<-as.numeric(speedLimCase$errE)
speedLimCase$errN<-as.numeric(speedLimCase$errN)
speedLimCase$waveMagE<- as.numeric(speedLimCase$waveMagE)
speedLimCase$waveMagN<- as.numeric(speedLimCase$waveMagN)


##########
#
#Creating the perfect case scenario where there should be visible correlation,
#this will choose the cases where the waves are the strongest
##########
# Selection from the speedLimCase of highest wave
maxWaveN<-speedLimCase[speedLimCase$caseClass==speedLimCase$caseClass[which.max((speedLimCase$EV)^2)],]
cor(as.numeric(maxWaveN$errN),as.numeric(maxWaveN$waveMagN))
cor(as.numeric(maxWaveN$errE),as.numeric(maxWaveN$waveMagE))

maxWaveE<-speedLimCase[speedLimCase$caseClass==speedLimCase$caseClass[which.max((speedLimCase$EU)^2)],]
cor(as.numeric(maxWaveE$errN),as.numeric(maxWaveE$waveMagN))
cor(as.numeric(maxWaveE$errE),as.numeric(maxWaveE$waveMagE))

### Selecting cases where the wave is above .3m/s
highWaves<-speedLimCase[speedLimCase$caseClass %in% speedLimCase$caseClass[which(speedLimCase$EV^2>.3^2|speedLimCase$EU^2>.3^2)],]

#### Creating different correlation values bewteen Error values and waves magnitudes for multitude of different subsets
#
#
#
#
###### Correlation with all ship messages ###### All Messages of the ship, (1)#############
#
#Removing all messages where the Time Diff is greater than 24 hours so no skewing occurs
waveOnly2<-waveOnly[waveOnly$timeDiff<86400,]
#Remove firt message of each trip
FirstMessage<-c()
oldTrip<-waveOnly2$mmsi[1]
for (i in 2:nrow(waveOnly)){
  if (waveOnly2$mmsi[i] != oldTrip){
    oldTrip<-waveOnly2$mmsi[i]
    FirstMessage<-c(FirstMessage,i)}
}
waveOnly_RM <-waveOnly2[-(FirstMessage),]


cor(as.numeric(waveOnly_RM$waveMagN),as.numeric(waveOnly_RM$errN), use = "complete.obs")
cor(as.numeric(waveOnly_RM$waveMagE),as.numeric(waveOnly_RM$errE), use = "complete.obs")

#
#
#
### Anchored vs Moving #### Anchored vs Moving (2)##############
#
#
#

anchored<-waveOnly_RM[waveOnly_RM$navigation_status==1,]

l2<-c()
oldTrip<-anchored$TripID[1]
for (i in 2:nrow(anchored)){
  if (anchored$TripID[i] != oldTrip){
    oldTrip<-anchored$TripID[i]
    l2<-c(l2,i)}
  else{oldTrip<-anchored$TripID[i]}
}
anchoredRM <-anchored[-(l2),]

moving<- waveOnly_RM[waveOnly_RM$navigation_status==0 ,]

cor(as.numeric(anchoredRM$waveMagN),as.numeric(anchoredRM$errN), use = "complete.obs")
cor(as.numeric(anchoredRM$waveMagE),as.numeric(anchoredRM$errE), use = "complete.obs")

cor(as.numeric(moving$waveMagN),as.numeric(moving$errN), use = "complete.obs")
cor(as.numeric(moving$waveMagE),as.numeric(moving$errE), use = "complete.obs")

#
#
#
### ROT vs No ROT######  (3) ##############
#
#
##Anchored

ROTtrip<-c()
waveOnly_RM$TripID<-as.character(waveOnly_RM$TripID)
waveOnly_RM$rot<-as.character(waveOnly_RM$rot)
for (i in 1:nrow(waveOnly_RM)){
```

```
  if (waveOnly_RM$rot[i] != '0.0'){
    ROTtrip<-c(ROTtrip,waveOnly_RM$TripID[i])
  }
}
uniqROTtrip<-unique(ROTtrip)

anchoredwithROT<- anchoredRM[anchoredRM$TripID %in% uniqROTtrip,]
anchoredoutROT<- anchoredRM[!(anchoredRM$TripID %in% uniqROTtrip),]

cor(as.numeric(anchoredwithROT$waveMagN),as.numeric(anchoredwithROT$errN), use = "complete.obs")
cor(as.numeric(anchoredwithROT$waveMagE),as.numeric(anchoredwithROT$errE), use = "complete.obs")

cor(as.numeric(anchoredoutROT$waveMagN),as.numeric(anchoredoutROT$errN), use = "complete.obs")
cor(as.numeric(anchoredoutROT$waveMagE),as.numeric(anchoredoutROT$errE), use = "complete.obs")

## Moving

movingwithROT<- moving[moving$TripID %in% uniqROTtrip,]
movingoutROT<- moving[!(moving$TripID %in% uniqROTtrip),]

cor(as.numeric(movingwithROT$waveMagN),as.numeric(movingwithROT$errN), use = "complete.obs")
cor(as.numeric(movingwithROT$waveMagE),as.numeric(movingwithROT$errE), use = "complete.obs")

cor(as.numeric(movingoutROT$waveMagN),as.numeric(movingoutROT$errN), use = "complete.obs")
cor(as.numeric(movingoutROT$waveMagE),as.numeric(movingoutROT$errE), use = "complete.obs")



######## No ROT and loooking at Gaps of 1-10 min ########### (4) ################

anchorGap<-anchoredoutROT[anchoredoutROT$timeDiff>60 & anchoredoutROT$timeDiff<600,] ##Anchored
movingGap<-movingoutROT[movingoutROT$timeDiff>60 & movingoutROT$timeDiff<600,] ## Moving

cor(as.numeric(anchorGap$waveMagN),as.numeric(anchorGap$errN), use = "complete.obs")
cor(as.numeric(anchorGap$waveMagE),as.numeric(anchorGap$errE), use = "complete.obs")

cor(as.numeric(movingGap$waveMagN),as.numeric(movingGap$errN), use = "complete.obs")
cor(as.numeric(movingGap$waveMagE),as.numeric(movingGap$errE), use = "complete.obs")


### Max Waves with <10 time gaps, wave > 0.3 m/s, 50 sequence ##### (5) #######
HeadingReq<- subset(waveOnly_RM, !is.na(waveOnly_RM$heading))
HeadingReq2<-HeadingReq[HeadingReq$timeDiff<600,]


HeadingReq2$caseClass<-cumsum(c(1, diff(HeadingReq2$Key) -1))#Givethe difference in Key values but makes repeating values remain the same
temp2<-rle(HeadingReq2$caseClass)
temp3<-which(temp2$lengths>50)
imax<- length(temp3)

# Gives the starting rows of the sequences greater then 50 in withHead table
temp4<-NA
for (i in 1:imax) {
  temp4[i]<- sum(temp2$lengths[1:temp3[i]])-temp2$lengths[temp3[i]]+1
}

caseKeys<-HeadingReq2$caseClass[temp4]
HeadingCase<-HeadingReq2[HeadingReq2$caseClass %in% caseKeys,]

HeadinghighWaves<-HeadingCase[HeadingCase$caseClass %in% HeadingCase$caseClass[which(HeadingCase$EV^2>.3^2|HeadingCase$EU^2>.3^2)],]

cor(as.numeric(HeadinghighWaves$waveMagN),as.numeric(HeadinghighWaves$errN), use = "complete.obs")
cor(as.numeric(HeadinghighWaves$waveMagE),as.numeric(HeadinghighWaves$errE), use = "complete.obs")

#### Wave Angle vs the Ships movement measurement ###### (6) ##################

waveAngle<-HeadingReq

waveAngle$errN<-as.numeric(waveAngle$errN)
waveAngle$errE<-as.numeric(waveAngle$errE)


waveAngle$waveDir<-NA
waveAngle$waveDir[1:nrow(waveAngle)]<-((atan2(waveAngle$EU[1:nrow(waveAngle)],waveAngle$EV[1:nrow(waveAngle)])*(180/pi))+360)%%360

# Wave magnitude
waveAngle$waveMag<-sqrt((waveAngle$EV^2)+(waveAngle$EU^2))

#Error Direction
x<-0
waveAngle$ErrorDir<-NA

waveAngle$ErrorDir[1:nrow(waveAngle)]<-((atan2(waveAngle$errN[1:nrow(waveAngle)],waveAngle$errE[1:nrow(waveAngle)])*(180/pi))+360)%%360

#Error Magnitude
waveAngle$errorMag<-sqrt((waveAngle$errN^2)+(waveAngle$errE^2))

###Angles between Error Vector, and Wave Vector

waveAngle$AngleBetween<-as.numeric(waveAngle$ErrorDir - waveAngle$waveDir)
#283 cases where the wave angle is NA because previous message had NaN Heading so needed to be subset
waveAngle2<-waveAngle[!is.na(waveAngle$AngleBetween),]

for (i in 1:nrow(waveAngle2)) {
    if (waveAngle2$AngleBetween[i] > 180) {
       waveAngle2$AngleBetween[i]<-waveAngle2$AngleBetween[i]-360
     }
    if (waveAngle2$AngleBetween[i] < (-180)) {
      waveAngle2$AngleBetween[i]<-waveAngle2$AngleBetween[i]+360
    }
}
waveAngle2$AngleBetween<-abs(waveAngle2$AngleBetween)  ### There are 573,544 cases with these requirements


### Selecting Perpedicular Cases

parralel<-waveAngle2[waveAngle2$AngleBetween>=175 | waveAngle2$AngleBetween<=5,] #32,703 cases
cor(as.numeric(parralel$waveMagN),as.numeric(parralel$errN), use = "complete.obs")
```

```
cor(as.numeric(parralel$waveMagE),as.numeric(parralel$errE), use = "complete.obs")


### Selecting Orthoganal Cases

perpendicular<-waveAngle2[waveAngle2$AngleBetween>=85 & waveAngle2$AngleBetween<=95,] # 33,088 cases
cor(as.numeric(perpendicular$waveMagN),as.numeric(perpendicular$errN), use = "complete.obs")
cor(as.numeric(perpendicular$waveMagE),as.numeric(perpendicular$errE), use = "complete.obs")


##### E-Odyn replicated scenario for ship messages ####### (7) #############
#Moving, SOG>6, Time Interval 1-2min, 0ROT# Selecting cases where sequence greater then 5 to weed out error

EODYN<-HeadingReq[HeadingReq$navigation_status==0,]
EODYN2<-EODYN[EODYN$sog>=6,]
EODYN3<-EODYN2[EODYN2$timeDiff<121,]
EODYN4<-EODYN3[EODYN3$rot=='0.0',]

EODYN4$caseClass<-cumsum(c(1, diff(EODYN4$Key) -1))#Givethe difference in Key values but makes repeating values remain the same
temp2<-rle(EODYN4$caseClass)
temp3<-which(temp2$lengths>5)
imax<- length(temp3)

# Gives the starting rows of the sequences greater then 5 in withHead table
temp4<-NA
for (i in 1:imax) {
  temp4[i]<- sum(temp2$lengths[1:temp3[i]])-temp2$lengths[temp3[i]]+1
}

caseKeys<-EODYN4$caseClass[temp4]
EODYNcase<-EODYN4[EODYN4$caseClass %in% caseKeys,]

UniqueClass<-unique((EODYNcase$caseClass))
minOnly<-EODYN4[1,]

for (i in 1:length(UniqueClass)){
  case<-EODYN4[EODYN4$caseClass==UniqueClass[i],]

  oldtime<-case$time_conv[1]
  for (j in 2:(nrow(case))){
    if ((case$time_conv[j] - oldtime) > 60){
      minOnly<-rbind(minOnly, case[j,])
      oldtime<-case$time_conv[j]
    }
  }
}

minOnly$errE2<-"na"
minOnly$errN2<-"na"
minOnly$waveMagN2<-"na"
minOnly$waveMagE2<-"na"
minOnly$timeDiff2<- c(0, diff(minOnly$time_conv))
imax <- nrow(minOnly)
minOnly$errE2[2:imax] <-
  round(
    (sin(minOnly$heading[1:(imax-1)]*pi/180)*(minOnly$speed[1:(imax-1)]*(minOnly$time_conv[2:imax]- minOnly$time_conv[1:(imax-1)])))
    -(minOnly$UTMlog[2:imax]-minOnly$UTMlog[1:(imax-1)])
    ,digits=2)

minOnly$errN2[2:imax] <-
  round(
    (cos(minOnly$heading[1:(imax-1)]*pi/180)*(minOnly$speed[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)])))
    -(minOnly$UTMlat[2:imax]-minOnly$UTMlat[1:(imax-1)])
    ,digits=2)

minOnly$waveMagE2[2:imax]<- round(minOnly$EU[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)]),digits=4)
minOnly$waveMagN2[2:imax]<- round(minOnly$EV[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)]),digits=4)

minOnly2<-minOnly[minOnly$timeDiff2<121&minOnly$timeDiff2>59,]

cor(as.numeric(minOnly2$waveMagN2),as.numeric(minOnly2$errN2))
cor(as.numeric(minOnly2$waveMagE2),as.numeric(minOnly2$errE2))

#### Ideal Case Scenarios ############## (8) ##########################
### NAv_stat = 0, SOG >6, ROT = 0, Time Intervals <120, 30 message sequence ###########
### Subset the Highest wave scenario for example, Subset High negative Correlation, Subset High POsitive Correlation, Subset no COrrelation######

IdealCase<-HeadingReq[HeadingReq$navigation_status==0,]
IdealCase<-IdealCase[IdealCase$sog>6,]
IdealCase<-IdealCase[IdealCase$rot=='0.0',]
IdealCase<-IdealCase[IdealCase$timeDiff<121,]

IdealCase$caseClass<-cumsum(c(1, diff(IdealCase$Key) -1))#Givethe difference in Key values but makes repeating values remain the same
temp2<-rle(IdealCase$caseClass)
temp3<-which(temp2$lengths>30)
imax<- length(temp3)

# Gives the starting rows of the sequences greater then 30 in dataset
temp4<-NA
for (i in 1:imax) {
  temp4[i]<- sum(temp2$lengths[1:temp3[i]])-temp2$lengths[temp3[i]]+1
}

caseKeys<-IdealCase$caseClass[temp4]
IdealCaseFinal<-IdealCase[IdealCase$caseClass %in% caseKeys,]

cor(as.numeric(IdealCaseFinal$waveMagN),as.numeric(IdealCaseFinal$errN), use = "complete.obs")
cor(as.numeric(IdealCaseFinal$waveMagE),as.numeric(IdealCaseFinal$errE), use = "complete.obs")

### Selecting Case with Highest waves from Ideal Cases
IdealMaxWave<-IdealCaseFinal[IdealCaseFinal$caseClass %in% IdealCaseFinal$caseClass[which.max(IdealCaseFinal$EV)],]

cor(as.numeric(IdealMaxWave$waveMagN),as.numeric(IdealMaxWave$errN), use = "complete.obs")
cor(as.numeric(IdealMaxWave$waveMagE),as.numeric(IdealMaxWave$errE), use = "complete.obs")

#Make List of correlations from each caseclass#
```

```
correlation<-data.frame(Case=caseKeys)

for (i in 1:length(caseKeys)){
  case<-IdealCase[IdealCase$caseClass==caseKeys[i],]
  correlation$TripID[i]<-case$TripID[1]
  correlation$NorthCor[i]<- round(cor(as.numeric(case$waveMagN),as.numeric(case$errN), use = "complete.obs"),digits = 4)
  correlation$EastCor[i]<- round(cor(as.numeric(case$waveMagE),as.numeric(case$errE), use = "complete.obs"),digits = 4)
  correlation$maxwavNor[i]<- max(case$EV,na.rm = T)
  correlation$maxwavEast[i]<- max(case$EU,na.rm = T)
}
### Selecting Strongest NEgative Correlation set ###
min(correlation$NorthCor,na.rm=T)
min(correlation$EastCor, na.rm=T)

StrongNeg<-IdealCaseFinal[IdealCaseFinal$caseClass==correlation$Case[which.min(correlation$EastCor)],]

plot(StrongNeg$UTMlog,StrongNeg$UTMlat,
     main = paste("Ship MMSI",StrongNeg$mmsi[1]),
     xlab = "UTM Longitude",
     ylab = "UTM Latitude",
     asp=1
)
lines(StrongNeg$UTMlog,StrongNeg$UTMlat)
points(StrongNeg$UTMlog[1],StrongNeg$UTMlat[1], col="red", pch = 19) # Where messages start


### Selecting Strongest Positive Correlation set ###
max(correlation$NorthCor,na.rm=T)
max(correlation$EastCor, na.rm=T)

StrongPos<-IdealCaseFinal[IdealCaseFinal$caseClass==correlation$Case[which.max(correlation$NorthCor)],]

plot(StrongPos$UTMlog,StrongPos$UTMlat,
     main = paste("Ship MMSI",StrongPos$mmsi[1]),
     xlab = "UTM Longitude",
     ylab = "UTM Latitude",
     asp=1
)
lines(StrongPos$UTMlog,StrongPos$UTMlat)
points(StrongPos$UTMlog[1],StrongPos$UTMlat[1], col="red", pch = 19) # Where messages start


### Selecting a randon case with no correlation  but decent high waves###

LowCor<-IdealCaseFinal[IdealCaseFinal$caseClass==864770,]

plot(LowCor$UTMlog,LowCor$UTMlat,
     main = paste("Ship MMSI",LowCor$mmsi[1]),
     xlab = "UTM Longitude",
     ylab = "UTM Latitude",
     asp=1
)
lines(LowCor$UTMlog,LowCor$UTMlat)
points(LowCor$UTMlog[1],LowCor$UTMlat[1], col="red", pch = 19) # Where messages start

# Go to CaseStudies.R and set up the Generation of Correlation Table of all the Cases and How their Correlation changes with time gaps
### Generating Case Scenarios and analyzing them
# Current Case Setting:Speed>=6knots, ROT=0, Waves present, time gaps no greater then 60min, sequence >=50
# These setting create: 778 cases -> 60,430 messages
#####
caseKeys<-caseKeys[!is.na(caseKeys)]
#Must generate a list of 50 case studies at random from the full 778

Rcases<-caseKeys[c(sample(1:length(caseKeys),50))]

CaseMessage<-speedLimCase[speedLimCase$caseClass %in% Rcases,]


#### Test for High Waves, comment out all
Rcases<-unique((highWaves$caseClass))
CaseMessage<-highWaves
######

#### Case set for final analysis, 50 Random###
#Apr 2018
Rcases<-caseKeys[c(sample(1:length(caseKeys),50))]
CaseMessage<-IdealCaseFinal
####

#Must run analysis on all the cases to test the correlation between wave velocity and error
correlation<-data.frame(Case=Rcases)
correlation$TripID<-"na"
correlation$NorthCor<-"na"
correlation$EastCor<-"na"
correlation$North1min<-"na"
correlation$East1min<-"na"
correlation$North2min<-"na"
correlation$East2min<-"na"
correlation$North5min<-"na"
correlation$East5min<-"na"
correlation$North10min<-"na"
correlation$East10min<-"na"
correlation$ErrorEast<-"na"
correlation$ErrorNorth<-"na"
correlation$CaliErrorEast<-"na"
correlation$CaliErrorNorth<-"na"
correlation$RealCalErrorEast<-"na"
correlation$RealCalErrorNorth<-"na"

ptm<-proc.time()

for (i in 1:length(Rcases)){
  tryCatch({
    case<-CaseMessage[CaseMessage$caseClass==Rcases[i],]
    correlation$TripID[i]<-case$TripID[1]
    correlation$NorthCor[i]<- round(cor(as.numeric(case$waveMagN),as.numeric(case$errN), use = "complete.obs"),digits = 4)
    correlation$EastCor[i]<- round(cor(as.numeric(case$waveMagE),as.numeric(case$errE), use = "complete.obs"),digits = 4)
```

```
  #1min Break
  if ((case$time_conv[nrow(case)]-case$time_conv[1])>60) {
    minOnly<-case[1,]
    oldtime<-minOnly$time_conv[1]
    for (j in 2:(nrow(case))){
      if ((case$time_conv[j] - oldtime) > 60){
        minOnly<-rbind(minOnly, case[j,])
        oldtime<-case$time_conv[j]
      }
    }

    minOnly$errE2<-"na"
    minOnly$errN2<-"na"
    minOnly$waveMagN2<-"na"
    minOnly$waveMagE2<-"na"
    imax <- nrow(minOnly)
    minOnly$errE2[2:imax] <-
      round(
        (sin(minOnly$heading[1:(imax-1)]*pi/180)*(minOnly$speed[1:(imax-1)]*(minOnly$time_conv[2:imax]- minOnly$time_conv[1:(imax-1)])))
        -(minOnly$UTMlog[2:imax]-minOnly$UTMlog[1:(imax-1)])
        ,digits=2)

    minOnly$errN2[2:imax] <-
      round(
        (cos(minOnly$heading[1:(imax-1)]*pi/180)*(minOnly$speed[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)])))
        -(minOnly$UTMlat[2:imax]-minOnly$UTMlat[1:(imax-1)])
        ,digits=2)

    minOnly$waveMagE2[2:imax]<- round(minOnly$EU[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)]),digits=4)
    minOnly$waveMagN2[2:imax]<- round(minOnly$EV[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)]),digits=4)

    correlation$North1min[i]<- round(cor(as.numeric(minOnly$waveMagN2),as.numeric(minOnly$errN2), use = "complete.obs"),digits = 4)
    correlation$East1min[i]<- round(cor(as.numeric(minOnly$waveMagE2),as.numeric(minOnly$errE2), use = "complete.obs"),digits = 4)

  }

  #2min Break
if ((case$time_conv[nrow(case)]-case$time_conv[1])>120) {
  minOnly<-case[1,]
  oldtime<-minOnly$time_conv[1]
  for (j in 2:(nrow(case))){
    if ((case$time_conv[j] - oldtime) > 120){
      minOnly<-rbind(minOnly, case[j,])
      oldtime<-case$time_conv[j]
    }
  }

  minOnly$errE2<-"na"
  minOnly$errN2<-"na"
  minOnly$waveMagN2<-"na"
  minOnly$waveMagE2<-"na"
  imax <- nrow(minOnly)
  minOnly$errE2[2:imax] <-
    round(
      (sin(minOnly$heading[1:(imax-1)]*pi/180)*(minOnly$speed[1:(imax-1)]*(minOnly$time_conv[2:imax]- minOnly$time_conv[1:(imax-1)])))
      -(minOnly$UTMlog[2:imax]-minOnly$UTMlog[1:(imax-1)])
      ,digits=2)

  minOnly$errN2[2:imax] <-
    round(
      (cos(minOnly$heading[1:(imax-1)]*pi/180)*(minOnly$speed[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)])))
      -(minOnly$UTMlat[2:imax]-minOnly$UTMlat[1:(imax-1)])
      ,digits=2)

  minOnly$waveMagE2[2:imax]<- round(minOnly$EU[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)]),digits=4)
  minOnly$waveMagN2[2:imax]<- round(minOnly$EV[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)]),digits=4)

  correlation$North2min[i]<- round(cor(as.numeric(minOnly$waveMagN2),as.numeric(minOnly$errN2), use = "complete.obs"),digits = 4)
  correlation$East2min[i]<- round(cor(as.numeric(minOnly$waveMagE2),as.numeric(minOnly$errE2), use = "complete.obs"),digits = 4)

}
  #5min Break
if   ((case$time_conv[nrow(case)]-case$time_conv[1])>300) {
  minOnly<-case[1,]
  oldtime<-minOnly$time_conv[1]
  for (j in 2:(nrow(case))){
    if ((case$time_conv[j] - oldtime) > 300){
      minOnly<-rbind(minOnly, case[j,])
      oldtime<-case$time_conv[j]
    }
  }
  minOnly$errE2<-"na"
  minOnly$errN2<-"na"
  minOnly$waveMagN2<-"na"
  minOnly$waveMagE2<-"na"
  imax <- nrow(minOnly)
  minOnly$errE2[2:imax] <-
    round(
      (sin(minOnly$heading[1:(imax-1)]*pi/180)*(minOnly$speed[1:(imax-1)]*(minOnly$time_conv[2:imax]- minOnly$time_conv[1:(imax-1)])))
      -(minOnly$UTMlog[2:imax]-minOnly$UTMlog[1:(imax-1)])
      ,digits=2)

  minOnly$errN2[2:imax] <-
    round(
      (cos(minOnly$heading[1:(imax-1)]*pi/180)*(minOnly$speed[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)])))
      -(minOnly$UTMlat[2:imax]-minOnly$UTMlat[1:(imax-1)])
      ,digits=2)

  minOnly$waveMagE2[2:imax]<- round(minOnly$EU[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)]),digits=4)
  minOnly$waveMagN2[2:imax]<- round(minOnly$EV[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)]),digits=4)

  correlation$North5min[i]<- round(cor(as.numeric(minOnly$waveMagN2),as.numeric(minOnly$errN2), use = "complete.obs"),digits = 4)
  correlation$East5min[i]<- round(cor(as.numeric(minOnly$waveMagE2),as.numeric(minOnly$errE2), use = "complete.obs"),digits = 4)
}
  #10 min Break
if ((case$time_conv[nrow(case)]-case$time_conv[1])>600) {
    minOnly<-case[1,]
    oldtime<-minOnly$time_conv[1]
```

```
      for (j in 2:(nrow(case))){
        if ((case$time_conv[j] - oldtime) > 600){
          minOnly<-rbind(minOnly, case[j,])
          oldtime<-case$time_conv[j]
        }
      }
      minOnly$errE2<-"na"
      minOnly$errN2<-"na"
      minOnly$waveMagN2<-"na"
      minOnly$waveMagE2<-"na"
      imax <- nrow(minOnly)
      minOnly$errE2[2:imax] <-
        round(
          (sin(minOnly$heading[1:(imax-1)]*pi/180)*(minOnly$speed[1:(imax-1)]*(minOnly$time_conv[2:imax]- minOnly$time_conv[1:(imax-1)])))
          -(minOnly$UTMlog[2:imax]-minOnly$UTMlog[1:(imax-1)])
          ,digits=2)

      minOnly$errN2[2:imax] <-
        round(
          (cos(minOnly$heading[1:(imax-1)]*pi/180)*(minOnly$speed[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)])))
          -(minOnly$UTMlat[2:imax]-minOnly$UTMlat[1:(imax-1)])
          ,digits=2)

      minOnly$waveMagE2[2:imax]<- round(minOnly$EU[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)]),digits=4)
      minOnly$waveMagN2[2:imax]<- round(minOnly$EV[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)]),digits=4)

      correlation$North10min[i]<- round(cor(as.numeric(minOnly$waveMagN2),as.numeric(minOnly$errN2), use = "complete.obs"),digits = 4)
      correlation$East10min[i]<- round(cor(as.numeric(minOnly$waveMagE2),as.numeric(minOnly$errE2), use = "complete.obs"),digits = 4)
  }
    ###Creating the Error Values based on each trip
  correlation$ErrorEast[i]<-sqrt(sum(as.numeric(case$errE)^2,na.rm = T))
  correlation$ErrorNorth[i]<-sqrt(sum(as.numeric(case$errN)^2,na.rm = T))


  correlation$CaliErrorEast[i]<-sqrt(sum((as.numeric(case$errE) + as.numeric(case$waveMagE))^2,na.rm = T))
  correlation$CaliErrorNorth[i]<-sqrt(sum((as.numeric(case$errN) + as.numeric(case$waveMagN))^2,na.rm = T))

  case$cog<-as.numeric(case$cog)
  correlation$RealCalErrorEast[i]<-sqrt(sum((as.numeric(case$errE) + (as.numeric(case$waveMagE)*(((1/7)*sin(case$cog*pi/180))+(cos(case$cog*pi/180)))))^2,na.rm = T
  correlation$RealCalErrorNorth[i]<-sqrt(sum((as.numeric(case$errN) + (as.numeric(case$waveMagN)*((sin(case$cog*pi/180))+((1/7)*cos(case$cog*pi/180)))))^2,na.rm =
  })
}

proc.time() - ptm


### Looking at certain cases

summary(as.numeric(correlation$North1min))
summary(as.numeric(correlation$North2min))
summary(as.numeric(correlation$North5min))
summary(as.numeric(correlation$North10min))
summary(as.numeric(correlation$East1min))
summary(as.numeric(correlation$East2min))
summary(as.numeric(correlation$East5min))
summary(as.numeric(correlation$East10min))

### Model Calibration Revised from modelCalibration.R Aug 16 2016
##
## April 2018

### Overall error in messages to message prediction initially (Error 1)
ErrorTestSample<-waveOnly_RM

#All Error East
E1<-sum(abs(as.numeric(ErrorTestSample$errE)),na.rm = T)
E1
mean(abs(as.numeric(ErrorTestSample$errE)),na.rm = T)
#All Error North
N1<-sum(abs(as.numeric(ErrorTestSample$errN)),na.rm = T)
N1
mean(abs(as.numeric(ErrorTestSample$errN)),na.rm = T)

### Calibrated error which includes wave movements into the equation (Error 2)

# New Error Estimation by adding Wave Magnitudes over time
ErrorTestSample$errErev<- as.numeric(ErrorTestSample$errE) + as.numeric(ErrorTestSample$waveMagE)
ErrorTestSample$errNrev<- as.numeric(ErrorTestSample$errN) + as.numeric(ErrorTestSample$waveMagN)

### New Collective Error
E2<-sum(abs(as.numeric(ErrorTestSample$errErev)),na.rm = T) #
N2<-sum(abs(as.numeric(ErrorTestSample$errNrev)),na.rm = T) #
#Percent Difference
(E2-E1)/E1*100
(N2-N1)/N1*100
# Error East has been decreased by 0.026% Error North increased by 0.899%


### Wave included but with a realistica calibration based on ship and wave directional orientation (Error 3)

## By adding the wave magnitudes to a physical model may enhance the accuracy to a greater degree
# an aspect ratio of a ship width vs length will 1:7 based off averages on panama, new panama, Malaca canal
# max ship sizes
ErrorTestSample$cog<-as.numeric(as.character(ErrorTestSample$cog))
ErrorTestSample$waveMagNrev<- as.numeric(ErrorTestSample$waveMagN)*((sin(ErrorTestSample$cog*pi/180))+((1/7)*cos(ErrorTestSample$cog*pi/180)))
ErrorTestSample$waveMagErev<- as.numeric(ErrorTestSample$waveMagE)*(((1/7)*sin(ErrorTestSample$cog*pi/180))+(cos(ErrorTestSample$cog*pi/180)))

ErrorTestSample$errErev2<- round(as.numeric(ErrorTestSample$errE) + as.numeric(ErrorTestSample$waveMagErev),digits = 2)
ErrorTestSample$errNrev2<- round(as.numeric(ErrorTestSample$errN) + as.numeric(ErrorTestSample$waveMagNrev),digits = 2)


E3<-sum(abs(as.numeric(ErrorTestSample$errErev2)),na.rm = T)
N3<-sum(abs(as.numeric(ErrorTestSample$errNrev2)),na.rm = T)
#Percent Difference
(E3-E1)/E1*100
(N3-N1)/N1*100
# Error East has been decreased by 4.296% Error North decreased by 3.741%
```

```
sums<-matrix(c(E1,N1,E2,N2,E3,N3),ncol=2)
colnames(sums)<-c('East Error','North Errror')
rownames(sums)<-c('Initial','Wave Calibrated','Direction Calibration')
SumTable<-as.table(sums)
SumTable$MeanEast<-mean(abs(as.numeric(ErrorTestSample$errE)),na.rm = T)

ModelCal<-CaseMessage[CaseMessage$caseClass%in%Rcases,]

### Ideal Case Scenario Break down and running model Calibration on the radom cases ###
round(cor(as.numeric(ModelCal$waveMagN),as.numeric(ModelCal$errN), use = "complete.obs"),digits = 4)
round(cor(as.numeric(ModelCal$waveMagE),as.numeric(ModelCal$errE), use = "complete.obs"),digits = 4)

mean(as.numeric(correlation$NorthCor), na.rm = T)
sum(as.numeric(correlation$NorthCor), na.rm = T)

mean(as.numeric(correlation$EastCor), na.rm = T)
sum(as.numeric(correlation$EastCor), na.rm = T)

summary(as.numeric(correlation$NorthCor))
summary(as.numeric(correlation$EastCor))


### Calibrate the Ideal Case Messeges and See if we can lower overall error

#Error East (squared variance)
E1<-sqrt(sum(as.numeric(ModelCal$errE)^2,na.rm = T))
E1
mean(abs(as.numeric(ModelCal$errE)),na.rm = T)
#All Error North (squared variance)
N1<-sqrt(sum(as.numeric(ModelCal$errN)^2,na.rm = T))
N1
mean(abs(as.numeric(ModelCal$errN)),na.rm = T)

#New Error Estimation by adding Wave Magnitudes over time
ModelCal$errErev<- as.numeric(ModelCal$errE) + as.numeric(ModelCal$waveMagE)
ModelCal$errNrev<- as.numeric(ModelCal$errN) + as.numeric(ModelCal$waveMagN)

### New Collective Error
E2<-sqrt(sum(as.numeric(ModelCal$errErev)^2,na.rm = T))
N2<-sqrt(sum(as.numeric(ModelCal$errNrev)^2,na.rm = T))

## By adding the wave magnitudes to a physical model may enhance the accuracy to a greater degree
# an aspect ratio of a ship width vs length will 1:7 based off averages on panama, new panama, Malaca canal
# max ship sizes
ModelCal$cog<-as.numeric(as.character(ModelCal$cog))
ModelCal$waveMagNrev<- as.numeric(ModelCal$waveMagN)*((sin(ModelCal$cog*pi/180))+((1/7)*cos(ModelCal$cog*pi/180)))
ModelCal$waveMagErev<- as.numeric(ModelCal$waveMagE)*(((1/7)*sin(ModelCal$cog*pi/180))+(cos(ModelCal$cog*pi/180)))

ModelCal$errErev2<- round(as.numeric(ModelCal$errE) + as.numeric(ModelCal$waveMagErev),digits = 2)
ModelCal$errNrev2<- round(as.numeric(ModelCal$errN) + as.numeric(ModelCal$waveMagNrev),digits = 2)


E3<-sqrt(sum(as.numeric(ModelCal$errErev2)^2,na.rm = T))
N3<-sqrt(sum(as.numeric(ModelCal$errNrev2)^2,na.rm = T))
#Percent Difference
(E3-E1)/E1*100
(N3-N1)/N1*100

#### Histogram and Charts ####
# March 2018
#
#
# ###
#First subset first message out
install.packages("ggplot2")
library(ggplot2)
#
FirstMessage<-c()
oldTrip<-shipping$mmsi[1]
for (i in 2:nrow(shipping)){
  if (shipping$mmsi[i] != oldTrip){
    oldTrip<-shipping$mmsi[i]
    FirstMessage<-c(FirstMessage,i)}
}
shippingRM <-shipping[-(FirstMessage),]

## Histogram of Time gaps
# Lets change the time into minutes
shippingRM$minGap<-shippingRM$timeDiff/60
shippingRM$hourGap<-as.numeric(shippingRM$timeDiff/3600)

under5min<-shippingRM[shippingRM$timeDiff<300,]
over5minUnder1day<-shippingRM[shippingRM$timeDiff>300 & shippingRM$timeDiff<86400,]
over1day<-shippingRM[shippingRM$timeDiff>86400,]

r <- hist(shippingRM$timeDiff, main="Shipping Time Differences Between Messages", ylab="Frequency", xlab="Time Difference (sec)")
barplot(r$counts, log="y", col="white", names.arg=r$breaks[-1],main="Shipping Time Differences Between Messages \n (Logarithmic)", ylab="Frequency (logarithmic)",

r <- hist(under5min$timeDiff, main="Shipping Time Differences Between Messages \n Under 5 min Apart", ylab="Frequency", xlab="Time Difference (sec)")
barplot(r$counts, log="y", col="white", names.arg=r$breaks[-1], main="Shipping Time Differences Between Messages \n Under 5 min Apart (Logarithmic)", ylab="Frequen

r <- hist(over5minUnder1day$hourGap, main="Shipping Time Differences Between Messages \n Over 5 min but under 24 hrs", ylab="Frequency", xlab="Time Difference (hou
barplot(r$counts, log="y", col="white", names.arg=r$breaks[-1], main="Shipping Time Differences Between Messages \n Over 5 min but under 24 hrs (Logarithmic)", yla

r <- hist(over1day$hourGap, main="Shipping Time Differences Between Messages \n Over 24 hrs", ylab="Frequency", xlab="Time Difference (hours)")
barplot(r$counts, log="y", col="white", names.arg=r$breaks[-1], main="Shipping Time Differences Between Messages \n Over 24 hrs (Logarithmic)", ylab="Frequency (Lo

### Histogram of Space gaps

shippingRM$distKM<-as.numeric(shippingRM$distDiff/1000)

under1km<-shippingRM[shippingRM$distKM<1,]
under30km<-shippingRM[shippingRM$distKM<30&shippingRM$distKM>1 ,]
over30km<-shippingRM[shippingRM$distKM>30,]

r <- hist(shippingRM$distKM, main="Shipping Distance Differences \n Between Messages", ylab="Frequency", xlab="Distance Difference (Kilometre)")
barplot(r$counts, log="y", col="white", names.arg=r$breaks[-1],main="Shipping Distance Differences \n Between Messages (Logarithmic)", ylab="Frequency (logarithmic
```

```
r <- hist(under1km$distDiff, main="Shipping Distance Differences \n  Between Messages Under 1km", ylab="Frequency", xlab="Distance Difference (metre)")
barplot(r$counts, log="y", col="white", names.arg=r$breaks[-1],main="Shipping Distance Differences \n Between Messages Under 1km (Logarithmic)", ylab="Frequency (l

r <- hist(under30km$distKM, main="Shipping Distance Differences \n Between Messages Under 30km", ylab="Frequency", xlab="Distance Difference (Kilometre)")
barplot(r$counts, log="y", col="white", names.arg=r$breaks[-1],main="Shipping Distance Differences \n Between Messages Under 30km (Logarithmic)", ylab="Frequency (

r <- hist(over30km$distKM, main="Shipping Distance Differences \n Between Messages Over 30km", ylab="Frequency", xlab="Distance Difference (Kilometre)")
barplot(r$counts, log="y", col="white", names.arg=r$breaks[-1],main="Shipping Distance Differences \n Between Messages Over 30km (Logarithmic)", ylab="Frequency (l


### Scatterplot of Space Gaps vs Time Gaps

plot(shippingRM$timeDiff,shippingRM$distDiff)

ggplot(under5min,aes(timeDiff,distDiff)) + geom_point() +
  xlab("Time Difference (seconds)") +
  ylab("Distance Difference(metres)") +
  ggtitle("Time vs Distance in 5 Minutes")

ggplot(over5minUnder1day,aes(timeDiff,distDiff)) + geom_point() +
  xlab("Time Difference (seconds)") +
  ylab("Distance Difference(metres)") +
  ggtitle("Time vs Distance in 1 day")

ggplot(over1day,aes(timeDiff,distDiff)) + geom_point() +
  xlab("Time Difference (seconds)") +
  ylab("Distance Difference(metres)") +
  ggtitle("Time vs Distance Over 1 day")

#####
# Creating a scatterplot of space over time with all ships, including 25m/s (90km/h,~50knots ) line of cut off Maximum distance in study area is 500km
# Maximum distance in study area is 500km which if at 25m/s would occur only at 20,000 seconds(5.55hrs)
testsample<-ships
testsample<- testsample[order(testsample$mmsi,testsample$time),]

testsample$timeDiff <- c(0, diff(testsample$time_conv))
cord.dec = SpatialPoints(cbind(testsample$longitude, testsample$latitude), proj4string = CRS("+proj=longlat"))  #Collecting the Long Lat points of the data
cord.UTM <- spTransform(cord.dec, CRS("+init=epsg:26711")) #Transforming points to UTM by Code for UTM zone 11 Nad 83
testsample$UTMlog<-cord.UTM$coords.x1
testsample$UTMlat<-cord.UTM$coords.x2

rm(cord.dec)
rm(cord.UTM)


testsample$distLat <- c(0, diff(testsample$UTMlat))
testsample$distLon<- c(0,diff(testsample$UTMlog))
testsample$distDiff<- sqrt((testsample$distLon^2)+(testsample$distLat^2))


FirstMessage<-c()
oldTrip<-testsample$mmsi[1]
for (i in 2:nrow(testsample)){
  if (testsample$mmsi[i] != oldTrip){
    oldTrip<-testsample$mmsi[i]
    FirstMessage<-c(FirstMessage,i)}
}
testsampleRM <-testsample[-(FirstMessage),]

limits<-testsampleRM[testsampleRM$timeDiff<20000,]
firstten<-testsampleRM[testsampleRM$timeDiff<600,]

ggplot(limits,aes(timeDiff,distDiff)) + geom_point() +
  xlab("Time Difference (seconds)") +
  ylab("Distance Difference(kilometres)") +
  ggtitle("Impossible values   n=3,832,252") +
  geom_segment(aes(x=0,y=0,xend=20000,yend=500000)) +
  scale_x_continuous(expand = c(0,0)) +
  theme(axis.text.y = element_text(angle = 45),plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(expand =c(0,0),breaks=c(0,100000,200000,300000,400000,500000),labels=c("0", "100","200","300","400","500"))

ggplot(firstten,aes(timeDiff,distDiff)) + geom_point() +
  xlab("Time Difference (seconds)") +
  ylab("Distance Difference(kilometres)") +
  ggtitle("Impossible Values First Ten Minutes  n=3,510,668") +
  geom_segment(aes(x=0,y=0,xend=600,yend=15000)) +
  scale_x_continuous(expand = c(0,0)) +
  theme(axis.text.y = element_text(angle = 45),plot.title = element_text(hjust = 0.5)) +
  scale_y_continuous(expand =c(0,0),breaks=c(0,50000,100000,150000,200000,250000, 300000),labels=c("0", "50","100","150","200","250", "300"))

### Generating Graphs and plots for the Case Scenarios randomly generated
plotpath<-("E:/Masters/MappingFiles/Graphs")

###


plot(minOnly$UTMlog,minOnly$UTMlat,
     main = paste("Ship MMSI",minOnly$mmsi[1]),
     xlab = "UTM Longitude",
     ylab = "UTM Latitude"
     )
lines(minOnly$UTMlog,minOnly$UTMlat)
points(minOnly$UTMlog[1],minOnly$UTMlat[1], col="red", pch = 19)

plot(minOnly$waveMagE2,minOnly$errE2,asp=1,
     main = paste("Ship MMSI",minOnly$mmsi[1],"\nError East vs Prediction Error East R =", correlation$EastCor[1]),
     xlab = "Wave movement eastward (m)",
     ylab = "Displacement eastward (m)"
)
abline(lm(as.numeric(minOnly$errE2)~as.numeric(minOnly$waveMagE2)))


path<-file.path(plotpath,paste("test",".jpg"),sep = "")
png(file=path)
plot(minOnly$waveMagN2,minOnly$errN2,asp=1,
     main = paste("Ship MMSI",minOnly$mmsi[1],"\nError East vs Prediction Error North R =", correlation$NorthCor[1]),
     xlab = "Wave movement northwardward (m)",
     ylab = "Displacement northward (m)"
```

```
)
abline(lm(as.numeric(minOnly$errN2)~as.numeric(minOnly$waveMagN2)))
dev.off()

ptm<-proc.time()
#length(Rcases)
pathConn<-paste(plotpath,"/LatexInputs.txt",sep="")
file(pathConn)
for (i in 1:length(Rcases)){
  tryCatch({
    case<-CaseMessage[CaseMessage$caseClass==Rcases[i],]

    ####2min Break####

    minOnly<-case[1,]

    path<-file.path(plotpath,paste("Case_",i,"_",case$mmsi[1],".jpg",sep = ""))
    png(file=path)
    plot(case$UTMlog,case$UTMlat,
         main = paste("Ship MMSI",case$mmsi[1]),
         xlab = "UTM Longitude",
         ylab = "UTM Latitude"
    )
    lines(case$UTMlog,case$UTMlat)
    points(case$UTMlog[1],case$UTMlat[1], col="red", pch = 19)
    dev.off()

    write(paste("\\includegraphics[width=0.33\\linewidth]{{",path,"}}",sep=""), file=pathConn, append=T)

    path<-file.path(plotpath,paste("Case_",i,"_","CorrelationEast",".jpg",sep = ""))
    png(file=path)
    plot(case$waveMagE,case$errE,asp=1,
         main = paste("Ship MMSI",case$mmsi[1],"\nError East vs Prediction Error East R =", correlation$EastCor[i]),
         xlab = "Wave movement eastward (m)",
         ylab = "Displacement eastward (m)"
    )
    abline(lm(as.numeric(case$errE)~as.numeric(case$waveMagE)))
    dev.off()

    write(paste("\\includegraphics[width=0.33\\linewidth]{{",path,"}}",sep=""), file=pathConn, append=T)

    path<-file.path(plotpath,paste("Case_",i,"_","CorrelationNorth",".jpg",sep = ""))
    png(file=path)
    plot(case$waveMagN,case$errN,asp=1,
         main = paste("Ship MMSI",case$mmsi[1],"\nError North vs Prediction Error North R =", correlation$NorthCor[i]),
         xlab = "Wave movement northward (m)",
         ylab = "Displacement northward (m)"
    )
    abline(lm(as.numeric(case$errN)~as.numeric(case$waveMagN)))
    dev.off()

    write(paste("\\includegraphics[width=0.33\\linewidth]{{",path,"}}",sep=""), file=pathConn, append=T)

    oldtime<-minOnly$time_conv[1]
    for (j in 2:(nrow(case))){
      if ((case$time_conv[j] - oldtime) > 120){
        minOnly<-rbind(minOnly, case[j,])
        oldtime<-case$time_conv[j]
      }
    }
    minOnly$errE2<-"na"
    minOnly$errN2<-"na"
    minOnly$waveMagN2<-"na"
    minOnly$waveMagE2<-"na"
    imax <- nrow(minOnly)
    minOnly$errE2[2:imax] <-
      round(
        (sin(minOnly$heading[1:(imax-1)]*pi/180)*(minOnly$speed[1:(imax-1)]*(minOnly$time_conv[2:imax]- minOnly$time_conv[1:(imax-1)])))
        -(minOnly$UTMlog[2:imax]-minOnly$UTMlog[1:(imax-1)])
        ,digits=2)

    minOnly$errN2[2:imax] <-
      round(
        (cos(minOnly$heading[1:(imax-1)]*pi/180)*(minOnly$speed[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)])))
        -(minOnly$UTMlat[2:imax]-minOnly$UTMlat[1:(imax-1)])
        ,digits=2)

    minOnly$waveMagE2[2:imax]<- round(minOnly$EU[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)]),digits=4)
    minOnly$waveMagN2[2:imax]<- round(minOnly$EV[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)]),digits=4)

    path<-file.path(plotpath,paste("Case_",i,"_","CorrelationEast2Min",".jpg",sep = ""))
    png(file=path)
    plot(minOnly$waveMagE2,minOnly$errE2,asp=1,
         main = paste("Ship MMSI",minOnly$mmsi[1],"\nError East vs Prediction Error East\n2Min R =", correlation$East2min[i]),
         xlab = "Wave movement eastward (m)",
         ylab = "Displacement eastward (m)"
    )
    abline(lm(as.numeric(minOnly$errE)~as.numeric(minOnly$waveMagE)))
    dev.off()

    write(paste("\\includegraphics[width=0.33\\linewidth]{{",path,"}}",sep=""), file=pathConn, append=T)

    path<-file.path(plotpath,paste("Case_",i,"_","CorrelationNorth2Min",".jpg",sep = ""))
    png(file=path)
    plot(minOnly$waveMagN2,minOnly$errN2,asp=1,
         main = paste("Ship MMSI",minOnly$mmsi[1],"\nError North vs Prediction Error North\n2Min R =", correlation$North2min[i]),
         xlab = "Wave movement northward (m)",
         ylab = "Displacement northward (m)"
    )
    abline(lm(as.numeric(minOnly$errN2)~as.numeric(minOnly$waveMagN2)))
    dev.off()

    write(paste("\\includegraphics[width=0.33\\linewidth]{{",path,"}}",sep=""), file=pathConn, append=T)

    ####5min Break######

    minOnly<-case[1,]
    oldtime<-minOnly$time_conv[1]
```

```
    for (j in 2:(nrow(case))){
      if ((case$time_conv[j] - oldtime) > 300){
        minOnly<-rbind(minOnly, case[j,])
        oldtime<-case$time_conv[j]
      }
    }
    minOnly$errE2<-"na"
    minOnly$errN2<-"na"
    minOnly$waveMagN2<-"na"
    minOnly$waveMagE2<-"na"
    imax <- nrow(minOnly)
    minOnly$errE2[2:imax] <-
      round(
        (sin(minOnly$heading[1:(imax-1)]*pi/180)*(minOnly$speed[1:(imax-1)]*(minOnly$time_conv[2:imax]- minOnly$time_conv[1:(imax-1)])))
        -(minOnly$UTMlog[2:imax]-minOnly$UTMlog[1:(imax-1)])
        ,digits=2)

    minOnly$errN2[2:imax] <-
      round(
        (cos(minOnly$heading[1:(imax-1)]*pi/180)*(minOnly$speed[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)])))
        -(minOnly$UTMlat[2:imax]-minOnly$UTMlat[1:(imax-1)])
        ,digits=2)

    minOnly$waveMagE2[2:imax]<- round(minOnly$EU[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)]),digits=4)
    minOnly$waveMagN2[2:imax]<- round(minOnly$EV[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)]),digits=4)

    path<-file.path(plotpath,paste("Case_",i,"_","CorrelationEast5Min",".jpg",sep = ""))
    png(file=path)
    plot(minOnly$waveMagE2,minOnly$errE2,asp=1,
         main = paste("Ship MMSI",minOnly$mmsi[1],"\nError East vs Prediction Error East\n5Min R =", correlation$East5min[i]),
         xlab = "Wave movement eastward (m)",
         ylab = "Displacement eastward (m)"
    )
    abline(lm(as.numeric(minOnly$errE)~as.numeric(minOnly$waveMagE)))
    dev.off()

    write(paste("\\includegraphics[width=0.33\\linewidth]{{",path,"}}",sep=""), file=pathConn, append=T)

    path<-file.path(plotpath,paste("Case_",i,"_","CorrelationNorth5Min",".jpg",sep = ""))
    png(file=path)
    plot(minOnly$waveMagN2,minOnly$errN2,asp=1,
         main = paste("Ship MMSI",minOnly$mmsi[1],"\nError North vs Prediction Error North\n5Min R =", correlation$North5min[i]),
         xlab = "Wave movement northward (m)",
         ylab = "Displacement northward (m)"
    )
    abline(lm(as.numeric(minOnly$errN2)~as.numeric(minOnly$waveMagN2)))
    dev.off()

    write(paste("\\includegraphics[width=0.33\\linewidth]{{",path,"}}",sep=""), file=pathConn, append=T)
    #####10 min Break #####

    minOnly<-case[1,]
    oldtime<-minOnly$time_conv[1]
    for (j in 2:(nrow(case))){
      if ((case$time_conv[j] - oldtime) > 600){
        minOnly<-rbind(minOnly, case[j,])
        oldtime<-case$time_conv[j]
      }
    }
    minOnly$errE2<-"na"
    minOnly$errN2<-"na"
    minOnly$waveMagN2<-"na"
    minOnly$waveMagE2<-"na"
    imax <- nrow(minOnly)
    minOnly$errE2[2:imax] <-
      round(
        (sin(minOnly$heading[1:(imax-1)]*pi/180)*(minOnly$speed[1:(imax-1)]*(minOnly$time_conv[2:imax]- minOnly$time_conv[1:(imax-1)])))
        -(minOnly$UTMlog[2:imax]-minOnly$UTMlog[1:(imax-1)])
        ,digits=2)

    minOnly$errN2[2:imax] <-
      round(
        (cos(minOnly$heading[1:(imax-1)]*pi/180)*(minOnly$speed[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)])))
        -(minOnly$UTMlat[2:imax]-minOnly$UTMlat[1:(imax-1)])
        ,digits=2)

    minOnly$waveMagE2[2:imax]<- round(minOnly$EU[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)]),digits=4)
    minOnly$waveMagN2[2:imax]<- round(minOnly$EV[1:(imax-1)]*(minOnly$time_conv[2:imax] - minOnly$time_conv[1:(imax-1)]),digits=4)

    if(nrow(minOnly)>2){
    path<-file.path(plotpath,paste("Case_",i,"_","CorrelationEast10Min",".jpg",sep = ""))
    png(file=path)
    plot(minOnly$waveMagE2,minOnly$errE2,asp=1,
         main = paste("Ship MMSI",minOnly$mmsi[1],"\nError East vs Prediction Error East\n10Min R =", correlation$East10min[i]),
         xlab = "Wave movement eastward (m)",
         ylab = "Displacement eastward (m)"
    )
    abline(lm(as.numeric(minOnly$errE)~as.numeric(minOnly$waveMagE)))
    dev.off()

    write(paste("\\includegraphics[width=0.33\\linewidth]{{",path,"}}",sep=""), file=pathConn, append=T)

    path<-file.path(plotpath,paste("Case_",i,"_","CorrelationNorth10Min",".jpg",sep = ""))
    png(file=path)
    plot(minOnly$waveMagN2,minOnly$errN2,asp=1,
         main = paste("Ship MMSI",minOnly$mmsi[1],"\nError North vs Prediction Error North\n10Min R =", correlation$North10min[i]),
         xlab = "Wave movement northward (m)",
         ylab = "Displacement northward (m)"
    )
    abline(lm(as.numeric(minOnly$errN2)~as.numeric(minOnly$waveMagN2)))
    dev.off()

    write(paste("\\includegraphics[width=0.33\\linewidth]{{",path,"}}",sep=""), file=pathConn, append=T)
  }
    })
}

proc.time() - ptm
```

```r
### Gap Analysis of Study Area

# There are 3,970,486 messages in the Longbeach data set
Gaps<- ships[order(ships$mmsi,ships$time),]
Gaps<- Gaps[!is.na(Gaps$mmsi),]
Gaps$timeDiff <- c(0, diff(Gaps$time_conv))

cord.dec = SpatialPoints(cbind(Gaps$longitude, Gaps$latitude), proj4string = CRS("+proj=longlat"))  #Collecting the Long Lat points of the data
cord.UTM <- spTransform(cord.dec, CRS("+init=epsg:26711")) #Transforming points to UTM by Code for UTM zone 11 Nad 83
Gaps$UTMlog<-cord.UTM$coords.x1
Gaps$UTMlat<-cord.UTM$coords.x2

rm(cord.dec)
rm(cord.UTM)

Gaps$distLat <- c(0, diff(Gaps$UTMlat))
Gaps$distLon<- c(0,diff(Gaps$UTMlog))
Gaps$distDiff<- sqrt((Gaps$distLon^2)+(Gaps$distLat^2))

# distDiff is the gap between messages
#Need to now remove the first message from each mmsi because their time diff and distance gap are based
# on the previous message of a different mmsi

FirstMessage<-c()
oldTrip<-Gaps$mmsi[1]
for (i in 2:nrow(Gaps)){
  if (Gaps$mmsi[i] != oldTrip){
    oldTrip<-Gaps$mmsi[i]
    FirstMessage<-c(FirstMessage,i)}
}
GapsRM <-Gaps[-(FirstMessage),]

#### Determineing the number of occurences when a Gap in time greater then 6 hours occur
GapsRM$refKey<- 1:nrow(GapsRM)

Hour6<- GapsRM[GapsRM$timeDiff>21600,]

### Of those which had a distance Gap greater then 5 km
FullGaps<-Hour6[Hour6$distDiff>5000,]

keyval<-c(FullGaps$refKey,FullGaps$refKey-1)
#######Taking the The start and end point of gaps as points to input into ArcMap
GapPoints<- GapsRM[GapsRM$refKey %in% keyval,]

write.csv(GapPoints,"E:/Masters/MappingFiles/GapPoints.csv",sep = ",",col.names = TRUE)
```

Appendix B

SQL code

```
porting CSV into Postgre

-- Need to Create  Table to hold data in PostSQL
        --CREATE TABLE command
-- Then create Colomn for each field in CSV
        -- create table shipdata (MMSI int, MESSAGE_ID int, REPEAT_INDICATOR int, TIME date, MILLISECOND int, REGION int, COUNTRY int,
              BASE_STATION int, ONLINE_DATA varchar, GROUP_CODE varchar,
SEQUENCE_ID int, CHANNEL char(1), DATA_LENGTH varchar, VESSEL_NAME varchar, CALL_SIGN varchar, IMO int, SHIP_TYPE int, DIMENSION_TO_BOW
       int, DIMENSIO_TO_STERN int, DIMENSIO_TO_PORT int,
DIMENSIO_TO_STARBOARD int, DRAUGHT int, DESTINATION varchar, AIS_VERSION int, NAVIGATION_STATUS int, ROT double precision, SOG double
       precision, ACCURACY int, LONGITUDE double precision,
LATITUDE double precision, COG double precision, HEADING varchar, REGIONAL int, MANEUVER int, RAIM_FLAG int, COMMUNICATION_FLAG int,
       COMMUNICATION_STATE int, UTC_YEAR int, UTC_MONTH int,
UTC_DAY int, UTC_HOUR int, UTC_MINUTE int, UTC_SECOND int, FIXING_DEVICE int, TRANSMISSION_CONTROL int, ETA_MONTH int, ETA_DAY int,
       ETA_HOUR int, ETA_MINUTE int, SEQUENCE varchar, DESTINATION_ID varchar,
RETRANSMIT_FLAG varchar, COUNTRY_CODE varchar, FUNTIONAL_ID varchar, DATA varchar, DESTINATION_ID_1 varchar, SEQUENCE_1 varchar,
       DESTINATION_ID_2 varchar, SEQUENCE_2 varchar,
DESTINATION_ID_3 varchar, SEQUENCE_3 varchar, DESTINATION_ID_4 varchar, SEQUENCE_4 varchar, ALTITUDE varchar, ALTITUDE_SENSOR varchar,
       DATA_TERMINAL int, MODE int, SAFETY_TEXT varchar,
NONSTANDARD_BITS int, NAME_EXTENSION varchar, NAME_EXTENSION_PADDING varchar, MESSAGE_ID_1_1 int, OFFSET_1_1 int, MESSAGE_ID_1_2 int,
       OFFSET_1_2 int, MESSAGE_ID_2_1 int, OFFSET_2_1 int,
DESTINATION_ID_A int, OFFSET_A int, INCREMENT_A int, DESTINATION_ID_B int, OFFSETB int, INCREMENTB int, DATA_MSG_TYPE int, STATIO_ID int,
        Z_COUNT int, NUM_DATA_WORDS int, HEALTH int, UNIT_FLAG int,
DISPLAY int, DSC int, BAND int, MSG22 int, OFFSET1 int, NUM_SLOTS1 int, TIMEOUT1 int, INCREMENT_1 int, OFFSET2 int, NUM_SLOTS2 int,
       TIMEOUT2 int, INCREMENT_2 int, OFFSET3 int, NUM_SLOTS3 int, TIMEOUT3 int,
INCREMENT_3 int, OFFSET4 int, NUM_SLOTS4 int, TIMEOUT4 int, INCREMENT_4 int, ATON_type int, ATON_name varchar, off_position int,
       ATON_status int, VIRTUAL_ATON int, CHANNEL_A int, CHANNEL_B int, TX_RX_MODE int, POWER int, MESSAGE_INDICATOR int,
       CHANNEL_A_BANDWIDTH int,
CHANNEL_B_BANDWIDTH int, TRANSZONE_SIZE int, LONGITUDE_1 double precision, LATITUDE_1 double precision, LONGITUDE_2 double precision,
       LATITUDE_2 double precision, STATION_TYPE int, REPORT_INTERVAL int,
QUIET_TIME int, PART_NUMBER int, VENDOR_ID varchar, MOTHER_SHIP_MMSI int, DESTINATION_INDICATOR int, BIANARY_FLAG int, GNSS_STATUS int,
       SPARE int, SPARE2 int, SPARE3 int, SPARE4 int)

-- Next loading the csv data into PostgreSQL

for /home/postgresdata/Ship_CSV (*csv) copy from DELIMITERS ',' CSV public.%~ships_data  > %~ships_data.sql
for %f (*sql) do psql -h myserver -d myDB -U postgres -f ships

ALTERNATE: create table shipdata (MMSI varchar, MESSAGE_ID varchar, REPEAT_INDICATOR varchar, TIME varchar, MILLISECOND varchar, REGION
       varchar, COUNTRY varchar, BASE_STATION varchar, ONLINE_DATA varchar, GROUP_CODE varchar,
SEQUENCE_ID varchar, CHANNEL varchar, DATA_LENGTH varchar, VESSEL_NAME varchar, CALL_SIGN varchar, IMO varchar, SHIP_TYPE varchar,
       DIMENSION_TO_BOW varchar, DIMENSIO_TO_STERN varchar, DIMENSIO_TO_PORT varchar,
DIMENSIO_TO_STARBOARD varchar, DRAUGHT varchar, DESTINATION varchar, AIS_VERSION varchar, NAVIGATION_STATUS varchar, ROT varchar, SOG
       varchar, ACCURACY varchar, LONGITUDE varchar,
LATITUDE varchar, COG varchar, HEADING varchar, REGIONAL varchar, MANEUVER varchar, RAIM_FLAG varchar, COMMUNICATION_FLAG varchar,
       COMMUNICATION_STATE varchar, UTC_YEAR varchar, UTC_MONTH varchar,
UTC_DAY varchar, UTC_HOUR varchar, UTC_MINUTE varchar, UTC_SECOND varchar, FIXING_DEVICE varchar, TRANSMISSION_CONTROL varchar, ETA_MONTH
        varchar, ETA_DAY varchar, ETA_HOUR varchar, ETA_MINUTE varchar, SEQUENCE varchar, DESTINATION_ID varchar,
RETRANSMIT_FLAG varchar, COUNTRY_CODE varchar, FUNTIONAL_ID varchar, DATA varchar, DESTINATION_ID_1 varchar, SEQUENCE_1 varchar,
       DESTINATION_ID_2 varchar, SEQUENCE_2 varchar,
DESTINATION_ID_3 varchar, SEQUENCE_3 varchar, DESTINATION_ID_4 varchar, SEQUENCE_4 varchar, ALTITUDE varchar, ALTITUDE_SENSOR varchar,
       DATA_TERMINAL varchar, MODE varchar, SAFETY_TEXT varchar,
NONSTANDARD_BITS varchar, NAME_EXTENSION varchar, NAME_EXTENSION_PADDING varchar, MESSAGE_ID_1_1 varchar, OFFSET_1_1 varchar,
       MESSAGE_ID_1_2 varchar, OFFSET_1_2 varchar, MESSAGE_ID_2_1 varchar, OFFSET_2_1 varchar,
DESTINATION_ID_A varchar, OFFSET_A varchar, INCREMENT_A varchar, DESTINATION_ID_B varchar, OFFSETB varchar, INCREMENTB varchar,
       DATA_MSG_TYPE varchar, STATIO_ID varchar, Z_COUNT varchar, NUM_DATA_WORDS varchar, HEALTH varchar, UNIT_FLAG varchar,
DISPLAY varchar, DSC varchar, BAND varchar, MSG22 varchar, OFFSET1 varchar, NUM_SLOTS1 varchar, TIMEOUT1 varchar, INCREMENT_1 varchar,
       OFFSET2 varchar, NUM_SLOTS2 varchar, TIMEOUT2 varchar, INCREMENT_2 varchar, OFFSET3 varchar, NUM_SLOTS3 varchar, TIMEOUT3 varchar,
INCREMENT_3 varchar, OFFSET4 varchar, NUM_SLOTS4 varchar, TIMEOUT4 varchar, INCREMENT_4 varchar, ATON_type varchar, ATON_name varchar,
       off_position varchar, ATON_status varchar, VIRTUAL_ATON varchar, CHANNEL_A varchar, CHANNEL_B varchar, TX_RX_MODE varchar, POWER
       varchar, MESSAGE_INDICATOR varchar, CHANNEL_A_BANDWIDTH varchar,
CHANNEL_B_BANDWIDTH varchar, TRANSZONE_SIZE varchar, LONGITUDE_1 varchar, LATITUDE_1 varchar, LONGITUDE_2 varchar, LATITUDE_2 varchar,
       STATION_TYPE varchar, REPORT_INTERVAL varchar,
QUIET_TIME varchar, PART_NUMBER varchar, VENDOR_ID varchar, MOTHER_SHIP_MMSI varchar, DESTINATION_INDICATOR varchar, BIANARY_FLAG varchar
       , GNSS_STATUS varchar, SPARE varchar, SPARE2 varchar, SPARE3 varchar, SPARE4 varchar);

COPY shipdata FROM '/home/databack/GSTS_historical_data_2012-11-011.csv'CSV HEADER;

-----FIRST CHANGE COLUMN TYPES SEE EMAIL

--- Data Cleaning for ships Database


--- Removing MMSI <200000000

DELETE FROM shipdata
WHERE shipdata.mmsi <200000000;
COMMIT;

--- Remove Null Geometries

DELETE FROM shipdata
```

```
WHERE message_id != 5 AND latitude IS NULL;
COMMIT;


-- Data splitting

SELECT DISTINCT ON (mmsi) * INTO shipdata_static  FROM shipdata
WHERE message_id = 5;
COMMIT;

SELECT * INTO shipdata_ships FROM shipdata
WHERE message_id = 1 OR message_id = 2 OR message_id = 3;
COMMIT;

--Deleting Duplicates

SELECT DISTINCT ON (MMSI,TIME, MILLISECOND) * INTO shipdata_ships2  FROM shipdata_ships;
COMMIT;

DROP TABLE shipdata_ships;
COMMIT;

-- Add Lat,Long fields
--ALTER TABLE shipdata_ships2
--ADD COLUMN Latitude double precision;

--ALTER TABLE shipdata_ships2
--ADD COLUMN Longitude double precision;


--UPDATE shipdata_ships2
-- SET latitude =  ST_Y("position");

--UPDATE shipdata_ships2
-- SET longitude =  ST_X("position");

-- Add Geometry
ALTER TABLE shipdata_ships2 ADD COLUMN gid serial PRIMARY KEY;
COMMIT;
ALTER TABLE shipdata_ships2 ADD COLUMN position geometry(POINT,4326);
UPDATE shipdata_ships2 SET position = ST_SetSRID(ST_MakePoint(longitude,latitude),4326);
COMMIT;

CREATE INDEX idx_ships_geom ON shipdata_ships2 USING GIST(position);


ALTER TABLE shipdata_ships2 DROP COLUMN COUNTRY, DROP COLUMN Millisecond, DROP COLUMN Region, DROP COLUMN Online_data, DROP COLUMN
      Group_Code, DROP COLUMN Sequence_ID, DROP COLUMN Channel, DROP COLUMN Data_Length, DROP COLUMN AIS_version, DROP COLUMN Regional,
      DROP COLUMN RAIM_flag, DROP COLUMN Communication_flag, DROP COLUMN Communication_state, DROP COLUMN Fixing_device, DROP COLUMN
      Transmission_control, DROP COLUMN Sequence, DROP COLUMN Destination_ID, DROP COLUMN Retransmit_flag, DROP COLUMN Country_code, DROP
       COLUMN Funtional_ID, DROP COLUMN Data, DROP COLUMN Destination_ID_1, DROP COLUMN Sequence_1, DROP COLUMN Destination_ID_2, DROP
      COLUMN Sequence_2, DROP COLUMN Destination_ID_3, DROP COLUMN Sequence_3, DROP COLUMN Destination_ID_4, DROP COLUMN Sequence_4, DROP
       COLUMN Altitude, DROP COLUMN Altitude_sensor, DROP COLUMN Data_terminal, DROP COLUMN Mode, DROP COLUMN Safety_text, DROP COLUMN
      Nonstandard_bits, DROP COLUMN Name_extension, DROP COLUMN Name_extension_padding, DROP COLUMN Message_ID_1_1, DROP COLUMN
      Offset_1_1, DROP COLUMN Message_ID_1_2, DROP COLUMN Offset_1_2, DROP COLUMN Message_ID_2_1, DROP COLUMN Offset_2_1, DROP COLUMN
      Destination_ID_A, DROP COLUMN Offset_A, DROP COLUMN Increment_A, DROP COLUMN Destination_ID_B, DROP COLUMN OffsetB, DROP COLUMN
      incrementB, DROP COLUMN data_msg_type, DROP COLUMN statio_ID, DROP COLUMN Z_count, DROP COLUMN num_data_words, DROP COLUMN health,
      DROP COLUMN unit_flag, DROP COLUMN display, DROP COLUMN DSC, DROP COLUMN band, DROP COLUMN msg22, DROP COLUMN offset1, DROP COLUMN
      num_slots1, DROP COLUMN timeout1, DROP COLUMN Increment_1, DROP COLUMN Offset2, DROP COLUMN Num_slots2, DROP COLUMN Timeout2, DROP
      COLUMN Increment_2, DROP COLUMN Offset3, DROP COLUMN Num_slots3, DROP COLUMN Timeout3, DROP COLUMN Increment_3, DROP COLUMN Offset4
      , DROP COLUMN Num_slots4, DROP COLUMN Timeout4, DROP COLUMN Increment_4, DROP COLUMN ATON_type, DROP COLUMN ATON_name, DROP COLUMN
      off_position, DROP COLUMN ATON_status, DROP COLUMN Virtual_ATON, DROP COLUMN Channel_A, DROP COLUMN Channel_B, DROP COLUMN
      Tx_Rx_mode, DROP COLUMN Power, DROP COLUMN Message_indicator, DROP COLUMN Channel_A_bandwidth, DROP COLUMN Channel_B_bandwidth,
      DROP COLUMN Transzone_size, DROP COLUMN Longitude_1, DROP COLUMN Latitude_1, DROP COLUMN Longitude_2, DROP COLUMN Latitude_2, DROP
      COLUMN Station_Type, DROP COLUMN Report_Interval, DROP COLUMN Quiet_Time, DROP COLUMN Part_Number, DROP COLUMN Vendor_ID, DROP
      COLUMN Mother_ship_MMSI, DROP COLUMN Destination_indicator, DROP COLUMN Bianary_flag, DROP COLUMN GNSS_status, DROP COLUMN spare,
      DROP COLUMN spare2, DROP COLUMN spare3, DROP COLUMN spare4;


### Code to select the "Anomolies" of when AIS turns off or loss of satellite signal ###


- Threshold of time = 24 hours # It lost 4 consecutive times of a sattelite passing


- Threshold on DIstance = 5 km

### PLACE INFORMATION FROM THE PREVIOUS ROW ONTO SAME LINE ###

SELECT loc_a.mmsi, loc_a.position AS startpoint, loc_b.position AS endpoint,
     loc_a.time AS starttime, loc_b.time AS endtime, loc_a.tripid AS startid, loc_b.tripid AS endid
  INTO cargogaps
  FROM
    (SELECT *, ROW_NUMBER() OVER (ORDER BY mmsi, "time") AS num FROM cargoshipsorder) AS loc_a
    JOIN
    (SELECT *, ROW_NUMBER() OVER (ORDER BY mmsi, "time") AS num FROM cargoshipsorder) AS loc_b
      ON (loc_a.mmsi = loc_b.mmsi AND
         (loc_a.num + 1) = loc_b.num)
 WHERE loc_a.time <> loc_b.time;


### SELECT THE ANOMOLIES INTO A NEW TABLE ### -- edited May 19 by colin to add x and y coords to table output

SELECT mmsi,ST_X(startpoint) as startpointx, ST_Y(startpoint) as startpointy,ST_X(endpoint) as endpointx, ST_Y(endpoint) as endpointy,
      starttime, endtime, startid, endid INTO cargogapsFixedCR FROM cargogaps WHERE (EXTRACT(EPOCH FROM endtime) - EXTRACT(EPOCH FROM
      starttime)) >= (24*60*60) AND SQRT(((ST_X(startpoint)- ST_X(endpoint))^2)+((ST_Y(startpoint)-ST_Y(endpoint))^2)) >= 5000;
```

Appendix C

Addendum

This section consists of the extra material and information suggested by the defence committee. Provided are the topics that needed to be addressed and expanded upon to give more insight into the thesis. Explained is the way the AIS message is generated and the path it takes into becoming a S-AIS message for this thesis. This path produces areas of uncertainty which are highlighted as to how they may have been created. Next the description of the significance values are expanded upon and how they were used in the thesis. Also other methods for viewing data relationships are suggested when working with large datasets. Then the addendum explains the physics involved when a wave hits a ship and the energy that is produced by the wave. Finally all the components that can attribute to the error vector are determined and why wave magnitude was the only contributing factor used for this thesis.

C.1 Diagram of AIS system's path way of signal and data uncertainty

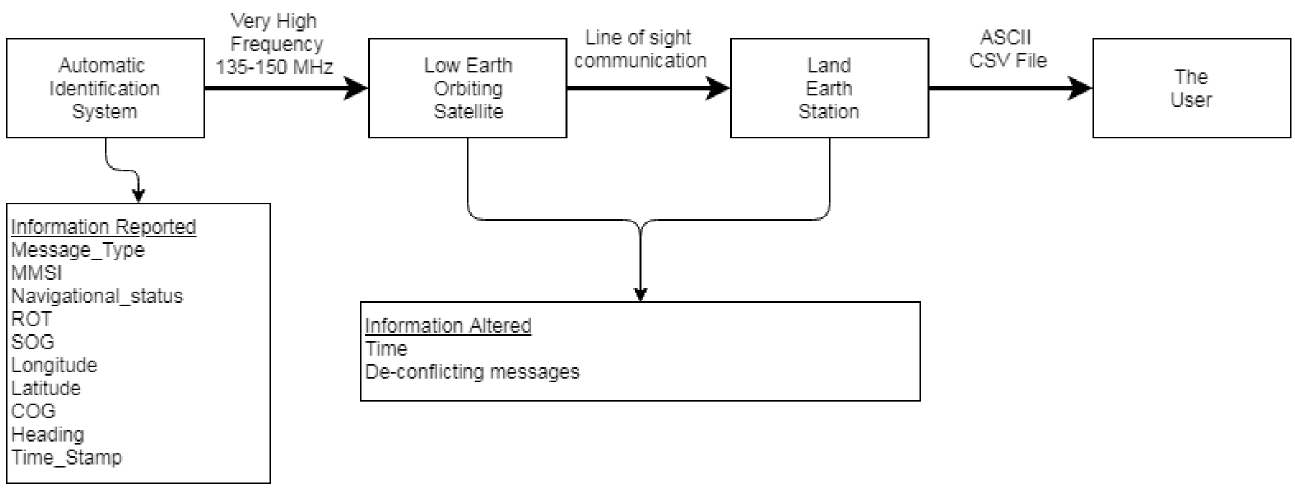

Figure C.1: The pathway from the ship's AIS signal getting to the user

Figure C.1 is designed to show information that was used in this study and provided through the AIS compared to what information may have been added afterwards through satellite or processing after the message was initially sent. The AIS on board a ship generates the ships message type, MMSI, navigational status, and heading. The ROT of the ship is provided by a rate of turn sensor that gives the information to the AIS. The SOG, longitude, latitude and COG are all provided by an on board Electronic Positioning System (EPS) that determines the ships current coordinates and calculates speed and direction. This information is then provided to the AIS so it can send the message. The last information the AIS sends is a time stamp that is only in UTC seconds that is generated by the electronic position fixing system (EPFS) (International Telecommunications Union, 2010). The EPS and EPFS on a ship are not all the same so the accuracy is different based on what positioning system is being used. The EPFS generates an accuracy values that is sent by the AIS which reports if the accuracy of the reported position is greater or less than 10 metres (International Telecommunications Union, 2010). This uncertainty in the data is of concern and needs to be considered in further research especially with message to message comparisons that travel at smaller distances. All this information should be continually broadcasted to nearby receivers such as LEO satellites.

When the AIS message is received by the satellite there are two possible options that occur. The satellite will store the AIS message and either wait until it is in line of sight of a Land Earth Station (LES) to send the messages, or the satellite will complete on board processing of the message and generate a time value for the now S-AIS message before storing and sending it to the LES. The time value is an important piece of information in this study since it was used as the assumed time when the report was sent. Since the time value is not generated by the AIS it subjected to uncertainty in its value. There is also chance of lost message that occurs when a satellite potentially reaches AIS message capacity. The satellite has limited memory to store messages and must be in line of sight of the LES to transmit those messages, this can lead to gaps in messages even when a satellite is present in an area (International Telecommunications Union, 2010). Satellites can also miss collecting AIS messages through Very High Frequency (VHF) interference from different reporting AIS devices. In high traffic areas there is greater interference that occurs which causes loss of data and uncertainty in AIS message sequences. When the S-AIS messages are sent to the LES it will use de-conflicting algorithms to remove messages that may be repeated or conflicting in some way (Carson-Jackson, 2012). Due to the companies keeping all this information hidden, since the algorithms and collecting process is proprietary, it is undetermined how the messages were processed before being received by a user adding more uncertainty to the user (Carson-Jackson, 2012). The S-AIS messages are then sent to a user in the form of ASCII files so they may be analyzed like in this thesis.

There are issues of uncertainty with working with HF radar wave data. Since the wave data consists of hourly averages it neglects effects that can be caused by larger rogue waves which can greatly disrupt the movement of a ship (Chakraborty, 2018). Large wave magnitudes can also disrupt the manoeuvres a ship can under go altering its travel as reported by Szelangiewicz, 2014. The uncertainty in both S-AIS and HF radar data must be looked into for further studies to find methods to minimize overall uncertainty.

C.2 Description of significance

In this thesis the use of the p-values of the linear regressions of correlating wave magnitudes and error vectors were statistically significant in every case because of the large sample sizes in all of the cases. Significance may have not been relevant in mentioning in this thesis when looking at sample sizes of this magnitude. Since the significance values are always strong it means the correlation test requires more attention to determine the relationship. In the thesis a Pearson correlation coefficient was used to measure the linear relation between wave magnitude and error in prediction. During the testing of correlation it may have been beneficial to use other methods of correlation testing like a non parametric correlation through Spearman rank-order. This may have helped determine trends in the data that did not show up with the Pearson correlation. Another way the error vectors and wave magnitudes could have been compared for a relation is through a probability plot, this would show if there is a form of normal distribution between the wave magnitudes and error vectors. With the misleading significance values it is more important to focus on the tests on finding relation between the data which we were unable to find through out this thesis.

C.3 Physics involved in wave impacts

The physics of a ships motion in waves can be described as seakeeping theory and the way the oscillating waves impact a ship is through six degrees of freedom called surge, sway, heave, heel, yaw, and pitch (Fossen, 2011). The waves act upon the wetted ship surface and the pressure applied by the wave will effect the motion of a ship (Nikolai Kornev, 2012). The waves acting on a ship will vary based on the period and velocity of a ship. “Short waves induce larger steady wave forces than long waves” (Ueno et al., 2000), and since there is always present short waves on the ocean the ship is continually effected. These waves effect the ship in its degrees of freedom which will change the inherit movement of the ship compared to a stable water condition. The force a wave can apply depends on the, water salinity, wave height, wave thickness, and speed. Seraphin et al., (2018) in Hawaii determined “a wave with a height of 2 metres and a wavelength of 14 metres breaking along 2 kilometres of coastline has approximately 45 kWh of energy. This is roughly equivalent to one gallon of gasoline”. A report created by cruise line Cunard calculated that the Queen Elizabeth 2, which has a gross tonnage of 70,327 tonnes, would travel 12.5 meters per gallon of fuel (Cunard, 2009). The Queen Elizabeth 2 has a wetted area of roughly 3000 m^2 so it would require 11 waves with a height of 2 metres to move the ship 12 metres. There are significantly more factors involved in determining the waves effectiveness in moving a ship in terms of drag, orbital motion of the waves, and ship resistance that make determining waves physical force on a object difficult.

C.4 What comprises of the error vector

When the error vector was calculated the prediction of the ship's future position was based on its dead reckoning. This takes the heading of a ship and current speed to determine a position over time, with no interference and initial position point accuracy then the predicted position should be the same as the next position message. The error vector is measured to be the displacement from the predicted position and reported position. This error is attributed to external forces acting on the position of the ship and in this thesis the only contributing factor that acted on this error was assumed to be wave movement. There are other factors though that the error vector is comprised of, the reason for using only waves was due to its magnitude relative to wind as studied in Szelangiewicz, 2014. Some of the errors that comprise of the error vector are wind, water density, AIS position error, physical ship information.

Waves contribute to the movement of a ship as discussed in Szelangiewicz, 2014 where a ship required to reduce max speeds by 20% to avoid rolling in high wind conditions. From the literature review it was determined though that wave velocity had a larger impact on the movement of a ship which is why wind was not used as well in this study. Another contributor to error in prediction can also be water density. To compare the water closer to the equator versus away from the equator the water temperature causes colder water to become more dense which will change the sea-keeping dynamics of a ship. In denser water a ship will float higher above the surface which can increase the force applied on the ship by wind. The next addition of error contribution is the AIS position error that come from the on board electronic position system of a ship. The ship's position system will have inherit error involved, there is a reported accuracy measurement in AIS messages that was not used in this study that will report if the provided position in less than or greater than 10 metres in position accuracy (International Telecommunications Union, 2010). When looking at point to point positional change this should have been taken into consideration. The last source that can be contributed to the error is unknown ship information such as rudder position or anchor type. This information is not provided in AIS messages and could provide some additional factors that effect the movement of a ship.

Bibliography

Abascal, Ana Julia, Sonia Castanedo, Vicente Fernández, and Raúl Medina (2012). "Backtracking drifting objects using surface currents from high-frequency (HF) radar technology". In: *Ocean Dynamics* 62.7, pp. 1073–1089. ISSN: 16167341. DOI: 10.1007/s10236-012-0546-4.

Balduzzi, Marco, Alessandro Pasta, and Kyle Wilhoit (2014). "A Security Evaluation of AIS Automated Identification System". In: *Proceedings of the 30th Annual Computer Security Applications Conference*, pp. 436–445. DOI: 10.1145/2664243.2664257.

Borkowski, Piotr (2017). "The ship movement trajectory prediction algorithm using navigational data fusion". In: *Sensors (Switzerland)* 17.6. ISSN: 14248220. DOI: 10.3390/s17061432.

Carson-Jackson, J. (2012). “Satellite AIS - Developing Technology or Existing Capability?” In: *Journal of Navigation* 65.July 2012, pp. 303–321. ISSN: 0373-4633. DOI: 10.1017/S037346331100066X.

Chakraborty, Soumya (2018). *Effects of Rogue Waves On Ships*. https://www.marineinsight.com/marine-navigation/effects-of-rogue-wave-on-ships/.

Chu, P. C., S. E. Miller, and J. a. Hansen (2013). “Fuel-saving ship route using the Navy’s ensemble meteorological and oceanic forecasts”. In: *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology* 12.1, pp. 41–56. ISSN: 1548-5129. DOI: 10.1177/1548512913516552. URL: http://dms.sagepub.com/cgi/doi/10.1177/1548512913516552.

Cunard (2009). *Queen Elizabeth 2 Technical Information*. https://web.archive.org/web/20090318012344/http://www.cunard.com/uploads/QE2_Tech.pdfl.

Dolinskaya, Irina (2012). “Optimal Path Finding in Direction,Location, and Time Dependent Environments”. In: *Naval Research Logistics* 59, pp. 325–339. DOI: 10.1002/nav.21492.

Dzvonkovskaya, Anna and Hermann Rohling (2010). “HF radar performance analysis based on AIS ship information”. In: *2010 IEEE (Institute of Electrical and Elec-*

tronics Engineers) National Radar Conference - Proceedings, pp. 1239–1244. ISSN: 10975659. DOI: 10.1109/RADAR.2010.5494430.

Fossen, Thor I. (2011). *Handbook of Marine Craft Hydrodynamics and Motion Control*.

Guichoux, Yann, Marc Lennon, and Nicolas Thomas (2016). "SEA SURFACE CURRENTS CALCULATION USING VESSEL TRACKING DATA". In: *Maritime knowledge discovery and anomaly detection workshop*. Joint Research Centre, Ispra (Italy), pp. 1–7.

Gunnar Aarsætherr, Karl and Torgeir Moan (2009). "Estimating Navigation Patterns from AIS". In: *Journal of Navigation* 62, p. 587. ISSN: 0373-4633. DOI: 10.1017/S0373463309990129.

Gurgel, Adikw and Ihrit Schlick (1996). "HF Radar WERA Application for Ship Detection and Tracking". In: *European Journal of Navigation* 7.3, pp. 18–25. URL: http://helzelgmbh.de/files/432/upload/Papers/EJN_ShipDetection-Tracking-2009.pdf.

Gurgel, K. W., H. H. Essen, and Kingsley (1999a). "High-frequency radars : physical limitations and recent developments". In: *Coastal Engineering* 37, pp. 201–218.

Gurgel, K. W., G. Antonischki, H. H. Essen, and T. Schlick (1999b). "Wellen Radar (WERA): A new ground-wave HF radar for ocean remote sensing". In: *Coastal En-*

gineering 37, pp. 219–234. ISSN: 03783839. DOI: 10.1016/S0378-3839(99)00027-7.

Gurgel, Klaus Werner, Thomas Schlick, Jochen Horstmann, and Salvatore Maresca (2010). "Evaluation of an HF-radar ship detection and tracking algorithm by comparison to AIS and SAR data". In: *2010 International Waterside Security Conference, WSS 2010*. DOI: 10.1109/WSSC.2010.5730246.

Harati-Mokhtari, Abbas, Alan Wall, Philip Brooks, and Jin Wang (2007a). "Automatic Identification System (AIS): Data Reliability and Human Error Implications". In: *Journal of Navigation* 60.03, p. 373. ISSN: 0373-4633. DOI: 10.1017/S0373463307004298. URL: http://www.journals.cambridge.org/abstract_S0373463307004298.

— (2007b). *Human Factors Aspect of Automatic Identification System (AIS) on the Ship's Bridge*.

Harlan, Jack, Eric Terrill, Lisa Hazard, Carolyn Keen, Donald Barrick, Chad Whelan, Stephan Howden, and Josh Kohut (2010). "The Integrated Ocean Observing System High-Frequency Radar Network: Status and Local, Regional, and National Applications". In: *Marine Technology Society Journal* 44, pp. 122–132. ISSN: 00253324. DOI: 10.4031/MTSJ.44.6.6.

Heij, Christiaan and Sabine Knapp (2015). "Effects of wind strength and wave height on ship incident risk : Regional trends and seasonality". In: *Transportation Research Part*

D 37, pp. 29–39. ISSN: 1361-9209. DOI: 10.1016/j.trd.2015.04.016. URL: http://dx.doi.org/10.1016/j.trd.2015.04.016.

Henningsen, Roar, Kjell Skjolsvik, Aage Andersen, James Corbett, and John Skjelvik (2000). *Study of Greenhouse Gas Emissions from Ships, Final Report to the International Maritime Organization*.

Heymann, Frank, Thoralf Noack, Paweł Bany, and Deutsches Zentrum (2011). *Plausibility analysis of navigation related AIS parameter based on time series*.

International Chamber of Shipping (2015). *ICS Kernel Description*. http://www.ics-shipping.org/.

International Maritime Organization (2016). *About IMO*. http://www.imo.org/en/About/Pages/Default.aspx.

International Telecommuniations Union (2010). “Technical characteristics for an automatic identification system using time-division multiple access in the VHF maritime mobile band”. In: *Lighthouse* 4, pp. 1–144.

International Telecommunications Union (2010). “Recommendation ITU-R M.1371-4”. In: *Lighthouse* 4, p. 144.

Lane, Richard O., David. a. Nevell, Steven D. Hayward, and Thomas W. Beaney (2010). “Maritime anomaly detection and threat assessment”. In: *2010 13th International Conference on Information Fusion*, pp. 1–8.

Last, Philipp, Christian Bahlke, Martin Hering-bertram, and Lars Linsen (2014). "Comprehensive Analysis of Automatic Identification System (AIS) Data in Regard to Vessel Movement Prediction". In: *The Journal of Navigation*. ISSN: 0373-4633. DOI: 10.1017/S0373463314000253.

Laws, K, J Vesecky, and J Paduan (2011). *Predicting the capabilities of ship monitoring by HF radar in coastal regions*.

Makridakis, Spyros and Michèle Hibon (2000). "The M3-Competition: results, conclusions and implications". In: *International Journal of Forecasting* 16.4, pp. 451–476. ISSN: 01692070. DOI: 10.1016/S0169-2070(00)00057-1. arXiv: /docs.google.com/a/deusto.es/file/d/0BzVwk6DUm3f4SXZnalVrTDdQV28/edit?pli=1 [https:].

Maulidi, Akhmad, Trika Pitana, Ketut Buda Artana, A A Bagus Dinariyana, Muhammad Badrus Zaman, Ahmad Agoes Masroeri, and Ricky Randall Sembiring (2014). "Database Integration Model for Automatic Identification System and Shipping Database In Real Time Traffic Monitoring". In: *Journal of Proceeding Series* 1, pp. 239–244.

Moreira, Lúcia, Thor I. Fossen, and C. Guedes Soares (2007). "Path following control system for a tanker ship model". In: *Ocean Engineering* 34, pp. 2074–2085. ISSN: 00298018. DOI: 10.1016/j.oceaneng.2007.02.005.

Natale, Fabrizio, Maurizio Gibin, Alfredo Alessandrini, Michele Vespe, and Anton Paulrud (2015). "Mapping fishing effort through AIS data". In: *PLoS ONE* 10.6, pp. 1–16. ISSN: 19326203. DOI: `10.1371/journal.pone.0130746`.

National Data Buoy Center (2017). *National Data Buoy Center*. `http://www.ndbc.noaa.gov/`.

National Oceanic and Atmospheric Administration (2013). *High Frequency Radar Surface Currents*. `https://tidesandcurrents.noaa.gov/hfradar/`.

Nikolai Kornev, Dr.-Ing. habil. (2012). *Ship dynamics in waves*. Universitat Rostock, pp. 1–117.

Pallotta, Giuliana, Michele Vespe, and Karna Bryan (2013). "Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction". In: *Entropy* 15, pp. 2218–2245. ISSN: 10994300. DOI: `10.3390/e15062218`.

Pallotta, Giuliana, Steven Horn, Paolo Braca, and Karna Bryan (2014). "Context-Enhanced Vessel Prediction Based On Ornstein-Uhlenbeck Processes Using Historical AIS Traffic Patterns : Real-World Experimental Results". In: *Information Fusion (FUSION), 2014 17th International Conference*, pp. 1–7. ISSN: 9788490123553.

Pearce, Fred (2009). *How 16 ships create as much pollution as all the cars in the world*. `http://www.dailymail.co.uk/sciencetech/article-1229857/How-16-ships-create-pollution-cars-world.html`.

Perez, Tristan (2007). *Modelling and Simulation of Marine Surface Vessel Dynamics*, pp. 1–16.

Pérez Arribas, F. (2007). “Some methods to obtain the added resistance of a ship advancing in waves”. In: *Ocean Engineering* 34, pp. 946–955. ISSN: 00298018. DOI: 10.1016/j.oceaneng.2006.06.002.

Ristic, B, B La Scala, and M Morelande (2008). “Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction”. In: *2008 11th International Conference on Information Fusion*, pp. 1–7. URL: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=4632190.

Sadowski, Jaroslaw and Agnieszka Czapiewska (2015). “Algorithms for Ship Movement Prediction for Location Data Compression”. In: *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation* 9, pp. 75 –81. ISSN: 2083-6473. DOI: 10.12716/1001.09.01.09.

Silveira, P A M, A P Teixeira, and C Guedes Soares (2013). “Use of AIS Data to Characterise Marine Traffic Patterns and Ship Collision Risk off the Coast of Portugal”. In: *the Journal of Navigation* 66, pp. 879–898. ISSN: 1469-7785. DOI: 10.1017/S0373463313000519.

Smith, Tristan, Eoin O'Keeffe, Lucy Aldous, and Paolo Agnolucci (2013). "Assessment of Shipping's Efficiency Using Satellite AIS data, Prepared for the International Council on Clean Transportation". In: March.

Szelangiewicz, Tadeusz (2014). "The influence of wind , wave and loading condition on total resistance and speed of the vessel". In: *Polish Maritime Research* 21.3, pp. 61–67.

Tsang, Edward (2015). *Combinatorial Explosion*. `http://cswww.essex.ac.uk/CSP/ComputationalFinanceTeaching/CombinatorialExplosion.html`.

Ueno, Michio, Tadashi Nimura, Hideki Miyazaki, and Koji Nonaka (2000). "Steady Wave Forces and Moment Acting on Ships in Manoeuvring Motion in Short Waves". In: *Journal of the Society of Naval Architects of Japan* 2000.188, pp. 163–172. DOI: `10.2534/jjasnaoe1968.2000.188_163`.

Vandecasteele, Arnaud and Aldo Napoli (2012). "An enhanced spatial reasoning ontology for maritime anomaly detection". In: *7th International Conference on System Of Systems Engineering - IEEE SOSE 2012*, pp. 247–252. DOI: `10.1109/SYSoSE.2012.6384120`.

Vesecky, John F., Kenneth E. Laws, and Jeffery D. Paduan (2010). "A system trade model for the monitoring of coastal vessels using HF surface wave radar and ship automatic identification systems (AIS)". In: *International Geoscience and Remote Sensing Sym-*

posium (IGARSS), pp. 3414–3417. ISSN: 2153-6996. DOI: 10.1109/IGARSS.2010.5650279.

Vespe, Michele, Harm Greidanus, and Marlene Alvarez Alvarez (2015). "The declining impact of piracy on maritime transport in the Indian Ocean: Statistical analysis of 5-year vessel tracking data". In: *Marine Policy* 59.December 2008, pp. 9–15. ISSN: 0308597X. DOI: 10.1016/j.marpol.2015.04.018. URL: http://linkinghub.elsevier.com/retrieve/pii/S0308597X15001001.

Vlachos, D. S. (2004). "Optimal Ship Routing Based on Wind and Wave Forecasts". In: *Applied Numerical Analysis & Computational Mathematics* 1.2, pp. 547–551. ISSN: 1611-8170. DOI: 10.1002/anac.200410018. URL: http://doi.wiley.com/10.1002/anac.200410018.

Wolff, Christian (2017). *Bragg-Scattering*. http://www.radartutorial.eu/07.waves/wa52.en.html.

Xiao, Fangliang, Han Ligteringen, Coen Van Gulijk, and Ben Ale (2015). "Comparison study on AIS data of ship traffic behavior". In: *Ocean Engineering* 95, pp. 84–93. ISSN: 0029-8018. DOI: 10.1016/j.oceaneng.2014.11.020. URL: http://dx.doi.org/10.1016/j.oceaneng.2014.11.020.

Zissis, Dimitrios, Elias K. Xidias, and Dimitrios Lekkas (2015). "Real-time vessel behavior prediction". In: *Evolving Systems*, pp. 1 –12. ISSN: 1868-6478. DOI: 10.1007/

s12530-015-9133-5. URL: http://link.springer.com/article/10.1007/s12530-015-9133-5{\#}.