2013-12

Traffic pattern detection using the Hough transformation for anomaly detection to improve maritime domain awareness

McAbee, Ashley S. M.

Monterey, California: Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

TRAFFIC PATTERN DETECTION USING THE HOUGH TRANSFORMATION FOR ANOMALY DETECTION TO IMPROVE MARITIME DOMAIN AWARENESS

by

Ashley S. M. McAbee

December 2013

Co-Advisor:	James Scrofani
Co-Advisor:	Murali Tummala
Second Reader:	David Garren

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE *Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
December 2013

3. REPORT TYPE AND DATES COVERED
Master's Thesis

4. TITLE AND SUBTITLE
TRAFFIC PATTERN DETECTION USING THE HOUGH TRANSFORMATION FOR ANOMALY DETECTION TO IMPROVE MARITIME DOMAIN AWARENESS

5. FUNDING NUMBERS

6. AUTHOR(S) Ashley S. M. McAbee

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number ____N/A____.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

Techniques for anomaly detection in the maritime domain by extracting traffic patterns from ship position data to generate atlases of expected ocean travel are developed in this thesis. An archive of historical data is used to develop a traffic density grid. The Hough transformation is used to extract linear patterns of elevated density from the traffic density grid, which can be considered the "highways" of the oceans. These highways collectively create an atlas that is used to define geographical regions of expected ship locations. Ship position reports are compared to the atlas of highways to flag as anomalous any ship that is not operating on an expected highway. The atlas generation techniques are demonstrated using automated information system (AIS) ship position data to detect highways in both open-ocean and coastal areas. Additionally, the atlas generation techniques are used to explore variability in ship traffic as a result of extreme weather and seasonal variation. Finally, anomaly detection is demonstrated by comparing AIS data from 2013 to the highways detected in the archive of data from 2012. The development of an automatic atlas generation technique that can be used to develop a definition of normal maritime behavior is the significant result of this thesis.

14. SUBJECT TERMS Maritime Domain Awareness, Hough Transformation, Anomaly Detection, Automated Information System, Pattern Extraction

15. NUMBER OF PAGES
105

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

THIS PAGE INTENTIONALLY LEFT BLANK

TRAFFIC PATTERN DETECTION USING THE HOUGH TRANSFORMATION FOR ANOMALY DETECTION TO IMPROVE MARITIME DOMAIN AWARENESS

Ashley S. M. McAbee
Lieutenant, United States Navy
B. S., United States Naval Academy, 2007

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

from the

NAVAL POSTGRADUATE SCHOOL
December 2013

Author: Ashley S. M. McAbee

Approved by: James Scrofani
Co-Advisor

Murali Tummala
Co-Advisor

David Garren
Second Reader

R. Clark Robertson
Chair, Department of Electrical and Computer Engineering

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Techniques for anomaly detection in the maritime domain by extracting traffic patterns from ship position data to generate atlases of expected ocean travel are developed in this thesis. An archive of historical data is used to develop a traffic density grid. The Hough transformation is used to extract linear patterns of elevated density from the traffic density grid, which can be considered the "highways" of the oceans. These highways collectively create an atlas that is used to define geographical regions of expected ship locations. Ship position reports are compared to the atlas of highways to flag as anomalous any ship that is not operating on an expected highway. The atlas generation techniques are demonstrated using automated information system (AIS) ship position data to detect highways in both open-ocean and coastal areas. Additionally, the atlas generation techniques are used to explore variability in ship traffic as a result of extreme weather and seasonal variation. Finally, anomaly detection is demonstrated by comparing AIS data from 2013 to the highways detected in the archive of data from 2012. The development of an automatic atlas generation technique that can be used to develop a definition of normal maritime behavior is the significant result of this thesis.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AIS	automated information system
AIVDM	received AIS data message
ASCII	American standard code for information interchange
COP	common operating picture
csv	comma separated value
GMM	Gaussian mixture model
GPS	global positioning system
KDE	kernel density estimator
MDA	maritime domain awareness
MMSI	maritime mobile service identifier
MSP	maritime security policy
nm	nautical miles
ROI	region of interest
VHF	very high frequency
WGS-84	world geodetic system established in 1984

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The research detailed in this thesis is a development of techniques that automate pattern extraction from large volumes of maritime vessel position data to enable region-based anomaly detection in the maritime environment. The extracted patterns, or *highways*, are detected using the Hough transformation and are compiled together into atlases of expected ocean traffic patterns that capture a geographical region-based definition of normal maritime behavior. The point-in-polygon approach facilitates geographical region-based anomaly detection of ships as compared to the generated atlases.

The atlas generation method can be broken into three key stages. In the first stage, a technique is developed to process position reports into a traffic density grid. In the second stage, a modified version of the Hough transformation commonly used in image processing is applied to identify linear patterns of elevated density in the traffic density grid. In the third stage, traffic density analysis along the highway is performed to define the highway's width. An iterative method enables the detection of less prominent highways until all significant traffic patterns have been extracted and compiled into an atlas of the area.

Anomaly detection is performed by comparing incoming ship position reports to the generated atlas. By using point-in-polygon solving techniques, each position report is identified as either "on," i.e., the report lies within a polygon defined by the width of the highway, or "off" a highway.

Archived automated information system (AIS) data from 2012 is used in this thesis to develop case studies of atlas generation and anomaly detection for validation of the developed techniques. The iterative process performed in an open ocean environment enables the detection of four significant highways crossing the southern Atlantic Ocean. More than 99.99% of the position reports received in the southern Atlantic Ocean during 2012 map to one of these four highways, even though they account for only 54% of the regional area.

The techniques developed in this thesis can be extended to detect non-linear highways as well. For example, in coastal regions, the highways can be detected by breaking the larger region into sub-regions and detecting the linear highways in each of the sub-regions, then using these highways as the piecewise-linear definition of a highway of some other shape.

Additionally, these atlas generation techniques can be used to define traffic variability in the maritime environment. As one example, some seasonal variability exists in ocean traffic patterns, and the comparison of spring, summer, fall, and winter atlases enables a measurement of this variability. As another example, the techniques developed in this thesis enable an automated assessment of how extreme weather systems like hurricanes impact ocean traffic patterns. A particular highway detected in the Caribbean Ocean disappeared as Hurricane Ernesto crosses its path in August 2012 but was reconstituted in the days afterward.

The anomaly detection techniques developed in this thesis can also be used to study ship behavior. As one example, the atlas generation techniques detect a highway along the western coast of Africa from data archived in 2012. If data in the same region from January 1, 2013 are compared to the detected highway, of the 148 ships that transit at least 150 nm during the day in question, only 24 fail to use the highway.

The significant contribution in this thesis is the exploration of a technique from image processing to the problem of maritime domain anomaly detection. The use of the Hough transformation, an image processing technique, to detect and quantify maritime vessel behavior patterns has not been observed before in literature. A second contribution is in the development of techniques to produce an atlas of ocean highways. The anomaly detection techniques explored in this thesis are also a significant contribution. Automating the initial determination of a vessel's geographic location as normal or abnormal can provide important first indications of a vessel's intentions without any requirement for expert analysis.

I. INTRODUCTION

The president of the United States issued National Security Presidential Directive NSPD-41 on December 21, 2004, which details the Maritime Security Policy (MSP) of the United States. It also established a committee to oversee the implementation of the National Plan to Achieve Maritime Domain Awareness (MDA) as one of eight supporting plans to the MSP [1]. In this document, MDA is defined as the effective understanding of anything associated with the global maritime domain that could impact the security, safety, economy, or environment of the United States [2]. The document emphasizes the need to identify threats in the maritime domain by integrating intelligence, surveillance, observation, and navigation systems into a common operating picture (COP) [2].

Data alone are not enough to achieve a useful COP; an ability to identify trends and differentiate anomalies is required for MDA objectives to be met [2]. There were more than 55,000 port calls in 2009 from nearly 7,000 different oceangoing vessels in the United States alone [3]. Rapid decision making and response in this busy environment requires automated methods to turn the volumes of raw data collected on these vessels into processed intelligence.

The research detailed in this thesis is a development of techniques that automate pattern extraction from large volumes of maritime vessel position data to enable anomaly detection. By doing so, this research directly contributes to the efforts for information exploitation detailed under the technology priorities in [2].

A. THESIS OBJECTIVE

The objective of this thesis is to contribute to MDA by developing an anomaly detection mechanism that identifies and labels a ship's behavior as normal or abnormal when compared to traffic patterns extracted from archived ship position data. The extracted patterns, or *highways*, are determined using the Hough transformation technique and are compiled together into atlases of expected ocean traffic patterns that capture a geographical-position based definition of normal maritime behavior.

The Hough transformation is used to build a model of expected vessel traffic based only on position reports, independent of any association to a particular vessel. The developed method was applied experimentally to an archive of historical data to demonstrate the concept of a complete anomaly detection architecture.

B. RELATED WORK

The related work for this thesis falls into two categories: that related to applications of the Hough transformation and that related to anomaly detection.

Hough transformation has been extensively used in the literature to identify linear regions in images. The Hough transformation is used in [4]–[8] to extract the location of roads from satellite images. Specifically, edge detection, the division of a larger region into a grid of small sub-regions, and optimal search techniques in combination with the Hough transformation to detect roads in each sub-region are used in [4], and then those detections are combined into a compilation of roads within the region of interest [4]. Alternatively, a wavelet transform to extract roads from remote imagery, even in the presence of noise is employed in [5]. The road detection problem is broken into two components in [6]: centerline extraction and width estimation by looking for parallel lines within a single image. Before the Hough transformation is applied in [6], Canny edge detection is used to preprocess the image so that the Hough transformation is effective. Region growing techniques to expand a single road into an understood network of roads are used in [8]. No literature was found using the Hough transformation in the maritime domain, but these references describe techniques for basic pattern extraction that are extended to the maritime domain for use in this thesis.

Anomaly detection in the maritime domain is a complex and multifaceted problem, and various approaches to building a robust anomaly detection technique are explored in the literature. An anomaly detection system for MDA that integrates statistical methods with qualitative or symbolic classification of ships to reduce analyst workload in the process of detecting illegal and hazardous activities on the oceans is proposed in [9]. The proposed system is based on interviews with maritime operators to identify real-world requirements and, if developed, would serve as a bridge between operators and surveillance systems to reduce analyst workload. Alternatively, the automatic development of a scheme of normal behavior patterns without the need for expert human input is a goal of this thesis.

An adaptive kernel density estimator (KDE) method is employed in [10] and [11] to develop a statistical model of expected vessel behavior based on position and speed. Alternatively, the KDE method is compared to a Gaussian mixture model (GMM) in [11], and it was found that KDE is superior to GMM. A statistical analysis of automated information system (AIS) data is employed in [10] to detect anomalies and predict future vessel behavior. Motion patterns are extracted from AIS data using KDE and are used to define normal behavior. Anomaly detection is based on ship motion compared to these predefined regions. A Gaussian sum tracking filter is employed to use the historical patterns to predict future movement. Both of these techniques use statistical methods to develop a model of past behavior based on individual vessel tracks, whereas patterns are extracted based on the position reports alone in this thesis.

Mechanisms for clustering, classifying, and detecting outliers in groups of vessels based on their behavior are developed in [12]. The method employs compression techniques to reduce the amount of data needed to understand vessel motion and similarity measurement powered by various alignment techniques for clustering, classification, and outlier detection. AIS archives provide the historical training data for the work described in [12] similarly to how they are used in the proof of concept for this thesis.

C. ORGANIZATION

Five chapters and two appendices are contained in this thesis. The background information related to AIS, the Hough transformation, and anomaly detection is covered in Chapter II. The pattern extraction techniques used in generating an atlas of normal behavior and how that atlas is used in anomaly detection are described in Chapter III. Specific details on how the techniques from Chapter III are employed in this research and selected case studies of atlas generation and anomaly detection are included in Chapter IV. A summary of key results and considerations for follow on work is provided in

Chapter V. Programming code in the Python language used in AIS data preprocessing is contained in Appendix A. The MATLAB programming code used to apply the Hough transform to maritime data is contained in Appendix B.

II. BACKGROUND

Three concepts that are central to the anomaly detection algorithm developed in this thesis are explored in this chapter. First, the data sources and formats available for developing MDA are discussed. Next, the Hough transformation and its potential applicability to improving MDA are broached. Finally, anomaly detection methods are examined.

A. AUTOMATED INFORMATION SYSTEM DATA

The pattern extraction and anomaly detection methods used in this thesis rely heavily on having accurate position reporting from vessels of interest. While these position reports could come from any means, including radar, satellite imagery, or observations recorded from trusted vessels, the development of AIS has created large archives of ship position reports that lend themselves well to research and development of MDA tools. The pattern extraction techniques could be used with any data source and more powerfully still with a fusion of multisource intelligence data sources, but only AIS data are used in this thesis as a proof of concept.

AIS was developed to provide ship operators with integrated displays of all ships within their very high frequency (VHF) radio range. It was conceived as a mechanism for improving safety at sea by enabling ships to clearly identify the other ships around them not just by location on a radar screen but also by specific name [13]. The system is now required by US Coast Guard regulation on all passenger vessels of more than 150 gross tons displacement or certified to carry more than 150 passengers-for-hire, all tankers, all vessels of more than 300 gross tons displacement, all sail boats over 65 feet in length, and all towing vessels over 26 feet in length [13]. Similar requirements have been implemented internationally. Estimates from 2012 indicate that more than 70,000 vessels are AIS equipped; projections indicate that as many as 150,000 vessels will be AIS equipped in the future [14].

Most AIS reports are transmitted at 12.5 watts, a signal strength that, coupled with their use of the VHF spectrum, enables them to be received by satellite-based collection systems. The data used for this thesis were collected by commercial satellite AIS collection companies that archive the data and sell it to customers for various research and operational management purposes. The data used in this research were collected by the exactEarth and ORBCOMM corporations. Both companies are using and expanding their own constellations of microsatellites, employing low earth orbits to provide global collection of AIS reports [15], [16].

AIS uses a self-organizing time-division multiple access scheme that enables users within radio range of one another to deconflict transmission. This can result in the receipt of as many as 7,500 messages per second at the satellite because there will be many local AIS deconfliction areas within a single satellite footprint [13], [14]. Each commercial company employs proprietary methods for satellite AIS detection, but the end product is an archive of position reports from all over the world [15]. These archives grow by approximately 300 MB every day as measured by the storage of this data in text formatted files.

There are a variety of AIS reports, but the most important to this thesis are the position reports that are transmitted every two to ten seconds by ships that are underway. These reports are transmitted in received AIS data message (AIVDM) formatted sentences with seven fields [17]. As an example, the message

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

is properly formatted with an AIVDM flag in the first field and the data payload in the sixth field. For position reports, this data payload contains 22 American standard code for information interchange (ASCII) characters that represent 128 bits of encoded data with four bits of padding. Each ASCII character represents six binary bits of information. The six bits are found by subtracting 48 from the standard value representation of the ASCII character so that ASCII characters “0” through “W” correspond to integer values 0 through 63 and binary values between 000000 and 111111. The bits are then related to various length data fields as displayed in Table 1. [17].

Table 1. The 16 AIS position report fields (after [13]).

Field	**Parameter**	**Number of bits**	**Units or Description**
1	Message ID	6	
2	Repeat indicator	2	Used by the repeater to indicate how many times a message has been repeated
3	User ID	30	A unique ID, usually the Maritime Mobile Service Identifier (MMSI)
4	Navigational status	4	Specific Codes for Underway, at anchor, restricted maneuverability and other navigational statuses
5	Rate of turn	8	degrees per minute
6	Speed over ground	10	1/10 knot
7	Position accuracy flag	1	

8	Longitude	28	1/10,000 min
			East = positive (as per 2's complement)
			West = negative (as per 2's complement).

9	Latitude	27	1/10,000 min
			North = positive (as per 2's complement)
			South = negative (as per 2's complement).

10	Course over ground	12	1/10 degree
11	True heading	9	Degrees
12	Time stamp	6	UTC second when the report was generated by the electronic position system
13	Special maneuver indicator flag	2	
14	Spare	3	Not used. Should be set to zero. Reserved for future use.
15	Receiver autonomous integrity monitoring flag	1	
16	Communication state	19	
	Number of bits	**168**	

The key fields for the data used in this thesis are 1, 3, 8, and 9. Position reports have a message ID of 000001, 000010, or 000011. The maritime mobile service identifier (MMSI) is a unique global identifier for ships. It can be used to correlate position reports into tracks for a particular ship or to cross-reference all information available for a ship from a fusion of data sources. The physical position report as determined by the global positioning system (GPS) using the world geodetic system established in 1984 (WGS-84) datum is located in fields 8 and 9. While a time stamp in seconds is located in Field 12, AIS systems only transmit the six binary digits representing the second that a report was transmitted. The receiver must use a local time reference to log the complete timestamp of the report. An example of how an AIVDM sentence can be decomposed into a position report is presented in Figure 1.

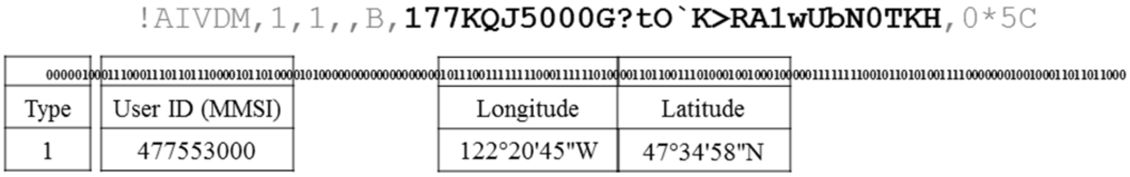


Figure 1. The data payload of AIVDM messages is stored in the sixth field in ASCII encoded binary that contains a message type, user ID, and position report.

B. THE HOUGH TRANSFORMATION

The human eye is capable of seeing how features are arranged in images, while computational algorithms are necessary to automate the same discernment [18]. For example, a viewer can discern specific patterns in the flow of maritime traffic when observing the map depicted in Figure 2 like the densely traveled paths between South America and Europe. These traffic patterns are to the oceans what highways are to the land. The techniques proposed in this thesis use the Hough transformation methods outlined in [18] and [19] and tailor those methods to accommodate data related to maritime vessel traffic so that the computer can extract the linear traffic patterns, or highways. Those extracted highways taken collectively into an atlas provide the computer with a template of normal vessel behavior to be used to comparatively identify anomalous behavior.

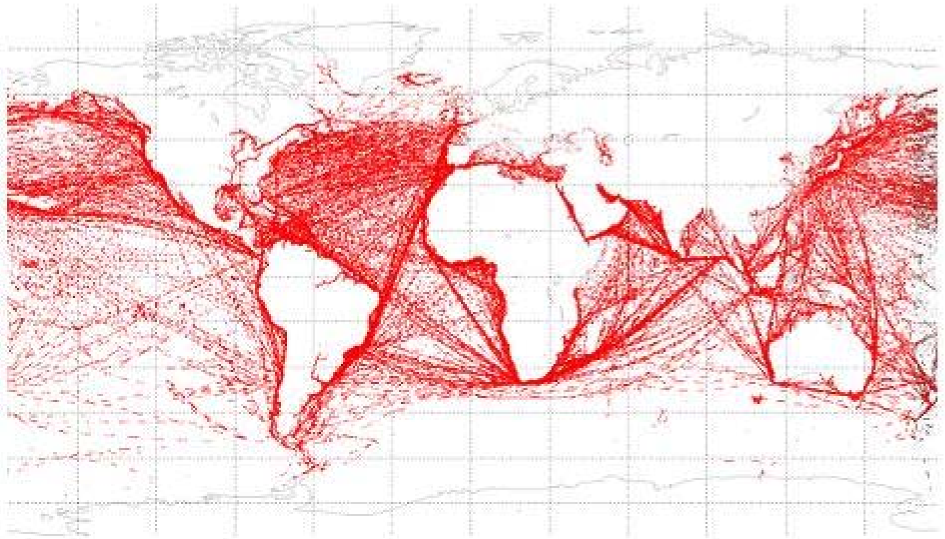


Figure 2. Traffic patterns are visible to the human eye when observing one million AIS position reports collected via satellite by the ORBCOMM constellation (from [16]).

1. Common Uses of the Hough Transformation

The Hough transformation, first published in 1962, is most commonly used to find alignments, or shapes, in images. The transformation can be adapted to detect any shape, but it has proved particularly useful in detecting linear patterns like those of roads [4]–[8]. Another example is using the Hough transformation to identify asbestos fibers by detecting and measuring the circular regions in electron diffraction patterns [18]. The transformation has also proved useful to robotics guidance and quality control applications because it can improve machine vision by enabling machines to gather measurements with sub-pixel accuracy [18].

2. Line Detection

Although the Hough transformation can be used to detect a variety of shapes in images, our work focused on the detection of linear regions only. As such, this description of the Hough transformation is tailored to describe straight line detections and is adapted from [18] and [19]. The Hough transformation is performed by translating a point in the reference axis system, or real space, to a set of definitions in the Hough space that exhaustively list the various straight lines that can be drawn through the point and how each of those lines relate to the origin of the reference axis system. To illustrate how the method is performed, a simple example is used to detect the co-linearity of the points P_1 through P_5 presented in Figure 3.

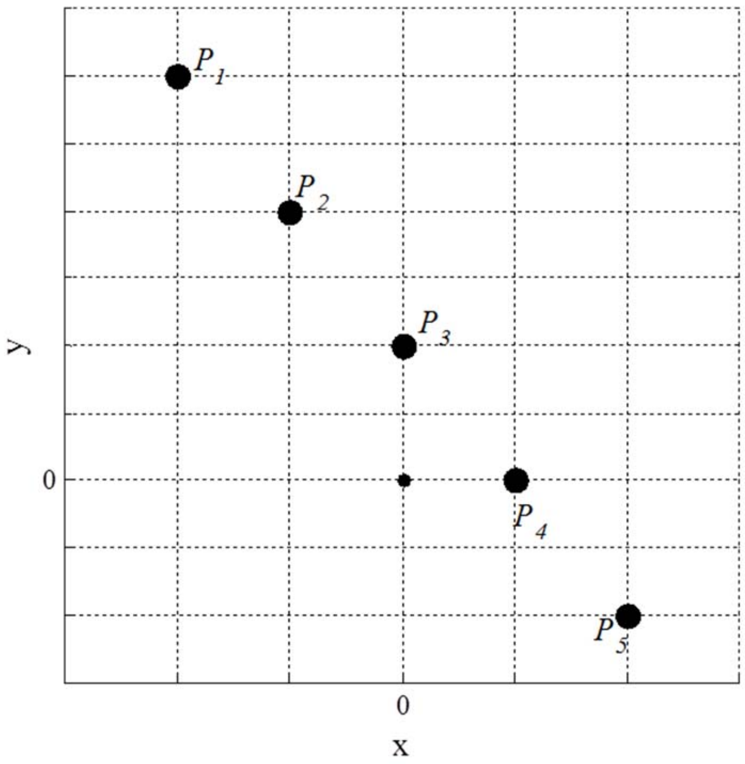


Figure 3. The Hough transformation is used to detect the co-linearity of points P_1 through P_5.

Every point P_n is transformed into a series of coordinates (d, θ). The first step in the process is represented in Figure 4. Every possible line $L_{\psi n}$ through P_n at an angle ψ with respect to a horizontal reference line is considered. Five examples of these lines are depicted in Figure 4. For each $L_{\psi n}$, the closest point of approach D between the origin of the reference axis system and $L_{\psi n}$ is identified in polar coordinates (d, θ) where d is the distance from the origin to D and θ is the angle between the horizontal axis through the origin and a line drawn from the origin to D.

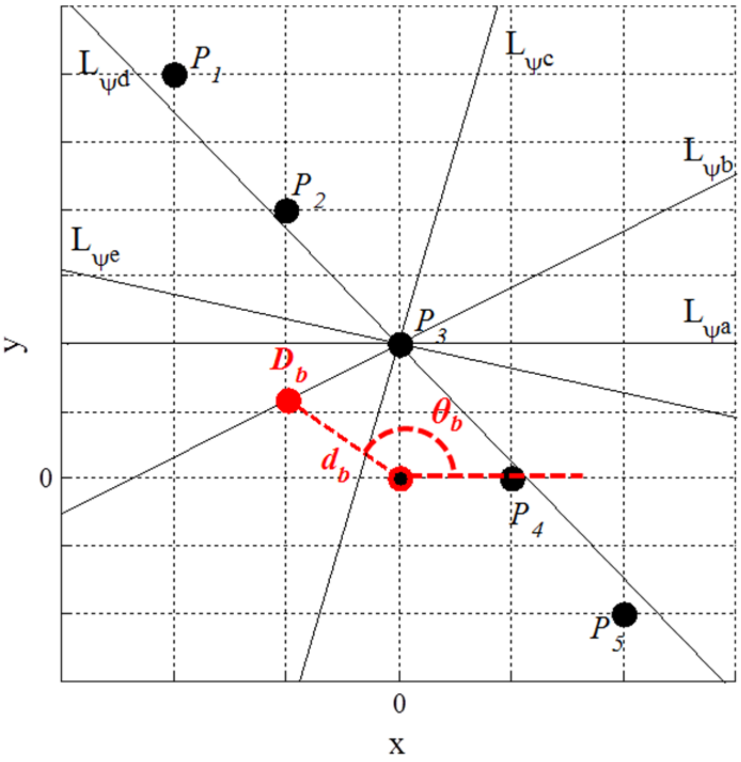


Figure 4. In the first step of the Hough transformation, the distance to the closest point of approach D must be found for every possible line through P_n.

The values of D_b, d_b, and θ_b are illustrated for P_3 and $L_{\psi b}$ in Figure 4. These values can be calculated by first defining Q to be a point that is one unit of distance away from P_n and along the same line that intersects P_n at angle ψ. The coordinates of Q are

$$(\mathrm{Q}_x, \mathrm{Q}_y) = (\mathrm{P}_{n,x} + \cos(\psi), \mathrm{P}_{n,y} + \sin(\psi)) . \tag{1}$$

Next, d can be found according to

$$d = \frac{\left|\det\begin{pmatrix} Q-P \\ O-P \end{pmatrix}\right|}{|Q-P|} \tag{2}$$

where O represents the origin located at (0, 0). The slope m can be found from

$$m = \frac{Q_y - P_y}{Q_x - P_x} \tag{3}$$

and the y-axis intercept b can be found from

$$P_y = mP_x + b . \tag{4}$$

Since D must also lie along the same line as P and Q, the coordinates of that point can be found from the simultaneous solution of

$$D_y = mD_x + b \tag{5}$$

and

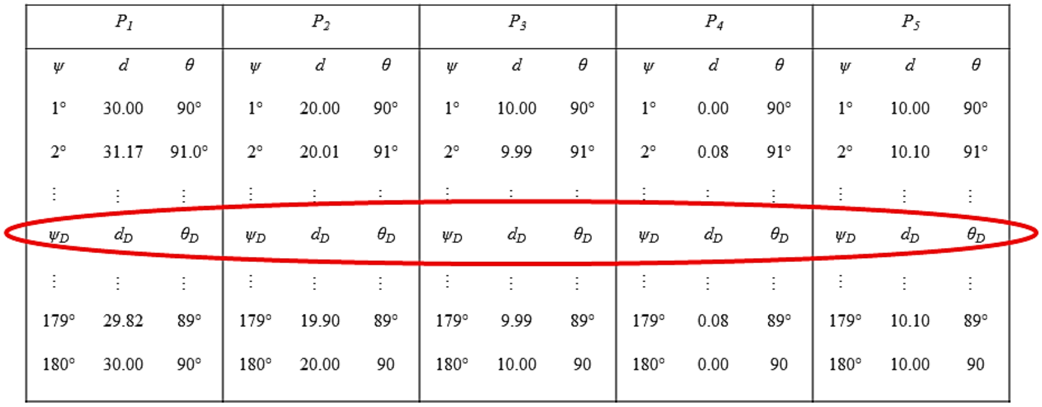
$$d = \sqrt{D_x^{\,2} + D_y^{\,2}}\,. \tag{6}$$

This enables the determination of θ from

$$\theta = \cos^{-1}\left(\frac{D_x}{d}\right). \tag{7}$$

This process is repeated for ψ = [1, 180] so that each point of interest P is expressed as a list of 180 (d, θ) pairs similar to Table 2. This collection of (d, θ) pairs is often considered the Hough space.

Table 2. The complete Hough space for the points in Figure 3 enables the determination of the common (d, θ) pair.



P_1			P_2			P_3			P_4			P_5		
ψ	d	θ	ψ	d	θ	ψ	d	θ	ψ	d	θ	ψ	d	θ
1°	30.00	90°	1°	20.00	90°	1°	10.00	90°	1°	0.00	90°	1°	10.00	90°
2°	31.17	91.0°	2°	20.01	91°	2°	9.99	91°	2°	0.08	91°	2°	10.10	91°
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
ψ_D	d_D	θ_D	ψ_D	d_D	θ_D	ψ_D	d_D	θ_D	ψ_D	d_D	θ_D	ψ_D	d_D	θ_D
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
179°	29.82	89°	179°	19.90	89°	179°	9.99	89°	179°	0.08	89°	179°	10.10	89°
180°	30.00	90°	180°	20.00	90	180°	10.00	90	180°	0.00	90	180°	10.00	90

Co-linear points have an identical entry of (d_D, θ_D) occurring in their coordinate sets at angle ψ_D. Visually, the presence of co-linearity results in an intersection at (d_D, θ_D) when (d, θ) plots for each P_n are overlaid as displayed in Figure 5.

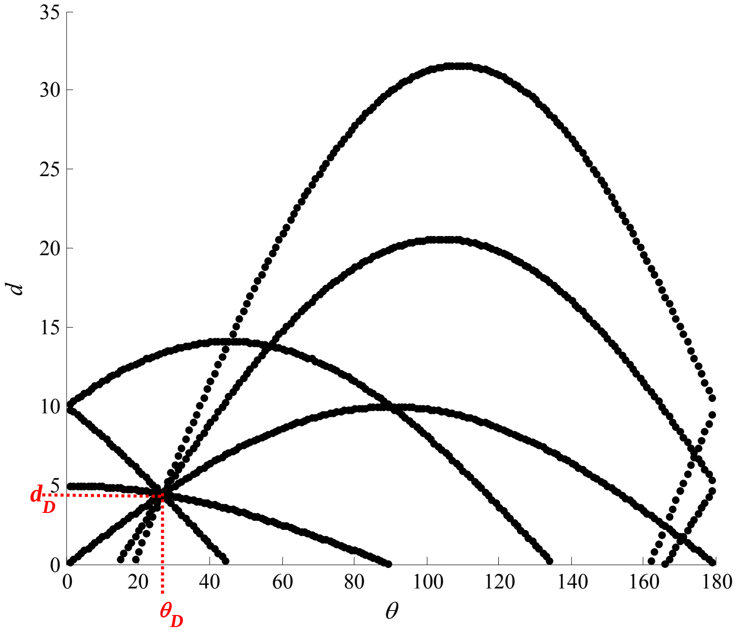


Figure 5. An intersection in (d, θ) plots for each P_n represents the detection of a co-linear region.

The final step is to return to the x-y coordinate plane and form the line through the co-linear point at angle ψ_D. The bounds of this line are found by identifying the extreme values of x and y from the set of coordinates that contain the set ($\psi_{D,}$ d_D, θ_D). Connecting these bounds results in the definition of a detected line as displayed in Figure 6.

In imagery applications, the intensity and color within a digital image are represented on a scale from [0, 255]. Co-linear points of interest with equal intensity represent the presence of a straight line in the image. For other applications, such as the MDA application investigated in this thesis, the data must be preprocessed so as to form a matrix similar to the digital representation of an image in order to be able to use the Hough transformation.

Other representations of the Hough transformation make use of the polar coordinate system to prevent infinite slope problems from hindering the detection of vertical patterns [18], [19]. In this thesis, the infinite slope problem was addressed by performing the Hough transformation on rotations of the same space.

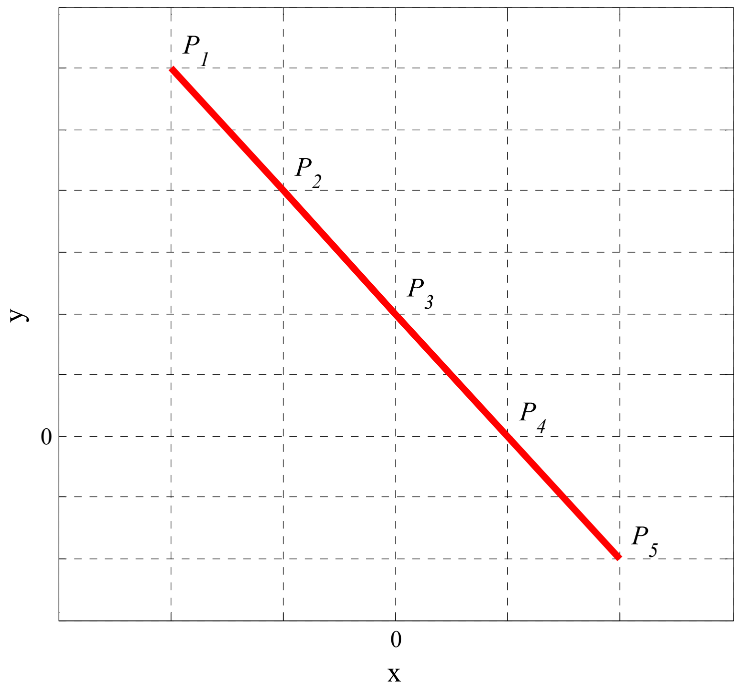


Figure 6. The final result of the Hough transformation is the definition of the line representing the co-linear regions of the image. In this example, the red line has been detected that describes the co-linearity of points P_1 through P_5.

C. ANOMALY DETECTION

Anomaly detection is the process of attempting to identify patterns in data that are unexpected or abnormal [20]. A direct approach to developing such a process is to draw clear boundaries around the region that describes normal behavior and identify anything outside of those boundaries as abnormal. Very few real world problems, however, lend themselves to clear definitions of every possible normal behavior. As such, most anomaly detection mechanisms are developed to solve very specific components of a larger analysis process. The broad application of that framework to MDA is examined and the portion of the larger MDA anomaly detection problem that this thesis works to solve is outlined in this section.

1. Classification of Anomaly Detection Systems

It is useful to consider a framework for comparing anomaly detection systems. One framework includes consideration of input data, type of anomaly, data labeling, and output [20].

The first aspect of an anomaly detection system is the nature of the input data that will be used [20]. For the MDA problem, multivariate data describing vessel traffic on the oceans are available. As has been described, position reports from AIS are one available data source, but even within AIS, ships also transmit static data reports containing detailed vessel information that could be of use in anomaly detection. Outside of AIS, other sources of MDA data include coastal and shipboard radar sensors, port records, ship registry records, satellite imagery, weather observations, buoy constellations, and an endless variety of other data associated with the seas. A robust anomaly detection system for MDA would need to absorb data from multiple sources.

The second aspect of anomaly detection is the type of anomaly the system will identify. Anomalies may be of point, contextual, or collective in nature. Point anomalies refer to the ability to identify a single data record as anomalous. Contextual anomalies are more complex in that they include a contextual case in which behavior is anomalous, whereas in a different context, identical behavior might be normal. Collective anomaly detectors identify sets of data that are, as a group, anomalous as compared to the entire data set [20]. All three types could be of use in MDA, but a truly exhaustive system needs both collective and contextual perspectives. Any recreational water area provides ready examples: taken as a point anomaly, it is not unusual for either a ski boat or a fishing boat to be in the middle of a lake. In a collective and contextual approach, comparing a set of observations of the two watercraft, it would be far more unusual for the ski boat to loiter in one spot than it would be for the fishing boat. Further, it would be abnormal for either to loiter in that same spot in the middle of a thunderstorm.

The third aspect of an anomaly detection system is to have a labeled data set where each record has already been identified as normal or abnormal. A completely exhaustive data set generally requires human expertise and can be difficult to obtain. As such, the system can be classified as *supervised, semi-supervised*, or *unsupervised* depending on the availability of labeled data, with supervised systems requiring a complete set of labeled data of both normal and abnormal behavior. Semi-supervised systems require a definition of normal behavior but not of abnormal. Unsupervised data sets require neither. Unsupervised systems are generally statistical in nature and overcome an absence of labeled data by assuming that normal is also the most statistically frequent behavior [20].

For MDA, thoroughly listing every possible normal and abnormal behavior is daunting and even potentially impossible. It is most likely that a complete MDA anomaly detection solution will involve some hybrid approach wherein human expertise has identified a set of normal and abnormal behaviors, but statistical methods are also used in detection to cover cases that may not have been considered.

The fourth aspect of anomaly detection systems is their output. These may be either qualitative or quantitative in nature. The qualitative, or *labeled*, solution is purely to identify behavior in a binary fashion as normal or abnormal, whereas the quantitative, or *score*, approach involves assessing some degree of normality on an established scale [20]. Both approaches could be useful in MDA, but the score-based approach affords more ability to infer meaning directly from the abnormal classification. In the recreational example, the score-based system might identify a loitering jet ski as less strange than a loitering jet ski in a thunderstorm, which could conceivably better enable emergency responders to prioritize their actions based directly on the anomaly detection system output.

2. The Point-in-polygon Problem

If expected behavior can be bounded within some geographic region, then qualitative anomaly detection becomes a problem of identifying whether detected behavior geographically lies within those boundaries. The following techniques for doing so are provided in [22]. To describe this technique, we determine if test point (x, y) lies within a polygon defined by two vectors v_x and v_y with the x and y coordinates of the outline of the polygon, respectively, as depicted in Figure 7.

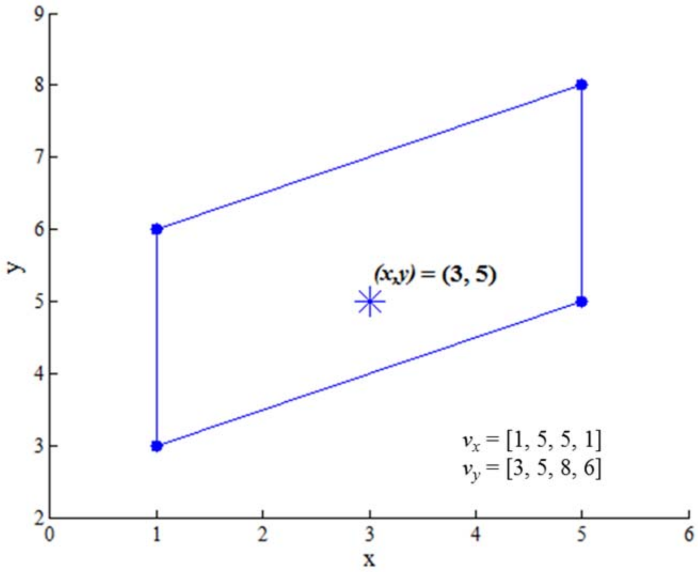


Figure 7. An example point-in-polygon problem begins with a point (x, y) and a polygon defined by vectors of coordinates v_x and v_y.

First, a reference coordinate system is set up with the test point at its center, translating the polygon definition to the new coordinate system as indicated by

$$\begin{aligned} v_x' &= v_x - x' \\ v_y' &= v_y - y' \end{aligned} \tag{8}$$

where

$$x' = \begin{bmatrix} x \\ x \\ \vdots \\ x \end{bmatrix} \tag{9}$$

and

$$y' = \begin{bmatrix} y \\ y \\ \vdots \\ y \end{bmatrix}. \tag{10}$$

The quadrant relative to the test point in which each of the points (v_x',v_y') rests is identified according to the reference values seen in Figure 8. The sequential list of each of the values moving around the polygon makes up the set of values *Qr*.

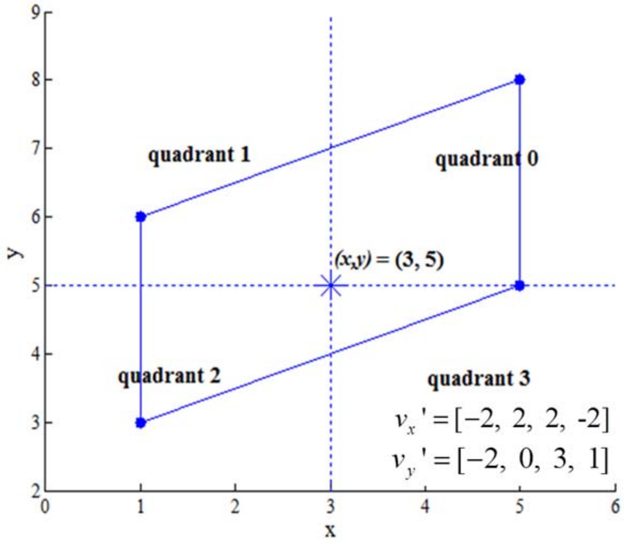


Figure 8. A numbered quadrant reference system relative to the test point enables a solution to the point-in-polygon problem.

For the example problem, *Qr* is determined by identifying the quadrant of the bottom left corner of the polygon, then the next corner moving counter-clockwise, and so on until all points of the polygon have been considered, resulting in *Qr* = [2, 0, 0, 1]. Next, the differences of vector *Qr* are calculated to create a new vector listing how many quadrants are between each consecutive point in the definition of the polygon. If a counterclockwise rotation is required to get from the quadrant of one point to the quadrant of the next, the difference has a negative sign. Clockwise rotations carry positive signs. For example, in a shift from the third quadrant to the zeroth quadrant, the difference is −1; a shift from the first quadrant to the third quadrant is a difference of +2. The difference vector of the example will be [2, 0, 1, 1]. If the sum of all of these differences is zero, then the test point is outside of the polygon. If the sum has any value other than zero, then the test point is inside of the polygon. Continuing with the example, we see that the sum of the difference vector is not zero, and thus, *(x, y)* lies inside of the polygon.

If the test point is changed to (3, 4), the reference axis changes to that displayed in Figure 9. In this alternate example, *Qr* = [1, 0, 0, 1]. The difference vector follows as [−1, 0, 1, 0]. This vector has a sum of zero, identifying the test point as lying outside of the polygon.

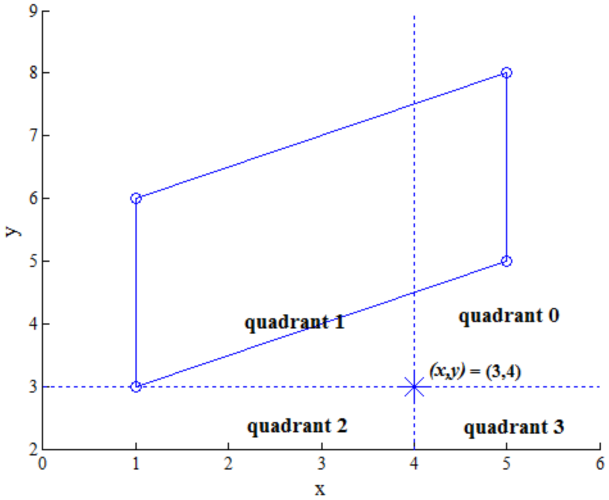


Figure 9. In a test case conducted with a point outside of the polygon, the reference quadrant system changes.

A geographic anomaly detection system for MDA can be instantiated with this point-in-polygon technique by using an archive of data to identify and bound the geographic regions of expected behavior. When a point of interest or a collection of points of interest are to be flagged as normal or abnormal, they are each compared to the polygon defining the region of expected behavior and flagged as *inside* and *normal* or *outside* and *abnormal.*

Background concepts necessary for the development of the techniques presented for anomaly detection in the maritime domain were examined in this chapter. The AIS system and its data collection process were discussed. Additionally, the Hough transformation, anomaly detection, and one solution to the point-in-polygon problem were described.

III. ANOMALY DETECTION TO IMPROVE MARITIME DOMAIN AWARENESS

One approach to improving MDA is to develop an automated system capable of highlighting anomalous ship tracks that do not fit the expected behavior patterns so that such ships can be flagged for follow-on analysis. One such system following the flow chart depicted in Figure 10 is developed in this chapter. The chapter is broken into two parts. Atlas generation is covered in the first; how the generated atlas is used for anomaly detection is explained in the second.

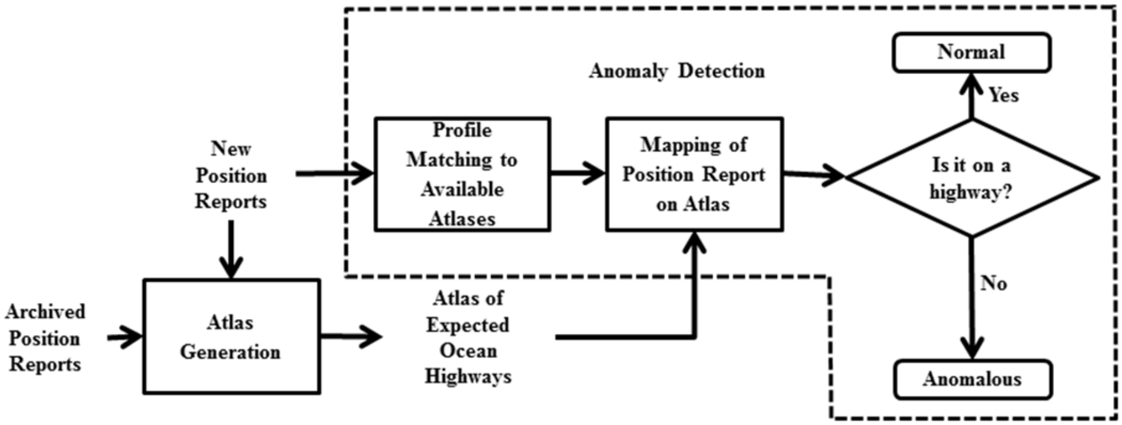

Anomaly Detection
New Position Reports
Profile Matching to Available Atlases
Mapping of Position Report on Atlas
Is it on a highway?
Yes
Normal
No
Anomalous
Archived Position Reports
Atlas Generation
Atlas of Expected Ocean Highways

Figure 10. The anomaly detection algorithm uses atlases generated from historical data to determine if real-time ship position data is normal or anomalous as compared to the historically observed behavior for an area.

Based on the classification system outlined in Chapter II, this system comprises a collective, position-based, unsupervised labeled anomaly detection system with some limited contextual cuing [20]. These factors are detailed in Table 3.

First, archived historical vessel tracking data are used to develop a variety of atlases that depict the expected behavior patterns on the oceans. Just as with terrestrial road maps, the atlas encompasses the collection of all expected traffic patterns, which are termed “highways.” Second, a comparison method is described for matching incoming data with available atlases and then determining whether or not a ship is on one of the highways within that atlas. The concepts used to implement each of these components are detailed in this chapter.

Table 3. The anomaly detection system can be classified according to the categories outlined by [20].

Aspect	Value	Description
Data Input	Multivariate	Geographic Position Reports Time Period of Interest Region of Interest Various Filtering Techniques
Type of Anomaly	Collective and Contextual (limited)	Collective in that the anomaly is based on a vessel track comprised of a collection of observations of geographic positions
Type of Data Set	Unsupervised	No human expert is involved in the system, but rather the statistical patterns that exist in an archive of data are assumed to represent normal behavior
Output	Labeled	A ship is either "anomalous" or "normal"

A. ATLAS GENERATION

Archives of data on ship tracks provide a wealth of information from which expected ship behavior over a variety of scenarios can be discerned. A common quality in all of the related work from Chapter I is that vessel tracks are preserved throughout the anomaly detection process. As an alternative technique, a common image processing technique is used in this thesis to develop an atlas of ship motion based on statistical analysis of position reports with no need for unique vessel identification.

The nine distinct steps depicted in Figure 11 outline the approach taken in this thesis to develop an atlas of expected ocean highways from archives of ship position reports.

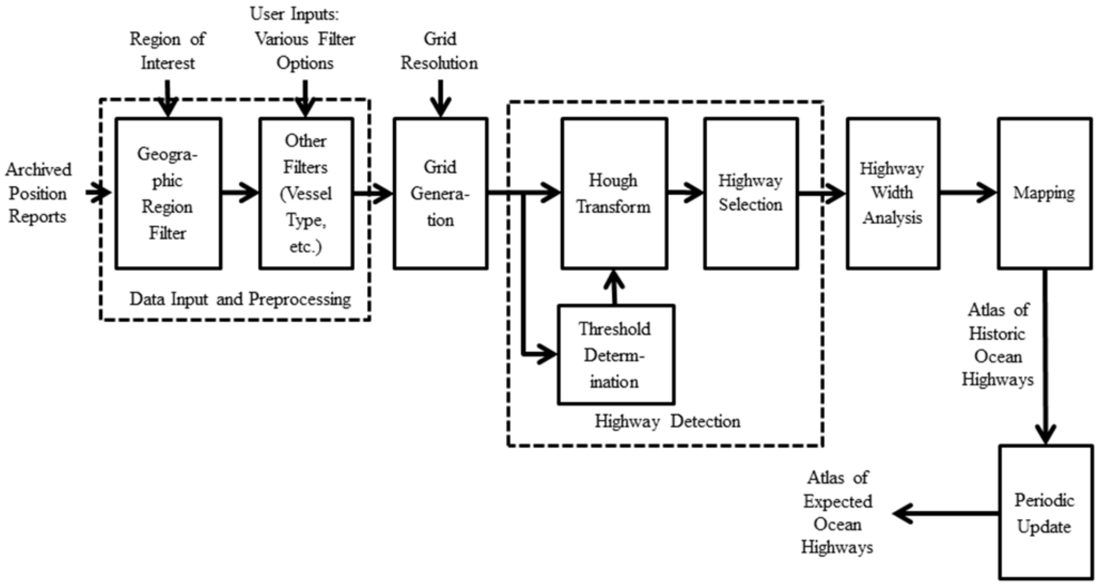


Figure 11. The atlas of historic ocean highways is tailored to a user's needs by taking input from the user on the region of interest, various filter options, and the required resolution of the area.

1. Data Input and Preprocessing

Preprocessing involves accepting data in the sensor output format and filtering and preparing that data so that follow-on steps can be correctly applied. This is the first step in the flow process outlined in Figure 11. The available archive of data is first filtered by timestamp to the window of interest and then filtered by latitude and longitude to the geographic area of interest. Data is then further filtered as necessary for other details, such as vessel type or flag of origin, depending on user preferences.

The key inputs to the method include an archive of position reports and three user inputs determining the region of interest, various filter options, and the grid resolution. The generic method enables the user to create an atlas; the various filtering options applied to the archive of data enable that atlas to be specifically matched to scenarios of interest. An archive of AIS position reports collected globally via satellite is used in this thesis, but the archive could just as easily be from any sensor capable of recording position reports. As one alternate example, in local recreational areas, it might be more useful to employ an archive of position reports collected via a coastal radar system since personal watercraft do not generally employ AIS.

The user also selects the region-of-interest (ROI). From the complete archive of position reports *G*, the data is filtered to include only the ROI as given by

$$G_{ROI} = G\{long_{\min} < G_x < long_{\max}; lat_{\min} < G_y < lat_{\max}\} \tag{11}$$

where G_{ROI} denotes the data only from the ROI, G_x and G_y represent the coordinates of each position report in longitude and latitude, and *long* and *lat* min and max represent the user defined boundaries to the ROI. Case studies in Chapter IV of this thesis exemplify the ability to select coastal regions or open-ocean areas.

Additionally, filters can be put in line to build the atlas from position reports of any category of data collected concurrently by the sensor recording position data. With the plethora of data transmitted in AIS reporting systems, atlases can be tailored to only indicate the historically expected highways employed by cargo vessels or only those employed by fishing vessels of less than a certain length, as just two examples. With other data sources and atlas applications, it might be more useful to filter data based on time of day or speed of vessels.

Once the original data set is filtered to meet the user's need, the data set is pared down to just the latitude and longitude pairs indicating ship position. No other data is used for atlas generation after the initial filtering occurs. In the case of AIS, the original position report has 16 fields of data as outlined in Chapter II, but only the sixth and seventh fields are used for atlas generation beyond this point.

2. Grid Generation

Grid generation is next in the process as we continue through the flow diagram outlined in Figure 11. The user also determines the grid resolution that should be employed. The highway detection method in follow-on steps uses the relative density of position reports in disparate sections of the ROI to detect the most frequently traveled routes in that area. Thus, the ROI must be divided into a grid with an equal number of rows and columns that can be used to develop a count of how many position reports occur in each region of the grid. The mesh size for this grid is determined by the user because the same grid size is not applicable in every scenario. The mesh size refers to the size of each region of the grid along the axis of longitude in minutes. This mesh size impacts the quality of the results. Many factors impact what the "right" value is for a given region. To better illustrate the impacts that might occur, four different grid sizes are used to produce the traffic density grids shown in Figure 12. The balance of finding a value that is fine enough without being too fine is explored in the next subsection.

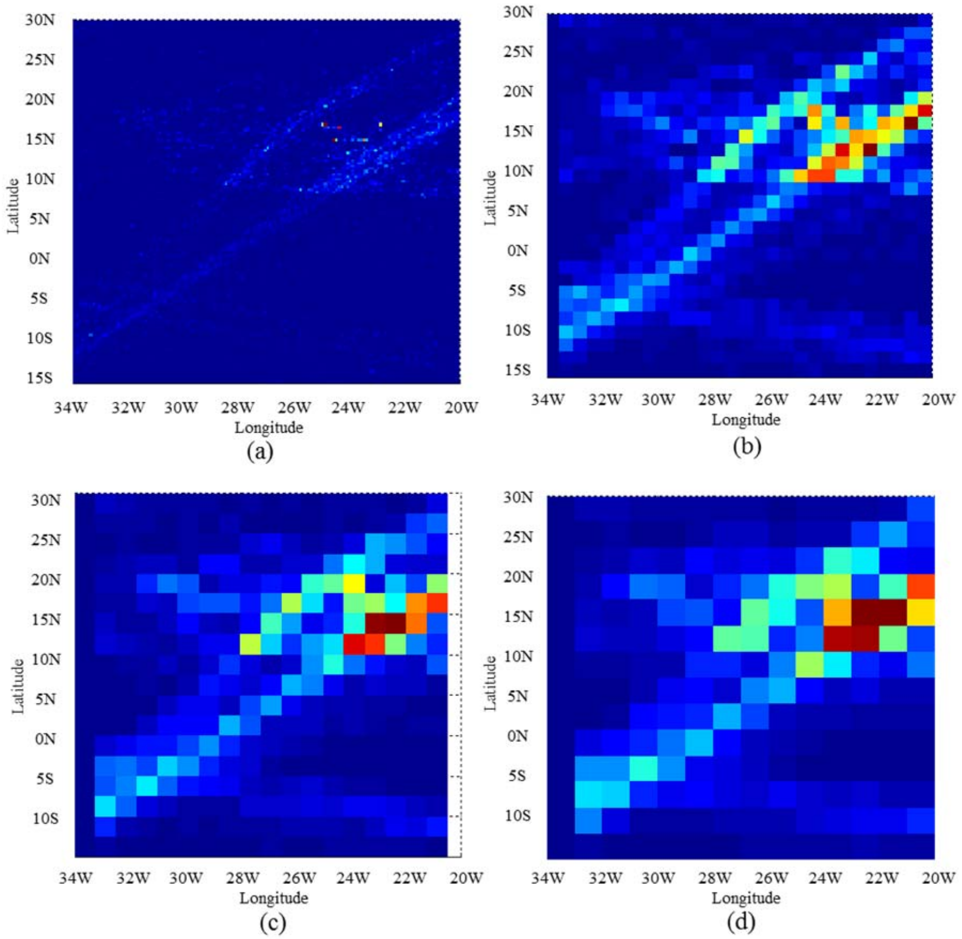


Figure 12.   Comparing grid sizes (a) 0.10 minutes, (b) 0.50 minutes, (c) 0.75 minutes, and (d) 1.00 minutes provides insight into how to select the best size. Reductions in grid size increase processing time and result in lost highways, yet larger grid sizes reduce the accuracy of highway placement.

a. Selecting a Fine Enough Grid Resolution

As in most digital applications, increasing resolution is generally desirable; however, there are some limits in this application. First, the significant figures measured by the position report source set a lower limit of what resolution is possible. For example, GPS reports in AIS are only received to the fourth decimal place in minutes of longitude. Setting a mesh size less than this results in mesh squares tied to position reports that can never exist in the data archive and creates false zero counts in the water space. Second, finer grids result in higher computational costs as the number of individual regions of the grid that must be considered grows. Third, less apparent associations may be overlooked. These missed highways are somewhat equivalent to the false negative problem in radar analysis – a highway may be overlooked when it actually does exist. To illustrate, if two ships are travelling on parallel courses with just 1.5 nautical miles (nm) of separation, or approximately 0.0250 minutes of longitude at the equator, and the grid is too fine, this may not be detected as a highway because of the zero traffic density regions between the two vessels. This impact becomes more pronounced at extreme latitudes where a degree of longitude represents a much smaller distance in nautical miles. For example, at the Arctic Circle, 1.5 nm of separation equates to 0.06 minutes of longitudinal separation. For reference, most vessels, with the exception of tugs and similarly specially equipped vessels, generally work to maintain separations of 1.5 nm for safety.

Although the regions are detected computationally rather than by observance, comparing the strength of the pattern that the human eye can detect as the size changes between the grids shown in Figure 12 is analogous to understanding how the grid size impacts the algorithm's ability to detect linear regions. The strength of the linearity of the trend line that crosses from the lower left hand corner to the upper right hand corner becomes more apparent as the grid size increases. With a grid size of only 0.1 minutes of longitude as in Figure 12(a), the trend is more difficult to visually distinguish from the low densities of the open ocean.

b. Selecting a Large Enough Grid Resolution

Selecting a mesh size that is too large can have negative impacts too. To begin, the grid size impacts the exactness of highway placement. Linear traffic patterns are detected from the grid, and the highway is overlaid across the associated grid squares. As a result, the placement has an inherent measurement accuracy of no more than +/- the diagonal length of a grid square. Further, nonexistent highways may be incidentally detected in a mesh that is too large. Equivalent to a false positive in the radar analysis realm, this relates to a highway being detected that does not really exist. By again observing the linear regions in the four grid options depicted in Figure 12, this point can be more clearly explained. At a grid size of 0.50 minutes of longitude as in Figure 12(b), the area of the ocean in the upper right corner appears to potentially have two linear regions that are somewhat parallel, separated by a low traffic density region. When the grid size is increased to 1.00 minutes of longitude as in (d), the low traffic density region between the two linear regions of higher traffic density is less apparent.

c. Final Grid Determination

Once a grid size Δ_x has been selected empirically, a vector of longitude grid region boundaries b_x is established according to

$$b_x = [Long_{\min}, Long_{\min} + \Delta_x, Long_{\min} + 2\Delta_x, ..., Long_{\max}] \qquad (12)$$

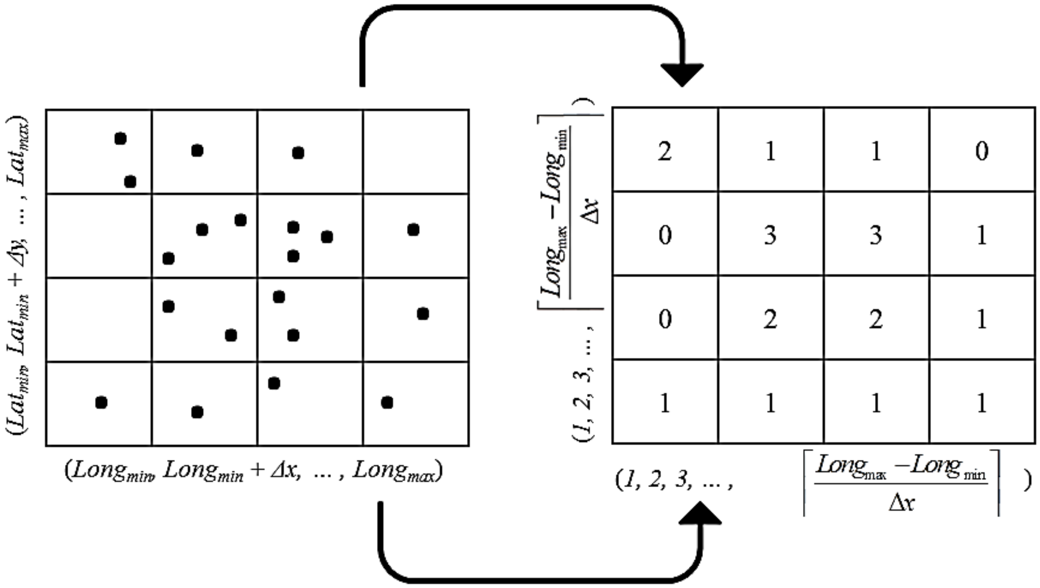
where the boundaries of the ROI are $[Long_{\min}, Long_{\max}]$ along the longitude of the ROI. The latitude steps are determined such that the grid is referenced over the ROI using an equal number of columns and rows through

$$b_y = [Lat_{\min}, Lat_{\min} + \Delta_y, Lat_{\min} + 2\Delta_y, ..., Lat_{\max}] \qquad (13)$$

where Δ_y is found from

$$\Delta_y = \frac{Lat_{\max} - Lat_{\min}}{\left\lceil \dfrac{Long_{\max} - Long_{\min}}{\Delta_x} \right\rceil}. \qquad (14)$$

The developed grid is used as an overlay above a map of ship position reports to develop a count of how many position reports exist in each region. An example of this process is depicted by Figure 13.


$(Lat_{min}, Lat_{min} + \Delta y, \dots, Lat_{max})$
$(Long_{min}, Long_{min} + \Delta x, \dots, Long_{max})$
$(1, 2, 3, \dots, \left\lceil \frac{Long_{max} - Long_{min}}{\Delta x} \right\rceil)$
$(1, 2, 3, \dots, \left\lceil \frac{Long_{max} - Long_{min}}{\Delta x} \right\rceil)$

2	1	1	0
0	3	3	1
0	2	2	1
1	1	1	1

Figure 13. The traffic density grid is formed by mapping position reports to the grid and counting how many position reports occur in each region of the grid to produce a matrix H of these counts. The grid has an equal number of columns and rows.

The matrix of these counts H, depicted on the right side of Figure 13, is normalized to generate a new matrix H_N by performing

$$H_N = \frac{H}{\max(H)}. \quad (15)$$

3. Highway Detection

Once the grid has been developed, the Hough transformation can be used to detect highways, which is the third major section of the flow diagram depicted in Figure 11.

a. Traffic Density Threshold Determination

Before the automatic detection of these co-linear densities can occur, a threshold for traffic densities of interest ρ_{TH} must be selected. The grid contains normalized traffic densities, so ρ_{TH} is some percentage of the maximum traffic density present in the ocean space. Any value can be selected, but in this thesis, the 75th percentile served as a reasonable starting point for most case studies. For comparison, the exact placement of the 75th percentile in a box plot of all traffic densities for two different grid sizes are depicted in Figures 14 and 15. This example uses a relatively empty open-ocean area with a strong diagonal traffic pattern. Within the figures, blue represents areas of low traffic density, with warmer colors of yellow and red indicating areas of higher traffic density. Using the percentile value rather than a specific traffic density couples ρ_{TH} to the previously determined grid resolution and empowers the algorithm to detect highways even as they begin to disappear to the human eye. The larger grid size depicted in Figure 14 requires a ρ_{TH} of 0.100 to consider the top quartile of regions for highway detection. The smaller grid size depicted in Figure 15 requires a ρ_{TH} of 0.025 to consider the top quartile of regions for highway detection.

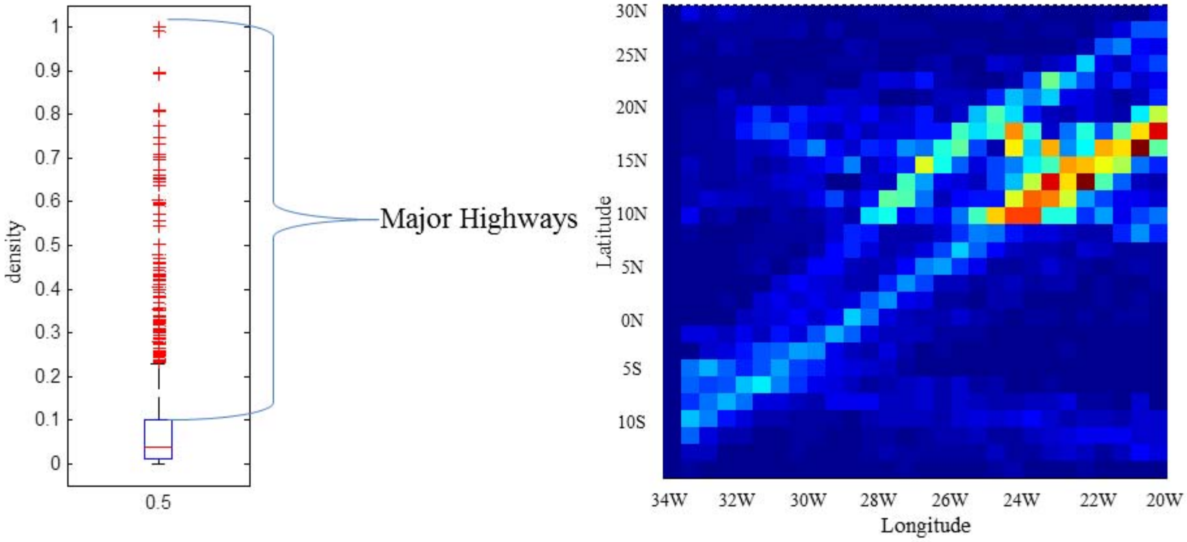


Figure 14. Looking at a box and whisker plot of ship densities for several grid sizes provides a mechanism for identifying the ship densities that occur in the most heavily travelled areas of a water space. With a grid size of 0.5 minutes of longitude, the 4th quartile densities range from 0.100 to 1.000.

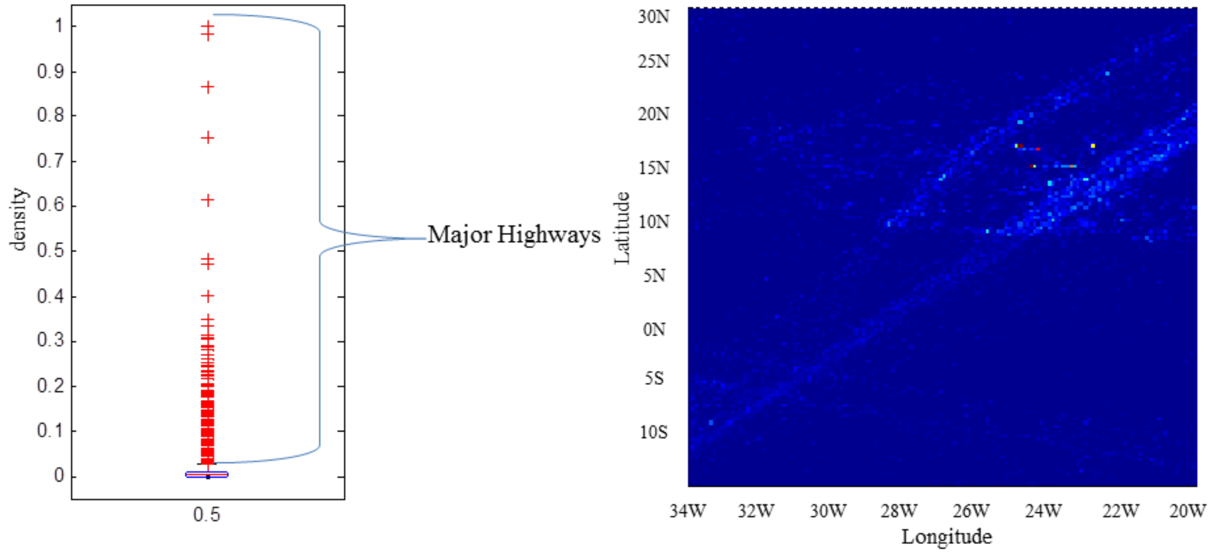


Figure 15. With a grid size of 0.1 minutes applied to the same area as depicted in Figure 14, the 4th quartile normalized densities now range from 0.025 to 1.000.

b. *Co-linear Threshold Determination*

In addition to the traffic density threshold, two other thresholds are used within the highway detection portion of the algorithm. As explained in Chapter II, elements of the grid matrix with a (d, θ) pair in common in the Hough space are co-linear. In the MDA application, tolerance factors are applied to both d and θ to enable highway detection in cases where a highway is more than one grid square wide or has slight disparities in slope over its length.

c. *Identification of the Dominant Trend*

Each region of the grid that meets or exceeds the traffic density threshold is transformed in the Hough space using the Hough transformation techniques described in Chapter II. To summarize, every region of the grid

$$R(x,y) > \rho_{TH} \quad , \qquad (16)$$

where x and y denote the row and column indices within the grid, is transformed into the Hough space where it becomes a set of values

$$S_{(x,y)} = \{(d, \theta)\} \tag{17}$$

that includes the (d, θ) pair of all possible straight lines through that grid region at indices (x, y). From the Hough space, co-linear regions are detected by identifying common (d, θ) pairs within each set *S* between different grid regions.

As the term is used in this thesis, the "dominant" trend refers to that linear region with the most co-linear grid regions that meet or exceed ρ_{TH}. The dominant trend pair (d_D, θ_D) can be found by performing a frequency study of (d, θ) occurrences across all grid regions. The most frequently occurring pair represents the dominant trend.

Although many applications directly use the straight-line definition derived from the Hough space to define the line detected in the real space, the maritime domain requires special consideration. Because highways may be of any width and, more specifically, because that width may change over the length of a highway, the direct result of the Hough space may not be the most comprehensive definition of the highway. Instead, a first-degree polynomial is fitted to the collection of all regions of the grid for which

$$S \cap (d_D, \theta_D) \neq 0 \tag{18}$$

is true. This polynomial fits the common form of

$$y = p_1 x + p_2 . \tag{19}$$

with p_1 describing the slope of the highway within the grid and p_2 describing where the line would intersect with the left edge of the grid. This polynomial best fits the collection of co-linear grid regions in the least-squares sense. An example of the placement of this centerline for a traffic density grid is displayed in Figure 16.

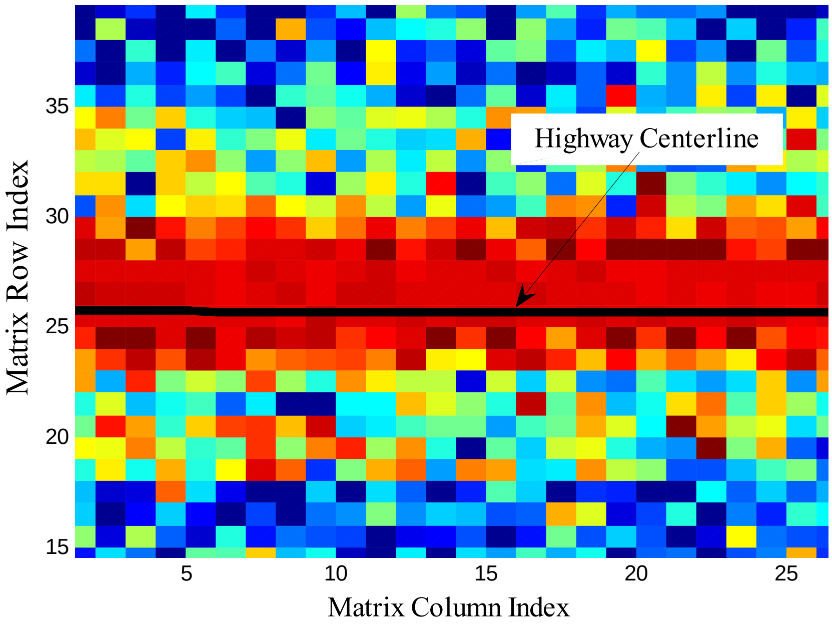


Figure 16. The highway centerline will be mapped across the traffic density grid and represent the location of the dominant trend.

Because the first-degree polynomial is of the form of Equation (17), it cannot be used to develop a model of highways that run vertically. Instead, vertical highways can be detected by rotating the grid 90 degrees clockwise before performing the Hough transformation and then rotating the resultant highways 90 degrees counter-clockwise before follow on steps are performed. This requirement to run the method in two orientations can be overcome by implementing a polar polynomial fitting technique, but this solution was not further explored within this thesis. One parameterization enables vertical line detection by using [18]

$$\rho = x\cos(\theta) + y\sin(\theta) \tag{20}$$

as an alternative form of Equation (17) where, rather than describing a line by its slope P_1 and intercept P_2, the line is described by the length ρ and angle θ of a normal to the line that runs through the origin.

4. Highway Width Analysis

Once the centerline of a highway is identified, a width study is performed to determine the left and right limits of the highway as seen from the process flow depicted in Figure 11. Terrestrial highways are relatively uniform in width, with changes occurring only as lanes are physically added to or removed from the infrastructure of the highway, but oceanic highways are not uniform since they are not generally physically bounded. As such, the width must be determined incrementally along the highway to best capture its variability.

The algorithm determines the width at each incremental step along the highway by considering the densities of the regions of the grid along a line perpendicular to the detected highway at the given step. The highway width is located at the index of the last region of the grid along that line that still meets or exceeds a traffic density threshold ρ_w. The width is different on either side of the highway centerline and at each incremental step.

The threshold ρ_w can be set to any value, but one technique for determining it comes from finding the parameters of mean μ and standard deviation σ for the normal distribution curve fitting a profile of the densities across the width of the highway at each incremental step. The location of the mean traffic density μ_i at each increment is found according to

$$\mu_i = \frac{\sum_j j\rho_j}{\sum_j \rho_j} \tag{21}$$

where i is the index of the incremental step along the highway centerline, j is index of the cross section of the highway width taken at a given increment, and ρ_j is the traffic density of the grid region at the index j. The standard deviation σ_i at each increment is found according to

$$\sigma_i = \sqrt{\frac{\sum_j j^2\rho_j}{\sum_j \rho_j} - \mu_i^2}\ . \tag{22}$$

The mean μ and standard deviation σ are found by taking the arithmetic mean where

$$\mu = \frac{\sum_i \mu_i}{i} \tag{23}$$

and

$$\sigma = \frac{\sum_i \sigma_i}{i} . \tag{24}$$

If ρ_w is set to define the highway as all regions that are within one standard deviation of the mean, then an adjusted value of threshold ρ_w' can be found directly from the normal distribution

$$f(x) = Ke^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{25}$$

from [21] where the factor

$$K = \frac{1}{\sigma\sqrt{2\pi}} . \tag{26}$$

The desired location of x is $(\mu - \sigma)$, or one standard deviation below the mean. This substitution results in the determination of ρ_w' through

$$\rho_w' = Ke^{-\frac{((\mu-\sigma)-\mu)^2}{2\sigma^2}} = Ke^{-\frac{1}{2}} . \tag{27}$$

One example of determining these values from the distribution of densities across the width of a highway is displayed in Figure 17. The profile of traffic densities across the highway width is used as a histogram of ship locations. The profile appears normal in shape, and a normal distribution is fit to it as displayed in red in Figure 17. The value of ρ_w' corresponds to the traffic density of the normal curve at the matrix row index where the normal curve has fallen to one standard deviation below the location of the mean, as indicated in Figure 17.

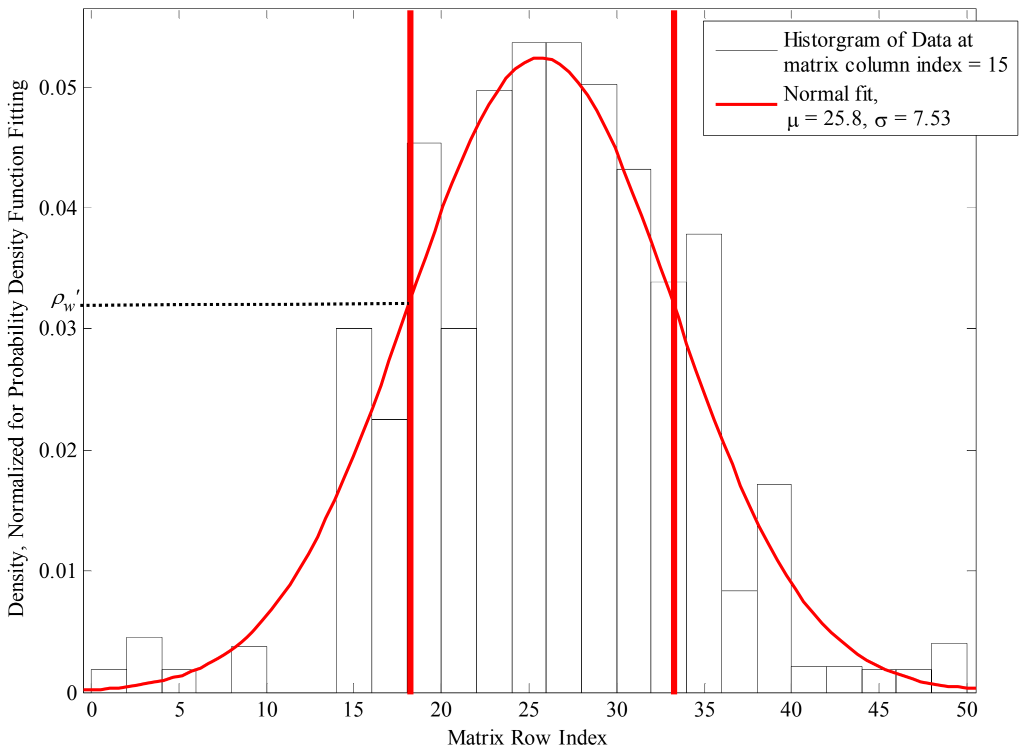


Figure 17. A normal curve can be fitted to the distribution of densities across the width of a highway.

Since the normal probability distribution comes from normalizing the distribution to have a unitary area under the curve by using the coefficient *K*, the final value of ρ_w then results from translating between ρ_w' and the true traffic density value using the scaling factor *K* from Equation (24). This enables the determination of ρ_w through

$$\rho_w = \frac{\rho_w'}{K}. \quad (28)$$

As an example, for a water space where $\rho_w = 0.643$, a demonstration of the incremental determination of width is displayed in Figure 18.

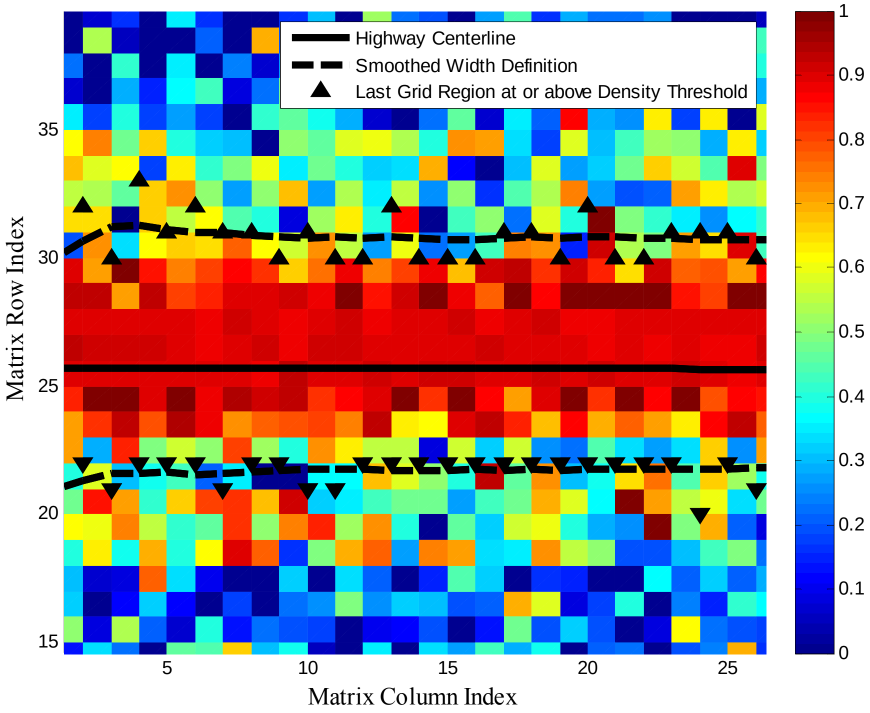


Figure 18. The width of a highway is determined by locating the last region of the grid moving out from the highway centerline that lies at or above $\rho_{w} =$ 0.643.

The extreme locations of the traffic density threshold in either direction are indicated with the triangles seen in Figure 18. Those limits are then smoothed into a defined width limit for the track using a moving average method as in

$$(w_{sx}, w_{sy}) = \left(\frac{\sum_{i=x}^{x+n} w_i}{n}, \frac{\sum_{j=y}^{y+n} w_j}{n} \right) \tag{29}$$

where n is the number of elements in the sum, also called the smoothing factor. Variables (w_{sx}, w_{sy}) and (w_i, w_j) represent the row and column indices within the grid of the smoothed and pre-smoothed widths, respectively. This result is plotted with dashed black lines in Figure 18.

5. Mapping

One of the final steps outlined in the flow diagram in Figure 11 involves adding the highway to a map of the ROI. The Hough transformation is performed on a space indexed according to the reference grid. For the end results to be universally meaningful, they must be translated back into latitude and longitude. This is performed through linear interpolation. For grid column references, interpolation to longitude is performed by calculating

$$l = Long_1 + (Long_2 - Long_1)\frac{x - x_1}{x_2 - x_1} \tag{30}$$

where l is the longitude being sought, $Long_1$ and $Long_2$ are the two known values of longitude correlating to x_2 and x_1, the grid column references to either side of the known value of x. For clarity, these values are all labeled in Figure 19.

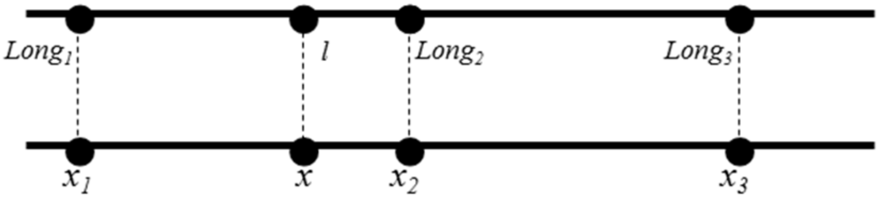


Figure 19. Interpolation is performed to transform grid reference values to values of latitude and longitude.

6. Trend Removal and Iterative Analysis

To enable the detection of less obvious linear trends in the data, the dominant trend is removed after it is discovered. Otherwise, secondary, tertiary, and further trends go undetected because they are overshadowed. The flow for this portion of the process is illustrated in Figure 20.

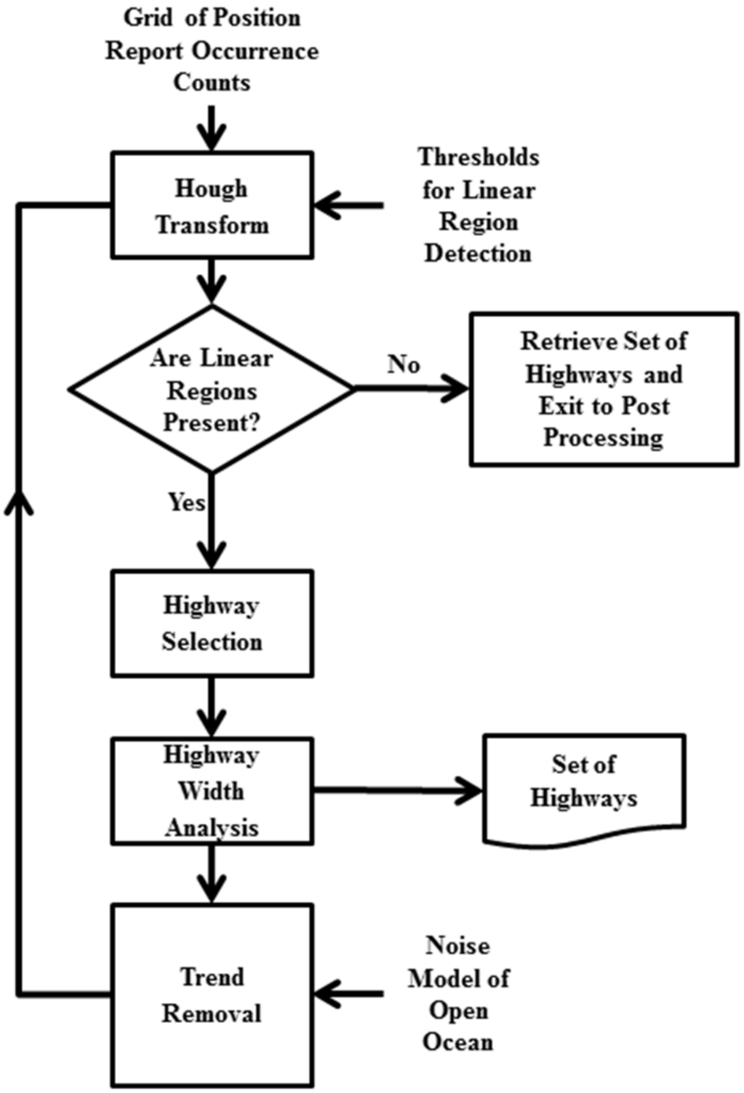

Grid of Position Report Occurrence Counts
Hough Transform
Thresholds for Linear Region Detection
Are Linear Regions Present?
No
Retrieve Set of Highways and Exit to Post Processing
Yes
Highway Selection
Highway Width Analysis
Set of Highways
Trend Removal
Noise Model of Open Ocean

Figure 20. The iterative use of the Hough transformation enables the detection of less prominent highways.

The grid space has been renormalized to the maximum traffic density of the secondary trend without removing the dominant trend from the water space shown in Figure 21. This renormalization only increasingly emphasizes the dominant trend. In the Hough space, the values of (d_D, θ_D) representing the primary trend occur so frequently that they overshadow any secondary trends. For proper consideration of subsequent trends, each trend must be effectively erased before the iterative use of the Hough transformation.

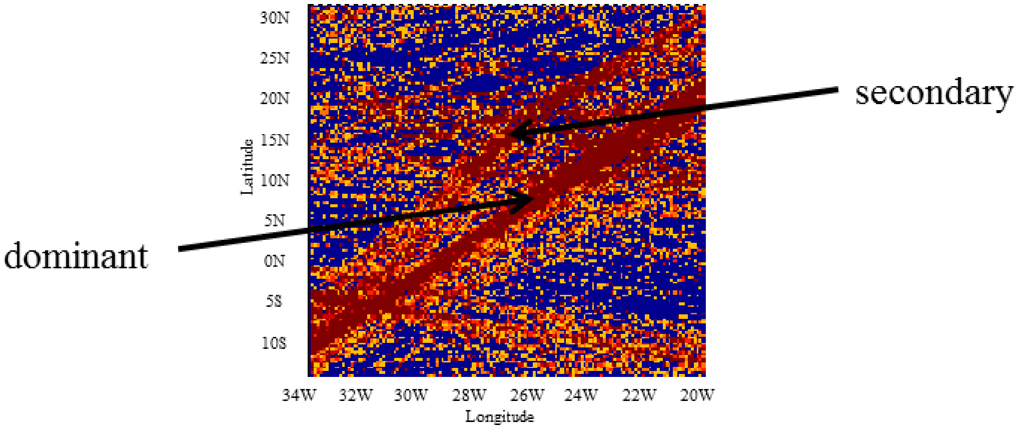


Figure 21. An example ocean space exemplifies why removal of the dominant trend is necessary before subsequent trends can be identified.

Simply removing the dominant trend and replacing those grid squares with zeros can unintentionally create new linear regions on the edges of the removed regions. Instead, the grid values that fall within a region bounded by the width definitions of the dominant trend are replaced by a pseudorandom value taken from the collection of all traffic density values occurring in the ROI. Once this replacement has occurred, the grid is renormalized to the new traffic density of interest, and the Hough transformation is performed again to identify the next-most-dominant trend in the region.

The four most dominant trends are iteratively identified for the Southern Atlantic Ocean in Figure 22. This case study is explored in more detail in Chapter IV, but this figure is included here to illustrate how the grid changes as trends are removed with each iteration of the highway discovery process. Replacing the dominant trend in (a) before renormalizing the grid enables the detection highway (b) that is travelled by fewer ships and, thus, has a lower traffic density than other highways in the water space. The same process continued enables the discovery of the highways in (c) and (d).

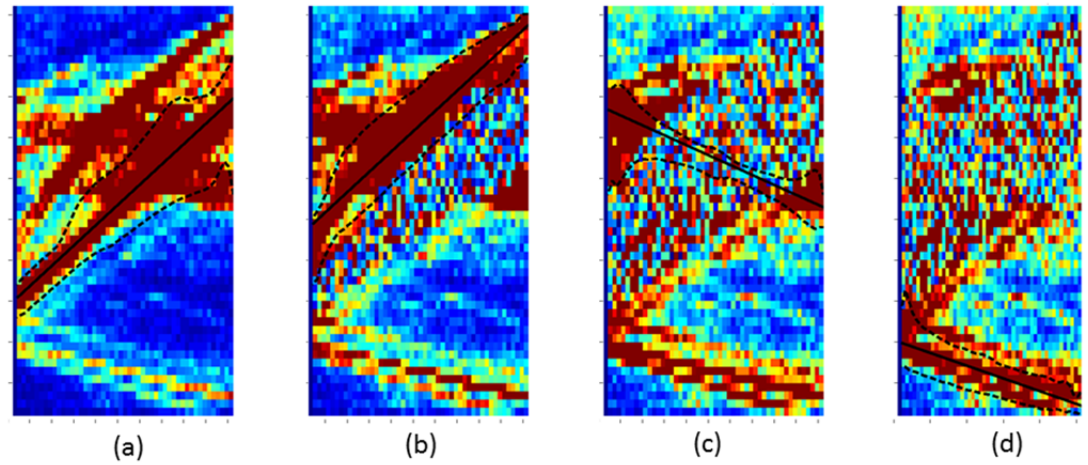


Figure 22. The four most dominant trends are detected in the Southern Atlantic Ocean from the reference grid of ocean densities.

This process can be continued until the water space is exhausted of all significant trends.

7. Post Processing Considerations and Outputs

The two products of the nine step method include two very similar items. First, an atlas of historic ocean highways is generated. Periodic reevaluations and updates must occur in step nine of the model because traffic patterns on the oceans are not static. This update process produces the final product: an atlas of expected ocean highways for the region of interest.

In both the historic and expected cases, the atlas includes a collection of historic highway definitions. Each highway definition is given to the user as three sets of coordinates in longitude and latitude pairs. The first set contains the reference line for the highway's centerline, while the second and third sets contain the lines defining the width of the highway to either side of the centerline. Because of its use of globally understood latitude and longitude, the format enables the highways to be imported into a variety of tools including mapping software.

B. ANOMALY DETECTION

The definitions of the highway are used for anomaly detection by considering whether a received position report lies inside or outside the limits of the highway. The widths of the highway are used to define a polygon starting along one side of the highway, crossing the highway perpendicularly at its end point, tracing down the other side of the highway in the opposite direction, and, finally, crossing the highway perpendicularly again to reconnect to the first side at the opposite end point.

Once this polygon has been defined, anomaly detection becomes a classic point-in-polygon problem. The method used to solve this problem in this thesis comes from [22] as described in Chapter II. To identify if a particular vessel of interest is anomalous, the available position reports associated with that vessel are each tested against the most applicable atlas of highways.

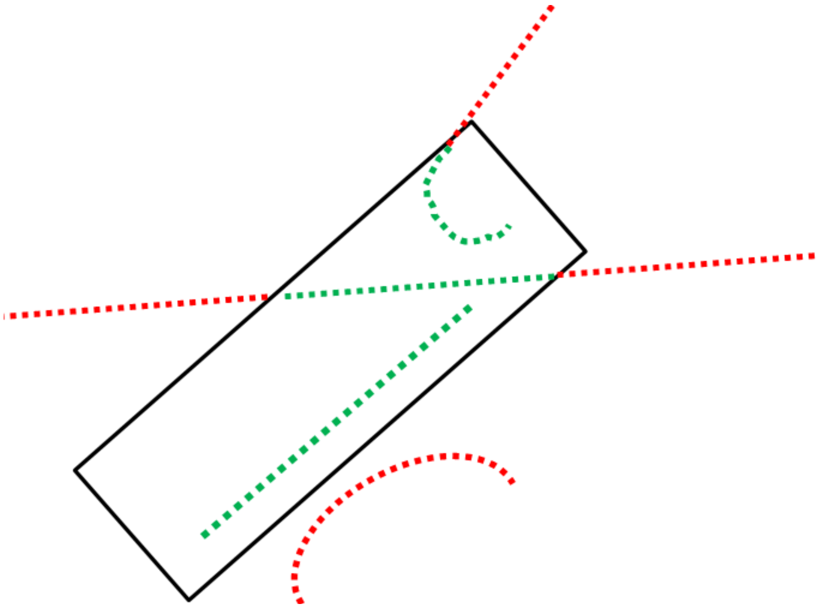


Figure 23. Anomaly detection within a geographic region, outlined in black, uses the point-in-polygon approach to highlight behaviors that are within the region in green and outside of the region in red.

A tolerance is set to determine how consistently a vessel must be located on a highway for it to be considered to be following the given route. This tolerance requires consideration of highway length and width coupled with the time interval over which the position reports are available. In some cases, this tolerance might be very high, requiring 100% of the position reports received to be on a highway for the vessel to prevent the vessel from being flagged as anomalous. One example of such a case is on lengthy transatlantic crossings, where vessels are frequently on the same great circle course for days at a time. In small coastal regions, however, some lower tolerance might be applied to consider ships entering or leaving a highway near the small harbors along coastal routes.

In the first section of this chapter, atlas generation from preprocessing through grid generation, highway detection, width determination, iteration, and post processing was covered. How that atlas is employed in anomaly detection was explained in the second section.

IV. RESULTS

The method described in Chapter III is applied to archived AIS data in this chapter. The implementation of the method, including programming techniques used, is discussed in the first section. A series of case studies meant to examine how atlas generation and anomaly detection occur in different scenarios is contained in the second.

To provide a proof of concept for the algorithm described in Chapter III, the PYTHON programming language was used for AIS data preprocessing functions and MATLAB was used to develop demonstrations of atlas generation and anomaly detection.

A. METHOD IMPLEMENTATION

Although any vessel position data source can be used to test the highway detection techniques, the large archives of AIS data available commercially make it a good candidate for initial investigations. A set of AIS position reports collected worldwide during 2012 was used in this thesis to study the technical considerations that arise in implementing the techniques developed in Chapter III.

1. Data Input and Preprocessing

AIS position reports are received in AIVDM sentence format as described in Chapter II. For this thesis, only the sixth field, the data payload field, needs to be decoded as all position related information is stored in this field as shown in Figure 1. Since this field is encoded in ASCII, a script written in the PYTHON programming language is used to translate the ASCII to binary and also to translate the binary values to a comma separated value (*csv*) text file that can be imported into MATLAB for the follow-on atlas generation methods. The code used to complete this translation is included in Appendix A. Each row of the *csv* contains ten fields of data. The column headings for these data fields are as follows:

- AIS Message Type
- Maritime Mobile Service Identity (MMSI)*

- Navigational Status
- Rate-of-Turn, degrees per minute
- Speed-Over-Ground, nautical miles per hour
- Latitude, Minutes*
- Longitude, Minutes*
- Course-Over-Ground, degrees
- True Heading, degrees
- Time Stamp, seconds Unix Time Code*

Once the *csv* has been created, it is imported into MATLAB for follow-on filtering and eventual atlas creation. Only those columns denoted with an asterisk (*) in the list above are used in atlas generation. The MMSI is preserved to enable filtering by the user, as the MMSI is a unique identifier for the ship that can be cross-referenced with other data sources to identify the ship and any related characteristics of the ship that are matters of public record including its country of registry, length, class, and in some cases even destination, cargo or crew manifests. As discussed in Chapter II, receiver stations for AIS log time of receipt. A time stamp field populated by the original satellite AIS receiver is also preserved to enable the user to select time periods of interest. Latitude and longitude are the crucial pieces of information as they are fed into the atlas generation algorithm to enable automatic highway detection as described in Chapter III.

2. Grid Generation

To create the traffic density grid, first a vector is created that starts at the western most boundary of the region and steps toward the eastern boundary in increments equal to the grid resolution. This process is outlined generically in Chapter III. To simplify the matrix operations performed in follow-on steps, a similar vector is created that steps incrementally from the southern boundary of the region of interest to the Northern boundary, but the steps are incremented such that an equal number of elements exist in both the longitude and latitude vectors. These two vectors are used to define the limits of each region of a grid.

Once the regions have been defined, a count of how many position reports occur in each region of the grid is taken. To do so, the position reports are sliced into bins of data falling between the incremental steps on the longitude vector. Next, the counts are populated into a traffic density matrix aligned with the longitude vector bins along the columns and the latitude vector bins along the rows. These counts are performed using MATLAB's *hist* function. Once the matrix is fully populated, it is normalized by the maximum occurring density count according to Equation (13) in Chapter III, resulting in a final traffic density matrix with values falling between zero and 1. This traffic density grid, when viewed using MATLAB's *surf* function, appears as displayed in Figure 24.

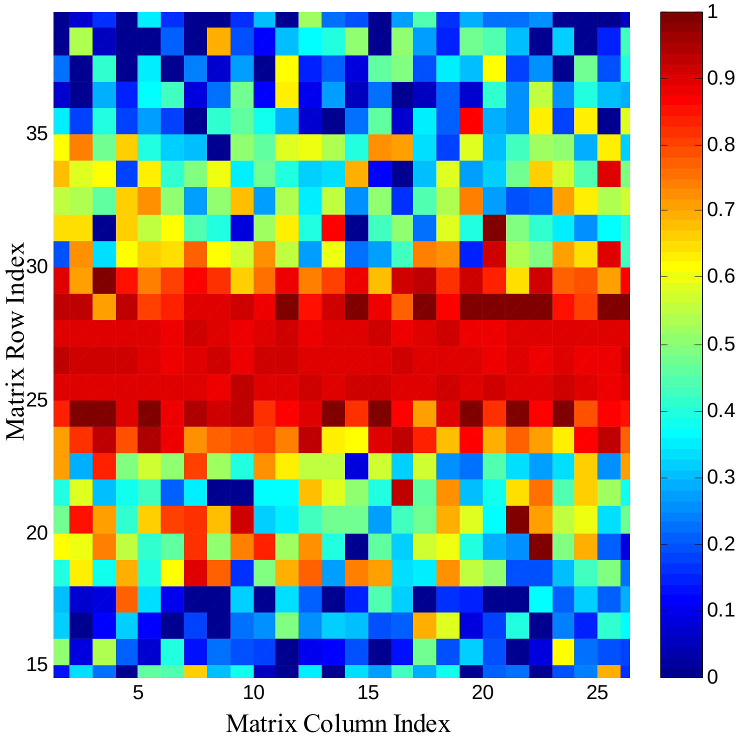


Figure 24. The traffic density grid contains values of ship position report densities between zero and one that can be displayed using MATALB's *surf* function. The color bar indicates that normalized traffic density values correspond to each color in the grid.

To the human eye, the highways are often apparent by observing this surface plot of the traffic density grid. The automatic detection of these highways is enabled by using the traffic densities in the way that the intensity and color would be used in image processing.

3. Highway Detection

As described in Chapter III, the Hough transformation is only performed on regions of the grid above ρ_{TH}. To do so, the traffic density grid is adjusted so that the maximum traffic density value of one indicates a traffic density at or above ρ_{TH}. The index of each of these regions is put through the Hough transformation described in Chapter II. The MATLAB script used to perform this transformation and to identify the value in the Hough space that occurs with the maximum frequency, which indicates the most dominant linear trend within the region, is included in Appendix B.

MATLAB's built in *polyfit* function is used to perform polynomial fitting to the indices within the traffic density grid of those regions that make up the dominant trend. This function is fully detailed in [23]. Essentially, *polyfit* finds the polynomial by first building a Vandermonde matrix,

$$V = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_{n-1} & 1 \\ x_n & 1 \end{bmatrix} \qquad (31)$$

where n is the number of grid regions contributing to the dominant trend and x represents the column indices of each of those regions.

Next, an orthogonal-triangular decomposition is performed resulting in two matrices, Q and R, such that Q is an n by two unitary matrix and R is a two by two upper triangular matrix such that

$$V = QR\,. \qquad (32)$$

The vector *p* of the coefficients of the best-fitting polynomial in the least squares sense is found by solving

$$p = \frac{R}{\mathrm{inv(Q)\,y}} \tag{33}$$

where *y* is the vector of row indices of each of the regions. The variable *p* contains two elements, p_1 representing the slope of the highway and p_2 representing the projected intersection of the highway with a line along the left vertical plane of the grid.

Additionally, the minimum and maximum column grid references from the set of all grid regions contributing to the dominant trend x_{min} and x_{max} are used to identify the extremities of the traffic pattern. The polynomial and extremities are used to create a model of the highway identifying how the traffic pattern is mapped to the grid. The model is identified by a series of (h_x, h_y) pairs found by first identifying the horizontal values h_x from

$$h_x = [x_{\min} : 0.1 : x_{\max}] \tag{34}$$

and then identifying the vertical components h_y from

$$h_y = p_1 h_x + p_2 \,. \tag{35}$$

4. Highway Width Analysis

As described in Chapter III, once a centerline has been found, the width is determined by moving out from the highway to either side and identifying where the traffic density first falls below the threshold. Of note, this can be implemented with a threshold value lower than that used to detect the highway itself. This application permits the edges of highways to have a lower traffic density than the centerline. Within MATLAB, a *while* loop is used to step outwards in the traffic density grid from the location of the highway. As soon as the traffic density tolerance is broken, that point (w_x, w_y) is recorded as the location of the boundary of the width of the highway. The MATLAB *smooth* function with a span of 50 is used to create the final width result (w_{sx}, w_{sy}) from

$$(w_{sx}, w_{sy}) = \left(\frac{\sum_{i=x}^{x+50} w_i}{50}, \frac{\sum_{j=y}^{y+50} w_j}{50} \right). \tag{36}$$

5. Trend Removal

To enable the detection of less dominant highways within the traffic density grid, highways that have already been detected are removed before the grid is normalized to the new traffic density of interest. The traffic density values of the regions of the traffic density grid within the boundaries of the highway width definitions are replaced according to

$$\rho' = G\{u\} \tag{37}$$

where ρ' is the replacement traffic density value, *G* is the set of all traffic density values occurring in the ocean region of interest, and *u* is a pseudo-randomly selected index taken from a uniform distribution between one and the number of densities in *G*. The value of *u* is determined by the built in *rand* function in MATLAB, which generates pseudorandom values between [0, 1] that are then scaled to the desired range of indices in *G*.

During the development of work in this thesis, we made an effort to build a statistical model of open-ocean traffic densities that could be used as a noise model in replacing the highway region, but the variability in different regions of the ocean and the availability of real data for the region of interest at this stage in atlas generation made use of such a model less desirable than randomly selecting a traffic density value from the collection of all values present in the traffic density grid. This method of implementation prevents the ocean noise model from needing to be tailored to the region of interest.

Once a highway has been removed, the process is iterated to subsequently detect less and less dominant regions.

6. Mapping

The centerline and width definitions are unique to the traffic density grid, but translating them back to a global latitude and longitude reference enables them to be used with a variety of display software or other follow on applications. Linear interpolation between grid reference values and the globally understood values of latitude and longitude is completed using MATLAB's *interp2* to complete this translation in accordance with the methodology outlined in Chapter III.

For this thesis, the figures displaying highway data overlaid above a map are generated using the standard MATLAB plotting tools coupled with *plot_google_map* from [24] and available via the MATLAB file exchange. This tool uses the Static Google Maps API to pull an image from Google maps and display it behind the active figure in MATLAB, aligning the map with the overlaid highway models by converting the map to WGS-84 datum coordinates. AIS and, thus, the atlases generated as part of this thesis also use WGS-84 datum for position reporting via GPS as described in Chapter II.

7. Anomaly Detection

Once an atlas has been defined, basic anomaly detection is performed by the MATLAB function *inpolygon* [25]. Based on the point-in-polygon problem described in Chapter II, *inpolygon* returns a logical 0 or 1 indicating that a point is either outside or inside of a polygon, respectively. The position reports available for a vessel of interest are each tested against the polygon defined by the highway width definitions. If a tolerable number of position reports are within the highway, then the ship is labeled normal. If not, then the ship is flagged as anomalous, indicating that it does not fit the expected traffic pattern of its geographic area.

B. CASE STUDIES

The thesis is not intended to draw conclusions regarding normal behavior of ships, but selected case studies were developed to exemplify the functionality of the highway determination and anomaly detection methods described in Chapter III.

The algorithm developed in this research is fairly generic and lends itself well to a variety of case studies. To demonstrate the potential uses of an atlas in a larger anomaly detection scheme, three models of employment are explored in this thesis. First, the complete algorithm with iteration is demonstrated using data from the southern Atlantic Ocean. This case exemplifies the capability of the algorithm when employed iteratively to find a collection of highways that are frequently traveled within a water space. Because the first case study focuses on a large open area of the ocean where transits in straight lines are common, the second case study explores the functionality of atlas development in coastal regions. Because of land and shoal water, ships are prevented from travelling in straight point-to-point transit fashion. Piecewise approximations of the coastal routes for two regions are presented where the linear Hough transformation results for adjoining regions are combined to provide a non-linear highway. Third, the use of pattern extraction to better understand how routes change over time is explored. As two of the many examples of factors affecting shipping routes, the impacts of seasonal variations and extreme weather systems are presented by comparing the highways detected in a given geographical region over different time intervals.

1. Demonstration of the Complete Method

The region of interest for this case study is the southern Atlantic Ocean. To ensure that only open-ocean traffic was observed, the region was bounded well outside of coastal waters. The exact region lies between the lines 15°S and 30°N in latitude and the lines of 30°W and 20°W in longitude. This area is outlined in black in Table 4. The data set of AIS position reports was collected via satellite over seven months in 2012. A sample of the position reports received are plotted in Figure 25 in blue to provide a visual idea of how traffic is distributed in the area. The grid size is 0.2 minutes longitude. Other general characteristics of the water space are listed in Table 5.

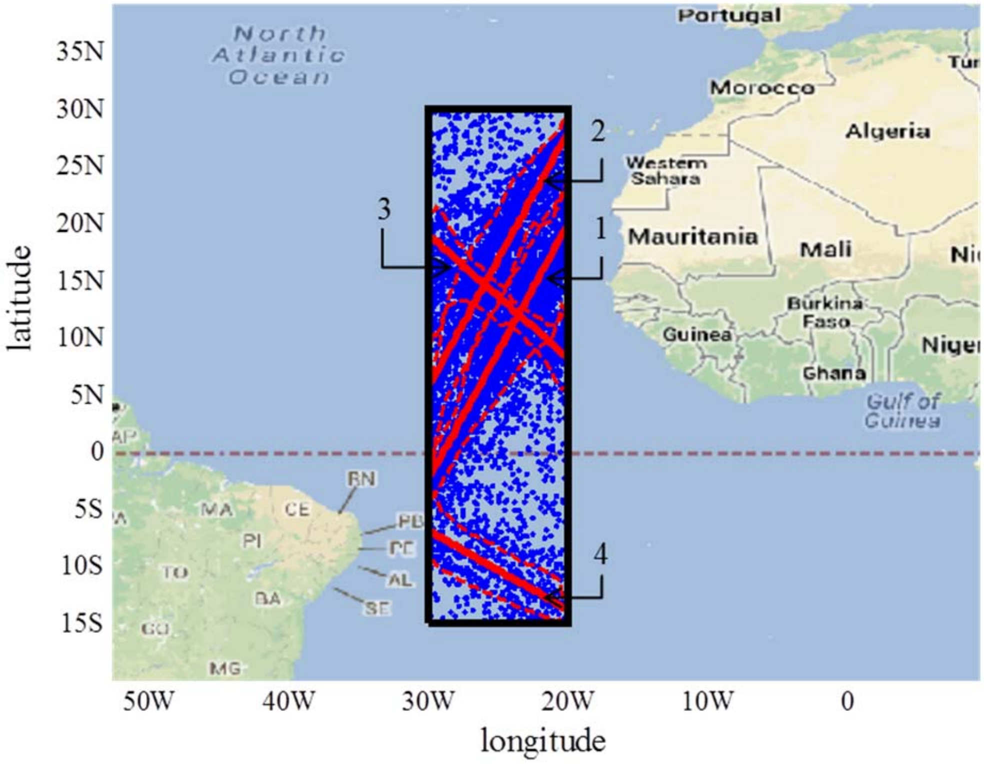


Table 4. The four most prominent highways in the southern Atlantic Ocean during 2012 are detected by the Hough transformation.

Table 5. The general specification of the water space used for the case study of the southern Atlantic Ocean during 2012 enable a comparison between "open-ocean" ship densities and "highway" ship densities listed in Table 7.

Specification	Value
Area of Region (nm^2)	1,600,300
Position Reports Received	2,143,706
Ship Density (reports per month per nm^2)	0.191

The Hough transformation is performed four times, with the recently identified trend removed after each iteration. This process is visually displayed in Figure 22 of Chapter III. Each iteration uses the new 75th percentile as the normalization factor, resulting in highway precedence as listed in Table 6.

Table 6. Traffic density thresholds used for the southern Atlantic Ocean during 2012 represent renormalization to the new 75th percentile over four iterations.

Rank	ρ_{th}
1	75.0
2	56.0
3	42.0
4	31.5

The highways are displayed in Figure 25 by a trio of red lines: a solid line along the centerline of maximum traffic density and a dashed red line to either side identifying the width limits. A general description of each of the highways in order of precedence is provided in Table 7. Additional specifications for each highway are listed in Table 8.

Table 7. The four most prominent highways in the southern Atlantic Ocean 2012 can be generally described by the route they take across the water space.

Rank	General Description	Eastern Most Point	Western Most Point
1	A direct route between the eastern most point of South America to Europe, passing east of Cape Verde	29° 48’ 00” W 02° 03’ 12” S	20° 00’ 00” W 19° 51’ 12” N
2	A direct route between the eastern most point of South America to Europe, passing west of Cape Verde	29° 48’ 00” W 06° 02’28” N	20° 00’ 00” W 27° 54’ 59” N
3	A direct route between North America to the southern tip of Africa, passing south of Cape Verde	29° 48’ 00” W 18° 48’ 35” N	20° 00’ 00” W 08° 21’ 33” N
4	A direct route between the eastern most point of South America to the southern tip of Africa	29° 48’ 00” W 07° 00’ 31” S	20° 00’ 00” W 13° 49’ 09” S

Table 8. Specifications for the four most prominent highways in the southern Atlantic Ocean during 2012 demonstrate how highway ship densities compare to the general water space described by Table 5.

Rank	Course	Mean Width (nm)	Highway Area (nm^2)	Position Reports "On Highway"	% of All Position Reports for Area	Ship Density on Highway (reports per month per nm^2)	Length (nm)
1	023.7°	485.6	2.837×10^5	1,046,434	48.81%	0.5269	1,436.5
2	023.1°	468.9	2.752×10^5	611,460	28.52%	0.3174	1,427.3
3	138.1°	279.9	1.601×10^5	382,764	17.86%	0.3416	855.2
4	125.3°	259.9	1.515×10^5	101,028	4.71%	0.0953	1,293.8
Total	---	---	8.705×10^5 (54% of total area)	2,141,686	99.99%	---	---

The propensity for maritime traffic to follow common routes, even on the unbounded open-ocean, can be discerned by considering the most dominant highway. The highway makes up just 17% of the total area of the region outlined in black in Figure 25, yet nearly half of all position reports received are located within the definition of this highway. In fact, 99.99% of all position reports received for the area fall on one of the four most dominant highways, even though they account for less than 50% of the total area of the region given the overlap between the first three highways. In seven months, only about 2,000 position reports were received outside of the areas described by the four highways. If this trend remains consistent, this Southern Atlantic region where more than 10,000 position reports occur daily would have a mean of fewer than ten that were off highway. This ability to highlight ships that are acting abnormally will better enable a human analyst to focus his or her efforts.

This case study also serves to demonstrate the importance of finding a method to discern traffic patterns that may not be readily apparent in the presence of the most dominant trends. The fourth trend has a ship density of only 0.0953 position reports per month per nm^2, which is lower than the ship density for the region as a whole, which lies at 0.191 position reports per month per nm^2. It might easily be ignored by a method that did not remove the more dominant trends after discovery.

2. Piecewise Highway Definition

The linear detection implementation may seem limiting to the usefulness of the algorithm, but it is not. The pattern extraction method used in the southern Atlantic can also be used to detect and approximately map non-linear highways by using a piecewise approach to a region. A key example of where this may be useful is in coastal areas in which geographic land features prevent direct straight-line routes from point to point. Essentially, the region of interest is split into sub-regions, the highways in each sub-region h_j are detected, and then those highways are concatenated together into h_T as in

$$h_T = [h_1, h_2, ..., h_j] \tag{38}$$

where j is the number of sub-regions used to divide up the coastal region. The complete highway h_T is then smoothed using a 5-point moving average operation into a single composite highway $h_T{}'$ across the entire region so that each element $h_{T_x}{}'$ of $h_T{}'$can be found from

$$h_{T_x}{}' = \frac{\sum_{i=x}^{x+5} h_{T_i}}{5} . \tag{39}$$

a. Eastern South American Case Study

For the Coastal South American Region between 15°S and 0° Latitude, the region requires five regions for a piecewise approximation of the coastal traffic pattern, outlined in black and labeled in Figure 25. The dominant trend in each of the other five regions is outlined with solid red lines. A sample of the position reports received for the area are plotted as a blue scatter plot to provide visual reference for traffic in the area.

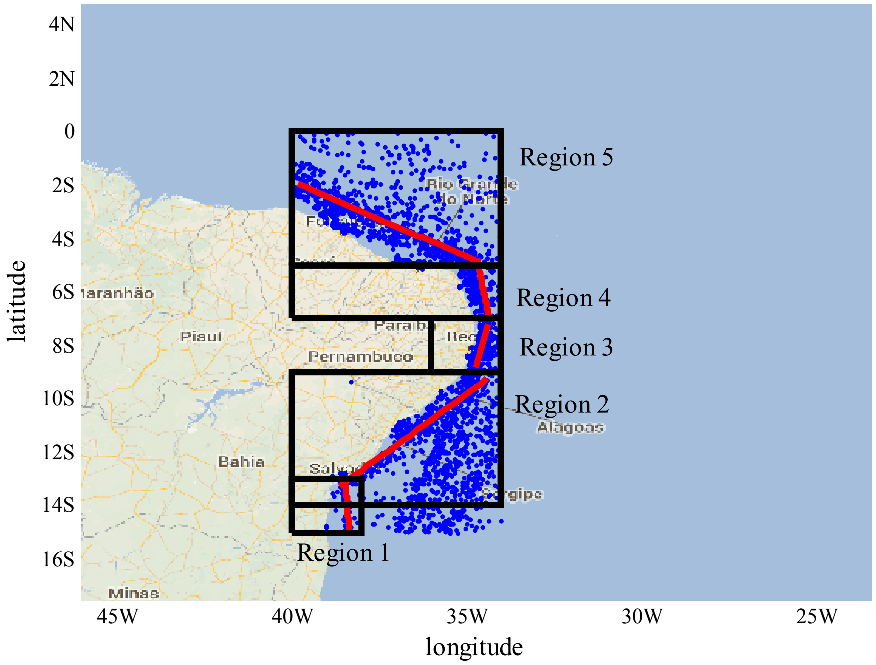


Figure 25. The dominant highways for each of five sub-regions of coastal South America depict the high traffic densities in coastal regions.

The Hough transformation was not performed iteratively for this case study. Instead, it was only performed once to compare the locations of the most dominant trend for each region. The first, second, and fourth regions require grid rotation for dominant trend discovery because of the near-infinite slope of the dominant trend in each of these areas. The second and fifth regions are analyzed in the standard north-south orientation. Good matchup occurs between the individual highways in each of these regions in that when moving from south to north, the next region's highway generally originates near the terminus of the previous region's highway.

The five local highways were smoothed into a single highway using a moving average operation with a step size of five data points. The same process was independently applied to the centerline of maximum traffic density and the left and right width definitions. The final result can be visualized in Figure 26. Additionally, the specifications of each of the five segments and of the final highway are listed in Table 9.

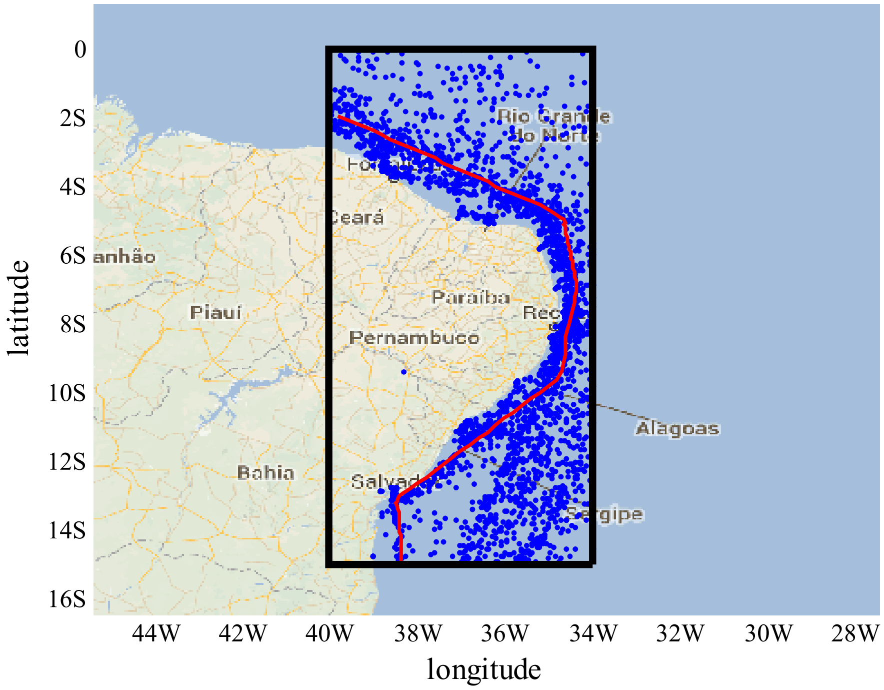


Figure 26. The results from the five sub-regions enable the development of a smoothed coastal highway representing the dominant traffic pattern along the coast of South America between 15°S and the equator.

Table 9. Specifications of the five piecewise segments and the final composite highway for coastal South America show how ship density can vary over the length of a coastal highway.

Segment	Ship Density in Segment (reports/nm^2)	Course	Mean Width (nm)	Length (nm)	Position Reports "On Highway"	Highway Area (nm^2)	% of All Position Reports in Segment Area	Ship Density on Highway (reports/nm^2)
1	0.04	356.9°	108.5	120.3	3,695	13,122	16.1%	0.04
2	0.56	045.8°	60.4	344.8	136,124	21,329	39.0%	0.91
3	0.15	011.9°	48.9	98.2	16,070	4,263	18.1%	0.54
4	0.13	351.1°	120.1	121.5	18,623	14,595	24.5%	0.18
5	0.62	119.7°	75.2	358.9	169,888	27,474	48.0%	0.88
Coastal Highway	0.31	n/a	97.5	788.9	326,727	87,732	34.5%	0.53

b. Western Africa Case Study

As a second demonstration of piecewise implementation of the highway detection techniques, a region on the coast of western Africa was selected. The region between 5°N and 30° N requires five regions for a piecewise approximation of the coastal traffic pattern as outlined in black in Figure 27. The dominant trend in each of the five regions is depicted in red. Regions 2 and 3 require rotating the grid 90° for discovery because of their near-infinite slopes in the north-south orientation. A sample of the position reports received for the area are plotted in blue to provide visual reference for traffic in the area.

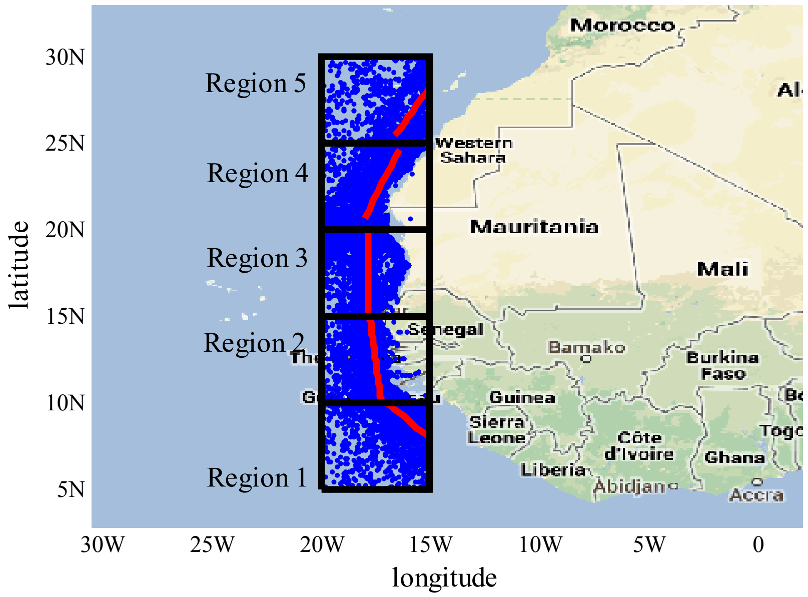


Figure 27. The dominant highways for each of five sub-regions of coastal western Africa are found by completing the pattern extraction method in each sub-region independently.

To provide a final highway definition, the five segments are smoothed into a single highway using the same method previously described and used with the South American region. The final result appears in Figure 28.

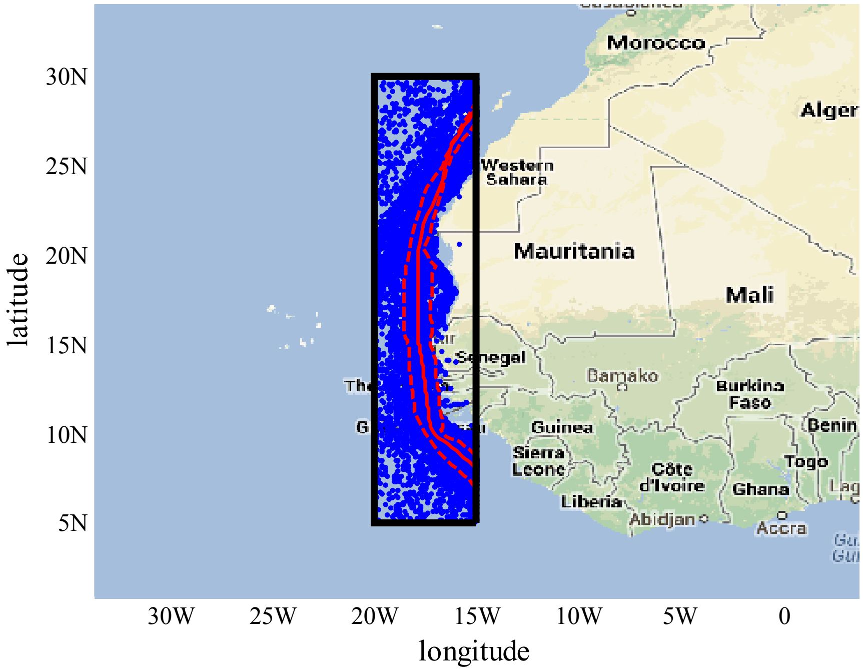


Figure 28. The smoothed coastal highway is mapped with a solid red line with its width defined by dashed red lines. This highway represents the piecewise combination over the five regions depicted in Figure 27.

One point that stands out in this case study is how thin the highway is in the area of the fifth region. Visually, based on the scatter plot of position report plotted in blue in Figure 29, it appears that the traffic is evenly spread and dense throughout the area. The specific details of each of the segments and the smoothed, complete highway are listed in Table 10. The algorithm identifies that while traffic may be spread across the fifth region, its most densely populated linear pattern identifies a highway that transits through the Canary Islands, passing between Gran Canaria at 27°58'N, 15°36'W and Fuerteventura at 28°20'N, 14°1'W.

This is one case where the human eye cannot find a distinct pattern in the region but the Hough transformation can, which demonstrates the effectiveness of this technique for detecting maritime highways.

Table 10. Five piecewise segments combine to form the final composite highway for coastal western Africa.

Segment	Ship Density (reports per month per nm^2)	Course	Mean Width (nm)	Length (nm)	Position Reports "On Highway"	Highway Area (nm^2)	% of All Position Reports in Segment Area	Ship Density on Highway (reports per month per nm^2)
1	0.94	135.0°	90.9	167.8	325,743	14,974	54.9%	3.11
2	1.91	354.0°	68.0	301.9	928,524	20,798	78.1%	6.38
3	1.55	000.5°	75.1	300.2	653,952	22,626	69.4%	4.13
4	1.85	020.6°	123.7	252.1	413,816	33,960	37.7%	1.74
5	1.46	027.1°	51.1	188.0	143,005	9,118	17.1%	2.24
Coastal Highway	1.54	n/a	68.9	1,298.5	2,517,997	81,863	54.0%	4.39

3. Highway Variability

The mapping of ocean highways requires periodic update for a variety of reasons. Some are universal to the maritime environment, but others are unique to the type of traffic of interest. Examples include extreme weather, seasonal variations, fluctuation in international markets and trade, and changes in law enforcement monitoring of an area.

Seasonal and weather related fluctuation are universal in impact and offer ground-truth case study opportunities for exploring how the algorithm developed in this research might be used to identify and understand variation in maritime traffic patterns. Conclusions drawn from weather might be adjusted for extension to provide insight into how inside information of market or law enforcement will impact traffic patterns, but these cases are not specifically developed as part of this research.

a. Hurricane Ernesto Case Study

This traffic analysis method provides a means for assessing the impact of extreme weather patterns on maritime traffic, which can assist in predicting the impact future weather events will have on traffic patterns in the area. Hurricane Ernesto moved on a generally westerly track through the Caribbean in early August of 2012 [26]. The National Hurricane Center graphic of the track of Hurricane Ernesto is shown in Figure 29.

The impact of the hurricane can be observed in the atlas generated daily over a three-day period, as displayed in Figure 30. The hurricane track for the day is overlaid in black above the Hough transformation detected highway in red. In this case, a particular highway of interest was selected from all highways found in the area. The highway is mapped using a normalized traffic density of interest of 0.80 and a traffic density tolerance of 0.50 for width determination. Essentially, the southern portion of the highway dissolves as the hurricane crosses its path and then reconstitutes within the next 24 hours as the hurricane moves further west.

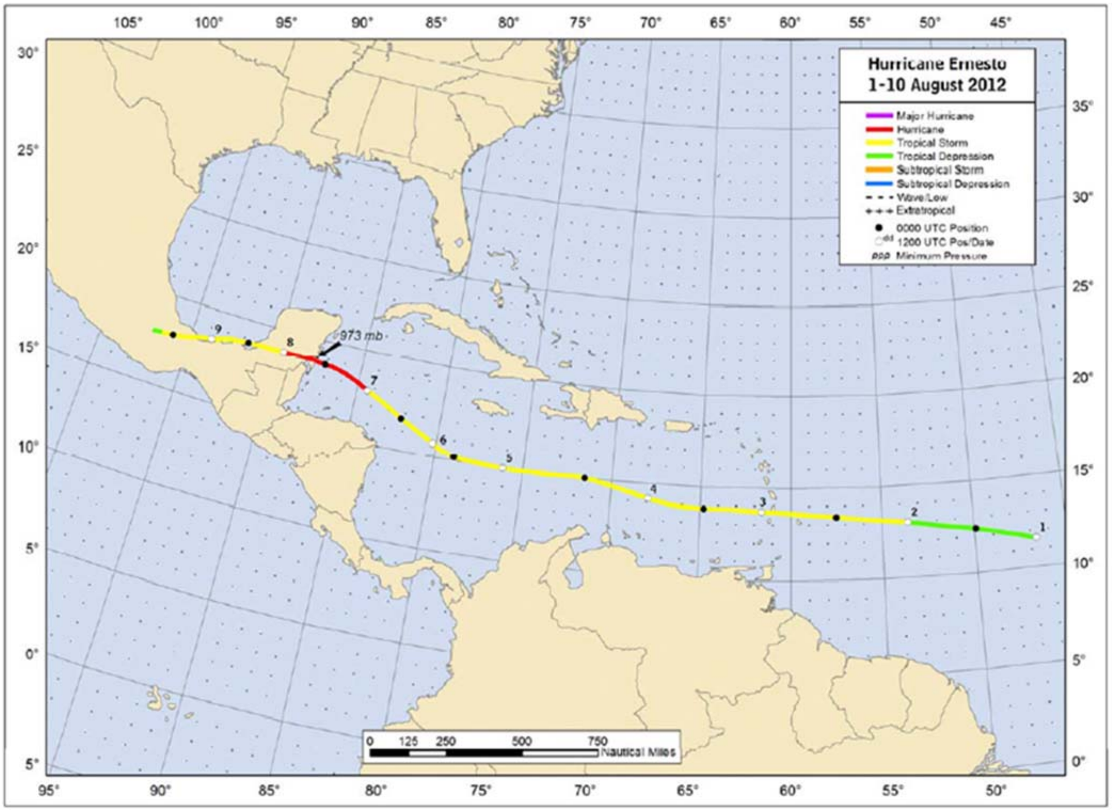


Figure 29. Hurricane Ernesto tracked East-to-West across the Caribbean in August 2012 (from [26]).

The specific measurements of this highway for five consecutive days are listed in Table 11. On August 5, 2012, the southern portion of the highway running along course 030° dissolves, leaving only the northernmost 130 nm of the complete 400 nm stretch intact. By August 7, 2012, the highway has essentially reformed to its dimensions previous to the storm, and the traffic density has resumed its pre-storm strength with just over one fifth of all position reports received for the area occurring along the 030° corridor.

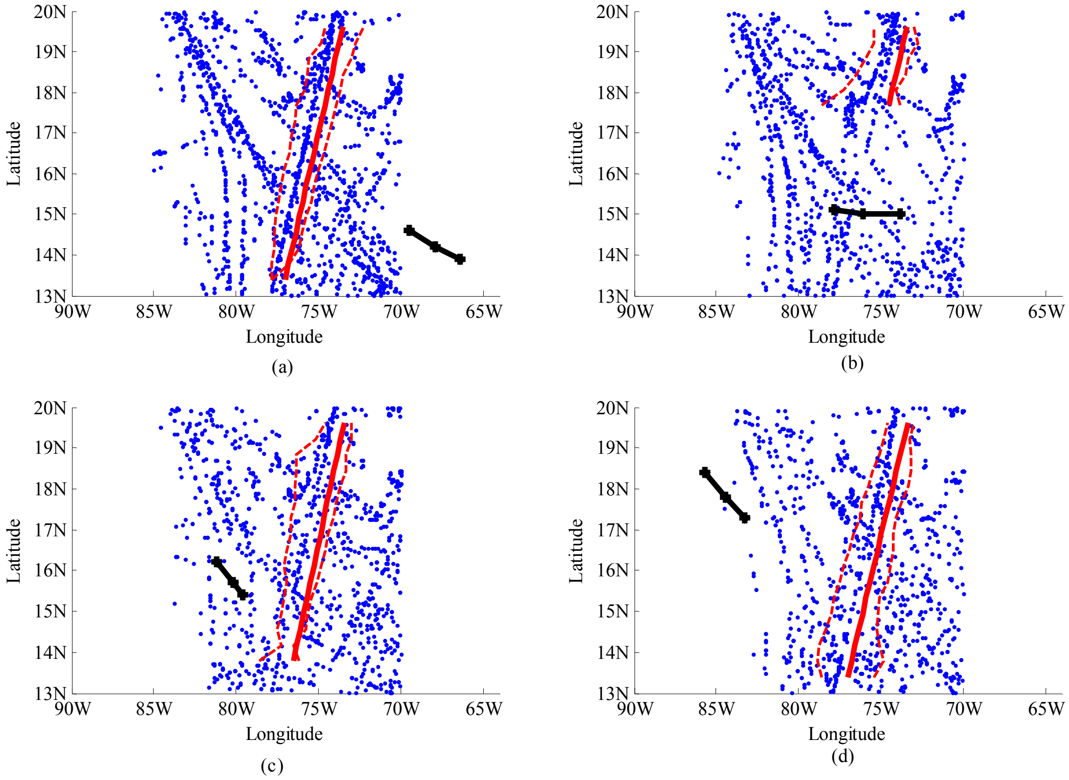


Figure 30. Hurricane Ernesto, plotted in black, caused a highway, plotted in red, to dissolve and then reform as Ernesto passed through the Caribbean over the days of (a) 04-05 August, (b) 05-06 August, (c) 06-07 August, and (d) 07-08 August 2012.

Although only a single case study, the persistence of this traffic pattern is evidence in support of the pursuit of an atlas-based anomaly detection method. The ocean may not have lane markings controlling ships' travel, but ships do remain in generally predictable and orderly routes. As soon as the storm had passed, travel along the highway resumed.

Table 11. The highway varied significantly before, during, and after Hurricane Ernesto in August 2012.

Time Period	Ship Density in Segment (reports/nm^2)	Course	Mean Width (nm)	Length (nm)	Position Reports "On Highway"	Approximate Highway Area (nm^2)	% of All Position Reports in Segment Area	Ship Density on Highway (reports/nm^2)
03 Aug 2012	0.0071	029.7	112.5	403.2	419	44,698	22.8%	0.0094
04 Aug 2012 BEFORE	0.0086	028.3	110.4	424.5	507	46,189	22.9%	0.0110
05 Aug 2012 DURING	0.0058	027.9	177.1	132.1	206	24,488	13.7%	0.0084
06 Aug 2012 AFTER	0.0059	026.8	123.8	392.2	235	49,423	15.3%	0.0048
07 Aug 2012	0.0054	029.4	152.26	428.8	342	66,982	24.3%	0.0051

The atlas method developed in this research provides an avenue for building insight into normal ship behavior, even in abnormal weather situations. One such behavior pattern may lie in the preemptive spike in traffic density just before the storm that can be observed along the highway. A plot of traffic density along the highway is presented in Figure 31.

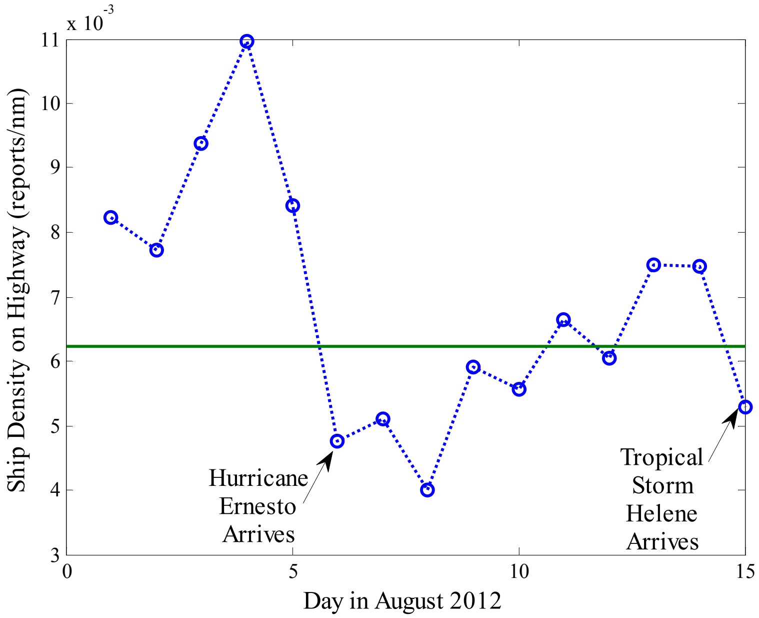


Figure 31. Hurricane Ernesto caused a drop in ship density on an ocean highway during August 2012.

Hurricane Ernesto struck this region on August 5, 2012 [26]. Traffic density on the highway mapped via the atlas method indicates that traffic spikes on August 4, remains high on the next day, and then sharply falls off when Hurricane Ernesto crosses it on August 6. As the hurricane moves across the area, traffic density drops below average and then begins to recover in the days after August 10 when the hurricane made land fall and moved inland [26], possibly permitting the coastal ports to resume their normal activities. The traffic density begins to recover, reaching another above average spike in traffic density, just before Tropical Storm Helene enters the same region on August 15, 2012 [27].

Although prudent mariners may already know that ships seek safe harbor just before bad weather, this algorithm enables analysts to qualify exactly what impact the storm will have on the most frequently traveled highways in an area.

b. Seasonal Variation in the Southern Atlantic Case Study

Hypothetically, it is unlikely that traffic remains perfectly static on the oceans. An atlas-based anomaly detection method is expected to require updates to the atlas at some frequency. Expanding the previously discussed southern Atlantic case study to observe the changes in the dominant and secondary highways over four quarters of 2012 provides some insight into the seasonal variability issues an ocean atlas would need to address. This case study will look at the changes over four quarters of 2012. The first quarter will be observed from January through March, with the second quarter falling from April through May, and so on for the third and fourth quarters.

The dominant trend over all four seasons of 2012 remained the direct route described in the case study discussed in Section B.1 of this chapter. The subtle variation in the dominant highway is displayed in Figure 32.

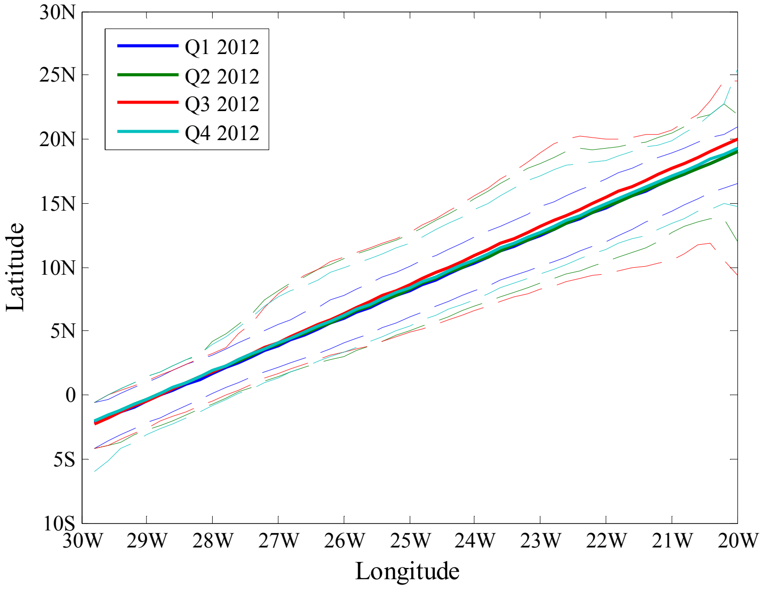


Figure 32. The most dominant highway in the southern Atlantic Ocean varies slightly as determined over four different quarters of 2012.

The trends are so close that the underlying map is discarded from the figure because it prevents the viewer from discerning the slight variations in slope and width that characterize each of the highways, as summarized in Table 11.

Table 12. A direct route from the eastern most point of South America to Europe, passing east of Cape Verde, is the most dominant highway in the southern Atlantic Ocean. Its characteristics change over 2012.

Quarter	Position Reports Received in Area of Interest	Ship Density (reports/ nm^2)	Course	Area (millions of nm^2)	Ship Density on Highway (reports /nm^2)	% of Total Reports	Length (nm)
Jan-Feb-Mar 2012	840,047	0.52	024.4	0.14	0.2953	34.2%	1,399
Apr-May-Jun 2012	1,000,242	0.62	024.6	0.26	0.2662	47.4%	1,390
Jul-Aug-Sep 2012	890,601	0.55	023.5	0.28	0.2188	47.5%	1,451
Oct-Nov-Dec 2012	1,134,816	0.70	024.3	0.23	0.2649	36.7%	1,405

This case study serves as evidence that some seasonal variation does occur; although, it may not be as pronounced in the most significant transoceanic highways as expected. Further work is needed to understand variation on a more local level where fishing seasons and harsh winter conditions may be more likely to impact coastal traffic. The research developed here can provide a method for how such variations might be studied by developing a mechanism for extracting the traffic patterns on any time scale and in a region of any size.

C. ANOMALY DETECTION TEST CASE

The use of the point-in-polygon technique described in Chapter II to enable anomaly detection along the western Africa coastal highway described in Section 2.b of this chapter is examined in this section. The highway is extracted from data collected during all of 2012. This case study assesses anomaly detection when comparing data from January 1, 2013, to the highway generated from the 2012 data. The track of each vessel is tested against the polygon defined by the width elements of the highway to identify whether or not each vessel is on the predefined coastal highway.

On January 1, 2013, 338 unique vessels are identified by MMSI in the archive of AIS data in the region bounded by the lines of longitude at 20°W and 13°W and the lines of latitude at 10°N and 26°N. This region and the highway previously developed are displayed in Figure 33.

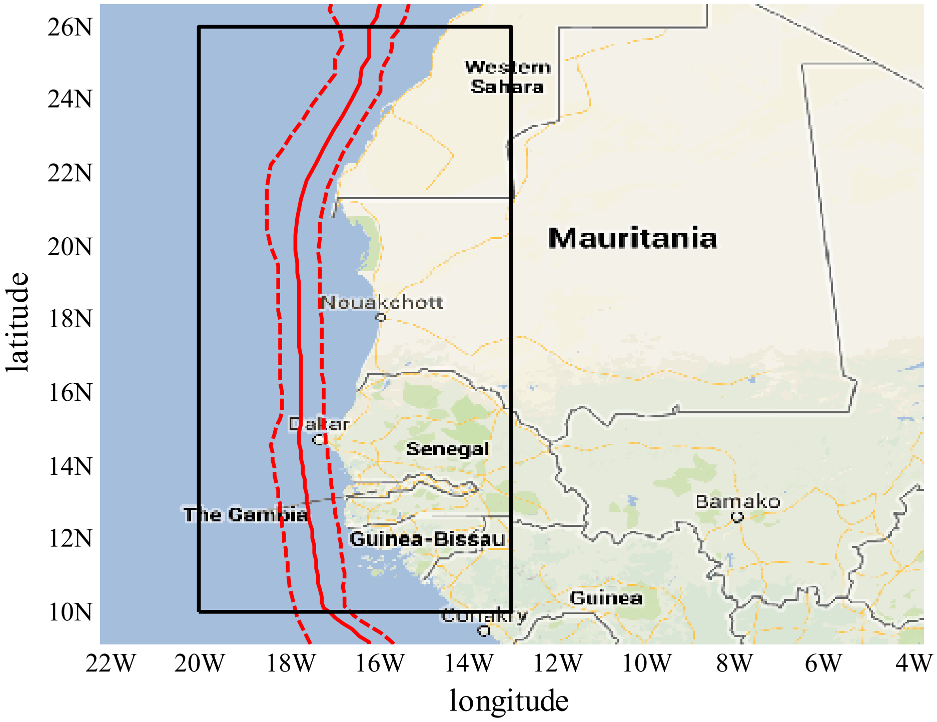


Figure 33. The highway detected in the western Africa case study in Chapter IV, Section B can be used in a coastal anomaly detection scheme in the region bounded by the black box.

If the available position reports for each of these 338 vessels are tested against the expected behavior for the area, 34.6% of the vessels, or 117 of them, are anomalous in that they are on the highway for fewer than 10% of their position reports. Alternately, 49.1% of the vessels, or 167 of them, are normal in that they are on the highway for more than 90% of their position reports. The remaining 54 vessels are on the highway some of the time but not all of the time. These results are summarized in Figure 34.

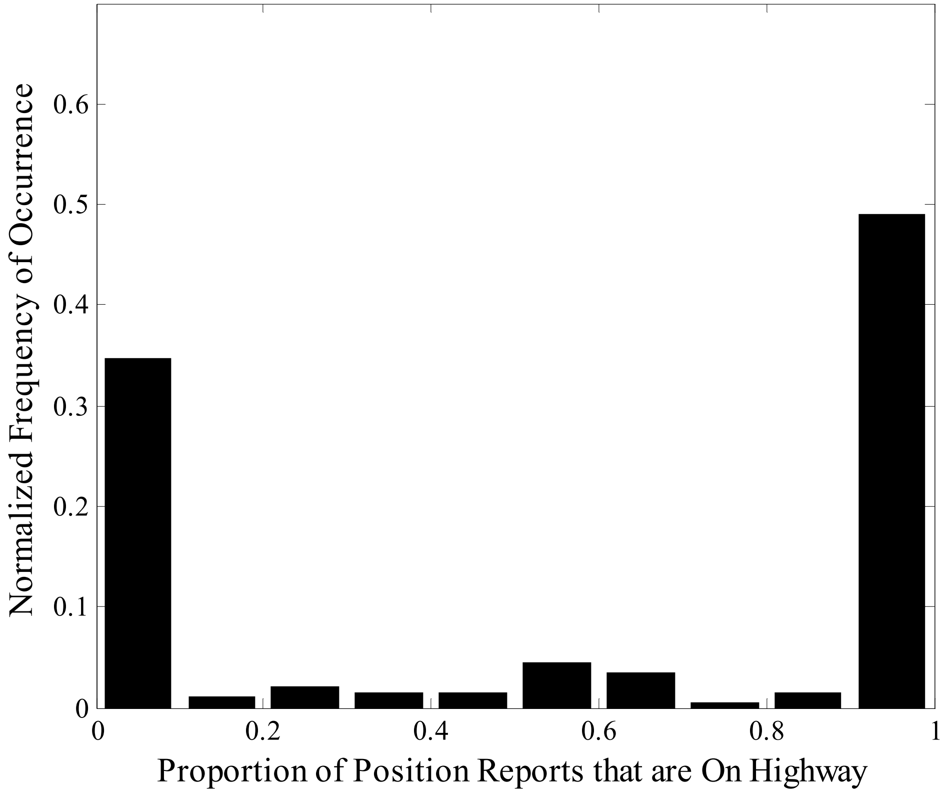


Figure 34. From the anomaly detection results, it can be seen that most ships are on the highway.

If only vessels that are transiting the area are considered, then the results are more telling. In the same ROI and same time window, 148 vessels are observed to move at least 150 nm during the day in question. Of them, 59.5% or 88 vessels are found to be on the highway more than 90% of the time, while the number of vessels exhibiting anomalous behavior drops to 18.2% or 27. In fact, only 24 vessels never use the highway at all. These data are displayed in Figure 35.

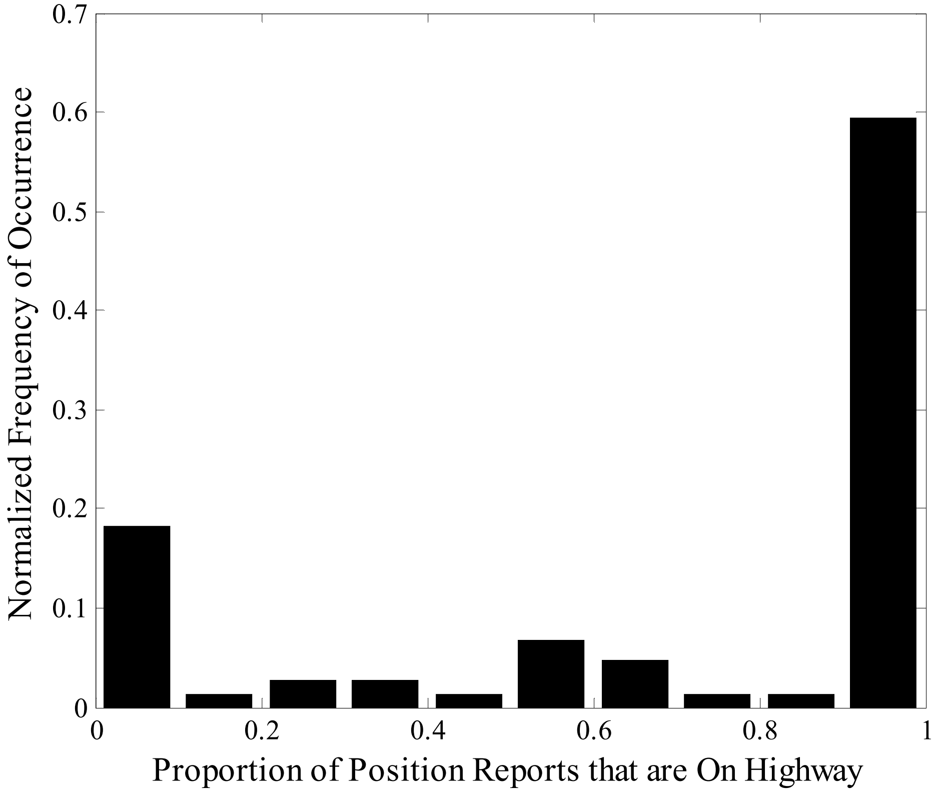


Figure 35. From the anomaly detection results for ships that travel at least 150 nm, it can be seen that a majority of ships are on the highway.

It is not the intent to draw conclusions regarding normal behavior of ships in this thesis, but rather selected case studies were developed to exemplify the functionality of the highway and anomaly detection methods described in Chapter III.

The methods outlined in Chapter III were expanded in this chapter with specific detail describing how the atlas generation and anomaly detection processes are completed using MATLAB. Test cases covering a variety of scenarios of atlas generation were outlined. Finally, a test case for anomaly detection was discussed.

V. CONCLUSIONS

A method for improving maritime domain awareness by developing an atlas of expected ocean traffic patterns was outlined and that atlas was used as the definition of normal within an anomaly detection scheme in this thesis.

An atlas generation method was developed beginning with a technique that preprocesses position reports into a traffic density grid. From the traffic density grid, a modified version of the Hough transformation was used to extract highways, which can be compiled and mapped into an atlas of expected ocean traffic patterns. An iterative method was developed to detect less prominent highways. Point-in-polygon problem solving was used to enable geographical region based anomaly detection of ship tracks as compared to the generated atlas.

A variety of specific case studies were developed in which the atlas generation and anomaly detection methods demonstrated usefulness to maritime domain awareness. As just one of several examples, the method enabled observations of how a hurricane transformed expected ocean traffic patterns as it traversed the Caribbean. The anomaly detection methods enabled analysis of transit versus local traffic in a coastal region.

A. SIGNIFICANT CONTRIBUTIONS

The most significant contribution in this thesis is the exploration of a technique from image processing to the problem of maritime domain anomaly detection. The use of the Hough transformation, an image processing technique, to detect and quantify maritime vessel behavior patterns has not been observed before in literature. This is a significant contribution to MDA because it enables the extraction of traffic patterns without requiring the preservation of specific vessel tracks, reducing the data processing and storage requirements necessary for behavior analysis.

A second contribution is in the development of techniques to produce an atlas of ocean highways. The extracted highways combine to form an atlas that provides a definition of established norms for information exploitation as required for national MDA efforts in accordance with [2].

The anomaly detection techniques explored in this thesis are also a significant contribution. An automatic initial determination of a vessel's geographic location as normal or abnormal is an important first step in assessing how much further analysis should be applied to determining a vessel's intentions.

B. RECOMMENDATIONS FOR FUTURE WORK

Several avenues for future work were opened in this thesis. The Hough transformation as used for this thesis was effective in detecting linear regions of interest within traffic patterns, but this could be expanded. First, the Hough transformation as used in this body of work must be run on two separate orientations of the ROI to overcome the problem of infinite slope. Future work could use polar coordinate techniques to enable detection of all linear patterns, regardless of slope, with only one ROI orientation required.

While the piecewise linear coastal highway development provided sound results, it required manual decision making for the division of the larger coastal region into smaller sub-regions. Techniques could be developed to automate this process.

The grid generation methods outlined in this thesis transform ship position data so that it becomes analogous to a digital image. In this work, the Hough transformation was explored, but other imagery processing techniques may also be able to contribute to MDA now that this technique for preprocessing the data into a traffic density grid has been observed to work.

As another avenue for research, statistical techniques could be developed to determine grid size without user input. In this thesis, grid generation was based on empirically-based human decisions. Automating the determination of this value is desirable. One possibility is more exhaustive analysis of maritime traffic patterns to understand optimum grid sizes based on distance from land.

AIS provided the large volume of reliable data necessary for algorithm development, but to truly improve MDA, new techniques should incorporate multiple data sources. The algorithm developed in this thesis was generally developed to be used with data from any source, but no other source was tested. Preprocessing techniques to absorb other data sources and verify their ease of incorporation to atlas generation and anomaly detection technique would increase the relevance of the techniques presented in this thesis. Because the techniques extract highways based only on position reports, the method outlined in this thesis could also provide a foundation for extracting a pattern from a fusion of data from different sensors so long as the disparate position measurement systems are aligned. Such data fusion has been identified as another important tenet of information exploitation [2].

The case studies demonstrated here have provided generic atlases of traffic patterns for all ships in general. Using additional filters during data preprocessing would enable the building of case studies with very specific atlases unique to the profile of the ships. Such atlases might prove more useful in anomaly detection because they would increase the contextual cuing used in identifying abnormal behavior.

In general, anomaly detection in the maritime environment has a multitude of unexplored possibilities. The method outlined in this thesis is a first means of detection based on statistical definitions of normal geographic position, but complete anomaly detection systems that used this method in series with other anomaly decision making tools would be more effective.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. AIVDM INTERPRETER

This appendix includes the PYTHON code used to translate AIS position reports in AIVDM data format to human-readable, comma separated list of the format [Message Type, MMSI, Navigation Status, Rate of turn, speed, longitude, latitude, course, true heading, timestamp]. Comments are in green text.

```
# Read in Line of 6-Bit ASCII encoded data

def getReport(asciiData):

    asciiData = list(asciiData)

    # Match ASCII Data to Binary Equivalent and Produce
    # Binary Data String
    upchar = ['0','1','2','3','4','5','6','7','8','9',':',
              ';','<','=','>','?','@','A','B','C','D','E',
              'F','G','H','I','J','K','L','M','N','O','P',
              'Q','R','S','T','U','V','W','`','a','b','c',
              'd','e','f','g','h','i','j','k','l','m','n',
              'o','p','q','r','s','t','u','v','w']

    downchar = ['@','A','B','C','D','E','F','G','H','I',
                'J','K','L','M','N','O','P','Q','R','S',
                'T','U','V','W','X','Y','Z','[','W',']',
                '^','-','.','/','0','1','2','3','4','5',
                '6','7','8','9',':',';','<','=','>','?']

    MMSIchar = ['0','1','2','3','4','5','6','7','8','9']


    valueString = [];
    for m in range(0, len(asciiData)):
        letter = asciiData[m]
        g = upchar.index(letter)
        g2 = bin(g)[2:]
        diff = '0'*(6 - len(g2))
        g3 = diff + g2
        valueString.append(g3)
    result = ''.join(valueString)
    print(result)
    # Determine message type
    try:
        messageType = int(result[0:6],2)            # Field 1
    except ValueError:
        return 0
    # Only position reports are handled by this function,
# return 1 if the entered report is not a position report
# of type (1, 2, or 3)
if (messageType >= 4) or (messageType < 1):
    return 1
else:
    repeat = result[6:8]
    MMSI = str(int(result[8:38],2) )            # Field 2
    NavStat = str(int(result[38:42],2) )        # Field 3


    ROT = str(int(result[42:50],2)   )         # Field 4



    speed = str(int(result[50:60],2)/10  )      # Field 5

    longitude1 = result[61:89]
    if longitude1[0] == '0':
        longitude = str(int(longitude1,2)/600000)
    else:
        idx = 0
        tempLong = ''
        while idx < len(longitude1):
            if longitude1[idx] == '0':
                tempLong += '1'
            else:
                tempLong += '0'
            idx += 1
        longitude=str(-1*int(tempLong,2)/600000) #Field 6

    latitude1 = result[89:116]
    if latitude1[0] == '0':
        latitude = str(int(latitude1,2)/600000)
    else:
        idx = 0
        tempLat = ''
        while idx < len(latitude1):
            if latitude1[idx] == '0':
                tempLat += '1'
            else:
                tempLat += '0'
            idx += 1
        latitude = str(-1*int(tempLat,2)/600000) #Field 7

    course = str(int(result[116:127],2)/10 )    # Field 8
    hdg    = str(int(result[127:137],2))        # Field 9
    timestamp = str(int(result[137:142],2))    # Field 10

    report =[str(messageType), MMSI, NavStat, ROT, speed,
              longitude, latitude,course, hdg, timestamp]
```

```

      return report
# END OF CODE
```

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. HOUGH TRANSFORMATION SCRIPT

This appendix includes the function written as a MATLAB script to perform the Hough transform on position report data, identify the dominant linear trend in the region, and return a highway definition describing a linear model of the dominant trend. Comments are in green text.

```
function [ highway, allPairsU, counts, pointsWithPair] =
HoughTransform_Maritime2( TestMatrix,...
    density_tol, slope_tol, dist_tol )
%This is a preliminary development for the Hough Transform applied to a
%density Matrix.

% TestMatrix is the input matrix of normalized density values for the
% region of interest
% density_tol is the user established density that should be used to
% identify a highway
% dist_tol is the tolerance a user would accept that would indicate two
% points are on the same line.
% tol_line is the tolerance for how many grid lengths away two points can
% be before they should be considered NOT part of the same line.

a = ((TestMatrix)');
d{size(a,1),size(a,2)} = [];
allPairs = [];
theta = [1:5:180];
for mm = 1:size(a,1)
    for nn = 1:size(a,2)
        if a(mm,nn) >= density_tol
            for q = 1:length(theta)
                Q1 = [mm,nn];
                Q2 = [mm + cosd(theta(q)), nn + sind(theta(q))];
                m = (Q2(2) - Q1(2))/(Q2(1) - Q1(1));
                b = -m*Q1(1) + Q1(2);
                P = [0,0];
                d{mm,nn}(q,2) = abs(det([Q2-Q1;P-Q1]))/norm(Q2-Q1);

                Dx = (sqrt(-Q2(2)^2 + 2*m*Q2(1)*Q2(2) +...
                    d{mm,nn}(q,2)^2*(m^2 + 1) - m^2*Q2(1)^2)...
                    - m*(Q2(2) - m*Q2(1)))/(m^2 + 1);

                d{mm,nn}(q,1) = real(acosd(Dx/d{mm,nn}(q,2)));
                allPairs = [allPairs;d{mm,nn}(q,1),d{mm,nn}(q,2)];
            end
        end
    end
end
%% Remove repeat entries in the table

allPairs(:,2) = round(allPairs(:,2)*1000)/1000;
allPairsU = unique(allPairs,'rows');

%% Identify the slope/tol pair that occurres most prevalently in the data
```

```
counts = zeros(size(allPairsU,1),1);
for i = 1:size(allPairs,1)
    for e = 1:size(allPairsU,1)
        diff = abs(allPairs(i,:) - allPairsU(e,:));

        if diff(1) < slope_tol && diff(2) < dist_tol
            counts(e) = counts(e) + 1;
        end

    end
end

[~, maxidx] = max(counts);

thePair = allPairsU(maxidx,:);
pointsWithPair = [];
for mm = 1:size(a,1)
    for nn = 1:size(a,2)
        list = d{mm,nn};
        for e = 1:size(list,1)
            list(e,:);
            thePair;
            [slope_tol, dist_tol];
            if sum(abs(list(e,:) - thePair) < [slope_tol, dist_tol]) == 2
                pointsWithPair = [pointsWithPair; mm,nn];  %#ok<*AGROW>
            end
        end
    end
end

P = polyfit(pointsWithPair(:,1),pointsWithPair(:,2),1);
startX = min(pointsWithPair(:,1));
stopX = max(pointsWithPair(:,1));

highways_x = startX:1:stopX;
highways_y = highways_x * P(1) + P(2);

highway = [highways_x; highways_y];
end
```

LIST OF REFERENCES

[1] G. W. Bush Administration. (Dec. 21, 2004). *Maritime Security Policy*, National Security Presidential Directive NSPD-41, [Online]. Available: https://www.fas.org/irp/offdocs/nspd/nspd41.pdf

[2] U.S. Department of Homeland Security (2005, Oct.). National plan to achieve maritime domain awareness. [Online]. Available: https://www.dhs.gov/national-plan-achieve-maritime-domain-awareness

[3] U.S. Department of Transportation Maritime Administration (2011, Feb.). U.S. water transportation statistical snapshot. [Online]. Available: http://www.marad.dot.gov/documents/US_Water_Transportation_Statistical_snapshot.pdf

[4] J. Rianto, "Road network detection from SPOT satellite image using Hough transform and optimal search," in *Asia-Pacific Conference on Circuits and Systems*, vol.2, pp. 177–180, 2002.

[5] X. Yang and G. Wen, "Road extraction from high-resolution remote sensing images using wavelet transform and Hough transform," in *5th International Congress on Image and Signal Processing,* Oct. 16–18, pp. 1095–1099, 2012.

[6] J. Guan; Z. Wang; and X. Yao, "A new approach for road centerlines extraction and width estimation," in *IEEE 10th International Conference on Signal Processing,* Oct. 24–28, pp.924–927, 2010.

[7] J. Cheng-Li; J. Ke-Feng; J. Yong-Mei; and K. Gang-Yao, "Road extraction from high-resolution SAR imagery using Hough transform," in *IEEE International Geoscience and Remote Sensing Symposium Proceedings.*, July 25–29, vol.1, pp. 336–339, 2005.

[8] D. Herumurti, K. Uchimura, G. Koutaki, and T. Uemura, "Urban road extraction based on hough transform and region growing," in *19th Korea-Japan Joint Workshop on Frontiers of Computer Vision,* Jan. 30 –Feb. 1, pp. 220-224, 2013.

[9] A. Holst, B. Bjurling, J. Ekman, A. Rudstrom, K. Wallenius, M. Bjorkman, F. Fooladvandi, R. Laxhammar, and J. Tronninger, "A Joint Statistical and Symbolic Anomaly Detection System: Increasing performance in maritime surveillance," in *15th International Conference on Information Fusion*, July 9–12, pp. 1919–1926, 2012.

[10] B. Ristic, B. La Scala, M. Morelande, and N. Gordon, “Statistical analysis of motion patterns in AIS Data: Anomaly detection and motion prediction,” in *11th International Conference on Information Fusion,* June 30 -July 3, pp. 40–46, 2008.

[11] R. Laxhammar, G. Falkman, and E. Sviestins, “Anomaly detection in sea traffic - A comparison of the Gaussian Mixture Model and the Kernel Density Estimator,” in *12th International Conference on Information Fusion*, July 6–9, pp. 756–763, 2009.

[12] G. de Vries and M. van Someren, “Machine Learning for vessel trajectories using compression, alignments and domain knowledge,” *Expert Systems with Applications*, vol. 39, pp. 13426–13439, 2012.

[13] U.S. Coast Guard Navigation Center. (n.d.) *Automatic Identification System overview* [Online]. Available: http://navcen.uscg.gov/?pageName=AISmain

[14] R. Challamel, T. Calmettes, and C. Gigot, “A European hybrid high performance Satellite-AIS system,” in *6th Advanced Satellite Multimedia Systems Conference and 12th Signal Processing for Space Communications Workshop,* pp. 246–252, 2012.

[15] exactEarth. (n.d.) *About exactAIS* [Online]. Available: http://www.exactearth.com/products/exactais/

[16] Orbcomm. (n.d.) *AIS* [Online]. Available: http://www.orbcomm.com/services-ais.htm

[17] E. Raymond. (2011, June). AIVDM/AIVDO protocol decoding. [Online]. Available: http://gpsd.berlios.de/AIVDM.html

[18] J. Russ, “Feature Specific Measurements” in *The Image Processing Handbook* (4th ed.). NewYork: CRC Press, ch. 9, pp. 491–496, 2002.

[19] R. Duda and P. Hart, “Use of the Hough transformation to detect lines and curves in pictures,” *Commun. of the Assoc. for Computing Machinery, Inc.*, vol. 15, no. 1, pp. 11–15, Jan. 1972.

[20] V. Chandola, A. Banerjee, V. Kumar, “Anomaly detection: A survey.” *ACM Comput. Surv*., vol. 41, no. 3, article 15, July 2009.

[21] W. Navidi, “Commonly Used Distributions” in *Statistics for Engineers and Scientists* (1st ed.). New York: McGraw-Hill, ch. 4, p. 231, 2006.

[22] K. Hormann and A. Agathos, “The point in polygon problem for arbitrary polygons,” *Computational Geometry*, vol. 20, no. 3, pp. 131–144, 2001.

[23] Mathworks. (n.d.) *polyfit* [Online]. Available: http://www.mathworks.com/help/matlab/ref/polyfit.html

[24] Z. Bar-Yehuda. (2010, May). plot_google_map [Online] Available: http://www.mathworks.com/matlabcentral/fileexchange/27627-plotgooglemap

[25] Mathworks. (n.d.) *inpolygon* in MATLAB, Release 2012b. Natick: The MathWorks, Inc., 2012.

[26] D. Brown. (2013, Feb.). Tropical Cyclone Report Hurricane Ernesto (AL052012). [Online]. Available: http://www.nhc.noaa.gov/data/tcr/AL052012_Ernesto.pdf

[27] L. Avila. (2012, Dec.). Tropical Cyclone Report Tropical Cyclone Helene (AL072012). [Online] Available: http://www.nhc.noaa.gov/data/tcr/AL072012_Helene.pdf

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California