Trajectory analysis using Automatic Identification System (AIS) in New Zealand Waters

Premalatha Sampath
(Student ID: 0784954)

This thesis is submitted as in partial fulfilment of Degree of Master of Computers and Information Sciences at the Auckland University of Technology

October 2012

Primary Supervisor : Dr. Dave Parry

Acknowledgements

I am very pleased to acknowledge the assistance and encouragement received from several people in the production of this thesis.

This thesis was completed in the Faculty of Design and Creative Technologies at AUT University in New Zealand. This thesis would not have been possible without the collaboration of many people, who in one way or another extended their valuable assistance.

First and foremost, I am heartily thankful to my primary supervisor, Dr Dave Parry, who has given tremendous support and guidance in my research over the last 15 months. His encouragement, creative ideas, constructive criticism, guidance and kind support from the initial to the final level enabled me to accomplish this research. This thesis would not have been possible without his support, supervision and help.

I also would like to thank Dr Subana Shanmuganathan from GeoInformatics Lab, who has kindly agreed to be my secondary supervisor. Her valuable inputs, suggestions, guidance and providing ArcGIS software for kind help during my thesis phase.

Next, I owe my deepest gratitude to Aaron Olphart of Kordia New Zealand providing Automatic Identification System (AIS) raw datasets to use in this thesis.

I would like to express my heartfelt thanks to colleagues at Whitireia New Zealand for their support. Without their support and encouragement, my study life at AUT would have been difficult. I would like to thank all of you.

I greatly appreciate the support of my husband, Natthan Kanthi, who also provided continual encouragement and took care of our two boys Benadiks Kanthi (5½ years) and Ryansun Kanthi (6 months) during my thesis.

Table of Contents

Table of Contents

Appendices

Table of Figures

Table of Tables:

Acronyms

AIS	Automatic Identification System
ECDIS	Electronic Chart Display and Information System
GPS	Global Positioning System
MMSI	Maritime Mobile Service Identity
NMEA	U.S.-based National Marine Electronics Association
SOG	Speed over ground (measured in knots).
SOLAS	Safety of Life at Sea.
UTC	Coordinated Universal Time.
VTS	Vessel Traffic Services.
IALA	International Association of Marine Aids to Navigation and Lighthouse Authorities (IALA)
IMO	International Maritime Organization
ETA	Estimated Time of Arrival (ETA)
RMC	Recommended Minimum Communication
TDMA	Time division multiple access
SOTDMA	Self-Organizing Time Division Multiple Access
ITU-R M	International Telecommunication Union (ITU) Radiocommunication Sector
NAVCEN	U.S. Coast Guard Navigation Center
CSTDMA	Carrier-sense Time Division Multiple Access

Glossary

Automatic Identification System (**AIS**) is an automatic tracking system used on ships and by vessel traffic services (VTS) for identifying and locating vessels by electronically exchanging data with other nearby ships and AIS Base stations.

ASCII. Acronym for American Standard Code for Information Interchange, a standard method of representing alphanumeric characters with numbers in a computer.

Azimuth. The horizontal direction or bearing of a celestial point from a terrestrial point, expressed as the angular distance from a reference direction. It is usually measured from 000° at the reference direction clockwise through 360°.

Decoder - a device which does the reverse operation of an encoder, undoing the encoding so that the original information can be retrieved. The same method used to encode is usually just reversed in order to decode. It is a combinational circuit that converts binary information from n input lines to a maximum of 2^n unique output lines.

Estimated time of arrival (ETA) - Estimated time of arrival at destination – UTC month/date hour:minute.

Greenwich time: Time based upon the Greenwich meridian as reference.

NMEA 0183 is a combined electrical and data specification for communication between marine electronic devices.

Path (trajectory): The trajectory used to move from an origin to a destination in space.

Schema: The structure in a database system that defines the relationship between tables and fields. The term is also used with XML, describing its structure and data types.

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning, except where due acknowledgement is made in the acknowledgements.

....

Signature

Trajectory analysis using Automatic Identification System (AIS) in New Zealand Waters

Abstract

Trajectory analysis is one of the most actively researched areas of spatio-temporal databases. Exploring and analysing large datasets of movement data has become a vital part of research in many disciplines and decision-making fields. The major challenge involved during the analysis process of trajectory data is to visualize, understand and extract meaningful patterns (Adrienko & Adrienko, 2011) out of millions of locations collected from **Automatic Identification Systems** (AIS) data points. AIS datasets are used in the maritime industry to assist in tracking and monitoring vessel movements. The ultimate aim of the study is to understand the characteristics of different types (Dodge, Weibel, & Forootan, 2009) of vessels using AIS movement data. The intention of the study is also to outline challenges encountered during this thesis and describe approaches taken to overcome them. AIS movement datasets are voluminous and are coded via a complex standard. Therefore, to conduct analysis on raw data to trajectory involved a two-phased methodological process. The first phase focused on development of a decoder to extract significant information from the raw data. The information extracted from movement data was then utilized to perform knowledge discovery in regard to dynamic objects. The second phase centred on trajectory analysis utilizing proposed spatio-temporal approach and clustering techniques. Each phase accounted for a part of the contribution made towards this thesis.

Phase 1 primarily focused on handling the large dataset and development of a decoder. Given a large dataset (2GB - 30 Million rows) of spatio-temporal movement observations, the goal was to perform a segmentation analysis into

sets of certain period of data. The objective of segmentation was to handle the datasets into manageable sizes to overcome the main memory issue. The purpose of the decoder is to decode the raw data to read the information signalled from the vessels. The decoder is developed in VB.net that could extract information such as **Maritime Mobile Service Identity** (**MMSI**), latitude, longitude, speed over ground (SOG) and time. The developed decoder has the capability to intake raw data in text files and output the decoded dataset in text files. As the decoded dataset is made available, it is imported to an Excel file. The most important challenge faced was clean the decoded data from noise, outliers and duplications. A filtering technique was utilized in the process of cleaning. Subsequently, the dataset is well prepared for the next phase of conducting trajectory analysis within the Geographical Information Systems (GIS). In Phase 2, trajectory analysis was conducted using a spatio-temporal approach that included methods such as projecting trajectories on ArcGIS, plotting velocity trend graphs (Guo, Liu, & Jin, 2010) and identifying clusters based on stop or start movements. Exploring and analysing large datasets of movement data is a central aspect of this research study. The research focused on methodological research aimed extracting the spatio-temporal information and patterns of moving objects.

The experimental results indicate that the proposed methods can be successfully applied to perform trajectory analysis on provided movement datasets of dynamic vessels.

1 Introduction

Modern positioning and identification technologies enable the collection of large amounts of data from moving objects. Data about movements of these objects are often called 'trajectories' in space and time. The movements are recorded as a series of geographic locations with respective time stamps. Exploring and analysing large datasets of movement data has become a vital part of research in many disciplines. There seems to be increased interest in many disciplines in the study of trajectory analysis. Such fields include behavioural sciences (Horne, Garton, Krone, & Lewis, 2007), ecology (Kubach, Scott, & Bulak, 2011), transportation modelling (Wang & Zimmermann, 2010), cartographic planning (Ping, Xinming, & Shengxiao, 2008), terrestrial change (Plumejeaud, Mathian, Gensel, & Grasland, 2011), urban transportation (Wolfson & Xu, 2010), marine traffic (Jiacaia, Qingshan, Jinxing, & Zheping, 2012) and ship movements (Willems, Wetering, & Wijk, 2011). To perform the analysis of trajectories requires the extraction of knowledge about these moving objects from massive datasets (Tsou, 2010; Willems, Hage, Vries, H.M., & Malaisé, 2010).

Analysis of trajectory data is becoming one of the most actively researched areas of spatio-temporal databases. The advent of modern positioning and identification devices has enabled the tracking of almost any type of moving object. A remarkable number of large datasets for these kinds of moving objects are available. The available data are diverse, ranging from domains such as the tracking of humans using GPS/smart devices, animal tracking using GPS collars, monitoring hurricane movements with the aid of satellite imagery, ship trajectories and movements using AIS, to name but a few. The availability of such repositories of these movement datasets brings challenges to the fields of knowledge discovery in-order to extract implicit and meaningful information from this data. The information could lead to the discovery of interesting patterns, to explore behaviour of moving objects (vehicles, animals, birds, humans) or spatio-temporal process (oil spills, catastrophic changes, hurricane movements).

Recognizing the growth in modern positioning and identification technologies, movement datasets from such technologies, have significantly increased research interest among the GIS community in developing methods, models and tools for the analysis of movement data, over the past few years. Accordingly, a large number of studies have been dedicated to the handling of movement data and their databases. This has been pioneered innovative approaches to exploit movement data in Knowledge Discovery (KD) process, namely trajectory data analysis (Spaccapietra, et al.,2008; Dodge, Weibel, & Forootan, 2009), pattern detection (Gidófalvi & Pedersen, 2009; Buchin et al, 2009; Laube & Purves, 2011), analogy behaviour (Etienne et al., 2010). Among the research exploring the similarities in the movement of multiple objects is challenging and emerging interest (Buchin et al., 2009; Etienne et al., 2010).

1.1 Motivation for Research

According to Spaccapietra, et al.,(2008) analysis of trajectory data aims to understand management of complex phenomena that are involved in moving objects. Dodge, Weibel, & Forootan (2009) conducted trajectory analysis from collected movement datasets of different vehicles in an urban city to identify similar trajectories. Buchin et al (2009) and Gidófalvi & Pedersen (2009) on trajectories regarding their movement pattern detection as the spatio-temporal expression. Etienne et al. (2010) investigated similar itineraries of moving objects and hence deduced its behaviour from massive collected movement datasets of moving vessels. None of these research efforts focused on the study of the functionality of identifying characteristics of different moving objects.

In addition, studying movement characteristics of an object in terms of essential movement parameters provides more relevant insight than simply the movement paths. The aim of this study is to be able to identify an object’s function with additional parameters to obtain detailed insight on the characteristics of the trajectories.

Previous studies have used different sources of data. Dodge et al., (2009), Buchin et al (2009) and Gidófalvi & Pedersen (2009) have utilised GPS datasets in their study, while Etienne et al. (2010) utilized Automatic Identification Systems (AIS) movement datasets of vessels. The movement dataset obtained from positioning and identification devices are raw datasets. These raw datasets are coded and analyzed. The size and characteristic feature of each dataset adds to the challenge of conducting trajectory analysis.

One major issue is how to optimize the information from movement datasets collected from identification devices. The ultimate aim of this thesis study is to optimize information obtained from identification devices of raw movement data in order to perform trajectory analysis.

Previously studies including Dodge et al., (2009), Buchin et al (2009), Gidófalvi & Pedersen (2009) and Etienne et al. (2010) have not focused on deducing movement characteristics of different types of a group of common entities. These studies in conducted in an urban context (Dodge et al., 2009), in the transportation context (Buchin et al, 2009 and Gidófalvi & Pedersen, 2009) and finally in a maritime context (Etienne et al., 2010). Although studies were conducted at various contexts, these studies not considered for optimization of information of raw data. The intention of the study is to understand the characteristics of different types of vessels using AIS movement data. This thesis is an attempt to explore the trajectories using a tool.

This thesis is a two part study. Firstly, to optimize information obtained from a dataset containing millions of items and secondly, to conduct trajectory analysis from a large dataset.

1.2 Case Study

To illustrate the concept of optimizing information from the raw movement data and to perform trajectory analysis on movement data, real raw movement datasets were used. The intention of this study is to investigate real movement datasets of maritime domains provided by **Kordia New Zealand** recorded between March 2011 and May 2011 from different vessels in New Zealand waters. These raw datasets were collected from Automatic Identification Systems (AIS) terminals. A large dataset (2GB - 30 Million rows) of spatio-temporal movement observations (in total 28 GB) has been utilized in this investigational study. The focus of the study is to make use of AIS raw movement dataset to conduct an explorative study on trajectory analysis. This study intendeds to identify the trajectories from different vessel types and then characterize those identified trajectories to a certain pattern.

1.3 Extracting implicit knowledge from movement parameters

In many applications, studying movement characteristics of objects in terms of essential movement parameters such as speed, rate of turn and direction provides more relevant insight than simply the movement paths. The trajectory, movement path is a geometrical abstraction over time and with movement parameters this conveys the physical notion. Hence, such parameters can hint with more insight on moving objects in a particular context.

One example, conveyance of vehicles that move along the same road, the movement path is normally similar for all vehicles; however, the movement parameter varies in space and time for each individual vehicle, due to topography of the road and other intrinsic factors such as traffic jams, Dodge et al (2009). Similar observations have been noticed from movements of vessels along the same itinerary (Etienne et al., 2010).

Although these studies have discussed movement parameters and their influencing factor in deducing behaviour of moving objects, they haven't focused on the characteristics of moving objects.

Commonly available techniques for knowledge discovery from movement datasets mostly rely on positional information of moving objects through time. In terms of trajectory representation, very little attention has been paid to other movement parameters. Recently a new generation of positioning, identifying and tracking technologies have emerged with advancement of in-vehicle sensors such as gyroscopes, accelerometers so that it is possible to capture a variety of movement parameters in addition to position information. These new sources of information make it possible to explore the movement data set in different dimensions. Hence, the ability of a new analysis technique that is capable of exploiting such information becomes a significant contributor to knowledge discovery from movement datasets.

1.4 Thesis Rationale

To perform trajectory analysis requires the extraction of knowledge about these moving objects from massive datasets (Tsou, 2010; Spaccapietra et al, 2008, Willems, Hage, Vries, & Malaisé, 2010). The major challenge involved during the analysis process of trajectory data is to understand and extract meaningful patterns out of millions of spatio-temporal data collected. Movement parameters enrich trajectories of moving objects with additional information about the characteristics of the movement. One objective of the study is to gain insight on the characteristics of moving objects using a speed parameter.

This thesis is a detailed description of the overall outline for knowledge discovery under this research study. The work focuses on two areas and their integration. On one side, **optimization of information** to extract hidden knowledge from raw datasets by designing an artefact (decoder) to decode into meaningful AIS messages and on the other side, with decoded AIS messages to conduct trajectory analysis is to identify the characteristics of various boats in New Zealand waterways using ArcGIS tool. The ultimate aim of this study is to be able to identify vessel types from trajectory information examined from AIS dataset.

1.5 Research Questions and Research Objective

There are two main objectives involved in this thesis study. The first aim is to optimize information by extracting hidden knowledge from raw movement datasets. The second aim is to conduct trajectory analysis with decoded movement datasets in order to identify the characteristics of different type of vessels. The analysis will include identifying vessel types from trajectory and characterising movement patterns of vessels. The third aim is to identify whether such moving objects do have specific patterns? To perform trajectory analysis and to answer the following research questions which are formulated to be investigated in this study:

RQ1: How can information be optimized from recorded AIS raw datasets (movement data) to perform trajectory analysis.

RQ1.1 What approach would be more appropriate to extract hidden knowledge from the AIS raw dataset?

RQ2: How should raw data be interpreted as a trajectory to perform trajectory analysis using the spatio-temporal approach.

RQ 2.1: How can spatio-temporal data be supported and managed in analysing the patterns of a moving object (different type of vessels) which characterise the movement pattern of vessels.

RQ 2.2: What are the constituting parameters of movement data essential for defining trajectory patterns?

In order to undertake this research, a literature review was first considered to understand the Automatic Identification System (AIS) and its application in maritime scenarios, format of AIS data, methods and approach to extract implicit knowledge from AIS datasets. A latter consideration was to investigate how this extracted information could be used to perform trajectory analysis. In order to perform trajectory analysis ArcGIS software was chosen. The knowledge discovered using AIS raw-data will help in analysing the type of vessel and the pattern movement of those vessels. A decoder was designed to decode AIS raw data of NMEA 0183 format into meaningful datasets. The process of knowledge discovery from raw data to meaningful spatio-temporal information is then presented and discussed in detail. Trajectory analysis was conducted using ArcGIS tool to address the research questions. Finally, the contributions and implications of the study in New Zealand waters and the challenges in handling real-time AIS raw data are discussed. The limitations of the current research and future research are addressed in the conclusion.

1.6 Research Contribution

The main contribution of this thesis was to discover the implicit spatio-temporal information in raw data and to optimize that information in order to perform trajectory analysis. A spatio-temporal approach was chosen to conduct trajectory analysis. In order to extract the implied spatio-temporal information from the raw data, a decoder was designed. The analysis is conducted using real-time raw datasets recorded and provided by Kordia New Zealand. The decoder developed in this research study identified the limitations of commercial off the shelf (COTS) software and freely available online tools for decoding a large amount of data.

1.7 Thesis Structure

This chapter examines the motivation for the current research study and outlines an overview of the research questions. The following chapters are structured to focus and address the research questions. Analysis was carried out accordingly in order to address this research study's objective (Chapter 5). The results are discussed in Chapter 6 followed by a list of limitations of this study and future recommendations are given in Chapter 7.

The entire process of research design is described in the following diagram (Figure 1.1).

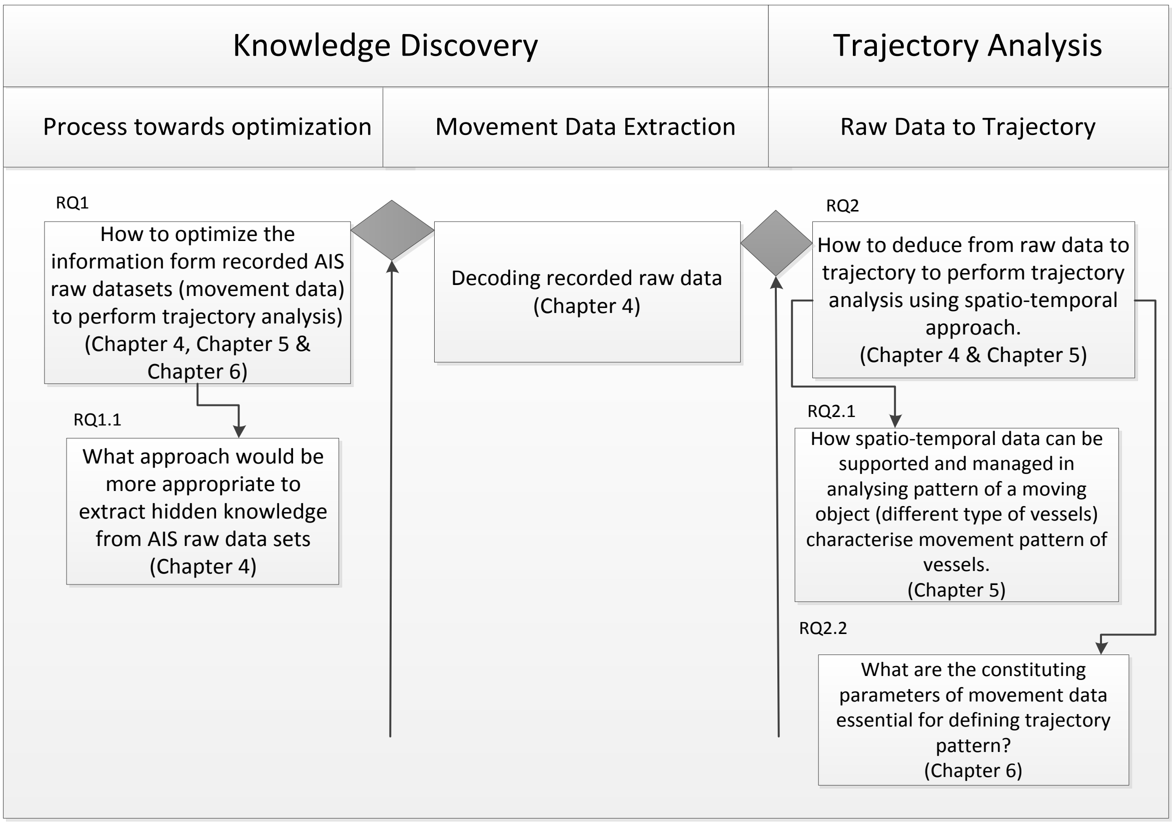


Figure 1.1: Research process for this thesis study

Chapter 2 discusses the literature on Automatic Identification System (AIS) and its applications, purpose of AIS in maritime navigation, how AIS works and how AIS datasets are beneficial in maritime activities, raw data recorded and collection process. As raw data plays a critical role in this research study, a brief discussion is conducted to study the format of AIS raw data implied spatio-temporal with speed parameter knowledge extracted from raw data.

A later part of this chapter will include a detail discussion on knowledge discovery in databases and on the choosing of the most appropriate model applied in this research study.

Chapter 3 In this chapter the research methodology is discussed as well as the rationale for selecting the Design Science approach. The introduction to research methodology and rationale to opt for Design Science has been discussed. This chapter explains how selected Design Science methodology is applied in the current research study. As Design Science methodology is a problem solving methodology, the guidelines used to help in conducting research are explained in detail. This research study is comprised of two phases: Knowledge Discovery or Data Extraction and Trajectory Analysis. The Knowledge Discovery phase explains the research process in data extraction and in the latter phase of trajectory analysis the various techniques applied are discussed in detail.

Chapter 4 explains in detail the decoder designed to decode the raw AIS data which are in NMEA 0183 format. The decoder is capable of importing a text file of raw data and exporting to decoded data (spatio-temporal data). Challenges in handling huge data sets and designing the decoder are discussed in detail. The decoded data set will be comprised of key attributes that help in performing the trajectory analysis. This chapter represents Phase 1 of this thesis study and lays the foundation to perform the next task of Trajectory Analysis (Phase 2).

Chapter 5 discusses the results of trajectory analysis conducted using the set of sample dataset. The following plan explains how the empirical study will be conducted and the results will be presented. The empirical study is a detailed trajectory analysis conducted using ArcGIS tool. The spatio-temporal analysis will be conducted to identify vessel types of trajectories and characterise movement patterns of vessels. An evaluation will be performed based on various criteria such as a similar type of vessel, pattern of vessel type or time versus space movements in a particular trajectory. The analysis results are intended to identify vessel types from trajectories and from the characterised movement pattern of vessels.

Chapter 6 presents the detail interpretation of the results from Chapter 5. Findings of the study are explained Section 6.3. This chapter also discusses the acquired knowledge using the real dataset of New Zealand maritime domain in relation to answering the research questions posed in Section 1.5. The issues incurred during this research study and challenges overtaken are also discussed.

Chapter 7 In this chapter the limitations of the current thesis are discussed. In addition recommendations are made as to the best practice methods in relation to handling real-time large dataset for research and industry practice. Then follows a description of the potential future work that can be considered in relation to detailed trajectory analysis. Recommendations aimed at enhancing the designed decoder in the future are presented. The thesis concludes with a brief summary of the whole research study.

2 Literature Review

This chapter reviews the available literature on trajectory analysis conducted in and out of New Zealand (Section 2.1 and Section 2.2). The next part of the chapter focuses on describing basic concepts of movement data and trajectory analysis. A detailed discussion was then carried out to understand the modelling approaches available from the last decade (Section 2.5). Characteristics of data acquisition systems and issues around these were studied (Section 2.7). A detailed discussion was carried out to understand the process involved in knowledge discovery (Section 2.8). Most importantly a review was conducted that focused on handling large datasets and how to employ them in an analytical study methodologically is further explained in Chapter 3. Classification of the AIS information were discussed (Section 2.9) from maritime industry manual. This facilitated as a gateway to understand the AIS information and design a decoder. The intension of designing a decoder is to extract the spatio-temporal information and then conduct the trajectory analysis. This will serve as a milestone towards answering the research question.

2.1 Introduction

Modern positioning and identification technologies enable the collection of large amounts of data from moving objects. Data about movements of moving objects are often collected as 'trajectories' in space and time. The movements are recorded as a series of geographic locations with respective time stamps. Exploring and analysing large datasets of movement data has become a vital part of research in many disciplines. There is increased interest in many disciplines in the study of trajectory analysis, such as in the behavioural sciences (Horne, Garton, Krone, & Lewis, 2007), ecology (Kubach, Scott & Bulak, 2011), transportation modelling (Wang & Zimmermann, 2010), cartographic planning (Ping, Xinming & Shengxiao, 2008), urban transportation (Wolfson & Xu, 2010), marine traffic (Jiacaia, Qingshan, Jinxing, & Zheping, 2012) and ship movements (Willems, Wetering, & Wijk, 2011). To perform trajectory analysis requires the extraction of knowledge about these moving objects from massive datasets (Tsou, 2010; Willems, Hage, Vries, H.M., & Malaisé, 2010).

According to Spaccapietra, et al., (2008) analysis of trajectory data aims to understand management of complex phenomena that involve moving objects. Dodge, Weibel, & Forootan (2009) focused on trajectory analysis using GPS datasets to classify trajectories based on similar patterns. Dodge et al., (2009) used the datasets that were collected from different types of moving objects such as cars, motorcycles, bicycles and pedestrians in an urban city. Buchin et al (2009) and Gidófalvi & Pedersen (2009) studied the detection of movement patterns from the trajectories of moving objects. Etienne et al. (2010) analysed similar trajectories of moving objects and deduced their behaviour from massive AIS datasets. Although transmitted AIS data collected from ships and base stations are in abundance, the methods for analysis and representation far more scarce (Dodge et al., 2009). Gudmundsson et al (2007) and Andrienko and Andrienko (2007) have commented on the lack of analytical concepts for increasing the amount of movement data collected from GPS tracking entities (persons, vehicles, or animals). Andrienko and Andrienko (2007) stressed that there is a need to find adequate methods for analysing movement data and extracting relevant information.

Extracting implicit and relevant spatio-temporal information from huge raw datasets and performing analysis with adequate methods was the motivating factor of this research. This thesis outlines discovering knowledge from the raw dataset and is followed by trajectory analysis performed using decoded raw data. The work focuses on two areas and their integration. Firstly, optimization of information to discover hidden knowledge from raw datasets by designing an artefact (AIS decoder) to decode the data into spatio-temporal AIS messages and secondly, with decoded AIS messages, to conduct trajectory analysis in order to identify the characteristics of moving pleasure boats, ferries and fishing boats in New Zealand waterways using ArcGIS. The ultimate aim of this study is to be able to identify vessel types from trajectory patterns from AIS data.

2.2 Previous Studies on Trajectory Analysis

The purpose of this section is to review recent research to gain insight on the application of domains, data types used and techniques. According to Lin & Su (2008) finding similar trajectories from moving objects are an interesting issue related to databases of moving objects. Some studies interested to identify pattern detection (Gidófalvi & Pedersen, 2009; Buchin et al, 2009). Most commonly used methods for identifying trajectories are based on geometric similarity. It is also interesting to study the functionality of identifying characteristics of different moving objects of same domain. It is important to take a close look at previously conducted studies to obtain knowledge on application of methods in conducting trajectory analysis.

Table 2.1 : Overview of previously conducted research studies on trajectory analysis

Research Studies	Description	Techniques/Model/Methods	Domain/Datatype
Brillinger, Preisler, Ager, and Kie (2004)	An exploratory data analysis of the trajectories describing movement behaviour of free-ranging animals	Developed a stochastic differential equation-based model	Free-ranging animals in particular deer and elk (Telemetry)
Laube and Purves (2006)	To evaluate extracted movement patterns.	Modeling relative movement patterns in order to search for seasonal migration of home coming pigeons.	By simulation on synthetic data
Giannotti and Pedreschi (2007)	Provides an overview of the history of analysing moving objects. Associated with time geography presented knowledge discovery of moving objects.	Spatio-temporal data mining techniques	GPS Datasets were used to analyse the scenario collected from 17,241 cars in Milan (Italy). The recorded datasets were taken between (duration of one week) Sunday to Saturday. The subset has been selected to consist of 6187 trajectories made on Wednesday 04/04/2007 between 6:00AM and 10:00AM.

Hornsby and Cole (2007)	Detecting movement patterns by analysis of different events.	Modeling the movement of objects from an event-based perspective	Vessel transit from offshore to the ferry landing
Dodge et al., (2009)	Study focused on trajectory analysis using GPS datasets to classify trajectories based on similar patterns.	Spatio-temporal approach and data mining technique.	Datasets used that were collected from different types of moving objects such as cars, motorcycles, bicycles and pedestrians in an urban city.
Etienne et al. (2010)	Detecting unusual behaviour of a moving vessel and studying vessels having similar trajectory patterns.	Spatio-temporal approach and data mining technique.	Unusual behaviours studied such as being ahead of schedule, delayed or veering to the left or to the right of the main route. Data sets used ships' positions recorded during 2 years around the Brest area, in France.
Demsar & Virrantaus (2010)	2D kernel density around 2D point data into 3D density around 3D polyline data	Spatial analysis of using space-time density novel type of spatio-temporal visualisation	Gulf of Finland
K. D. de Vries(2012)	Designed Simple Event Model	Model designed to study events. Maritime – vessel navigation data.	Aggregation of web data (marine traffic.com)

From the above tabulation (Table 2.1) it reveals that research study conducted by Brillinger, Preisler, Ager, and Kie (2004) and Etienne et al. (2010) in detecting movement behaviour demonstrates the application of the spatio-temporal approach. The detailed study of Giannotti and Pedreschi (2007) emphasises the process of knowledge discovery to analyse the history of moving objects. It is also illustrated that techniques of modelling (Vries K. D. de, 2012) and visualization approaches (Demsar & Virrantaus, 2010) over spatio-temporal analysis is predominant in these studies. These studies postulate that the spatio-temporal approach is affirmed in the trajectory analysis of any kind of moving object. It also highlights that there is a lack of study on the functionality of identifying characteristics of different moving objects of same domain.

2.3 Similar Studies conducted in NZ

This section explores similar studies previously conducted in New Zealand. Auckland University has conducted a research study on homing pigeons using Global Positioning Devices datasets (Laube, Dennis, Forer, & Walker, 2007). Research conducted on vegetation communities examined ground-based permanent photo-point images. This was a joint study between University of Otago and CSIR–NRE, Earth Observation Research Group, South Africa (Michel, Mathieu, & Mark, 2009). The former is a spatio-temporal analysis approach on trajectories whereas the latter is a spatial analysis (Object Oriented approach) using large-scale aerial photographs or satellite images. From the review clearly shows that there is a lack of academic study on trajectory analysis in a New Zealand context.

In order to undertake this investigational study, it is also important to conduct a literature review on modelling techniques, knowledge discovery in databases, data collection, and unfolding the characteristic of collected data.

2.4 Definitions and Basic Concepts of Trajectory

Data about the movements of objects are represented as trajectories in space and time. In other words, the movement of an object is recorded as a series of geographic locations with respect to time. Various definitions were given on the trajectory of a moving object. A trajectory represents a moving object's travelled path along with time instants at every position along the path (Vazirgiannis & Wolfson, 2001). Macedo, et al., (2008) described the movement of an object with respect to time and physical position as a trajectory.

A moving object is an object whose position and/or extent changed over time (Erwig, Güting, Schneider, & Vazirgiannis, 1999). The movement of the object was mapped by using time and space, indicating location at different points in time. The sequence of time-stamped locations visited by a moving object formed that object's trajectory, where the trajectory represented the path taken by an object together with the time instants at which the object was at any position along the path (Vazirgiannis & Wolfson, 2001).

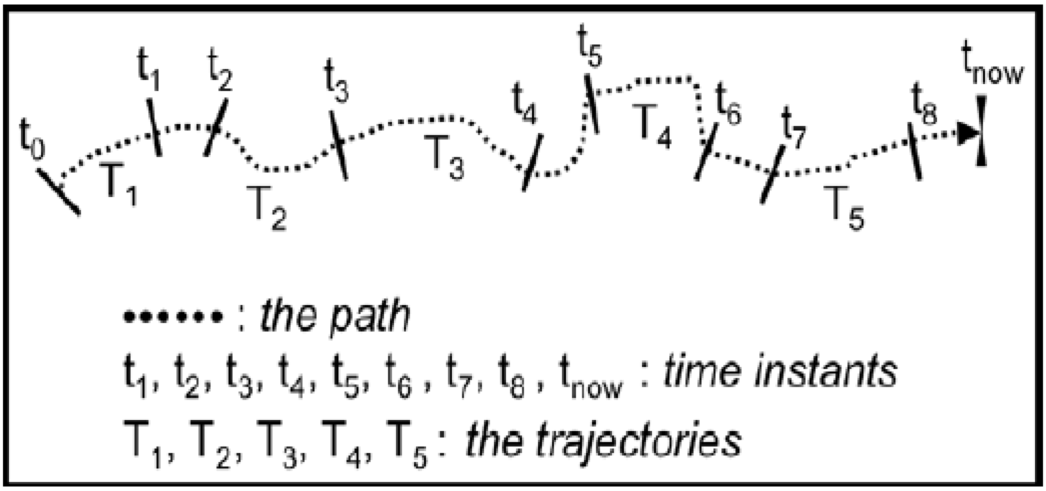


Figure 2.1: A Spatio-temporal path for a moving object and its trajectories with time instants (Source: Spaccapietra, et al.,2008)**.**

The paths of moving objects are represented as trajectories. These trajectories are comprised of a sequence of positions in the two-dimensional geographic environment of location with time stamps, ie. T= $\{(x_1, y_1, t_1), (x_2,y_2,t_n),.......,(x_n, y_n, t_n)\}$ for some n, such that (x_1,y_1) measures the geographic location of a moving object at time t_1. Figure 2.1 illustrates the trajectory of a moving object with represent to spatial coordinates *(x, y)* and *t* represents time.

2.4.1 Transformations of movement data: From raw data to trajectory

For modelling concepts to represent trajectories, basic concepts of trajectory data need to be presented conceptually. This is the objective of this section, where a formal definition of a trajectory and Interpolation are discussed. Then, through description of sample points to Interpolation, the process of building a trajectory from the set of positional (spatio-temporal) data of the moving ships/ferries/vessels will be discussed. Finally, different trajectory approaches will be discussed to choose the most appropriate approach to apply in this research study.

2.4.2 What is a trajectory?

Advancement in technology and wide use of wireless communication devices has induced the growth of research on spatio-temporal databases. In particular, spatio-temporal databases that deal with moving objects involved in various different application areas and often tailored to a particular specific type of moving object. The movement of a moving object is called a **Trajectory.**

Taking into account all the above discussed features, an intuitive definition of a trajectory of any moving object can be formally described as:

$$I \subseteq R \to R^2 : t \mapsto (t) = (\alpha_x(t), \alpha_y(t)) \text{ ----------------------- } 2.4.1$$

where trajectory T is the graph of a continuous mapping from $I \subseteq R$ to R^2 (in a two dimensional plane)

$$T = \{(\alpha_x(t), \alpha_y(t), t) \mid t \in I\} \subset \mathrm{R}^2 \times \mathrm{R} \text{ ---------------------------- } 2.4.2$$

2.4.3 From sample points to trajectories

The foremost restriction using sample data to perform any trajectory computation should contain the sample points. Therefore, all points (x_i, y_i, t_i) in the sample points should have $(x_i, y_i, t_i) = (\alpha_x(t_i), \alpha_y(t_i), t_i)$.

This becomes rather trivial so that sample points are ordered in time, ie if *(i<j)* then

$(t_i < t_j)$. This order will be preserved along the trajectory. Secondly, for given sample data, there are an infinite number of trajectories which are connected. These trajectories are by no means unique. The quest to find a suitable curve connecting the "dots" or the "sample points" is called interpolation. Interpolation is also a chosen technique to deal with missing values.

Although the movement of an object is continuous, AIS technology allows sample data to be collected at discrete instance of time from the object's position. Interpolating these sample data, the object's movement can be exacted. Linear interpolation is considered the simplest approach. The sample data collected at the discrete instance becomes the end points of a line of segments of polylines. The movement of an object is represented in 3-dimensional space by an entire polyline that consist of spatial *(x, y)* and temporal *(t)* coordinates.

2.4.3.1 Applying the interpolation method to sample points

The purpose of this section is to review Interpolation technique. It is the approach followed in most of the studies to analyse trajectory. This is an attempt to understand and consider applying in this study.

The chosen interpolation method is Linear Interpolation. Linear Interpolation is the fastest and easiest among all other interpolation methods (Macedo, et al., 2008, p. 125). The basic idea of linear interpolation is to connect the sample points with straight lines, though the linearity expressed as equal jumps in time leads to equal jumps in space. For example, fragments or segments between the points (x_i, y_i, t_i) an $(x_{i+1}, y_{i+1}, t_{i+1})$ is formally given by:

$$(x, y, t) = (x_i, y_i, t_i) + \frac{t - t_i}{t_{i+1} + t_i}(x_{i+1} - x_i, y_{i+1} - y_i, t_{i+1} - t_i) \quad \text{----------2.4.3}$$

The following Table 2.2 highlights existing studies on Trajectory Modelling.

Table 2.2 : Overview of existing studies on trajectory models.

	Reference	Interpolation Technique	Analysis approach	Special features	Constraints	Domain	Dataset used
1	Van Kreveld and Luo (2007)	Discrete	Spatio-temporal	Similarity, considered sub-trajectory similarity	Specifying minimum duration, outliers and sensitive to noise		
2	Frentzos et al., (2007)	Discrete with linear interpolation	Spatio-temporal		Considered removing noise from dataset and outliers		
3	Pelekis et al. (2007)	Discrete (Linear Interpolation)	Spatio-temporal	Considering speed and direction	Geometric shape and finding intersection points		
4	Trajecevski et al. (2007)	Discrete	Spatio-temporal		Same duration, ignored speed and high complexity		
5	Ding et al. (2008) (ωDF)	Discrete	Spatio-temporal		Ignored speed, removed noise from the dataset and outliers		
6	Lee et al. (2008)	Discrete and continuous	Spatio-temporal	dynamic time warping, maximum speed	Specifying maximum duration	Animal movement, hurricane, vessel navigation	GPS & AIS

7	Buchin et al. (2009)	Continuous	Spatio-temporal	Sub-trajectory similarity, variable time-shift and a variable length	Specifying minimum duration, removed noise from the dataset and outliers		
8	Etienne et al. (2010)	Discrete	Spatio-temporal	Applied Spatial shift	Outliers, used filters, removed noised from dataset,	Maritime	AIS
9	Gerben K. D. de Vries(2012)	Simple Event Model	Spatio-temporal	Piecewise linear segmentation, speed	Outliers,	Maritime – vessel navigation data	aggregate web data (marine traffic.com)

From the Table 2.2, it is clear that interpolation technique is necessary to obtain more accurate knowledge on a sample dataset. Further, it is acknowledged that information could be lost or distorted during large dataset handling and sometime noise removed from the dataset could lose their accuracy.

The following Section 2.5 is a discussion on various modelling approaches applied in previous studies. This Section is a detail description on trajectory modelling and choosing the more appropriate approach for the chosen case study.

2.5 Modeling approaches for Trajectories

According to Spaccapietra, et al., (2008) modelling trajectories fall in 3 categories: They are Spatiotemporal data models, Constraint data models and moving objects data models. The first categories do not recommend specific concepts for trajectories, however they can be used to represent trajectories. The last category, moving object data model is an attempt specially developed for the modelling and querying moving objects. It is applied to modelling and querying of trajectories. Silva. P & Santos (2010) have reviewed spatio-temporal databases for moving objects. A detailed literature review was conducted to understand the data models that support moving objects.

To transform raw data to trajectory data, a popular method proposed by (Spaccapietra, et al., 2008) is called "a stop and move" approach. This is a formal definition for modeling the trajectory data. The idea of "stop and move" approach has been widely adopted or expanded depending on applications, for example identifying stop patterns for frequent points in some regions (Baglioni, Macêdo, Renso, Trasarti, & Wachowicz, 2009), regulating the interval in places using stop (Palma, Bogorny, Kuijpers, & Alvares, 2008) and applying more restrictions to stops and movement with reference to geographic perspective (Zheni, Frihida, Ghezala, & Claramunt, 2009).

The aforementioned approach “Stop and Move” approach is ideal for the discussion on scenarios that moving objects that stop frequently or in regular intervals of time, whereas continuously moving objects like vessels, ferries or boats will that be suitable to apply this approach. At times, vessels like **naval research vessels, pilot vessels and fishing vessels** whose speed always varies and frequently stop and start their journey this approach is more appropriate. After a thorough analysis based on various vessel functionality, the approach “Stop and Move” is chosen for case study.

2.5.1 Origins of Spatio-Temporal Databases

Glancing back at the history of spatio-temporal data models, it was apparent how spatio-temporal database evolved.

The Spatial Database was focused on modelling, querying and integrating geometric and topological information in databases, whereas Temporal Database concentrated on modelling, querying and recording the evolution of facts under different notions of time, that is current and past states of the real world (Chomicki, Toman, & Böhlen, 2001).

However, in reality, space and time are rarely independent, hence integrating space and time meant dealing with time- varying geometries which gave birth to the spatio-temporal database.

2.5.2 Various data models of moving objects

2.5.2.1 Time-stamped data models

Three time-stamping approaches had been proposed to incorporate time into relational databases during the period of 1980s and the beginning of 1990s. They were time-stamping individual tuples (Tansel, et al., 1993) and time-stamping individual attributes (Chomicki, et al., 2001).

At the same time, in order to enhance GIS’s spatial data models by inclusion of models of time, various research projects were undertaken. Among them the Discrete model, in which time was measured at certain points or intervals and variation was discontinuous between these points. Hence the structure of points in time were isomorphic to the natural numbers, versus the continuous model in which time structure was isomorphic to real numbers and thus contained no gaps.

A ***tuple time-stamping approach*** had often been used to expand the schema of a relation by explicit temporal attributes that were used to describe the validity period of a whole relation (Erwig & Schneider, 1999). In data models, any single attribute change led to duplicating the whole tuple and, as a result, multiple relations contained information about an object. Hence continuous change of spatial objects could not be properly modelled using this approach (Erwig & Schneider, 1999).

The ***Snapshot Model*** proposed by Armstrong in 1988, had every layer as a set of temporally harmonized units of theme and the model represented states of one layer at different times. One of the main characteristics of this model was that layers at different time were not necessarily different (Armstrong, 1988).

Langran (1988) proposed the ***Space-time Composite Model*** in 1988, which overlapped multiple snapshot layers into one spatio-temporal composite layer. Theoretically, it described the attribute change of spatial objects during a time interval. For updates, the object needed to be reconstructed and so accordingly, geometrical and topological relationships also needed to be updated (Langran & Chrisman, 1988). In this model the major problem was that only either space or time could be variable at a time but not both. As a result, although the model allowed the recording of change, it failed to capture movement.

Beller et al's Model called the ***Temporal May Set*** (TMS) (1991) was a collection of layers representing time instances at which an event had taken place (Beller, et al., 1991). Unlike the Snapshot Model, this model required layers to be different at different times and as a result, the whole set showed subsequent states of one spatial unit during a specified period (Beller, et al., 1991).

In 1994 Rafaat et al., discussed spatial and thematic attribute changes in a GIS and as a result proposed a relational method to accessed spatial and temporal topologies (Raafat, Yang, & Gauthler, 1994). Using the tuple-time-stamping approach for various time intervals, the attribute changes of an object were recorded, until it was destroyed or became another entity. States of a single data layer were represented by a set of relations, containing topological attributes and a positional relation (Abraham & Roddick, 1999).

The spatio-temporal object model proposed by Worboys in 1994, defined the spatio-temporal objects as spatio-bi-temporal complexes. In this model the spatial properties were represented by basic complexes and the temporal properties were represented using bi-temporal elements attached to all components of spatial features (Worboys, 1994). Even though this model allowed the capturing of change in attributes in both time and space, it could only represent discrete attribute change.

2.5.2.2 Process-based data models

Whigham (1993) had proposed a dual ordered hierarchical structure where time and events were represented in their own hierarchy and placed on a spatial background reference. Events were placed in a hierarchy with location reference, and then a time hierarchy was linked with the events hierarchy to establish the time frames for the individual events (Whigham, 1993). The links could be used to show casual relationships between events (Abraham & Roddick, 1999).

Event Spatio-temporal Data Model by Peuquet & Duan (Donna J. Peuquet, 1994; D. J. Peuquet & Duan, 1995) differentiated between three representations of spatio-temporal data with respect to specific analytical tasks. In this model, a triad of object-based, location-based and time-based data was defined (Worboys, 1994) . The events were sequentially recorded where each event was associated with a list of all changes that occurred since the last update (D. Peuquet & Wentz, 1994). Although this model had capabilities and efficiency to support both spatial and temporal queries, it suffered from lack of a comprehensive definition.

The ***Object-oriented Data Model*** of Raper & Livingstone (Raper & Livingstone, 1995) had every spatio-temporal phenomenon represented by a set of form, process and material objects. Its attributes were three dimensional locations and one dimensional time. This model, however, had difficulty in dealing with area data and topological relationships (Yuan, 1996.) even though it could handle point spatial objects well.

2.5.2.3 Integrated spatio-temporal data models

2.5.2.3.1 Conceptual data models

a. **Data models are based on time-dependent functions**

   In this model, space and time were integrated into a single data model using time-dependent functions, where each element consisted of a geometric value, a date and a behavioural function (Yeh & Cambray, 1995).

b. **Data models are based on abstract data types**

   Guiting & Schneider in 2000 had proposed an interesting abstract data model for moving objects, in which spatio-temporal phenomena were represented as attribute data types, where appropriate operations were associated with these data types to provide an abstract spatio-temporal type extension to a DBMS data model and query language. The most important operator was a type constructor ***r***, which mapped any given primitive data type ***'a'*** into a function type *r (a)* with semantics *r(a) = time – a*, which allowed to define the data type r(point) that represented the position of a point and data type r(region) that represented the geometry of a region at any time (t). Hence a moving object was represented by a sequence of (x, y, t, $a_1 \ldots a_n$) where x and y referred to the object's positions at time t and $a_1 \ldots a_n$ stood for possible attributes. This approach provided complete support for both discrete and continuous spatio-temporal objects.

   In this model the semantics for data types and operations were defined, then appropriate data structures and algorithms were defined for them independently. To support time-dependent geometries, a mechanism called **lifting** was used, in which all operations already defined over non-temporal types were uniformly and consistently made applicable to corresponding temporal types. Finally special operations were offered for temporal types r(a) with functional values. The advantage of this model was that the object representation was closer to reality than others.

c. Data models based on unified modelling language

Price et al., in 1999 developed a spatio-temporal extension to UML by adding a set of constructs that could be applied to classes, attributes and associations. This extension was based on spatio-temporal symbols and a certain section was used to describe the semantics of spatio-temporal properties (Price, Tryfona, & Jensen, 2000). It covered a wide range of spatio-temporal semantics with well-defined syntaxes but the issue of constraints that could be applied on associations or classes were not discussed (Friis-Christensen & Jensen, 2001).

In 2000, Brodeur et al developed a geospatial repository compatible with ISO standards for geographic information using an UML based visual model (Brodeur, Bedard, & Proulx, 2000). The spatio-temporal characteristics were defined at class and attribute levels but lacked the formal definition of constraints and spatio-temporal operators (Friis-Christensen et al, 2001).

d. Data models based on modelling unit

Tryfona and Jensen in 1999 developed the spatio-temporal extension to the ER model using abstractions and modelling units. They defined a set of conceptual structures called STER Spatio Temporal ER model (Tryfona & Jensen, 1999) in which time-stamped (spatial) objects as well as time-stamped tuples which is associated with temporal and spatial icons to objects, attributes and relationships. The advantage of this model was that it offered explicit support for describing and modelling time-varying relationship between associated spatial objects (Price, et al., 2000).

e. **Data models for multidimensional data**

In 2004 Jensen et al., proposed a multidimensional data model for location-based services. The assumption here was that moving objects were active agents that were able to disclose their positional information to location-based services. As a result a number of requirements for a multidimensional data model for location-based services were identified which included explicit and multiple hierarchies in dimensions, partial containment, non-normalized hierarchies, different level of granularity, many-to-many relationships between facts and dimensions and handling impression of data from different dimensions. The association of the proposed data model with algebra provided a basis for formally defining a query language (Jensen, Kligys, Pedersen, & Timko, 2004).

f. **Data models for multimedia data**

For multimedia applications, multimedia data model defined moving spatio-temporal relationships where the relationship between two moving objects was defined by three components – the topological relationship between the two objects, the directional relation between two objects and the time interval that these two relationships were holding (Ozsu & Li, 1998).

2.5.2.3.2 Computational data models

In 2003 Speicys et al developed a computational data model where data structures could be used as a basis for data storage and query processing as well as to model road networks, moving and stationary objects. It included a two-dimensional approach as well as a graphic representation of road networks and static and moving objects (Speičvcys, Jensen, & Kligys, 2003).

2.5.2.3.3 Logical and physical data models

In 1999 Wolfson and his colleagues introduced a system architecture that consisted of three levels. The first level was an object-relational Database Management Systems, which stored moving object data in the form of a sequence of time-stamped positions. The second level was a GIS that was responsible for storing, querying and manipulating spatial objects. The third layer, Databases fOr **MovINg Objects** tracking (DOMINO), contained temporal attribute that predicated and offered support for inherent uncertainty of moving object data (Wolfson, Sistla, Xu, et al., 1999). Later this model was modified to support uncertainty of moving object data and the deviation between a moving object's actual location (mobile) and its location as stored in the database (stationary). If the object was stationary, its location attribute was a (x, y) coordinate pair. If the object was mobile, its location attribute had six sub-attributes – the pointer to a line object represented by the network segment on which an object was moving, the location and time at which the object started its movement, the direction in which the object travelled, the speed at which the object travelled, and finally, an uncertainty measure, which could have been either a constant or a function of time, representing the threshold of the location deviation. Further contribution led to a probabilistic model and an algorithm for query processing where the location of a moving object was a random variable. The density function of this variable was determined using object location at any point in time and uncertainty derived from the database (Wolfson, Sistla, Chamberlain, & Yesha, 1999).

2.5.2.3.4 Moving Object Database

Researchers have been studying in Moving Objects Databases (MOD) have represented issues of trajectories aiming to keep track of object locations. It is necessary to manage time dependent locations of moving objects such as mobile phone users, cars, ships, animals. Therefore the **moving point** is the basic abstraction, if interested in extent (for example, spread of vegetation, oil spill, group of people), then it is a discussion about **moving regions.**

Wolfson et al., (1999) had proposed a data model called MOST that allows storing the motion vector rather than the objects' position in the database. This approach avoided a very high volume of updates on database.

In 2003 Chen et al developed a prototype called CAMEL (Continuous, Active Monitor Engine for Location-based services) for moving object database by offering a high performance location management engine (Chen, Rao, Yu, & Liu, 2003).

Table 2.3: Evolution Summary of Spatio-temporal models

	Approach	Model	Explanation/Description
Spatio Temporal	Entity-Relationship (ER)	ST USM (Khatri, Ram, & Snodgrass,	An annotation-based approach focusing on nontemporal and nongeospatial aspects (i.e., "what") and followed by augmenting the conceptual schema with geospatiotemporal

		2004)	annotations (i.e., "when" and "where").
		STER (Tryfona & Jensen, 1999)	STER - given in terms of the underlying ER model. The ER model is used as the concrete context for presenting the constructs. The semantics of the resulting spatiotemporal ER model, STER, is given in terms of the underlying ER model. STER is accompanied by a textual counterpart, and a CASE tool based on STER is being implemented, using the textual counterpart as its internal representation.
	Object-oriented	Perceptory (Badard, Larrivee, Proulx, & Nadeau, 2004; Chen, 2011; Larrivée, Bédard, & Pouliot, 2005)	defines spatial, temporal plug-ins for visual languages (PVL) to included in any existing database design tool.
		Extended spatio-temporal UML (Price, Tryfona, & Jensen, 2000; Ram, Snodgrass, Khatri, & Hwang, 2001)	A conceptual modeling language using object-oriented Unified Modeling Language (UML). UML is extended to capture semantics of spatio-temporal data with simplicity and clarity. The fundamental modelling constructs are spatial, temporal, and thematic.
		OMT-G (Borges, Davis, & Laender, 2001)	Provides primitives for modeling the geometry and the topology of spatial data, object multiple views and spatial relationships. Also, includes tools to specify transformation processes. This adds advantage to OMT-G compared with other existing models.

		STOQL (Huang & Claramunt, 2002)	STOQL adding spatial and object dimensions to OQL. This is built on extending the existing data model, OMDG.
		Spatio-temporal ODMG (Camossi, Bertolotto, Bertino, & Guerrini, 2003)	A multigranular spatio-temporal data model, extended from ODMG model with multiple spatial and temporal granularities. This provides a framework for mapping the movement of an entity.
		Tripod (Griffiths, Fernandes, Paton, & Barr, 2004)	This object model utilizes a method called **'history'** to maintain knowledge about entities that change over time.
		MADS (Parent, Spaccapietra, & Zimányi, 2006)	An object + relationship spatiotemporal conceptual data model. Objects characterized by properties (attributes and methods). Spatiality and temporality – associated at the different structural levels: object, attribute and relationship.
	Geographic information/ geomatics ISO TC 211	ISO 19107 - Spatial Schema (ISO/TC211, 2003)	Defining a set of spatial data types(vector) and operations for geometric /topological spaces.
		ISO 19108 - Temporal Schema (ISO/TC211, 2002)	Defining a set of temporal datatypes as well functions to describe spatial events associated with time.
	Constraints Logic-based approach (2D representation	The Linear Constraint Model (Grumbach, Su, & Tollu, 1995)	Polynomials used to model data restricted to being linear.

	using mathematical terminology)	The Differential Geometry Model (Grumbach, Rigaux, & Segoufin, 2001; Perona, Porporato, & Ridolfi, 2000)	Represented model using differential geometry. The order of the derivatives being used to describe direction, velocity and acceleration.
	Moving Object database	MOD (Sistla, Wolfson, Chamberlain, & Dao, 1997) MOST Wolfson et al., (1999)	Research studies focusing on modelling trajectories as spatio-temporal objects (***moving objects***) and their databases called Moving Object Databases (MOD).

The literature review highlights that the research in spatial, temporal or spatio-temporal analysis were actively been carried out at various levels of complexity and for the different requirements of the application. After an extensive research on trajectory model it is evident that conceptual data models apparently provides first-hand information on building a data model.

2.6 Classification of moving object applications

Applications dealing with time-varying data are classified by the following categories and based on the type of change they accommodated.

Category 1: Applications that are concerned with changes of non-spatial characteristics of objects (e.g. land parcels in a cadastral information system).
Category 2: Applications in which the position of objects continuously changes (e.g. cars moving on a road network).

Category 3: Applications with objects that integrate changes in the above case as well as changes in their extent (moving regions). In this case mostly happens in environmental applications (e.g. monitoring water pollution caused by oil spills).

Table 2.4 : Categories of applications dealing with time-varying data.

Application Type	Example	Type of Change	Reference
Applications that deal with changes of non-spatial characteristics of objects.	Land parcels in a cadastral information system	Discrete	(Çag˘das & Stubkjær, 2011)
Applications that deal with objects that continuously changes	Cars moving in a road network.	Continuous	(Giannotti, Nanni, Pedreschi, & Pinelli, 2009)
Applications that deal with objects that continuously change as well as changes in their extent. This type of change mostly happens in the study of environmental applications.	Monitoring sea water pollution caused by ships dumping waste. For example, oil spills in sea water.	Continuous	(Kubach, Scott, & Bulak, 2011)

The first category dealt with discrete phenomena whereas the second and the third handle continuous change. Meratnia (2005) added her argument stating movements or changes of continuously moving objects in terms of travel. "Medium" is a new terminology been introduced to study how object positions are thought to travel. According to Meratnia (2005) medium may or may not impose restrictions on movements. Thus resulting in three scenarios of object movement as mentioned below. The focus was on continuous change, where the 'medium' via which the objects were thought to travel resulted in three scenarios.

- Free movement in 2D or 3D space (e.g. animal migration).
- Restricted movement in 2D or 3D space (e.g. ships along coastlines).
- Restricted movement on 2D or 3D networks (e.g. car movements).

Obviously, there are scenarios where movements occur in combination, eg., movements in shopping malls. Shoppers' movements can be seen as free movements, since shoppers can freely move in space in any desired direction or any manner, whereas their movements are restricted within hallways, pathways, mall walls, shelves and so on. It is also identified that moving objects can be stationary due to special circumstances of movements. Stationary circumstances may perhaps include physical obstacles (broken roads, accidents, traffic jams), temporary constraints. On the movements interfere for a very short time interval (stops at traffic lights) also semi-permanent constraints on the movement or could be of personal decisions (waiting in car to pick-up, stopping at the petrol shop).

The above discussed categories and scenarios are the movements that occur in combination in continuously moving objects in terms of travel. The following section will be a discussion on data acquisition methods of a moving object whose characteristics is to generate a large volume of dataset.

2.7 Data acquisition methods

There are several concerns related to data modelling of data acquisition method. This has direct effect on the data types to be supported by the data model. The other concern is that, it requires the data model to provide a strategic approach to handling imprecise and inaccurate data.

Data acquisition methods for moving objects can be categorized on the following basis:

- **Analog Method:**
  In the analog method, the moving objects themselves take note of whereabouts they are in terms of time and location.
- **Digital Method:**
  In the Digital Method, two methods of monitoring are used to take note of the moving objects.

- **Passive Monitoring:** In this method, objects need to have identification. In passive monitoring, objects are not associated with sensors. As objects are not associated with sensors internally, they have to be monitored externally. For instance, radar used in air traffic control. Another example - photo cameras used in the monitoring and analysis of car traffic.

- **Active Monitoring:** In this method, moving objects have a positing device installed on-board. The positioning device can be used for either purpose mentioned below:

    - To record a sequence of time-stamped positions: The recorded sequence of time-stamped positions or observations can be either synchronized or independent. When all moving objects are observed at the same order instance. Global Positioning Systems (GPS) and cellular phones are examples of this category. Obviously, accuracy of the obtained data depends on the technology that is used.

    - To record speed, acceleration and/or direction at any particular time. Function-based data models are the most commonly used technique. The changes are recorded instantaneously with any change that can be detected. Later on, the recorded values are used to find the location of the object.

Data acquired through different methods are voluminous dataset. The process of extracting required information from huge voluminous data is called “Knowledge Discovery” and the analytical reasoning process is called “Sensemaking”. The following section describes the literature review conducted on Knowledge Discovery and Sensemaking.

2.8 Knowledge Discovery and Sensemaking

Knowledge Discovery in Databases (KDD) is the broader context in data mining. It is a process of identifying or discovering useful and yet undiscovered implicit knowledge from real-world data. Knowledge Discovery in databases is an iterative process consisting of (1) data preparation and cleaning (2) data extraction (3) analysis and interpretation. This is a complex task comprised of data exploration, analysing and interpretation through a process of knowledge discovery (Riveiroa, Falkmana, Ziemkea, & Kronhamnb, 2009).

To perform the above mentioned tasks a theoretical framework was chosen for understanding the analytical reasoning process called ***Sense-making*** *(Bodnar, 2005; Pirolli & Card, 2005)*. This consists of a series of tasks such as information gathering, representation of the information in a schema, insight through the manipulation of the schema and obtaining knowledge. These tasks were carried out as a series of steps undertaken to perform the analysis.

Klein, Moon & Hoffman (2006) followed the steps of Bodnar (2005) who framed the "Think Loop Model". This is derived from series of cognitive task analysis. The Figure 2.2 represents data flow and process flow. The flow of data shows the transformation from raw information to reportable results. The process involves the following steps:

- External data sources - the raw evidence of information.
- *Shoebox* - represents a subset of the external data that is relevant for processing.
- Evidence file - refers to snippets extracted from the shoebox
- *Schema* - representation of the information to draw conclusions
- Hypothesis - tentative representation of these conclusions with supporting arguments
- Presentation - present other work product used for communication.

Nanni, Trasarti, Renso, Giannotti, & Pedreschi (2010) have proposed an 8 step framework particularly designed to accommodate their GeoPKDD system project to extract knowledge out of massive trajectory datasets (GPS). This framework mainly focused on privacy-aware knowledge discovery, which is not suitable to apply in this research study as this is out of scope in this study. This research study mainly focused on knowledge discovery then follows trajectory analysis.

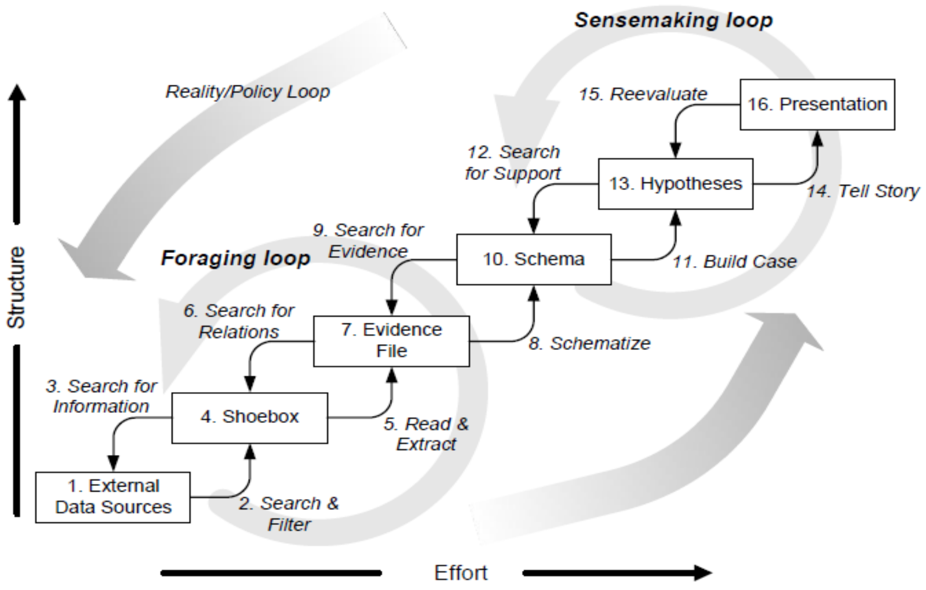


Figure 2.2: "Think Loop Model" combining foraging and sense-making loops

(Source: Riveiroa et al., 2009).

A close analysis of the above mentioned process is adapted to transform raw data into trajectory data according to the thesis requirement. Following these procedures, Klein, Moon & Hoffman (2006) notices that sense-making follow the progression:

data → information → knowledge → understanding

In this research study of trajectory analysis, the suggested progression will be,

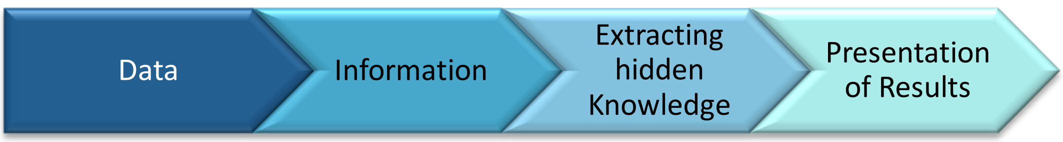


Figure 2.3: Progression in this research study (Adapted from Klein et al., 2006)

Klein et al., (2006) notional model has identified a more appropriate model to perform this research study on trajectory analysis compared with Nanni et al (2010). Indeed, a mixed approach appears to be more suitable as both models are not on their own adequate enough to deal with large raw dataset analytical study.

The following section will give an account of the Automatic Identification Systems (AIS) acquaintance in the maritime industry. The section will also describe how does AIS works and the implicit information of AIS.

2.9 Automatic Identification Systems (AIS) & its application

The use of Automatic Identification systems (AIS) in maritime monitoring started from 2004 (IALA, 2004). The advent of global navigation satellite systems (GNSS), particularly with AIS is extensively used in large ships of over 300 tonnes and passenger vessels. AIS is mainly used for collision avoidance and as a navigation aid, as well as for search and rescue operations.

AIS allows vessels to electronically broadcast ship data at regular intervals. These data include vessel identification, position, course speed and other crucial information related to the vessel. Each vessel is identified with its own unique identification number called Maritime Mobile Service Identity (MMSI). The vessel is installed with AIS which transmits continuously providing comprehensive and detailed data sets for individual vessels. The received data from vessels are in the National Marine Electronics Association (NMEA) format.

These raw data are comprised of critical information that could be applied to a range of vital applications, such as vessel monitoring, maritime surveillance (Feixiang, 2011), security, search and rescue, traffic management, collision avoidance (Mou, Tak, & Ligteringen, 2010). Additional value-added services that supplement the AIS message are also considered in extracting traffic density plot, monitoring vessel location and movement, Area of Interest, historical track analysis or similar trajectory analysis. Such information is essential in assisting with the investigation and identification of oil-spills from suspect vessels (Eide, Endresen, Brett, Ervik, & Røang, 2007), verifying that vessel traffic has complied with navigation rules or with the evaluation of the need for additional navigational aids.

The above mentioned applications provide the opportunity for more refined vessel movement and detailed data sets. This also provides challenges in data management and analysis. The study also takes the opportunity to discuss challenges involved in optimizing information from NMEA 0183 format raw data which is decoded into meaningful AIS messages. The motivation and rationale is the lack of study on trajectory analysis on New Zealand vessels/ships.

2.9.1 How does AIS Works?

Improvements made to navigational standards include:

GPS: In 1994, GPS was declared operational with 18 satellites that transmit signals to a GPS receiver. This obviously enabled a vessels location, velocity and direction to be determined.

ECDIS: Electronic Chart Display and Information System provided a computer based alternative to paper charts. In 1996, IMO introduced ECDIS standards, which means that future SOLAS regulations could possibly approve replacing paper charts with ECDIS.

AIS: An Automatic Identification System became possible with the advent of transponders. In 2004, AIS became a requirement for SOLAS vessels.

Among numerous security regulations that came into effect after September 2004, one of the main requirements was commercial marine vessels to be fitted with Automatic Identification Systems (AIS). AIS provides a means for vessels to transmit data electronically, including vessel identification, position, speed and course with Vessel Traffic Services (VTS). The data is shared with other vessels too. AIS utilizes Global Positioning Systems (GPS) technology in conjunction with shipboard sensors in addition to VHF radio communication equipment. This helps navigation information to be exchanged electronically. Vessel identifiers such as the vessel name and VHF call sign are made available during initial equipment installation program. This information is included along with location information together with the ship's global navigation satellite system receiver and gyrocompass (IALA, 2004). AIS is used by marine vessels in coordination with VTS to monitor vessel location and movement. AIS transmitters normally send data every 2 to 10 seconds depending on moving the vessel's speed while underway and every 3 minutes while the vessel is at anchor. These data include the following attributes:

Table 2.5: Mandatory attributes from AIS transmitter

- Vessel's Maritime Mobile Service Identity (MMSI) - an unique, 9-digit identification number.
- Navigation status - "at anchor", "underway using engine(s)", or "not under command"
- Rate of turn - right or left, 0 to degrees per minute
- Speed over ground - 0 to knots (normally 0.1 knot resolution)
- Position Accuracy
- Longitude and Latitude - 1/10,000 minute;
- Course over ground - relative to north to 0.1 degree
- Time stamp - Coordinated Universal Time (UTC) time accurate to nearest second when the data originated or generated.

In addition, further details are broadcast every 6 minutes, whether the vessel is underway or at anchor.

Table 2.6: Additional attributes from AIS transmitter

- IMO - International Maritime Organization's Ship identification number (IMO) - a 7 digit ship's registration number
- Call Sign-International radio call sign assigned to vessel by its country, upto 7 character.
- Vessel Name - 20 Characters to signify the name of the vessel.
- Ship/Cargo - Type of the vessel
- Vessel Dimension - to nearest meter
- Type of position system - GPS/Differential Global Positioning Sysytems (DGPS)/Long Range Navigation (LORAN) - C.
- Location of positioning system's antenna (Vessel onboard)
- Draught of Ship - 0.1 meter to 25.5 meters;
- Destination - mention to length of 20 characters
- Estimated Time of Arrival (ETA) - ETA at destination mentioned in UTC date hour:minute.

2.9.2 Classification of AIS information

The data is received by an AIS receiver and relayed to a computer-based navigation system. Such received data are encoded in NMEA sentences (64-bit plain text). The information received can be classified into 2 categories namely, Ship information and Voyage information (NAVCEN, 2010).

Table 2. 7: Classification of AIS Information.

Ship information	***Voyage information***
Ship name	Current latitude
Type of vessel	Current longitude
Call sign	Speed over ground
MMSI number	Course over ground
IMO number	Heading
Registration	Rate of turn
Length	Navigational status
Beam	Destination
Draft	Estimated Time of Arrival (ETA)

According to International Maritime Organization (IMO, 2011) ship information can be further divided into Static and Dynamic information,

- Static Information - vessel's name, IMO number, MMSI number, dimensions.
- Dynamic Information - vessel's position, speed, current status, course and rate of turn.
- Voyage - specific information - destination, estimated time of arrival (ETA) and draught.

Data received by AIS unit are encoded in NMEA sentences (64-bit plain text). A sample is shown below:

```
!AIVDM,1,1,,B,1INS<8@P001cnWFEdSmh00bT0000,0*38
```

Figure 2.4: NMEA format (64-bit plain text) data received from AIS unit

The original dataset is provided and includes a combination of the following information, a timestamp, a talkerID and a SentenceID. The snapshot (Figure 2.5) shows the original dataset.

	NMEA_ID	TimeDate	RawSentence	TalkerID	SentenceID
1	1	2011-03-19 00:00:00.550	$PSHI,001,46.47459443,,01448,32972,175956.32,0,1742,-...	P	SHI
2	2	2011-03-19 00:00:00.580	!AIVDM,1,1,,A,19NWqsP2C><Lh8?a7wJe;bL006sd,0*0F	AI	VDM
3	3	2011-03-19 00:00:00.580	$GPZDA,,,,,,*48	GP	ZDA
4	4	2011-03-19 00:00:00.580	!AIVDM,1,1,,A,15RUV`0023<Nmil`3ui00;?n0@S9,0*12	AI	VDM
5	5	2011-03-19 00:00:00.580	$PSHI,001,31.95769142,,05327,31956,065627.63,0,1198,-...	P	SHI
6	6	2011-03-19 00:00:00.580	!AIVDM,1,1,,A,15RUV`0023<Nmil`3ui00;?n0@S9,0*12	AI	VDM
7	7	2011-03-19 00:00:00.580	$PSHI,001,46.50148233,3,65224,31433,175956.35,0,1743,...	P	SHI
8	8	2011-03-19 00:00:00.580	$PSHI,001,00.37873544,,65108,24711,000000.20,0,0014,-...	P	SHI
9	9	2011-03-19 00:00:00.580	!AIVDM,2,1,3,2,E1c2;q@b44ah4ah0h:2ab@70VRh6@:Ukl>...	AI	VDM
10	10	2011-03-19 00:00:00.580	!AIVDM,2,2,3,2,1CQ1A83PCAH0,4*10	AI	VDM
11	11	2011-03-19 00:00:00.580	!AIVDM,1,1,,A,18OocJ501q<IG3MdAN`co9Un0802,0*20	AI	VDM
12	12	2011-03-19 00:00:00.580	$PSHI,001,29.57771903,,64767,31358,214844.80,0,1109,-...	P	SHI
13	13	2011-03-19 00:00:00.597	!AIVDM,1,1,,A,143V0p7014<HtNQWG1DQciL00<00,0*55	AI	VDM
14	14	2011-03-19 00:00:00.597	$GPZDA,,,,,,*48	GP	ZDA
15	15	2011-03-19 00:00:00.597	!AIVDM,1,1,,B,67`B160PH1m@wsp0003000@10H0,2*49	AI	VDM
16	16	2011-03-19 00:00:00.597	$GPZDA,,,,,,*48	GP	ZDA
17	17	2011-03-19 00:00:00.597	!AIVDM,1,1,,B,181580001w<JJ>qalcMPV0Sn00Rr,0*38	AI	VDM
18	18	2011-03-19 00:00:00.597	$PSHI,001,00.03192694,,01778,34173,000000.03,0,0001,-...	P	SHI
19	19	2011-03-19 00:00:00.597	$GPGGA,,,,,,0,00,,,,,,,*66	GP	GGA
20	20	2011-03-19 00:00:00.597	$GPGGA,,,,,,0,00,,,,,,,*66	GP	GGA
21	21	2011-03-19 00:00:00.597	$GPZDA,,,,,,*48	GP	ZDA
22	22	2011-03-19 00:00:00.597	$GPVTG,,,,,,,,,N*30	GP	VTG
23	23	2011-03-19 00:00:00.597	$GPVTG,,,,,,,,,N*30	GP	VTG
24	24	2011-03-19 00:00:00.597	$GPGGA,,,,,,0,00,,,,,,,*66	GP	GGA
25	25	2011-03-19 00:00:00.610	$GPZDA,,,,,,*48	GP	ZDA
26	26	2011-03-19 00:00:00.610	$GPGGA,,,,,,0,00,,,,,,,*66	GP	GGA
27	27	2011-03-19 00:00:00.610	$GPVTG,,,,,,,,,N*30	GP	VTG
28	28	2011-03-19 00:00:00.610	$GPZDA,,,,,,*48	GP	ZDA
29	29	2011-03-19 00:00:00.610	$GPVTG,,,,,,,,,N*30	GP	VTG
30	30	2011-03-19 00:00:00.610	$GPGGA,,,,,,0,00,,,,,,,*66	GP	GGA

Figure 2.5: The original raw dataset provided by Kordia New Zealand.

2.10 Summary

This chapter reviewed the available literature on trajectory analysis conducted in and out of New Zealand. Chapter 2 discussed in detail the Automatic Identification System (AIS) within the maritime domain, its significant usage in maritime situations, a detailed study on AIS data format, its characteristic features, how the data is recorded and collected. The recorded dataset was always a raw data. These raw data are voluminous, noise and erroneous. The basic components of building trajectory data models were discussed.

These raw dataset have to be decoded to extract the hidden knowledge. Various knowledge discovery processes were investigated. The decoder is identified as gateway to obtain the movement data to perform the analysis. Exploring and analysing large datasets of movement data is a central aspect of this research study.

This chapter reviews the available literature on trajectory analysis conducted in and out of New Zealand (Section 2.1 and Section 2.2). The next part of the chapter focuses on describing basic concepts of movement data and trajectory analysis. A detailed discussion was then carried out to understand the modelling approaches available from the last decade (Section 2.5). Characteristics of data acquisition systems and issues around these were studied (Section 2.7). A detailed discussion was carried out to understand the process involved in knowledge discovery (Section 2.8). Most importantly a review was conducted that focused on handling large datasets and how to employ them in an analytical study methodologically is further explained in Chapter 3. Classification of the AIS information were discussed (Section 2.9) from maritime industry manual. This facilitated as a gateway to understand the AIS information and design a decoder. The intension of designing a decoder is to extract the spatio-temporal information and then conduct the trajectory analysis. This will serve as a milestone towards answering the research question.

3 Research Methodology

3.1 Introduction

This chapter presents the research methodology and paradigm used to fulfil the planned research objectives of this study. It offers a brief overview of the research problem statements, the research approach and the chosen methodology applied to undertake this thesis study. The research objective is to study the trajectory movement of various kinds of vessels through optimizing the information collected from an AIS terminal. The main objective of optimizing this information is to analyse and describe the trajectory characteristics of the moving objects using a trajectory analysis model. The spatio-temporal analysis method has been selected to study the trajectories of different vessel types. In order to analyse the trajectory of a moving object and optimize the data it is necessary to extract the knowledge (information) and make useful and valuable information from the large datasets. To undertake this research and answer the research questions, an empirical research approach was used.

3.1.1 Roadmap and research phases

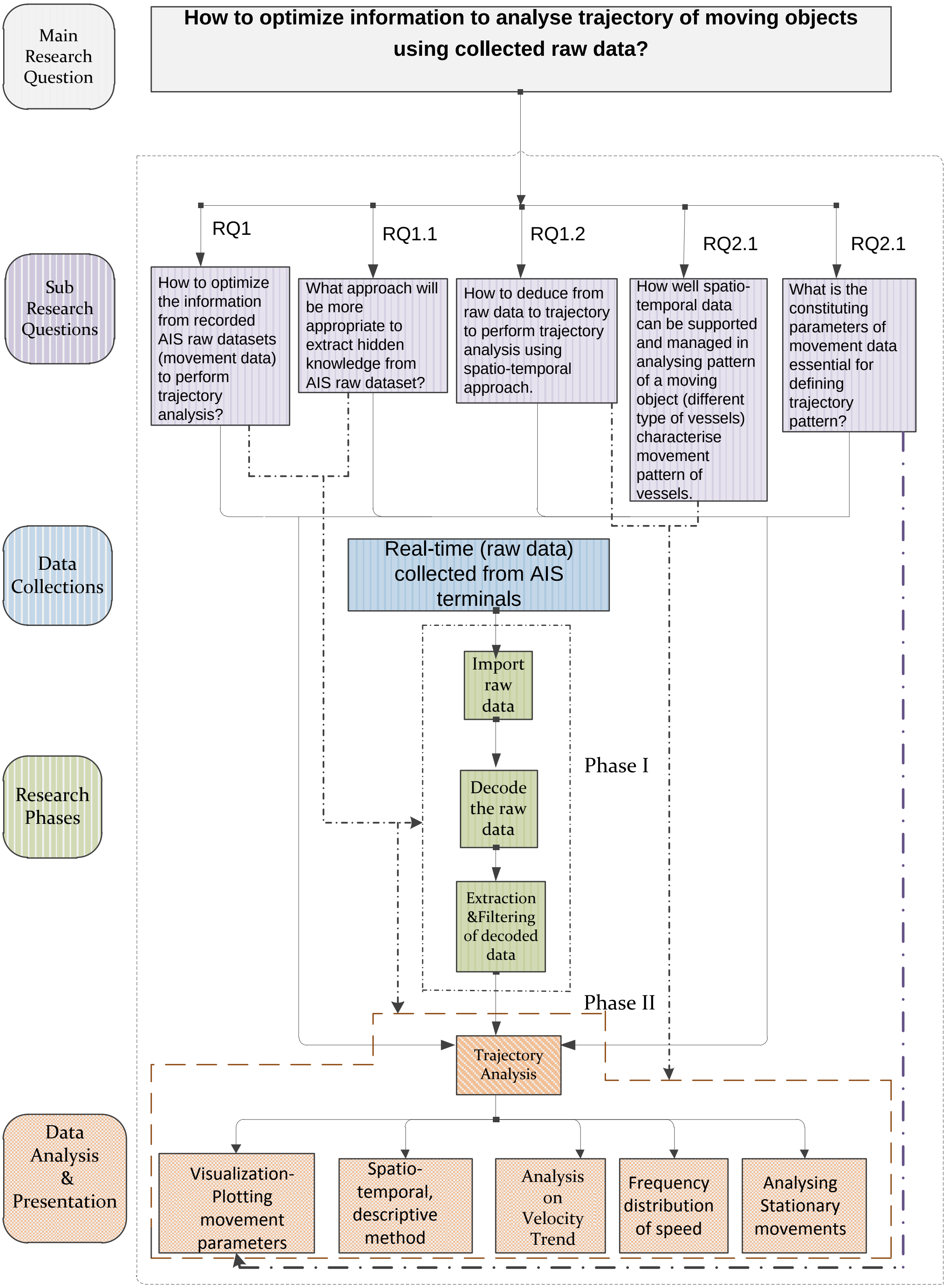


Figure 3. 1 : Roadmap of this research study.

3.2 Design Science As Research Approach

Design Science is a technically orientated field. According to Peffers, Tuunanen, Rothenberger, & Chatterjee (2008) Design Science is a multi-paradigmatic area of research that provides valuable knowledge about the discipline of information systems. The main tasks involved in Design Science are the creation, construction and evaluation of artefacts or models. The evaluation process determines how well the artefact has performed. Any research study in Design Science begins with the awareness of a problem. The idea or proposal could be improved upon at a later stage of the problem based on requirements.

The Design Science framework has been chosen to evaluate the application designed, developed and deployed in a trajectory analysis. Recent studies show (Alturki, Gable, & Bandara, 2011) that Design science (DS) research has become an accepted approach for research in the IS discipline (Iivari, 2007; W. Kuechler & Vaishnavi, 2008). The Rationale and goal of Design Science research is utility (Hevner et al, 2004). As Design Science deals with research through the building and evaluation of artefacts is to identify the application requirements. Hevner et al., (2004) designed a framework for evaluating activities carried in an IS environment. Frameworks are designed to evaluate or justify activities to identify weaknesses in theory or artefact and processes in order to refine the defect. The process of refinement and reassessment is typically discussed in future research scope or directions (Section 7.5).

3.2.1 Guidelines of Design Science Approach

In any Design Science research initiative, a problem or need is first identified and the search for a solution creates the need for an artefact. The research must generate an output which is satisfactory to the interests of the audience and the researcher. At the turn of the millennium, Hevner et al (2004) raised the importance of Design Science within the wider Information Systems community.

Subsequently, other researchers (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2008; Kuechler & Vaishnavi, 2008; Ostrowki et al., 2011) have proposed ideal guidelines for the research requirements that should be involved in Design Science.

A research study can start with the awareness of a problem, followed up by suggested solutions and supported by knowledge from a related field. This can be used to create a proposal and then a tentative design plan. A proposal or design plan could consist of a novel or creative idea, improvements to any original design can then be made in the later development stages.

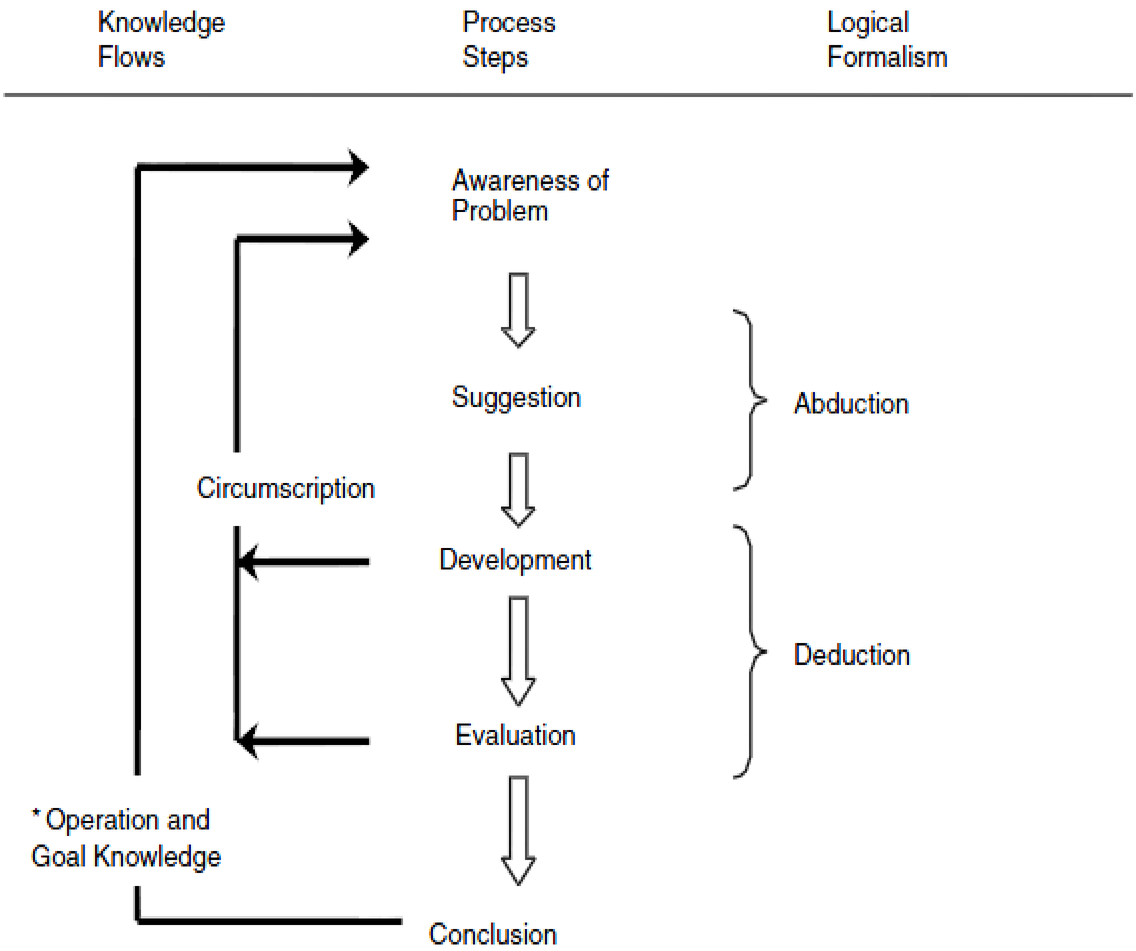


Figure 3. 2: Design Science Research Methodology
Source: (Vaishnavi & Kuechler, 2008, p. 20)

3.2.2 Comparative study of Design Science guidelines

The Design Science methodology is basically considered on three approaches namely, research problem, design of an artefact, evaluation and iterative process. The steps outline by **Vaishnavi & Kuechler (2008)** above in Figure 3.2 and Peffers et al. (2008) coincide with the guidelines provided by Hevner et al. (2004) except for the stage the process called “demonstration”. Demonstration is embedded within the design evaluation of an artefact. Demonstration also provides rigorous research and supports the decision to iterate the process of undertaking research. Hevner et al. (2004) proposed the guideline of a research contribution is a part of the communication process (Peffers et al., 2008) or deriving conclusion (Vaishnavi & Kuechler, 2008) in the other two discussed approaches.

Table 3.1 : Comparison of guidelines of Design Science Approach

#	**Hevner et al. (2004)**	**Vaishnavi & Kuechler (2008)**	**Peffers et al. (2008)**	**Ostrowki et al. (2011)**
1. Artefact Design	Artefact Design	Development of Artefact	Design and development	Identify problem and motivation
2. Problem Relevance	Problem relevance – Important and relevant problems.	Awareness of problem	Problem identification and motivation	Meta-design
3. Evaluation	Evaluation of artefact	Evaluate	Evaluation, demonstration	Ex post artificial evaluation
4. Contribution	Research contributions	Conclusion – contribution part of the conclusion	Part of communication	Design practice
5. Rigor	Researching Rigorously	Suggestions, development	Demonstration	-
6. Process	Design - search process	Suggestions – Suggestions with iterative process	Design method, demonstrated with evaluation and iteration if necessary	Ex post Naturalistic evaluation
7. Communication	Communication	Part of conclusion	Communication	Demonstration

However, Ostrowki et al. (2011) entirely presented a different guideline compared to Hevner et al. (2004), Vaishnavi & Kuechler (2008) and Peffers et al. (2008). Ostrowki et al. (2011) formulated a different set of guidelines beginning with identifying the problem, presenting the meta-design followed by conducting an artificial evaluation. Conducting evaluation is the only step all these four approaches agree on. As the part of evaluation, research rigor (Hevner et al., 2004) while in other approaches is brought about by the suggestion (Vaishnavi & Kuechler, 2008) and demonstration suggested by Peffers et al (2008). In contrast, Ostrowki et al. (2011) does not commented on the research rigor step. Otherwise, all the four studies agree closely on 7 step guidelines of research Design Science methodology.

3.2.3 Evaluation of Artefacts

Evaluation is a crucial part of the research process in Design Science. For IT products controlled experiments can be conducted to assess the overall reliability and performance of the product. This is part of the process required to integrate the product with a business environment. Product evaluation can be both qualitative and quantitative. Quantitative evaluation can be usefully applied when newer artefacts are being compared with older ones or with previous versions. The same applies when new and original artefacts are produced in order to address reoccurring or past problems. Qualitative evaluation can be used when an artefact is designed to solve a new or previously unaddressed problem.

Evaluation is observing and measuring how well the artefact supports a solution to the problem or proposed project. The activity involves comparing the objectives of a solution to actual observed results. This comparison is mainly performed from use of the artefact in the demonstration. It requires knowledge of relevant metrics and analysis techniques (Peffer et al. 2008; Pries-Heje, Baskerville, & Venable, 2008).

When an evaluation has been completed the designer can choose whether to make further improvements to the artefact. Design is partially a search process conducted in order to identify the most suitable solution to a given problem. Overall, researchers in all scientific fields seek to innovate and create knowledge from the research they carry out. The knowledge created as a result of Design Science research also needs to be communicated to audiences through an appropriate output. An article in an academic journal helps to validate the credibility of the completed research. An additional step which also becomes part of the evaluation process is the demonstration of the finished artefact. Demonstration is a vitally important stage when concerned with the release of new I.T products.

3.2.4 Research Rigor & Search Process

Design is a process that is concerned with the invention or creation of new and original items or artefacts. Design Science is a science of the artificial and is concerned with knowledge relating to man-made objects or exacting hidden knowledge. It is this distinction which sets Design Science apart from other fields of science, such as natural science.

Design is a process which involves a series of thoughts and activities by which connection between an artefact (problem) and various possibilities and solutions are studied and implemented (Miller, 2004). Theorists working in a variety of scientific and other fields have defined research in general to be an activity that attempts to understand a particular phenomenon and the way in which a phenomenon functions. A phenomenon can typically be described as a behaviour or set of behaviours that the researcher finds interesting and therefore wishes to investigate. Research allows us to predict the behaviour of these phenomenon(s) or the outcome of a particular process.

As mentioned earlier, Design Science is a multi-paradigmatic area of research that provides valuable knowledge about the discipline of information systems (Peffers et al., 2008). Design science can be defined as knowledge of material culture or artificial world counterbalanced by knowledge of science (technologies) and knowledge of humanities (values, skills, expertise). It can be understood in 2 different ways: Firstly, Design Science as an accumulation of knowledge obtained through an organized and wholly systematic application of experimental (scientific) methods to design type problems which is equated with the positivist paradigm. Secondly, Design Science generated and accumulated through systematic application of both scientific methods and other research approaches such as constructive and critical methods.

Previously researchers from a variety of fields in both natural science and design science have been found to generally agree on the methods required for researching their phenomenon of interest. These groups of researchers can generally be described as paradigmatic research communities. In other newly emerging research communities there can be some overlap and variation in areas of interest and the methods of investigation used. These groups of researchers can be described as multi-paradigmatic or pre-paradigmatic research communities. Information systems are one field of IT that could be seen as a good example of a multi-paradigmatic research community.

3.2.5 The Outputs of Design Science Research

Researchers within the Information Systems research community have often expressed different opinions about what the ideal focus of research outcomes should be. Whilst some have focused on the production of finished artefacts and how well these artefacts function, others have instead opted to focus on the research process itself and the various elements involved in this process.

Design Science attempts to change the state of the world through the introduction of new and novel artefacts. Design Science researchers are therefore normally open to new ideas and alternative views of the world. According to Peffers et al., (2008) Design Science is a discipline oriented towards creation of successful artefacts. These views are in contrast to the positivist view common in natural science, which supports a more fixed view of the world. Hevner & Chatterjee (2010) stressed that newly designed artefacts should provide value to their users instead of adding attractiveness. This has been strongly suggested and stressed by designers of IT-based digital technology products as designed artefacts play a crucial role in problem solving and achieving improvements (Hevner & Chatterjee, 2010).

The final results of any Design Science research are likely to combine sets of both models and methods. Additionally the outcomes of design science research may be concerned with the production of a real life artefact. Such artefacts or inventions are normally produced with a specific use or purpose in mind rather than simply being created for research purposes.

Table 3.2 : Comparative Study - Design and Design Science process elements for Design Science research method (DSRM) in Information Systems.

Common Design Process Elements	Jay F. Nunamaker et al (1991)	Rossi & Sein (2003)	Walls et al. (2004)	Hevner et al. (2004)	Cole et al. (2005)	Peffer et al. (2008)
Problem identification and motivation	Construct – conceptual framework model.	Identifying need – problem motivation	Meta-requirements, develop theories (kernel)	Key and relevant problems	Identify the need	Identify the problem & motivation
Objectives	Develop a system architecture	Build	Requirements	Implicit “relevance”	Identify the objective/aim	Define objective of a solution
Design and Development	Design and build the proposed system	Design and build the system	Design method, meta design	Iterative search process, Design Artifact	Build the system	Design & Demonstration development
Demonstration	Experiment and observe	Conduct experiment	-	-	-	Evaluation

Evaluation	Evaluate the system	Evaluate	Process testable design/product hypothesis testing	Evaluate	Evaluate	-
Communication	-	-	-	Communication	-	Communication

3.3 Design Science research methodology process model

All six papers discussed in Table 3.2 have contributed ideas for process elements. These research studies (Nunamaker et al.,1991; Rossi & Sein., 2003; Walls et al. 2004; Hevner et al., 2004; Cole et al., 2005; Peffer et al., 2008) in Information System (IS) and other disciplines have contributed ideas for process elements. These papers have contributed some component in the initial stages from research to define a research problem. Nunamaker et al. (1991) and Wall et al. (2004) emphasized theoretical bases, whereas researchers from engineering background focused more on applied problems. Cole et al. (2005) suggested the requirements/need for problem enumeration whereas Rossi and Sein (2003) advocated the need for identification. Hevner et al., (2004) affirms that Design Science research should stress important and relevant problems. The importance of Design Science within the wider information systems community has raised after Hevner et al.,(2004) contribution towards Design Science methodology. Following the guidelines based on those representative papers, Peffers et al. (2008) has synthesized the process elements and components of Design Science research methodology (DSRM). The synthesized process model consisting of six activities in sequence (Figure: 3.3).

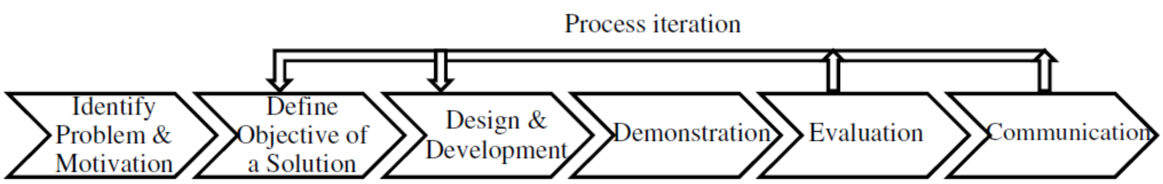


Figure 3.3: Design science research methodology process model

(Source: Peffers et al. 2008)

3.4 Foundations of Design Science in Information Systems (IS)

Information Systems (IS) is a discipline that by its very nature is directly linked and related to other areas of information technology. These different fields of IT can include hardware, software, databases and communication technology. Depending on the perspective of the researchers involved, the IS discipline is concerned with the design of technical systems combined with the need to design technically enabled social systems. Some have argued that in regard to these so called *soft factors* (Montealegre & Keil, 2000), the social systems are an essential part of the design process with regards to information systems whilst others have argued that the design of information systems should be concerned simply with the technical aspect of design.

Information System (IS) is a discipline which involves a complex web of relationships and interactions between technical systems or technologies (data processing such as storing, retrieving, and displaying) and human dimensions (social, cultural, political and organisational factors). The main contributing factor to the Information System (IS) failure is ignoring the soft factors (human dimensions) in the design (Montealegre & Keil, 2000). At the turn of the millennium, Hevner et al's (2004) publication raised the importance of Design Science within the wider information systems community.

3.4.1 Design Science As Research Area in IS

The chief factor in the growth of Design Science as a research area within the field of Information Systems was the need to complement the largely behavioural orientation of Information Systems research with more design orientated research. This change was part of an attempt to increase the relevance of Information Systems overall.

Design Science, alternatively known as Design Theory was first developed in the early nineties, and for most of the remainder of the decade it remained a largely niche field. Some researchers and scholars have expressed doubt over the philosophical repercussions and lack of overall clarity regarding the understandings and endeavours involved with Design Science. Often these arguments have been concerned with the differences between high quality professional design and Design Science research (Gibson and Arnott, 2007).

Early Design Science research efforts were largely grounded in the scientific traditions of positivism. More recent developments in Information Systems have pointed towards further favouring the adaption of Design Science methods, focusing increasingly on the overall technical aspect.

3.4.2 Design As Interdisciplinary study

There are varying definitions of what the real meaning of design actually is. The word design generally refers to complete objects or artefacts that are created intentionally by the human hand. This focus on completed design artefacts has been called the *product view* of design (Marxt and Hacklin, 2005). When describing the overall meaning of design, other theorists have alternatively chosen to focus on the actual design process itself rather than the finished product (Hevner et al., 2004).

Design cannot be limited to one discipline as it is an interdisciplinary study. Design in itself can be seen as a distinct area of scholarship in the same way that science and humanities are distinct areas of scholarship. Design can be seen to embody a level of knowledge equal to that of science or humanities. Overall design can be defined as the science of material culture or the artificial world.

The study of Design Science can also include the study of the principles, practices and the procedures of design. The aforementioned observations about Design Science all have strong implications for Information Systems. Within the IS community there is also some discussion about whether the focus of design should be on the product, the design process, the design practice or on all of these.

This becomes evident that Design Science methodology is suitable for Information System studies. It is also affirming that Design Science is well suited for interdisciplinary study. The thesis study is an interdisciplinary study between Spatio-temporal database management systems and Geographic Information Systems (GIS) using the case-study of New Zealand vessel trajectory analysis. After conducting a rigorous study on the Design Science approach and its suitability in applying it in Information Systems studies, the following section discusses them in detail in this research process which is been discussed in detail in Section 3.3.

3.5 Application of Knowledge Discovery process

A detail discussion has been conducted in Chapter 2 under the section Knowledge Discovery process and Sense Making (Section 2.8). In Section 2.8 presents a detail review has been conducted on various studies (Bodnar., 2005; Klein et al., 2006, Giannotti and Pedreschi., 2008) of Knowledge Discovery. The purpose of this section is to conduct the comparison between these studies to set the guidelines for analytical study to apply during experimentation.

There has been proliferation of knowledge discovery process of identifying or discovering useful and yet undiscovered implicit knowledge from the real-world data. This process coincides with large datasets at various domains ranging from government, scientific study and industry. These datasets linked together across several attributes of place, measurements, location, discipline and time. In particular, very large databases consist of rich information in terms of attribute depth along with many records of objects. These attributes and objects are represented to play a great role in the development of knowledge discovery.

According to Giannotti and Pedreschi (2008) the knowledge discovery process usually involves experimentation, iteration and many design decisions. As discussed earlier in Section 2.8, although Klein et al (2006) process is adapted to transform raw data into trajectory data, it is later noticed that the combination findings from of **Klein et al(2006), Bodnar (2005) and Giannotti et al (2008) are** applicable to this research study. Table 3.3 illustrates a comparative study conducted on the Knowledge Discovery process of previously mentioned studies. Table 3.3 also demonstrates how each of the steps of knowledge discovery process were applied systematically in this analytical study.

Table 3.3: Comparative Study and demonstration of Knowledge discovery process in this thesis study.

Klein et al (2006)	Bodnar (2005)	Giannotti and Pedreschi (2008)	Steps applied in this thesis study
Data	*External data sources*- the raw evidence of information.	Understanding of the knowledge domain	Data - Maritime domain -Selection of dataset from around 28 GB raw dataset -Shoebox method chosen to select datasets.
	Shoebox - represents a subset of the external data that is relevant for processing.	Selecting a data set/ focusing on a sub-set of variables /selecting data samples	
		Data cleaning and pre-processing	
Information	*Evidence file* - refers to snippets extracted from the shoebox	Data reduction and transformation	Information -Cleaning raw dataset and select -Decoding selected raw dataset using developed decoder -outlined decoded dataset
	Schema - representation of the information to draw conclusions		

Knowledge	Hypothesis - tentative representation of these conclusions with supporting arguments	Knowledge extraction • Pattern discovery • Classification • Clustering	Extracting hidden knowledge -Trajectory formation - Characteristics of moving vessel -Pattern discovery
Understanding	Presentation - present other work product used for communication.	Evaluating /validating the results/ Consolidating discovered knowledge	Presentation of results -Evaluation - Understanding patterns

3.5.1 Use of Problem Structuring Methods in Design Science approach

This section explains the bridging between practitioners and academic communities in filling the gap through the Design Science approach.
Keys (2007) had a strong belief in Problem Structuring Methods (PSM) which is a canon of knowledge bases that guide and support decision makers to solve technological problems. PSM recognises and supports practices which help to first decide what the problem is rather than concentrating on finding the best possible solution (Rosenhead & Mingers, 2001). Examples already available which shows the use of PSMs mainly consists of object and realization knowledge and that producing practice/process knowledge will further facilitate the understanding of how PSMs are to be used.

Modelling a Design Science approach for the use of PSMs will improve the accessibility of knowledge about standard practices and the expertise of analysts working to solve technological hurdles. Design Science involves the construction of theory and systems suitable for practical purposes. The Design science paradigm generates technological rules that are valid and reliable. Within Design Science, technological rules are validated by field-testing in case studies and grounding against a suitable theoretical framework (Keys, 2007). There are some differences that always exist as double hurdle problems in management studies between the importance of academic research methodologies and practical relevance (B. Kuechler & Vaishnavi, 2011). Knowledge obtained by Design Sciences is valuable to both practitioners (technological rules) and academic communities (rules being field-tested and grounded). Design Science contributions are found at the junction connecting description-driven and prescription-driven research.

Field-testing involves applying technological rules in a practical domain by developers of rules and also by third parties to provide external assessment. Grounding enables professionals to interpret why an observed outcome results from a particular action and provides rich understanding of various situations. Grounding helps to form theoretical rules of thumb and field-testing allows those rules to be justified by practical application. There are some well-developed field-tested and grounded rules available for the users of PSM but they are not very clearly presented (Keys, 2007). They exist as 'thick descriptions' (van Aken, 2005) that explore and provide explanations of real situations. Process knowledge should be developed (field-tested and grounded) and presented as technological rules to complement object and realization knowledge already available. A repository of object-design, realization design and process design will help professionals achieve expertise in the use of PSMs. In the current research study, in order to build an artefact for decoding raw data of NMEA format to spatio-temporal data, the guidelines of Design Science methodology are applied.

3.6 Case Study: Research Problem and motivation

Exploring and analysing large datasets of movement data has become a vital part of research in many disciplines. A significant number of large datasets of movement data is thus available. Automatic Identification Systems (AIS) are used in maritime monitoring as a means for littoral states to receive information on vessels operating near their coasts. AIS equipped vessels electronically broadcast information continuously at regular intervals about vessel information including its identity, position, course and speed. The importance of AIS, how it works, classification of AIS and its applications were discussed in detail in Chapter 2.9. AIS information becomes a critical tool to discover hidden knowledge related to the global maritime environment that could affect the safety, security, economy or environment of New Zealand. This study is conducted using AIS data provided by Kordia New Zealand recorded between Feb 2011- May 2011. A Spatio-temporal analysis approach is selected to conduct this study. This paper focuses on achieving the optimization of the information to discover hidden knowledge from the raw datasets. The hidden knowledge implies the spatio-temporal information extracted from the raw-dataset. The purpose of optimizing information by conducting trajectory analysis is to identify the characteristics of different type of moving vessels in New Zealand waterways. The knowledge discovered can also be applied in other aspects such as prediction of arrival and destinations, safety, security and additional navigational aids in maritime.

3.7 Design Science Approach applied in this research study

The Table 3.4, shows following the guideline of Design science approach in this research study.

Table 3. 4: Design Science Guidelines applied in this research study.

Guideline	Description	Compared with designed approach
1. Design an Artefact	a. Research based on Design-Science must produce a practical artefact.	To perform analytical study on trajectory of moving vessels, first and foremost the raw data has to be decoded. An artefact to decode movement information from raw data has been designed using VB.net. To design the decoder, standard procedures and methods are followed according to the IALA guidelines on AIS (IALA, 2004). Analysis carried out using ArcGIS software.
	b. Artefact in form of model or a method.	Artefact designed to support NMEA format raw data and capacity to intake between 1 GB -1.5 GB file size. This artefact has a major advantage to handle large file size and is user-friendly.

2. Problem Relevance	***Objective of Design-Science:*** a. To develop technology-based solution b. Problem Relevance	This approach is entirely technology based. The approach is followed by developing the artefact using in object-oriented design model. An artefact is designed to decode the raw dataset to read the spatio-temporal information and then later utilize to analyse trajectories to identify the characteristics of different type of moving vessels. There are three problem relevance identified. First, decoding raw data to readable spatio-temporal data. Second, handling large voluminous data set. Third, to conduct trajectory analysis using decoded data exported to ArcGIS tool. The research is focused on New Zealand waters.
3. Design Evaluation	***Rigorously demonstrated:*** a. Utility, quality and efficiency of the design. b. Evaluations done with well-executed method	Utility: i. To decode raw dataset in large voluminous file sizes. ii. To study the trajectory characteristics and pattern of moving vessels in New Zealand Waters. To closely and precisely study the trajectory pattern of moving object particularly 9 selected different types of vessels which is discussed in detailed in Section 5.3. Evaluation is conducted on the following criteria i. Analysis on velocity trend. ii. Frequency distribution of speed parameter. iii. Stationary Movements of vessels anchored in harbour. Hevner et al., (2004) states that an efficient analysis and evaluation method must be experimental, analytical, and descriptive.
4. Research Contributions	***Clear and Verifiable Contributions:*** In order to be	**Design Tool:** An artefact is designed to solve the two-fold

	effective, the Design-Science research must provide contributions very clearly and verifiable in design artefact, designing foundation and design methodologies. Artefacts are methodologies followed for system development, design tools and even prototype designing.	problem. i. To decode raw data of NMEA format into readable spatio-temporal database. ii. During the process of decoding large voluminous of data need to be handled. Methodology: Thus the developed artefact helped to decode the raw data into spatio-temporal data that could be used for analysis using ArcGIS tool. With the help of ArcGIS, a spatio-temporal approach is undertaken to conduct the following analysis. (i) Analytical (ii) Descriptive – Informed argument and scenarios (Hevner et al, 2004). The above mentioned analysis will be conducted to study the trajectories of 9 different types of vessels. Evaluations will be conducted on scenarios such as period of return trip, Spatial Analysis, Frequency distribution on speed parameter - mean, median and mode, Analysis on Stationary movements.
5. Research Rigor	***Relies on application of rigorous methods:*** Design-science highly recommends and relies on application with rigorous method evaluation in construction as well as designing.	According to Hevner et al (2004) the designed artefact should work effectively. This artefact to analyse trajectory behaviour spatio, temporal and spatio-temporal aspects. The main requirement of Design science is rigour and relevance. The design of artefacts is based on research conducted on data models of trajectory model. This model particularly designed considering maritime domain. Evaluation will be performed carefully on decoded datasets to analyse various type of trajectories formed by moving vessels.
6. Design as Search Process	***Utilizing availability:*** a. The search for an effective artefact requires utilizing	Design definitely requires a search process to come up with an effective solution to the problem. In this approach, extensive research work has been carried by studying

	available means to reach desired ends Hevner et al., (2004). b. As well as satisfying laws in the problem environment.	various similar works conducted in military, freight vessels etc. The supporting literatures reviewed and articles rationality confirms the study of trajectory analysis of moving objects has got involved additional complexity of handling spatio-temporal database challenges. In addition, to conduct analysis using such spatio-temporal data needs understanding and knowledge. Supporting Spatio-temporal datasets: ArcGIS supports extensively spatio-temporal datasets and plots the positional coordinates recorded from AIS devices. Providing various analysis feature: ArcGIS also provides a rich set of analytical features. These features help to conduct an evaluation with rigour, which becomes as one of the critical component of Design Science. The problem definition is well analysed and framed the research question to satisfy the chosen case study.
7. Communication of Research	The processed information has to be presented in effective method that reaches all kind of audiences - technology and management.	Design-Science research places emphasis on the importance of the problem as well as novel and effective approach solution for the artefact. The chosen approach, emphasis to analyse the trajectory of a moving object in context of New Zealand maritime datasets. This gives more focus on different type of vessels entering and leaving New Zealand maritime zone. This provides a close study on activities performed by various kinds of boats. The artefacts designed are effectively to decode the raw datasets collected from AIS devices plus analyse trajectory movements of various types of vessels discussed in Section 5.3.

The following section is a discussion on workflow of proposed two-phased study.

3.8 Workflow of Two-Phased Study

The objective of this section is to provide a detailed picture of the overall data collection and analysis process for the thesis study. The aim of this research study is “how to optimize information to analyse the trajectory of moving objects using the raw data obtained from AIS terminals fitted in vessels”.

In order to conduct analysis from raw data to trajectory involves a two-phased methodological process. Phase 1 primarily focused on handling the large dataset and developing a decoder. Given a large dataset (2GB - 30 Million rows) of spatio-temporal movement observations segmentation technique applied to obtain dataset of certain time period, later decoding the data then finally process of cleaning the noise filled dataset. Phase 2, trajectory analysis is conducted using spatio-temporal approach that included series of methods such as projecting trajectories on ArcGIS, plotting velocity trend graphs and clustering techniques. The Figure 3.4 presented a detail picture of the workflow of the overall thesis study.

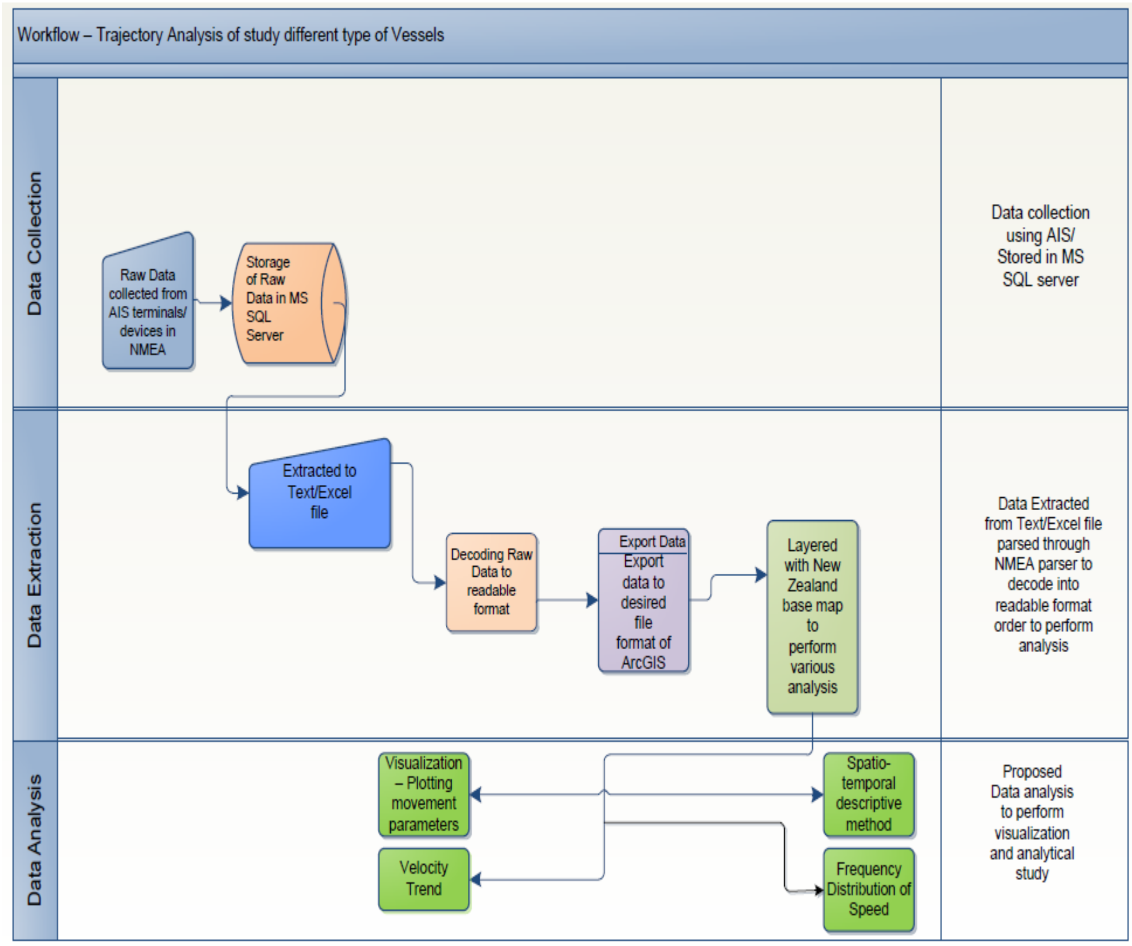


Figure 3. 4 : Illustration of Workflow of this trajectory analysis study.

There is a series of steps involved in this research activity during the process of trajectory analysis.

Automatic Identification System (AIS) data collection used on ships and by Vessel Traffic Services (VTS) mainly is to identify and locate vessels electronically. The data has been exchanged with other nearby ships and AIS Base stations. AIS information supplements the data supplied from the marine radar which is collected by New Zealand Maritime. Data are information collected to track real-time vessel information such as ship identity, purpose, course and speed, primarily in New Zealand Waters. AIS is intended to assist in the tracking and monitoring of vessel movements. This serves as a great advantage to maritime authorities. The datasets been used in this thesis study were collected between March 2011 and May 2011. The total size of the dataset provided by Kordia is around 28GB (7CDs – 4GB each).

To elaborate on the process steps involved from raw data to trajectory analysis, the following is a detail description is presented below.

3.8.1 Process of Knowledge Discovery (Phase 1)

This is the first phase of the study to open the original dataset from MSSQL server 2008 and then decode the raw movement data. This stage or phase also focused on handling the large dataset and developing a decoder. Given a large dataset (2GB - 30 Million rows) of spatio-temporal movement observations, the goal is to perform segmentation into sets of certain period of data. The objective of this segmentation process is to be able to handle the datasets in manageable sizes and to overcome the issues of insufficient memory.

Step 1: - ***Decoding Raw Data:-*** The original dataset is provided by Kordia New Zealand as raw data in NMEA 0183 format. The raw data is exported to MS SQL Server 2008. The datasets are raw AIS information in NMEA format and are time stamped with Timestamp, TalkerID and SentenceID details. To select a certain period of data, a SQL query is conducted in order to be able to select valid rows with TalkerID and Timestamp. Selection of certain period of data is demonstrated in Appendix E.

Step 2: The selected set of raw AIS sentences that belongs to certain period of data are saved as a text file. The next step involves decoding the raw data using the developed decoder. The decoded dataset is collected as an OUTPUT text file. The purpose of the decoder is to decode the raw data to read the information transmitted from the vessel. The decoder was developed in VB.net and can extract information such as the **Maritime Mobile Service Identity** (**MMSI**), latitude, longitude, speed over ground (SOG) and timestamp.

Step 3: The OUTPUT text file of each database has to be imported to an Excel sheet in order to perform the next level of tasks that involve data cleaning procedures.

Step 4:- *Data Cleaning*: One of the major challenges involved in handling raw datasets is cleaning them to ensure the removal of noise and invalid data.

Step 5: ***Data Selection:*** A subset of the decoded dataset is selected to perform analysis and imported as an Excel file of a manageable size in order to avoid exceptional memory errors. The original data was 2GB, and such huge datasets could not be handled as a single file because the system would encounter exceptional memory errors and would not be able to respond to queries.

Step 6: In this step exports the cleaned decoded dataset is exported to the ArcGIS tool to conduct **the** trajectory analysis.

There was a particular focus on the speed parameter and other space (Longitude and Latitude) and time attributes. Another important attribute used to identify particular vessels is MMSI. Other attributes may be utilized in future studies. The decoded datasets were thus obtained and taken to the next level to conduct trajectory analysis.

3.8.2 Trajectory Analyses (Phase 2)

Trajectory analysis was conducted using a GIS tool (ArcGIS) to study and understand the trajectory patterns of different types of vessels. First the decoded dataset was projected on the GIS tool. Then a 2D graph of speed over time was plotted against to study their movement trend using speed parameter. Following a detailed study of a frequency distribution of speed, a particular vessel type was selected and thoroughly investigated on speed profile. Also a spatio-temporal approach using selected a clustering technique was applied to study the vessel’s anchored positions and provide insights into the vessels particular characteristics. The objective of applying various methods such as plotting trajectories on a GIS tool and studying speed profiles over time and anchored positions was to rigorously investigate and identify the characteristic features of different types of vessel movements.

3.8.3 Contributions and Communication

According to Hevner et al. (2004) evaluating the artefact satisfactorily when the research process is concluded, it should be communicated as a research output. Further, mentioning the research contributions and including new knowledge is the final phase of the Design Science approach. The results of the experiments provide the indication on "how well" the prototype is performed. However, if the research outputs are not acceptable and satisfactory, the experiment needs to be repeated to obtain better suggestions to improve the artefact and evaluation. Otherwise it could become a future research along with along with looking at opportunities for improvement of the current research. This thesis resulting from the current research study of encoding raw data and presenting results and is itself a communication of the research process and research outcome.

3.9 Summary

This chapter outlined and discussed the research approach and process to conduct the current research study using Design Science. The main research question "How to optimize information to analyse trajectory of moving objects using raw data collected from AIS terminals?" was applied to an artefact. The process of conversion or encoding of raw data was outlined. The research process was analysed following the steps proposed by **Ostrowki et al. (2011).** Further, performed a comparative study with guidelines of Design Science proposed by Hevner et al. (2004), Vaishnavi & Kuechler (2008) and Peffers et al. (2008).

The process of data collection (raw AIS data) and encoding raw data was described. The main purpose of the research study is to perform trajectory analysis using decoded data. The rationale for the study on trajectory analysis is to have better information of vessel movements and their patterns. The following chapters cover in depth how an artefact is built to encode raw AIS data, perform trajectory analysis using encoded data and evaluation conducted on velocity plots, frequency distribution using speed parameter and stationary movements. The evaluation is performed through empirical observations and served as the development of research findings.

4 Decoding Raw Data – Knowledge Discovery Process (Phase I)

The objective of this chapter is to describe how to develop a decoder was developed to decode the raw datasets of NMEA 0183 format in order to extract spatio-temporal movement information from raw data. The purpose to develop a decoder is important, as this serves as a first step towards to conduct trajectory analysis and being able to answer the research questions.

This chapter also explores the issues related to handling large raw datasets. A detail exploration will be conducted on character of raw data and fields involved. Based on a detailed study of ITU-R M.1371 standard and IALA Technical Clarifications on Recommendation (ITU-R M.1371-1), a decoder is developed to decode raw data recorded from the vessels. A detail exploration has also been conducted on the process of filtering noise from decoded datasets.

4.1 Importance of Datasets

Most scientific research based study is based on the gathering and analysis of data. Data becomes the central and vital component of every research study.
The data in a scientific research study could be categorized into two, namely: Real time data sets and simulated data sets. Data collected from sensors over a period of time from a moving object is categorized as real time data. Simulated datasets are generated in a controlled environment using a model.

Challenges were always enfold in analysing both type of data sets. Real time data sets need to be cleaned and processed before performing any analytical study whereas in simulated datasets thus generated using a model has to be adjunct to or substitute to actual model after thorough study. According to (Spaccapietra, et al., 2008) to simulate typical movement data of a moving object, it is important to have detailed information about the trajectory of the object, in addition to information related to environmental conditions.

The availability of relevant data is a critical success factor for empirically based research (Dodge, Weibel, & Forootan, 2009). Relevant data of a particular domain is more important to perform a rigorous study. In order to perform a rigorous study, real data sets are more suitable as they exactly inform the real situation and condition. Especially, studying the trajectory of a moving object such as a car in a busy city road during peak hours or a high speed craft sailing towards its destination during off-peak time are examples.

4.2 Challenges handling large raw datasets

The importance of using real datasets in an analytical study has been stressed in various studies (Dodge, Weibel, & Forootan, 2009; Spaccapietra, et al., 2008). However, challenges are always incurred at various levels. One of the major challenges with handling a large raw dataset is its size. Another commonly faced challenge is the type of the data – raw data. In practice, sensors used in moving objects can be used to make available location and time information. These sensors emit raw data. Mostly real time datasets are raw datasets. In a real world application, large sets of raw data are collected. It is a difficult task to extract information and later process it to extract meaningful patterns from such massive data sets. The task involves following, a series of steps which have to be carried out systematically ranging from decoding, filtering, selection of data subset and analysis performed using an appropriate analytical tool.

4.3 A close look at previous studies

In this section, a thorough investigation is conducted on previous similar studies. The intention of this study is to investigate the usage of simulated data and real time data used in previous and similar studies. An investigation conducted on previous studies is an attempt to understand the application of type of dataset been used in those studies.

Table 4.1 : Investigation conducted on previous similar research studies.

Research Studies	Domain	Data used	Real-time dataset used
Brillinger, Preisler, Ager, and Kie (2004)	Wild-life free-range moving animals in particular deer and elk	Telemetry	✓
Laube and Purves (2006)	Urban Walking patterns	By simulation	X
Giannotti and Pedreschi (2007)	Urban domain GPS car datasets.	Real-time GPS dataset.	✓
Hornsby and Cole (2007)	Maritime domain (Vessel transit from offshore to the ferry landing)	Dataset collected from ferries	✓
Etienne et al. (2010)	Maritime Domain (Data sets collected from ships during 2 years around the Brest area.)	Data sets collected from ships.	✓
(Demsar & Virrantaus, 2010)	Maritime Domain (Gulf of Finland)	Data sets collected from ships.	✓
Gerben K. D. de Vries (2012)	Maritime – vessel navigation data	Aggregation of web data (marine traffic.com)	✓

From the Table 4.1 investigation, the importance of utilizing real-time datasets in order to ensure reliability of the outcomes of the study can be deduced.

Another interesting study has been conducted by Si & Aung (2011) on GPS devices of **RMC sentence** using NMEA protocol. This study undertook to extract positional information (latitude and longitude) for navigation and positional information (Si & Aung, 2011), whereas this research study is focused on !AIVDM sentence of NMEA protocol of AIS raw data. As discussed earlier a raw dataset has to be decoded and then an analysis has to be carried out using the ArcGIS tool.

4.4 Steps involved in decoding raw data

To perform any analytical study the first step involved is the decoding of raw data. A further challenge includes the size of the datasets. In the context of vessel trajectory case study, the recorded raw dataset from Automatic Identification System (AIS) were used. These datasets are large and complex to decode. It is a difficult task to extract meaningful patterns from large AIS datasets as they need to be decoded and coined to trajectory data. Further decoding of raw data to extract in order to remove noise and misread data from the observations. Additional challenge involved in characterization and comparison (Guo, Liu, & Jin, 2010) from different types of vessels and boats.

4.4.1 Opening Raw Datasets

AIS is intended to assist the tracking and monitoring of vessel movements. This serves as a great advantage to maritime authorities. These AIS messages are raw data. A sample of raw data has been provided by Kordia New Zealand. The datasets being used in this thesis study were recorded March 2011- May 2011. The total size of the dataset is around 28GB (7CDs – 4GB each).

Raw datasets are saved in **.mdf** format. The provided raw data is exported to MS SQL Server2008. The datasets are raw AIS information with time stamped along with TalkerID and SentenceID. The raw data is saved in NMEA format.

Table 4.2: Illustration of the dataset size (in GB) and rows. Original dataset provided by Kordia New Zealand.

SNO	Name of the Database	File Size	Total no.of rows in each Database	After selecting !AIVDM sentences
1.	DBW9	2GB	31556311	7428087
2.	DBW10	2GB	30030064	6431881
3.	DBW11	1.98GB	30569527	6758447
4.	Dbw12	2GB	30274507	6844678
5.	DBW13	2.4GB	30422013	6801562
6.	DBW14	2GB	30837331	6938492
7.	DBW15	2GB	30559526	6432882
8.	DBW16	2GB	30801061	6421981
9.	DBW17	2GB	26730230	5775511
10.	DBW18	2GB	22160407	5210036
11.	DBW19	2GB	30151872	6419352
12.	DBW20	2GB	30256812	6533954
13.	DBW21	2GB	30246812	6533853

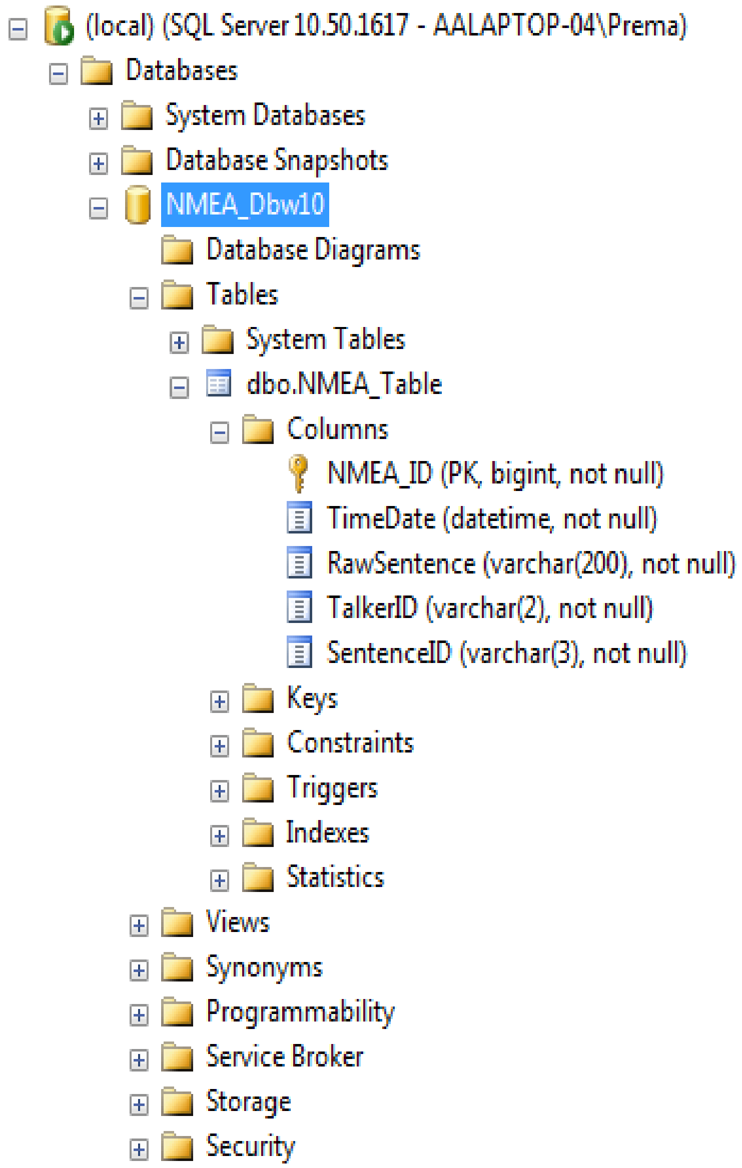
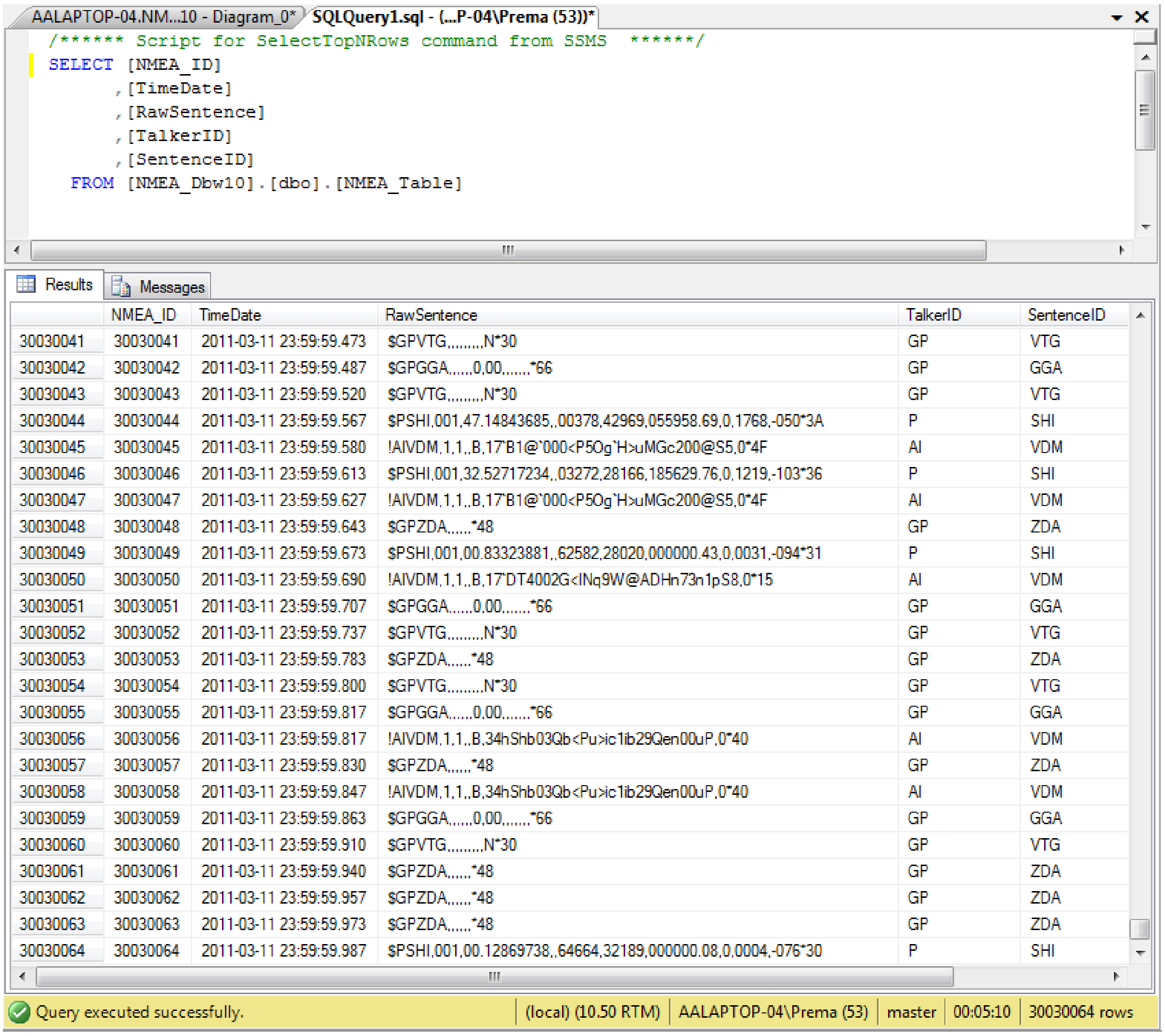


Figure 4.1: Screenshot display of raw data opened in MS SQL Server 2008.

The original file is of .mdf file format used MS SQL Server R2 2008 to open them. These files were able to be open only using an Admin account. The next step is to select only !AIVDM format sentences from a particular file. In order to select !AIVDM format sentences, the following SQL statement was utilized (shown in the Figure 4.2).



AALAPTOP-04.NM...10 - Diagram_0* SQLQuery1.sql - (...P-04\Prema (53))*

```sql
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT [NMEA_ID]
      ,[TimeDate]
      ,[RawSentence]
      ,[TalkerID]
      ,[SentenceID]
  FROM [NMEA_Dbw10].[dbo].[NMEA_Table]
```

Results | Messages

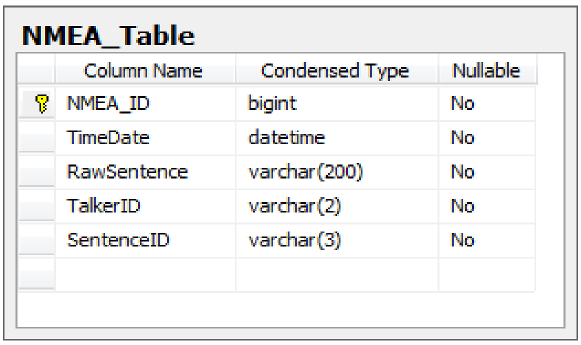
	NMEA_ID	TimeDate	RawSentence	TalkerID	SentenceID
30030041	30030041	2011-03-11 23:59:59.473	$GPVTG,,,,,,,,,N*30	GP	VTG
30030042	30030042	2011-03-11 23:59:59.487	$GPGGA,,,,,,0,00,,,,,,*66	GP	GGA
30030043	30030043	2011-03-11 23:59:59.520	$GPVTG,,,,,,,,,N*30	GP	VTG
30030044	30030044	2011-03-11 23:59:59.567	$PSHI,001,47.14843685,,00378,42969,055958.69,0,1768,-050*3A	P	SHI
30030045	30030045	2011-03-11 23:59:59.580	!AIVDM,1,1,,B,17`B1@`000<P5Og`H>uMGc200@S5,0*4F	AI	VDM
30030046	30030046	2011-03-11 23:59:59.613	$PSHI,001,32.52717234,,03272,28166,185629.76,0,1219,-103*36	P	SHI
30030047	30030047	2011-03-11 23:59:59.627	!AIVDM,1,1,,B,17`B1@`000<P5Og`H>uMGc200@S5,0*4F	AI	VDM
30030048	30030048	2011-03-11 23:59:59.643	$GPZDA,,,,,,*48	GP	ZDA
30030049	30030049	2011-03-11 23:59:59.673	$PSHI,001,00.83323881,,62582,28020,000000.43,0,0031,-094*31	P	SHI
30030050	30030050	2011-03-11 23:59:59.690	!AIVDM,1,1,,B,17`DT4002G<INq9W@ADHn73n1pS8,0*15	AI	VDM
30030051	30030051	2011-03-11 23:59:59.707	$GPGGA,,,,,,0,00,,,,,,*66	GP	GGA
30030052	30030052	2011-03-11 23:59:59.737	$GPVTG,,,,,,,,,N*30	GP	VTG
30030053	30030053	2011-03-11 23:59:59.783	$GPZDA,,,,,,*48	GP	ZDA
30030054	30030054	2011-03-11 23:59:59.800	$GPVTG,,,,,,,,,N*30	GP	VTG
30030055	30030055	2011-03-11 23:59:59.817	$GPGGA,,,,,,0,00,,,,,,*66	GP	GGA
30030056	30030056	2011-03-11 23:59:59.817	!AIVDM,1,1,,B,34hShb03Qb<Pu>ic1ib29Qen00uP,0*40	AI	VDM
30030057	30030057	2011-03-11 23:59:59.830	$GPZDA,,,,,,*48	GP	ZDA
30030058	30030058	2011-03-11 23:59:59.847	!AIVDM,1,1,,B,34hShb03Qb<Pu>ic1ib29Qen00uP,0*40	AI	VDM
30030059	30030059	2011-03-11 23:59:59.863	$GPGGA,,,,,,0,00,,,,,,*66	GP	GGA
30030060	30030060	2011-03-11 23:59:59.910	$GPVTG,,,,,,,,,N*30	GP	VTG
30030061	30030061	2011-03-11 23:59:59.940	$GPZDA,,,,,,*48	GP	ZDA
30030062	30030062	2011-03-11 23:59:59.957	$GPZDA,,,,,,*48	GP	ZDA
30030063	30030063	2011-03-11 23:59:59.973	$GPZDA,,,,,,*48	GP	ZDA
30030064	30030064	2011-03-11 23:59:59.987	$PSHI,001,00.12869738,,64664,32189,000000.08,0,0004,-076*30	P	SHI

Query executed successfully. | (local) (10.50 RTM) | AALAPTOP-04\Prema (53) | master | 00:05:10 | 30030064 rows

Figure 4.2: Screenshot illustrating a successfully executed query of Dbw10 (30030064 rows).

4.4.2 Raw data – Data type

The raw dataset data types are as follows:

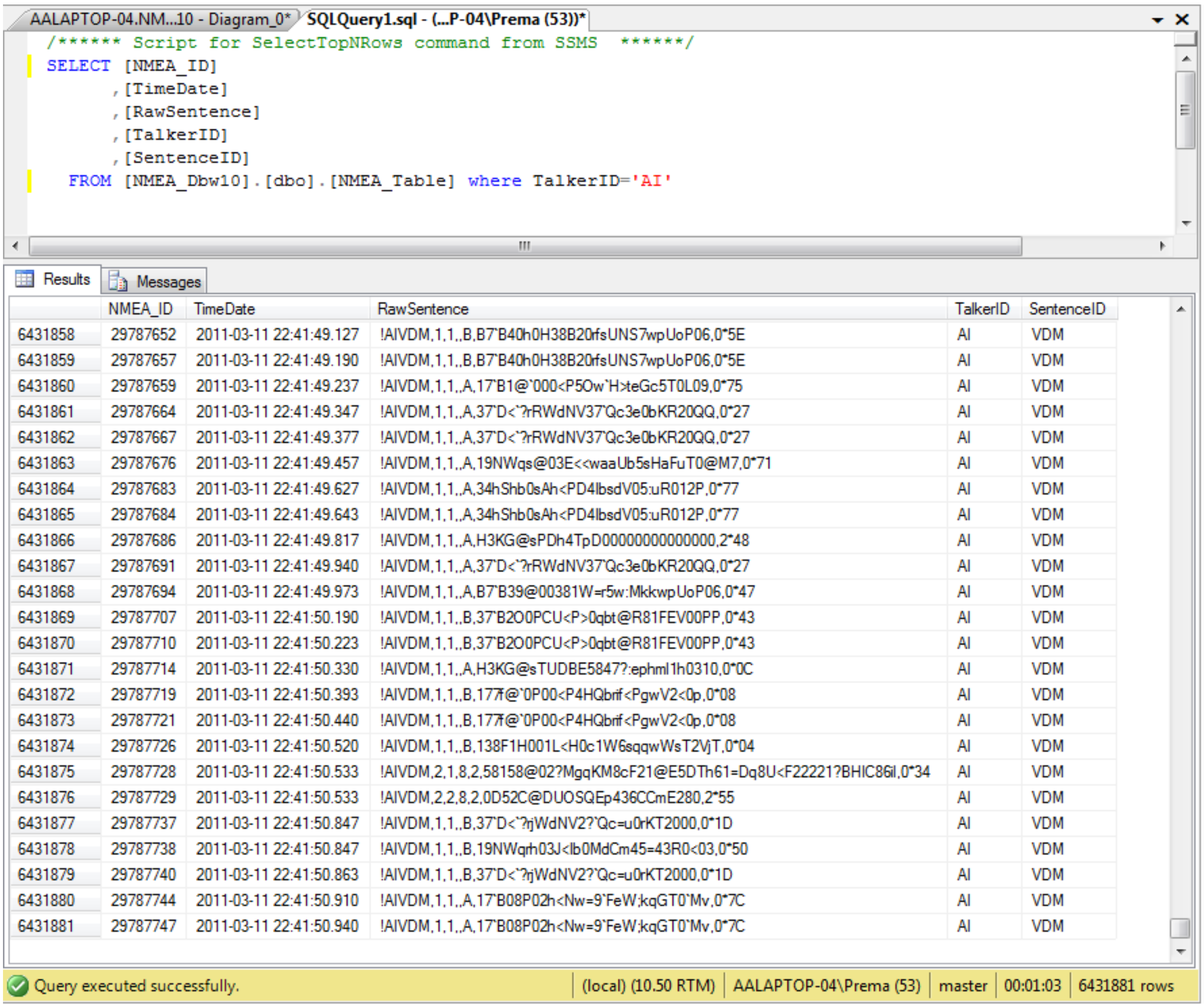


NMEA_Table

Column Name	Condensed Type	Nullable
NMEA_ID	bigint	No
TimeDate	datetime	No
RawSentence	varchar(200)	No
TalkerID	varchar(2)	No
SentenceID	varchar(3)	No

Figure 4.3: Depicts data type of provided raw dataset.

In particular selecting TalkerID='AI' provides VDM sentences i.e., raw sentences that starts with !AIVDM.




```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT [NMEA_ID]
      ,[TimeDate]
      ,[RawSentence]
      ,[TalkerID]
      ,[SentenceID]
  FROM [NMEA_Dbw10].[dbo].[NMEA_Table] where TalkerID='AI'
```

	NMEA_ID	TimeDate	RawSentence	TalkerID	SentenceID
6431858	29787652	2011-03-11 22:41:49.127	!AIVDM,1,1,,B,B7`B40h0H38B20rfsUNS7wpUoP06,0*5E	AI	VDM
6431859	29787657	2011-03-11 22:41:49.190	!AIVDM,1,1,,B,B7`B40h0H38B20rfsUNS7wpUoP06,0*5E	AI	VDM
6431860	29787659	2011-03-11 22:41:49.237	!AIVDM,1,1,,A,17`B1@`000<P5Ow`H>teGc5T0L09,0*75	AI	VDM
6431861	29787664	2011-03-11 22:41:49.347	!AIVDM,1,1,,A,37`D<`?rRWdNV37`Qc3e0bKR20QQ,0*27	AI	VDM
6431862	29787667	2011-03-11 22:41:49.377	!AIVDM,1,1,,A,37`D<`?rRWdNV37`Qc3e0bKR20QQ,0*27	AI	VDM
6431863	29787676	2011-03-11 22:41:49.457	!AIVDM,1,1,,A,19NWqs@03E<<waaUb5sHaFuT0@M7,0*71	AI	VDM
6431864	29787683	2011-03-11 22:41:49.627	!AIVDM,1,1,,A,34hShb0sAh<PD4lbsdV05:uR012P,0*77	AI	VDM
6431865	29787684	2011-03-11 22:41:49.643	!AIVDM,1,1,,A,34hShb0sAh<PD4lbsdV05:uR012P,0*77	AI	VDM
6431866	29787686	2011-03-11 22:41:49.817	!AIVDM,1,1,,A,H3KG@sPDh4TpD00000000000000,2*48	AI	VDM
6431867	29787691	2011-03-11 22:41:49.940	!AIVDM,1,1,,A,37`D<`?rRWdNV37`Qc3e0bKR20QQ,0*27	AI	VDM
6431868	29787694	2011-03-11 22:41:49.973	!AIVDM,1,1,,A,B7`B39@00381W=r5w:MkkwpUoP06,0*47	AI	VDM
6431869	29787707	2011-03-11 22:41:50.190	!AIVDM,1,1,,B,37`B2O0PCU<P>0qbt@R81FEV00PP,0*43	AI	VDM
6431870	29787710	2011-03-11 22:41:50.223	!AIVDM,1,1,,B,37`B2O0PCU<P>0qbt@R81FEV00PP,0*43	AI	VDM
6431871	29787714	2011-03-11 22:41:50.330	!AIVDM,1,1,,A,H3KG@sTUDBE5847?:ephml1h0310,0*0C	AI	VDM
6431872	29787719	2011-03-11 22:41:50.393	!AIVDM,1,1,,B,177f@`0P00<P4HQbrif<PgwV2<0p,0*08	AI	VDM
6431873	29787721	2011-03-11 22:41:50.440	!AIVDM,1,1,,B,177f@`0P00<P4HQbrif<PgwV2<0p,0*08	AI	VDM
6431874	29787726	2011-03-11 22:41:50.520	!AIVDM,1,1,,B,138F1H001L<H0c1W6sqqwWsT2VjT,0*04	AI	VDM
6431875	29787728	2011-03-11 22:41:50.533	!AIVDM,2,1,8,2,58158@02?MgqKM8cF21@E5DTh61=Dq8U<F22221?BHIC86il,0*34	AI	VDM
6431876	29787729	2011-03-11 22:41:50.533	!AIVDM,2,2,8,2,0D52C@DUOSQEp436CCmE280,2*55	AI	VDM
6431877	29787737	2011-03-11 22:41:50.847	!AIVDM,1,1,,B,37`D<`?rjWdNV2?`Qc=u0rKT2000,0*1D	AI	VDM
6431878	29787738	2011-03-11 22:41:50.847	!AIVDM,1,1,,B,19NWqrh03J<lb0MdCm45=43R0<03,0*50	AI	VDM
6431879	29787740	2011-03-11 22:41:50.863	!AIVDM,1,1,,B,37`D<`?rjWdNV2?`Qc=u0rKT2000,0*1D	AI	VDM
6431880	29787744	2011-03-11 22:41:50.910	!AIVDM,1,1,,A,17`B08P02h<Nw=9`FeW;kqGT0`Mv,0*7C	AI	VDM
6431881	29787747	2011-03-11 22:41:50.940	!AIVDM,1,1,,A,17`B08P02h<Nw=9`FeW;kqGT0`Mv,0*7C	AI	VDM

Figure 4. 4: Screenshot showing a selection of TalkerID='AI' chosen from !AIVDM sentences. The number of rows selected is 6431881 rows.

Using the SQL statement the TalkerID = 'AI' is selected to retrieve only! AIVDM sentences. Selecting or highlighting a raw sentence column and timestamp column, the entire set is exported to a text file. While exporting, the columns were swapped as follows: Raw sentences TimeStamp.

4.4.3 Exporting raw dataset from MS SQL server R2 2008 to text file of manageable size

Another consideration to be taken into account while exporting to text file is size. The reason for exporting as text files in a manageable size is because the original file is 2GB. The selection of only !AIVDM sentences from 2GB sized file, returns only of 88MB of data. The selected !AIVDM data is used for remaining stage of the decoding and analysis. The large datasets have imposed challenges in handling and managing them. A detailed description is provided in the Appendix [E] that demonstrates the approach taken to split the large sized text file to manageable size text files.

4.4.4 A closer look on !AIVDM sentence layers

The following table explains the raw data items and their purpose.

!AIVDM,1,1,,B,177KQJ5000G?tO`K>RA1wUbN0TKH,0*5C

Table 4.3 :Explanation on 7 fields of !AIVDM sentence.

Field No:	Item Found	Description
Field 1	!AIVDM	Identifies this as an AIVDM packet.
Field 2	1 (in this example)	Represents the count of fragments in the currently accumulating message. Each sentence payload size is limited to NMEA 0183’s 82 character maximum. At times it is required to split up a payload over several fragment sentences while reaching its maximum.
Field 3	1 (in this example)	The fragment number of this sentence. A sentence with a fragment count of 1 and a fragment number of 1 is complete in itself.
Field 4	(empty in this example)	A sequential message ID for multi-sentence messages.
Field 5	B (in this example)	A radio channel code. AIS use duplex from two VHF radio channels. AIS Channel A is 161.975Mhz (87B); AIS Channel B is 162.025Mhz (88B).
Field 6	177KQJ5000G?tO`K>RA1wUbN0TKH (in this example)	The data payload.

Field 7	0 (in this example) *	The number of fill bits required to pad the data payload to a 6 bit boundary, ranging from 0 to 5. Equivalently, subtracting 5 from this tells how many least significant bits of the last 6-bit nibble in the data payload should be ignored. *-separated suffix (*5C) is the NMEA 0183 data-integrity checksum for the sentence, preceded by "*". It is computed on the entire sentence including the AIVDM tag but excluding the leading "!".

The following section is a detail explanation on Field 6 and Field 7. The characteristic features of Class A and Class B transponders which mentioned as Field 5 is explained in the Appendix E.

4.5 Procedure to decode !AIVDM sentence

The following section will discuss in detail the series of steps involved in decoding a raw sentence. The process of decoding and interpreting the contents of the encapsulated string is a three step process:

Table 4. 5: Illustration on 3 step process of decoding raw dataset

Step 1: The string symbols are converted into corresponding 6 binary representation. **Step 2:** The binary strings are organised and then parsed by following the rules mentioned in the referenced document ITU-R M.1371-1, table 4. 6. **Step3:** Using the referenced document, rules are used and applied to convert the binary strings into the relevant information.

A detail explanation is provided in the following section.

Step 1: Conversion from symbols to binary bits

The procedure of conversion of the VDM decoding begins with the first symbol of the string. The procedure to identify the AIS message is carried out from the first six bits of the binary message data. This is simply identified with the message number of the decimal equivalent of the binary number. In this case, 000001 = message 1. These 6 bits occupy the bit position specifically 1 to 6. On the left most, ‘0’‘s are in position 1 and on the right, most ‘1’s are in the position 6.

Step 2: Organizing the Binary Message Data:

As discussed in the previous step identification of the AIS message is done based on the first 6 bits of binary message data, the next process involves applying parameters using Table 4.6, ITU-R M.1371-1. The “number of bits” is established the bits will be that applied to each of the parameters using the table 2, ITU-R M.1371-1 as reference.

Table 4. 6: ITU-R M.1371-1 of prescribed parameters

S.No	Type of Message	Bits
1.	Message ID	1-6
2.	Repeat Indicator	7-8
3.	User ID	9-38
4.	Navigation status	39-42
5.	Rate of turn	43-50
6.	Speed On Ground (SOG)	51-60
7.	Position accuracy	61
8.	Longitude	62-89
9.	Latitude	90-116

10.	COG	117-128
11.	True Heading	129-137
12.	UTC second when report generated	138-143
13.	Regional Application	144-147
14.	Spare	148
15.	RAIM Flag	149
16.	Communications State	150-168

However, in this study the motivation is to use MMSI, Latitude, Longitude and Speed on Ground (SOG). The remaining parameters will be considered in future studies.

Once the above mentioned ordering of bits has been established, it is important to always keep the same for a “message 1”. For each of the referenced AIS message table the challenge is to keep the same order. For example: if bits 1 through 6 were 010100 (6 bit representation) after the decoding process was complete, then the VDM message identified would be message 20 ($010100_2 = 20_{10}$). The procedure of organising the decoded binary message Data involves the following process:

- Identification of the message number
- Organise and parse the binary bites using the appropriate message table.

Step 3: Interpretation of the Decoded Binary Strings

The final process of the conversion involves the decoded bits to indicate the useful information. The two step process involved is:

- Organise bits accordingly to prescribed parameter of
- The parameters described in information table ITU-R M.1371-1 (Table 4.6)

The following session depicts a series of steps involved in developing a decoder.

4.5.1 Algorithm to decode raw data to meaningful movement data (messages)

Step 1. To check whether the given sentence isValid or not

Step 1.1. If the sentence isValid // the sentence that starts with !AIVDM

Step 1.1.1. Start with the !AIVDM array of ascii characters

Step 1.1.2. Convert to the 6 bit binary value.

Step 1.1.3. Convert this byte array to a 6 bit bitstream

Step 1.1.4. concatenate all decoded binary values into one long message string.

Step 1.1.5 Convert Binary to Decimal. Converted Decimal will correspond the following bit position and meaning

// Representation of bit position and their meaning.

Step 1.1.4.1.	0-5	Class A Report
Step 1.1.4.2	6-7	Repeat Indicator
Step 1.1.4.3	8-37	MMSI
Step 1.1.4.4	38-41	Navigation Status
Step 1.1.4.5	42-49	Rate of Turn
Step 1.1.4.6	50-59	Speed over ground
Step 1.1.4.7	60	Position Accuracy
Step 1.1.4.8	61-88	Longitude
Step 1.1.4.9	89-115	Latitude
Step 1.1.4.10	116-127	Course over ground
Step 1.1.4.11	128-136	True Heading (HDG)

else

Step 1.2 Exit

4.6 Trajectory Model

As discussed earlier in chapter 2 under Section 2.4.3 and after investigating the !AIVDM layers, a domain class diagram is illustrated below to assist in next stage of the process to develop a decoder.

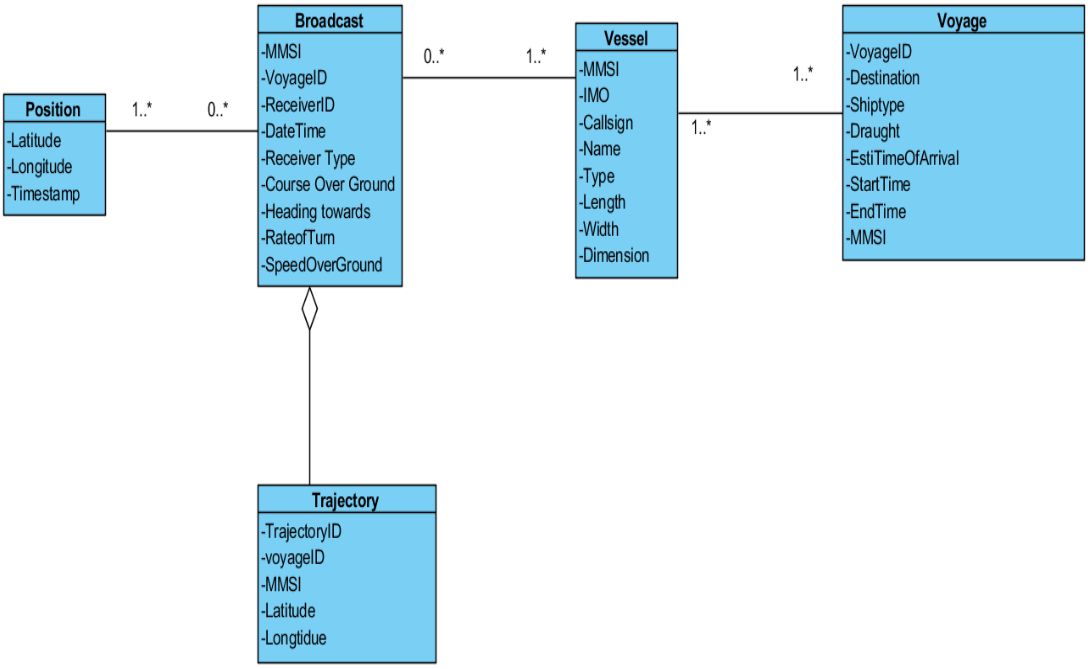


Figure 4. 5: Trajectory model - class domain diagram

4.7 Designing AIS Decoder

The decoder is designed using **VB.net** with **file handling** concepts. The raw data is made available in a text file format and the output is also collected in text file format. The reason a text file was chosen is that they are easy to handle, with the advantage of being able to control the text file size and the flexibility to export to various tools.

The following code explains the step involved to convert the 6 bit binary value (step 1.1.2)

Table 4.7: Code (VB.net snippets) instructs the procedure of conversion of 6 bit binary value

```
If Len (whole_string) = 168 Then

            'its a single sentence

            szArr = whole_string.ToCharArray()

            message_type = ""

            For I = 0 To 5

                message_type = message_type & szArr(I)

            Next

            ' MsgBox(message_type)

            message_type_number =
BinaryToDecimal(message_type)

            ' MsgBox(message_type_number & " " & ais_data)
```

The next set of coding explains the procedure of conversion to Binary.

Table 4.8 : Code (VB.net snippets) outlines the procedure conversion to Binary

```
'Converting to Binary
        For i = 0 To Len(frr) - 1
            'MsgBox(szArr(I))
            holding = Asc(frr(i)) - 48
            If holding > 40 Then holding = holding - 8
            'MsgBox(holding)
            If (holding And 32) = 32 Then bit = 1 Else bit = 0
            longstr = bit
            If (holding And 16) = 16 Then bit = 1 Else bit = 0
            longstr = longstr & bit
            If (holding And 8) = 8 Then bit = 1 Else bit = 0
            longstr = longstr & bit
            If (holding And 4) = 4 Then bit = 1 Else bit = 0
            longstr = longstr & bit
            If (holding And 2) = 2 Then bit = 1 Else bit = 0
            longstr = longstr & bit
            If (holding And 1) = 1 Then bit = 1 Else bit = 0
            longstr = longstr & bit
            whole_string = whole_string & longstr
            'MsgBox(longstr)
        Next i
```

This part of the coding explains the conversion procedure from Binary to Decimal number.

Table 4.9: Code (VB.net snippets) instructs conversion procedure from Binary to Decimal

```
Public Function BinaryToDecimal(ByVal Binary As String) As
Long

    Dim n As Long

    Dim s As Integer


    For s = 1 To Len(Binary)

       n = n + (Mid(Binary, Len(Binary) - s + 1, 1) * (2 ^ _

          (s - 1)))

    Next s

    ' ListBox2.Items.Add(Binary & "  " & n)

    BinaryToDecimal = n

 End Function
```

4.7.1 Dealing with Timestamp Attributes

Mathematically, time is identified as a continuous set of linear order and distances between the elements. The elements are defined as moments sometimes referred as positions in time. Compared to positions in space, it is necessary to have a reference system for the specification of moments in data. In most cases, temporal reference is carried out on the basis of the standard Gregorian calendar and standard division method: of a day into hours, hours into minutes and so on (Andrienko et al., 2008). Normally, time of the day is specified according to the time zone of the place, using Greenwich Mean Time (GMT). GMT is tagged along with collected data termed as TimeStamp.

4.7.2 Conversion of positional coordinates

The formula depicts the logic behind the conversion of binary to latitude and longitude coordinates.

Coordinates:

Degrees Minutes Seconds to Degrees Minutes.m
Degrees = Degrees
Minutes.m = Minutes + (Seconds / 60)

Degrees Minutes.m to Decimal Degrees

.d = M.m / 60

Decimal Degrees = Degrees + .d

(Or)

Decimal Degrees = Degrees + minutes/60 + seconds/3600

The geographical coordinates of New Zealand 41° 0' 0" S / 174° 0' 0" E

Table 4.10: Representation of Latitude and Longitude.

Latitude:	**North +**
	South –
Longitude:	**East +**
	West –

The below code illustrates the conversion of latitude and longitude data sets.

Table 4. 11: Outline presented depicts the technique and formula applied in conversion of latitude and longitude data.

```
Longitude = ""

'Lat - bit 89 for 27 bits (a signed binary number, divide by 600000)

'Lon - bit 61 for 28 bits (a signed binary number, divide by 600000)


          latnum = latnum / 600000

          longnum = longnum / 600000

'Calculating Longitude -Decoding binary bits to decimal        '
MsgBox(zz & "zz long is")

        Longitude = ""

         zz = 0

            For I = 62 To 88

              Longitude = Longitude & szArr(I)

              zz = zz + 1

            Next

'Calculating Latitude -Decoding binary bits to decimal

      Latitude = ""

            zz = 0

            For I = 90 To 115

              Latitude = Latitude & szArr(I)

              zz = zz + 1

            Next
```

The above coding is developed to calculate Longitude and Latitude from the bit position 62 to 88 and 90 to 115 respectively.

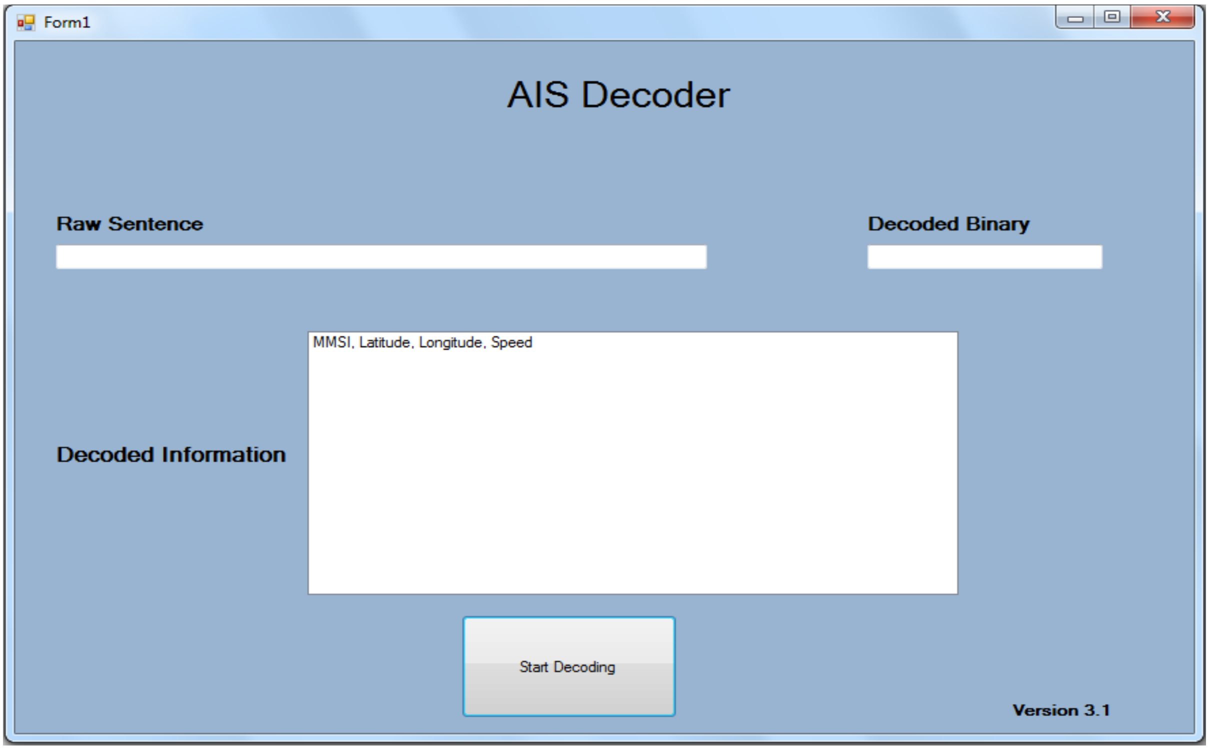


Figure 4. 6: Screenshot illustrating the developed AIS decoder.

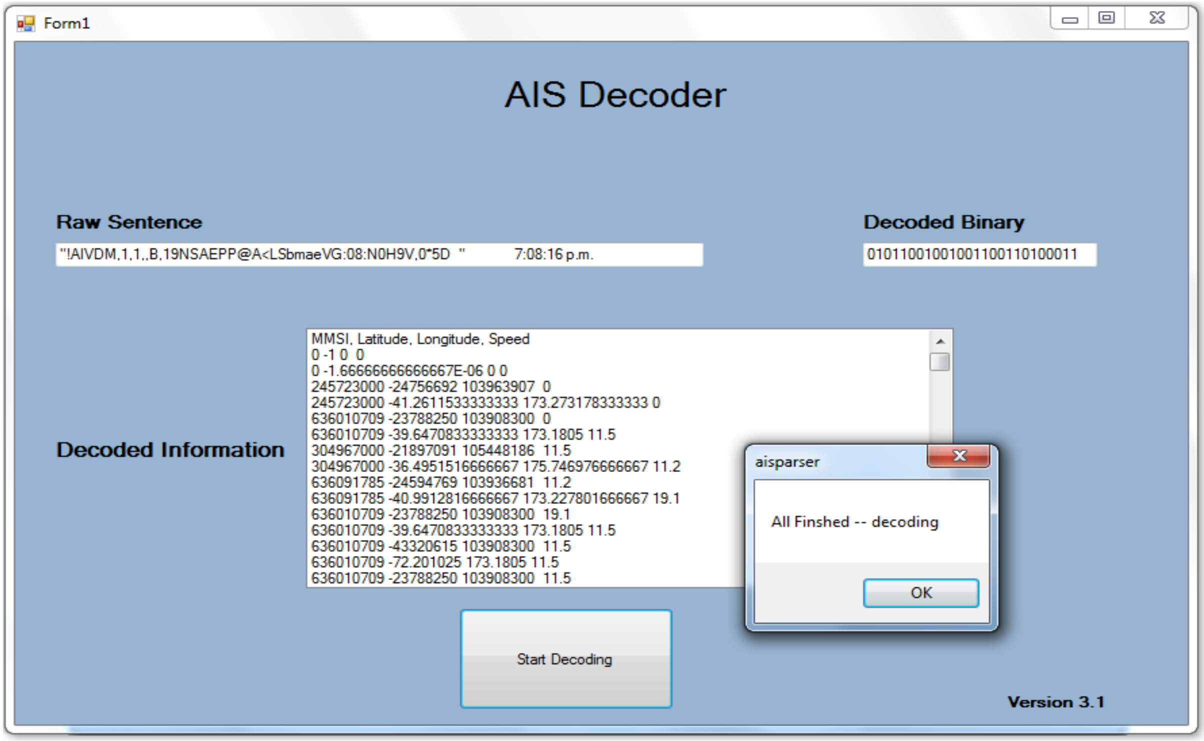


Figure 4. 7: Screenshot showing AIS decoder once has been finished decoding.

Step one: Once the text file is ready, the next step involved is to decode the raw data. Again the decoded information is collected as an OUTPUT text file. The OUTPUT text file as the following decoded information:

MMSI	Latitude	Longitude	Speed	Rawdata	Timestamp

Dbw19_12.05.2011_OUTPUT.txt

```
512000399 -36.7948766666667 174.678178333333 8.8 "!AIVDM,1,1,,A,17`B1ShPAH<OW?obtR0eGrRr08@`,0*09  "    12/05/2011 0:01
512131000 -41.39155 174.733116666667 14.7 "!AIVDM,1,1,,B,17`Iwf0vjC<OoEu`D<SSc2nl0404,0*13  "   12/05/2011 0:01
512131000 -41.39155 174.733116666667 14.7 "!AIVDM,1,1,,B,17`Iwf0vjC<OoEu`D<SSc2nl0404,0*13  "   12/05/2011 0:01
512131000 -70.4565583333333 174.733116666667 14.7 "!AIVDM,1,1,,B,37`B08P02r<MmAq`JQh23Abp00bi,0*4D  "   12/05/2011 0:01
512131000 -41.39155 174.733116666667 14.7 "!AIVDM,1,1,,B,37`B08P02r<MmAq`JQh23Abp00bi,0*4D  "   12/05/2011 0:01
305646000 -36.7879166666667 174.80785 11.8 "!AIVDM,1,1,,A,14SO9d541n<P=?=btjDeFrpn0D0T,0*08  "  12/05/2011 0:01
305646000 -36.7879166666667 174.80785 11.8 "!AIVDM,1,1,,A,14SO9d541n<P=?=btjDeFrpn0D0T,0*08  "  12/05/2011 0:01
512175000 -45.4892633333333 170.969163333333 14 "!AIVDM,1,1,,B,17`LcV002<<>`WmUv8R`G6ht08A2,0*53  " 12/05/2011 0:01
512002959 -41.284 174.7835 2.1 "!AIVDM,1,1,,A,17`B;Sh00E<P66a`H8`=SUlr0HA3,0*15  "  12/05/2011 0:01
512002959 41.2839983333333 174.7835 2.1 "!AIVDM,1,1,,B,B7`B<?h0637@v;ru:jVbSwfUoP06,0*5B  " 12/05/2011 0:01
512000741 -36.8411183333333 174.766781666667 7.6 "!AIVDM,1,1,,B,17`B2qOP1<<P1=;brm`Ah?vp0404,0*06  "    12/05/2011 0:01
512000741 -36.8411183333333 174.766781666667 7.6 "!AIVDM,1,1,,B,17`B2qOP1<<P1=;brm`Ah?vp0404,0*06  "    12/05/2011 0:01
305646000 -36.7879166666667 174.80785 11.8 "!AIVDM,1,1,,A,14SO9d541n<P=?=btjDeFrpn0D0T,0*08  "  12/05/2011 0:01
636091785 -41.0604766666667 173.287395 1.6 "!AIVDM,1,1,,B,19NWuRB00@<I?jc`PDPUbp2r00ST,0*76  "  12/05/2011 0:01
636091785 41.060475 173.287395 1.6 "!AIVDM,1,1,,A,B5`V6600>s7qlProEhtwcwg5WP06,0*24  "  12/05/2011 0:01
636091785 -70.7876316666667 173.287395 1.6 "!AIVDM,1,1,,A,37`B03kP@K<HsUA`NQfndUFp00wP,0*77  "  12/05/2011 0:01
477880800 -35.9284166666667 174.584616666667 2.8 "!AIVDM,1,1,,A,177gQp000L<O;mUcL@iQ8Q0n0<08,0*39  "    12/05/2011 0:01
477880800 -35.9284166666667 174.584616666667 2.8 "!AIVDM,1,1,,A,177gQp000L<O;mUcL@iQ8Q0n0<08,0*39  "    12/05/2011 0:01
273000000 -36.892565 174.781716666667 0 "!AIVDM,1,1,,A,144FV@0P00<P5U=bpu3@0?vur0Rd,0*7A  " 12/05/2011 0:01
273000000 -36.892565 174.781716666667 0 "!AIVDM,1,1,,A,144FV@0P00<P5U=bpu3@0?vur0Rd,0*7A  " 12/05/2011 0:01
355929000 -46.59125 168.348785 0 "!AIVDM,1,1,,B,15CL;b0000<2`sgUEiheUTRv2<01,0*53  "    12/05/2011 0:01
512000743 -36.888225 174.989618333333 0 "!AIVDM,1,1,,B,17`B2qwP00<Q2OGbq7>@0?vr0404,0*63  " 12/05/2011 0:01
512000743 -74.9598833333333 174.989618333333 0 "!AIVDM,1,1,,B,37`B1F?007<P<Rubu6cWkAPv016@,0*65  "  12/05/2011 0:01
565760000 -35.5086216666667 176.255433333333 14.5 "!AIVDM,1,1,,A,18KS@00PBA<VmEIcc`bo9Efv0404,0*2E  "   12/05/2011 0:01
512000200 -36.6396166666667 174.776428333333 102.3 "!AIVDM,1,1,,B,17`B0j0Oww<P423c2=qf49fv0HAl,0*4E  "  12/05/2011 0:01
512000200 -75.2084916666667 174.776428333333 102.3 "!AIVDM,1,1,,B,37`B1F?007<P<Rubu6cWkAPv016@,0*65  "  12/05/2011 0:01
538003456 -39.982255 173.287148333333 2.8 "!AIVDM,1,1,,A,181580000LdI?f3a7kUsma`t08AN,0*1C  "   12/05/2011 0:01
```

Figure 4.8 : Decoded information after passing through an AIS Decoder. The above screenshot shows decoded information that includes MMSI, Latitude, Longitude, Speed, Rawsentence and Timestamp.

Step Two: The OUTPUT text file of each database has to be been imported to an excel sheet in order to perform the next level task to conduct the trajectory analysis.

Step Three: The next step involves the process of data cleaning. Data cleaning is an important process to remove noise and inconsistent data in the data mining process. At last after cleaning the dataset is ready to conduct the trajectory analysis.

	A	B	C	D	E	F
1	MMSI	Latitude	Longitude	Speed	Date	Time
2	512000048	-36.93313833	174.7853483	0	13/05/2011	12:02:00 a.m.
3	512000048	-36.93313833	174.7853483	0	13/05/2011	12:02:00 a.m.
4	512000743	-36.88991167	174.9897567	0	13/05/2011	12:02:00 a.m.
5	512000743	36.88991	174.9897567	0	13/05/2011	12:02:00 a.m.
6	512000743	36.88991	174.9897567	0	13/05/2011	12:02:00 a.m.
7	512131000	-41.29431667	174.8562667	14.9	13/05/2011	12:02:00 a.m.
8	512131000	41.294315	174.8562667	14.9	13/05/2011	12:02:00 a.m.
9	512131000	41.294315	174.8562667	14.9	13/05/2011	12:02:00 a.m.
10	355929000	-46.59123667	168.3487433	0	13/05/2011	12:02:00 a.m.
11	538003725	-39.47248333	176.9207717	0.1	13/05/2011	12:02:00 a.m.
12	538003725	39.47248167	176.9207717	0.1	13/05/2011	12:02:00 a.m.
13	538003725	39.47248167	176.9207717	0.1	13/05/2011	12:02:00 a.m.
14	538003725	-72.375625	176.9207717	0.1	13/05/2011	12:02:00 a.m.
15	538003725	-39.47248333	176.9207717	0.1	13/05/2011	12:02:00 a.m.
16	512000050	-40.776	174.3376667	12.2	13/05/2011	12:02:00 a.m.
17	512000200	-36.66698167	174.8523017	7.2	13/05/2011	12:02:00 a.m.
18	512000200	-36.66698167	174.8523017	7.2	13/05/2011	12:02:00 a.m.
19	512000048	-36.93313833	174.7853483	0	13/05/2011	12:02:00 a.m.
20	512000048	36.93313667	174.7853483	0	13/05/2011	12:02:00 a.m.
21	512000048	-36.93313833	174.7853483	0	13/05/2011	12:02:00 a.m.
22	512219000	-35.83556667	174.49635	0	13/05/2011	12:02:00 a.m.
23	512219000	35.835565	174.49635	0	13/05/2011	12:02:00 a.m.
24	512000322	-41.28136333	174.7814	0	13/05/2011	12:02:00 a.m.
25	512000742	-36.88977167	174.9896383	0	13/05/2011	12:02:00 a.m.
26	512000742	36.88977	174.9896383	0	13/05/2011	12:02:00 a.m.
27	512000742	36.88977	174.9896383	0	13/05/2011	12:02:00 a.m.

Figure 4. 9: Illustration of an imported Excel file with MMSI, Latitude, Longitude, Speed, Date and Time attributes.

The Decoded raw data output is collected in text file format. The option of collecting decoded information has the advantage that it could be exported to other file formats and interoperability. It is observed to have flexibility running in other applications of ArcGIS components such as ArcMap and ArcGlobe.

The raw data collected in text file has to be exported to excel file to perform necessary statistical calculations.

4.8 Summary

The dataset provided by Kordia New Zealand was reviewed. The challenges encountered during the opening a large dataset was discussed. The problem encountered in handling large sized dataset is discussed and approach taken towards was explained in detail. The rationale to select small subsets of data from a large dataset was discussed. A tool (decoder) has developed to decode the raw data using VB.net was discussed.

A decoder was developed to perform the following:

1. To convert from raw data to 6 bit binary values.
2. To convert decimal values to corresponding bit positional values.
3. The converted binary values were associated to corresponding bit values following ITU-R M.1371-1 guidelines.
4. The Maritime Mobile Service Identity (MMSI) was extracted.
5. The Latitude and Longitude positional values were extracted.
6. The Speed over Ground (SOG) data of a vessel were extracted.

After decoding the required parameters from the raw data, the next step in filtering processes were presented in detail. The cleaned dataset will be used to perform the trajectory analysis is described and evaluated in Chapter 5 using ArcGIS tool.

5 Trajectory Analysis (Phase 2)

This chapter presents a detailed description of the development of the methods used in the knowledge discovery from movement data relating to moving objects. The main motivation of Chapter 5 is to conduct spatial-temporal analysis using an ArcGIS tool. There are four main steps involved in this chapter. The Section 5.1.1 was to introduce the two clustering techniques and discuss a suitable approach to this study. The first step is a discussion on the selection of a GIS tool to perform the trajectory analysis of a moving vessel. The second step involves a thorough discussion on the analysis conducted the plotted data from the decoded data set (discussed in Chapter 4) using ArcGIS and a trend graph which illustrates speed versus time. As well, a spatio-temporal approach using the selected clustering techniques is also discussed in detail. The third step, investigates the direction and speed of the vessels. In this step the frequency distribution of the speed of a particular vessel type is selected and thoroughly investigated using a speed profile. In the final step, a study is conducted on those vessels which are anchored (stationary) using a clustering technique based on a speed parameter. The purpose of this speed parameter is to detect the overall amount of time that vessels are anchored in one position.

An evaluation will be performed on the frequency distribution created using speed parameter and stationary locations. This evaluation process helps to identify the interesting patterns representing through the knowledge discovery process. This process is also identifies on characteristics of vessel type produced by moving vessels.

The results will be presented after conducting rigorous analysis. The analysis was conducted using various approaches such as the spatio-temporal approach, clustering techniques based on **Stops** and **Moves** of trajectories, analysing the trends of speed versus time, investigating speed profile and speed based clusters. This analysis is considered to be a key part of the study. In addition this chapter aims to highlight the most pertinent outcome of this thesis and to a range of options for future research.

5.1 Methods and techniques applied in conducting trajectory analysis

In order to perform knowledge discovery on exacted data to identify vessel type characteristics and patterns ***Clustering Based Stops and Moves of Trajectories*** (CB-SMoT) methods was applied. A thorough investigation is conducted and the rationale for selecting **CB-SMoT** against **Intersection based Stops and Moves of Trajectories** (IB-SMoT) is discussed in the next following Section 5.1.1.

5.1.1 Two facets of trajectory analysis

According to Spaccapietra et al., (2008) trajectories are composed of two facets, namely:

Geometric facet: A single continuous subset of the path covered by the moving object's position during the entire lifespan of the object with spatial reference along with time. It can be expressed as $[t_{begin}, t_{end}] \rightarrow$ space.

Semantic facet: This information conveys the meaning of the trajectory exclusively to the application domain. The semantic facet is made up of different components. The different components of semantics are:

Stops, Moves, Begin and End

Stops: A vessel, ship or boat that is brought to a halt in a specific location for a certain period of time.

Moves: A vessel, ship or boat that starts its motion from a stationary position.

Begin and End: Vessel owners and companies are interested in monitoring and managing their vessels during each trip, for example the vessels starting location and destination. Trajectories were marked with a beginning and end respectively indicating the start and destination locations.

To conduct analysis on sample points becomes more complex with very little or no semantics. Applying the approach of **Stops** and **Moves** (Spaccapietra et al., 2008) to trajectories, therefore adds meaningful and interesting semantics to the decoded data. Hence this could make the process of interpretation much simpler to handle during the process of interpretation.

In general, stops are considered the most important part of a trajectory analysis, while moves are the movements of the vessel between each stop. This approach has supported various application domains such as bird migration (Laube et al. 2007) and traffic management. Traffic management studies are very similar to the vessel trajectory studies where stops can be traffic lights, a speed camera or controllers, passing roundabouts and other urban traffic control factors.

Two different methods have been discussed and developed are instantiated towards “stops” and “moves” movement approach. They are **Intersection based Stops and Moves of Trajectories** (IB-SMoT) (Alvares, et al., 2007) and Clustering Based Stops and Moves of Trajectories (CB-SMoT) (Palma, Bogorny, Kuijpers, & Alvares, 2008). The first method, Intersection based Stops and Moves of Trajectories (IB-SMoT) is basically stops that have been generated based on the intersection of the trajectory sample points with geographic objects at a specific geographic location. The "stop" should take minimum threshold of time. For example, recreational activities on a city tour.

Another method ***Clustering Based Stops and Moves of Trajectories*** (CB-SMoT) is a clustering method entirely based on the **speed** variation of the trajectory. This method seems to be more appropriately applied on fishing boats, ferries and high speed crafts. The CB-SMoT method mainly based on generating clusters where speed is much lower than a given threshold, for instance speed = 0 means the particular vessel is anchored. Generating clusters based on a speed parameter is good as it provides the characteristics of a vessel or boat. In addition, the method is complemented with additional information such as the pattern detection and activity based profile. According to Rocha et al (2010) and Palma (2008), the CBSMoT method is remarkably useful when speed plays an essential role.

In order to study Stops and Moves of trajectories, a sophisticated tool is necessary to answer spatial locations and temporal details (t) such as when the vessel was seen at a particular location (x, y).

5.1.2 Display of spatio-temporal data

Obviously, the decoded data set is comprised of spatial information and temporal information, which is collectively known as Spatio-temporal data. Spatio-temporal data involves three main attributes – latitude, longitude represented as (x, y) and time (t).

Spatio-temporal data represents the three basic attributes of an entity or a real world event. An entity includes the existence, place and nature with dimensions of time, location in terms of a coordinate system.

- The time dimension ***t*** represents the existent time of an entity or event '***e'***. These data answer the question of WHEN does an event happen?
- The location dimension of ***(x, y)*** represents the place of an entity or event. This defines the coordinates of where the object was. These data answer the question of WHERE?
- The theme dimension describes the nature of an entity or event. These data answer the questions of WHAT?

Using the above mentioned basic attributes, different questions about movement data can be formulated.

Table 5.1: Question Formulation based on movement data in respect of time and location using spatial and temporal attributes (Adapted from N. Andrienko, Andrienko, Pelekis, & Spaccapietra, 2008)

General Questions	Category	Location	
		Elementary	Synoptic
Time	Elementary	Where was an particular entity '***e'*** at time ***t*** moment?	What was the distribution (spatial) of all entities at moment time ***t***?
	Synoptic	How did entity ***e*** move during the time period -	How did all entities move during the time

		t1 to ***t2***?	period from ***t1*** to ***t2***?
Complex Questions	Category	Location	
		Elementary	Synoptic
Time	Elementary	How did an entity e change the location from time t1 to time t2?	How did entities e1 and e2 in the time period from t1 to t2 (in terms of movement pattern)?
	Synoptic	What is the difference in times when an entity ***e*** visited places ***p1*** and ***p2***?	How does the movement pattern of entity ***e*** in time interval ***t1*** to ***t2*** differ from the movement behavior in time interval from ***t3*** to ***t4***?

The main objective of using spatio-temporal attributes is to get a better insight as to the behaviour of the moving object movement details. The decoded data set has spatio-temporal data and it is essential to have a powerful tool to analyse such datasets.

5.2 Tool used to conduct the analysis

5.1.2 ArcGIS

According to ESRI (2010) ArcGIS is a "computer based system" that stores geographically referenced data. This allows for the application of a wide range of information processing that includes manipulation, analysis and modeling. ArcGIS 10 also provides for map display and production. This tool can be used for computerized mapping and spatial analysis. Further, ArcGIS provides functionality to capture, store, query, analyze, display, and output geographic information. This information about the earth's surface are then disseminated.

After studying the features of ArcGIS tool it becomes evident that this tool could be ideal to display information of a vessel's lifeline or geometry of its movement paths. This tool can be used to conduct spatio-temporal analysis. A detail features of GIS data types was given in the **Appendix C.**

5.1.2.1 ArcMap

ArcGIS's mapping application is ArcMap which allows one to create maps, query attributes, analyze relationships spatially and temporally. ArcMap uses ArcCatalog, an application mainly involved in organizing spatial data contained in one's personal computer or various other locations. In other words, it is briefly described as a data management application. ArcCatalog also enables the search, previewing and adding of data to ArcMap. It provides further assistance in managing metadata and setting up address based services called geocoding.

5.1.2.2 ArcGlobe

ArcGlobe is a 3D visualization application. This allows large amounts of geographic information system data to be projected on a globe's surface. The ArcGlobe environment provides a unique way to view and analyze displayed GIS data. Data that is spatially referenced is displayed on a 3D globe surface in its true geodetic locations. Furthermore with true representation of geodetic references, one can manipulate the globe to investigate and analyze its data. This manipulation can involve viewing the globe as a whole or in regions. Another advantage of ArcGlobe is that it provides viewing data covering a global extent or zooming seamlessly to high detail. The strength of ArcGlobe is its power to handle voluminous data while performing analysis. This allows efficient displaying and querying of raster data which is integrated into the ArcGIS geodatabase (3D Analyst and ArcGlobe, 2012).

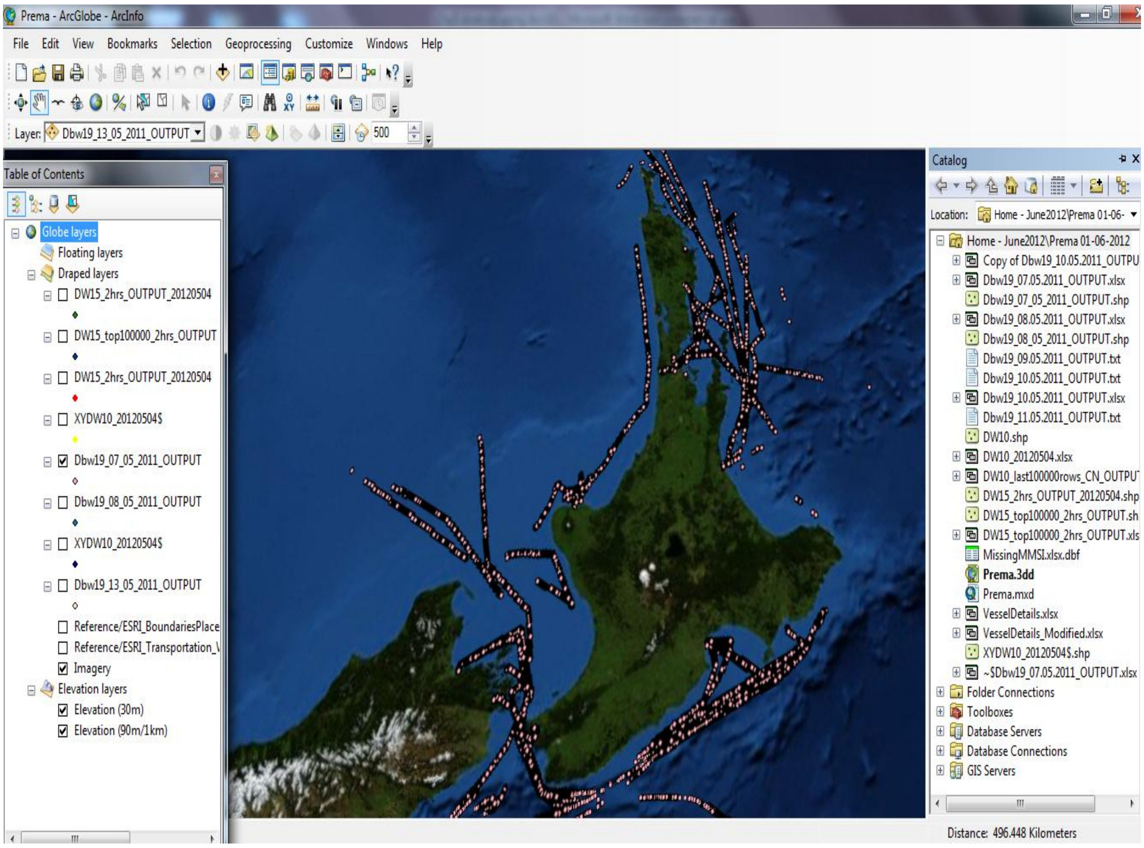


Figure 5.1 : ArcGlobe displays the trajectory of vessel movements dated 7/05/2011

Using ArcGIS's tools ArcMap and ArcGlobe, the trajectory analysis will be conducted on specific types of vessels.

5.3 Types of Vessels

As the case study is about moving vessels, it is important to understand the different types of vessels used in maritime activity. According to the International Maritime Organization (IMO), the ship and vessel types are classified as follows:

- **Passenger ship:** A passenger ship – with the capacity of carrying more than 12 passengers (SOLAS I/2).
- **Fishing Vessel:** A fishing vessel – commercially utilized to catch fish, whales, seals and other living resources of the sea **(**SOLAS I/2).
- **Nuclear ship:** A **nuclear ship** exclusively provided with a nuclear power plant (SOLAS I/2).

- **Bulk carrier** (**SOLAS IX/1.6**) a ship constructed to serve various purposes. Generally with a single deck, on topside tanks and with hopper side tanks in cargo space. This is primarily intended to carry dry cargo in bulk and includes different types such as ore carriers and combination carriers.
- **Bulk carrier (SOLAS XII/1.1)** which primarily carries dry cargo in bulk. This includes such types as ore and combination carriers.
- **Oil tanker** (SOLAS 74) a ship primarily constructed to carry oil in bulk in its cargo spaces. This includes combination carries called "NLS tanker" and any gas carrier.
- **Cargo ship** – a ship primarily designed for the carriage of general cargo with a multi-deck or single-deck hull.
- **High-speed craft** (SOLAS X/1.2):- A high speed water vessel for civilian use. These high speed crafts are also popularly called Fastcraft or Fast ferry. Mostly these craft serves as passenger ferries. The larger crafts are also known as catamarans can carry cars and other vehicles. High speed crafts are capable of a maximum speed in meters per second (m/s) equal to or exceeding 3 times.
- **Mobile offshore drilling unit** (MODU) a vessel capable of drilling operations such as for the exploration or exploitation of resources underneath the seabed. Such resources include liquid or gaseous hydrocarbons, sulphur or salt (SOLAS IX/1).
- **Special purpose ship** (SPS) means a mechanically self-propelled ship which by reason of its function carries on board more than 12 special personnel (IMO, 2012).

Other than the above mentioned vessel types there are various small to medium sized boats that are included in the data set that Kordia New Zealand provided.

- **Tug:** A boat that maneuvers vessels by towing, stopping or disabled vessels. These tugs are mainly used in a crowded harbor or in a narrow canal when the vessels were not able to move by themselves, such as disabled ships or oil platforms. As these boats are built for special purposes, they are powerful for their size and strongly built. Some tug boats are used as ice breakers and sometimes as salvage boats.

- **Anchor handling vessel:** This vessel is mainly used to handle anchors for oil rigs. In addition, it can act as a tower at times, tow oil rigs to particular locations and assist them in anchoring. At times these vessels act as Emergency Rescue and Recovery Vessels (ERRV).

- **Trawler:** A trawler is basically used for commercial fishing and is designed to operate as a fishing trawler. The name is coined out of the popular method of fishing called **trawling**. Basically, trawling refers to fishing vessel designed to use trawl nets that are used to catch large volumes of fish.

- **Pleasure Craft:** A pleasure craft or pleasure boat is mainly used for personal, family activity and occasionally is used as a recreational boat.

- **Pilot Vessel:** This kind of boat is particularly used to transport maritime pilots between land and inbound or outbound ships.

- **Naval/naval auxiliary vessel:** These vessels are designed basically to support a number of roles for the navy such as replenishment, transport, repair, research and other naval operations. These vessels are also used in support roles in harbours. This is considered as another critical support role.

- **Ro-ro/passenger ship:** Ro-ro is also known as Roll-on/roll-off. These ships are vessels that are exclusively used to carry wheeled cargo supporting the transportation of cars and trucks across oceans.

- **Research/survey vessel:** These kinds of vessels are used for mapping and research purposes.

5.4 Plotting trajectories via ArcGIS

In order to implement the system design in ArcGIS 10, the trajectory data model proposed in Chapter 2 geo- database is used. The geo database data model which is comprehensively designed to support geo referential data along with time, is used to store and manage all data sets. This is to emphasize on spatio-temporal approach 2D space (longitude and latitude) with an orthogonal third dimension of time (1D). A 3D feature displaying spatial coordinates along with time is useful and, ArcGlobe facilitates visualization to understand plotted trajectories.

5.4.1 Data Export

One major challenge to represent and visualize spatio-temporal features in ArcGIS 10 is the time attribute. Time values of a spatio-temporal feature must have a similar format as other values in the 3D feature. It is essential to maintain the exact similar format while exports time attribute into different file(s) formats without exploiting its data format. This is to be certain that changes between the format of a time value in the real world and the format in a computer system are compatible. Accordingly on this based defined time representation, the implementation of spatio-temporal feature representation established from an Excel to geodatabase.

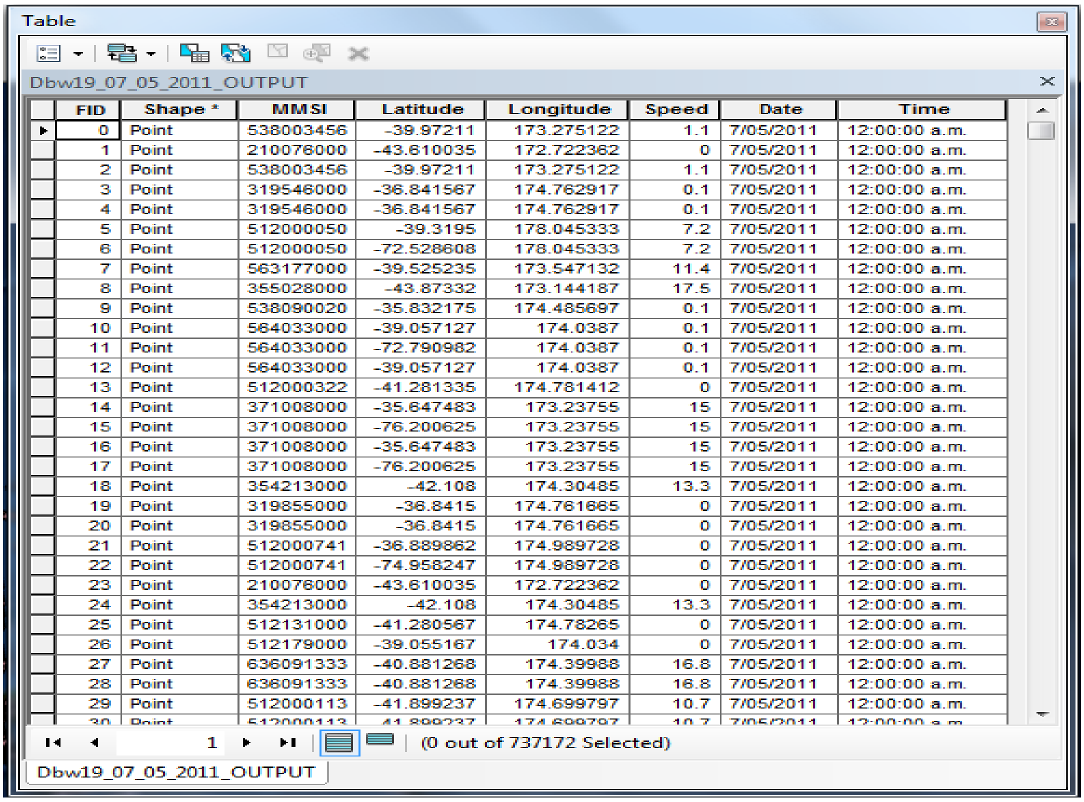
5.4.2 Save as .SHP files

The extension in ArcGIS allows a user to export coordinates in Excel and convert to a format to save as a Shape file to perform further investigations in ArcGIS.

Table 5.2: Datatypes used in ArcGIS tool to conduct trajectory analysis (analytical study).

Field Name	Description	Datatype
FID	The unique indicator index for sample points that starts	ObjectID

	from 1.	
Shape	The geometry type of spatial feature in this geographic database table. This uses a **point** data type since the AIS sample data are of discrete forms.	Point
MMSI	Maritime Mobile Service Identity - a unique identification number of a vessel.	Double
Latitude	X coordinate value of the vessel at a specific time	Double
Longitude	Y coordinate value of the vessel at a specific time	Double
Speed	Speed of the moving vessel at a specific location of a specific time and expressed in **Knots.**	Double
Date	Data collection date from AIS	Date
Time	Data collection time from AIS	Time



Table

Dbw19_07_05_2011_OUTPUT

FID	Shape *	MMSI	Latitude	Longitude	Speed	Date	Time
0	Point	538003456	-39.97211	173.275122	1.1	7/05/2011	12:00:00 a.m.
1	Point	210076000	-43.610035	172.722362	0	7/05/2011	12:00:00 a.m.
2	Point	538003456	-39.97211	173.275122	1.1	7/05/2011	12:00:00 a.m.
3	Point	319546000	-36.841567	174.762917	0.1	7/05/2011	12:00:00 a.m.
4	Point	319546000	-36.841567	174.762917	0.1	7/05/2011	12:00:00 a.m.
5	Point	512000050	-39.3195	178.045333	7.2	7/05/2011	12:00:00 a.m.
6	Point	512000050	-72.528608	178.045333	7.2	7/05/2011	12:00:00 a.m.
7	Point	563177000	-39.525235	173.547132	11.4	7/05/2011	12:00:00 a.m.
8	Point	355028000	-43.87332	173.144187	17.5	7/05/2011	12:00:00 a.m.
9	Point	538090020	-35.832175	174.485697	0.1	7/05/2011	12:00:00 a.m.
10	Point	564033000	-39.057127	174.0387	0.1	7/05/2011	12:00:00 a.m.
11	Point	564033000	-72.790982	174.0387	0.1	7/05/2011	12:00:00 a.m.
12	Point	564033000	-39.057127	174.0387	0.1	7/05/2011	12:00:00 a.m.
13	Point	512000322	-41.281335	174.781412	0	7/05/2011	12:00:00 a.m.
14	Point	371008000	-35.647483	173.23755	15	7/05/2011	12:00:00 a.m.
15	Point	371008000	-76.200625	173.23755	15	7/05/2011	12:00:00 a.m.
16	Point	371008000	-35.647483	173.23755	15	7/05/2011	12:00:00 a.m.
17	Point	371008000	-76.200625	173.23755	15	7/05/2011	12:00:00 a.m.
18	Point	354213000	-42.108	174.30485	13.3	7/05/2011	12:00:00 a.m.
19	Point	319855000	-36.8415	174.761665	0	7/05/2011	12:00:00 a.m.
20	Point	319855000	-36.8415	174.761665	0	7/05/2011	12:00:00 a.m.
21	Point	512000741	-36.889862	174.989728	0	7/05/2011	12:00:00 a.m.
22	Point	512000741	-74.958247	174.989728	0	7/05/2011	12:00:00 a.m.
23	Point	210076000	-43.610035	172.722362	0	7/05/2011	12:00:00 a.m.
24	Point	354213000	-42.108	174.30485	13.3	7/05/2011	12:00:00 a.m.
25	Point	512131000	-41.280567	174.78265	0	7/05/2011	12:00:00 a.m.
26	Point	512179000	-39.055167	174.034	0	7/05/2011	12:00:00 a.m.
27	Point	636091333	-40.881268	174.39988	16.8	7/05/2011	12:00:00 a.m.
28	Point	636091333	-40.881268	174.39988	16.8	7/05/2011	12:00:00 a.m.
29	Point	512000113	-41.899237	174.699797	10.7	7/05/2011	12:00:00 a.m.

1 (0 out of 737172 Selected)

Dbw19_07_05_2011_OUTPUT

Figure 5.2: An example of a geodatabase table which shows the demonstrates fields with time attribute.

5.4.3 Spatio-temporal Trajectories

As earlier discussed in Chapter 2 regarding a definition of Trajectory and its analysis, this chapter revisits that definition in with more analytical depth through in ArcGIS. The original meaning of a trajectory denotes the changing position of a moving object in a 2D geographical space along with time as in a 3D space (eg., the trajectory of a plane, vessel, boat or vehicle). Such data used is described as spatio-temporal trajectories if spatial coordinates are used to express the position of the travelling object (Spaccapietra, et al., 2008). Most often, the moving object is geometrically represented as a point. For example, this could be a person, an animal, a car, a ship or a train. Some moving object may have a surface or volume geometry in which both aspects of change in position and in shape and may correspond to the trajectory. This thesis study will consider analyse trajectories using spatio-temporal data generated by moving object represented as points. The moving objects selected to conduct analysis were 9 different types of vessels. They were fishing vessels, Naval/Research vessel, High speed craft, Ro-ro cargo vessels, Pleasure boats, Pilot vessels, Ferry vessels, Passenger boats, Ferry boats and Container vessels.

5.4.4 Steps involved on plotting trajectories on ArcGIS

Once an appropriate method of trajectory analysis is selected, the next step is to export data to the ArcGIS tool for analysis. Earlier in Chapter 4, a detailed discussion described and evaluated the conversion of raw data to decoded data. Decoded data has undergone a series of steps within in the filtering process. Once the data set is filtered and cleaned, it is now ready for analysis. This section will discuss in detail the steps involved in the conversion of the Excel file format to .shape (.SHP) format to support analysis in ArcGIS.The following steps describe the process of converting into a .SHP file and plotting the vessel activties on ArcMap.

Step 1: Run ArcMap and open a document

Step 2: Click to add Data button and select the required Excel file to perform analysis.

Step 3: Conversion of Excel file to .shp format

Step 3.1: Select specific a file to be converted and right click “select option XY coordinates".

Step 3.2: A dialog box prompts the user to specify the fields for the X and Y coordinates. Select X coordinates for latitude and Y coordinates for longitude accordingly.

Step 3.3: A spatial reference is selected to describe the Coordinate System. The coordinate system selected is GCS_WGS_1984.

Step 3.4: Select OUTPUT path of the file to be saved.

Step 4: Once the .SHP file is ready, drag and drop on to the ArcMap display area.

Step 5: Save the changes applied to the file.

Following the above mentioned steps, the overview of the trajectories projected on ArcGIS is shown in Figure 5.3 and Figure 5.4.

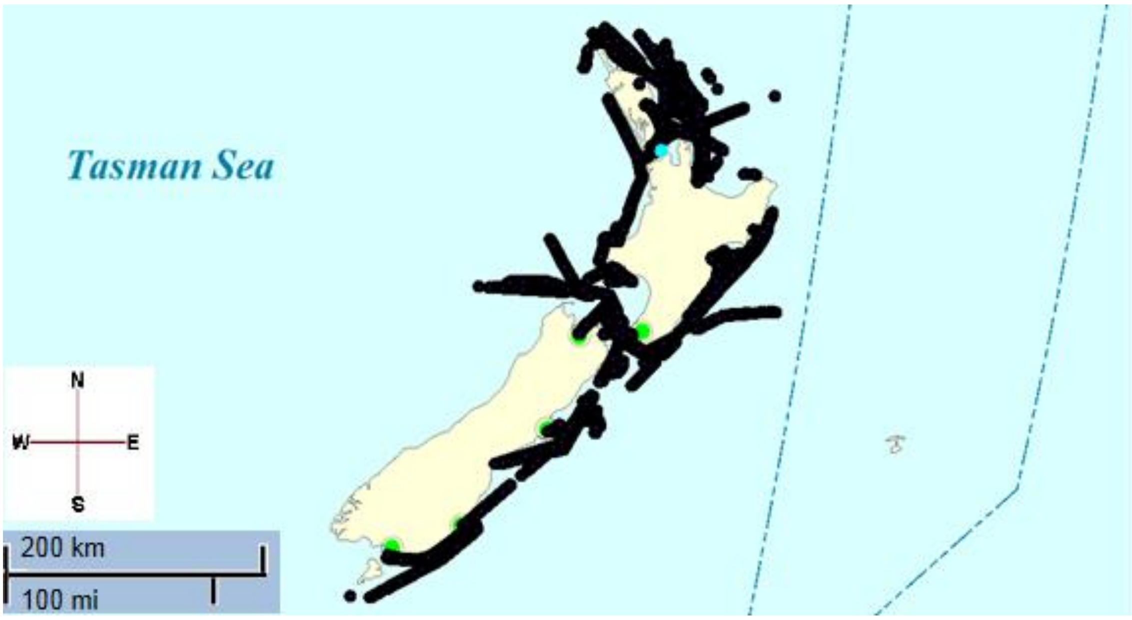


Figure 5.3: Overview of trajectories of vessels displayed using ArcMap - dated 7/05/2011 (North Island) between 00.02 and 23.59 (24 Hours)

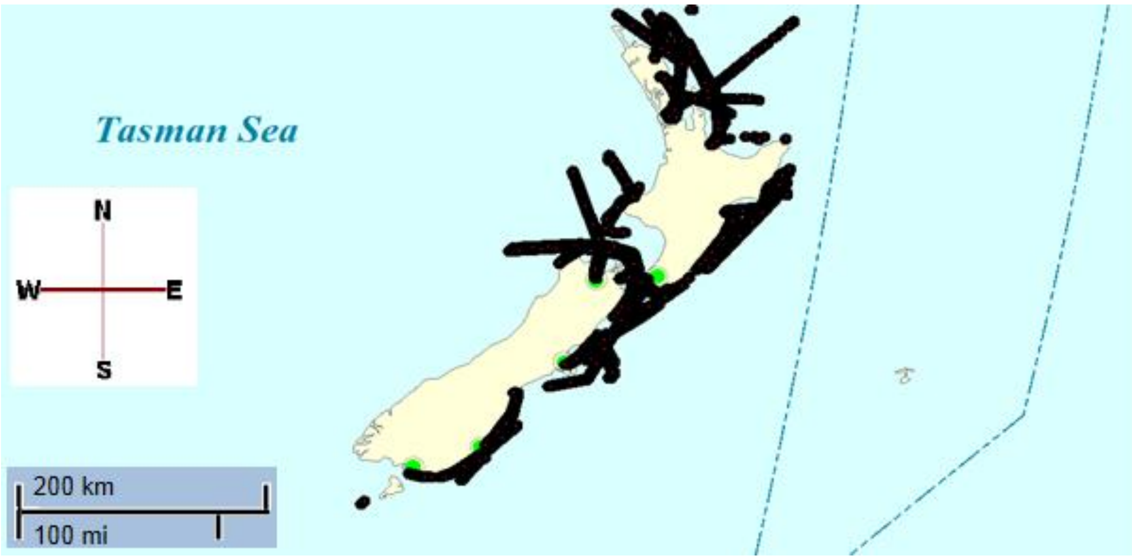


Figure 5.4 : Overview of trajectories of vessels displayed using ArcMap - dated 8/05/2011 (South Island) between 00.04 and 23.59 (24 Hours).

An overview of the trajectory data projected dates 7/05/2011 and 8/05/2011 is now presented. These dates fall on weekends, Saturday and Sunday. The Figure 5.4 and Figure 5.3 represent the trajectory dated 7/05/2011 and 8/05/2011 respectively. This is to conduct a comparative study between two days about the trajectories projected by different types of moving vessels.

Table 5.3 : Details on the number of vessels and total number of records generated for each date.

Date	Number of ships/vessels/boats	Number of points/ rows (after cleaning)
7/05/2011	80	737172
8/05/2011	85	778185
9/05/2011	81	730143
10/05/2011	95	873388
12/05/2011	82	740144
13/05/2011	86	735463
9/04/2011 12am – 2am	52	64263

(2 hours)		
11/03/2011 21.30 – 23.59 hours (2 hours)	56	77481

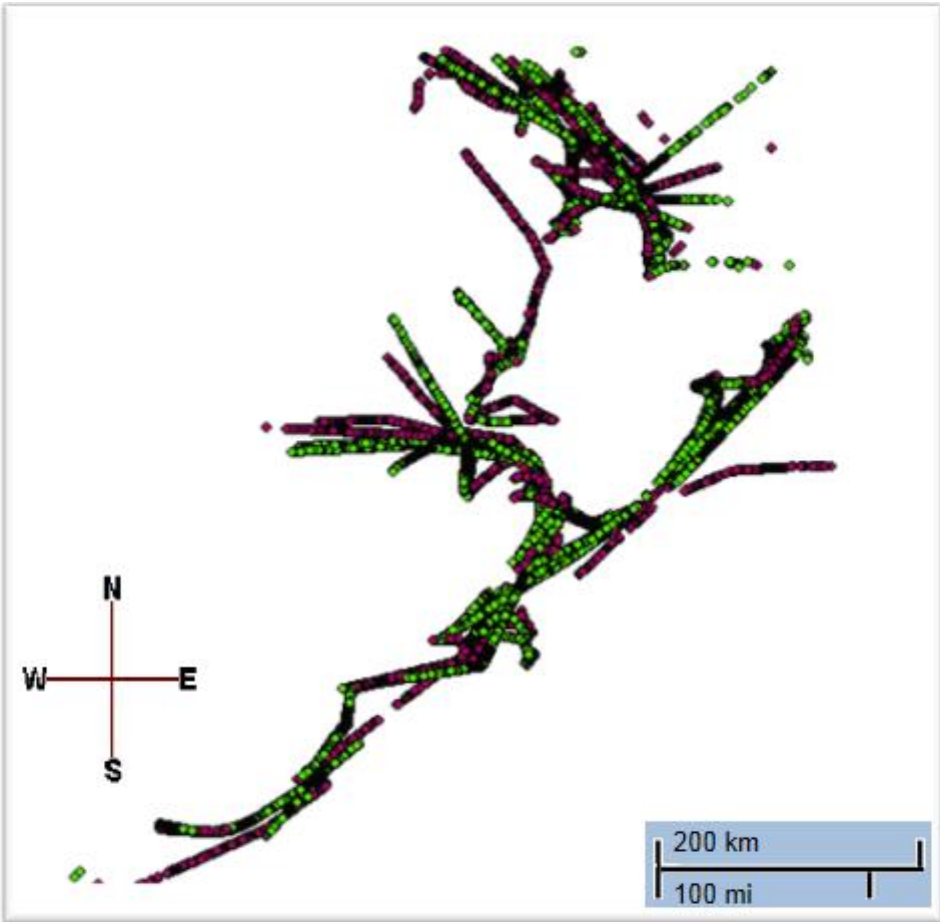


Figure 5.5: Illustration using ArcMap - North Island with green shaded trajectory dated 7/05/2011 (Saturday) followed by pink shaded trajectory dated 8/05/2011 (Sunday).

Based on details of Table 5.3, the trajectory points were directly proportional to the number of vessels. The Figure 5.5 illustrates, the routes used by the vessels in these two days (7/05/2011 and 8/05/2011) were similar. The trajectory of the vessels in these two days was comparatively very similar. It is also noticed that fishing boats has taken other than the usual route to carry out fishing.

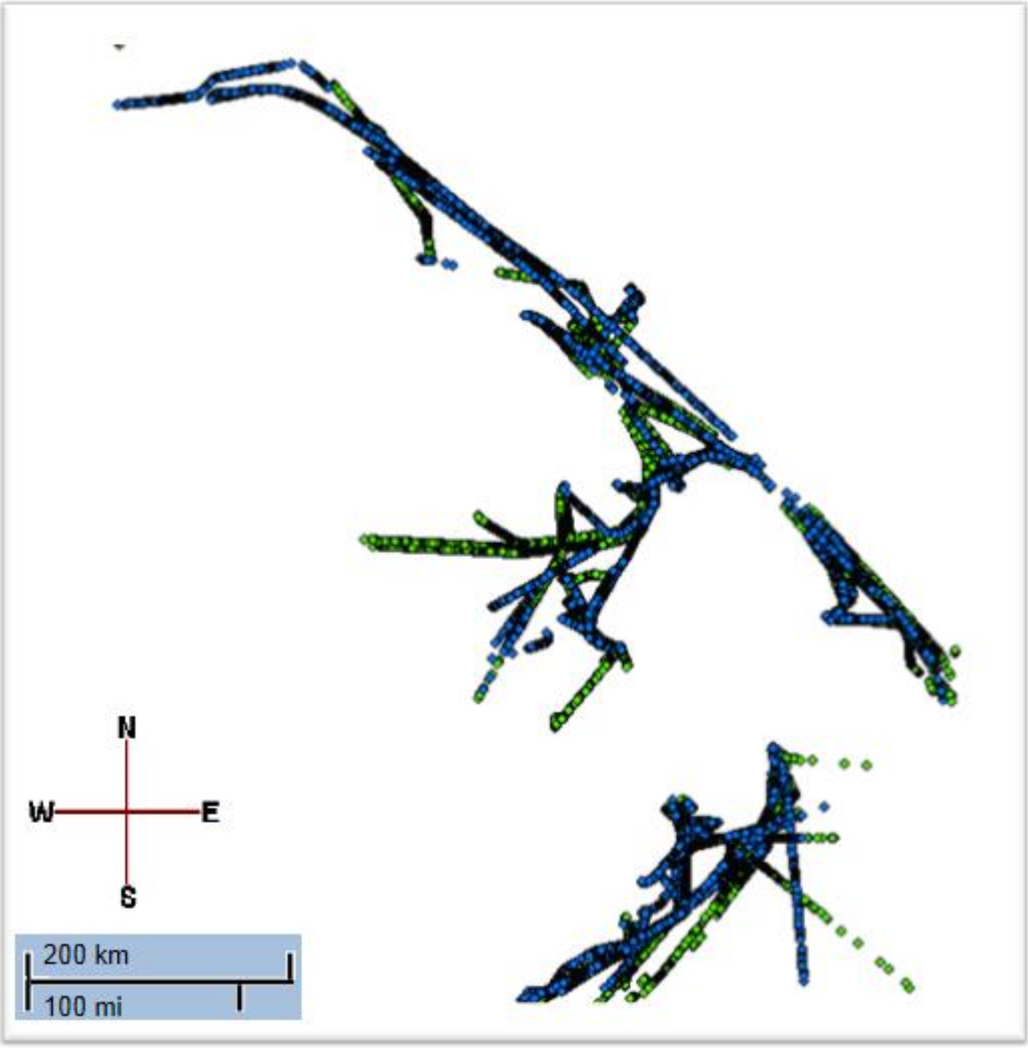


Figure 5.6: ArcMap illustrating South Island with blue shaded trajectory dated 7/05/2011 (Saturday) followed by green shaded trajectory dated 8/05/2011 (Sunday).

The Figure 5.6 illustrates, in South Island, trajectory density is much lesser compared to North Island (Figure 5.5). The shaded blue trajectory indicates trajectory formed by different type of vessels dated 7/05/2011 whereas, green shaded trajectory dated 8/05/2011 (Sunday). The density and number of trajectories show that Sunday with less number of vessel movements compared to 7/05/2011 (Saturday).

5.5 Summary Velocity and Movement Data

A summary of data collected from one large dataset (DW19) was used to identify some characteristics of the velocity profiles.

Table 5.4 : Movement and number of vessel types

Vessel Type	Percentage of readings where vessel is moving	Number of Vessels of this type in this dataset
Anchor handling vessel	98.95%	2
Bulk carrier	89.26%	22
Cargo Ship	96.08%	8
Cement carrier	50.55%	3
Container ship	88.11%	20
Drill Ship	82.32%	1
Ferry	41.22%	1
Fish factory	6.96%	2
High Speed Craft	52.53%	3
Law Enforcement Vessel	93.81%	1
LPG tanker	99.04%	2
Naval/naval auxiliary vessel	95.87%	7
Oil products tanker	36.78%	3
Oil/chemical tanker	99.99%	7
Passenger ship	99.95%	2
Pilot Vessel	54.74%	2
Reefer	99.49%	2
Research/survey vessel	85.83%	3
Ro-ro cargo	85.82%	1
Ro-ro/passenger ship	55.40%	2

Sailing Vessel	20.52%	2
Suction hopper dredger	49.64%	1
Trawler	84.84%	9
Tug	35.74%	5
Vehicle carrier	93.67%	1

Data was combined for each vessel of a type. In terms of the velocity profile, the mean and standard deviation of the velocity measured over each vessel type collection was also calculated.

Table 5.5: Mean Velocities of Different vessel classes in DW19 Dataset

Mean Velocity (Knots)	Vessel Type	Number of records	Std Deviation of Velocity
24.91	High Speed Craft	32444	8.74
20.19	Naval/naval auxiliary vessel	9061	29.48
19.39	Law enforcement	13267	13.93
18.76	Pilot Vessel	26009	8.43
17.31	Vehicle carrier	12455	1.72
17.04	Ro-ro/passenger ship	17367	4.34
17.00	Reefer	12701	3.28
16.68	Ferry	3918	4.23
15.36	Ro-ro cargo	21347	3.54
15.09	Container ship	177424	4.55
13.38	Cement carrier	21465	1.86
12.87	Oil/chemical tanker	54299	2.91
11.88	Cargo Ship	31049	3.59
11.10	Passenger ship	13826	3.77
10.30	Oil products tanker	12111	3.94
10.19	Bulk carrier	53960	5.25

8.70	Research/survey vessel	8308	1.95
8.47	Pleasure Craft	4633	0.35
5.56	Suction hopper dredger	1376	2.37
5.29	Tug	12536	3.06
5.26	LPG tanker	3193	4.50
4.72	Anchor handling vessel	19849	5.14
4.58	Trawler	32105	2.35
2.70	Sailing Vessel	6060	4.35
2.52	Tug/supply vessel	15271	4.49

By itself this data does not give a great insight, but it appears to confirm that the data recovery has been successful to this point, by confirming common-sense assumptions, such as the notion that a Sailing vessel is slower than a High Speed vessel for example.

5.6 Analysis based on the Spatio-temporal Approach

5.6.1 A. Aggregation

Aggregation is a method dealing with large amounts of data, when it can potentially become difficult to perform an investigation in full detail. Aggregation is also a way to distill large amounts of data out of fine-detail 'noise'. In aggregation, analysing speed data about individual vessels based on movements or events are transformed into a statistical summary which details such characteristics as minimum, maximum, average, median, sum and count. To perform spatial and temporal aggregation for elementary analysis with respect to the set of moving vessels is out of scope of this study.

According to Andrienko and Andrienko (2010), there are three main approaches which could be used for the aggregation of movement data.

Table 5. 6: Three main approaches for the aggregation of movement data (Adapted from Andrienko and Andrienko, 2010).

	Approach	Description	Other Similar Studies
#1	Spatial (S-), temporal (T-) or attribute (A-) proximity	Trajectories are divided into segments- viewed as a set of discrete movement events i.e., geographic locations with respective timestamps. **(S · T- or S · T)**	Slingsby et al. (2008), Zhao et al. (2008)
#2	Trajectory-based: a route-based (R) aggregation, often performed by clustering	Trajectories are aggregated in their entirety based either on their similarity in geographic, temporal or attribute space (or a combination of both)	Rinzivillo et al. (2008) Wilson (2008)
#3	Based on origin and destination S · S or S · S · T · T	The movement is viewed as a vector between the two locations, thus ignoring the route between origin and destination spatial locations.	Guo (2007); Andrienko and Andrienko (2010)

The following section describes how spatial and temporal aggregation of movement data is undertaken as a part of the analysis. The technique used in the next section is trajectory-based: a route-based as well as on origin and destination events were considered along with Stops and Moves approach.

5.5.1 B. Temporal Analysis

Temporal analysis along with spatial analysis provides better insights on any application domain. Simple temporal mapping is built. The map is now time aware and set by a “time-slider” control.

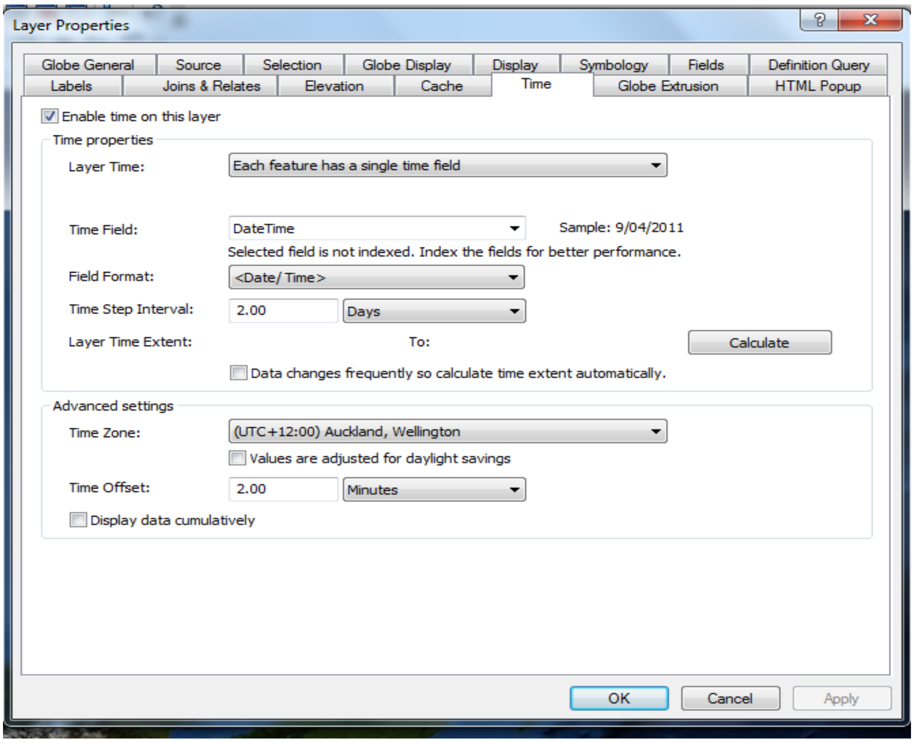


Figure 5.7: Temporal Analysis performed on hourly dataset dated 9/04/2011.

Figure 5.8(a)

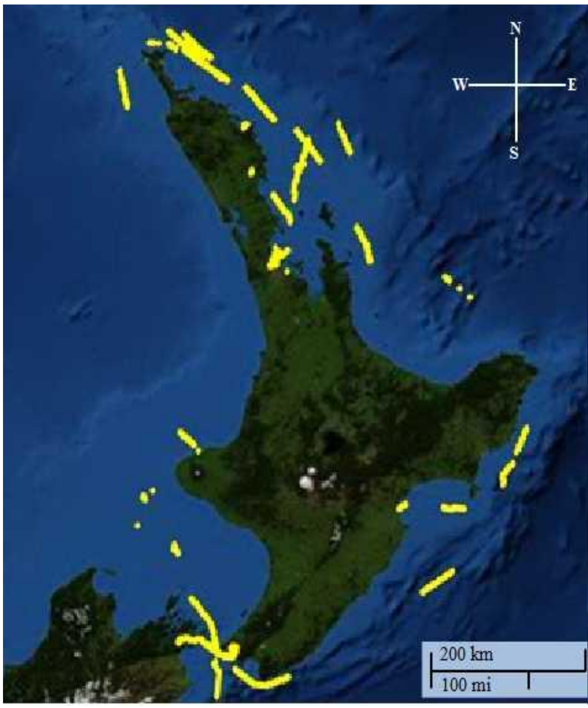


Figure 5.8(b)

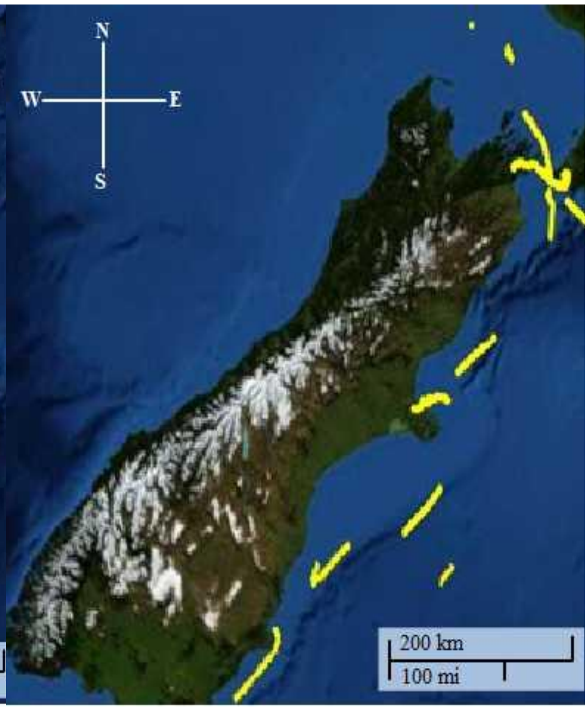


Figure 5.8: (a) Outline of temporal analysis demonstrating trajectories around North Island, dated 9/04/2011, time between 12am and 2am. (b) Temporal analysis performed around South Island, dated 9/04/2011 between 12am and 2am.

5.6.2 Spatial Analysis – Describing velocity plots

As overall trajectories of daily vessel movements were plotted on ArcMap or ArcGlobe, the next step was to conduct an analysis on each trajectory produced by each vessel based on the speed parameter. This is an attempt to answer the research question, "what are the constituting parameter(s) of movement data essential for defining trajectory patterns"? It is an attempt to find out that the additional parameter does provide more and enrich information towards the knowledge discovery process. In particular, selecting various vessel types discussed in Section 5.6 is taken into consideration for an analysis of speed over time.

To enhance the selection of a particular vessel or specific vessel type, a query generator built-in in ArcMap or ArcGlobe will assists in opening the geo-database. The query generator executes a query using SQL commands with an user-friendly interface. An overview of different methods and approach applied in this research study in listed in the Table 6.3.

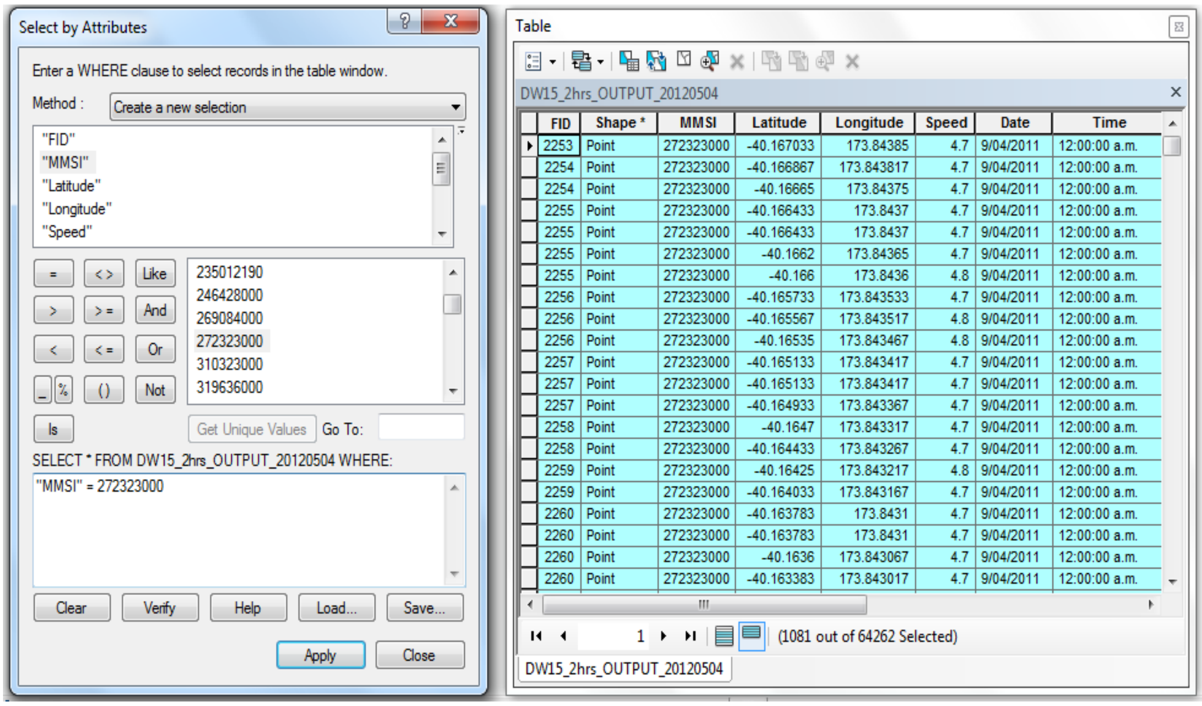


Figure 5.9: Screenshot illustrating the selection of a specific fishing trawler vessel (MMSI – 272323000).

5.5.2.1 Fishing Vessel

An investigation was carried out on the fishing boat conducted based on the data sets provided by Kordia New Zealand. The study focused on identifying spatio-temporal trends and velocity displacement characteristics with regard to time.

Figure 5.10: Illustration of the trajectory of a fishing vessel (dated 09/04/2011) in the Tasman Sea.

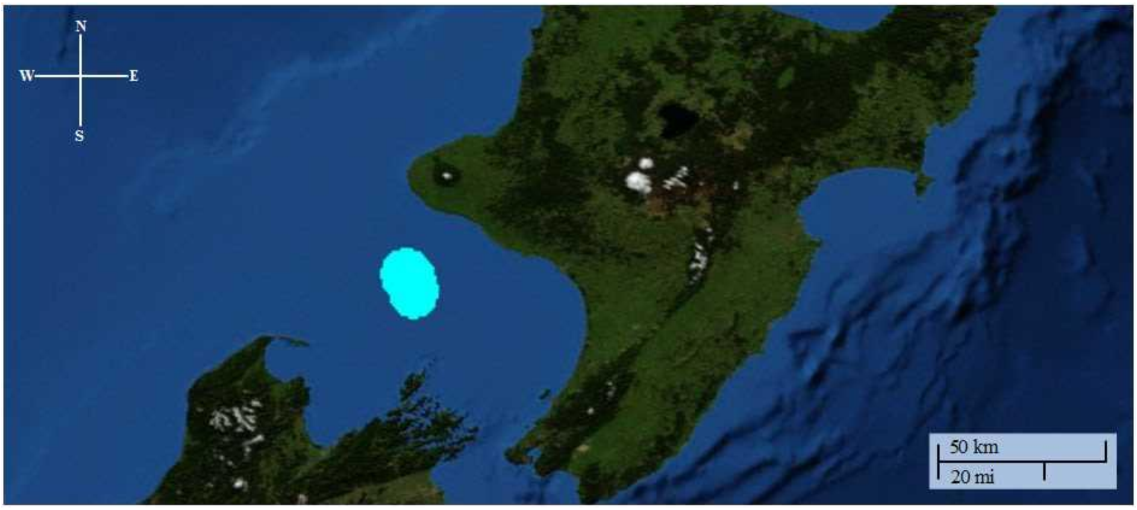


Figure 5.11: In the Tasman Sea, the highlighted trajectory belongs to PROFESSOR MIKHAYLO ALEKSANDROV (Fishing vessel) MMSI 272323000. This highlighted trajectory represents 2 hours of travel by the fishing vessel (MMSI 272323000) on 09/04/2011.

Figure 5.12: Graph displays the fishing vessel's trend movement expressed over Time (hr:min) vs Speed (in knots).

This particular fishing vessel shows constant speed in movement for nearly an hour. Then it suddenly slows down and then later started fluctuating in speed. The variation in speed is shown for nearly 45 minutes before returning to normal speed, and then finally back to its previous normal speed. This variation in speed and unsteady motion demonstrates fishing would have happened during a period of time. The steady motion or constant speed demonstrates the vessel planned to stop at a particular location for fishing.

Table 5.6a: Spatio-temporal analysis of PROFESSOR MIKHAYLO ALEKSANDROV (Fishing vessel) MMSI 272323000 dated 9/04/2011.

Facet of trajectory	Latitude	Longitude	Speed (Knots)	Time	Location
Moves	-40.167033	173.84385	4.7	12.00am	Tasman Sea
Moves	-40.165733	173.84353	4.7	12.01	Tasman Sea
Moves	-40.16425	173.84217	4.8	12.06	Tasman Sea
Moves	-40.163383	173.843067	4.7	12.08am	Tasman Sea
Moves	-39.9959	174.2481	4.8	12.20	Tasman Sea

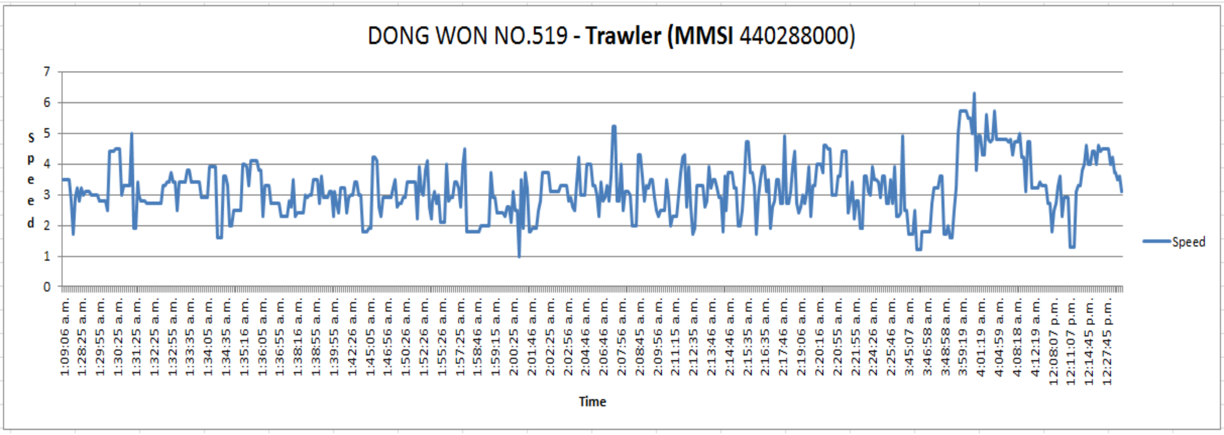


Figure 5.13: Trawler movements of a fishing vessel Dong Won No. 519 (MMSI 440288000) demonstrating speed over time dated 11/05/2011.

The above Figure 5.13 shows the varying speeds of a vessel over a period of 4 hours. The speed of the vessel is inconsistent and fluctuates constantly.

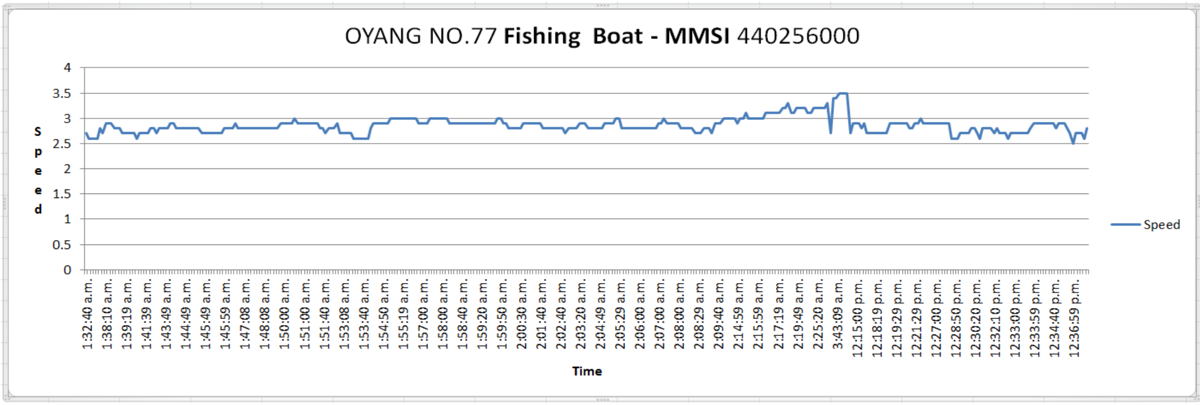


Figure 5.14: Illustrating the velocity of fishing boat OYANG No.77 (MMSI 44025600) demonstrating time (hr:min:sec) over speed (Knots) dated 11.05.2011

An investigation is carried out on another fishing boat (MMSI 440890000 - 530 Dong Won) fishing vessel) that shows a close similar trend in its movements during the phase of sailing between Dundein and Ruapuke Island (Figure 5.15). The trend of the time axis shows it has been moving for a long time without any stop-over during that period of time. From the Figures 5.13 and 5.14 the speed of these fishing boats was between 2.5 to 4 knots. It is evident that the average speed is very minimal.

The spatio-temporal analysis conducted on another fishing vessel (MMSI 272221000) demonstrates that it was anchored at Nelson port, dated on 9/03/2011. Furthermore it reveals that only a few signals were made available.

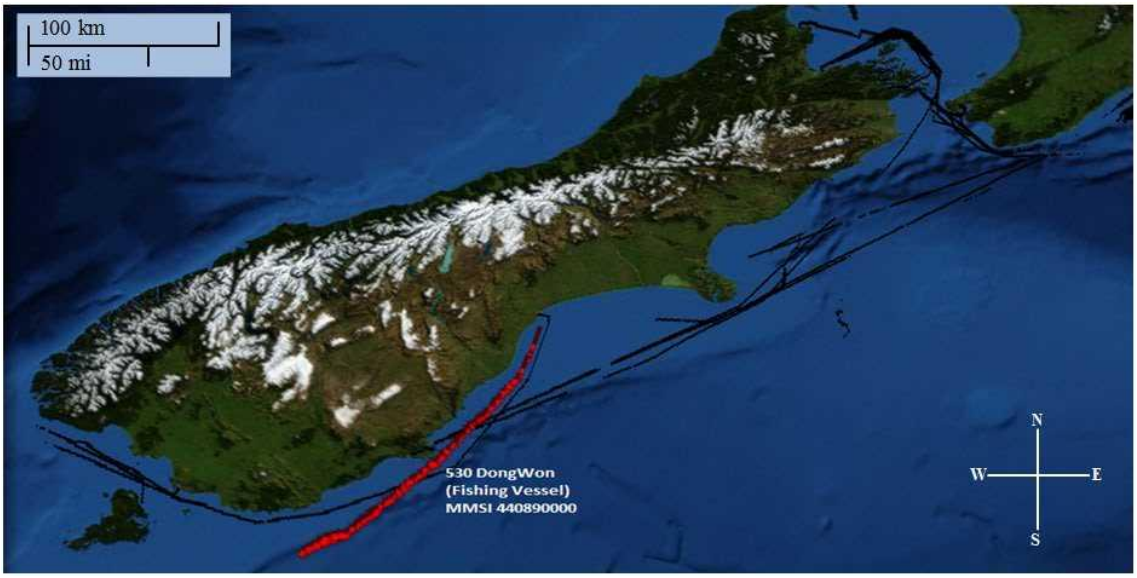


Figure 5.15: The highlighted (red coloured) line illustrates the trajectories produced by fishing vessel 530 Dong Won (MMSI 440890000) near the South Island.

5.5.2.2 High speed Craft

According to the type of boats discussed in Section 5.6, high speed crafts travel is 3 times the speed of the normal boats or ferry. These boats were used for personal use or carrying passengers. To investigate its speed characteristics and type of usage, a high speed vessel was selected called clipper3. Clipper 3 (MMSI 512000741) dated 11/05/2011 and 13/05/2011 was used for analysis as proof about their source, destination, trajectory and speed parameters were discussed below is shown in the Figure: 5.16.

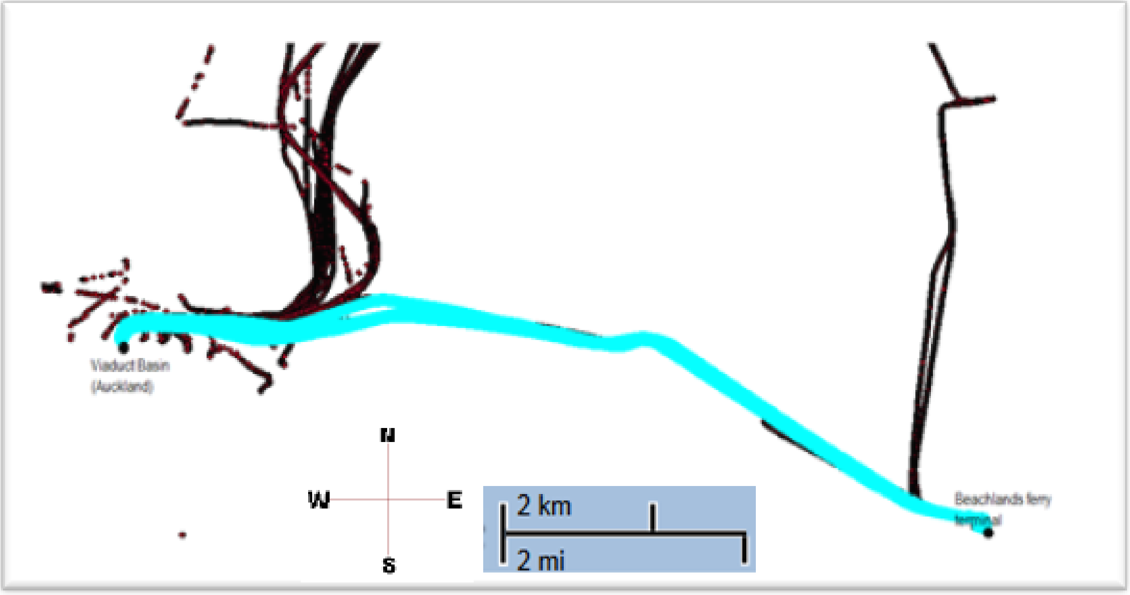


Figure 5.16: Trajectory of Clipper 2 (MMSI 512000742) dated 13/05/2011 illustrated using ArcMap.

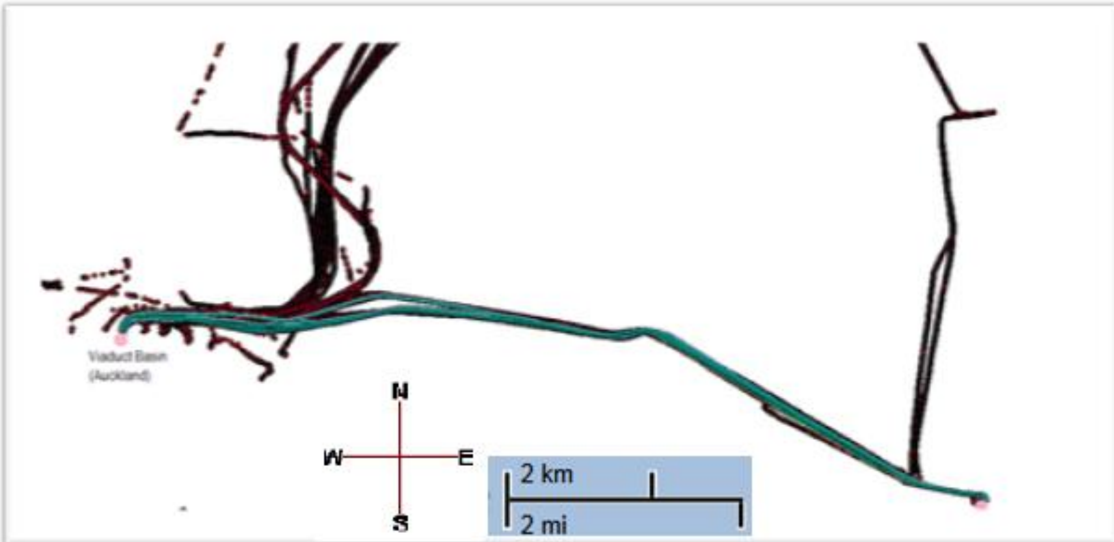


Fig 5.17: Clipper3 (MMSI 512000741) dated 11/05/2011 illustrated using ArcMap tool

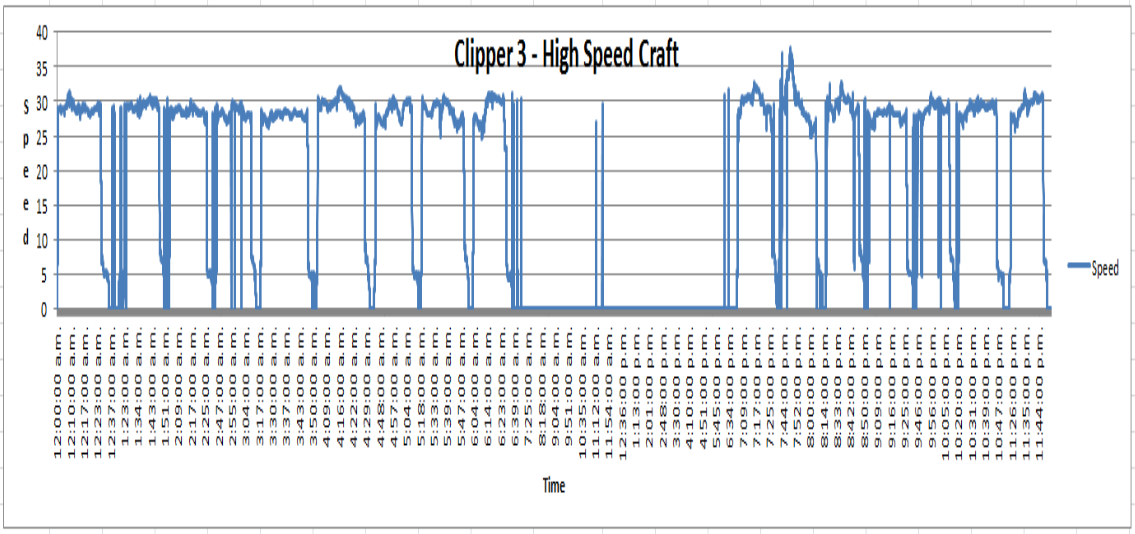


Figure 5.18: Trend showing Clipper 3 – the speed trends of a high speed craft during a weekday dated 10/05/2011.

Table 5.7 : Spatial Temporal Analysis of Clipper 3 which travelled between Auckland and Beachland (Pine Harbour) dated 12/05/2011.

Facet of trajectory	Latitude	Longitude	Speed (Knots)	Time	Location	Arrival/ Destination
Stops	-36.8882	174.987	5.1	5:10:00 a.m.	Beachland	Arrival
Moves	-36.8899	174.9897	1.1	5:20:00 a.m.	Beachland	Destination
Stops	-36.8427	174.7662	1.2	5:52:00 a.m.	Auckland	Arrival
Moves	-36.8394	174.7837	27.5	6:04:00 a.m.	Auckland	Destination
Stops	-36.888	174.9858	4.8	6:29:00 a.m.	Beachland	Arrival

The speed parameter clearly demonstrates its high speed character with an average speed of 30 knots.

5.5.2.3 Naval Research Vessels

Naval research vessels have several roles to play including research. One particular vessel known as Resolution Naval/Research (MMSI 512000200) vessel has been selected for this study. The trajectories projected on the ArcMap tool of this vessel (Figure 5.19) shows the nature of the way in which vessel wandered without any particular route or path. The nature of its trajectory also reveals the fact that it was exploring or possibly undertaking a survey routine. This analysis took a closer look at the trajectory of this naval and research vessel which started from Stanley Bay near Devonport. North of Long Bay the trajectory illustrates that the vessel is on a specific mission or research purpose. It also showed that 2371 rows were selected out of 735463 rows dated 13/05/2011 from GIS database to study this naval research vessel.

Figure 5.19: Illustration of the naval research vessel called RESOLUTION (MMSI 512000200) trajectory dated on 13/05/2011.

5.5.2.4 Ro-Ro Cargo Vessels

As discussed earlier in Section 5.3, Ro-ro is also known as Roll-on/roll-off. These are vessels that are exclusively used to carry wheeled cargo and so support transportation of cars and trucks across oceans.

The above figure 5.20 shows a vessel as being both stationary and in state of movement. Although not constant, the changes in speed are not that dramatic. Generally, the speed of the vessel is 20 knots and steadily drops down towards a stop or anchoring in a harbour.

The Arahura Ferry normally has a travelling route scheduled from Lambton Harbour (Wellington) to Picton Harbour. It is also noticed in Figure 5.21 and from the Table 5.8, that the average resting or anchored time is 95 minutes. The average travel time between Lambton Harbour and Picton Harbour is appropriately 3 hours and 5 minutes.

Table 5.8: Illustrating spatial-temporal details of a Ro-Ro cargo travelling to and fro between Lambton Harbour and Picton Harbour dated 9/05/2011

Facet of trajectory Stops/Moves	Latitude	Longitude	Time	Location
Stops	-41.2566667	174.3993333	12:01:00 a.m.	Lambton Harbour
Stops	-41.2663333	174.7876667	1:46:00 a.m.	Lambton Harbour
Moves	-41.2661667	174.7876667	2:45:00 a.m.	Lambton Harbour
Stops	-41.2838733	174.0061617	5:44:00 a.m.	Picton Harbour
Moves	-41.2835	174.0066667	6:38:00 a.m.	Picton Harbour
Stops	-41.2661667	174.7876667	9:43:00 a.m.	Lambton Harbour
Moves	-41.2661717	174.7876383	10:39:00 a.m.	Lambton Harbour
Stops	-41.2838333	174.0061667	1:43:00 p.m.	Picton Harbour
Moves	-41.2661667	174.7876667	2:45:00 a.m.	Lambton Harbour
Stops	-41.2838733	174.0061617	5:44:00 a.m.	Picton Harbour
Moves	-41.2835	174.0066667	6:38:00 a.m.	Picton Harbour
Stops	-41.2661667	174.7876667	9:43:00 a.m.	Lambton Harbour
Moves	-41.2661717	174.7876383	10:39:00 a.m.	Lambton Harbour
Resting time		Approximately 95mins		
Travel time		Approximately 3 hrs 5 mins		

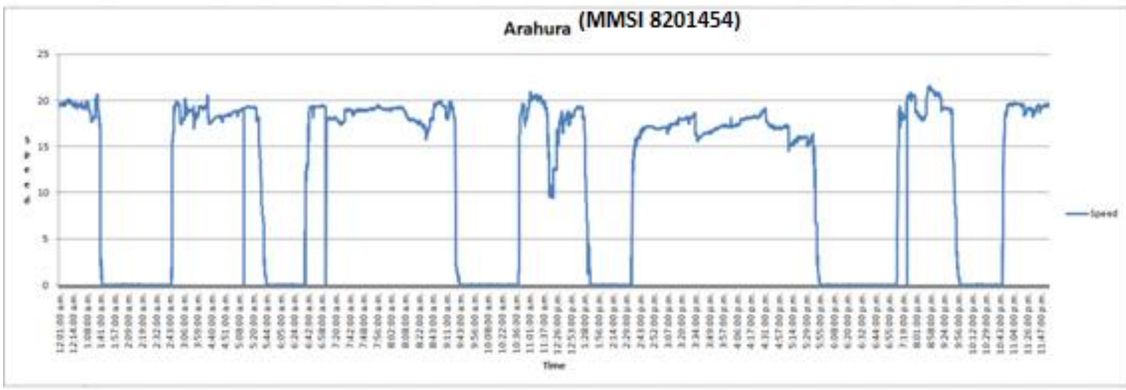


Figure 5. 20: Illustrating the movement trends of a Ro-Ro cargo vessel Arahura (MMSI 8201454). Speed (knots) of a Ro-Ro vessel measured against time (hr:min:sec) dated 9/05/2011.

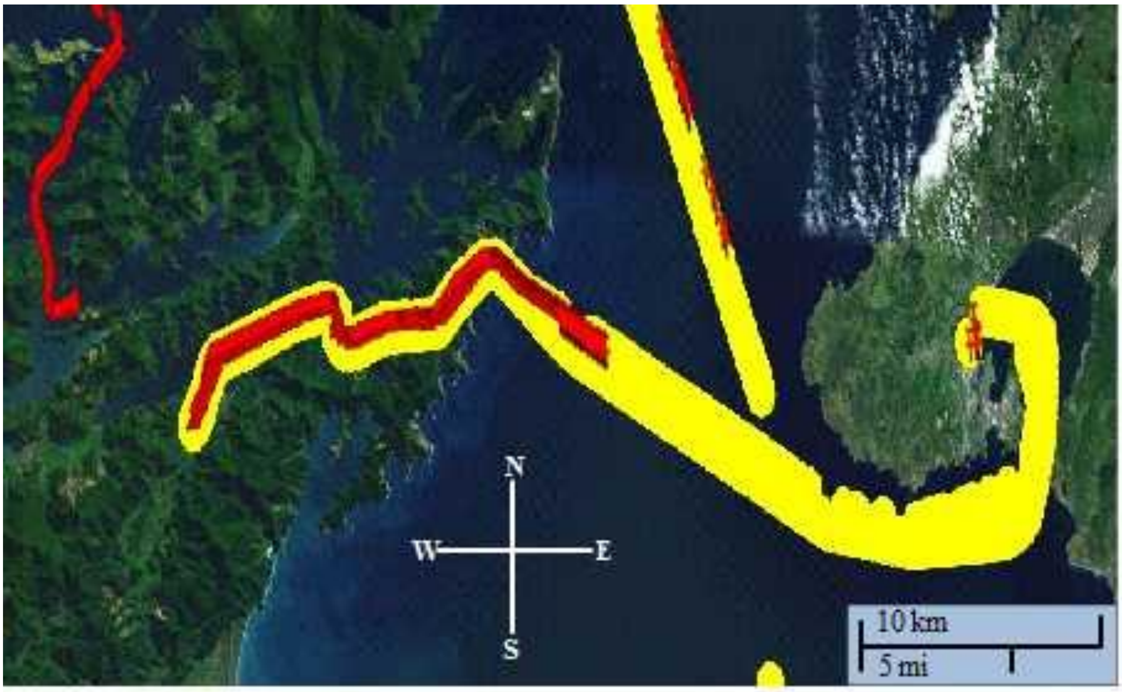


Figure 5. 21: Trajectory of a Ro-Ro cargo vessel called Straitsman (MMSI 512002959) between Picton Harbour and Lambton Harbour. The yellow shade represents the route and the red shade represents the trajectory of the moving Ro-Ro cargo.

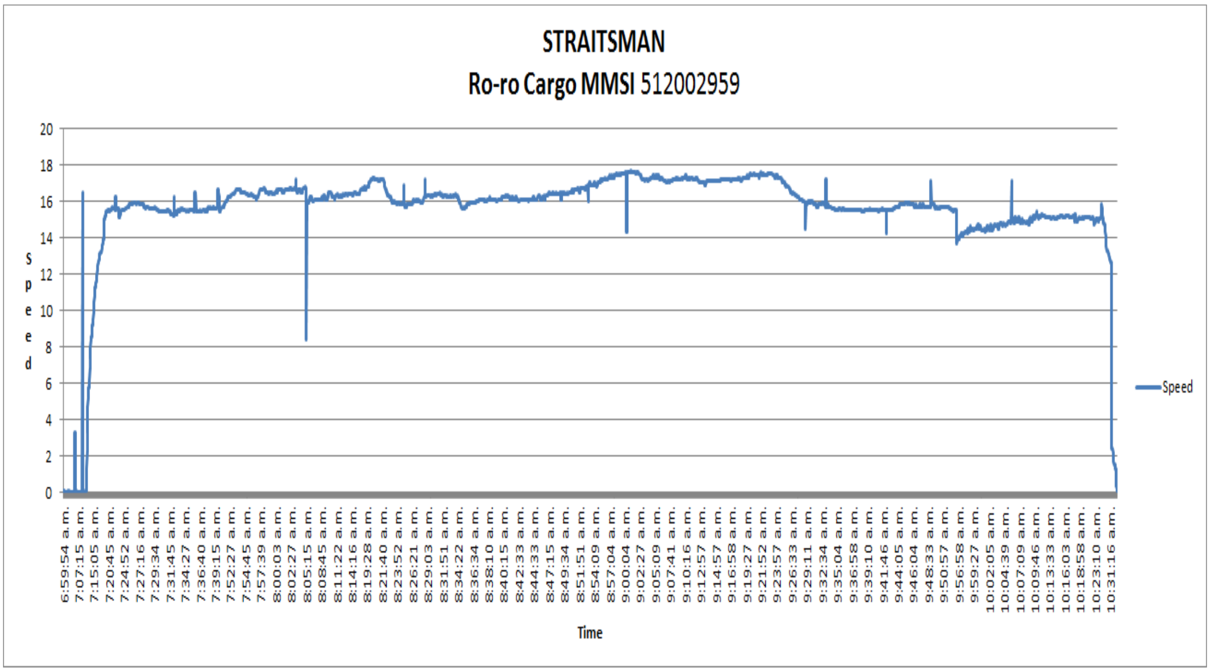


Figure 5.22: A Ro-Ro cargo vessel data set illustrating vessel speed (Knots) over time (hr:min:sec). This illustration shows the vessel movements occurred between 6.59am and10.30 am, and more dated 11/05/2011 (Wednesday).

Table 5.9a: Illustrating the spatial temporal details of another Ro-Ro cargo vessel Straitsman (MMSI 512002959) travelling route between Picton Harbour and Lambton Harbour dated 11/05/2011.

Name of the location	Location		Time	Status	Start from
	Latitude	Longitude			
Picton Harbour	-41.28333333	174.00583	7:06:34 a.m.	Stops	Picton harbour
Picton Harbour	-41.27816667	174.01066	7:14:54 a.m.	Moves	Picton Harbour
Lambton Harbour	-41.28033333	174.78333	10:52:51 a.m.	Stops	Lambton Harbour

Again the rigorous analysis is required to find the trajectory of the vessel Straitsman Ro-Ro cargo during a weekend. One of weekend data sets selected is dated 8/05/2011 (Figure 5.23). The vessel analysis shows a totally 24874 (rows) records were created by this particular Ro-Ro cargo vessel. The measurement of speed over time shows a more constant pattern. A comparison is made with the 11/05/2011 dataset (Figure 5.22) of the same Ro-Ro cargo vessel and thus also shows the ***same pattern***. For example, the trips between ports, 3 hours approximately was spent on their trips between ports. It is also observed that the ports and the routes were alike, whereas a discussion spent resting or anchored time shows a variation in time. For instance, in the early hours of the journey the anchored time is appropriately 1 hour (between 4am to 5am) whereas during the busy hours in the morning the analysis shows approximately 2 hours of anchored time. The analysis (Figure 5.23) shows that the trips were scheduled at regular interval of time and the time spent resting or anchored were also noticed at regular periodic intervals.

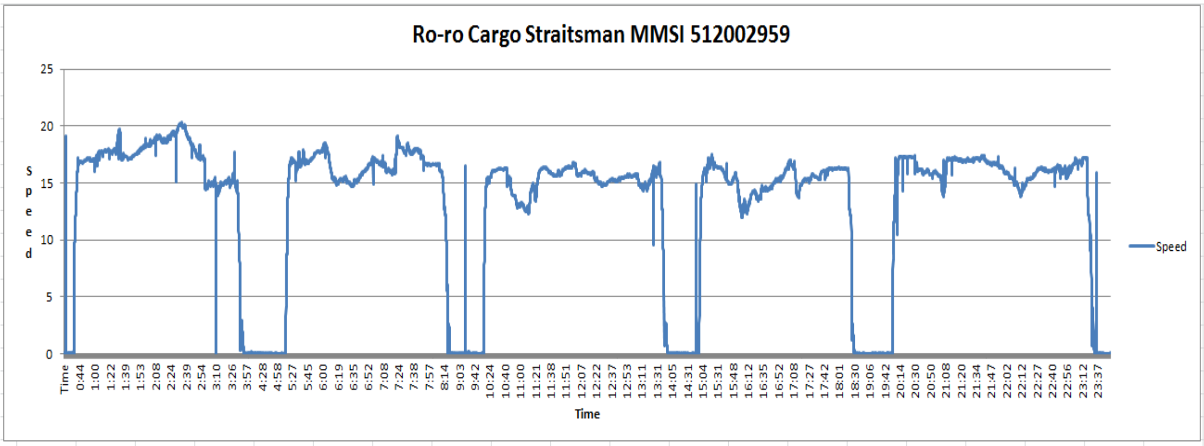


Figure 5. 23: Straitsman Ro-Ro cargo dataset illustrating trend pattern speed (knots) over time (hr:min) dated 8/05/2011 (Sunday)

5.5.2.5 Pleasure Boats

Pleasure boats are a kind of vessel exclusively used for personal, family and occasionally for recreational purpose.

Two pleasure boats were taken into consideration to perform analysis. Boat #1- Aquarius and Boat # 2 – Ammonite.

Regarding Boat #1- Aquarius, the graph displays velocity over time and shows its variation. The variation is terms of speed, measured in Knots. This demonstrates the variation in speed which have could been because it was involved in recreational or site seeing activities. Further, the analysis shows this craft has been in operation near the **Tryphena area.** This craft started from Auckland and took the Auckland-Matiatia route.

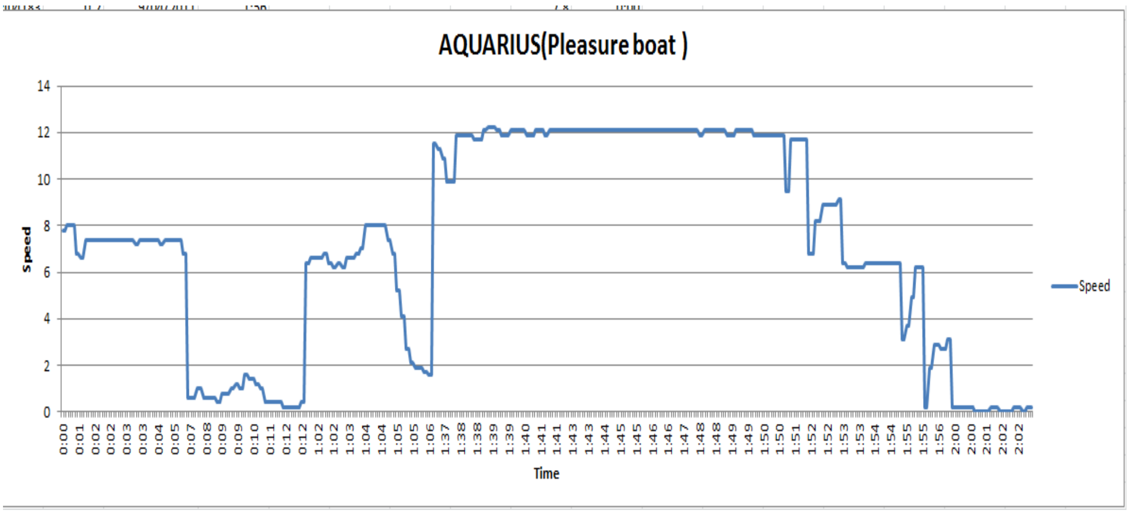


Figure 5. 24 : Route taken by Aquarius (Pleasure boat) between 12.00am to 2.03 am on 7/05/2011. Speed measured in knots over time (hr:min)

The Figure 5:25 depicts the most popular routes taken by various ferries, passenger ships, pleasure crafts and many smaller vessels. The routes passed through between Motutapu Island, Waiheke Island and Motuihe Island. Another popular route normally taken by various vessels is the route between Auckland and Motuihe Islands. Another similar route is Auckland to Tryphena Harbour (Great Barrier Islands).

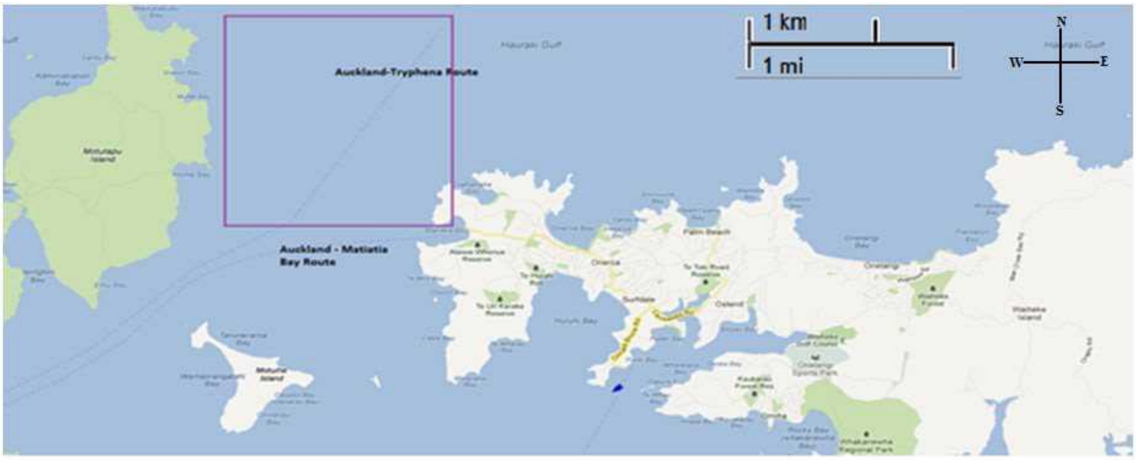


Figure 5.25: Popular routes taken by Pleasure boats (Adapted from: Marinetraffic.com).

Boat #2: About 4634 rows of data were created by Ammonite pleasure boat (MMSI 126282899) on 8/05/2011 from midnight till 7.08am. Boat #2 shows has started from Paihia. The analysis shows the boat was between Paihia and Russell and headed towards Bay of Islands and the middle of the ocean. Investigating the trend pattern of Ammonite Pleasure Craft (Figure 5.26) demonstrates it has been more constant in its speed between 8 knots and 9 knots. According to the velocity of the craft, it is constant and steady.

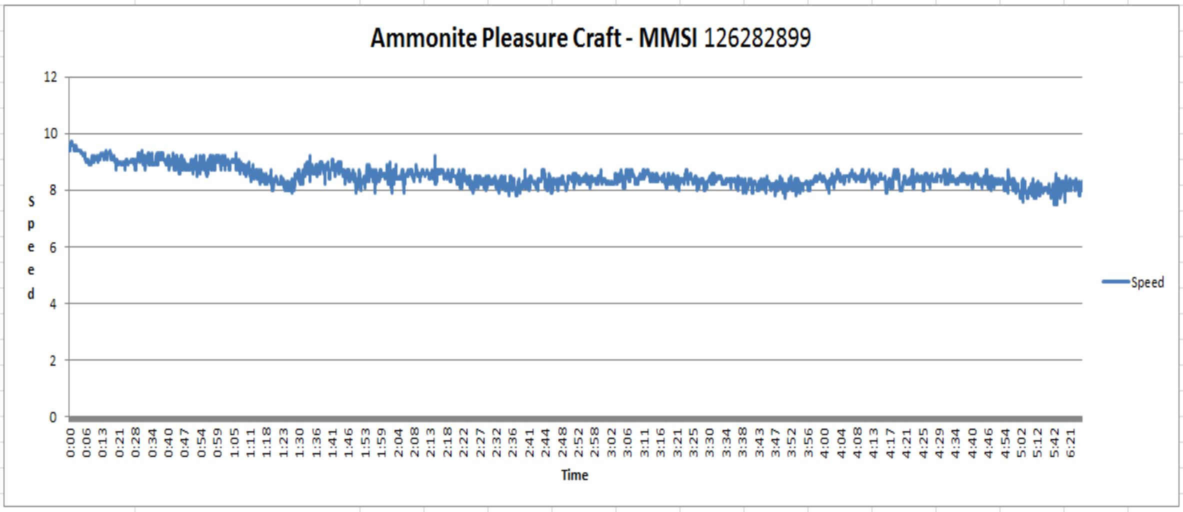


Figure 5. 26: Demonstrating the velocity (speed in knots) of the vessel against time (hr:min)

Table 5.9b: Table illustrating the spatial coordinates referring to location and speed along with time

Name of the location	Location		Speed	Time	Status	Start from
	Latitude	Longitude				
Between Paihia and Russell	-35.2833	174.10482	9.2	12:00:00 a.m	Moves	Near Paihia
Towards Bay of Islands	-35.199258	174.132231	9.6	12:36:00 a.m.	Moves	
Out in ocean	-35.01757	174.155508	8.4	1:52:00 a.m.	Moves	
	-34.85351	174.1541683	8.5	3:03:00 a.m.	Moves	
	-34.67741	174.1571817	8	4:19:00 a.m.	Moves	
	-34.5494266	174.1460433	8.2	5:15:00 a.m.	Moves	
	-34.338835	174.1013633	8.2	6:51:00 a.m	Moves	

Later readings were not available to study this vessel. This could be due to the vessel being out of range of AIS. To verify the location coordinates and the trajectory produced by this particular vessel, the coordinates with speed and time were plotted on ArcGIS. The highlighted trajectory shows it is continuous and moves on in the ocean (Figure 5.27).

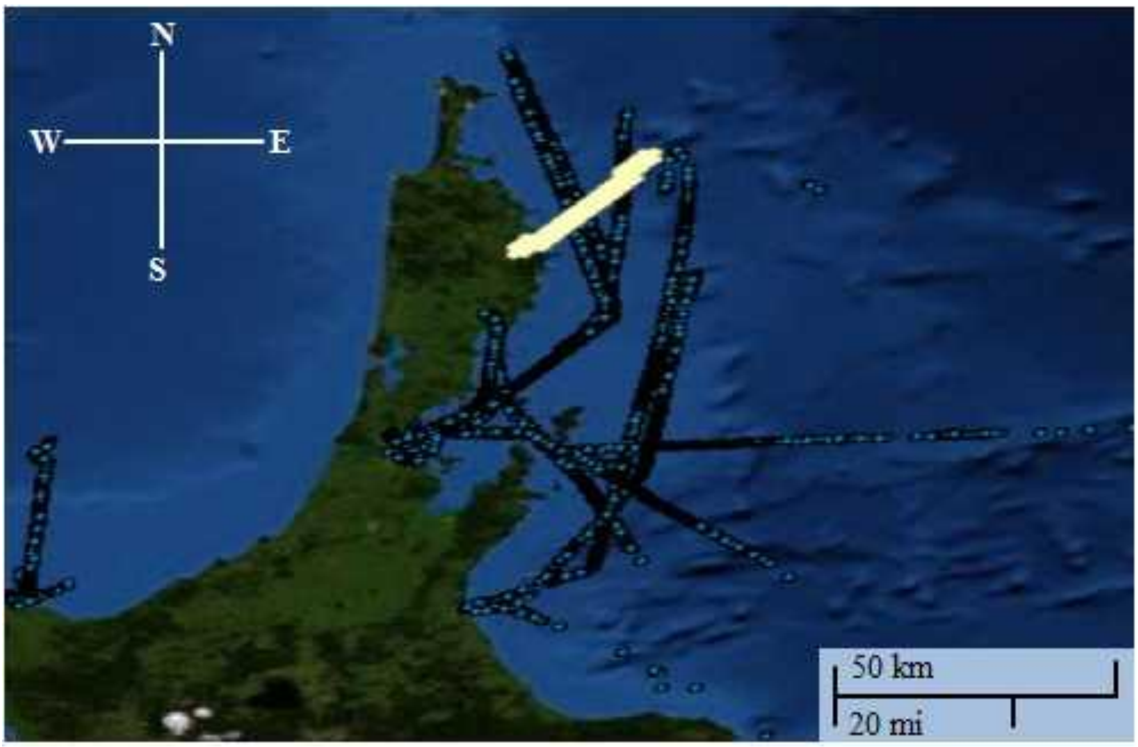


Figure 5. 27: Illustration to highlight the trajectory produced by a pleasure boat called Ammonite (MMSI 126282899)

5.5.2.6 Pilot Vessel

A pilot vessel is exclusively used to transport maritime pilots between land and inbound or outbound ships. This kind of role is carried out during the time of need and the specific purpose. This vessel will be called into operation during times of need.

In terms of Wellington Pilot (Figure 5.28) vessel's speed, it is for steady nearly an hour and then slowly it started varying its speed. From 7.06am to 7.11am the trend slows down, this could be because the pilot vessel has reached its designated vessel or ship for the assigned task. Taking a close look at the following trend informs that the pilot vessel has stopped for approximately 20 minutes. Later, its speed instantaneously rose to 14 knots. This could be once the assigned task has been completed. This particular vessel is found operating in the Cook Strait region.

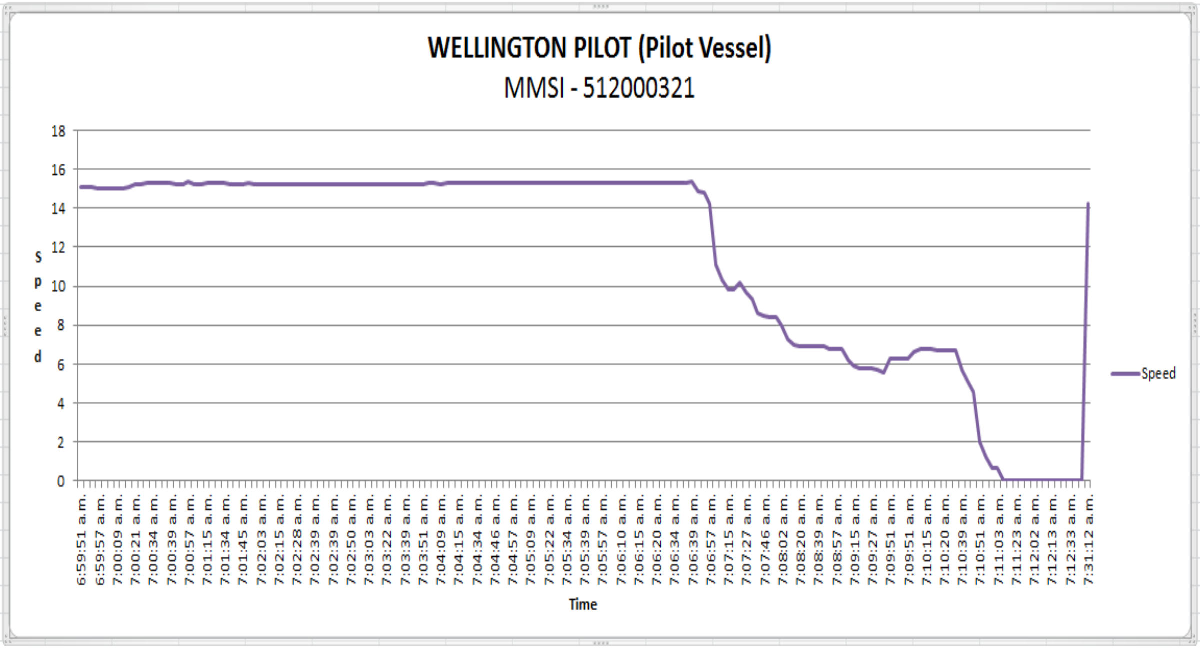


Figure 5.28: A graph illustrating the speed measured in knots against time (hr:min:sec) trend of a Pilot Vessel. Dated 11/05. Time: 6.59 am to 7.31 am.

On 8/05/2011 it is noticed that this pilot boat is frequently in use between Wellington Harbour and in Fitzroy Bay (Figure 5.30 and Table 5.10).

Table 5.10: Illustrating spatio-temporal details with speed parameter (knots) dated 8/05/2011.

Latitude	Longitude	Speed (Knots)	Time (hr:min)	Location
-41.3422	174.8454	15.1	7:26	Between Fitzroy bay
-41.2845	174.8334	15.9	10:28	near Wellington harbour
-41.2812	174.7817	5.3	19:22	Wellington harbour
-41.3768	174.8296	15.6	20:09	Wellington harbour

The Table 5.10 and Figure 5.30 demonstrate the spatio-temporal analysis and trend graph respectively for the Pilot Vessel (Wellington Pilot). These studies eventually verify the results between the spatio-temporal analysis and the velocity trend analysis.

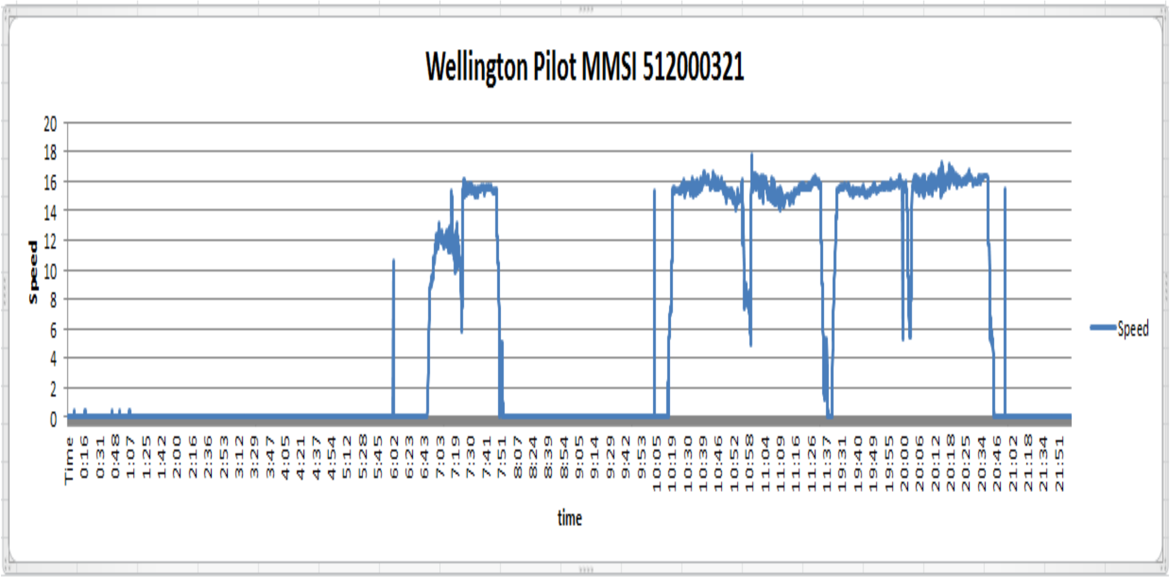


Figure 5.29: Graph illustrating the trend of speed (in knots) over time (hh:mm) of a Pilot Vessel. Time: 00.16am to 21.51 (9.51 pm) dated 8/05/2011.

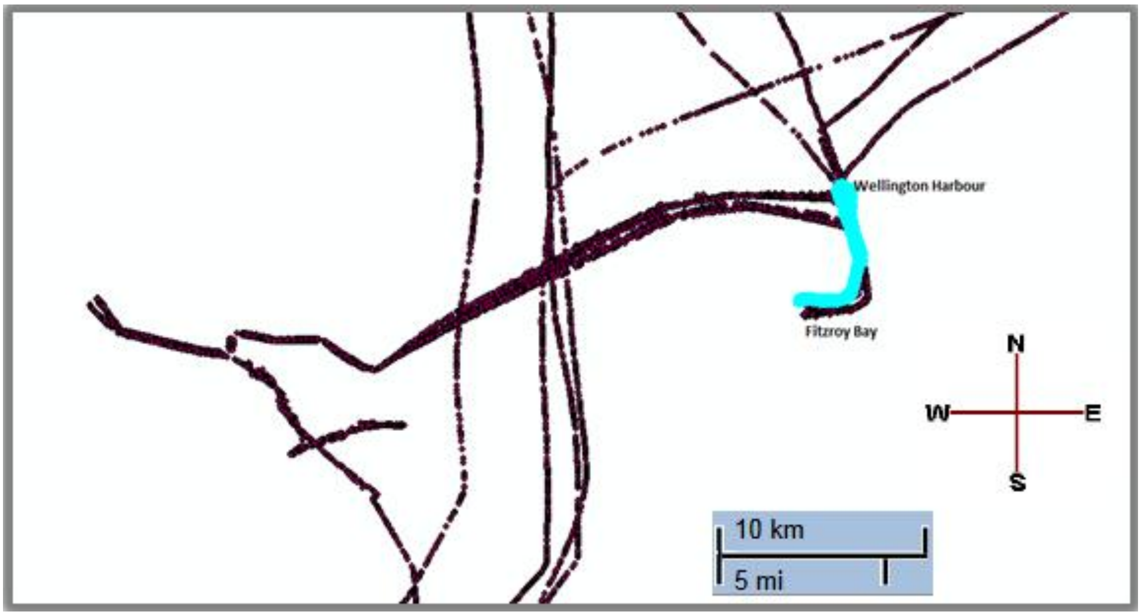


Figure 5.30: The trajectory of the Pilot Wellington MMSI 512000321 (between Wellington harbour and Fitzroy harbour) dated 8/05/2011.

Another pilot vessel Akarana (MMSI 512000344) is used to for study the trend of speed over time. The trend points out approximate similarity compared to the with the Wellington Pilot vessel. The irregular nature of both the speed and stop/move patterns indicate that the vessel is influenced by needing to provide ad-hoc services. Further analysis of spatial locations indicates that it is on service between Hauraki Bay (near McKenzie Bay) and Auckland.

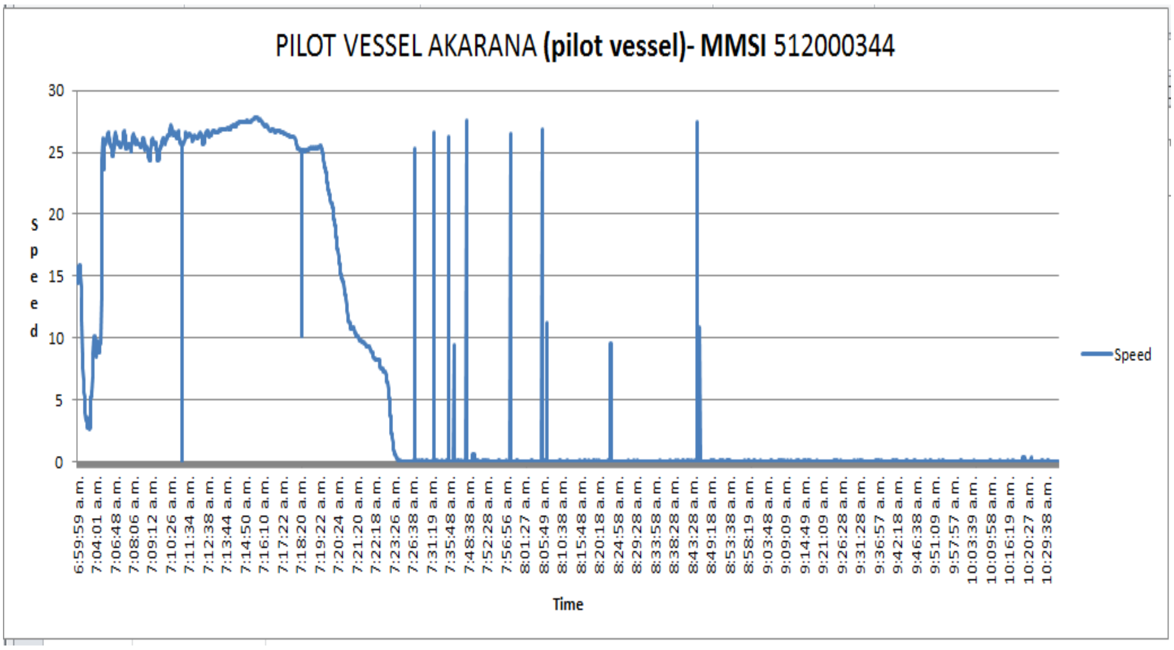


Figure 5.31a: Another example of a Pilot Vessel Akarana illustrating a similar trend of speed (in knots) over time (hh:mm:ss).

Table 5.11: Table illustrating the spatial coordinates of a pilot vessel (MMSI 512000344)

Name of the location	Location		Start Time	Status	Start from	Destination
	Latitude	Longitude				
Hauraki Bay (near Mckenzie Bay)	-36.76100167	174.81582	7:06:30 a.m.	moves	Un-known	Auckland
Auckland	-36.84503667	174.78293	7:47:58 a.m.	stops		

5.5.2.7 Ferry Boats

The ferry boats were frequently noticed between Wellington Harbour and Picton Harbour. The ferries were directed using the route of Cook Strait and narrowly passed through Arapawa Island to reach Picton Harbour (Figure 5.31b). In addition, it is also noticed that this route is more frequently used and much populated with various types of boats, ferries and ships. In other words, the Cook Strait route is a densely populated route.

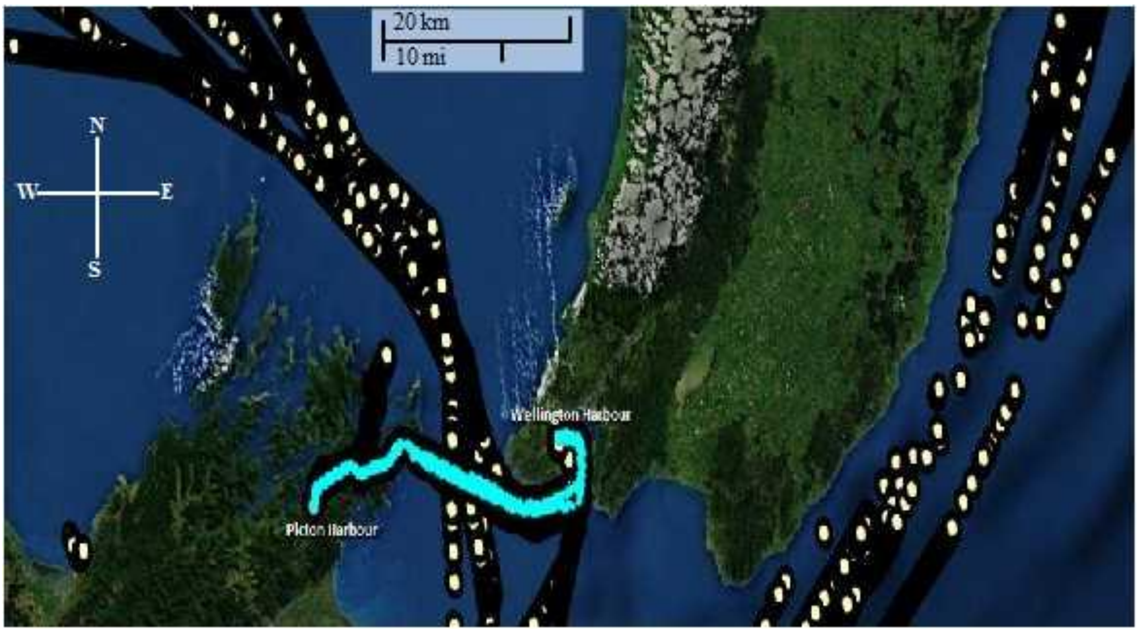


Figure 5. 31b: Illustrating the trajectory taken by a ferry boat from Wellington Harbour to Picton Harbour.

5.5.2.8 Passenger Boats

A passenger boat's (Figure 5.33) trajectory shows its path between Mahau sound and Pelorus Sound. This trajectory illustrates that this particular boat was been involved in a site seeing activity or field trip tour based role. Another example has been used to to illustrate a similar role played by another similar ship, the Spirit of New Zealand. The trend graph of speed (in knots) over time (hh:mm) noticed to be irregular pattern (Figure 5.32). This illustrates its role as site seeing or field trip activity.

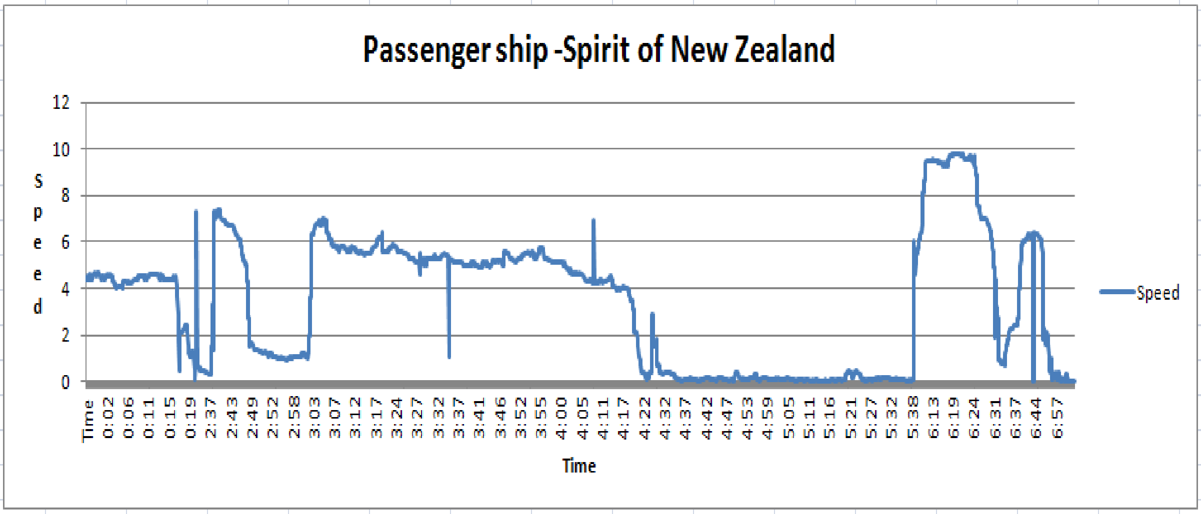


Figure 5.32: Trend showing speed (in knots) versus time (hr: mins) of a passenger ship.

Another interesting trend noticed is utilizing much anticipated routes, in additional to exploring activities.

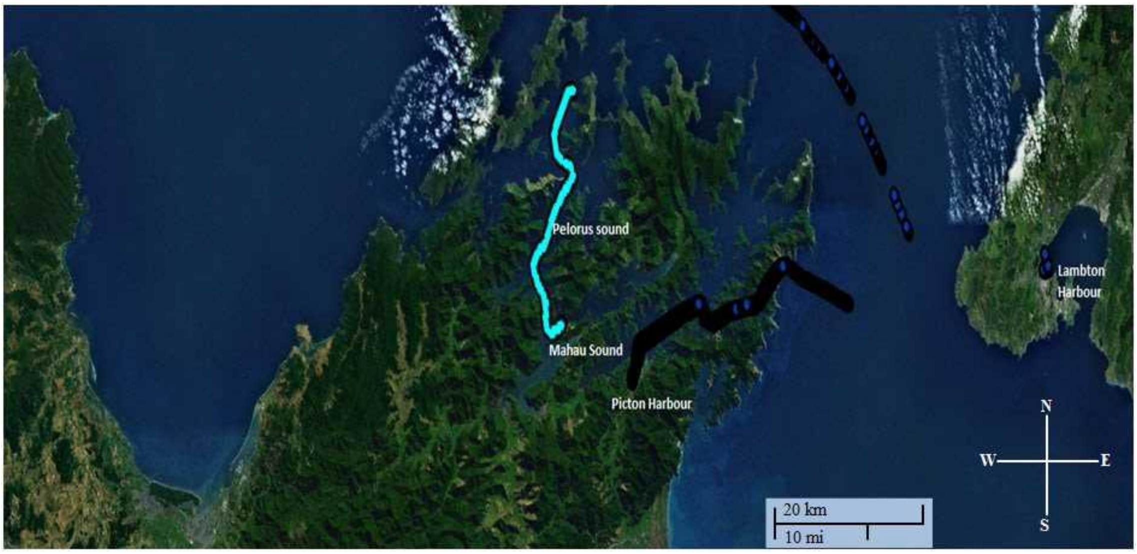


Figure 5. 33: A trajectory analysis of a passenger boat between Mahau Sound and Pelorus Sound

5.5.2.9 Container Ships

Container ships are also known as cargo ships and are primarily designed to carry multi-purpose goods in single or multi decks. They carry loads in ‘truck-size’ containers.

To analyse the characteristics of container ships, Spirit of Endurance (MMSI 512042000) is taken as an example. Out of the selected dataset, dated on 11/05/2011 and 12/05/2011, the records show the following inferences: The Day 1, dated 11/05/2011, had only a few records recorded at Wharangi Port which indicated the vessel was anchored.

Day 2 – dated 12/05/2011, a thorough investigation was carried out to analyse the trajectory of this container ship. The trend in this graph clearly shows its speed was constant over a period of 8 hours (appropriately). The speed is well maintained between 15 and 20 knots.

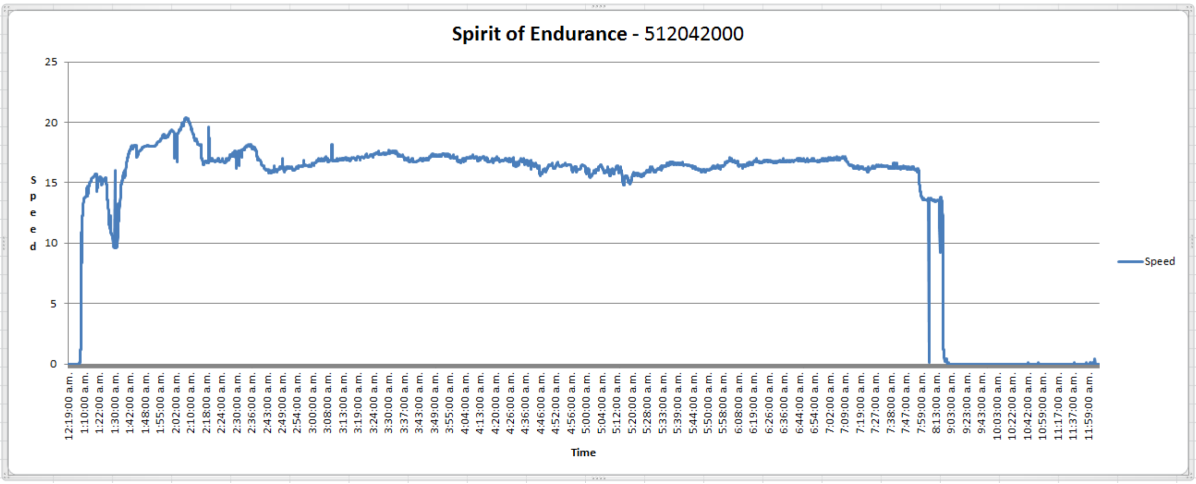


Figure 5.34: Illustrating the speed (knots) of a container vessel over time (hr:min:sec). Dated 12/05/2011 temporal detail: 12.19am to 12 am (appropriately 12 hours)

Further analysis shows that in a journey of 7 hours and 45 minutes (appropriately) this container ship has not taken the normal route of ferries or passenger boats. The reason might be the size and mass of the vessel, also so as not to interfere in the passenger or ferry routes. Furthermore, considering the safety reasons, the routes are separated. A detail tabulation (Table 5.12) depicts the spatio-temporal analysis conducted on the same set of data to verify its location against time details.

Table 5.12 : An Illustration of the spatial coordinates referenced along with the temporal (hr:min) details. Speed is measured in knots.

Latitude	Longitude	Speed (knots)	Time (hr:min:sec)	Location
-41.27857833	174.78995	0	12:56:00 a.m.	Lambton Harbour, Wellington
-41.27721333	174.8118633	13.8	1:12:00 a.m.	Departure
-41.36802667	174.8098017	18.1	1:44:00 a.m.	Fitzroy Bay
-41.37817333	174.7122783	18.8	1:59:00 a.m.	Moving towards Cook Strait
-41.19977833	174.53889	16.2	2:48:00 a.m.	Away from normal route of ferries
-41.05122167	174.4774017	17.3	3:22:00 a.m.	Towards Port Gore
-40.851645	174.31263	17.1	4:11:00 a.m.	East of Nukuwaiata Islands
-40.66303833	174.0347367	16.3	5:13:00 a.m.	East of Stephens Islands
-40.70169333	173.8744483	16.3	5:45:00 a.m.	North of Stephens Islands
-40.84762333	173.6801317	16.7	6:31:00 a.m.	West of D Urville Islands
-40.96529	173.53348	17	7:05:00 a.m.	Tasman Bay
-41.15812667	173.3098233	13.6	8:02:00 a.m.	Towards Nelson (near Wakapuaka)
-41.26078	173.2698083	0.9	8:57:00 a.m.	Arrives Port Nelson
-41.26107667	173.269715	0	11:36:00 a.m.	Port Nelson, NELSON

5.5.3 Frequency Distribution – Speed parameter

Speed is considered as an additional parameter to analyse the trajectory of a moving object. The speed parameter provides additional information and knowledge in addition to the trajectory which is also analysed. As the speed parameter facilitates to an understanding of the nature of the trip, for example, pre-assigned tasks or task orientation, besides, the time of travel whether off-peak or peak time. The speed parameter has also facilitated the study of the completion of the trip in the estimated time. Another inference noticed was the route taken as frequently travelled route. While the speed is zero, this indicates the vessel is anchored or stopped for a reason.

Overall, the speed parameter helps to characterise the type of a vessel with smooth trajectory. For example, high-speed craft average speed is between 25 knots - 28 knots.

The objective of using speed information is to have a frequency distribution graph to calculate mean, median, mode, range and standard deviation. These statistics become a critical factor when conducting quantitative analysis.

5.5.3.1 Analysis conducted on fishing vessels

After a fishing vessel is studied in a spatial and temporal analysis, the same vessel is again used to study its speed parameter. The descriptive statistical analysis is conducted using the graphing functionality of the ArcGIS tool. The outcome shows the minimum and maximum speed of this fishing vessel is 1.8 and 7.5 knots respectively. Studying the graph, illustrates the variation in speed values. Most times, the speed range is between 4 and 4.6 knots. This characterises the fishing vessel, speed in a task-oriented role of itinerant searching for right area for fishing or fishing related activities.

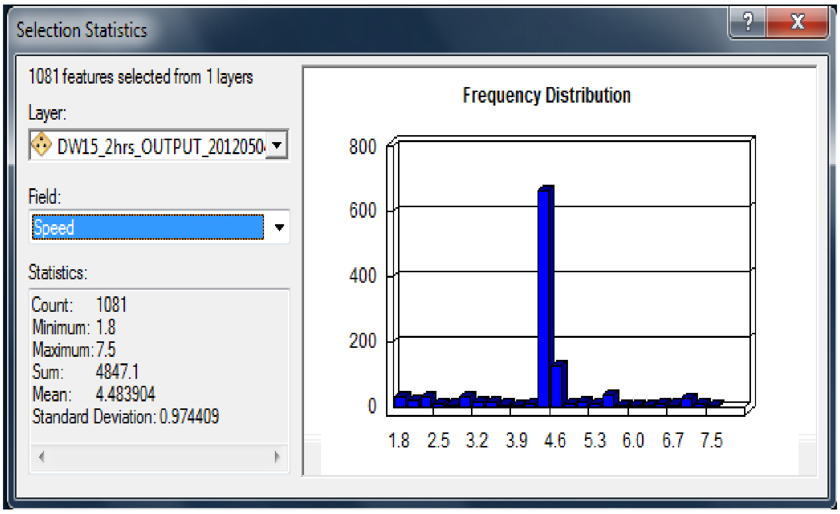


Figure 5.35 : An illustration of the speed (measured in knots) statistics of a Fishing factory called PROFESSOR MIKHAYLO ALEKSANDROV (MMSI 272323000)

5.5.3.2 Naval/Research Vessel

To study the speed parameter of a naval research vessel, the vessel Resolution Naval/Research vessel (MMSI 512000200) is selected. The study reveals (Figure 5.36) that the minimum and maximum speeds were 2.2 and 11.8 knots respectively. Mostly, the speed remained between 9 knots and 10 knots. Therefore, the mean speed is 8.6 knots. Additionally, the graph shows that this research vessel was on the move continuously during its mission.

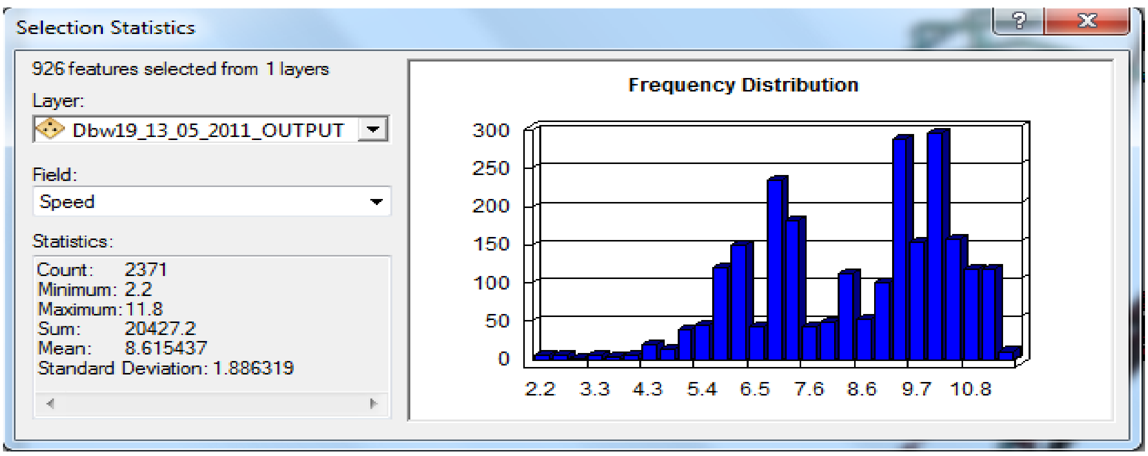


Figure 5.36: A Frequency Distribution of a naval research vessel (MMSI 512000200).

5.5.3.3 Pleasure Boat

To study a pleasure boat's frequency distribution, previously analysed pleasure boat (MMSI 12682899) was selected. The graph (figure 5.37) illustrates that, the particular boat is on the move continuously, and travelling at a more constant speed of 11 knots. The average mean speed shows 8.8 knots and this is maintained at a constant speed. The few counts of speed at 'zero' means not much time was spent in a stationary position. The speed profile of this pleasure boat helps to comprehend two facets: the first is it is constant movement and it has an almost similar speed profile. The Figure 5.38 demonstrates the frequency distribution of a pleasure boat dated 13/05/2011. This statistical study performed using ArcGIS tool.

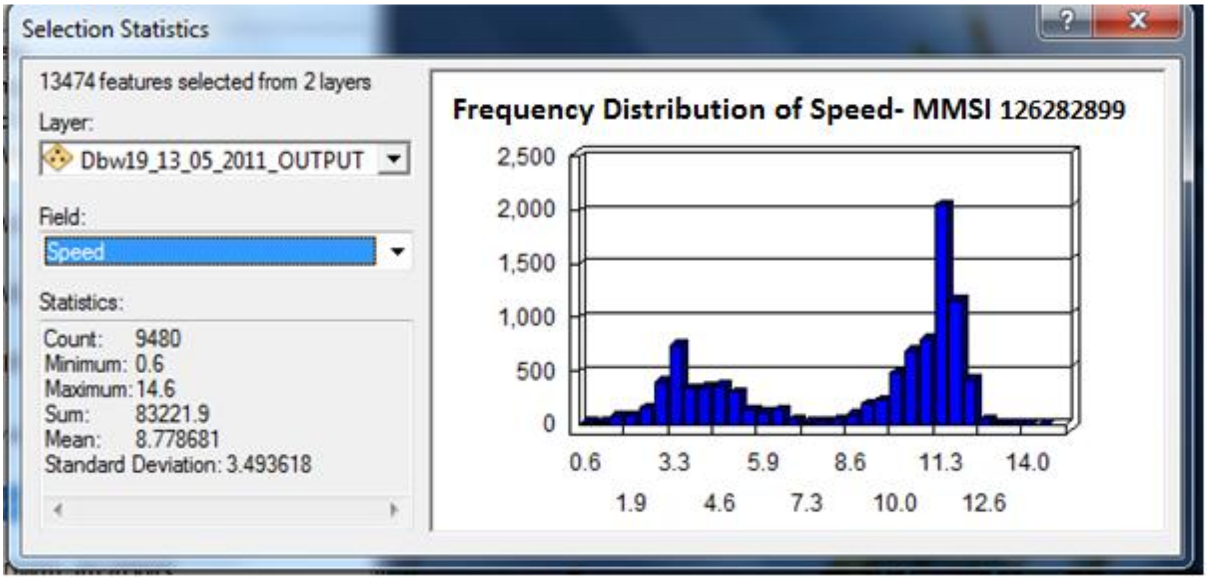


Figure 5.37: Frequency Distribution of a pleasure boat (MMSI 1262282899). Speed measured in knots.

5.5.3.4 Passenger Boat

The frequency distribution of a particular passenger boat (MMSI 512201000) has been selected to analyse its speed profile to understand and comprehend the characteristics of this type of moving boat. This graph shows (Figure 5.38) that, this passenger boat has spent most of its time in a stationary position, dated on 13/05/2011. Another week day's record inferred that it exhibits the same pattern of stationary position for the same time. This is concluded from number of counts at speed= 0. Generally speaking on the speed profile of a vessel, their minimum and maximum speeds were between 0 knots and 9.7 knots respectively.

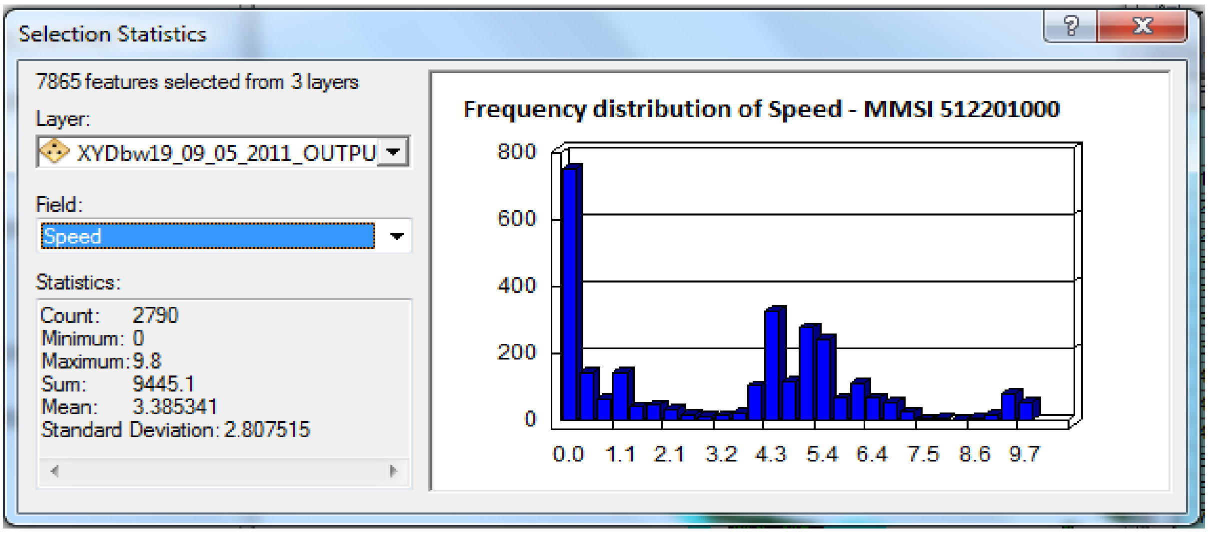


Figure 5.38: Frequency Distribution (speed measured in knots) of a passenger boat.

5.5.3.5 High Speed Craft

A close investigation was conducted to study a high speed craft. In the dataset provided by Kordia New Zealand, it is noticed there were three popular high speed craft namely Clipper 1, 2 and 3. According to IMO (IMO, 2010) these craft were crafts capable of a maximum speed in metres per second (m/s) equal to or exceeding 3.7 times. The graph illustrates (Figure 5.39) the speed of this high speed craft is 34.5 knots which is 3 times more than the average speed of a passenger or pleasure boat. This quality itself underpins and supports the point of high speed craft.

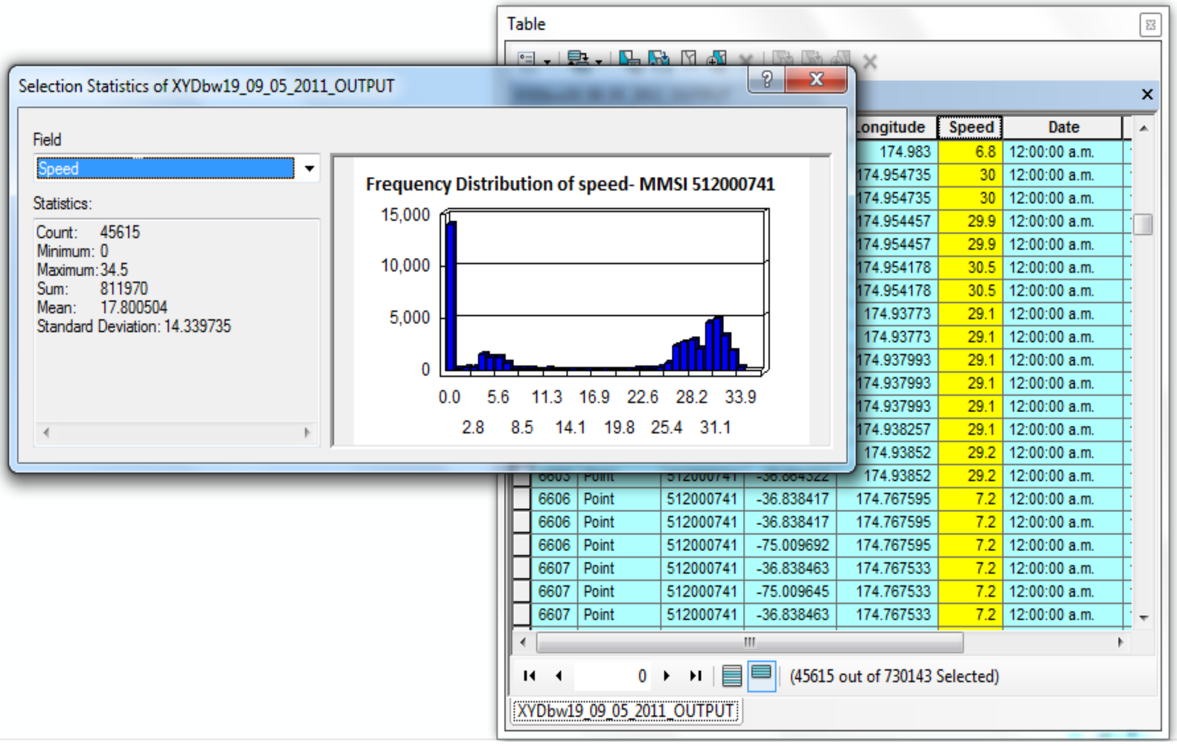


Figure 5.39: Frequency Distribution (speed measured in knots) of a high speed craft.

5.5.3.6 Pilot Vessel

In order to illustrate the speed profile of a pilot vessel called Wellington Pilot, (MMSI 512000321) is selected. Previously, this pilot vessel was discussed to analyse the spatio-temporal displacement in Section 5.8.2.6. The rationale for selecting the same vessel is based on the fact that it would provide further insight to understand its characteristics. The graph illustrates (Figure 5.40) that the maximum speed of this pilot vessel is 17.8 knots and its average mean speed is 5.6 knots. It is also noticed that the sum of the counts of speed zero conveys that at most times of the day, it is in stationary position at a particular location.

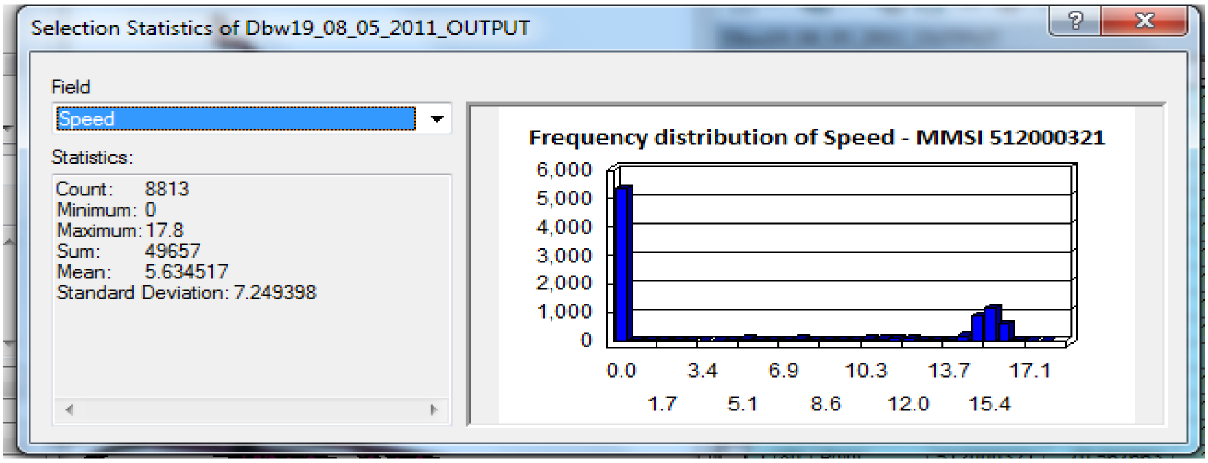
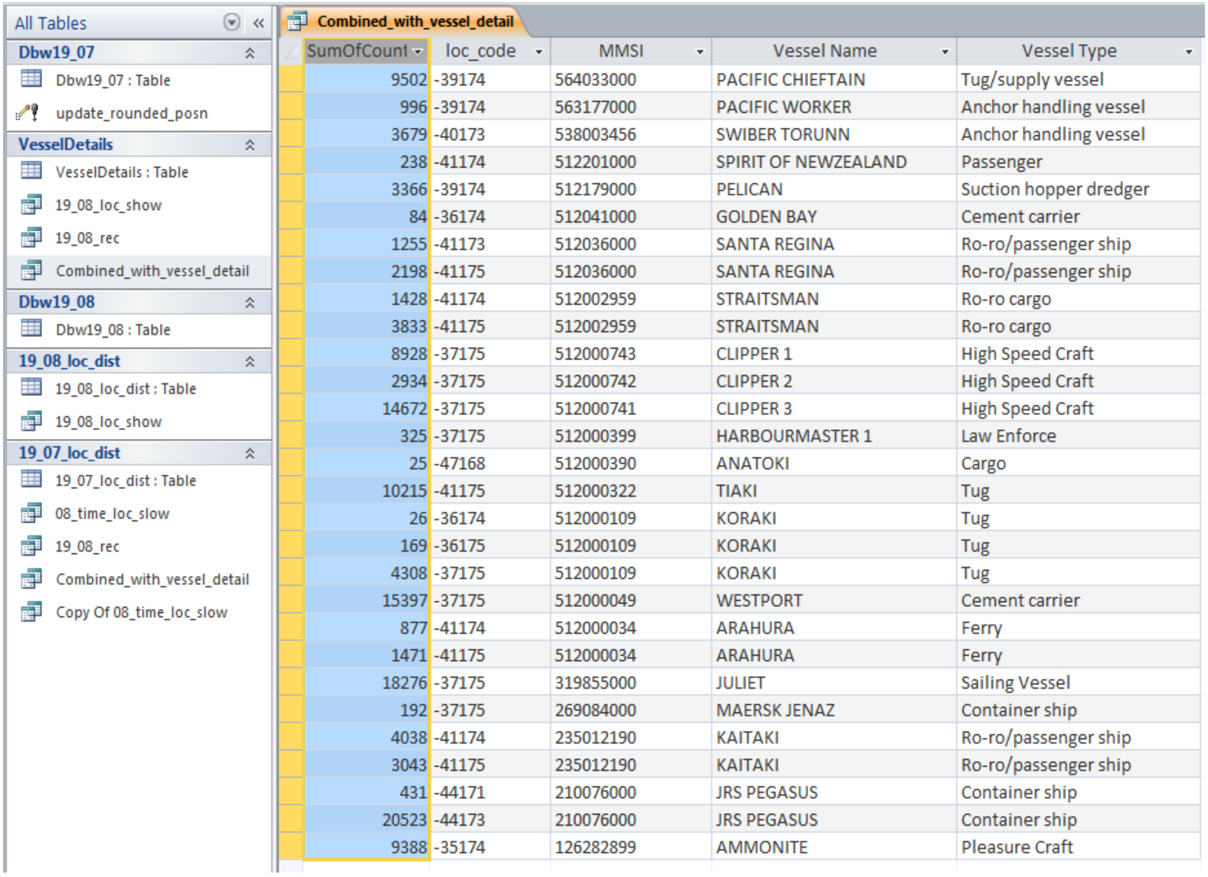


Figure 5.40: Frequency Distribution (speed measured in knots) of a Pilot Vessel called Wellington Pilot (MMSI 512000321).

5.5.4 Stationary Locations

To affirm the identified characteristics of a vessel type one final endeavour is undertaken. The final stage of the analysis is to identify the duration spent in ports in harbour at anchored. The analysis is conducted on the duration spent in ports. The motivation is to understand how long these vessels have been anchored in harbours and relate this to their roles. In order to understand a small utility was created using MS Access. This utility calculates the sum of the Counts of a particular vessel whose speed is less than 1. The location coordinates are rounded to the nearest value, loc_code. The other attributes were used such as MMSI, Vessel Name and Vessel Type in order to identify the vessels.



SumOfCount	loc_code	MMSI	Vessel Name	Vessel Type
9502	-39174	564033000	PACIFIC CHIEFTAIN	Tug/supply vessel
996	-39174	563177000	PACIFIC WORKER	Anchor handling vessel
3679	-40173	538003456	SWIBER TORUNN	Anchor handling vessel
238	-41174	512201000	SPIRIT OF NEWZEALAND	Passenger
3366	-39174	512179000	PELICAN	Suction hopper dredger
84	-36174	512041000	GOLDEN BAY	Cement carrier
1255	-41173	512036000	SANTA REGINA	Ro-ro/passenger ship
2198	-41175	512036000	SANTA REGINA	Ro-ro/passenger ship
1428	-41174	512002959	STRAITSMAN	Ro-ro cargo
3833	-41175	512002959	STRAITSMAN	Ro-ro cargo
8928	-37175	512000743	CLIPPER 1	High Speed Craft
2934	-37175	512000742	CLIPPER 2	High Speed Craft
14672	-37175	512000741	CLIPPER 3	High Speed Craft
325	-37175	512000399	HARBOURMASTER 1	Law Enforce
25	-47168	512000390	ANATOKI	Cargo
10215	-41175	512000322	TIAKI	Tug
26	-36174	512000109	KORAKI	Tug
169	-36175	512000109	KORAKI	Tug
4308	-37175	512000109	KORAKI	Tug
15397	-37175	512000049	WESTPORT	Cement carrier
877	-41174	512000034	ARAHURA	Ferry
1471	-41175	512000034	ARAHURA	Ferry
18276	-37175	319855000	JULIET	Sailing Vessel
192	-37175	269084000	MAERSK JENAZ	Container ship
4038	-41174	235012190	KAITAKI	Ro-ro/passenger ship
3043	-41175	235012190	KAITAKI	Ro-ro/passenger ship
431	-44171	210076000	JRS PEGASUS	Container ship
20523	-44173	210076000	JRS PEGASUS	Container ship
9388	-35174	126282899	AMMONITE	Pleasure Craft

Figure 5.41: A utility designed to study the stationary position of a vessel. This screenshot illustrates the 'Sum of Count' of the rounded value of Location code for each particular vessel.

A query has been created to obtain the results of the above mentioned values. In this analysis, weekend data and two week days of week data is used. The following section describes this screenshot in detail.

Table 5.13 : A query generated to select vessel details and the sum of count to study anchored for a lengthy period of time.

```
SELECT Sum([19_07_loc_dist].CountOfMMSI) AS
SumOfCountOfMMSI, [19_07_loc_dist].loc_code,
[19_07_loc_dist].MMSI

FROM 19_07_loc_dist

WHERE ((([19_07_loc_dist].Speed)<1))

GROUP BY [19_07_loc_dist].loc_code, [19_07_loc_dist].MMSI

ORDER BY [19_07_loc_dist].MMSI DESC;
```

The most number of sumOfcountOfMMSI column is observed and selected those rows for investigation. A sample chosen from the result is shown below.

For instance, high speed craft Clipper 3 is observed to be stationed or anchored for a long period of time. Compared with other high speed crafts shows clipper 3 is less utilized. The most frequently used high speed boat has a less number of sum counts in stationary locations. The next day records were more than in 11233 counts, it eventually became clear that the particular vessel has been anchored for a long period of time.

Table 5.14: Analysis on the amount of timed spent Anchored (maximum counts).

Analysis on Anchored for lengthy period of time (maximum counts).					
SumOfCount OfMMSI	**Rounded loc_code**	**Date**	**MMSI**	**Vessel Name**	**Vessel Type**
8928	-37175	7/05/11	512000743	CLIPPER 1	High Speed Craft
2934	-37175	7/05/11	512000742	CLIPPER 2	High Speed Craft
14672	-37175	7/05/11	512000741	CLIPPER 3	High Speed Craft
10215	-41175	7/05/11	512000322	TIAKI	Tug
9388	-35174	7/05/11	126282899	AMMONITE	Pleasure Craft
11233	-37175	8/05/11	512000741	CLIPPER 3	High Speed Craft
10047	-37175	8/05/11	512000109	KORAKI	Tug
3770	-41175	8/05/11	512000034	ARAHURA	Ferry
18182	-37175	8/05/11	319855000	JULIET	Sailing Vessel

Ferry boats were observed to be anchored for a limited period of time or short period of time, as they were used frequently. Another reason might be that these are boats which were scheduled for frequent services. Sailing vessels would need a thorough examination before their next sailing. This could be the reason for a lengthy anchored time.

Table 5.15, on the next page, illustrates the least number of counts of records whose speed is less than 1. Having a speed of less than 1 therefore, this confirms that the vessels are anchored at the dock or harbour.

Table 5.15: Vessels with speed of less than 1 indicating anchored for minimum time period (minimum counts).

Vessels Anchored for short hours (minimum counts).					
SumOfCountOf MMSI	Rounded loc_code	Date	MMSI	Vessel Name	Vessel Type
238	-41174	8/05/11	512201000	SPIRIT OF NEWZEALAND	Passenger
26	-36174	8/05/11	512000109	KORAKI	Tug
4038	-41174	8/05/11	235012190	KAITAKI	Ro-ro/passenger ship
3043	-41175	8/05/11	235012190	KAITAKI	Ro-ro/passenger ship

From the Table 5.13 and Table 5.14, it is unable to make a conclusion on the findings that whether a particular vessel type is normally anchored for such a period of a time. Otherwise, the anchored time of these vessels could be based on services they provide or it could be on that particular day's event. It is unable to conclude based these findings however, it sheds some insight on type of vessels been at anchored at harbour or at ports.

5.6 Summary

This chapter has focused on the analysis conducted on the trajectories of moving vessels. The analysis was performed rigorously and was to the motivated by a desire to answer to the research questions. One of the key rationales for conducting such a rigorous analysis was to characterise the different types of vessels. Different approaches were chosen to achieve this. Firstly in order to locate moving vessel is trajectory, the decoded data set was exported to ArcGIS for analysis of the plotting data. Secondly, after the trajectory points were plotted using the ArcGIS tool and queried through an attribute table using SQL commands to identify a particular vessel and its trajectory coordinates, the vessel's movements at a location at a particular time were investigated using spatio-temporal analysis. Moreover, plotting on ArcGIS and conducting an analysis using the spatio-temporal approach has provided an insight in to the trajectory of each patterns revealing such as continuous and smooth, discontinued and uneven, short-tripped or itinerant movements. Thirdly, speed versus time has been graphed to study in order to understand the characteristics of the different types of vessel movements. Studying the trend of speed over time has allowed this researcher to comprehend stationary (anchored) period or whether the vessel was moving on (underway). Fourthly, a thorough study was conducted using the utility developed exclusively to study the anchored period at a particular location. This assisted the researcher to understand the period of time the vessel was being stationary. This analysis was carried out to understand the characteristics of these vessels and in particular how long had the vessel been anchored in relation to their type and therefore their role.

6 Discussion and Addressing Research Questions

This chapter provides a broad discussion on the overall results of this thesis. This chapter also revisits research questions that were posed in Chapter 1.

This research study involves developing the motivation for and design of an artefact to decode raw data, a series of analysis conducted using decoded data and demonstrates with results followed by evaluation. The research study was conducted using the design science methodology as explained by Peffers et al. (2008). Because of the very large datasets AIS data management is complex in areas where large numbers of vessels are being studied. However, this thesis deals with single vessel tracking only.

The outline of this chapter is as follows: Section 6.1 is a detailed discussion on evaluation criteria used, followed by the next Section 6.2 goes back over the methods introduced in the previous chapters together with their findings. The much anticipated part that has been added in order to answer the research questions regarding trajectory patterns is explained in detailed in Section 6.3.3. A detailed discussion on descriptive analysis (Section 6.4), followed by revisiting research questions is discussed in Section 6.5. Later, the challenges encountered (Section 6.6) during this thesis study and strengths of the research methods (Section 6.7) used and analysis are discussed towards the end of this chapter.

6.1 Research Questions and Criteria

The main research question is to optimize the information from recorded AIS raw datasets (movement data) to perform trajectory analysis were initiated in Chapter 4 and Chapter 5. Chapter 4 is made up of a detailed report on understanding NMEA 0183 format of AIS raw data, designing a decoder using VB.net and then obtaining the generated decoded data. The entire process of decoding the raw data is designed to ensure optimization of the information, which involves conducting trajectory analysis using various evaluation criteria, is Chapter 5. This chapter evaluates the trajectories plotted on ArcGIS of different types of vessels using quantitative analysis on of the speed parameter. The results thus obtained through quantitative analysis of each vessel type served as a milestone in the research study. This is discussed in detail in Section 6.4.

6.2 CB-SMoT Method

Basically two different kinds of clustering approaches were discussed in detail to investigate the trajectory analysis of moving vessels. They have been developed to instantiate the model of stops and moves namely ***CB-SMoT*** and ***IB-SMoT***. Out of these two methods, inspired by ***CB-SMoT*** model was chosen and its significance was discussed in detail in Section 5.1.1. A clustering approach called *Clustering Based Stops and Moves of Trajectories* (CB-SMoT) was applied to vessel types based on speed and is described in Section 5.1.1. The first step is to identify vessels that were anchored. The second step is to perform the trajectory analysis that would assist in investigating a vessel's characteristics. The chosen clustering approach was adopted to use in this study to investigate the **stops** and **moves** movement of the vessels. This characterises the change of the speed during arrivals and departures. This signifies the moment-related characteristics (Andrienko et al., 2008, p.19) of the different type of vessels. Thirdly, the rationale for selecting the CB-SMoT approach was that it allowed the research question to be addressed. This approach is applied thoughtfully in Section 5.5.2 in the spatial analysis used to investigate the facets of the trajectory of the vessels. Furthermore, the **CM-SMoT** approach is also applied to investigate stationary vessels anchored in ports and harbours.

6.3 Main Findings

This section elaborates on the main findings that been inferred the from trajectory analysis conducted using ArcGIS software which was described in Chapter 5.

Movement parameters are measurable quantities in terms of location, speed and time. Movement parameters were categorized into spatial (location-oriented), temporal (time-based) and spatial and temporal dimension. In Section

5.8 a detailed discussion is provided based on the approach of *Clustering Based Stops and Moves of Trajectories* which was incorporated into the spatio-temporal approach. *S*electing this approach has three fold advantages to study the moment-related characteristics of the different type of vessels. First, to observe the duration of each type of vessel spends at a particular location. Secondly, to obtain insight on the movement data of each vessel type during its journey moving. Thirdly, this approach contributes to describing the particular characteristics of a vessel type.

6.3.1 Number of moving objects involved

The analysis was carried out using a week's dataset dated between 7/05/2011 to 13/05/2011 and two randomly selected datasets dated 9/04/2011 and 11/03/2011. A subset of the dataset dated 9th April 2011 was selected for the 2 hour period from 12am to 2am. Subsequently, another subset of data dated 11th March 2011 was also chosen. This second subset covered nearly 2.5 hours of the original dataset. This detailed information was tabulated in Section 5.5.

Table 6.1 : List of vessels identified per day from the provided dataset.

Date	Number of ships/vessels/boats
7/05/2011	80
8/05/2011	85
9/05/2011	81
10/05/2011	95
12/05/2011	82
13/05/2011	86
9/04/2011 12am – 2am	52
11/03/2011 21.30 – 23.59 hours	56

In total 125 vessels (approximately) and 9 different types of vessels were studied. The purpose of and the role played by each type of vessel were discussed in Section 5.6. The nine different types of vessels discussed were fishing vessels, Naval/research vessel, high speed craft, Ro-Ro cargo vessels, pleasure boats, pilot vessels, ferry vessels, passenger boats, ferry boats and container vessels.

6.3.2 Different approaches to analysing Trajectories

As the decoded data was exported to the ArcGIS software, a systematic analysis is carried out to investigate the characteristics of the different types of vessels.

The procedure followed is described below:

1: Movement data was plotted on the ArcGIS tool

Step 1.1: A Query is generated to select a particular vessel type.

2: A spatio-temporal approach was conducted on plotted movement data.

3: A graph is plotted between speed versus time to study the trend.

4: A frequency distribution graph is generated based on the speed of each vessel type.

4.1: An analysis is conducted on the speed profile to study the frequency distribution of each vessel type.

5: An investigation is conducted on the stationary movements of each vessel type.

In steps 3, 4 and 5, the speed parameter has played a key role in characterising a moving vessel. The speed parameter served as a primitive movement parameter. These trajectory types or patterns are produced by the moving objects at a particular instance of time and location.

Various trajectory patterns have been identified during this analysis. Significant patterns such as meeting, concentration, constancy in direction, full synchronization, wandering, isolation, convergence, spatial concentration and moving clusters were noticed.

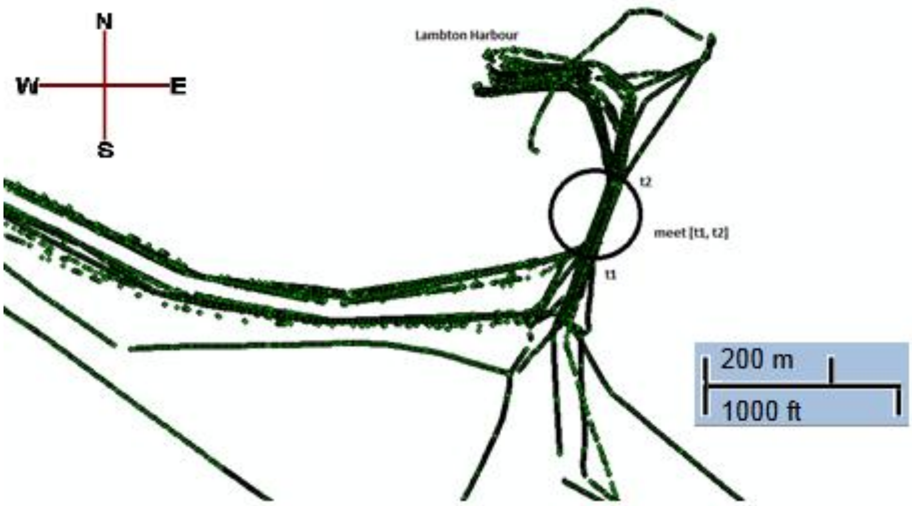
The most popular or frequently used routes observed are between Picton and Lambton Harbour, between Auckland and Beachlands, between Lambton and Nelson Harbour.   The most densely populated harbour was identified as Auckland Viaduct.

6.3.3 Movement Patterns

This section highlights the movement patterns noticed during the investigation.

6.3.3.1 Meet

According to Dodge et al (2009) a pattern that consists of a set of moving objects that stay within a specific radius in a certain time interval between t1 and t2 is termed as a **'meet'**. When the moving objects stay stationary for a certain period of time in clusters, they form a stationary cluster. It is also noticed that there are two types of variants of 'meet' - fixed meet and varying meet. Fixed meets are depending on whether the objects stay together whereas varying meets depending on changes in the meeting region.



.

Figure 6.1: Illustration of Movement pattern described as a 'Meet'.

6.3.3.2 Concentration

It is noticed when moving objects are concentrated into a certain instance of time this is termed a spatial concentration (Dodge et al., 2008).

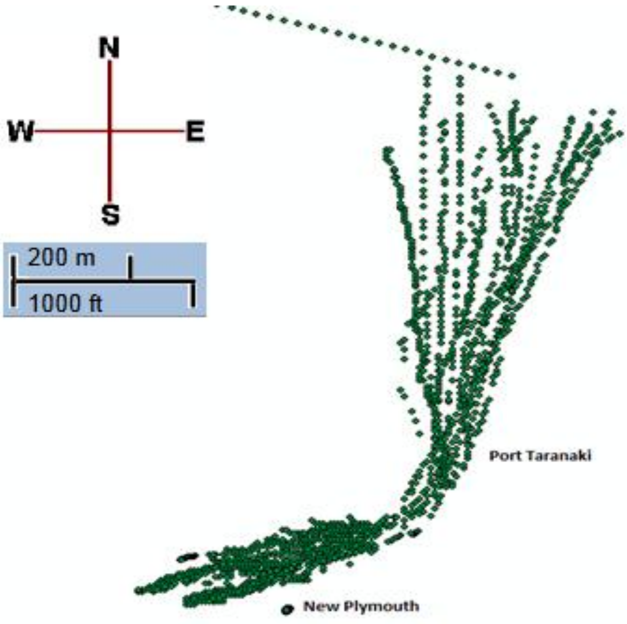


Figure 6.2 : Illustration showing a concentration pattern towards ports.

6.3.3.3 Moving in Clusters

Moving in clusters is described as occurring when sets of objects move along in the same direction close to each other during a specific time interval. It must be noted that, it is not necessary for the group of moving objects to remain the same, but individual objects may enter and leave while the cluster is moving (Figure 6.3).

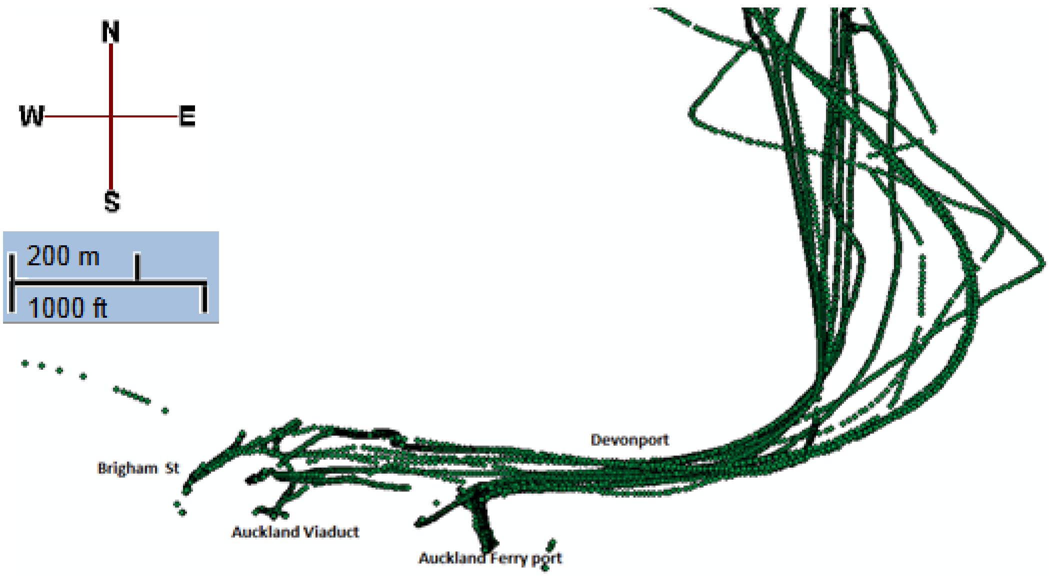


Figure 6.3: Illustration showing vessels moving in clusters. Movement patterns were observed in AUCKLAND viaduct region.

Another type of noticeable movement pattern has also been identified during investigating the trajectory analysis. Some interesting patterns occur such as convergence (near ports), wandering from the trawler or even totally isolated from the trawler.

- The study relieved some of the interesting movement patterns namely, ***moving clusters***, ***concurrence***, ***repetitive and isolated***, ***temporal synchronization***, ***concurrence*** and ***meet***.
- Sometimes, influences of other agents can show an emergence of certain patterns in the domain's specific applications. For instance, analysis conducted on fishing boats revealed that the presence of fish made the vessel make repetitive or isolated patterns.
- It is also inferred that the influence of these patterns can be based on purely spatial or temporal or they can be of mixed of both.

6.3.3.4 Higher density movements towards the ports

It was also observed that more densely populated movements occur at ports and towards or nearby entrances of ports (Figure 6.4).

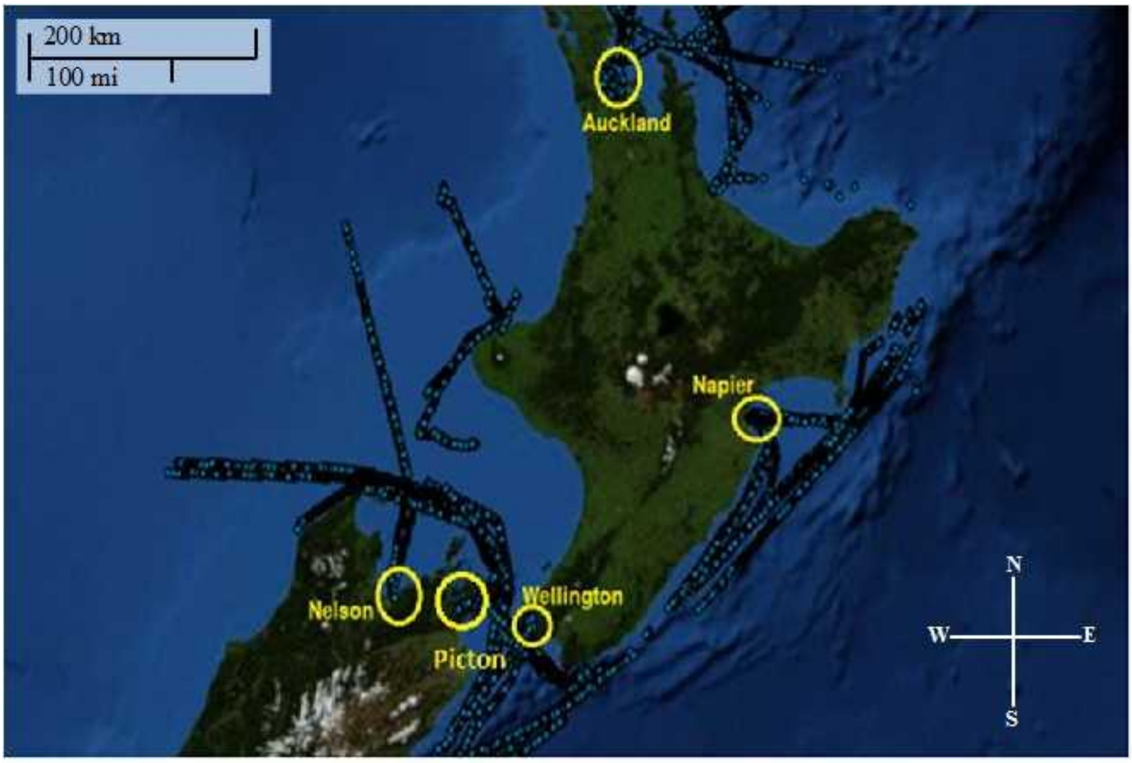


Figure 6.4: Ports with higher density of traffic in some North and South Island ports.

Figure 6.4 shows higher density movements towards the ports and in the ports. From the analysis, it is observed that Auckland is more densely populated than other ports.

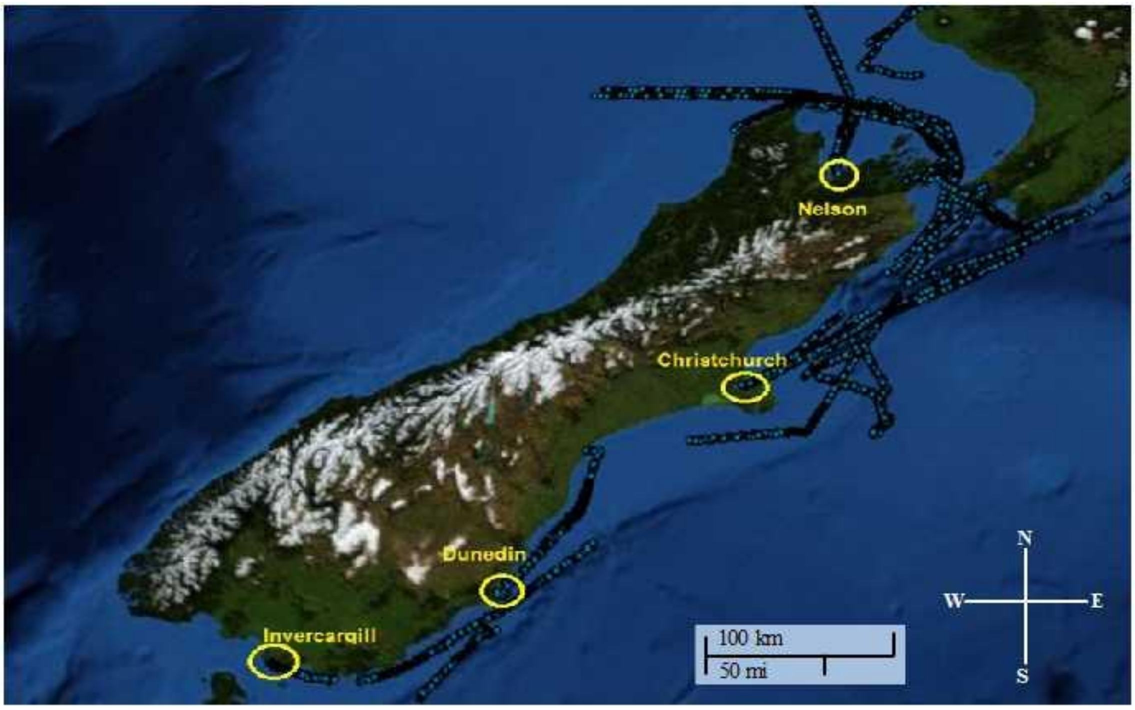


Figure 6.5: Higher density of traffic movements at ports in the South Island.

6.4 Descriptive Analysis

This section attempts to provide a broad discussion on the overall results of datasets utilized in this thesis. This section represents the consolidated results descriptive analysis to study the characteristics of the different type of vessels.

Table 6.2 below represents the minimum, maximum, mean and median were studied elaborately based on a particular one day dataset of vessels already that were discussed Section 5.5.2 and Section 5.5.3.

Table 6.2: Descriptive statistics on the speed shown for different types of vessels.

Type of Vessel	Min	Max	Mean	Median
Ro-Ro (Straitsman)	0	20.3	13.3922	16.7
Pleasure boat (Aquarius)	7.5	9.7	8.474	8.4
Pilot Vessel (Wellington Pilot)	0	17.8	5.67700	0
Fishing (Professor Mikhaylo Aleksandrov)	1.8	7.5	4.483904	4.6
Ferry (Aratere)	0	19.7	15.25395	18.6
Speed boat (Clipper 3)	0	34.5	17.800	4.5

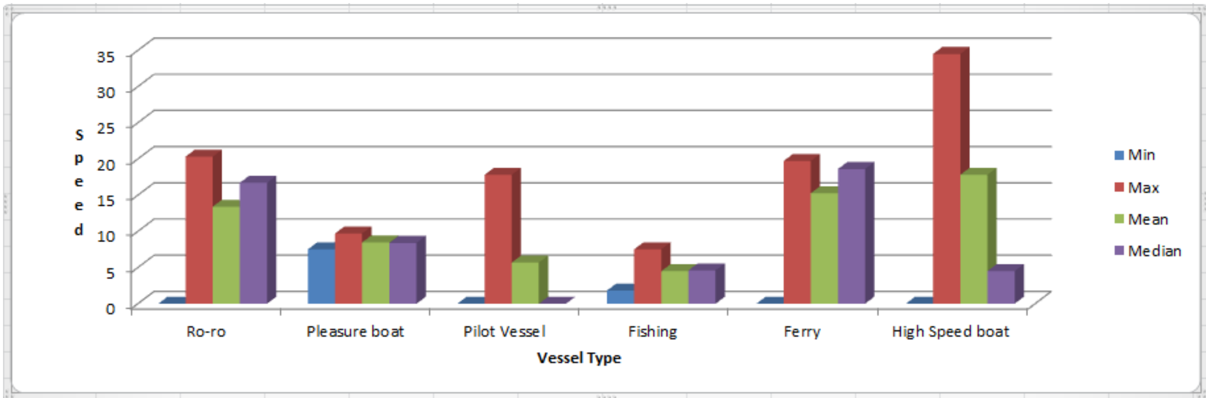


Figure 6.6a: Comparison of the minimum, maximum, mean and median velocity of different vessel types.

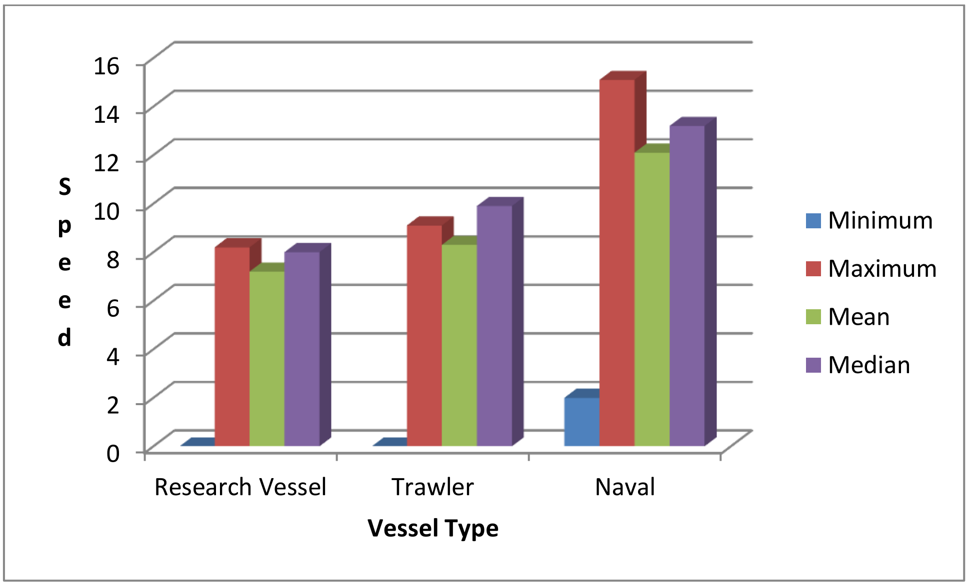


Figure 6.6b: Comparison of the minimum, maximum, mean and median velocity of Research vessel, Trawler, Naval vessel.

The speed parameter was identified as one of the main contributing attributes towards identification of vessel type. From Table 6.6a, the average speed of a speed craft to passenger boat is 3 times faster. From Table 6.6b, it is noticed that the average speed of naval vessel was approximately 14 knots. The research vessel and trawler noticed to be much similar in their speed trend (Figure 6.6b). Although, the maximum and minimum speed of Ro-Ro vessel, Pilot vessel and Ferry vessel were noticed to be approximately 20 knots and 0 knots. The speed profile of the fishing vessel shows approximately 6 knots, this reveals that the vessel very much under fishing.

Table 6.3: Descriptive statistics on the speed shown for Research Vessel, Trawler, Naval of vessels.

Type of Vessel	Minimum	Maximum	Mean	Median
Research Vessel (SONNE)	0	8.2	7.2	8
Trawler (OYANG NO.75)	0	9.1	8.3	9.9
Naval (Resolution Research vessel)	2	15.1	12.1	13.2

6.5 Revisiting the Research Questions

This chapter **revisits** the initial **research questions (Chapter 1)** to re-focus to check on whether the research questions were answered.

RQ1: How can the information from recorded AIS raw data sets (movement data) be optimised when performing trajectory analysis?

The investigation on the above research question started with reviewing the general literature on Automatic Identification System (AIS) the data format of raw data and the attributes involved in raw data. The purpose of the review was to explore the spatio-temporal information from the raw datasets. Chapter 1 outlined the plan for conducting the thesis. Chapter 2 outlined a detailed discussion on trajectory concepts, AIS attributes and knowledge discovery. In order to perform such trajectory analysis, it was necessary to decode the raw data. As the raw data set is compressed information in a NMEA 0183 format are meaningless in its raw form, there was an inevitable need to develop a decoder. A thorough exploration was conducted to study the attributes involved in NMEA sentences. The study showed that there were 16 different attributes involved (Section 4.5). As this research study focused on spatial, temporal and speed parameters the other 13 attributes were ignored. The decoder was identified as a gateway to obtain the movement data in order to perform the trajectory analysis. A decoder was developed using VB.net. The VB.net program was developed so that it could accept large volumes of raw data set in text files, process the raw data to decode it into meaningful data and produce an output file in text file format. Once the decoded datasets were available, it was important to clean, remove the gaps and noise. The decoded attributes were spatial coordinates, temporal facts along with speed parameter formulated as multifaceted attribute. Chapter 4 contains a detailed discussion on the development of the decoder, followed by description of the cleaning process. Once the cleaning process was completed and the data set ready, conducting trajectory analysis involving multifaceted attributes was the next challenge of this research study. Conducting an analysis of the trajectories produced by the vessels using AIS data from the vessels is the ultimate aim of this study. The purpose of the analysis is to characterise the type of vessels by analysing their movement characteristics. During the series of steps taken from data extraction to data analysis, hidden knowledge was discovered. This entire process is termed 'knowledge discovery'. Thus to optimize information from recorded AIS raw data sets (movement data) to conduct trajectory analysis requires a systematic approach. A systematic approach is applied in multi-methodological methodology framework of Design Science.

RQ 1.1: What approach will be more appropriate to perform knowledge discovery from the AIS raw dataset?

Knowledge discovery in moving objects databases is a complex task of data exploration, analysis and interpretation. A comparative study was conducted between different models to decide which would be the most appropriate model for the chosen case study. In the experiment phase, there were various steps involved to extract hidden knowledge from a large AIS raw datasets. The notational model of Klein et al (2006) notional model was selected and applied. A detailed discussion was conducted in Chapter 3 that demonstrated the systematic approach which would be applied to perform knowledge discovery from a large data set. It is apparent that this systematic method has delivered a rigid approach to the adopted methodology. Although the Klein et al (2006) model can be adapted to transform raw data into trajectory data, it has later noticed that the models of Klein et al (2006), Bodnar (2005) and Giannotti et al (2008) were able to be applied to this research study. Therefore, a mixed approach was adopted towards the knowledge discovery process in this study.

RQ 1.2: How can the characteristics of the trajectories of vessels be deduced from raw data using the spatio-temporal model approach?

The process of conducting an analysis of raw data to trajectories of vessel is basically on decoded datasets. The Analysis was conducted by four different methods. The first method was the plotting of movement data that consisted of MMSI, latitude, longitude and speed of the ArcGIS tool to project the trajectories. The second approach was to conduct spatio-temporal analysis with **stop and move approach.** The third approach was to plot velocity trend graphs to study speed versus time, time, followed by a fourth approach which involved generating a frequency distribution graph based on the speed of each vessel type. Finally the fifth approach was to investigate stationary movements of each vessel type. An investigation of the velocity trends studied speed versus time. This approach focused on temporal granularities with respect to movement parameter speed, as these attributes report on movement trends. It was challenging to plot a 3D graph. It was identified to be a challenge to plot a 3D graph of spatial coordinates with temporal and speed parameters. The second approach of applying descriptive methods studying spatial and temporal coordinates with **stop** and **move approach** was chosen to understand the trajectories using more than one approach.

Table 6. 4: Overview of different methods and approach applied in this research study.

SI.No	Name of the Approach	Explanation	Method
1	Plotting on ArcGIS	Plotting movement parameter that involves - MMSI, latitude, longitude, temporal and speed	Visualization
2	Spatio-temporal Analysis	Studying spatial and temporal coordinates with **stop** and **move**	Spatio-temporal, descriptive method
3	Velocity Trend	To study speed versus time	Time versus the movement parameter speed. Graphical method
4	Frequency distribution of Speed	To study the speed of each vessel type	Visualization, Quantitative approach
5	Stationary Movements	To study stationary movements of vessels stayed in port for a certain	Spatio-temporal approach. Quantitative approach.

		time period	

The advantage of using various approaches is to validate the results obtained from one approach and compare this with another approach. The validation provides much more rigid ratification of results. Later, an investigation was conducted on the trajectory patterns where collective movements were studied elaborately. Ultimately, different approaches were used along with the spatio-temporal approach. It was observed that there needs to be a systematic approach in conducting quantitative analysis. Consequently, selecting relevant real movement data at the appropriately selected amount of samples were applied to the proposed methods, to evaluate the results. This analysis provided evidence that each vessel type has different characteristics and different patterns.

RQ 2.1 How well can spatio-temporal data be supported and managed in analysing patterns of moving object characteristics of movement pattern?

Movement parameters are categorized into spatial (location-oriented), temporal (time-based) and spatial and temporal dimension. These movement parameters contribute vital information about moving objects that has positional information with temporal granularities. Therefore, it becomes important to conduct spatio-temporal analysis to understand the characteristics of the moving objects.

The different approaches taken to understand the characteristics of the moving object (vessel) that focused on spatio-temporal analysis and stationary movements at a particular location were dedicated to study a particular vessel's position at a particular location. This analysis is conducted to identify the duration spent in ports or anchored in a harbour. The rationale is to understand the characteristics of these vessels and how long the vessels spend anchored depending on their type and role. For instance, it is observed that passenger boats and ferries spend a minimum amount of time in ports, whereas other types of boats for example pilot vessels, tug and bulk containers stayed much longer in port. Furthermore, tugs and high speed crafts were used based on a specific request or necessity, so the duration spent in **port varied**. Hence this is discussed in Section 5.9.

Although analysis was conducted quantitatively, the movement patterns are demonstrated using visualization, graphical and descriptive method.

To visualise the patterns formed by the moving objects, the visualization approach is taken. This is presented in the ArcGIS tool. The study relieved some of the interesting movement patterns namely, ***moving clusters***, ***concurrence***, ***repetitive and isolated***, ***temporal synchronization***, ***concurrence*** and ***meet (Section 6.3)***. This research has identified and formalized important features characterizing moving objects. The feature of characterizing these moving objects was using the speed parameter of movement. Quantitative methods have been used to extract such features from the raw data with the aim of transforming the trajectories.

The key element of this method was to generate a movement parameter profile (Section 5.8.3 and Section 5.8) based on the function of the movement parameter speed over time. Time, an essential dimension of the movement is embedded and can be accompaniment by analysis to give a better insight about characteristics of the moving object (vessel).

As discussed in Section 6.8 about movement pattern, it is also observed that classification of patterns can be generic (Dodge et al, 2008) and domain-specific patterns.

RQ 2.2 What is the constituent parameters of movement data essential for defining a trajectory pattern?

In response to the above research question, the main objective of this thesis study is to fully utilise movement parameters. Each movement parameter describes the unique dimensions of the moving object. Further it supports the analysis in revealing particular characteristics of such moving objects. Movement parameters are the key quantifiers describing dynamic objects. The methods are applied in this study along with the movement parameter in order to study and understand the characteristics of different kinds of vessel movements. The speed parameter with spatial location coordinates and temporal granularities played a vital part during the analysis. In the analysis phase, there were five different approaches. Each approach was intended help to identify ‘characteristics of the moving objects’ (vessel types) utilizing various methods.

The speed parameter was identified as one of the main contributing attributes towards identification of vessel type. For instance, the average speed of a speed craft to passenger boat is 3 times faster. However, in terms of the spatio-temporal aspect the speed parameter has to be combined along with space and time parameters. This describes the multi-dimensional aspect of a dynamic object. The study of the multi-dimensional characteristics of a trajectory eventually leads to a better understanding of the patterns. A thorough description of the patterns defines the characteristics and individualities of the vessel type. Finally, the speed is not the only parameter that has helped define trajectory pattern, yet space and time parameters have also complimented for the defining of a trajectory pattern.

6.6 Challenges

Exploring and analysing large data sets of movement data was a vital part of this research study.

The first challenge encountered was developing a decoder. As this study have two-phases: the first phase of this research focused on the development of a decoder while the second phase focuses on trajectory analysis. The challenge involved to understand NMEA 0183 format of AIS raw data, designing a decoder using VB.net and then obtaining the generated decoded data. The analysis was conducted on the decoded data set after parsing this data through the decoder. The aim of decoding the raw data into meaningful movement data is the preparation for subsequent analyses. The decoded raw data have provided the insight of hidden information and enhanced the understanding in order to model on trajectories.

The second challenge involved in handling real movement datasets is usually very complex and challenging to do so properly. The challenge mainly includes cleaning and filtering. A commonly encountered issue with real datasets is that they contain a lot of noise and gaps. Moreover, a raw data set adds another level of complexity as these datasets are required to be cleaned after decoding. However, it is important to consider the real movement datasets in the evaluation of knowledge discovery and analysis. In this research study maritime domain (raw) datasets of vessel movements were utilized.

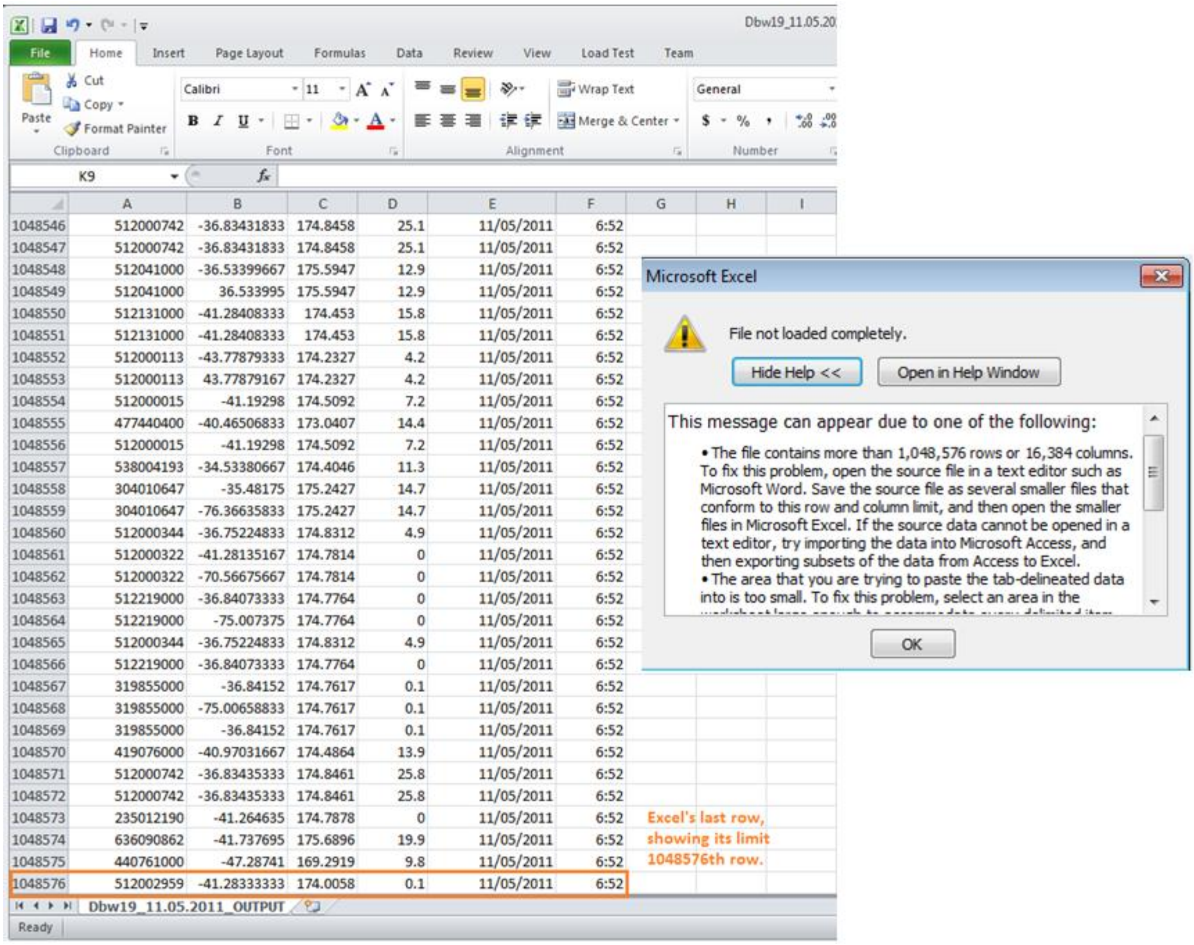


Figure 6.7: File exceeds more than 1048576 rows during export from text file to Excel.

The third challenge encountered was very large movement datasets. During analysis exporting to excel rows exceeded their threshold limit. One of the challenges faced was importing data given that Excel has limited capacity for handling large datasets. During the process of importing (large sized file) decoded text file to Excel was that it had attained its maximum threshold. This occurred while handling the large data set file Dbw_19_11.05.2011 dated 11.05.2011, therefore this file had to be divided into 4 different portions containing approximately 264375 rows each. Each file had approximately 6 hours' worth of data set from the particular day (11.05.2011). Thus large dataset issues were managed at various phases of the analysis.

6.7 Strengths of the method

6.7.1 Movement parameter

The main objective of this thesis is utilization of movement parameters. Each movement parameter describes one dimension of the moving object. Further it supports in revealing the particular characteristics of such moving objects. Therefore, movement parameters are the key quantifiers describing dynamic objects. The methods selected to apply in this study including the movement parameter depends on the aim of characterising different types of vessel and different kinds of movements. For instance, the average speed of high speed boats is 28 knots, while a passenger boat's average speed varied between 8.0 knots - 9.7 knots. Each of these selected parameters can give an insight into one dimension of the movement which characterises each vessel type.

6.7.2 Approach

Clustering Based Stops and Moves of Trajectories (CB-SMoT) is the chosen approach based on the speed variation of the trajectory. This method seems to be more appropriate for use on fishing boats, ferry boats and high speed crafts. The CB-SMoT method is basically used where speed is much lower than a given a given threshold, for instance speed = 0 means the particular vessel is anchored. In addition, it complements additional information such as pattern detection and an activity based profile. According to Rocha et al. (2010) and Palma (2008) this method is useful when the speed plays an essential role.

6.7.3 Real movement datasets

A review of the previous literature (van Kreveld and Luo, 2007; Buchin et al., 2009, Frentzos et al., 2007; Pelekis et al., 2007) only evaluated their theory, frameworks or paradigms on synthetic and simulated data sets. It is obvious the effectiveness of a proposed framework or paradigm on simulated data sets were not sufficiently determined. However, this study used real datasets. The handling of real movement datasets is usually very complex and challenging. The challenge mainly involves cleaning and filtering. A commonly encountered issue with real data set is that they contain a lot of noise and gaps. However, it is important to consider the real movement datasets in the evaluation of knowledge discovery and analysis.

6.7.4 Different approaches in conducting trajectory analysis

Utilization of various methods actually provides a rigorous approach towards evaluation of results. The first method is the plotting of movement data that consist of MMSI, latitude, longitude and speed of ArcGIS tool to project the trajectories. The second approach was concerned with the study of spatial and temporal coordinates with **stops** and **moves** in ***descriptive*** methods. The third approach was to plot velocity trend graphs to study speed versus time, followed by the fourth approach frequency distribution graph generated based on the speed of each vessel type. Finally the fifth approach was to investigate the stationary movements of each vessel type.

6.7.5 Mixed approaches on Knowledge discovery process

It is also observed that in order to perform trajectory analysis from large real-time raw data sets using the mixed model approach is more appropriate. In Chapter 2 various models were discussed such as those of Klein et al(2006), Bodnar (2005) and Giannotti et al (2008). Each model has its own features that are more suitable to apply in the process knowledge discovery.

6.8 Summary

This chapter contains a detailed discussion on the analysis conducted and described in chapter 5. The highlight of this chapter is to list all the findings discussed during the process of analysis. In addition, the various identified patterns of movements projected by dynamic vessels were identified. An evaluation was conducted on speed parameters in descriptive form in order to understand their characteristics in terms of speed. This chapter revisited the research questions that were introduced in Chapter 1 and each question was reviewed to ensure it had been answered. The challenges encountered and strategies taken to overcome those challenges were discussed in Section 6.6. Finally, the strengths of the methods that were applied towards this thesis study were discussed in detail.

7 Research contribution, Limitations, Recommendations and Future Work

This thesis presented research on the optimization of information from recorded AIS raw datasets (movement data) to perform trajectory analysis. In response to the research objectives, a two phase methodological study was conducted. The first phase focused on the development of a decoder to extract perform knowledge discovery from raw movement data about dynamic objects. The second phase centred on conducting trajectory analysis which involved utilizing proposed spatio-temporal approach. The general aim of this chapter was to highlight the main contributions (Section 7.1), the most relevant outcomes identified in this study and draw attention to the outlook on future research (Section 7.5).

7.1 Main Contributions

In response to the research objectives, this methodological research study is conducted in two stages. This methodological study was conducted to perform knowledge discovery about the movement parameters utilised. This research mainly contributes to knowledge discovery from movement data, with the development of a quantitative approach that enables to identify characteristics of vessel type and trajectory patterns to be identified collectively. The methodological study utilises decoded movement parameters directly recorded by new generations of mobile sensor technologies.

As this thesis pursued a two-phased methodological research study towards the optimization of recorded AIS raw datasets to conduct trajectory analysis, each stage accounts for a part of the contributions made towards this investigational study. Accordingly, this thesis has achieved the following contributions.

Phase 1: In this phase a decoder was developed to decode the raw movement data in order to extract implicit movement information of the moving vessel. The decoder was developed using VB.net software based on the concept of file-handlers. This enabled the decoder to accept input and output as text file format. Therefore, text files have the major advantage of being able to be imported to other applications to conduct analysis in other tools. Consequently, it was considered to be efficient and effective throughout the study because they made it easier to handle large datasets. The approach to develop a decoder particularly focused on the AIS NMEA data format. A data model was presented in Section 4.6 based on the analysis of AIS parameters followed by applying an algorithm to explain the series of steps involved in the process of converting from raw movement data to meaningful movement data. Finally, the decoder was developed using VB.net to *extract* hidden information of MMSI, latitude, longitude and speed. This phase actually established an important role in the knowledge discovery process for decoded raw datasets.

Phase 2: This phase exclusively focused on performing the analysis based on decoded datasets. The analysis was conducted using five different approaches along with proposed the spatio-temporal approach. The first method was to plot movement data that consisted of MMSI, latitude, longitude and speed on ArcGIS tool to project the trajectories, plotting velocity trend graphs to study speed versus time, spatio-temporal approach with stops and moves clusters. Frequency distribution graphs were generated based on the speed of each vessel type, and an investigation into the stationary movements of each vessel type was also conducted. An investigation on trajectory patterns, velocity trends, frequency distribution on speed and stationary movements were conducted. Later, an investigation was conducted on trajectory patterns where collective movements were studied elaborately.

It is necessary to note that this thesis study was conducted using the above mentioned methods on real movement dataset of a maritime domain. Sample data included movement data of different types of vessels recorded between March 2011 and May 2011. The experimental result of this thesis indicates that the proposed methods could be successfully applied in support of performing trajectory analysis of dynamic objects.

7.2 Insights

In response to the research questions, research was conducted using an iterative process (**Vaishnavi & Kuechler, 2008; Peffers et al. 2008)** of awareness of the problem, development of an artefact (decoder), and demonstrated with evaluation. The actual research process helped to reveal valuable hidden information through discovering knowledge from real movement data (AIS dataset) of a maritime domain. The real datasets were AIS raw datasets from different types of dataset.

This research study started off with the initial proposal of choosing either one of a number of knowledge discovery approaches. The selected knowledge discovery approaches were Klein et al (2006), Bodnar (2005) and Giannotti et al (2008). However, after starting the experiment it was observed that more than one approach was required to conduct the knowledge discovery process effectively. The mixed approach was taken into consideration due to the complexity of the study as it involved handling of very large real movement datasets.

It was discovered that there needed to be a systematic approach in conducting quantitative analysis. Consequently, real movement data was selected to reduce the data-handling issues with a subset of samples for evaluation. The analysis provided evidence that each vessel type had different characteristics of vessels and different patterns. For instance, movement patterns of fishing vessels were isolated and repetitive, whereas the movement pattern of ferries or passenger boats have pre-assigned paths between ports.

One major insight gained from this was that thesis movement parameters can enrich exploration in studying and knowledge discovery approaches with valuable information in understanding movements. The valuable information from movement data were obtained from positional temporal granularities and speed. Various approaches provided a detailed insight on characteristics of vessel movement and thus formed trajectories. Hence, an investigation was conducted on trajectory patterns where collective movements were studied in detail.

Furthermore, working with a large set of real movement data during this research, it was observed that there were complications and pitfalls. The real movement datasets were raw data. Out of the large dataset, a subset of data (Shoebox) was selected that was relevant for decoding, processing and then conducting analysis. The main pitfall observed was the lack of benchmark rules in the selection of samples from large datasets and confidence that the subsets were truly representative.

7.3 Comparison with Other Studies

Most studies conducted on moving objects have concentrated on behaviour and movement similarity assessments. However, this study focused on a different perception of moving objects as described to conduct trajectory analysis and identifying characteristics of each type of vessel.

Laube et al. (2004; 2007) in his works has utilized three movement parameters of position, time and direction and these were involved in the formalization of a flock pattern. Dodge et al. (2011a) discussed their work with movement parameters such as speed, azimuth, time, turning angle and acceleration.

Laube at al. (2007) has suggested four lifeline context operators are needed to quantify movement parameters in order to find movement patterns. Dodge et al (2008; 2011) has proposed two groups of movement parameters, namely primitive and secondary derivatives. The primary derivatives such as position and time, are ***instantaneous*** operators, Dodge et al (2011). This instantaneous operator directly helps in measuring the position of the moving object. The advancement in sensor technology has helped to find derivatives such as speed, acceleration and direction promptly.

Contrary to Dodge et al (2011) work, Spaccapietra et al. (2008) does not include movement parameters. Their work movement is modelled as a sequence of ***moves*** and ***stops.*** These ‘moves’ and ‘stops’ movements are considered as events and consist of two facets, namely a geometric facet and a semantic facet. A Geometric facet describes the spatio-temporal particulars of the moving object and the semantic facet describes the application-oriented and semantic characteristics features of the positional details.

The highlight of this research work is the involvement of the movement parameters such as speed, positional coordinates and time. In addition, **stops** and **moves** sequences which prove to be more effective in defining complex patterns such as moving objects. Consequently, the semantic facet helps to identify events (start, stop and move) and characters along the trajectories. Such information can be useful in the identification of the characteristics of moving vessel types. However, the semantic facet is not elaborated, as it is not within the scope of this thesis study.

The definition of a pattern is highly influenced by two aspects: spatial scale and temporal granularity (Wood and Galton, 2009). Dodge et al (2011) have used spatial scale and temporal granularity. Sampling granularity controls the 'trade-off' between the amounts of information used in an analysis. Especially while dealing with a large real-time AIS dataset with 30 Million rows in one file, there are challenges involved in handling the analysis. Originally 13 files each with 30 million rows were provided, it would have been a time consuming task to perform an analysis on such large files of 2 GB (approximately). Sampling of a subset was used in conducting scientific studies with large datasets. During the sample selection, choosing the appropriate amount of information was very important. Under sampling may cause information loss whereas oversampling may cause information redundancy. Therefore, it affects the interpretation of the analysis. After careful consideration, spatial scale and temporal granularity were considered to be important facets in the identification a of vessel type and its movement patterns.

This is an early attempt to study trajectories exclusively focused in New Zealand waters. The motivation was to identify patterns over time, patterns over moving objects (vessels); identify the moving object (vessel) type and characteristics of the moving object. Hence relationships between patterns reveal some of the dimensions and different patterns of each vessel category were not investigated.

Andrienko and Andrienko (2007) have recognized three categories of patterns: similarities, differences and arrangements. Dodge et al (2011) have classified these as generic and behavioural patterns. Wood and Galton (2009) have introduced classification as collectives based on category. In this thesis study on selection of vessels was performed from the collection as whole as a set and number of objects. The relationship is made between the number of objects and the number of rows recorded on that particular day.

7.4 Limitation of the Study

This section details the various limitations of the study. The raw dataset used in this thesis has recorded AIS vessel movement data between March 2011 and May 2011. Although the main focus of the study is to analyse movement patterns it also endeavours to explicitly capture the space-time structure in data in order to meaningfully analyse the dynamic objects. This thesis report is not a detailed study on the obtained dataset due to time and scope constraints. Moreover, the research questions investigation is limited to a specific domain. Furthermore, other domains are not investigated. Some of the other limitations and pragmatic constraints are:

7.4.1 Other parameters of AIS raw data

The various identified occurrences (Section 6.3.3) of patterns can be influenced by a number of various factors which are not within the scope of this study. However, the patterns characterise the type of trajectory the vessel produced in conducting their study. Dodge et al (2008) discussed one of the factors as the mass of the moving object. This thesis study has not included the mass parameter in discussion. Although at present, the discussion has ignored mass parameter due to scope and space constraints, there are probabilities to include this parameter in future studies.

In future, some of the important attributes such as rate of turn and direction will be considered to get more insight into the characteristics of moving objects and their patterns.

7.4.2 Other influencing factors

These can include such as spatial constraints, as topography and natural barriers as well as tidal effects and changes in the weather. Yet again these aspects are not covered in this study, although these might be considered in future studies.

7.4.3 Limited to maritime domain

This research has focused on only the maritime domain. This was driven by the availability of large datasets in this area.

7.4.4 Artefact Designed: Decoder

The artefact was designed with a limited purpose in mind. Its chief role was to decode movement data in order to exhibit MMSI, spatial coordinates of latitude and longitude, temporal granularity details and speed. This designed decoder does not support other parameters found in the decoding process.

7.4.5 Type of Vessel

In this thesis, 16 different types of vessels were introduced in Section 5.6, however only 9 different types of vessels were considered for analysis (Section 5.8.3 and 5.8.4) and discussed in Chapter 6. The 9 different types of vessels were High Speed Craft, Passenger Boats, Ferries, Pilot Vessels, Pleasure Boats, Naval/research Boats, Fishing Boats, Container Ships, Ro-Ro Cargo vessels were identified. Although other types of vessels were also identified, they were not included in the discussion due to time constraints. The other types of boats were reefer, wood chip carrier, oil products tanker, multi-purpose offshore vessels, law enforcement vessels, cement carriers, livestock carriers and suction hopper dredger. The previously mentioned types of vessels might be considered in future studies.

7.5 Outlook for Future Research

The main objective was to conduct a spatio-temporal analysis which endeavoured to explicitly capture the space-time structure in data in order to meaningfully analyse dynamic objects. Various approaches have been used to carry out the analysis the trajectories and to conduct movement pattern analysis. Although, the study was confined to the maritime domain, the following Section may present opportunities for future research as described in the Sections between 7.3.1 to 7.3.4.

7.5.1 Other parameters of AIS raw data

As discussed earlier in Section 7.4.1, if mass, rate of turn and direction parameters these vital parameters can provide a profound insight to study and understand were included the characteristics of the vessel types. Rate of turn and direction can be obtained from the provided raw dataset whereas extracting the mass parameter may not be implicit and only gained explicitly.

7.5.2 Other influencing factors

Spatial constraints namely topography and natural barriers such as tidal effects and changes in the weather can influence movement data. Yet again this kind of information has to be collected explicitly and processed with movement data. Though GIS applications have capabilities to handle diverse data, it is necessary to pre-process prior in an investigational study. Major challenges will continue to occur around merging and pre-processing methods.

7.5.3 Limited to maritime domain

As discussed earlier, this study has handled a raw AIS dataset raw. Dodge et al. (2008, 2011) has concentrated on multi-domain datasets such as eye movement, hurricanes and transportation. Another attempt to exploit movement data collected from Global Positioning Systems (GPS) and obtained from vehicles will be experimented on to analyse the trajectory characteristics of vehicles on land.

7.5.4 Automation Process

In future studies, the clustering technique will be performed by using an appropriate algorithm. This will allow larger datasets to be scanned for examples of unusual or unexpected behaviour, and should assist in data reduction.

7.5.5 Artefact

In future studies, the extraction of the other sets of parameters will be considered strongly. The other set of parameters would be rich source of information to conduct other analytical study.

7.6 Recommendations

7.6.1 Usage of real movement datasets

From the review of the previous literatures (van Kreveld and Luo, 2007; Buchin et al., 2009, Frentzos et al., 2007; Pelekis et al., 2007), it is obvious the effectiveness of proposed framework or paradigm is not sufficiently determined due to usage of simulated movement datasets. This research study is a practical evidence to demonstrate that it is possible to perform knowledge discovery out of raw spatio-temporal data. It seems reasonable to suppose that, it is a good practice to conduct knowledge discovery from real-movement dataset.

7.6.2 Mixed Approach towards knowledge discovery process

It is observed during this research study that the mixed approach was necessary towards the knowledge discovery process. As real datasets are complex to handle, it is recommended to utilise a mixed approach in the knowledge discovery process.

7.6.3 Mixed approach towards analysis

A complex research method that involves a large volume of real movement datasets, it is a good practice to present results using different approaches.

7.6.4 Choosing relevant movement parameters

It is also identified that use of the relevant movement parameters is critically important. As relevant movement attributes contribute knowledge discovery process. The evidences were noticed in various other research works (Buchin et al., 2009, Frentzos et al., 2007; Pelekis et al., 2007; Dodge et al. 2009). This study also experienced and acknowledged the fact that movement parameters are highly influential in trajectory analysis of any moving object.

7.6.5 Requires Methodological Approach

A systematic approach is an important strategy in a research project or study that involves various phases, methods, process and steps. Therefore, it is highly recommended to follow systematic that has complex process involved.

7.6.6 Requires Benchmark in samples

It is important and highly recommended to set a benchmark in selecting sample datasets from the large datasets to perform trajectory analysis.

7.7 Conclusion

The main objective of this thesis was to **optimize the information from recorded AIS raw datasets (movement data) to perform trajectory analysis**. This involved a two phased study. The phase 1 involved research on the optimization of information from recorded AIS raw datasets (movement data). Later the process involved the utilization of movement data to perform trajectory analysis. The movement data used in this thesis were real datasets provided by Kordia New Zealand. In response to the other research objectives, a two phase methodological study was conducted. The first phase focused on the development of a decoder to extract perform knowledge discovery from raw movement data about dynamic objects. The second phase centred on trajectory analysis of utilizing proposed clustering techniques. The systematic and methodological approach taken to conduct this research was able to answer the research questions framed in the Chapter 1. The experimental results of this thesis indicate that the proposed methods could be successfully applied in support of performing trajectory analysis on other dynamic objects. This research has identified and formalized important features characterizing the moving objects. The feature of characterizing these objects was using the speed parameter of movement. Quantitative methods have been used to extract such features from the raw data with the aim to conduct trajectory analysis although still conveying the important movement features.

Although this thesis has outlined its findings, contributions, recommendations and challenges in contrast it has its own limitations. However, this research identifies the opportunity to optimize information from recorded AIS raw datasets (movement data) to conduct trajectory analysis, in particular to have applied the analysis in New Zealand waters. It is to be noted that AIS is not a complete record of every vessel, as not all vessels use this technology and the devices may not always be turned on.

References:

Abraham, T., & Roddick, J. F. (1999). Survey of Spatio-Temporal Databases. *Springerlink, GeoInformatica., 3*(1), 61-99. doi:10.1023/A:1009800916313

Adrienko, N., & Adrienko, G. (2011). Spatial Generalization and Aggregation of Massive Movement Data. *IEEE Transactions on Visualization and Computer Graphics, 17*(2), 205-219. doi:10.1109/TVCG.2010.44

Aken, J. E. V. (2005). Management Research as a Design Science: Articulating the Research Products of Mode 2 Knowledge Production in Management. *British Journal of Management, 16*(1), 19-36.

Alturki, A., Gable, G. G., & Bandara, W. (2011). A Design Science Research Roadmap. *Springerlink, 6629*(2011), 107-123. doi:10.1007/978-3-642-20633-7_8

Alvares, L. O., Bogorny, V., Kuijpers, B., Fernandes, J. A., Moelans, B., & Vaisman, A. (2007). Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems Symposium conducted at the meeting of the Geographic Information Systems doi:10.1145/1341012.1341041

Andrienko, G., Andrienko, N., & Wrobel, S. (2007). Visual analytics tools for analysis of movement data. *ACM, 19*(2). doi:10.1145/1345448.1345455

Andrienko, N., Andrienko, G., Pelekis, N., & Spaccapietra, S. (2008). Basic Concept of Movement Data. In F. Giannotti & D. Pedreschi (Eds.), *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery* (pp. 30-32). Berlin, Heidelberg: Springer-Verlag.

Armstrong, M. P. (1988). Temporality in spatial databases.*American Society for Photogrammetry and Remote Sensing.* Symposium conducted at the meeting of the GIS/LIS: Accessing the World, Falls Church, Virginia, USA.

Badard, Y., Larrivee, S., Proulx, M.-J., & Nadeau, M. (2004). Modeling geospatial databases with plug-ins for visual languages: A pragmatic approach and the impacts of 16 Years of research and experimentations on Perceptory. In *Conceptual Modeling for Advanced Application Domains* (Vol. 3289 pp. 17-30). New York: Springer: Berlin Heidelberg.

Baglioni, M., Macêdo, J. A. F. d., Renso, C., Trasarti, R., & Wachowicz, M. (2009). Towards Semantic Interpretation of Movement Behavior *Advances in GIScience - Springerlink*, 271-288. doi:10.1007/978-3-642-00318-9_14

Beller, A., Giblin, T., Le, K. V., Litz, S., Kittel, T., & Schimel, D. (1991). *A temporal GIS prototype for global change research.* Paper presented at the meeting of the GIS/LIS: Accessing the World, Atlanta, Georgia, USA.

Bodnar, J. W. (2005). Making Sense of Massive Data by Hypothesis Testing Symposium conducted at the meeting of the International Conference on Intelligence Analysis

Borges, K. A. V., Davis, C. A., & Laender, A. H. F. (2001). OMT-G: An object-oriented data model for geographic applications. *GeoInformatica, 5* (3), 221-260.

Brillinger, D. R., Preisler, H. K., Ager, A. A., & Kie, J. G. (2004). An exploratory data analysis (EDA) of the paths of moving animals. *Journal of Statistical Planning and Inference, 122*(1-2), 43–63. doi:http://dx.doi.org/10.1016/j.jspi.2003.06.016

Brodeur, J., Bédard, Y., & Proulx, M.-J. (2000). *Modelling geospatial application databases using UML-based repositories aligned with international standards in geomatics*. Paper presented at the meeting of the Proceedings of the 8th ACM international symposium on Advances in geographic information systems doi:10.1145/355274.355280

Buchin, K., Buchin, M., Kreveld, M. V., & Luo, J. (2009). *Finding long and similar parts of trajectories.* Paper presented at the meeting of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems doi:10.1145/1653771.1653813

Çag˘das, V., & Stubkjær, E. (2011). Design research for cadastral systems. *Elsevier Computers, 35*(2011), 77-87. doi:10.1016/j.compenvurbsys.2010.07.003

Camossi, E., Bertolotto, M., Bertino, E., & Guerrini, G. (2003). A multigranular spatiotemporal data model. Symposium conducted at the meeting of the 11th ACM international symposium on Advances in geographic information systems doi:10.1145/956676.956689

Chen, S. (2011). Visual Modeling and Representations of Spatiotemporal Transportation Data: An Object-Oriented Approach. Symposium conducted at the meeting of the 2011 International Symposium on Computer Science and Society (ISCCS), Kota Kinabalu. doi:10.1109/ISCCS.2011.70

Chen, Y., Rao, F., Yu, X., & Liu, D. (2003, 21–24 January 2003). CAMEL: A Moving Object Database Approach for Intelligent Location Aware Services *Springer-Verlag.* Symposium conducted at the meeting of the 4th IEEE International Conference on Mobile Data Management (MDM'03), Melbourne, Australia. doi:10.1007/3-540-36389-0_23.

Chomicki, J., Toman, D., & Böhlen, M. H. (2001). Querying ATSQL databases with temporal logic. *ACM Transactions on Database Systems (TODS), 26*(2), 145 - 178. doi:10.1145/383891.383892

Cole, R., Purao, S., Rossi, M., & Sein, M. (2005). Being Proactive: Where Action Research Meets Design Research Symposium conducted at the meeting of the International Conference on Information Systems (ICIS), Las Vegas, Nevada, USA.

Demsar, U., & Virrantaus, K. (2010). Space–time density of trajectories: exploring spatio-temporal patterns in movement data. *International Journal of Geographical Information Science, 24*(10), 1527-1542. doi:10.1080/13658816.2010.511223

Ding, H., Trajcevski, G., & Scheuermann, P. (2008 16-18 June 2008). *Efficient Similarity Join of Large Sets of Moving Object Trajectories*. Paper presented at the meeting of the TIME '08. 15th International Symposium on Temporal Representation and Reasoning., Montreal, QC doi:10.1109/TIME.2008.25

Dodge, S., Weibel, R., & Forootan, E. (2009). Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems, 33*(1), 419-434. doi:10.1016/j.compenvurbsys.2009.07.008

Dodge, S., Weibel, R., & Laube, P. (2011, 27 April 2011 - 29 April 2011). Trajectory similarity analysis in movement parameter space Symposium conducted at the meeting of the 19th GIS Research UK Annual Conference 2011, Portsmouth.

Dodge, S., Weibel, R., & Lautenschütz, A.-K. (2008). Towards a taxonomy of movement patterns. *Journal Information Visualization 7*(3), 240-252. doi:10.1057/palgrave.ivs.9500182

Eide, M. S., Endresen, O., Brett, P. O., Ervik, J. L., & Røang, K. (2007). Intelligent ship traffic monitoring for oil spill prevention: risk based decision support building on AIS. *Marine Pollution Bulletin, 54*(2), 145-148. doi:10.1016/j.marpolbul.2006.11.004

Erwig, M., Güting, R. H., Schneider, M., & Vazirgiannis, M. (1999). Spatio-Temporal Data Types: An Approach to Modeling and Querying Moving Objects in Databases. *Journal of Geoinformatica, 3*(3), 269–296. doi:10.1023/A:1009805532638.

Erwig, M., & Schneider, M. (1999, 1-3 September). *Developments in spatio-temporal query languages*. Paper presented at the meeting of the Proceedings of 10th International Workshop on Database and Expert Systems Applications. , Florence, Italy. doi:10.1109/DEXA.1999.795206

ESRI. (2010). Retrieved 29th Oct 2012, from http://www.esri.com/

Etienne, L., Devogele, T., & Bouju, A. (2010). *Spatio-temporal trajectory analysis of mobile objects following the same itinerary*. Paper presented at the meeting of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.,

Feixiang, Z. (2011). *Mining ship spatial trajectory patterns from AIS database for maritime surveillance*. Paper presented at the meeting of the 2nd IEEE International Conference on Emergency Management and Management Sciences (ICEMMS), Beijing doi:10.1109/ICEMMS.2011.6015796

Frentzos, E., Gratsias, K., Pelekis, N., & Theodoridis, Y. (2007). Algorithms for Nearest Neighbor Search on Moving Object Trajectories. *GeoInformatica, 11*(2), 159-193 doi:10.1007/s10707-006-0007-7

Friis-Christensen, A., & Jensen, N. T. C. S. (2001, 9–10 November). Requirements and research issues in geographic data modeling. *ACM Press.* Symposium conducted at the meeting of the 9th International Symposium on Advances in Geographic Information Systems (ACMGIS'01), Atlanta, Georgia, USA.

Giannotti, F., Nanni, M., Pedreschi, D., & Pinelli, F. (2009). Trajectory pattern analysis for urban traffic. Symposium conducted at the meeting of the the Second International Workshop on Computational Transportation Science. doi:10.1145/1645373.1645381

Giannotti, F., Nanni, M., Pinelli, F., & Pedreschi, D. (2007, August 12-15, 2007). *Trajectory pattern mining*. Paper presented at the meeting of the Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, San Jose, California, USA

Giannotti, F., & Pedreschi, D. (2008). Mobility, Data Mining and Privacy: A Vision of Convergence. In F. Giannotti & D. Pedreschi (Eds.), *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery.* (pp. 3-4): Springer Publishing Company.

Gibson, M., & Arnott, D. (2007). The Use of Focus Groups in Design Science Research. Symposium conducted at the meeting of the Proceedings of the 18th Australasian Conference on Information Systems, University of Southern Queensland, Toowoomba, Australia.

Gidófalvi, G., & Pedersen, T. B. (2009). Mining long, sharable patterns in trajectories of moving objects. *GeoInformatica, 13*(1), 27-55. doi:10.1007/s10707-007-0042-z

Griffiths, T., Fernandes, A. A., Paton, N. W., & Barr, R. (2004). The tripod spatio-historical data model. *Data & Knowledge Engineering, 49*(1), 23 - 65. doi:10.1016/j.datak.2003.08.003

Grumbach, S., Rigaux, P., & Segoufin, L. (2001). Spatio-Temporal Data Handling with Constraints. *Journal Geoinformatica, 5*(1). doi:10.1023/A:1011464022461

Grumbach, S., Su, J., & Tollu, C. (1995). Linear constraint query languages expressive power and complexity. *Springerlink, 960*(1995), 426-446. doi:10.1007/3-540-60178-3_96

Gudmundsson, J., Kreveld, M. V., & Speckmann, B. (2007). Efficient detection of motion patterns in spatiotemporal data sets. *GeoInformatica, 11*(2), 195–215.

Guo, D., Liu, S., & Jin, H. (2010). A graph-based approach to vehicle trajectory analysis. *Journal of Location Based Services, 4*(3-4), 183-199. doi:10.1080/17489725.2010.537449

Hevner, A., & Chatterjee, S. (2010). Introduction to Design Science Research In *Design Research in Information Systems* (Vol. 22, pp. 1-8). doi:10.1007/978-1-4419-5653-8_1

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quaterly, 28*(1), 75-105.

Horne, J. S., Garton, E. O., Krone, S. M., & Lewis, J. S. (2007). Analyzing animal movements using brownian bridges. *Ecological Society of America, 88*(9), 2354–2363. doi:doi:10.1890/06-0957.1

Hornsby, K. S., & Cole, S. (2007). Modeling Moving Geospatial Objects from an Event-based Perspective. *Transactions in GIS, 11*(4), 555–573. doi:10.1111/j.1467-9671.2007.01060.x

Huang, B., & Claramunt, C. (2002). Stoql: An ODMG-based spatio-temporal object model and query language. Symposium conducted at the meeting of the 10th International Smposium on Spatial Data Handling(SH-02).

IALA. (2004). *IALA Technical clarifications on recommendation ITU-R M. 1371-1*: International Association of Marine Aids to Navigation and Lighthouse Authorities. Retrieved from http://www.iala-aism.org/iala/publications/publications.php?LeTypePub=13

IMO. (2011). *International Maritime Organization*. Retrieved from http://www.imo.org/ourwork/safety/navigation/pages/ais.aspx

ISO/TC211. (2002). *ISO/TC211: Geographic Information and Temporal Schema. ISO 19108:2002*.

ISO/TC211. (2003). *ISO/TC211: Geographic Information and Spatial Schema. ISO 19107:2003.*

Jay F. Nunamaker, J., Chen, M., & Purdin, T. D. M. (1991). Systems development in information systems research. *Journal of Management Information Systems, 7*(3), 89-106.

Jensen, C. S., Kligys, A., Pedersen, T. B., & Timko, I. (2004). Multidimensional data modeling for location-based services. *The International Journal on Very Large Data Bases, 13*(1). doi:10.1007/s00778-003-0091-3

Jiacaia, P., Qingshan, J., Jinxing, H., & Zheping, S. (2012). An AIS data Visualization Model for Assessing Maritime Traffic Situation and its Applications. Symposium conducted at the meeting of the 2012 International Workshop on Information and Electronics Engineering

Keys, P. (2007). Developing a Design Science for the Use of Problem Structuring Methods. *Springerlink, 20*(4), 333-349. doi:10.1007/s11213-007-9066-4

Khatri, V., Ram, S., & Snodgrass, R. T. (2004). Augmenting a Conceptual Model with Geospatiotemporal Annotations. *IEEE Transactions on Knowledge and Data Engineering, 16*(11), 1324 - 1338. doi:10.1109/TKDE.2004.66

Klein, G., Moon, B., & Hoffman, R. R. (2006). Making Sense of Sensemaking 2: A Macrocognitive Model. *IEEE Intelligent Systems, 21*(5), 88 - 92. doi:10.1109/MIS.2006.100

Kreveld, M., & Luo, J. (2007). The Definition and Computation of Trajectory and Subtrajectory Similarity. Symposium conducted at the meeting of the Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems (GIS '07) doi:10.1145/1341012.1341068

Kubach, K. M., Scott, M. C., & Bulak, J. S. (2011). Recovery of a temperate riverine fish assemblage from a major diesel oil spill. *Wiley Online Library -Freshwater Biology, 56*(1), 503–518. doi:10.1111/j.1365-2427.2010.02517.x

Kuechler, B., & Vaishnavi, V. (2011). Promoting Relevance in IS Research: An Informing System for Design Science Research. *International Journal of an Emerging Transdiscipline, 14*(1), 125-138.

Kuechler, W., & Vaishnavi, V. (2008). The emergence of design research in information systems in North America. *Journal of Design Research, 7*(1), 1-16. doi:10.1504/JDR.2008.019897

Langran, G., & Chrisman, N. R. (1988). A Framework For Temporal Geographic Information. *Cartographica, 25*(3), 1-14. doi:10.3138/K877-7273-2238-5Q6V

Larrivée, S., Bédard, Y., & Pouliot, J. (2005). How to Enrich the Semantics of geospatial databases by properly expressing 3D objects in a conceptual model.*Springer.* Symposium conducted at the meeting of the On The Move to Meaningful Internet Systems Workshops. doi:10.1007/11575863_122

Laube, P., Dennis, T., Forer, P., & Walker, M. (2007). Movement beyond the snapshot – Dynamic analysis of geospatial lifelines. *Computers, Environment and Urban Systems, 31*(5), 481-501. doi:10.1016/j.compenvurbsys.2007.08.002

Laube, P., & Purves, R. S. (2006). An approach to evaluating motion pattern detection techniques in spatio-temporal data. *Computers, Environment and Urban Systems, 30*(3), 347–374. doi:http://dx.doi.org/10.1016/j.compenvurbsys.2005.09.001

Laube, P., & Purves, R. S. (2011). How fast is a cow? Cross-Scale Analysis of Movement Data. *Transactions in GIS, 15*(3), 401–418. doi:10.1111/j.1467-9671.2011.01256.x

Lee, J.-G., Han, J., Li, X., & Gonzalez, H. (2008). TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering. *Journal Proceedings of the VLDB Endowment, 1*(1), 1081-1094. doi:10.1145/1453856.1453972

Lei Zhao, P. J., Lanlan Zhang, Huaishuai Wang and Sheng Lin. (2011). Developing an Oracle-Based Spatio-Temporal Information Management System *Springerlink, Volume 6637*(2011), 168-176. doi:10.1007/978-3-642-20244-5_16

Lin, B., & Su, J. (2008). One Way Distance: For Shape Based Similarity Search of Moving Object Trajectories. *SpringerLink, 12*(2), 117-142.

Livari, J. (2007). A Paradigmatic Analysis of Information Systems As a Design Science. *Scandinavian Journal of Information Systems, 19*(2), 39-64.

Macedo, J. A. F. d., Vangenot, C., Othman, W., Frentzos, E., Kuijpers, B., Ntoutsi, I., et al. (2008). Trajectory Data Models. In F. Giannotti & D. Pedreschi (Eds.), *Mobility, Data Mining and Privacy - Geographic Knowledge Discovery* (pp. 123-150). Berlin Heidelberg: Springer-Verlag.

Marxt, C., & Hacklin, F. (2005). Design, product development, innovation: all the same in the end? A short discussion on terminology. *Journal of Engineering Design 16*(4). doi:10.1080/09544820500131169

Meratnia, N. (2005). *Towards database support for moving object databases* (Ph.D Thesis). University of Twente.

Michel, P., Mathieu, R., & Mark, A. F. (2009). Spatial analysis of oblique photo-point images for quantifying spatio-temporal changes in plant communities. *Applied Vegetation Science, 13*(2), 173–182. doi:10.1111/j.1654-109X.2009.01059.x

Miller, W. R. (2004, Feb 2, 2012). *Definition of Design*. Retrieved from http://static.userland.com/rack4/gems/wrmdesign/DefinitionOfDesign1.doc

Montealegre, R., & Keil, M. (2000). De-escalating Information Technology Projects: Lessons from the Denver International Airport. *MIS Quaterly, 24*(3), 417-447. doi:10.2307/3250968

Mou, J. M., Tak, C. v. d., & Ligteringen, H. (2010). Study on collision avoidance in busy waterways by using AIS data. *ScienceDirect, 37*(5-6), 483–490. doi:10.1016/j.oceaneng.2010.01.012

Nanni, M., Trasarti, R., Renso, C., Giannotti, F., & Pedreschi, D. (2010). Advanced Knowledge Discovery on Movement Data with the GeoPKDD system. Symposium conducted at the meeting of the Proceedings of the 13th International Conference on Extending Database Technology. doi:10.1145/1739041.1739129

NAVCEN. (2010). *Automatic Identification System Overview.* : U.S. Coast Guard Navigation Center. Retrieved from http://www.navcen.uscg.gov/

Ostrowski, Ł., Helfert, M., & Hossain, F. (2011). A Conceptual Framework for Design Science Research. *Journal of Springer-Verlag Berlin 90*(1), 345-354.

Ozsu, M. T., & Li, J. Z. (1998). *Modeling and querying multimedia data*. University of Alberta Edmonton, Alta., Canada.

Palma, A. T., Bogorny, V., Kuijpers, B., & Alvares, L. O. (2008). A clustering-based approach for discovering interesting places in trajectories. *ACM.* Symposium conducted at the meeting of the 2008 ACM symposium on Applied computing doi:10.1145/1363686.1363886

Parent, C., Spaccapietra, S., & Zimányi, E. (2006). *Conceptual Modeling for Traditional and Spatio-Temporal Applications: The MADS Approach*. New York: Springer-Verlag, Berlin Heidelberg.

Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2008). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems, 24*(3), 45-77. doi:10.2753/MIS0742-1222240302

Pelekis, N., Kopanakis, I., Ntoutsi, I., Marketos, G., & Theodoridis, Y. (2007, April 17-20). *Mining Trajectory Databases via a Suite of Distance Operators*. Paper presented at the meeting of the Proceedings of the 2007 IEEE 23rd International Conference on Data Engineering Workshop, doi:10.1109/ICDEW.2007.4401043

Perona, P., Porporato, A., & Ridolfi, L. (2000). On the trajectory method for the reconstruction of differential equations from time series. *Springerlink, 23*(1), 13-33. doi:10.1023/A:1008335507636

Peuquet, D. J. (1994). It's about time: A conceptual framework for the representation of temporal dynamics in geopraphic information systems. *Annals of the Association of American Geographers, 84*(3), 441-461. doi:10.1111/j.1467-8306.1994.tb01869.x

Peuquet, D. J., & Duan, N. (1995). An Event-Based Spatiotemporal Data Model ESTDM for Temporal Analysis of Geographical Data. *International Journal of Geographical Information Systems, 9*(1), 7-24. doi:10.1080/02693799508902022

Ping, Y., Xinming, T., & Shengxiao, W. (2008). Dynamic Cartographic Representation of Spatio-temporal data. *International Society for Photogrammetry and Remote Sensing. , 37*(B2), 7-12.

Pirolli, P., & Card, S. (2005). The sensemaking process and leverage points for analyst technology. Symposium conducted at the meeting of the International Conference on Intelligence Analysis (IA'05)

Plumejeaud, C., Mathian, H., Gensel, J., & Grasland, C. (2011). Spatio-temporal analysis of territorial changes from a multi-scale perspective. *International Journal of Geographical Information Science, 25*(10), 1597–1612. doi:10.1080/13658816.2010.534658

Price, R., Tryfona, N., & Jensen, C. S. (2000). Extended Spatiotemporal UML: Motivations, Requirements and Constructs. *Journal of Database Management. Special Issue on UML, 11*(4), 14–27. doi:10.4018/jdm.2000100102

Pries-Heje, J., Baskerville, R., & Venable, J. (2008). Strategies for design science research evaluation Symposium conducted at the meeting of the ECIS Processdings

Raafat, H., Yang, Z., & Gauthler, D. (1994). Relational spatial topologies for historical geographical information. *International Journal of Geographical Information Systems 8*(2), 163-173. doi:10.1080/02693799408901992

Ram, S., Snodgrass, R., Khatri, V., & Hwang, Y. (2001). DISTIL: A design support environment for conceptual modeling of spatio-temporal requirements. *Springerlink:Conceptual Modeling, 2224*(2001), 70-83. doi:10.1007/3-540-45581-7_8

Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N., & Andrienko, G. (2008). Visually driven analysis of movement data by progressive clustering. *Journal Information Visualization, 7*(3), 225-239. doi:10.1057/palgrave.ivs.9500183

Riveiroa, M., Falkmana, G. o., Ziemkea, T., & Kronhamnb, T. (2009). Reasoning about anomalies: a study of the analytical process of detecting and identifying anomalous behavior in maritime traffic data. In W. J. Tolone & W. Ribarsky (Chair), Symposium conducted at the meeting of the Proceedings of SPIE. doi:10.1117/12.818117

Rocha, J. A. M. R., Oliveira, G., Alvares, L. O., & Bogorny, V. (2010, 7-9 July 2010). *DB-SMoT: A Direction-Based Spatio-Temporal Clustering Method*. Paper presented at the meeting of the 2010 5th IEEE International Conference Intelligent Systems (IS), London. doi:10.1109/IS.2010.5548396

Rosenhead, J., & Mingers, J. (2001). A new paradigm of analysis. In J. Rosenhead & J. Mingers (Eds.), *Rational analysis for a problematic world revisited* (pp. 1-19). Chichester: John Wiley and Sons.

Rossi, M., & Sein, M. K. (2003). Design Research workshop : A proactive research approach Symposium conducted at the meeting of the 26th Information Systems Research Seminar in Scandinavia

Saltenis, S., & Jensen, C. S. (2002). Indexing of Moving Objects for Location-Based Service. *IEEE Computer Society Washington.* Symposium conducted at the meeting of the 18th International Conference on Data Engineering ICDE '02.

Si, H., & Aung, Z. M. (2011). Position Data Acquisition from NMEA Protocol of Global Positioning System. *International Journal of Computer and Electrical Engineering, 3*(3), 353-357.

Sistla, A. P., Wolfson, O., Chamberlain, S., & Dao, S. (1997). *Modeling and Querying Moving Objects*. Paper presented at the meeting of the Proceedings of the Thirteenth International Conference on Data Engineering ICDE '97, Birmingham , UK doi:10.1109/ICDE.1997.581973

Spaccapietra, S., Parent, C., Damiani, M. L., Macedo, J. A. D., Porto, F., & Vangenot, C. (2008). A Conceptual view on trajectories. *ScienceDirect, 65*(1), 126-146. doi:10.1016/j.datak.2007.10.008

Speičvcys, L., Jensen, C. S., & Kligys, A. (2003). Computational data modeling for network-constrained moving objects. Symposium conducted at the meeting of the Advances in geographic information systems. doi:10.1145/956676.956692

Trajcevski, G., Ding, H., Scheuermann, P., Tamassia, R., & Vaccaro, D. (2007, November 07-09). *Dynamics-aware similarity of moving objects trajectories*. Paper presented at the meeting of the Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems, Seattle, Washington. doi:10.1145/1341012.1341027.

Tryfona, N., & Jensen, C. S. (1999). Conceptual Data Modeling for Spatiotemporal Applications. *Geoinformatica, 3*(3), 245 - 268. doi:10.1023/A:1009801415799

Tsou, M. C. (2010). Discovering Knowledge from AIS Database for Application in VTS. *Journal of Navigation, 63*(03), 449-469 doi:10.1017/S0373463310000135

Vaishnavi, V. K., & William Kuechler, J. (2008). Suggestion and Development Patterns. In *Design Science Research Methods and Patterns: Innovating Information and Communication Technology* (pp. 119-156). Boston, MA, USA Auerbach Publications.

Vazirgiannis, M., & Wolfson, O. (2001, 12–15 July). A Spatiotemporal Model and Language for Moving Objects on Road Networks.*Springer-Verlag.* Symposium conducted at the meeting of the 7th International Symposium on Advances in Spatial and Temporal Databases, Redondo Beach, California, USA.

Vries, G. K. D. d. (2012). *Kernel methods for vessel trajectories* (Dissertation). University of Amsterdam.

Walls, J. G., Widermeyer, G. R., & Sawy, O. A. E. (2004). Assessing Information System Design Theory in Perspective: How Useful was our 1992 Initial Rendition. *Journal of Information Technology Theory and Application 6*(2), 43-58.

Wang, H., & Zimmermann, R. (2010). Processing of continuous location-based range queries on moving objects in road networks. *IEEE Transactions on Knowledge and Data Engineering, 1*(99), 1-15. doi:10.1109/TKDE.2010.171

Whigham, P. A. (1993, 19-22 September). *Hierarchies of space and time* Paper presented at the meeting of the Spatial Information Theory A Theoretical Basis for GIS Marciana Marina, Elba Island. Italy. doi:10.1007/3-540-57207-4_13

Willems, N., Hage, W. R. V., Vries, G. D., H.M., J., & Malaisé, J. V. (2010). An integrated approach for visual analysis of a multisource moving objects knowledge base. *International Journal of Geographical Information Science 24*(10). doi:10.1080/13658816.2010.515029

Willems, N., Wetering, H. V. D., & Wijk, J. J. V. (2011). Evaluation of the Visibility of Vessel Movement Features in Trajectory Visualizations. *Wiley Online 30*(3), 801-810. doi:10.1111/j.1467-8659.2011.01929.x

Wood, Z., & Galton, A. (2009). A taxonomy of collective phenomena. *Journal Applied Ontology, 4*(3-4), 267-292

Wolfson, O., Sistla, A. P., Chamberlain, S., & Yesha, Y. (1999). Updating and Querying Databases that Track Mobile Units. *Distributed and Parallel Databases - Special issue on mobile data management and applications., 7*(3). doi:10.1023/A:1008782710752

Wolfson, O., Sistla, A. P., Xu, B., Zhou, J., Chamberlain, S., Yesha, Y., et al. (1999 5–7 July). Tracking Moving Objects Using Database Technology in DOMINO.*Springer-Verlag.* Symposium conducted at the meeting of the 4th International Workshop on Next Generation Information Technologies and Systems., Zikhron-Yaakov, Israel.

Wolfson, O., & Xu, B. (2010). Spatio-temporal databases in urban transportation. *IEEE Data Engineering, 33*, 1-8. doi:10.1.1.165.3023

Worboys, M. F. (1994). A Unified Model for Spatial and Temporal Information. *The Computer Journal, 37*(1), 26-34. doi:10.1093/comjnl/37.1.26

Yeh, T. S., & Cambray, B. d. (1995). Modeling highly variable spatio-temporal data. Symposium conducted at the meeting of the 6th AustraliAsian Database Conference.

Zheni, D., Frihida, A., Ghezala, H. B., & Claramunt, C. (2009). A Semantic Approach for the Modeling of Trajectories in Space and Time. *Springerlink, 5833*(2009), 347-356. doi:10.1007/978-3-642-04947-7_41

Appendix A

Vessels identified from the chosen sample dataset.

MMSI	VesselName	VesselType
5120004	NZ SCHOOL FISHERIES	Other
126282899	AMMONITE	Pleasure Craft
210076000	JRS PEGASUS	Container ship
211216200	SONNE	Research/survey vessel
212865000	CMA CGM MATISSE	Container ship
228230700	BENIGUET	Unspecified
235012190	KAITAKI	Ro-ro/passenger ship
235012190	KAITAKI	Vehicles carrier
235076734	SJN NORDIC	Bulk carrier
235079608	DORADO DISCOVERY	Research/survey vessel
235679000	BRITISH MERLIN	Crude oil tanker
244848000	DONAUGRACHT	Cargo
245723000	ATLANTIC KLIPPER	reefer
248092000	DANSHIP BULKER	Bulk carrier
248333000	ECE NUR K	Oil/chemical tanker
257594000	TEXAS	Vehicles carrier
266259000	FEDORA	Vehicles carrier
269083000	MAERSK JAUN	Container ship
269084000	MAERSK JENAZ	Container ship
272204000	IVAN GOLUBETS	Fish factory
272204000	IVAN GOLUBETS	trawler
272221000	ALEKSANDR BURYACHENK [UA]	Fish factory
304010647	CHRISTA RICKMERS	Cargo
305245000	BBC COLORADO	Cargo
319546000	KAORI	Sailing Vessel
319855000	JULIET	Sailing Vessel
351240000	PAIWAN WISDOM	Bulk carrier
352258000	STOLT RINDO	Oil/chemical tanker
352503000	BOTANY TROUBADOUR	Oil/chemical tanker
353243000	CALIFORNIA MERCURY	Container ship
353534000	ULTRAMAR 1	passenger ship
354213000	ORIENT HOPE	Bulk carrier
354367000	ALAM SERI	Bulk carrier
354615000	SCAN BULKER	Bulk carrier
355028000	MSC AMERICA	Container ship
355163000	GREEN HOPE	Bulk carrier
355242000	ORIENT EXPLORER	Research/survey vessel
355883000	KING SUGAR	Bulk carrier
355906000	LOUISE BULKER	Bulk carrier

356543000	GLORIOUS SUNRISE	Cargo
356817000	HARU VERDY	reefer
357151000	GOLDEN MERMAID	Oil/chemical tanker
357507000	TPC WELLINGTON	Bulk carrier
357767000	AQUILA EXPLORER	Multi purpose offshore vessel
370560000	GLOBAL BAY	Bulk carrier
370765000	HONGBO	Oil/chemical tanker
371008000	VICTOIRE	LPG tanker
371352000	MAEA	LPG tanker
371797000	TRANS FUTURE 6	Vehicle carrier
431700210	SHOFUKU MARU NO.78	Fishing
440288000	DONG WON NO.519	trawler
440761000	SUR ESTE NO.709	Fishing vessel
440890000	530DONGWON	Fishing vessel
441769000	OYANG NO.75	Trawler
441769000	OYANG NO.75	trawler
477049800	MOUNT OWEN	Bulk carrier
477071100	KOTA PERMATA	Container ship
477125800	POS KNIGHT	Bulk carrier
477143300	MAERSK RADFORD	Cargo
477235000	CAPE SCOTT	Bulk carrier
477440400	KWANGTUNG	cargo
477621300	YE CHI	Oil products tanker
477647800	MCC DHAKA	Container ship
477661100	PACIFIC LOHAS	Cargo
477728800	CAP MANUEL	Container ship
477746600	PORTLAND BAY	Container ship
477815800	SPRING SKY	Vehicles carrier
477892600	TENG HANG	Bulk carrier
477958600	MAERSK ABERDEEN	Container ship
477958600	MAERSK ABERDEEN	Container ship
512000015	THOMAS HARRISON	Trawler
512000030	SEAMOUNT EXPLORER	trawler
512000034	ARAHURA	Ferry
512000048	MILBURN CARRIER II	Cement carrier
512000049	WESTPORT	Cement carrier
512000050	KAKARIKI	Oil products tanker
512000109	KORAKI	Tug
512000113	INDEPENDENT 1	Trawler
512000200	RESOLUTION	Naval/naval auxiliary vessel
512000200	RESOLUTION	Naval/naval auxiliary vessel
512000300	MANAWANUI	Naval/naval auxiliary

		vessel
512000321	WELLINGTON PILOT	Pilot Vessel
512000322	TIAKI	Tug
512000327	BREAM BAY	Tug
512000344	PILOT VESSEL AKARANA	Pilot Vessel
512000390	ANATOKI	Cargo
512000399	HARBOURMASTER 1	law force
512000741	CLIPPER 3	High Speed Craft
512000742	CLIPPER 2	High Speed Craft
512000743	CLIPPER 1	High Speed Craft
512002959	STRAITSMAN	Ro-ro cargo
512004173	TE MARU	tug
512036000	SANTA REGINA	Ro-ro/passenger ship
512041000	GOLDEN BAY	Cement carrier
512042000	SPIRIT OF ENDURANCE	Container ship
512071000	ARATERE	Ferry
512153000	HMNZS CANTERBURY	Naval/naval auxiliary vessel
512154000	ROTOITI	Naval/naval auxiliary vessel
512157000	PUKAKI	Naval/naval auxiliary vessel
512158000	TAUPO	Naval/naval auxiliary vessel
512169000	AORAKI	tug
512175000	TOREA	Oil/chemical tanker
512179000	PELICAN	Suction hopper dredger
512201000	SPIRIT OF NEWZEALAND	passenger ship
512219000	AWANUIA	Oil products tanker
512392000	F.V. OCEAN PIONEER	Other
538003236	WREN	Bulk carrier
538003456	SWIBER TORUNN	Anchor handling vessel
538003725	TPC SAMJIN	Bulk carrier
538004124	TPC VANCOUVER	Bulk carrier
538004193	NIGHTHAWK	Bulk carrier
538070349	SURI	Pleasure Craft
538080008	CV 9	Pleasure craft
538090413	MARE TRAVELLER	Bulk carrier
563026000	VHOEGH CHENNAI	Vehicles carrier
563177000	PACIFIC WORKER	Anchor handling vessel
563945000	SOUTHERN CROSS	Container ship
564033000	PACIFIC CHIEFTAIN	Tug/supply vessel
564443000	ALAM SAKTI	Bulk carrier
565153000	ADVANCE II	Oil/chemical tanker

636014031	SOFRANA TOURVILLE	Container ship
636014934	NOBLE DISCOVERER	Drill Ship
636090862	MAERSK DUNBAR	Container ship
636090863	MAERSK DUNCAN	Other
636090976	ANL BINBURRA	Container ship
636091171	JPO SCORPIUS	Container ship
636091333	VALENTINA SCHULTE	Container ship
636091887	CAP MONDEGO	Container ship
636096640	TREMONIA	Bulk carrier
512000058	TANGAROA	trawler

Appendix B

Screenshots of AIS Raw data

```
2011-04-15 14:57:03.390 !AIVDM,2,2,6,0,888888888880,4*6D
2011-04-15 14:57:03.407 !AIVDM,2,1,5,0,E7`B15t0h1:Wdh@@@@@@@@@@@@@6@9rNmO09H10888QJ``88,0*3B
2011-04-15 14:57:03.407 !AIVDM,2,2,5,0,888888888880,4*6E
2011-04-15 14:57:03.423 !AIVDM,1,1,,B,E7`B17t77oL@@@@@@@@@@@@@@@@6@7CHmNDE810888QJ8`8888888888880,0*06
2011-04-15 14:57:03.453 !AIVDM,1,1,,B,E7`B18L77oHp@@@@@@@@@@@@@@@6@9cnmN0dH10888Qr<`8888888888880,0*14
2011-04-15 14:57:03.517 !AIVDM,1,1,,A,17`B0<P01f<L1=iWN9C8oVv40@2D,0*0A
2011-04-15 14:57:03.533 !AIVDM,1,1,,B,E7`B18dW7oHph@@@@@@@@@@@@@@6@9mQmMpeh10888Qr8`8888888888880,0*23
2011-04-15 14:57:03.627 !AIVDM,2,1,0,0,E7`B18dW7oHph@@@@@@@@@@@@@@6@9mQmMpeh10888Qr8`88,0*6A
2011-04-15 14:57:03.627 !AIVDM,2,2,0,0,888888888880,4*6B
2011-04-15 14:57:03.627 !AIVDM,2,1,8,0,E7`B17t77oL@@@@@@@@@@@@@@@@6@7CHmNDE810888QJ8`88,0*47
2011-04-15 14:57:03.627 !AIVDM,2,2,8,0,888888888880,4*63
2011-04-15 14:57:03.627 !AIVDM,1,1,,B,19NWtH@02v<QHPK`:I1TISN8082H,0*37
2011-04-15 16:32:03.313 !AIVDM,1,1,4,1,87`B16@0@uq0,0*4C
2011-04-15 16:32:03.343 !AIVDM,1,1,,A,17`B0<P01f<Jf@IWF1:8m7420D01,0*76
2011-04-15 16:32:03.377 !AIVDM,1,1,,B,13Ib6;8000<P8;ic2re1wnN60V20,0*7D
2011-04-15 16:32:03.390 !AIVDM,1,1,,A,87`B17h0@uq@oU05mjhP000000000000000000000000000000000000000,2*63
2011-04-15 16:32:03.423 !AIVDM,1,1,,B,13Ib6;8000<P8;ic2re1wnN60V20,0*7D
2011-04-15 16:32:03.440 !AIVDM,1,1,5,1,87`B16h0@uq0,0*65
2011-04-15 16:32:03.453 !AIVDM,1,1,,A,87`B18@0@uq<tE07gjhP000000060H00000000=j8000000006600000000,2*1A
2011-04-15 16:32:03.517 !AIVDM,1,1,,A,87`B18P0@uq;FE07o:hP000000000000000000000000000000000000000,2*46
2011-04-15 16:32:03.610 !AIVDM,1,1,,A,87`B19@0@uq9u508CRhP000000000000000000000000000000000000000,2*5D
2011-04-15 16:32:03.673 !AIVDM,1,1,7,1,87`B18@0@uq0,0*41
2011-04-15 16:32:03.673 !AIVDM,1,1,6,1,87`B17h0@uq0,0*67
2011-04-15 16:32:03.673 !AIVDM,1,1,,A,87`B1:00@uq7O506RBhP000000000000000000000000000000000000000,2*15
2011-04-15 16:32:03.690 !AIVDM,1,1,8,1,87`B18P0@uq0,0*5E
2011-04-15 16:32:03.720 !AIVDM,1,1,,A,87`B1:P0@uq6dm05=JhP000000000000000000000000000000000000000,2*63
2011-04-15 16:32:03.890 !AIVDM,1,1,0,1,87`B1:00@uq0,0*34
2011-04-15 16:32:03.890 !AIVDM,1,1,9,1,87`B19@0@uq0,0*4E
2011-04-15 16:32:03.890 !AIVDM,1,1,1,1,87`B1:P0@uq0,0*55
2011-04-15 16:32:04.330 !AIVDM,1,1,2,1,87`B16P0@uq0,0*5A
2011-04-15 16:32:04.343 !AIVDM,1,1,,A,87`B16P0@uqChE052BhP000000000000000000000000000000000000000,2*39
2011-04-15 16:32:04.360 !AIVDM,1,1,,A,19NWuIh02a<@s7IVD<QPchV:0H21,0*1F
2011-04-15 16:32:04.360 !AIVDM,1,1,,A,87`B17@0@uqB;E05KjhP000000000000000000000000000000000000000,2*2B
2011-04-15 16:32:04.533 !AIVDM,1,1,3,1,87`B17@0@uq0,0*4A
2011-04-15 16:32:04.547 !AIVDM,1,1,,A,87`B18h0@uq;M508@JhP000000000000000000000000000000000000000,2*55
2011-04-15 16:32:04.750 !AIVDM,1,1,4,1,87`B18h0@uq0,0*6A
```

Appendix B.1: Screenshot illustrating that !AIVDM sentence need to be removed.

	NMEA_ID	TimeDate	RawSentence	TalkerID	SentenceID
1	7	2011-04-09 00:00:01.517	!AIVDM,1,1,,B,67`B164PH1m@wsp0003000@10H0,2*4D	AI	VDM
2	16	2011-04-09 00:00:01.533	!AIVDM,1,1,,B,139KWB301D<RdRA`3M`A8Q9l0@09,0*55	AI	VDM
3	22	2011-04-09 00:00:01.550	!AIVDM,1,1,,B,139KWB301D<RdRA`3M`A8Q9l0@09,0*55	AI	VDM
4	24	2011-04-09 00:00:01.550	!AIVDM,1,1,,A,15Dj?N002g<L:`qdGfm5I4En00S5,0*5E	AI	VDM
5	26	2011-04-09 00:00:01.550	!AIVDM,1,1,,B,B7`B;Ph0Ck8:IKrhD992gwP5oP06,0*6C	AI	VDM
6	28	2011-04-09 00:00:01.550	!AIVDM,1,1,,B,138Fl:002F<d<eua1o3`G6el0400,0*1F	AI	VDM
7	29	2011-04-09 00:00:01.550	!AIVDM,1,1,,B,B7`B;Ph0Ck8:IKrhD992gwP5oP06,0*6C	AI	VDM
8	31	2011-04-09 00:00:01.563	!AIVDM,1,1,,B,37`B0<@3Av<JpiQ`iu@CsC:000wP,0*17	AI	VDM
9	33	2011-04-09 00:00:01.563	!AIVDM,1,1,,B,37`B0<@3Av<JpiQ`iu@CsC:000wP,0*17	AI	VDM
10	40	2011-04-09 00:00:01.580	!AIVDM,1,1,,B,19NWp@m0?w<tSF0l4Q@>493v8T00,0*2F	AI	VDM
11	41	2011-04-09 00:00:01.580	!AIVDM,1,1,,B,18KS8<001p<>ojmUl:rbq`in0H0L,0*1A	AI	VDM
12	44	2011-04-09 00:00:01.580	!AIVDM,1,1,,B,H7`B2:10Thu@000000000000000,2*4A	AI	VDM
13	58	2011-04-09 00:00:01.673	!AIVDM,1,1,,B,H7`B2:4UoCD1B00J=1kjmi000000,0*49	AI	VDM
14	62	2011-04-09 00:00:01.673	!AIVDM,1,1,,B,35BBdR5000<P847`HC=a@bon0PTi,0*0E	AI	VDM
15	66	2011-04-09 00:00:01.690	!AIVDM,1,1,,A,37`D<`?2BOdO0si`JhF=Lb`020W1,0*78	AI	VDM
16	67	2011-04-09 00:00:01.690	!AIVDM,1,1,,A,37`D<`?2BOdO0si`JhF=Lb`020W1,0*78	AI	VDM
17	71	2011-04-09 00:00:01.690	!AIVDM,1,1,,B,140WUH002;<Qfmic7Ki9C7Kl0<03,0*4F	AI	VDM
18	75	2011-04-09 00:00:01.703	!AIVDM,1,1,,B,140WUH002;<Qfmic7Ki9C7Kl0<03,0*4F	AI	VDM
19	76	2011-04-09 00:00:01.703	!AIVDM,1,1,,B,37`B1ShP@@<P08Qbs9g80RD4011Q,0*68	AI	VDM
20	77	2011-04-09 00:00:01.703	!AIVDM,1,1,,B,37`B0<@3Av<JpiQ`iu@CsC:000wP,0*17	AI	VDM

Appendix B.2: Screenshot that illustrates repeated AIS datasets.

Appendix C

GIS Features:

The main objective of using spatio-temporal attributes is to get good insight of moving object movement details. Therefore, the decoded data set has spatio-temporal data and it is essential to have a powerful tool to handle such datasets.

5. 2 Why GIS?

In recent times, research on moving-object data analysis has been fostered by new identification technology for collecting, storing location-aware data for monitoring and analysis purpose. These technologies have made available a massive repository of spatio-temporal data. Therefore, such data has to be exported to a tool to perform suitable analytical study. The quest for selecting an appropriate tool and performing suitable analysis will be the main discussion of this chapter.

Movement data collected from real time objects have enriched with information. Such information exhibits to understand their spatial relationships, movement dynamics and activity patterns. To study these features of movement data a tool that performs spatial, temporal analysis also supports storage of reality information of movement, Geographic Information system was considered.

5.2. 1 Geographic Information Systems – (GIS) is a database that is used for very specific sorts of information about reality. This information was organized in a way that would prove useful in the real world. It was developed based on the model of the world and was represented as a Raster or Vector Reality.

5.2. 2 Raster Reality: In raster reality the system displays, locates and stores graphical data by using a matrix or grid of cells. A unique reference coordinate represented each pixel either at a corner or the centre. Each cell or pixel has a discrete attribute data assigned to it. Raster data resolution was dependent on the pixel or grid size and may vary from sub-meter to many kilometres. Since these data were two-dimensional, GIS stored different types of information such as forest cover, soil type, and land use. Generally, raster data required less processing than vector data, but consumed more computer storage space.

5.2.3 Vector Reality: In vector reality the system displayed graphical data as points, lines, or curves, or areas with attributes. Cartesian coordinates (x and y) and computational algorithms of the coordinates defined points in a vector system. Lines or arcs were a series of ordered points. Areas or polygons were also stored as ordered lists of points, by making the beginning and end points with the same node; the shape was closed and defined. Vector systems were capable of very high resolution and graphical output was similar to hand-drawn maps. The system worked well with azimuths, distances and points, but required complex data structures and was less compatible with remote sensing data. Vector data required less computer storage space and therefore, maintaining topological relationships was easier in this system.

However, the data modelled in either system could be converted into the other making them flexible for use.

Method	Advantages	Disadvantages
Raster	• Simple data structure • Compatible with remotely sensed or scanned data • Simple spatial analysis procedures	• Required greater storage space on computer • Depending on pixel size, graphical output may be less pleasing • Projection transformations were more difficult • More difficult to represent topological relationships
Vector	• Requires less disk storage space • Topological relationships are readily maintained • Graphical output more closely	• More complex data structure • Not as compatible with remotely sensed data • Software and hardware were often more expensive • Some spatial analysis procedures

	resembled hand-drawn maps	may be more difficult. • Overlaying multiple vector maps was often time consuming.

Table5. 2: Different data type of GIS.

5.3 Organizing Attribute Data: GIS provided a linkage between spatial and non-spatial data. Location information was linked to specific information in a database through the use of ID number and then stored in separate files under this ID number. There were three types of files where data could be stored. They were Flat Files (spreadsheets), Hierarchical Files and Relational Files.

5.3.1 A Flat File or spreadsheet was a simple method for storing data where all records have the same number of fields and individual records have different data in each field with one field serving as a key to locate a particular record.

5.3.2 Hierarchical files stored data in more than one type of record, usually described as a "parent-child, one-to-many" relationship. Here one field was a key to all records and hence data does not have to be repeated. This system allowed records with similar attributes to be associated together as the records were linked to each other by a key field in a hierarchy of files. One record lead to another in a relatively descending pattern where each record except for the master record, had a higher level record file linked by a key field "pointer".

5.3.3 Relational Files connected different files or tables (relations) by the use of a common link of data, to join or associate records where information was stored with the help of matrices of tables. As long as the tables had a common link they could be combined by the user to form new inquires and data output. As queries were not limited by a hierarchy of files but were based on relationships from one type of record to another, they were flexible and most popular database model for GIS. They were particularly suited for SQL (structured query language).

Structure	Advantages	Disadvantages

Flat Files	• Fast data retrieval • Simple structure and easy to program	• Difficult to process multiple values of a data item • Adding new data categories required reprogramming • Slow data retrieval without the key
Hierarchical Files	• Adding and deleting records was easy • Fast data retrieval through higher level records • Multiple associations with similar records in different files	• Pointer path restricts access • Each association required repetitive data in other records • Pointers required large amount of computer storage
Relational Flies	• Easy access and minimal technical training for users • Flexibility for unforeseen inquiries • Easy modification and addition of new relationships, data and records • Physical storage of data could change without affecting relationships	• New relations could require considerable processing • Sequential access was slow • Method of storage on disks impacts processing time • Easy to make logical mistakes due to flexibility of relationships between

	between records	records.

Table 5.3: Different level of file features existing in Geographic Information System (GIS).

5.4 Additional Features of GIS tools:

GIS had power to record more than location and simple attribute information and because of this spatial relationships could be examined based on location as well as functional and logical relationships. Databases were designed to represent, model and store information about these relationships as per requirement for particular applications.

Spatial Relationships were as follows:

- Absolute and relative location
- Distance between features
- Proximity of features
- Features in the neighbourhood of other features
- Direction and movement from place to place
- Boolean relationships of “and”, “or”, “inside”, “outside”, “intersecting”, non-intersecting”.

5.4.1 Functional Relationships included information about how geographic features were connected and interacted in real-life terms. For example a road network might be classified functionally from the largest superhighway down to the most isolated rural road or suburban cul-de-sac based upon their role in the overall transportation system.

5.4.2 Logical Relationships involved “if-then” and “and-or” conditions that must exist among features stored in the dataset. For example no land may be permitted to be zoned for residential use if it lies within a river’s five-year flood plain.

A useful relationship was maintained in many spatial databases with the help of topology, as GIS manipulated analgised and used topological data in determining data relationships. Spatial relationships in a GIS were determined by mathematics of connectivity or adjacency of points or lines. Topological data structure logically was determined exactly by means of nodes (topological junctions) how and where points and lines connected on a map. The shape of an arc or polygon was determined by the order of connectivity, where the computer stored the information in various tables of the database structure. By storing information in a logical order, relationship between missing information was readily apparent e.g. a line segment of a polygon.

Arc-node topology was developed several decades ago as a convenient way of storing information and it is therefore useful to see how this topology was represented by the connections of points, lines and regions, which were coded into the database. The location of nodes which represented endpoints, intersections of lines and boundaries were recorded. Based upon these nodes, “arcs” were defined. The arcs’ endpoints were assigned direction by arrowheads. The starting point of the vector was the ‘from node’ and the destination was the ‘to node’. As long as the direction was recorded and stored in the database, the orientation of a given vector could be assigned in either direction. By tracking the orientation of arcs, it was possible to use this information to establish routes from node to node. Hence the movement from one node to another could be located by the necessary connections in the database.

Polygons defined by arcs could be traced around its area in a clockwise direction recording the component arcs and their orientations. When the arc was followed in reverse orientation, the tracing was assigned a negative sign in the database. Each arc had a polygon recorded on the left and right side of its direction of orientation. If an arc was on the edge of the study area, it was bounded by the “universe”.

After investigating thoroughly the benefits, advantage and additional features of GIS tool clearly demonstrates its strength on conducting analytical study on spatio-temporal attributes. Therefore, ArcGIS tool is selected for this research study as the study involves disseminated information about earth information.

Appendix D

Source code – AIS Decoder

```
// Program to decode raw AIS data to spatio-temporal movement data.

// Movement data includes spatial data - latitude/longitude/speed.

// Temporal data includes timestamp details adjoined with raw data.


Public Class AIS_form


    Private Sub AISBtn1_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles Start_Decode_Btn.Click



        'Declaring StreamWriter Object

        Dim file As System.IO.StreamWriter



        'Variable Declaration - Key variables: latitude, longitude, MMSI and
speed

        Dim holding, latnum, longnum, mmsinum, speed As Double

        Dim ais_data As String

        Dim szArr As Char()

        Dim frr As Char()



        'Declaration of decode elements

        Dim longstr As String

        Dim eightbit, fred As String

        Dim bit, z, zz As Integer

        Dim whole_string, full_str As String

        Dim message_type_number As Integer

        Dim message_type, mmsi, Longitude, Latitude As String
```

```
        'Set up and display the OpenFile diaglog

        'File opened with the specified folder directory

        'file ready to open and read

        FileOpen(1, "C:/PremaWorks/Data/11.05_6pm_12pm.txt", OpenMode.Input,
OpenAccess.Read)

        'Set up and display the OpenFile diaglog

        'File opened with the specified folder directory

        'file ready to open and write

        file =
My.Computer.FileSystem.OpenTextFileWriter("C:/PremaWorks/Data/11.05_6pm_12pm_OUTP
UT.txt", True)


        'To check End of File

        Do While Not EOF(1)

            ais_data = LineInput(1)

            full_str = ais_data

            ' save full_str

            ais.Text = full_str

            ais.Refresh()




            'To Check whether the input string starts with !AIVDM

            'Also to check whether !AIVDM is not null

            If InStr(ais_data, "!AIVDM,1") > 0 Then

                ais_data = Trim(Mid(ais_data, InStr(ais_data, "!AIVDM,1")))

                ' ais_data = Mid(ais_data, 15)

                'find the message header data

                z = 0
```

```
'to identify the strings that is been separated with commas
    'After!AIVDM 5 commas are in a sentence
    eg:!AIVDM,1,1,,A,17`B2qgP4`<PmpibqhM4rww`08OA,0*2C
        For i = 1 To 5
            z = z + InStr(Mid(ais_data, z + 1), ",")
            ' MsgBox(z)
        Next i

        'last one is start of sentence
        z = z + 1
        ais_data = Mid(ais_data, z)
        ais_data = Mid(ais_data, 1, InStr(ais_data, ",") - 1)
        ' MsgBox(ais_data & "ais_data is")
 'to copy the characters in this instance to an array of Unicode character
        frr = ais_data.ToCharArray()
        whole_string = ""
        holding = 0

       'Converting to Binary
        For i = 0 To Len(frr) - 1
            'MsgBox(szArr(I))
            holding = Asc(frr(i)) - 48

            If holding > 40 Then holding = holding - 8
            'MsgBox(holding)
            If (holding And 32) = 32 Then bit = 1 Else bit = 0
            longstr = bit
            If (holding And 16) = 16 Then bit = 1 Else bit = 0
            longstr = longstr & bit
            If (holding And 8) = 8 Then bit = 1 Else bit = 0
```

```
                longstr = longstr & bit
                If (holding And 4) = 4 Then bit = 1 Else bit = 0
                longstr = longstr & bit
                If (holding And 2) = 2 Then bit = 1 Else bit = 0
                longstr = longstr & bit
                If (holding And 1) = 1 Then bit = 1 Else bit = 0
                longstr = longstr & bit

                whole_string = whole_string & longstr
                'MsgBox(longstr)
                ' TMPCHAR = Mid(ais_data, I, 1)
                ' /* Push It Into The Temporary Array */
                ' MsgBox(TMPCHAR)
                ' TMPARRAY(I) = TMPCHAR

    Next i
            '  MsgBox(whole_string)
            'RECONVERT

            z = 0
            eightbit = ""
            For I = 1 To 167
                eightbit = eightbit & Mid(whole_string, I, 1)
                z = z + 1
                If z = 8 Then
                    z = 0
                    '  MsgBox(eightbit)
                    '  MsgBox(BinaryToDecimal(eightbit))
          ' MsgBox(eightbit & " " & BinaryToDecimal(eightbit) & " " &
Chr(BinaryToDecimal(eightbit)))
                    eightbit = ""
```

```
        End If
    Next



    ' MsgBox(Len(whole_string))
    If Len(whole_string) = 168 Then
        'its a single sentence
        szArr = whole_string.ToCharArray()

        message_type = ""
        For I = 0 To 5
            message_type = message_type & szArr(I)
        Next

        ' MsgBox(message_type)
        message_type_number = BinaryToDecimal(message_type)
        ' MsgBox(message_type_number & " " & ais_data)

        'When a message string message String is a valid string
        If message_type_number = 1 Then
            mmsi = 0

            'Bit position from 8 to 37 represents MMSI
            ' MMSI- Maritime Mobile Service Identity
           For I = 8 To 37
                mmsi = mmsi & szArr(I)
            Next

            'Calculating Longitude -Decoding longitude
            Longitude = ""
```

```
                Lat - bit 89 for 27 bits (a signed binary number, divide by
600000)

                        'Lon - bit 61 for 28 bits (a signed binary number, divide
by                       600000)

                        zz = 0

                        For I = 62 To 88

                            Longitude = Longitude & szArr(I)

                            zz = zz + 1

                        Next


                        'Calculating Latitude -Decoding latitude

                        ' MsgBox(zz & "zz long is")

                        Latitude = ""

                        zz = 0

                        For I = 90 To 115

                            Latitude = Latitude & szArr(I)

                            zz = zz + 1

                        Next

                        '  MsgBox(zz & "zzis")

                        'sog = ""


                        'Calculate Speed over ground

                        For I = 50 To 59

                            sog = sog & szArr(I)

                        Next


                    End If


                    'Longitude -- Conversion of Binary to Decimal values


                    mmsinum = BinaryToDecimal(mmsi)
```

```
            If Val(szArr(61)) = 1 Then

                Longitude = twos_comp(Longitude)
                longnum = BinaryToDecimal(Longitude) + 1
                longnum = longnum * -1
            Else
                longnum = BinaryToDecimal(Longitude)

            End If

            'Longitude -- Conversion of Binary to Decimal values

            If Val(szArr(89)) = 1 Then
                Latitude = twos_comp(Latitude)
                latnum = BinaryToDecimal(Latitude) + 1
                latnum = latnum * -1
            Else
                latnum = BinaryToDecimal(Latitude)
            End If
ListBox_Display.Items.Add(mmsinum & " " & latnum & " " & longnum & " raw "
& speed)
fred = "latitude " & latnum & " longitude " & longnum & "  raw data"



            'file.WriteLine(fred)
            latnum = latnum / 600000
            longnum = longnum / 600000

            ' MsgBox(latnum & " " & longnum)
            '  MsgBox(longnum & "longnum is")
```

```
                    'If longnum > 180 Then longnum = longnum - 180
                    'If latnum > 90 Then latnum = latnum - 90
                    'If szArr(61) = "0" Then latnum = 90 - latnum
                    'If szArr(89) = "0" Then longnum = 90 - longnum
                    'longnum = BinaryToDecimal(Longitude) / 600000
                    ListBox_Display.Items.Add(mmsinum & " " & latnum & " " &
                    longnum & " " & speed)
                    fred = mmsinum & " " & latnum & " " & longnum & " " & speed &
                    " " & full_str
                    'file =
My.Computer.FileSystem.OpenTextFileWriter("d:/data/ais/output.txt", True)
                    file.WriteLine(fred)
                End If
                message_type = Latitude
                TextBox_Display.Text = message_type
                'MsgBox(mmsi)
                'MsgBox(Latitude)
                'MsgBox(Longitude)

            End If
            'next line of file

        Loop

        MsgBox("All Finshed -- decoding")
        FileClose(1)
        file.Close()
    End Sub
    'Function to Convert Binary to Decimal
```

```
Public Function BinaryToDecimal(ByVal Binary As String) As Long
    Dim n As Long
    Dim s As Integer

    For s = 1 To Len(Binary)
        n = n + (Mid(Binary, Len(Binary) - s + 1, 1) * (2 ^ _
            (s - 1)))
    Next s
    ' ListBox2.Items.Add(Binary & "  " & n)
    BinaryToDecimal = n
End Function

'Two's complement for signed numbers
Public Function twos_comp(ByVal Binary As String) As String
    Dim s As Integer
    Dim twoscomp As String
    twoscomp = ""

    For s = 1 To Len(Binary)
        If (Mid(Binary, Len(Binary) - s + 1, 1)) = "1" Then
            twoscomp = "0" & twoscomp
        Else
            twoscomp = "1" & twoscomp
        End If
    Next s
```

Appendix E

Handling large AIS dataset

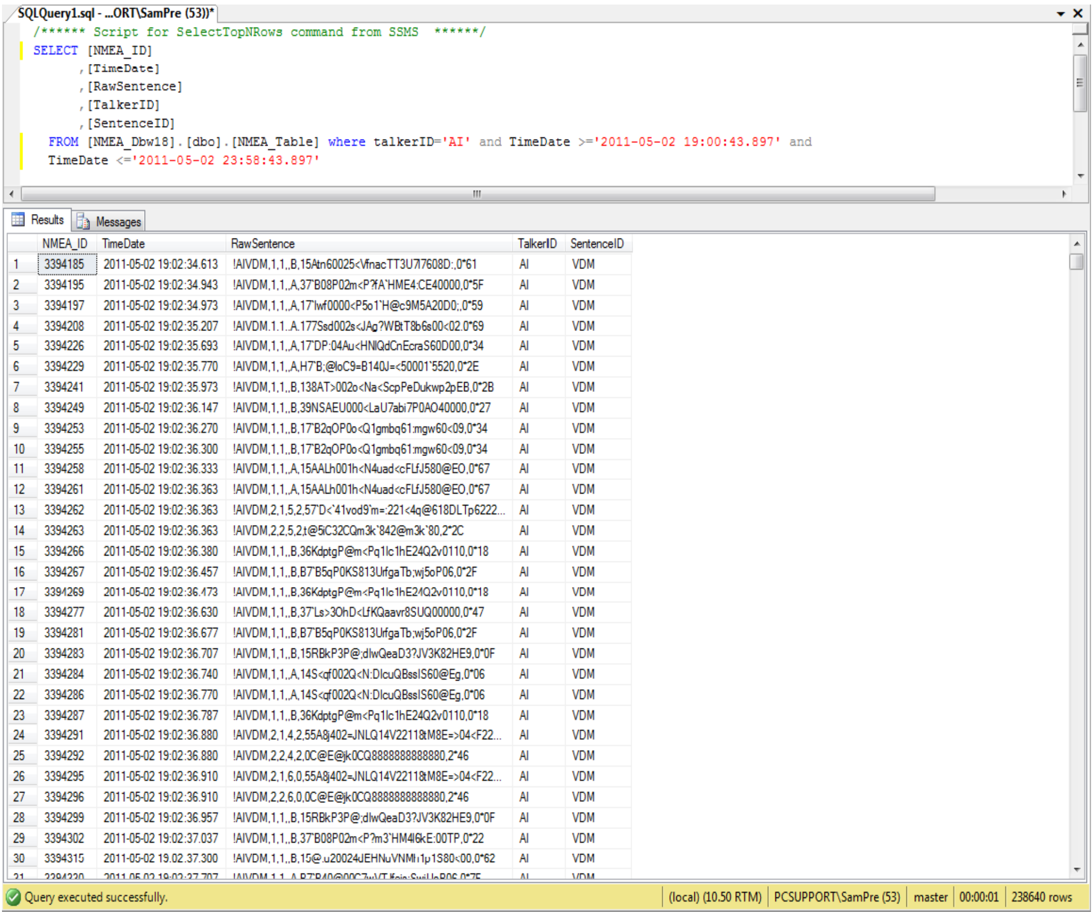


Figure E.1: Screenshot showing a selection of TalkerID='AI' and TimeDate chosen from !AIVDM sentences from DBW18. This demonstrates the approach taken to split the large sized text file to manageable size text files.

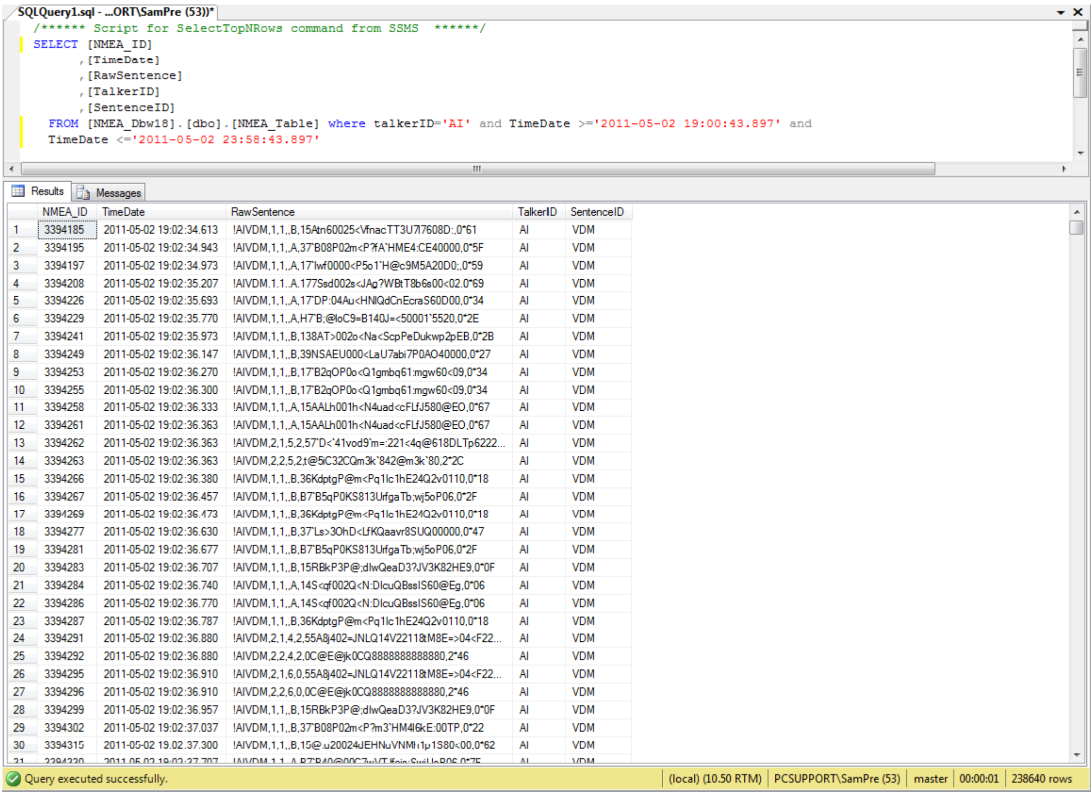


Figure E.2: Screenshot showing a selection of TalkerID='AI' and TimeDate chosen from !AIVDM sentences from another database. This demonstrates the approach taken to split the large sized text file to manageable size text files.