CHARACTERIZING RECREATIONAL BOATING PATTERNS BASED ON GPS TRAJECTORY POINTS

by

YAN WU

Submitted
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

Major Subject: Industrial Engineering

at

Dalhousie University

Halifax, Nova Scotia   March 2007

Your file Votre référence
ISBN: 978-0-494-27653-2
Our file Notre référence
ISBN: 978-0-494-27653-2

DALHOUSIE UNIVERSITY

To comply with the Canadian Privacy Act the National Library of Canada has requested that the following pages be removed from this copy of the thesis:

Preliminary Pages
- Examiners Signature Page
- Dalhousie Library Copyright Agreement

Appendices
- Copyright Releases (if applicable)

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AR	Aspect Ratio
CI	Coverage Index
CCG	Canadian Coast Guard
DPA	Douglas -Peucker Algorithm
DFS	Furthest Distance From Shore
GPS	Global Positioning System
HRM	Halifax Regional Municipality
MARIN	Maritime Activity and Risk Investigation Network
MARIS	Maritime Activity and Risk Investigation System
MAUP	Modifiable Areal Unit Problem
MDA	Multivariate Discriminant Analysis
MDPA	MARIN Douglas-Peucker Algorithm
MS	Mean Speed
MTA	Mean Turning Angle
NB	New Brunswick
NAD	Normalized Absolute Deviation
NS	Nova Scotia
PFD	Personal Floatation Devices
POB	Persons on Board
PWC	Personal Watercraft
TD	Total Distance Travelled
USCG	United States Coast Guard

Acknowledgements

I would like to express my sincere appreciation and gratitude to my supervisor, Dr. Ronald Pelot, associate professor in the department of Industrial Engineering and Faculty of Management at Dalhousie University, whose knowledge, experience and guidance have been crucial throughout the course of this study. I could always benefit from the discussions I had with him on various technical issues when I faced a cross-road.

This thesis project is supported by the Canadian Coast Guard (CCG), the National Search and Rescue Secretariat (NSS), and the GEOIDE National Centre of Excellence (NCE). This research relies on excellent research assistance by members of the MARIN (Maritime Activity & Risk Investigation Network) research group in the Department of Industrial Engineering, Dalhousie University, in particular Christine Delbridge, Andrea Cameron, Lauren Plummer and R.C. Hilliard. Finally, we are grateful to the anonymous boaters who volunteered to employ our GPS units to monitor their vessels' movements.

I also wish to thank Dr. Carl-Louis Sandblom and Dr. John Kozey of Department of Industrial Engineering, for their help, advice, time and interest in reviewing this work. Special appreciation is extended to Dr. Irene Abi-Zeid from Université Laval, for agreeing to serve on the examining committee.

Last, but not least, I would like to thank my parents, my husband and rest of my family for their support and patience through the course of this study.

Abstract

Recreational boating is a very popular activity in Canada. Consequently, its associated risk is appreciable. Although the study of recreational boating is a very important element of maritime risk analysis, little spatial information is available on recreational boating movements. However, a better understanding of the patterns in Canada could be important for coastal safety and security, two key issues that motivated this project

For this study, Global Positioning System (GPS) data points were collected for a sample of recreational boating trajectories of four types of boats, namely canoes, kayaks, motorboats and sailboats, and in two environments, coastal and river. The GPS data were then examined to find spatial patterns by establishing and detecting each trajectory's important movement features. Based on these patterns, trajectories of different boat types were simulated to help evaluate recreational boating traffic in the context of a recognized maritime risk of incidents.

Aside from the critical steps of assiduous data cleaning for preparation of features extraction, the other indispensable process is dedensification of the data. Using GPS units to gather data results in a large number of points for a single trajectory, but not all of these points are significant for pattern analysis. Dedensifying is the process of removing such unnecessary points while retaining turns. A modified MARIN Douglas-Peucker algorithm was developed to accomplish this purpose. Furthermore, in order to overcome the limitations of setting a single pre-specified tolerance value for such an algorithm, this study advances an objective and context-specific method to select the best dedensified trajectory for any given boat trip.

After these preparations, evaluations were conducted showing that eight attributes adequately represent the patterns, and algorithms were developed to calculate them: total distance travelled, mean speed, maximum speed, maximum five percent speed, mean turning angle, coverage index, aspect ratio and furthest distance from shore. Classification of boat types models for both the individual study areas and the combined geographic areas were constructed based on univariate ANOVA tests, multivariate discriminant analysis and other statistics techniques. The classification rates of the linear models, which only retain independent variables with significant discriminating power, exceed 80% accuracy.

It was found that one can discriminate and classify between different boat types to varying extents based on these movement patterns' attributes. Moreover, it was also shown that there are some differences in vessel movements between the two areas, but most of the patterns are not dependent on location.

This study developed procedures for a novel application: spatial pattern analysis for recreational boating based strictly on GPS trajectory points. The results of this study provide insight into recreational boat movement characteristics and exposure levels to advance the research on risk analysis associated with this activity, improving accident prevention and search and rescue resource planning. Moreover, the results of this research can help detect boat types and abnormal movements based solely on tracking, which may prove useful for coastal security.

1. Introduction

Recreational boating is a popular pastime in Canada, including summer pursuits at the cottage, commonly occurring throughout fishing seasons, and increasingly involving newer activities such as sea kayaking and adventure tours. The variety of vessel types, activities, and geographic milieus associated with boating present innumerable sources of hazards for this inherently risky pastime, and yet it appears that there is some commonality in many of the causes and outcomes of boating incidents. Therefore, it is a useful endeavour to attempt to identify and quantify some aspects of recreational boating to assist with the development of a risk model, which can then provide insight for better prevention and mitigation measures in this area. To that end, this research proposes to characterize certain types of boating activity movements as an important aspect of an overall risk analysis. Furthermore, distinguishing between boat classes based on movement patterns can provide useful information for coastal security.

In this chapter, the context for the work is presented, including noting the importance of recreational boating in this country, and the key drivers of related risks. Pertinent recreational boating research is reviewed to categorize three different types of primary data. The advantages and disadvantages of each data source are evaluated. Within the context of the notable magnitude of recreational boating risks and the deficiencies in previous research, the objectives and the scope of this study are presented and defined. The final section provides an outline of the entire thesis.

1.1 Background

1.1.1 Background on Recreational Boating Activities and Associated Risks

The Maritime Activity and Risk Investigation Network[1] (MARIN) designed a pan-Canadian survey to address recreational boat ownership and usage patterns, which was administrated by Leger Marketing, a professional survey company. The questionnaire is attached in appendix 1. The survey was conducted over the telephone from February 17th to the 22nd, 2004. Using Statistics Canada census data, the results were weighted according to geographic location, age, and gender to ensure a sample representative of the entire adult population in Canada. The maximum margin of error for this study is $\pm 6.2\%$, 19 times out of 20. Of the 1,501 Canadians contacted by Leger, 818 adults agreed to participate. Of these, 251 participants owned a boat and were qualified to answer all of the questions prepared by MARIN. Thus, thirty-one percent of respondents (251/818) own some type of boat. Extrapolating that to the entire Canadian households of 8,701,700 (Statistics Canada, 2005), roughly 2,700,000 (31%*8,701,700) households own at least one recreational boat. Comparable estimates from other sources suggest that as many as 10 million Canadians participate in recreational boating (vessels under six meters long) each year in Canadian waters, such as the study by Groff and Ghadiali (2003) using data provided by Canadian Red Cross. It can be concluded that recreational boating is a very popular leisure activity among Canadians. This is reasonable given the plentiful waterways around and within Canada, as well as the relatively high spending

[1] Maritime Activity and Risk Investigation Network, a research group led by Dr. Ronald Pelot of Dalhousie University, and supported by the Canadian Coast Guard (CCG), National Search and Rescue Secretariat (NSS), GEOmatics for Informed DEcisions (GEOIDE NCE), and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

power and interest in outdoor activities. In addition, increased exposure to many media sources entices people to partake in diverse recreational boating activities.

Because of the magnitude of recreational boating activities, it is not surprising that the associated incidents comprise a serious problem. Both drowning and near-drowning incidents arise from recreational boating, as well as other types of injuries such as being hit by the propeller after falling overboard. All the incidents accounted for in this regard are unintentional and near-drowning is defined as "when a drowning victim is rapidly resuscitated and survives to reach hospital" (Canadian Red Cross Society, 2001).

A primary source of information about recreational boating incidents is a database called the Canadian Surveillance System for Water Related Fatalities, which was established by Canadian Red Cross, the Royal Lifesaving Society of Canada and the National Association of Coroners in 1991. The drowning accidents are investigated by coroners or medical examiners and assigned a code representing the external cause of injury. The available data spans 1991 to 1999, yielding an average of 140 recreational boating drownings every year in Canada (Canadian Red Cross Society, 2003). However, it is important to note that because boating drownings are frequently misclassified, according to the National Drowning Report (Canadian Red Cross Society, 2001), it is suspected that the data underestimates the total number of actual drowning deaths by up to 43%.

Aside from drownings, available information indicates an average of 104 recreational boating-related near-drownings resulting in hospitalization every year in Canada (Canadian Red Cross Society, 2001), however this figure is widely assumed to be a gross underestimation. In fact, some victims of near-drowning incidents who are conscious and breathing on their own after being rescued or resuscitated, may not be immediately transported to a hospital for the needed medical attention and thus they do not appear in the database. In order to acquire better data with which to quantify the number of people who are resuscitated after nearly drowning, the Water Incident Research Alliance (WIRA) was founded, a coalition of groups to systematically record non-fatal water incidents and fatalities in Canada based on voluntary reporting. WIRA's mission also involves actively compiling information and preparing summary reports on this topic (WIRA, 2005). With all these efforts, more precise information about recreational boating incidents may be obtained.

As with any accident studies, the personal costs associated with incidents are incalculable. However, in order to generate an estimate of how severe this recreational risk is, several methods have been used to quantify the costs, including economic and social burden.

The human capital approach estimates the indirect economic costs of boating drownings based on the average annual wage, activity participation rate, average employment rate, real wage growth rate, discount rate, and age/sex specific mortality rates. Applying this method Groff and Ghadiali (2003) calculated the total economic burden of drowning deaths in Canada to be $30,156,533 for 1999.

Note that this estimate of economic burden included only the drowning cases coded with cause of injury codes E830 and E832, where E stands for External Causes. 830 represents accidents to watercraft causing submersion, which includes submersion and drowning due to boat overturning, boat submerging, falling or jumping from a burning ship, falling or jumping from a crushed watercraft, vessel sinking, and other accident to a watercraft; the digits 832 corresponds to other accidental submersion or drowning in a water transport accident, which includes submersion or drowning as the result of an accident other than an accident to the watercraft, such as falls from a gangplank or ship, being thrown overboard by the motion of the vessel, or washed overboard. Cases in which submersion or drowning occurs when a swimmer or diver voluntarily jumps from a boat are excluded (Public Health Data Standards Consortium, 2006). Nevertheless, due to incomplete E-code recording and improper coding for many cases, this estimation conducted by Groff and Ghadiali (2003) was underrated. Checking the Visual Surveillance Report (Canadian Red Cross Society, 2001), there were 122 recreational boating drowning deaths in Canada in 1999 with only 86 drowning cases coded with either E830 or E832. So the human capital approach did not capture all of the indirect economic burden. Moreover, it does not reflect other substantial direct costs, such as those related to medical treatment on site, damage to the recreational boats, and search and rescue (SAR) efforts. The cost of SAR activities including operating expenditures and capital expenditures exceeded $70 million in 1998/1999 spent by the Canadian Coast Guard (CCG, 2003), not to mention the provincial expenditures on SAR for inland rivers and lakes.

In order to value both indirect costs and direct costs, Groff and Ghadiali (2003) estimated that the total average annual cost of all recreational boating drownings in Canada for 1991 and 1992 was about $80 million per year. This calculation is based on the work of Rice et al. (1989) for the U.S., assuming that the average costs per drowning (including both the indirect costs of loss of productivity as well as direct costs for medical treatment, funeral services, etc.) would be similar in the U.S. and Canada.

In addition, Groff and Ghadiali (2003) calculated the potential years of life lost to represent the social burden. Using an average age of death for 1999 of 75 for both males and females (Statistics Canada, 1999), subtracting the actual age at death from the expected age of death, then multiplying this figure by the number of deaths, a grand total of 2,767 potential years of life were lost to Canadians in 1999 due to boating drownings (Groff and Ghadiali, 2003). For those who have accidents but survive, even a small amount of aspirated water can cause damage to the lungs which could ultimately lead to serious respiratory difficulties or even death if not medically treated (Golden and Tipton, 2002), and some near-drowning victims sustain brain damage due to a lack of oxygen (The Canadian Red Cross Society, 2001). Moreover, these tragic events also have a devastating and long lasting impact on the family members and friends.

All these analyses and estimations are incomplete and conservative, but nevertheless they highlight the recreational boating risks.

1.1.2 Recreational Boating Research Overview

Given the large number of recreational boaters, as well as the number of related incidents and associated losses, stakeholders (e.g. CCG) have realized that the problem is a serious one, thus research efforts have been increased to save lives and property.

There are three main types of data sources for studying recreational boating and its associated risks. One primary mechanism is surveys, where the subjects can be the general public, the boating community, experts, researchers, or sellers and manufacturers. Many surveys have been carried out by various organizations since it is a common and easily applied methodology. Lentnek et al. (1969) conducted an interview survey at 15 Ohio lakes as early as 1969 to acquire data for a spatial analysis of activity-specific boating. Donnelly et al. (1986) analyzed a typology of recreational boating-related activities based on a statewide random survey of Maryland boaters in 1986. McAvoy et al. (1990) described the entire population of Minnesota registered owners' boating patterns and behavior based on a survey of 2,490 boat owners. Peterson (1991) addressed the shape and extent of market areas for recreational facilities by investigating recreational boaters in three Great Lakes states. Tarrant and English (1996) constructed a crowding-based model of social carrying capacity for whitewater boating use through an on-site survey on the Nantahala River in North Carolina in the summer of 1994. However, we should use these results carefully because they derive from different years, locations and research purposes. Moreover, applying different methodologies such as the use of phone surveys or spot surveys, or choosing from different populations like the general public or boat owners, might yield inconsistent results. For example the probability that boaters wear lifejackets may equal 0.47, 0.33, or 0.21 according to various researchers and organizations conducting differing surveys (Groff and Ghadiali, 2003). MARIN has also conducted its own specific surveys. MARIN has designed a recreational boat owners' survey administered at various locations across Canada during boats shows (Appendix 2), as well as the aforementioned telephone survey for our own specific research aims (Pelot et al., 2004a; Pelot et al., 2004b). Another important aspect of surveys is the return rate. Sidman (2004) figured out that the return rate was approximate 20 percent based on his previous surveys. Thus a comprehensive survey typically takes a long time, which may impact a study's timeliness. MARIN's pan-Canadian phone survey on recreational boating habits yielded an effective return rate of 17% (251 qualified participants from 1501 targeted). Moreover, a question arises whether there is any bias between these two groups: those refusing to complete the survey versus those willing to answer. This is easy to verify by testing the demography homogeneity of these two groups.

$H_0 : p_{1j} = p_{2j} \quad j = 1,2,...,J$

$H_a : H_0$ *is not true*

The null hypothesis H_0 of homogeneity states that the proportion of individuals in demographic category j is the same between the group not willing to answer (p_{1j}) and the group that answered (p_{2j}), and that this is true for every category. The categories in our case are age, occupation, number of children, education, income and language. Table 1-1 is an example of a chi-square test of independence between the categories of age bracket and survey status (i.e. Answer and No Answer). Each cell in the table contains the number for each age bracket and survey status combination, along with the row percentage, column percentage, and total percentage of observations falling in that cell. The results of the test for independence indicate that the percentage of observations in each cell is significantly different from the survey status of the total row percentage and total column percentage. Thus, there is an interaction between the age and survey status, which influences the number of individuals who answered. It can be concluded confidently that the two groups are different (i.e. bias exists) as a function of age because all of the *p*-values (i.e. for each demographic variable) are so small, indicating significant differences in the proportions (Table 1-2).

Table 1-1 Cross-tabulation of Age and Survey Status

age	status Answered	No Answer	RowTotal
18-24	112* 0.6 0.14 0.075	75 0.4 0.11 0.05	187 0.12
25-34	166 0.61 0.2 0.11	107 0.39 0.16 0.071	273 0.18
35-44	180 0.56 0.22 0.12	140 0.44 0.2 0.093	320 0.21
45-54	161 0.54 0.2 0.11	139 0.46 0.2 0.093	300 0.2
55-64	102 0.51 0.13 0.068	99 0.49 0.14 0.066	201 0.13
65-74	58 0.46 0.071 0.039	69 0.54 0.1 0.046	127 0.085
75 above	33 0.4 0.041 0.022	49 0.6 0.071 0.033	82 0.055
Unknown	2 0.18 0.0025 0.0013	9 0.82 0.013 0.006	11 0.0073
ColTotal	814 0.54	687 0.46	1501

Chi^2 = 24.69039 d.f.= 7 (p=0.0008608823)

```
*+----------+
 |N         |
 |N/RowTotal|
 |N/ColTotal|
 |N/Total   |
 +----------+
```

Table 1-2 *P*-values of Chi-square test

Demographic Variable	Gender	Age	Occupation	Number of children	Education	Income
p-value	0.062	0.001	0.007	0.023	0.006	0.000

Generally speaking, a survey is an easy method to obtain data for a study, hence it is often used. But it is a time-consuming process and often results in a low return rate and sometimes suffers from bias. Furthermore, different survey methodologies and targeted populations may yield different results.

The second type of data source derives from information recorded by various organizations. For example, boat sales, licensing data, and some organization statistics can provide general indications of ownership and usage. Siderelis et al. (1995) used a random household-based sample from a registered database of boat owners living in the region surrounding the Catawba River Basin in North Carolina to build a boating choice model for the valuation of lake access. Unfortunately, not all recreational boats have licenses in Canada, as only power boats with horsepower of 10 hp or more are required to have one (Great Lake Commission, 2000). Furthermore, these databases are inaccurate and poorly maintained. Most organization statistics databases apply primarily to large types of boats such as commercial fishing boats and ships. Alternatively, the System of Information for Search and Rescue (SISAR) incident database collected by the Canadian Coast Guard is useful to provide accident information on all types of vessels (Uremovich and Pelot, 2002), and more than half of the recorded incidents are associated with recreational boating. From the SISAR database, the date, time, location, type of accident and other information are available for each incident, which provides a reliable record of recreational incidents over several years.

The third data source is boating activity observation by air spotting, ground spotting, water-based spotting, or remote sensing such as satellite detection or global positioning system (GPS) tracking. Pelot et al. (2000) developed a methodology on how to carry out aerial observation missions on boating in the Bay of Fundy, and subsequently conducted some pilot studies. Other organizations like the Canadian Red Cross used ground-spotting to figure out the percentage of boaters wearing personal flotation devices, and establish the frequency or number of boating activities in particular locations. Observations using tracking devices like GPS can yield accurate information about boat positions and movements. The global positioning system is a satellite-based navigation system relying on satellites in precise orbits approximately 11,000 miles above earth. Anyone can use GPS anywhere anytime in all weather conditions for free. They are increasingly employed for navigation and recording spatial information in many current applications. For example, Magee and Denys (2005) used GPS to obtain accurate velocity assessment of rowing skiffs to determine how the velocity of a skiff responds to the crew's combined power and rowing technique, which provides useful feedback for coaches and athletes. But few studies have been conducted applying GPS data to recreational boating patterns for risk analysis.

1.2 Objectives of This Study

1.2.1 General and Specific Objectives

It can be concluded from the introduction that recreational boating risks have engendered a range of intriguing research problems and instigated some actions due to its potentially severe impacts, and yet many unanswered aspects remain such as the spatial patterns. Generally speaking, this study applies the third data collection method as the primary data source to characterize recreational boating patterns based on GPS trajectory points. These patterns are evaluated for two complementary purposes in this study. The first is to compare movement patterns across different recreational boat types, and the second is to contrast patterns of different recreational boat types at different geographic locations. More details are given as follows:

Objective 1: Characterizing patterns of different recreational boat types

It is possible that many factors influence the trajectory characteristics of recreational boats including the type of vessel, shore geography, weather conditions, operator traits, and outing purpose (such as fishing, touring/cruising, commuting, partying, hunting, whitewater sports, swimming/diving/snorkeling, water-skiing/tubing/, organized activity, etc.). However, the aim of this study is whether the type of recreational boat solely based on path data (position and velocity) can be determined without reference to the geography, weather and operator factors. That is, those other factors are hypothesized as extraneous, or that the fundamental boating movements are essentially independent of those factors. Objective 1 can help detect abnormal vessel movements through remote sensing, which can be an important contribution for coastal safety and security. Moreover, the results of this study are valuable for realistically simulating recreational boats' trajectory in a GIS model, which is crucial for recreational boating traffic analysis. This work addresses the issue of whether different movement patterns' trajectory can be generated for different recreational boat types.

For example, given 5 paths in Figure 1-1, we would like to be able to determine with some level of confidence which trajectories are canoes (C1, C2, C3), and which are sailboats (S1, S2).

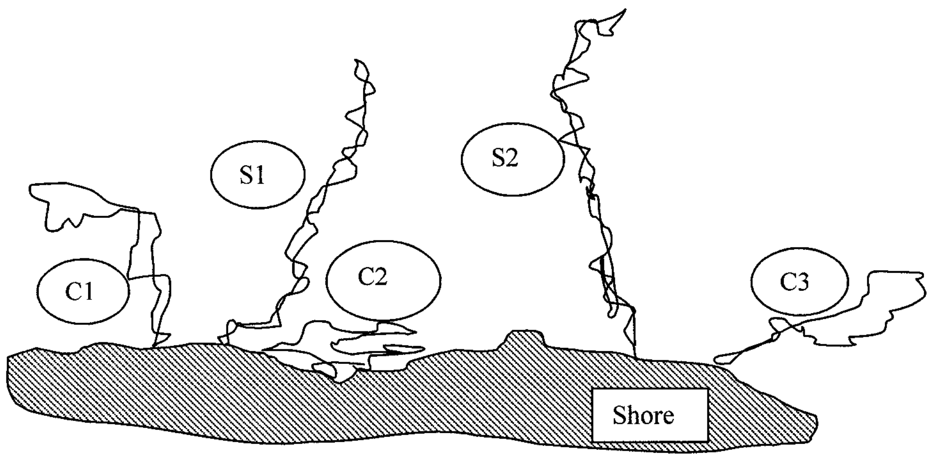


Figure 1-1 Patterns of Trajectories of Different Types of Boats

Objective 2: Characterizing patterns of different recreational boat types at different geographic locations

Canada has extensive and varied coastlines, exhibiting marked differences in geographic characteristics. To extend the pattern analysis beyond one specific location, another objective of this study is to determine whether the same range of recreational boat types have the same movements' patterns in different geographic settings. For example, as illustrated in Figure 1-2, are there any significantly different characteristics in kayak movement patterns between Shore 1 and Shore 2? Though this examination, it can be determined whether the results of this study are only applicable to the specific study areas, due to their unique features, or whether the outcomes can be generalized.

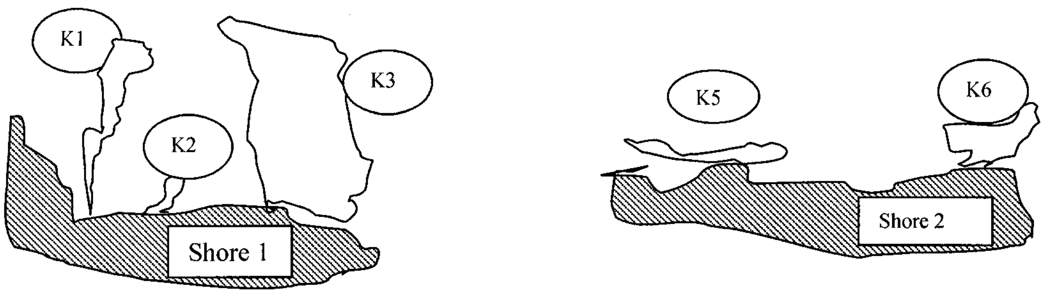


Figure 1-2 Patterns of Trajectories at Different Locations

In order to attain the two general objectives, there are specific objectives that must be met. They are listed below in the processing sequence used throughout this analysis.

1) To acquire GPS trajectory data, essential for realistic input. This study's primary data is indirect observation through GPS.

2) To process the original data before pattern feature extraction. Besides the critical steps of assiduous cleaning for preparation of the features extraction, the other indispensable process of this study is dedensification. The term dedensification describes the outcome of this operation, which is a trajectory with fewer GPS points while keeping all the indispensable features from the original trajectory.

3) To extract and calculate boat movement attributes according to previous trajectory studies and visual observations.

4) To evaluate the discriminating power of these attributes. Statistical tests are applied to every attribute to determine whether the attribute can significantly distinguish different recreational boat types or differentiate between different geographic locations.

5) To classify boating movements based on selected attributes. Models are constructed to determine the recreational boat type solely based on the GPS trajectories' movement features.

6) To simulate recreational boating trajectories for different boat types. Based on the discrimination and classification analyses, it can be established whether different boating trajectories should be simulated for different boat types. If yes, different recreational boat types' trajectories can be simulated based on the patterns characterized in this study. This can be applied to investigate maritime traffic densities and patterns for different areas of application. Moreover, the classification model might be used to assess the value of the simulation algorithm.

1.2.2 Anticipated Benefits

Having outlined the objectives of this study both generally and specifically, the broader context and benefits of achieving these objectives will now be presented. In order to construct a spatial and temporal risk model for recreational boating, two key aspects must be developed and explored. One is the exposure to risk, and the other is a classification scheme to distinguish between activities. Evaluating the patterns of recreational boating provides valuable information regarding these two issues. Consequently, recreational boating risk analyses can be improved. This section concentrates on outlining the role of exposure measures and classification models, as well as the possible impacts of this work.

1.2.2.1 Exposure Measures

Exposure is a typical measure of activity level. There are two main reasons to include exposure in spatial and temporal risk modelling of recreational boating. One is that exposure is an important risk factor. The more exposure a boat has, the more risk it faces. Another is that creating a recreational boating risk model requires activity levels as a baseline to establish relative risk measures. Conversely, the incident data alone could be analyzed to make recommendations to Coast Guard about hotspots (areas containing high numbers of boating activities and incidents) or trends, but those results might lead them astray if traffic is omitted from the analysis. The number of incidents and the rate of incidents in a particular area can reveal very different concerns.

There are many different potential measures for exposure, ranging from a rough measure whereby boating activity levels are assumed to be proportional to population, to more specific information such as the number of boat owners or boating participants. Other related information sources could serve the purpose such as the number of trips taken, the number of unique vessels in an area, the number of licensed vessels or the detailed number of boating-hours, and the total distance of trips. Considering incident rates for example, there was an average of 244 drowning and near-drowning incidents per year in Canadian waters based on Canadian Red Cross Society (2003) data from 1991-1999, although it should be noted these values are likely significantly underestimated. Table 1-3 shows diverse incident rates derived from different exposure measures.

Table 1-3 Incident Rates based on Various Exposure Measures

Exposure Measure	Incident Rate
Canadian Population	7.68×10^{-6} per person
Boat ownership	9.05×10^{-5} per household boat owner
Boating participation	2.44×10^{-5} per boater
Number of trips	2.87×10^{-6} per trip
Duration of boating trip	6.52×10^{-7} per boating hour
Trip distance	3.47×10^{-8} per km

Although the average incident numbers were derived from 1991-1999 data for this study while the boating behavior survey and population data are from 2004-2005, it nevertheless generates reasonable estimates and demonstrates the relative advantages and disadvantages of different measures of exposure. Using population as the underlying exposure measure is the easiest way to estimate an incident rate. Based on the Canadian population of 31,788,635 (Statistics Canada, 2005) the incident rate is 7.68×10^{-6} (244/31,788,635). Nonetheless, disadvantages of this estimation method are obvious: not all Canadian citizens own a boat, and of those owning certain kinds of boats, not all of them go boating every year. Thus, the results may be misleading. The MARIN phone survey revealed that 31 percent (251/818) of Canadian households do indeed own some type of recreational boat. As of January 2004, Canada had 8,701,700 households according to Statistics Canada; thus, at that time there were roughly 2,700,000 (31%* 8,701,700) households possessing a boat. Hence the incident rate based on household boat ownership is 9.05×10^{-5}. By using the rate of boat household ownership among Canadians as the underlying estimate of exposure, a more accurate calculation of incident rate may be achieved than that resulting simply from total population statistics. For instance, according to the Canadian Red Cross Society, the rate of drowning in Canada was nearly twice as high as that in the United States. However, it is important to note that these drowning rates were calculated based upon population rather than boat ownership or boating participation or frequency. The International Boat Industry (2005) estimate of the per-capita boat ownership rate was relatively high in Canada (1:5), but considerably less prevalent in U.S. (1:16). With these ratios, we can also compare incident rate based on boat ownership.

$$\because \frac{R_C^P}{R_A^P} = 2, \qquad \frac{B_C}{P_C} = \frac{1}{5}, \qquad \frac{B_A}{P_A} = \frac{1}{16},$$

$$\therefore \frac{R_C^B}{R_A^B} = \frac{N_C / B_C}{N_A / B_A} = \frac{N_C / (\frac{P_C}{5})}{N_A / (\frac{P_A}{16})} = \frac{5N_C / P_C}{16N_A / P_A} = \frac{5}{16} * \frac{R_C^P}{R_A^P} = \frac{5}{16} * 2 = \frac{5}{8}$$

R_C^P is the rate of drowning based on population in Canada, R_C^B is the rate of drowning based on boat ownership.
R_A^P is the rate of drowning based on population in the United States, R_A^B is the rate of drowning based on boat ownership.
B_C is the number of boat owners in Canada, B_A is the number of boat owners in the United States.
P_C is the population of Canada, P_A is in the population the United States.
N_C is the number of drownings in Canada, N_A is the number of drownings in the United States.

The derivation presents two contradictory conclusions. Based on population, the occurrence of drownings is much more serious in Canada than that in the U.S., however if based on boat ownership, the inverse is true.

Even using boat ownership as the exposure measure, the problem of usage frequency still exists, because it does not reveal how many household residents used those boats, and at what frequency. That is why using participation is a more accurate exposure measure than both population and ownership. Estimates suggest that as many as 10 million Canadians participate in recreational boating each year on Canadian waters (Groff and Ghadiali, 2003). Based on this concept, the incident rate is 2.44×10^{-5} (244/10,000,000). However, boating participants might have more than one boating experience per year. Strictly speaking, exposure is dependent on the amount of time during which there is a potential for an incident to occur, which directly relates to the number of boating trips taken. As shown in Figure 1-3, almost three quarters (74%) of

Canadians who own a boat say that they take between 1 to 10 boat trips in their most active boating months of the year (Pelot et al., 2004b), which is usually from June to September. So the overall average of this given range is approximately 8.5 trips during the most active boating month. Assuming that the number of trips during the least active boating months is negligible, the number of incidents per trip is consequently 2.87×10^{-6} (244/10M/8.5). Based on these calculations, the number of trips taken seems to provide a better measure of exposure, compared to population, boat ownership or participation rates.

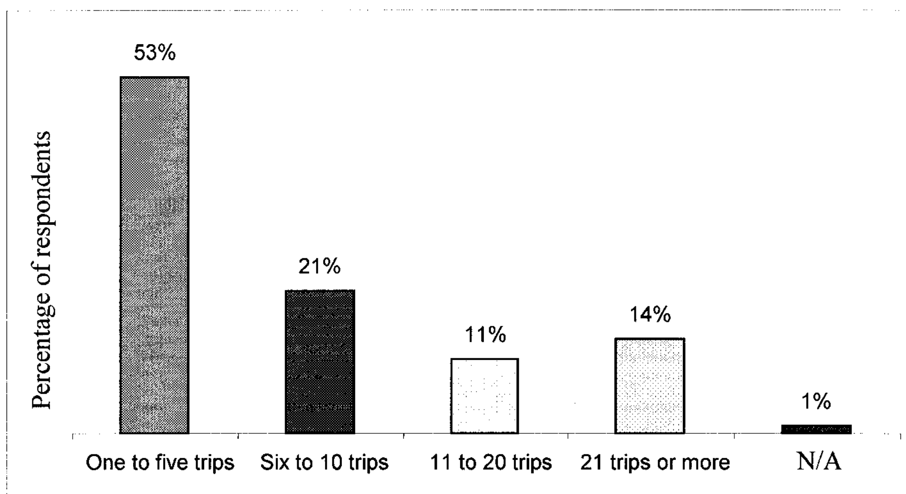


Figure 1-3 Number of Trips Taken in Peak Boating Season (Summer) by Phone Survey Participants

Respondents of the phone survey were read a list of options from which to estimate the length of time spent during a typical boat trip. Half of the people surveyed gave a response of equal to or less than two hours for total trip time. Another 35 percent indicated that their average boating trips were in the range of three to six hours. An additional 14 percent estimated that their typical trips last seven or more hours in duration. From these acquired statistics, the mean trip length across all recreational boat types was estimated at 4.4 hours (Pelot et al., 2004b). Therefore, the number of incidents per hour of boating is 6.52×10^{-7} (244/10M/8.5/4.4). Similarly, in order to consider trip distance, 46 percent indicated that they typically travel at speeds below 10 km/h (2.78m/s). Approximately another quarter (24%) of the respondents travel at average speeds ranging from 10 – 25 km/h (2.78m/s – 6.94m/s). Considering all reported speed values, the calculated mean trip speed is equal to 18.8 km/h (5.22m/s) (Pelot et al., 2004b), from which the average trip distance can be estimated using the mean speed and mean duration, which yields an incident rate per kilometre travelled equal to 3.47×10^{-8} (244/10M/8.5/4.4/18.8).

Compared to the estimations using the average number of trips taken, the total number of hours spent boating and the total distance travelled are more accurate and detailed exposure measures. Although these calculated rates are ideal in theory, in practice it is very difficult to obtain accurate estimates of trip duration and speed from surveys because boaters do not typically monitor these attributes. For duration, sometimes boaters just stop for resting, sightseeing or lunch. It is very hard to estimate the period during which movements take place. As for speed, it is even harder to establish a good estimate. Therefore, when calculating incident and drowning rates, the number of trips is a commonly applied measure of exposure.

In this study, we will use GPS data to obtain more accurate exposure measures. Because GPS can record time and boat position, with a typical accuracy for civilian GPS position of about 15 meters (Wikipedia, 2005), it results in more precise exposure measures for recreational boating risk analysis.

1.2.2.2 Boating Activity Classification

The second issue that must be considered to proceed with a spatial and temporal risk model is to develop some protocols for grouping activities. "Maritime recreation" is a very broad term. In theory it could encompass anything from a cruise ship voyage, to deep-sea sports fishing, to scuba diving (Walsh, 1991). In the present study, we restrict ourselves to outings using specific types of pleasure boats, with explicit regard to the purpose of the trip. Because of the broad definition of maritime recreation, the vessel type classifications are diverse as well. Table 1-4 provides some examples of classification schemes for different surveys and research aims.

Table 1-4 Different Boat Types Classification Schemes

Recreational Boat Owners Survey (Pelot et al., 2004a): PWC Kayak/Canoe Pedal Boat Zodiac Rowboat Houseboat Sailboat (with auxiliary power) Sailboat (with auxiliary power) Motorboat (with cabin) Motorboat (without cabin) Inflatable Raft Other	*Small Vessel Inventory in Canada:* Canoe, kayak, paddle boat PWC Sailboat<=6m and <10hp Sailboat<=6m and >=10hp Sailboat>6m and <10hp Sailboat>6m and >=10hp Motorboat<10hp Motorboat>=10hp and <=6m Motorboat>=10hp and >6m and <=8m Motorboat>=10hp and >8m
US Coast Guard Accident Report: Open Motorboat Cabin Motorboat Auxiliary Sail Sail Only Rowboat Canoe/Kayak Personal Watercraft Pontoon Boat House Boat Other	*Toronto Boat Show Stats (2003):* Runabout Cruiser Sailboat Fishing Boat Canoe/Kayak/Pedal High performance PWC or Jet Boat Pontoon or Desk Motor Yacht

Commercial Recreation Data (Pelot et al, 2004c):	*Bay of Fundy Spotting Handbook (Pelot et al., 2000):*
Jet Boat PWC Kayak/Canoe Sailboat (without auxiliary power) Sailboat (with auxiliary power) Motorboat with inboard motor Motorboat with outboard motor Non-powered, non-sail boat Houseboat Zodiac Fishing Boat Hovercraft Submarine/Submersible Inflatable Raft Other	Tour boat Motorboat (no cabin) Motorboat (cabin) Sailboat PWC Kayak Canoe Rowboat Sailboard Less than 20ft 20ft-40ft 40ft-65ft Over 65ft

There are many reasons to create groups for analysis. Often the groupings are established *a priori*. Sometime it is necessary to group things to restrict the number of distinct categories to a reasonable quantity, or else the analysis may be too complicated, or the results too confusing for effective action. Another reason is to avoid too small of a sample size. Otherwise, the number of some uncommon vessel types might be too small to carry out reliable statistics tests. In such case, these uncommon types need to be combined with other similar types according to some criteria. For example, with a limited sample size, trying to contrast uses of different canoes is likely impossible and not that informative, but grouping all canoes together to compare with kayaks could be fruitful. Most classification criteria involve vessel type, length and horsepower. None of these schemes distinguish based on vessel movement characteristics, which is a primary goal of this thesis. The approach then is to group elements *a posteriori* based on the outcomes of measured attributes. In this case the classification depends on how well individual cases can be discriminated based on the available attributes, and what the purpose of study is. It has been determined in this study that grouping boats according to their movement characteristics serves the dual goals of modelling vessel tracks in GIS systems, as well as providing pattern-based approaches for remote identification of vessel types for coastal security and safety.

Generally speaking, this study will examine GPS and other data sources, to find patterns of recreational boating activity, especially spatial patterns, which would yield insight into recreational boating exposure. Good exposure measures can help us perform risk analysis of recreational boating more accurately by obtaining activity level. Moreover the outcomes of this study can help detect boat types and abnormal movements, which will prove useful for coastal security, and better anticipation for response planning in the case of potential maritime casualties. Also using the pattern results from raw trajectory analyses, this study provides valuable input for simulating a representative trajectory of a single boat stochastically, as well as overall traffic representations assuming that adequate information is available on the distribution of outings.

1.3 Scope of Project

The scope of this study in terms of time period and study area is defined below. The types of recreational boats analyzed in this study are also provided.

1.3.1 Study Time Period

The amount of recreational boating activity is different across years and seasons, and even varies between weekdays and weekends. The frequency of boating is generally higher in the summer, daytime and weekends. The annual weather pattern in Canada is one evident reason for the variability. Another is that boaters have more leisure time and vacations in the summer and on weekends. Moreover, it is more pleasant and safer to boat during the daytime because of light and temperature conditions. The boat owners survey (Pelot et al., 2004a) addressed this issue by requesting participants to estimate the number of trips they took in each month of the preceding year. It is clear that boating is most frequent during the summer season, with the maximum number of trips occurring in July and August, as shown in Figure 1-4. Therefore, for this study, the GPS data was collection was performed during the summers of 2004 and 2005, but across different times and dates in order to obtain sufficient data.

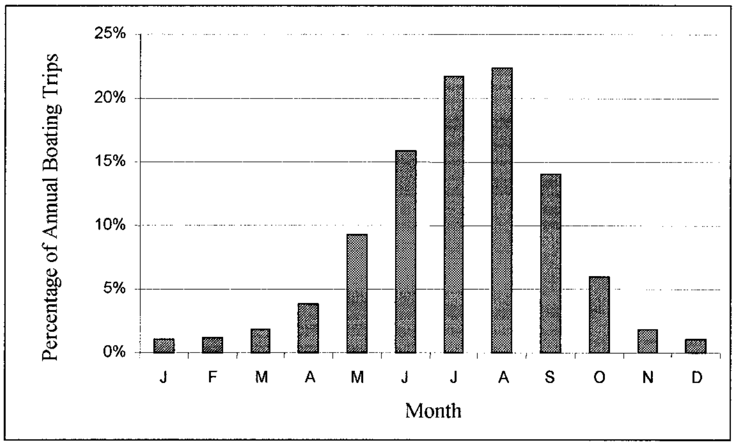


Figure 1-4 Percentages of Annual Boating Trips in Each Month of the Year

1.3.2 Study Areas

The study areas are navigable areas within CCG responsibility, mainly coastal areas around Halifax in Nova Scotia (NS), and the Saint John River area around Fredericton in New Brunswick (NB). These two areas bear different geographic characteristics, which can help us to address the second objective (i.e. the geographic location influence on boating movements) of this thesis. From the trajectories of these two areas, we will determine whether the boating patterns between these two areas are different. If not, that means geography is not a significant factor in influencing the patterns of recreational boating, at least within the scope of this study. If geography is not relevant, the boat type could be detected based only on the patterns of the GPS trajectories, a result which could be generalized to other regions. Otherwise, if the patterns are significantly different in the two areas, one can hypothesize that geography is an important consideration for vessel movement patterns.

1.3.3 Vessel Types

For the purposes of this study we will only conduct research on the following boat types:

- Canoe (C)
- Kayak (K)
- Motorboat (M)
- Sailboat (S)

Although some classification schemes may combine kayaks and canoes into the same class as shown in Table 1-4, we deliberately separate them to check if the spatial patterns are different, whereas they may seem superficially similar (for example they are both slower and are human powered compared with other boat categories). Sailboats are unique among these four types because they are more influenced by a natural outside force – wind. Motorboats are distinct from the others because of the speed which can be attained through their motors, although sailboats often use auxiliary power as well. Because these four types of boats represent a wide range of characteristics, they are deemed sufficient for the purposes of this study and other types are excluded.

It should also be noted that there are different sub-types of vessels because of varying length and horsepower of boats. Taking motorboats for example, intuitively higher horsepower boats are usually faster than less powerful ones. Some of this information was tracked by asking study participants to fill out an auxiliary form (Appendix 3), giving information about their boat characteristics and some other relevant features which cannot be ascertained through the GPS data, such as whether the motor was used during a sailing trip.

1.4 Outline of this Thesis

The recreational boating situation in Canada was presented earlier in Chapter 1, which included the popularity of this activity and its associate risk. The importance of this risk was highlighted by noting the indirect and direct economic and social costs established by other researchers. Although pertinent studies have been carried out to examine recreational boating activities relying on three principal data source types, survey results, various organizations' databases, and direct observation, no complete spatial and temporal risk model has been constructed for the purpose of recreational boating risk analyses. Two of the hurdles for achieving this are to establish accurate exposure measures and reasonable classification schemes. In section 1.2.2, it was explained why these two issues matter and how this research attempted to solve them. The context, study period, study areas and vessel types have also been defined in Chapter 1 to provide a clear scope within which the aims of this project are tackled.

Three important subjects are presented in the literature review in Chapter 2. The first involves the status of risk analyses of recreational boating, mainly those concerned with finding risk factors by applying simple statistical methods. The second aspect concerns exposure estimations for boat types other than recreational boats. The reasons why those techniques (i.e. calculating exposure from relevant databases) cannot be employed in this case are explained. Because this study centres on recreational boating GPS trajectory analysis, a review of other trajectory-based investigations is also presented in Chapter 2.

Chapter 3 illustrates the main process flow of this analysis, from data to information and then to knowledge. Section 1 and section 2 are about data collection and cleaning, as well as sample size assessment. This original data preparation part is crucial for the whole work although it is less technical than the following two stages. Without sufficient solid data acquisition the study cannot proceed. However, having data is almost useless if no information can be extracted from the data. So the next section is about attributes extraction and calculation which can serve to achieve the objectives of the research. Finally, the last section in chapter 3 is about knowledge formation, turning the calculations into insight.

The details of computing key attributes of recreational boat movements are presented in Chapter 4. Moreover, nine variables are ultimately derived from these five attributes. Among these attributes, turning angle data can engender an interesting feature: mean turning angle, to represent movement characteristics. However, to achieve this calculation required the development of a new line-simplification procedure, which is sufficiently complex that its derivation is deferred to Chapter 5. Furthermore, as described in section 1 of Chapter 5, these methods are based on existing geomatics concepts, used to simplify shorelines for example. Since dedensifying a recreational boating trajectory is somewhat different than simplifying other types of lines, this study proposes two modified algorithms, of which one (MDPA) is then employed for further analysis. Besides the line simplification algorithm, a context-specific algorithm was also devised to select the best simplified trajectory to calculate mean turning angle. The details of these methods are fully described and interpreted in Chapter 5.

All the work described above results in well-defined and prepared variables as inputs for the analyses presented in Chapter 6. This chapter applies univariate and multivariate analysis of recreational boating trajectories, and answers the two questions presented at the beginning of this study: 1) Do different boat types have significantly different boating characteristics? 2) Does geographic location influence boating characteristics significantly? Finally, classification models for boat types are also constructed in this chapter, with the ensuing assessment of results.

It is well-known that traffic estimation is indispensable for a comprehensive maritime risk analysis. The MARIS software can simulate traffic for ships, commercial tour boats, ferries and fishing boats (Pelot, 2001). Based on the outcome of this work, recreational boating trajectories for different boat types can also be simulated in MARIS. The simulation methodology and results are presented in Chapter 7. Moreover, the simulation limitations are noted as well.

Finally, a summary and conclusions appear in Chapter 8, and recommendations for future work are also discussed.

2. Literature Review

This section serves to assemble several published and unpublished analyses and theses relevant to the objectives of this study. The information presented is useful to serve as a starting point for the present research in establishing methodologies and procedures. The first section is about pertinent risk analysis for recreational boating suggesting that thorough studies about this activity have not been completed. One key ingredient for recreational boating traffic analysis is exposure, which has been explained in the preceding section on objectives, so the second part of this chapter is to review methodologies employed to estimate exposure for maritime non-recreational boating activities. Those calculations are much easier to conduct because of their relatively well-maintained data sources. Conversely, recreational boating activity lacks such data, and in addition the complexity of movement of recreational boats makes the exposure calculation very difficult. However, inspired by the concepts outlined in the third section on trajectory analysis, the rudiments of this project are formed.

2.1 Risk Factors of Recreational Boating

Recreational boating risk is evidently a very serious problem as described in the previous chapter. Lots of research has been conducted to determine which factors contribute to the risk, including which measures for alleviating such risks could be implemented. The methodologies used for finding risk factors are mainly surveys and incident databases, and most of the analysis methods are simple statistics tests.

In general, there are six categories of factors: vessel and equipment, boating circumstances, environmental conditions, operator behavior, personal characteristics and government and organizational aspects (such as the Canadian Coast Guard and Transport

Canada). Some obvious factors have been studied, but other more challenging ones have not. No literature has presented a complete overview of these potential risk factors; however some of them nevertheless provide evidence and suggestions for introducing regulations and laws. These factors will be explained below, but in no particular preference.

Intuitively, the first factor is the vessel itself, which defines or constrains boating characteristics and activities (Ministry of Natural Resources, 1990; Pelot et al, 1997; Breen and Hejzlar, 2000; Pelot, 2000b; McCarthy and Talley, 2001; Torres, 2001). For example, Personal Watercrafts (PWCs) are seldom used for fishing and virtually always go very fast, otherwise boaters cannot enjoy the high speed which is the principal feature of PWCs. Conversely, human-powered boats such as canoes and kayaks provide a completely different experience. This is one reason why boat types must be classified by group in most surveys in order to differentiate between recreational boating features. For instance the recreational boat owners' survey and the pan-Canadian phone survey, both of which were conducted by MARIN, defined more than 10 different boat types (Pelot et al, 2004a; Pelot et al, 2004b). In relating boat types to risk possibility using historical data of Canadian National Surveillance System for Water-Related Fatalities database from 1991 to 2000, the Canadian Red Cross Society (2003) reported that powerboats are more dangerous, accounting for slightly over half of the incidents (51%), especially the small open powerboats less than 5.5m in length have the highest drowning risk potential (38%). Canoe-related fatalities are salient as well. They made up a 22% share of all drownings based on the 10-year average. Figure 2-1 shows the composition of the pie chart for different boat types' incidents. Evidently boat type is definitely a very important factor.

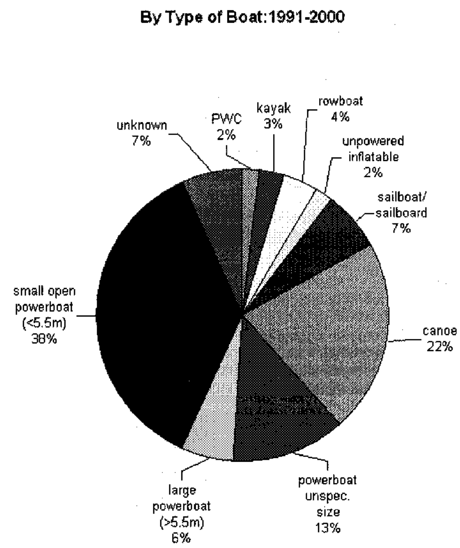


Figure 2-1 Recreational Boating Drowning Deaths in Canada by Type of Boat
(Canadian Red Cross, 2003)

Besides boat type, there are many other factors related to the vessel and its equipment, such as the age of the boat, the damage pattern (Breen and Hejzlar, 2000) and maintenance as an indicator of boat condition. Other than that, equipment on board such as navigation equipment (e.g. GPS), communication radio and safety equipment (e.g. lifesaving devices and fire extinguishers), sometimes play an important role in saving people's life, and such factors should be accounted for in related studies. As previously described, exposure is another significant factor, which can be reflected through the amount of boating time, the total distance travelled or the number of trips the boat has taken per time period. However, this factor has hardly been studied because of the difficulty of obtaining reliable measurements.

The second factor is boating circumstances, which places extra emphasis on the boat's situation during the boating activity, for example the distance from shore, the number of persons on board (POB) and the boating activity (Ministry of Natural

Resources, 1990; Pelot et al, 1997; Breen and Hejzlar, 2000; Pelot et al, 2000b; McCarthy and Talley, 2001). All of these may be indirectly associated with boating safety. Some operations for specific activities might be harmless themselves, for example reeling in a fish by standing near the shipboard, but such practices often generate unnecessary risk. Some researchers and organizations realized this, thus they conducted some pertinent investigations (Breen and Hejzlar, 2000; Ministry of Natural Resources, 1990; Pelot et al, 1997). The Ministry of Natural Resources' office of recreational boating (1990) studied recreational boating fatalities in Ontario from 1980 to 1987, and found that 33.5% of the fatal accident victims were fishing or intending to fish, which was the largest group. Pleasure boating as the purpose of trip followed with 20.7%. Other purposes like hunting and commuting caused a relatively smaller number of deaths. In addition, the compatibility between vessel type and activity is another aspect (Breen and Hejzlar, 2000). Because different boat types are more suitable for certain boating activities, the two should be compatible otherwise there is a greater possibility of incurring an accident.

The environmental factors such as geographic features and weather conditions comprise the third major risk factor to investigate. There are lots of studies about this, for example those on visibility (Breen and Hejzlar, 2000), which could be influenced by fog (McCarthy and Talley, 2001), the time of boating (i.e. daytime or night), or inadequate lights and markers (Uremovich, 2002). Based on the Canadian National Surveillance System for Water-Related Fatalities, 20% of deaths happened during twilight or after dark (Canadian Red Cross Society, 2001). The speed and direction of the wind cannot be neglected either (Pelot et al., 1997; McCarthy and Talley, 2001), because 32% of all recreational boating drownings occurred during strong wind condition. Moreover, waves accounted for 28% of the incidents in 1999 (Canadian Red Cross Society, 2001). Water conditions such as rapids, flow strength (i.e. current and tide) and depth are other important elements, especially given the cold water in Canada where the ocean temperatures can be as low as 10°C even in summer, and lakes and rivers even colder at certain times of the year. Recreational boating drowning deaths that took place in cold or extremely cold water comprised 47% of all the deaths in 1999 (Canadian Red Cross Society, 2001). Ice, reefs, rocks and underwater obstacles are physical geographical hazards (Ministry of Natural Resources, 1990; Pelot et al., 1997; Pelot, 2000a). In a broad sense, boating traffic level itself is another element of the environmental factors, especially near shore where the congestion of boat traffic may often be dense. The activities of other boats (Ministry of Natural Resources, 1990) and the speed of other boats (Environics Research Group Ltd., 1998) can adversely impact a specific vessel.

Human error is always a significant factor in safety analyses, if not the principal one. Hence the fourth risk factor which has been studied relates to behavioral factors. Some boaters are more risk-taking than others, practicing unsafe actions (Pelot et al., 1997; Uremovich, 2002; Groff and Ghadiali, 2003) and/or neglecting boat maintenance (Ministry of Natural Resources, 1990; Environics Research Group Ltd., 1998; Uremovich, 2002). Among the behavioral factors, speeding, alcohol use, and the ignoring or refusing to use personal floatation devices (PFD) have been extensively examined (Ministry of Natural Resources, 1990; Environics Research Group Ltd., 1998; McCarthy and Talley, 2001; Uremovich, 2002; Groff and Ghadiali, 2003). Based on the Canadian National Surveillance System for Water-Related Fatalities database, it was found that approximately nine in 10 of the drowning victims, were not properly wearing a PFD (Canadian Red Cross Society, 2001).

The fifth risk factor concerns personal attributes. Amongst the many factors studied, males accounted for 90% of all recreational boating drownings in 1999 (Groff and Ghadiali, 2003), especially males between the ages of 25 and 45 (Environics Research Group Ltd., 1998; McCarthy and Talley, 2001; Groff and Ghadiali, 2003). With respect to cultural groups, the aboriginals' drowning rate in 1999 was eight times higher than that of non-aboriginal Canadians, even though their whole population accounts for only 3% of the Canadian population (Groff and Ghadiali, 2003). Other personal factors studied include religion, education, personal income, boating experience, physical conditions like eyesight, and swimming skill (Ministry of Natural Resources, 1990; Pelot et al., 1997; Environics Research Group Ltd., 1998; Breen and Hejzlar, 2000; Wang, 2000; McCarthy and Talley, 2001; Uremovich, 2002; Groff and Ghadiali, 2003).

The final risk factor concerns government and organizations. For example the wearing of PFDs is not mandatory in Canada, although Groff and Ghadiali (2003) suggested mandatory-wear legislation based on their research. They postulated that such regulations would decrease the mortality rates due to drowning. Similar laws, such as minimum operator age, mandatory boating education (Pelot et al., 1997; Environics Research Group Ltd., 1998; Breen and Hejzlar, 2000; Wang, 2000; McCarthy and Talley, 2001; Uremovich, 2002; Groff and Ghadiali, 2003), and organizational programs related to operation and safety training (Wang, 2000), have been evaluated. Introducing such laws and regulations also require the corresponding means to enforce them (Environics

Research Group Ltd., 1998; Groff and Ghadiali, 2003), otherwise, those laws and regulations still cannot contribute to mitigate the boating risk.

The factors mentioned above are not proven to be risk factors that significantly influence the recreational boating, because all of the analyses above are based on simple summaries of specific data instead of statistical tests. Taking the personal factor "gender" for example, in order to target education or prevention, the statistics presented in these previous studies do not explicitly prove that males are riskier boaters, because that would depend on the amount of boating activity by each gender. As another example, countering Groff and Ghadiali's evidence for the mandatory wearing of PFD, Wang (2000) reported that mandatory wearing of personal floating devices did not significantly reduce boating accidents based on recreational boating accidents and fatalities from 1990 to 1994 in different states in the United States, which have diverse state laws on wearing PFD issue. One possible reason is that boaters tend to undertake riskier actions because they feel safer with the PFD on. These individual studies contribute to the body of literature on boating safety, but none comprise a comprehensive risk analysis.

2.2 Exposure Measures for Some Other Boat Types

It would be expected that the amount of exposure affects the likelihood of encountering potential hazards, some of which may provoke accidents. The higher the level of exposure, the more risk associated with the boating activity. In order to examine the relationship between exposure and risk for preventing or decreasing the damage resulting from incidents, quantifying the exposure, or level of boating activity, is necessary. Hence, the literature related to exposure research is presented here. It is more difficult to measure traffic on open water (in particular the ocean) than it is on land because roads define the travel paths, or at least confine the paths. Measuring recreational boating exposure is even harder than other boat types (i.e. non-recreational) which often are tracked to some degree. In this section, we first introduce some methodologies used to calculate exposure for other categories of boats, and then point out why these methodologies are not ideal for recreational boats, as well as note some shortcomings of these methodologies.

Uremovich (2002) used the Eastern Canada Vessel Traffic Services Zone (ECAREG) and Arctic Canada Traffic Zone (NORDREG) trip dataset to calculate exposure for the risk analysis of merchant shipping traffic in the Canadian Atlantic region. These vessel monitoring systems record departure port, destination port, intermediate Canadian reporting way-points and vessel characteristics for trips that either enter or leave a port within Atlantic or Eastern Arctic Canadian waters. Using these movement location points for each trip, a novel track generation algorithm (Hilliard and Pelot, 2002) connected the spatially referenced points while avoiding land, and then calculated the lengths of the trips within the defined study area, which included Atlantic Canadian waters and a portion of Eastern Arctic Canadian waters, also extending into the St. Lawrence River.

Uremovich (2002) also postulated that the level of risk increases with the level of exposure. He calculated merchant ship exposure using three different measures: the number of trips within the study area, the number of unique (distinct) vessels within the study area, and the sum of the trip distances within the study area (km). In order to compare these three measures, Uremovich analyzed and examined the annual relationship between the three activity exposure measures for the period 1988-99. He found that there were strong linear relationships between the annual number of unique vessels transversing these waters and annual number of trips, and between the annual sum of distance of trips (km) and the annual number of trips. Because of the high collinearity among these exposure measurements, it does not matter which one is used for analysis. Since it is easiest to count number of trips via the database, and the fact that this measure is commonly applied by others, it was selected as the measure of exposure in that study for risk analysis of merchant ships.

Shields (2003) analyzed exposure measures for the assessment of fishing vessel risk in the Bay of Fundy. The dataset for fishing boats used for his study is referred to as Catch-Effort (CE) maintained by the Department of Fisheries and Oceans (DFO), which documents the ports, boat characteristics and fishing activity throughout the Atlantic region. The fishing activity is referred to as an effort, which represents one defined attempt at landing fish. Shields (2003) designed four different measures of exposure: number of trips taken, number of fishing efforts made, number of vessels licensed, and number of boat-hours. All of this information could be extracted from the CE database directly. Unfortunately, he could not prove the hypothesis that the probability of incident is proportional to the amount of exposure measured by Trips, Vessels, Efforts, or Vessel-hours. He even found a negative relationship between incidents and all exposure measures in the Bay of Fundy. Does that mean the reasonable assumption of a functional relationship fails for fishing vessels? By further examination of the data and consultation with DFO, it was found that there was an increase in the number of vessels being recorded in CE database within the last several study years because the data collection technique and criteria have been changed, so the exposure measures could not be compared on a reasonable basis.

Shields' (2003) study reinforced how important the data source is, and not to use it without fully appreciating its limitations. So when Shahrabi (2003) studied fishing vessel risks in the Canadian Maritimes area, he chose Zonal Interchange Files (ZIF) as his data source. Unlike Uremovich (2002) who calculated the trip's distance as an exposure measure, Shields calculated the number of boat-hours because some of the reported fishing efforts were not geo-referenced. It is impossible to connect points to generate trajectories without knowing the latitude and longitude of the locations, thus it is impossible to estimate distance. Shahrahi also faced the same problem. However, he randomly assigned a latitude and longitude for each fishing catch effort within the reported Northwest Atlantic Fisheries Organization (NAFO) Division according to the vessel's destination on the fishing grounds on a particular date. Therefore, a fishing trajectory could be generated through the track generation process, which was also used in Uremovich's study, and then a representative distance could be calculated correspondingly.

The studies introduced above in this section have all attempted to develop trip-based analyses to estimate the possibility of incidents. Shahrabi (2003) also tried a grid-based approach, whereby the measure of exposure was the density of traffic in each grid cell. Such analysis is more useful for finding hazardous areas. Thus, different study objectives and data sources lead to various constraints on generating useful measures of exposure.

The dispersion algorithm, which Shahrabi (2003) used to randomly assign within a feasible area a geo-referenced effort location to generate tracks for fishing vessels, is interesting but may require more validation, especially for the purpose of grid-based analyses. Extending the idea of concept of a visit-day introduced by Pelot (2000b), Torres (2001) applied this as the exposure measurement for fishing vessel risk analysis in the Canadian Atlantic region using the ZIF database. One visit-day is defined as a particular fishing vessel reporting in a particular area on a particular date, regardless of how many catch-efforts it makes or how long it spends in the area. With the available ZIF data, a visit-day is a good measure for analysis, although it does not reflect the duration or distance of the trip.

Note that the data sources used for exposure calculations are typically existing databases from organizations, which often have some drawbacks. One is consistency, such as the CE data set used by Shields (2003) which has separate databases for each region of Atlantic Canada: Newfoundland, the Gulf, and Scotia-Fundy. Each database is structured and coded somewhat differently, although they record similar information. Another drawback is the dependence on the collection procedure, which might result in unreliable information, hence surprising conclusions, such as the negative relationship between exposure and incidents explained by Shields (2003). Furthermore, the available data perhaps cover only a period of several years, which leads to a relatively limited sample size, particularly for trend analysis. These issues can prevent one from drawing robust conclusions based on statistical test results.

With regards to recreational boats, the primary problem is that there are no databases that record the activities of recreational boats. Even if there were, the trajectory generation algorithm (Hilliard and Pelot, 2002) is a poor choice for recreational boating traffic simulation, because this algorithm is based on shortest distance criterion (with land avoidance). This assumption may apply reasonably well to commercial ships that usually take the shortest distance between the launch and destination ports. However, recreational boating typically involves a great deal of meandering as it is generally not destination oriented, and the trajectory generation algorithm cannot account for different patterns associated with various recreational boat types.

2.3 Trajectory Analysis

Although few studies exist on recreational boating trajectories for risk analysis purposes, in this section research on track patterns in related fields will be introduced, which not only initiated this project but also inspired some new methodologies and techniques.

Smith (1974) described and analyzed European Thrushes' food search paths to gain knowledge of their good search behaviors. The existence of environmental heterogeneity, such as food distribution, influences search behavior in a complicated way. In order to minimize such complications, the study area was an urban park in central Oxford with relatively uniform conditions. Typically, several thrushes would appear on the meadow in that park almost synchronously, remaining for less than thirty minutes in the absence of disturbance, and their foraging movements were divided up into a series of natural units for convenient study: the successive moves and the turns made between them. The study meadow was designed into 6×5 array of grid squares. Each side of the square was 4.57 meters, and the corners of it were marked by colour-coded pegs and the geometrical centre with a numbered peg. One observer spoke into a tape recorder about the bird's successive movements' position, time and direction. The other observer provided commentary on the feeding and other behaviors of the bird onto the same tape. So the tracks and behaviors were separately recorded this way, and the tracks were plotted on graph paper except for those lasting less than one minute. With 69 track maps, Smith (1974) calculated 1) the duration of the moves and the pauses; 2) move length; 3) angle of turn; 4) overall speed of movement; and then examined the biological significance of the observed movement patterns. For example he used chi-squared sequence tests to examine whether moves or turns of given size classes occurred in random sequences or if there were any regularities in the sequences.

Wiens et al. (1995) studied patterns of insect movement in micro-landscape mosaics. Beetles, grasshoppers and harvest ants were chosen as this study's subjects in two 25 m^2 study areas, one with bare ground and low grass and the other containing a mixture of grasses, cactus, and low shrubs. Researchers tracked individual insect for 5 to 30 minutes in these two different areas using small numbered flags to mark the locations at 5 to 30 seconds time intervals. Unlike Smith (1974) who plotted the tracks, Weins (1995) generated pathway maps electronically. He derived scale-dependent measures such as the straight-line distance from beginning to end of a pathway to characterize movement in terms of absolute distance, and he found the scale-dependent pathway measures varied significantly among species and with variations in the spatial heterogeneity of the mosaic. Additionally, a scale-independent measure, fractal dimension, was calculated to quantify the complexity or tortuosity of a pathway.

The word fractal is a compression of the words fraction and dimensional, and expresses the idea that a line may be somewhere between one and two dimensional, with fractal dimension (O'Sullivan and Unwin, 2003). The fractal dimension of movement pathways over a plane surface lay between 1.0 (a straight line) and 2.0 (Brownian motion) (Dicke and Burrough, 1988; Milne, 1991). The process to determine the fractal dimension is illustrated in Figure 2-2. Supposing the dashed line is a trajectory, a yardstick l_0 measures the length of the trajectory as 5 times in Figure 2-2(a), but 11 times if the length of the yardstick decreases to half of the length of yardstick l_0 as shown in Figure 2-2(b), and 23 times if the length of the yardstick used in Figure 2-2(c) decreases to a quarter of the original one. Thus

$$\frac{l_j}{l_{j+1}} = 2, \ but \ \frac{n_{j+1}}{n_j} > 2$$

Note that the ultimate yardstick measure in each case has a high possibility of not being equal to l_0, l_1 or l_2, exactly (i.e. it is virtually always less than the length of yardstick), hence producing a coarse measurement. Assuming the ratio of the number of segment lines at any two scales is in constant relation to the ratio of the lengths of the yardsticks, then

$$\frac{n_{j+1}}{n_j} = \left[\frac{l_j}{l_{j+1}}\right]^D \quad 1 \le D \le 2 \qquad [2\text{-}1]$$

where D is the fractal dimension common to the two measure scales.

D can be derived by rearranging Equation [2-1] as

$$D = \frac{\log({}^{n_{j+1}}/_{n_j})}{\log({}^{l_j}/_{l_{j+1}})} \qquad [2\text{-}2]$$

In practice, using a Richardson Plot (Figure 2-3) showing at least three yardstick lengths and the resulting yardstick counts, the fractal dimension D can be estimated by linear regression.

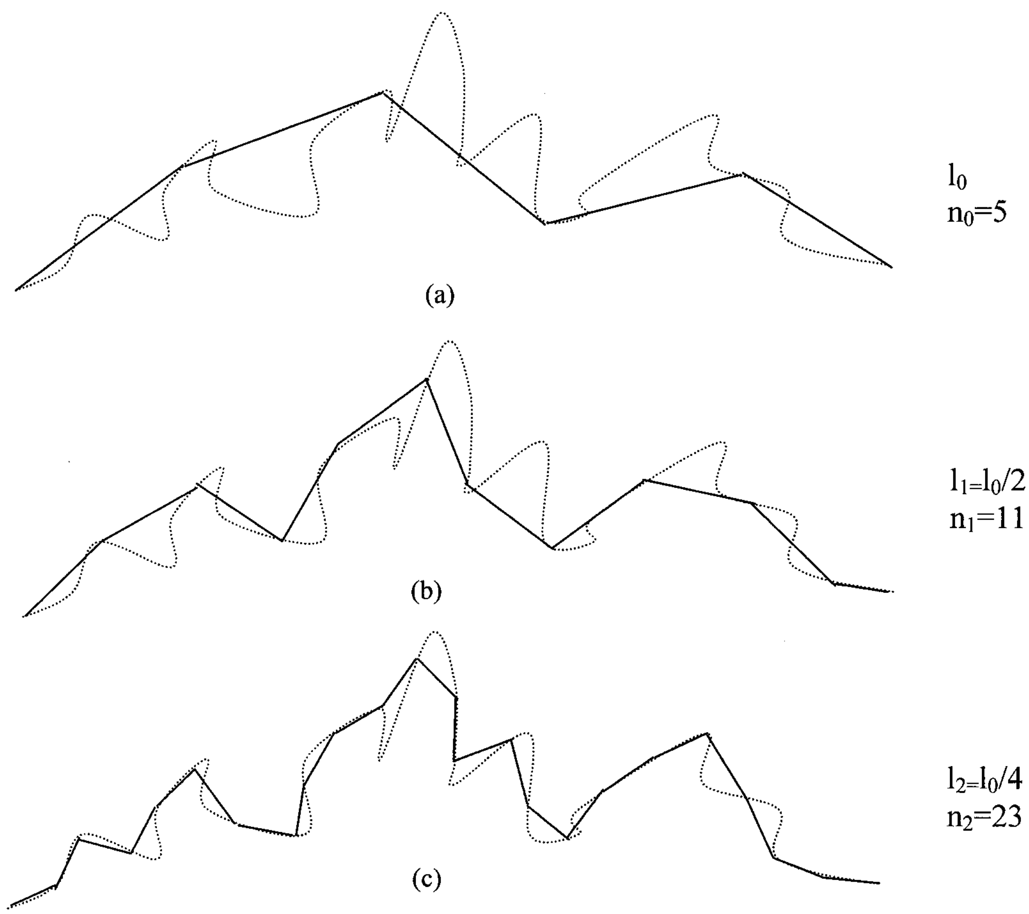


Figure 2-2 Fractal Dimension Determination

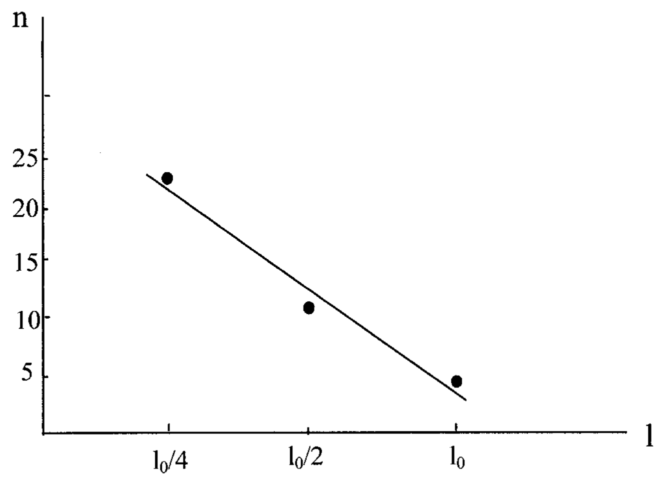


Figure 2-3 Richardson Plot

Fractal dimensions provide a way to assess similarities or differences in how organisms respond to heterogeneity that is independent of differences in body mass, physiology, diet, life history, or agility (Wiens et al., 1995), however fractal dimensions do not give a complete representation of movements by themselves, that is why Wiens et al. (1995) also used scale-dependent measures.

Studies of birds and insects constricted within an easily observable area have drawn on measures such as fractal pattern, movement rate, length, duration, direction and turning angle to quantify movement paths (Smith, 1974; Dicke and Burrough, 1988; Milne, 1991; Turchin, 1991; Wiens et al., 1995), however those approaches had not been used to understand behavioral patterns of far-ranging organisms such as mammals over seasons simply because of the logistical limitations of obtaining continuous, accurate location data (Koenig et al., 1996). With the advance of technology however, Poole et al. (2000) and Apps et al. (2001) used radio-telemetry or other remote monitoring devices to group all animal locations within one behavioral category. Johnson et al. (2002) applied

GPS collars to record frequent and accurate relocation of large mammals, caribou, and reconstruct their movement paths. These paths gave information about the frequency of movement events and the rate of movement of each event, which led to insight on the relationship among movement behaviors, land-cover type, energetic costs of movement, as well as season and predation risk. Hunter (2005) from the Mobile Multi-Sensor Systems Research Group at University of Calgary also used GPS collars with a camera to track grizzly bears' location and record their behaviors, as well as the environment. A spatial-temporal model was built to understand the resource requirements of grizzly bears through these GPS data. The above methods to monitor animal movement behaviors and resources can be extended to fishermen's search behavior and fish sources, which can help to describe and analyze fisheries, as well as improve management regulations. Dreyfus-Leon (1999) made an individual-based model of fishermen's search behavior with neural networks and reinforcement learning.

Besides these trajectory analyses to establish the relation between animal movement behaviors and resources, trajectories analysis have also been used for safety purposes. Johnson and Hogg (1996) and Makris and Ellis (2002) used video surveillance of pedestrian trajectories to aid the recognition of unusual behavior, identified as atypical motion. More advanced than previous research for surveillance and event recognition, which relied on known scenes where objects tend to move predefined ways (Howarth and Buxton, 1992), Johnson and Hogg (1996) chose an open pedestrian scene in their experiments, where pedestrians were free to walk. A fixed camera tracked uniquely labeled objects frame by frame to update their positions at a fixed rate, so an object's trajectory could be described in terms of a sequence of flow vectors including position and velocity information. Applying this information, probability density functions of both instantaneous movements and partial trajectories within a scene were developed to recognize atypical movements and to flag possible incidents. Makris and Ellis (2002) tracked pedestrian pathways from video sequences of nature outdoor scenes over long time periods as well, and then automatically extracted a mean path which was defined as the most frequently used path. The mean path represented a typical movement and could be used as support to recognize unusual movements and patterns of behaviors. Moreover, the mean path provided an efficiently compressed method to encode and annotate individual tracks to construct a log of movement patterns over long periods of time.

Maritime trajectory analyses have been developing gradually as in the other fields. The United States Coast Guard has trajectory analysis specialists, for which one responsibility is the estimation of the movement and behavior of spills based on visual observations, remote sensing information, computer modelling, observed and predicted tidal patterns as well as current and weather data (USCG, 2005). Other maritime studies have been carried out around harbour areas, where the traffic is intense, especially from small crafts. Pasquariello et al. (1998) developed techniques to detect little craft from among other objects with raw image sequences acquired by radar, which could help deck officers and navigators to avoid collisions. Liowski et al. (2000) applied radar images as well to extract bearing, speed, and distance between ships to determine a safe trajectory for a ship, or more precisely to determine an effective area, as defined by Goodwin (1975) to be the area around the ship which a navigator would like to keep clear with respect to other ships or stationary objects.

Another interesting application of trajectory analysis arises in certain sports. Little and Gu (2001) acquired trajectory data by employing parsed video to obtain individual shots in tennis or hockey, and then key frames from each shot were extracted and merged into coherent groups. They isolated two elements from a trajectory: the path in the image and the speed at which the body moved along the curve, which represented the object motion. Unlike other trajectory research described above, all trajectory attributes such as distance, speed and angle were extracted from raw data sources, the path and speed curves were obtained from smoothed data by Little and Gu (2001). Because they thought only corners and high curvature points aided object recognition, they used a local approximation algorithm developed by Chang et al. (1991) to retain such points. The main algorithm depends on the distance of a point from the segment connecting its two neighboring points. Points where this measure exceeds a threshold are retained, otherwise discarded. This idea is very similar as the well-known Douglas-Peucker algorithm, which will be introduced in detail in Chapter 5. The features that Little and Gu (2001) derived were invariant to scaling and rigid motions and preserved local features, which could assist a coach to search for certain patterns of movement in a tennis or hockey match.

As to other sports such as rowing, skiing and bicycling, coaches have tried to understand, analyze and predict the impact that the crew's effort and movement have on the motion of the equipment. Martin and Bernfield (1980) used a camera operating at 24 Hz (24 frames per second) and photogrammetric techniques to record the position of a 1976 U.S. Olympic eight-oared crew, and then they calculated the velocities trying to find the effect of stroke rate on the velocity. But the camera must be set stationary on the bank to achieve required accuracy, thus the crew's performance could only be reliably assessed across a 200 meters range. Because GPS is a position and timing tool, researchers can use it to determine both velocities and accelerations for sports training purpose. Moreover, GPS is very accurate, which enables one to determine velocities from as slow as a few millimetres per annum (Beavan et al., 1999) to supersonic speeds (Haering, 1998). A detailed explanation on the operation of GPS will be presented at the beginning of the next chapter. GPS has been used to track downhill skiers during training for competitive races to provide position, velocity and acceleration information (Skaloud et al., 2001). Lambert and Santerre (2004) used GPS to monitor the performance of canoeists over a competition course. In a similar study, Magee and Denys (2005) applied kinematic GPS to record two rowing trials' positional data at a 10 Hz (10 measurement per second) frequency rate, which led to sufficient quality and quantity of trajectory data to measure even small dynamic changes in the skiff's motion.

Trajectory analysis is not a novel research method, as demonstrated through these diverse applications, and yet there is room for much development in this field. From the beginning, trajectory data was obtained by human observations. With the advancement of technology, cameras and video cameras are employed to acquire trajectory data, and presently remote sensing such as radar and GPS are applied to acquire more and more precise data. Trajectory analysis covers many research fields such as zoology, ecology, sports and safety, backed by the techniques of image computing, pattern recognition, geomatics and simulation. This study aims to find patterns of recreational boating trajectories based on GPS data.

3. Methodology and Techniques

Succinctly, this research is about pattern analysis of recreational boats' GPS trajectory. This chapter presents the general processes applied to the work: 1) how to acquire GPS data and the pretreatment of the data; 2) how to infer interesting and important features' information from original GPS data; 3) how to detect boating patterns for different boat types. The following sections will present these issues sequentially. However, the emphasis in this chapter is put on the data acquisition and preparation, with every detail thoroughly explained. Only general introductions are presented about information extraction and knowledge attainment, with those topics elaborated in later Chapters 4, 5 and 6.

3.1 Data Acquisition

The first GPS (Global Positioning System) satellite was launched by the U.S. Military in February 1978 (Dykes, 1999), and now there are 24 satellites transmitting back to Earth. A GPS receiver can lock these signals and process them to triangulate a precise location on the globe. The procedure is demonstrated in Figure 3-1.

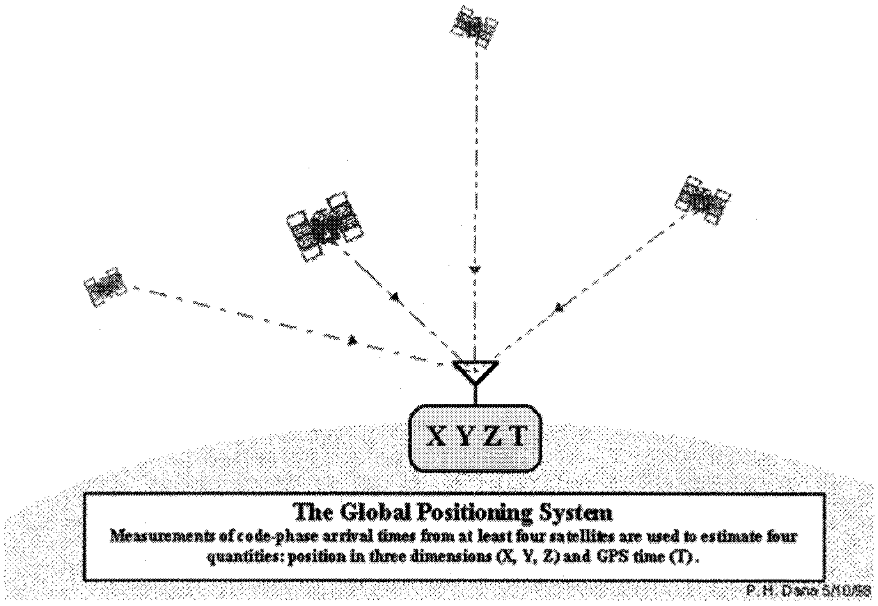


Figure 3-1 Global Positioning System (Dana, 1997)

The satellite signal contains three pieces of information: the satellite number, its position in space and the signal origination time. The GPS unit receives that signal, and records the receipt time, then compares it with the satellite signal origination time to determine how far away that particular satellite is. Taking Figure 3-2(a) for example, the GPS unit measures d1 from one satellite, so the points on the circle C1 whose radius is d1 are all possible locations. With a second satellite, the possible positions are reduced to 2 points, corresponding to the intersects of these two circles: C1 and C2. So if we want to know the longitude and latitude of a position, a third satellite is necessary as shown in Figure 3-2(b). Hence with a minimum of three satellites, a GPS receiver can determine a latitude/longitude position, and with a minimum of four satellites, it can determine altitude as well. The more satellite signals received, the more accurate the measurements.

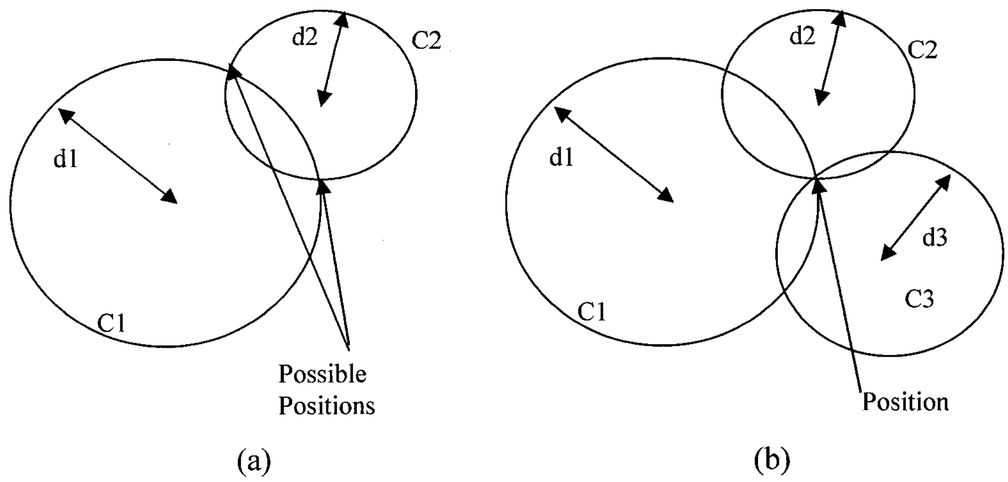


Figure 3-2 Working Algorithm of GPS

During the summers of 2004 and 2005, volunteers who were willing to carry one of the 11 GARMIN GPS 76 marine navigator units during their recreational boating outings were found at boating spots and boating clubs around the two study areas, referring to the coastal region around Halifax in Nova Scotia (NS), and the Saint John River area around Fredericton in New Brunswick (NB). These two areas exhibit different geographic configurations, which are required to address the second thesis objective on location comparisons. Figure 3-3 is a picture of GARMIN GPS 76, which is designed to provide precise GPS positioning with less than 15 meters error. However, the study concentrates on the relative distances of the GPS points within a trajectory, and these absolute positioning errors will not therefore affect the trajectory-based calculations. It could however cause some error when calculating distances from shore, but within the context of this study that magnitude of error is not problematic. Moreover, the volunteers were asked to complete an auxiliary form with supplemental information (Appendix 3). For instance, the characteristics of their boat such as horsepower and length, information about themselves including boating experience and whether they were formally trained, and finally some information about their outing habits such as whether they used the motor during sailing.

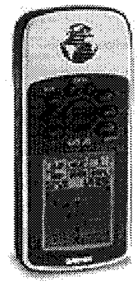


Figure 3-3 GARMIN GPS 76

Each file of GPS trajectory data collected is named according to a specific structure as follows:

- The first letter defines the boat type
    - c: canoe
    - k: kayak
    - m: motorboat
    - s: sailboat
- The next four numbers are the date the trip was taken, in the form MM-DD
- The next two numbers are the Coordinated Universal Time (UTC) of that boat's departure on its trip, in 24 hours format
- The last two numbers represent the ID# of the GPS used for that trip.

For example c-0713-10-08 denotes a canoe trajectory collected on July 13 starting at 10 AM from GPS unit 08.

In order to obtain reliable results, a suitable sample size must be established. The sample size depends upon the minimum detectable difference of interest, the acceptable probability of rejecting a true null hypothesis (alpha), the desired probability of correctly rejecting a false null hypothesis (power), and the variability within the population(s) under study. Aside from these statistical considerations, reality dictates that a study is also constrained by resources and budget.

Initial GPS data collection began in the summer of 2004 around Halifax Regional Municipality (HRM). This yielded a total of 39 trajectories: 16 canoe trajectories, 2 kayak trajectories, 1 motorboat trajectory and 20 sailboat trajectories. This whole data set provides enough information to get a rough idea about how many trajectories would be needed to achieve the research goals. According to the mean speed and its deviation, the following sample sizes were calculated for comparing between pairs of boat types, setting α equal to 0.8 and β equal to 0.1:

Table 3-1 Estimated Equal Sample Size

Two-Tail / One-Tail	Canoe	Kayak	Motorboat	Sailboat
Canoe		58	4	5
Kayak	42		4	6
Motorboat	3	3		8
Sailboat	4	5	6	

Above the diagonal in Table 3-1 are the sample size results used a two-tail calculation for two sample tests of normal means and assuming equal sample sizes based on mean speed. It shows that many canoe and kayak trajectories are needed if we want to distinguish between them. This is predictable because canoes and kayaks are very similar, and if we are determined to distinguish between them based solely on mean speed, a large sample size would be obligatory. At least 58 canoe and kayak trajectories would be needed to tell the difference between them with $\alpha = 0.8$ and $\beta = 0.1$, which is difficult to obtain with limited resources. However, it is usually true that the mean speed of these four boat types increases in the order of canoe, kayak, sailboat and motorboat (as shown later), so a one-tail test can be used to calculate the sample size, resulting in a smaller requirement than that calculated by the two-tail test. The results are listed below the diagonal in Table 3-1. As expected, the necessary sample size required is decreased, but the number of canoe and kayak trajectories is 42 respectively, which is still too large for practical purposes.

It was observed that sailing is a relatively more popular boating activity in the NS study area, hence implying that it is easier to acquire sailboat trajectories. As a matter of fact, 20 sailboat trajectories were collected in the initial data collection phase, which corresponds to the largest sample size for that vessel type. This fact led to treating the sailboat as the basis of comparison for the other boat types, to allow for the fact that there are different levels of difficulty to get samples from different subjects (i.e. boat types). Following the recommendation by Kutner et al. (2005), which is setting the sample size of the base comparison subject to be twice as large as for the other subjects, the sample size for sailboat was assumed to be double that of the other boat types in order to improve the precision of the three pair-wise comparisons. The outcomes are listed in Table 3-2. No matter which test is used, one-tail or two-tail, the sample sizes are practical considering statistical, resource and budget aspects.

Table 3-2 Sample Size Estimation with Sailboat as baseline

	Canoe		Kayak		Motorboat	
Sailboat (Two-Tail)	6	3	7	4	15	8
Sailboat (One-Tail)	4	2	5	3	11	6

3.2 Data Preparation

Cleaning the GPS trajectories from raw data is necessary, such as when Smith (1974) discarded tracks lasting less than one minute while plotting the European thrushes' food search path on graph paper. In this study, five cleaning operations are required: irrelevant point deletion; accompanying trajectories deletion; elimination of travel pauses; erroneous point deletion; and trip delineation.

First, the trajectories outside the study area, which is constrained to navigable waters within CCG responsibility, are deleted. Because MARIN distributed its GARMIN GPS 76 units to volunteer boaters, there were no explicit controls on where they chose to go. For instance, the data file k-0716-18-04 is not in the Saint John River, nor in any of the study areas, so it was discarded. Additionally, any points on land at the beginning or end of the trip were removed because sometimes boaters left the GPS unit on when they weren't on the water.

The second scenario to be considered arose when more than one boat with a GPS unit went out together. Figure 3-4 demonstrates one of the examples of this occurrence. The solid line trajectory is k-0710-19-06 and the dashed line is k-0710-19-11 travelling in NB. As seen from the name of these two files, both are kayaks that went out on July 10^{th} starting at 19:00, and the trajectories are almost completely overlapping, showing that these two boats were accompanying each other. If both of the trajectories were retained for analysis, the results would be affected since their paths are not independent. Therefore, the less experienced one is kept for further analysis because the more experienced boaters are assumed to accommodate the less experienced ones. In this case for example, the boater driving k-0710-19-11 has 10 years experience while the other one has 12 years boating practice, so the k-0710-19-11 would be kept. Such ancillary information could be drawn from the auxiliary GPS survey. Following this cleaning criterion, the sample size of kayaks in NB dropped from 13 to 8. Although the number of good samples is diminished, accuracy is gained because of the resulting random sample.

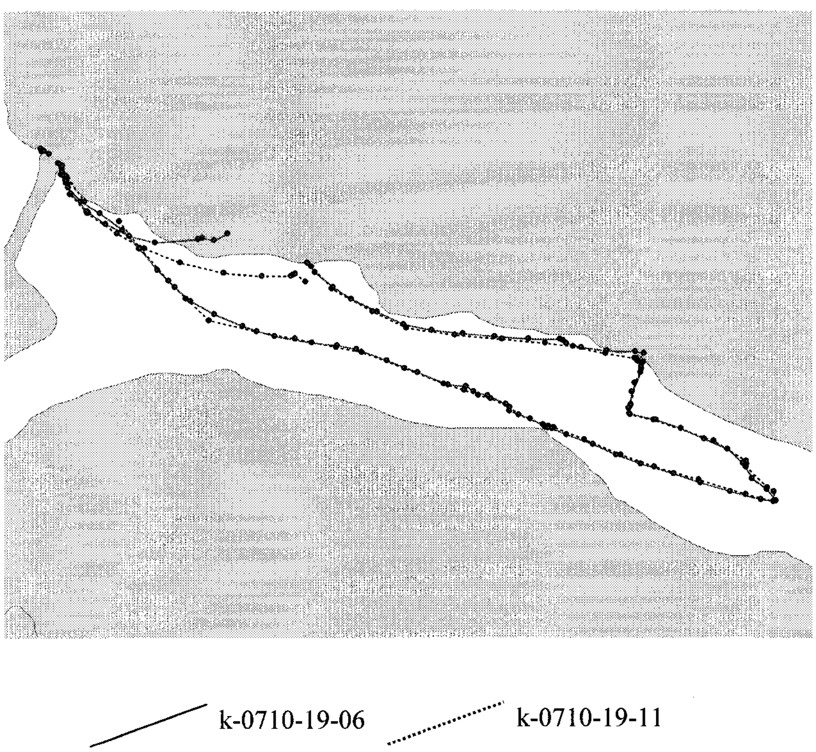


Figure 3-4 Accompanying Trajectories

The third operation begins with calculating the speed between each pair of successive points for each trajectory consisting of n points, yielding n-1 segment speeds. Because we are only interested in the movements of the boat, for every segment where the speed was less than 0.1 meter/second we deleted the second endpoint, since this was interpreted as a stopping or resting situation (the two points are not necessarily coincidental though, due to drifting). This threshold was established by calculating the mean speed using the original data, and then excluding the segment speeds less than 0.1 m/s, 0.3 m/s and 0.5 m/s respectively (Table 3-3). With the stops or near-stops removed, the average speed is obviously higher, but the values associated with different thresholds were not that different. As more points were removed, the standard deviation of speed for each boat type tightened but at the expense of fidelity to the original movement, which may affect other attributes, such as total distance travelled. Since there is no ideal resolution to this issue, the minimum threshold examined of 0.1 m/s was chosen to achieve the desired purpose with the least impact possible. Please note the data used are the trajectories collected in 2004 at NS.

Table 3-3 Different Thresholds for Eliminating Pauses in the Travel

	Original Data			**Segment Speed $\geq$ 0.1 m/s**			**Segment Speed $\geq$ 0.3 m/s**			**Segment Speed $\geq$ 0.5 m/s**		
	Mean Speed	Standard Deviation	Number of Points	Mean Speed	Standard Deviation	Number of Points	Mean Speed	Standard Deviation	Number of Points	Mean Speed	Standard Deviation	Number of Points
Canoe	0.501	0.450	53	0.757	0.418	49	0.898	0.373	42	0.970	0.344	37
Kayak	0.697	0.540	190	1.049	0.498	179	1.192	0.437	162	1.265	0.382	150
Motorboat	2.358	2.167	193	2.776	2.150	191	2.832	2.110	186	2.856	2.086	183
Sailboat	1.702	0.946	285	1.984	0.898	277	2.097	0.869	269	2.197	0.843	260

Conversely, we also deleted the endpoint for each segment having an extremely high speed. This situation arises due to inaccurate recording, as GPS units require at least 3 satellite signals for accurate positioning, but occasionally fewer than three signals are received, leading to erroneous results. Although there is no foolproof way to isolate these anomalies, an unreasonably high speed was deemed to correspond to a signal that jumped a large distance over an extremely short time span. A rough rule to define such an outlier in this study is any endpoint whose segment speed is above six standard deviations from the mean speed. In order to avoid the difficulty caused by outliers, they were deleted in the cleaning procedure.

Finally, there was a need to establish what constitutes a trip. For example, does a sailboat on a 3-day outing comprise a single trip for the purposes of pattern analysis? If a motorboat consistently returns to a dock to change passengers, should it be regarded as a single trajectory or multiple trips? Since there is no standard criterion, for this study MARIN defined a trip as an outing that is more or less continuous (i.e. no more than a 1-hour break) and normally involves returning to the origin (Pelot et al., 2004a). Although the definition is subjective, it is reasonable for such research, and this operating criterion is consistent with MARIN's prior studies. Therefore, we divided one original trajectory into several trips where the duration between two points was greater than one hour.

To illustrate, canoe trip c-0713-10-08 composed of 68 points was cleaned as follows. First of all, 2 points on land were deleted. There was no time gap larger than one hour in this case. The speed of each segment is shown in the following histogram (Figure 3-5), including 3 segments with speed <0.1 m/s to be eliminated, and one high-speed outlier, which was then also deleted.

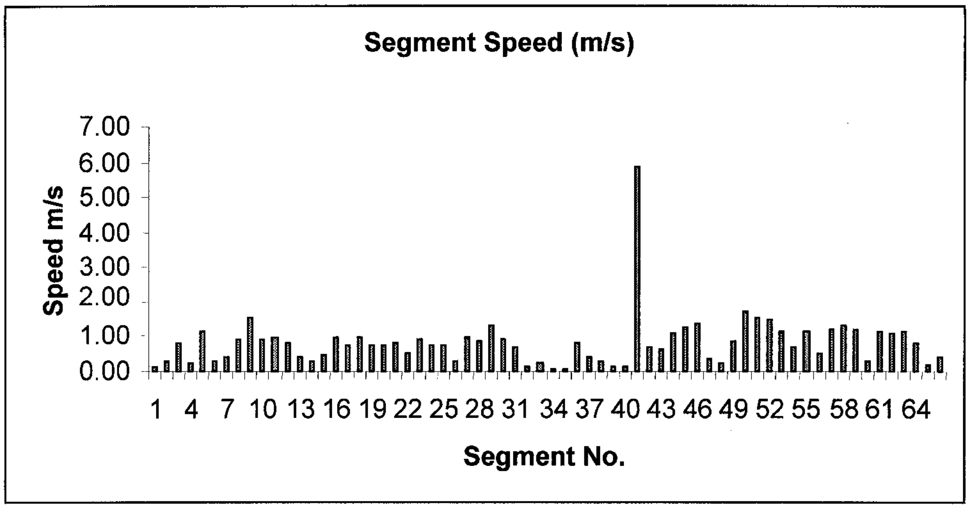


Figure 3-5 Segment speeds of a raw GPS trajectory

Table 3-4 is the summary of sample sizes of trajectories after all of the cleaning operations are completed. Table 3-5 is the detailed data for the sample size at different locations in different years. The subsets are used for specific aims at different stages of the study. For example, the initial 2004 HRM samples were applied for estimating the desirable sample size. The 2004 NS dataset was used to determine the threshold for eliminating pauses in the travel as explained above, and to develop algorithms elaborated in later chapters. After more data were obtained in 2005, the algorithms were validated using this subset, and comprehensive analyses were achieved. Despite a significant data collection effort, due to various reasons such as bad weather and people's willingness to participate, the number of trajectories are not ideal, especially in NB, but they still provide some indication about the spatial patterns of different boat types at different locations. More importantly, the data is useful to develop new recreational boating GPS trajectory pattern analysis methodologies, which is the principal contribution of this research.

Table 3-4 Sample Size of Number of Trajectories

Sample Size	**NB**	**NS**	**Combined**
Canoe	4	17	21
Kayak	8	21	29
Motorboat	15	10	25
Sailboat	5	47	52
Sum	32	95	127

Table 3-5 Sample Size Collected in 2004 & 2005

	2004			**2005**	
	NS		**NB**	**NS**	**NB**
	HRM	Outside HRM			
Canoe	16			1	4
Kayak	2	17		2	8
Motorboat	1		1	9	14
Sailboat	20			27	5

3.3 Information Extraction and Knowledge Attainment

Roger's famous saying is "we are drowning in information and starving for knowledge" (Borgelt and Kruse, 2002). If obtaining knowledge of spatial patterns is the ultimate goal, first of all it is necessary to extract and calculate trajectory attributes from the GPS data, and then statistical analysis can be employed to discover what differences exist between categories of boats with regards to attributes extracted from the trajectories. Characterizing boats can be based on movement trajectory in a static view, but velocity is also another important factor for classifying the type of vessel. It is obvious that canoes and kayaks are relatively slow, while sailboats and motorboats are relatively fast. The velocity factor can perhaps be best used to exclude some boat types, for instance high speed precludes non-motorized vessels. From the literature review of trajectories analysis, it was found that speed, turning angle, and total distance travelled were typically selected. Moreover, by visually observations for the trajectories, there are some differences between bounding box and furthest distance from shore. Thus, they are all selected as the attributes to be extracted and calculated. Among them, the speed and turning angle are dynamic, while the rest are static. The algorithms used for extracting all of these attributes are derived in Chapter 4; hence, this section is simply an overview of the content. If no differences among the four boat types could be detected according to these attributes, other characteristics which would give more information, such as distance from the launch point, might have to be included.

Progressively, from the lowest comprehension level based on GPS data, to the intermediate information level involving boat movement features, to arrive the highest comprehension level of spatial knowledge of boat patterns, the procedures for this study can be shown in Figure 3-6.

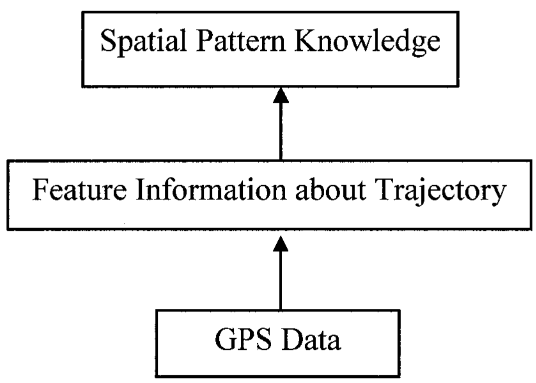
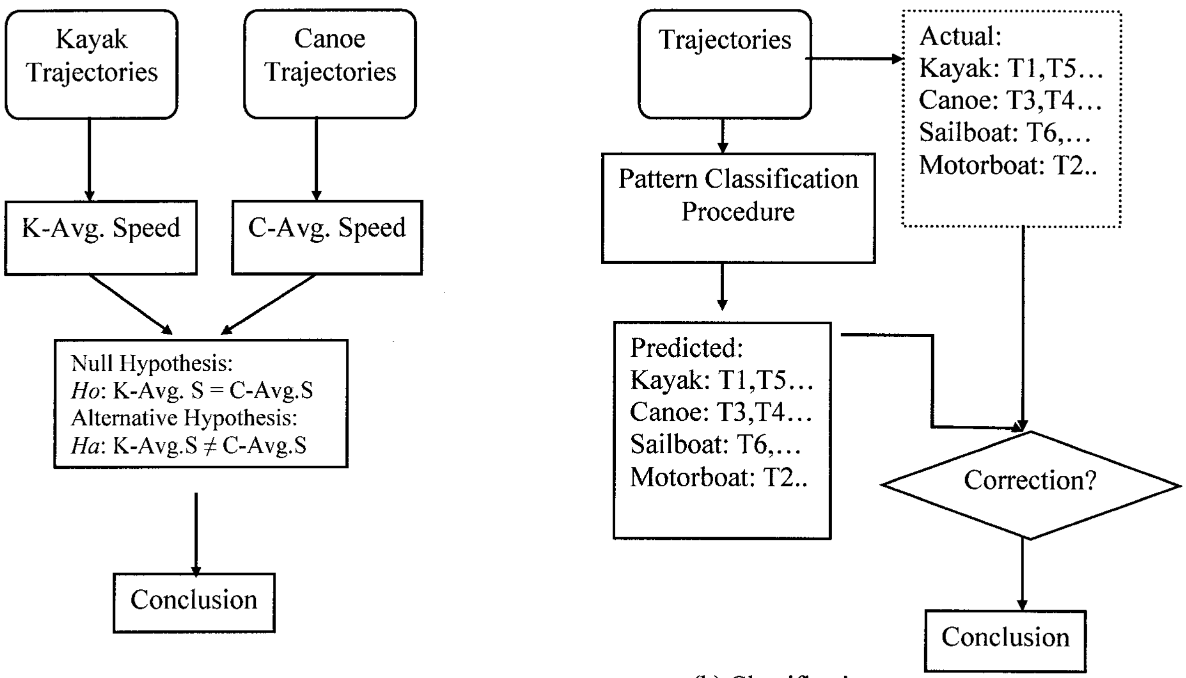


Figure 3-6 Three Levels of Comprehension

As illustrated in section 3.1 and 3.2, the lowest level GPS data could be acquired, and the second level involving boating feature information could be determined using the approaches which will be explained in detail in Chapter 4. The principal aim of this study is to acquire knowledge about the classification of recreational boats based solely on trajectory movements. First, we must establish whether different boats have different patterns. If there is no distinction among them, classification cannot be achieved since they would then all belong to the same class according to their movement characteristics. That is why this project has two specific goals which are clearly listed in the objective section: to check whether different boat types have significantly different behaviors with regards to attained attributes, and to identify the class to which a boat belongs based on its attributes. The possibility of discrimination leads to the possibility of classification.

The concepts of discrimination and classification are exemplified in Figure 3-7. The most significant difference between these two methods is whether the boat types are known *a priori*. Known boat types are necessary to perform discrimination analysis. For instance, sample trajectories of canoes and kayaks are obtained and the boat type for each trajectory is known, therefore boating features such as the average speed can be extracted from these sample trajectories and tested to arrive at a conclusion whether canoes and kayaks can be differentiated based on their average speed, as shown in the flowchart in Figure 3-7(a). Likewise, conclusions based on other attributes can be obtained.

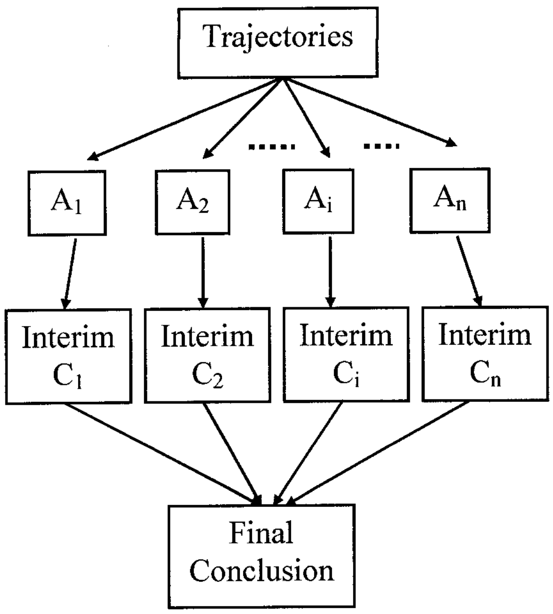
Conversely, the classification process shown in Figure 3-7 (b) assumes unknown boat types, using certain pattern classification procedures to predict a type, and then compares to the actual boat type to examine if the classification method works well. If the conclusion is bad, in other words if the classification method misclassified most of the samples, there are two possibilities. One is that the classification method is not ideal, and the other is that there is no significant difference among the samples, so a good classification scheme cannot be determined. In this study, since the discrimination procedure is applied first, if no significant distinctions between boat types are found, the classification won't work. Consequently, a bad performance of the classification procedure would be attributed to an unfit pattern classification procedure.



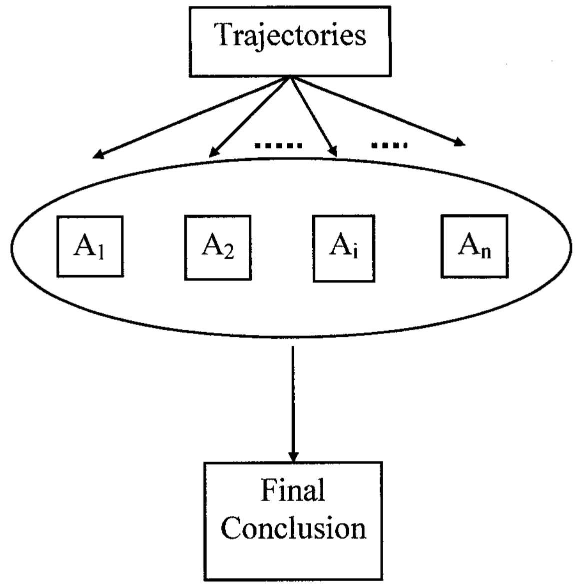
(a) Discrimination (b) Classification

Figure 3-7 Discrimination Process versus Classification Process

In order to accomplish the discrimination and classification analyses, there are two different processes which could be followed. One is univariate analysis as shown in Figure 3-8(a). As the name indicated, the analysis depends on every single attribute from A_1 to A_n, which are extracted from the sample trajectories. A final conclusion is derived from every single interim conclusion. Conversely, multivariate analysis considers all of the attributes together, and then arrives at a final conclusion directly as Figure 3-8(b) demonstrates.



(a) Univariate Analysis



(b) Multivariate Analysis

Figure 3-8 Univariate and Multivariate Analyses

Specific methods of univariate analysis and multivariate analysis for the purpose of discrimination and classification will be expounded in Chapter 6 in their respective sections, making the context and content of these procedures more clear.

4. Attributes Extraction and Inferred Variables

In this chapter, the methods for extracting and calculating attributes will be explained. Some of these attributes have been collected through surveys. The results from the GPS data will be compared to those from MARIN surveys, to demonstrate that the GPS data can give much more precise information. Moreover, surveys cannot yield estimates for some important trajectory attributes such as bounding box, turning angle and segment length, which only can be calculated from GPS data.

4.1 Speed

As reviewed in section 2.3 on Trajectory Analysis, almost every previous study applied a speed attribute for its specific goal. Of course, speed is an important attribute of recreational boating trajectories for pattern analysis. Presumably, the type of recreational boat can partially be differentiated according its speed, or at least limited to particular categories. For example, if the speed is very fast, at least we can be sure that it is not a canoe or kayak. In order to check this proposition, speed variables for different types of recreational boats are analyzed to ascertain whether or not the speed is significantly different across different boat types. If true, the results could be valuable for remote detection of the type of recreational boats in the coastal security realm.

Mean speed is the average of all segment speeds derived from a trajectory that has been cleaned. Maximum speed is the fastest segment speed. Because extreme values such as maximum speed are generally unstable for analysis purposes, and given the accuracy limitations of GPS units, we also calculate the "maximum 1/20 speed", as defined below:

Step1: find the fastest 5% segment speeds

Step2: if the number in step 1 is less than 3,

Then find the 3 fastest segment speeds,

Else skip to step 3.

Step3: $max_{1/20}$ speed = mean speed of selected fastest segment speeds.

According to the kayak club at the University of Washington, beginners are usually somewhat slower, paddling at 1.03 to 1.29 meter per second; moderate kayak boaters typically travel at 1.54 meters per second; and experienced boaters can reach 2.51-2.57 meters per second (The University Kayak Club, 2004). According to our GPS points, the mean speed of kayaking in coastal areas is 1.09 m/s, and 1.01 m/s in the river; the max speed is 2.13 m/s in coastal waters, and 2.19 m/s in the river; and finally, the $max_{1/20}$ speed is 1.88 m/s in coastal areas and 1.83 m/s in the river. All of these speed data extracted from the GPS boating trajectories are in general agreement with the above cited expert opinion.

Table 4-1 shows the *p*-value of the paired t-test between max speed and $max_{1/20}$ speed for each boat type in both study areas. The null hypothesis H_0 is that the means are same. Setting the critical significance level at 0.10, only kayaks in NB did not show significant difference between these two attributes. Although the two attributes are significantly different, we will nevertheless use $max_{1/20}$ speed for proposed analysis because it is more reliable than a single max speed value.

Table 4-1 Paired t-test results for max speed and max 1/20 speed

	NB					NS				
P-value	Canoe	Kayak	Motorboat	Sailboat	All	Canoe	Kayak	Motorboat	Sailboat	All
	0.059	0.11	0.068	0.052	0.0041	0.0044	0	0.091	0	0

Although speed is a useful attribute, surveys cannot provide an exact mean speed or maximum speed. Most questionnaires ask respondents to select from pre-defined speed ranges, thus it is a rough estimation. For example, the MARIN phone survey (Pelot et al., 2004b) reported that 46 percent of respondents thought their usual boating speed was under 10 km/h (2.78m/s), and almost one quarter of respondents (24%) said they maintain speeds between 10-25 km/h (2.78m/s-6.94m/s). The overall mean comes to 18.8 km/h (5.22m/s) when considering the ranges of speed values reported. Comparing to the mean speed extracted accurately from GPS trajectory points (Table 4-2), only motorboats' speed falls in the second range in Figure 4-1 (i.e. 2.78m/s-6.94 m/s), while all the others fall into in the first range, which is less than 2.78 m/s. These results are reasonable since motorboats can reach higher speeds. However, the data resulting from the survey are obviously higher than that calculated from the GPS data, perhaps because the survey included more than the four boat types evaluated in this study, many of which were fast moving types of boat. Therefore, the boaters' mean speed estimates were somewhat higher. Moreover, the speed from GPS is calculated based on the whole trajectory including all the segment speeds (except for those less than 0.1m/s), while survey data would likely be based on times when the boaters were moving quickly.

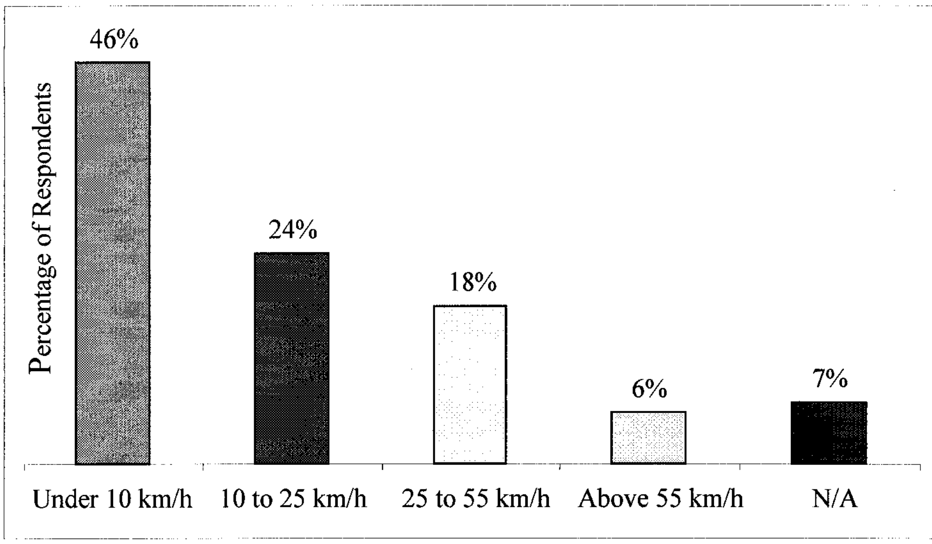


Figure 4-1 Speed Travelled according to Phone Survey

Table 4-2 Mean Speed Extracted from GPS Data

	Canoe (m/s)	Kayak (m/s)	Motorboat (m/s)	Sailboat (m/s)	Total (m/s)
Coast	0.88	1.09	3.98	2.06	1.83
River	1.17	1.01	3.78	1.92	2.47

4.2 Total Distance

Total trip distance is calculated based on the cleaned trajectories as was done with speed. This attribute exhibits a very large range. The shortest travel distance was only 355 meters by a canoe in NS, while the longest one was 87,648 meters by a motorboat in NB. If we establish 5000 meter categories, with all total travel distances above 45,000 meters assigned to the highest category 10 (Table 4-3), it is found that the most common range is category 1, with 36.2% of the vessels travelling within that range, with almost three quarters (74%) of the vessels falling into the first three ranges.

Table 4-3 Categories of Total Distance

Category	Lower Bound (m)	Upper Bound (m)	Proportion
1	0	5,000	0.362
2	5,000	10,000	0.260
3	10,000	15,000	0.118
4	15,000	20,000	0.071
5	20,000	25,000	0.063
6	25,000	30,000	0.032
7	30,000	35,000	0.016
8	35,000	40,000	0.024
9	40,000	45,000	0.024
10	45,000	infinity	0.032

It is very difficult to ask survey respondents to estimate the total distance they travelled. Considering this, MARIN did not ask this question during the phone survey, but MARIN did ask how fast the respondent usually travelled in the most often used boat and how long s/he typically spent on the water each trip. The overall mean speed came to 18.8 km/h (5.22m/s) when considering the ranges of speed values reported, and the overall mean duration was estimated at 4.4 hours for this survey (Pelot et al., 2004b). Roughly, the mean distance for a typical trip would be 83,000 (18.8km/h*4.4h) meters, which is only a little bit shorter than the longest trip recorded by GPS. One reason for this is that the estimations were from all kinds of boats which were included in the survey. The other is that the calculation from the survey to get the attribute of total distance travelled is not ideal. Another survey question which can provide coarse information about total distance travelled is to ask the respondent the launch point and the destination point, and then calculate the distance between these two locations without considering the sinuosity of recreational boating, which would likely produce an estimate not even close to the precise distance travelled. GPS can give us much more accurate information: the actual distance that the boat travelled, not just an estimated range.

4.3 Bounding Box

Motivated by the observations that different patterns of recreational boating movement would lead to different bounding box sizes and shapes, the use of this analytical tool was adopted, as has been done in previous research (e.g. Erol and Kossentini, 2003). A bounding box as defined in this study is the smallest rectangle which encompasses a whole boating trajectory.

In order to determine a bounding box, first of all it is necessary to locate the two points on the trajectory which are separated by the longest distance. For example, sailboat trajectory s-0623-18-09 (Figure 4-2) has the longest distance from Point 4 to Point 192. Note that the longest distance does not necessarily include the launch point. Then finding the maximum distance to this straight line laterally in each direction gives point 224 with distance w1 on one side, and point 182 with distance w2 on the other. Finally, the bounding box with length L and width W (= w1 + w2) is generated.

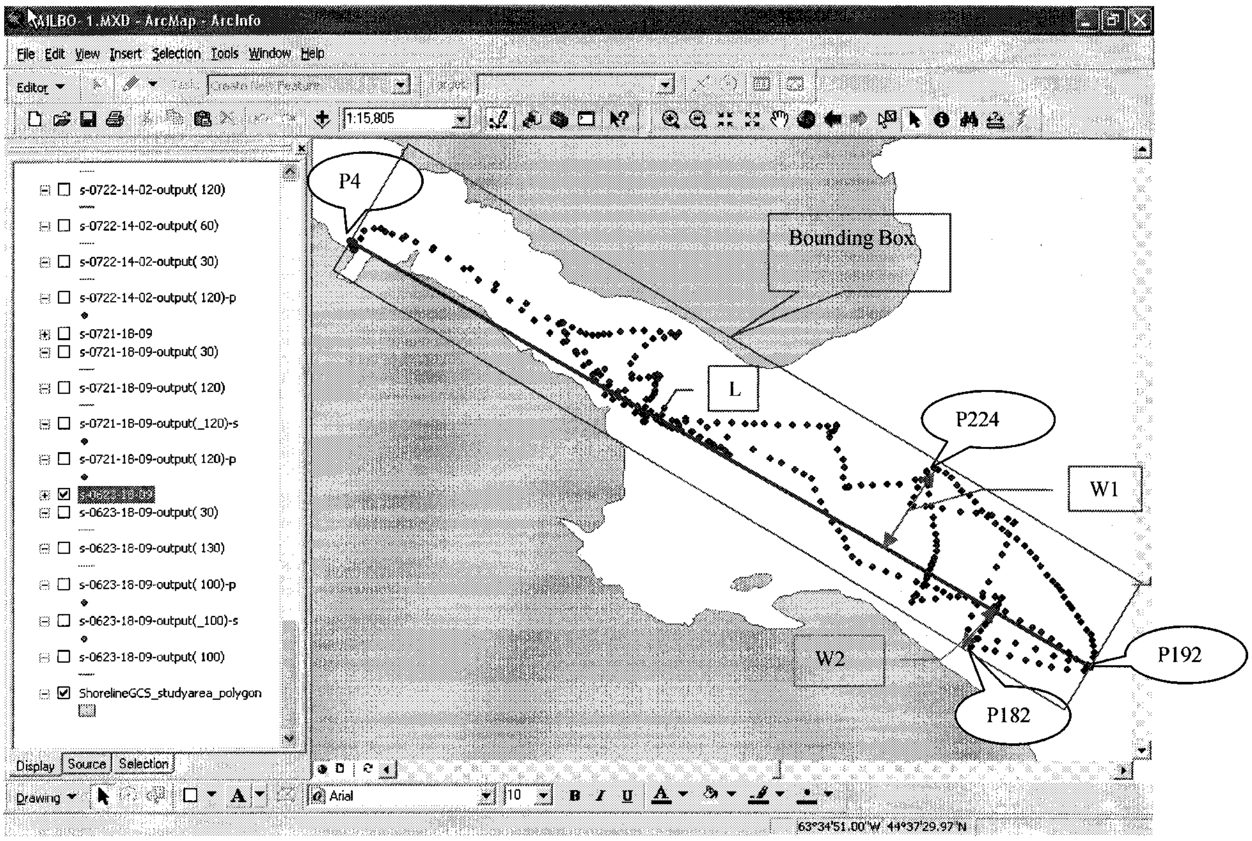
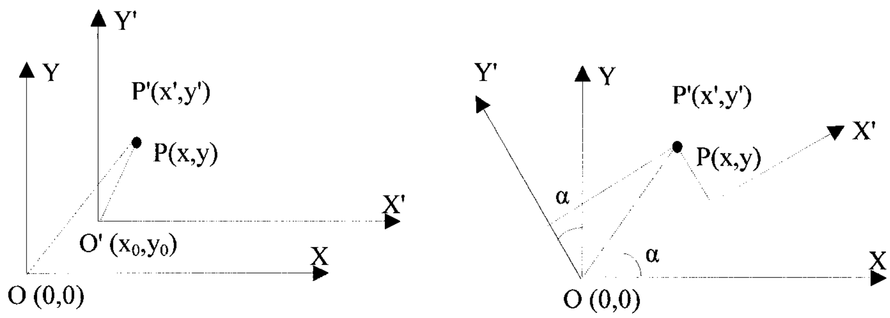


Figure 4-2 Bounding Box

The principle of establishing the bounding box is simple, but calculating the lateral distances involves coordinate transformations. One is a shift transformation as shown in Figure 4-3(a), and the other is a rotation transformation demonstrated in Figure 4-3(b).



(a) Shift Transformation (b) Rotation Transformation

Figure 4-3 Coordinate transformations

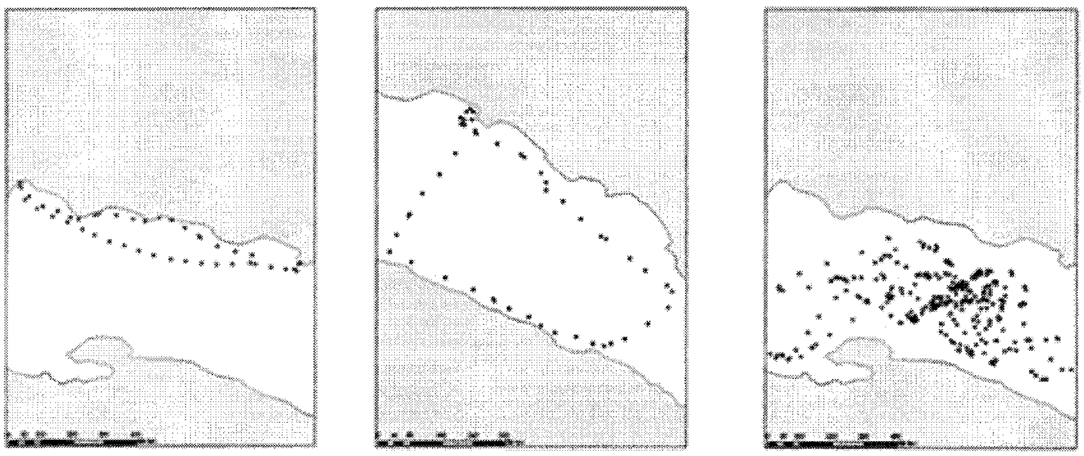
The shift transformation moves the original point O (0,0) to O′ (x_0,y_0). So the point P (x,y) in the original coordinates will shift to the new position in the new shifted coordinates P′(x′,y′), where $x'=x-x_0$, $y'=y-y_0$. If rotating the original coordinate system α degrees, P(x,y) rotates to P′(x′,y′) as displayed in Figure 4-3(b), and

$$x' = \cos\alpha * x + \sin\alpha * y;$$

$$y' = -\sin\alpha * x + \cos\alpha * y.$$

Applying these coordinate transformation operations, the coordinate origin point should be placed on the left point of the longest distance straight line, in the example (Figure 4-2) that would be the point p_4. Then by rotating the coordinates, the x-axis becomes coincident with the longest distance straight line. The next step is to find the highest point above the x-axis (p_{224}) and lowest point (p_{182}) below it. The longest distance will be the length of the bounding box and the sum of the respective distances of the highest and lowest points from the long axis centerline will be the width of the bounding box.

Given the length and width of the bounding box, we can infer the following variable: aspect ratio, which equals the width divided by the length of the bounding box. This attribute expresses whether the travel trajectory is narrow or wide. For example, in Figure 4-4 the aspect ratio of canoe trajectory c-0707-08-03 is 0.18, while that of c-0721-11-04 is 0.88. It is obvious that the second path was broader compared to the first one, which remained close to shore.



(a) c-0707-08-03 (b) c-0721-11-04 (c) s-0722-14-10

Figure 4-4 Boating Trajectories

Considering the bounding box and total distance metrics together, another variable called coverage index can be derived from the perimeter of bounding box divided by the total distance travelled. The coverage index of the sailboat trajectory shown in Figure 4-4 (c) is 0.2, while that of the canoe trajectories in Figure 4-4 (a) and (b) is 1.18 and 1.42 respectively. This attribute represents whether the trajectory is complicated, including lots of back-and-forth movements. The smaller the variable's value is, the more complex the trajectory. These attributes cannot be assessed through a survey, while GPS data provides detailed spatial and temporal boating information.

Although Dicke and Burrough (1988) and Wiens et al. (1995) used fractal dimensions for characterizing tortuosity of animal trails, fractal dimension analysis is not suitable for characterizing complexity for boating trajectories. An important characteristic of fractals is that of self-similarity, which means when examined at increasingly larger resolution, increasing amounts of details are resolved that are scaled versions of the variation seem at lower resolution (Mandelbrot, 1977; Mandelbrot, 1983; Voss, 1984). However, for the kinds of variation found in natural phenomena, exact geometric self-similarity is unlikely (Dicke and Burrough, 1988). Li et al. (2005) also deemed that fractal dimension could not be considered as a fixed value (i.e. D in Equation [2-1] is not common to the two measure scales). Therefore, our study proposes the concept of a coverage index to represent the tortuosity of boating trajectories. Moreover, another advantage of the coverage index variable is its ease of calculation.

4.4 Furthest Distance from Shore

Aside from the aspect ratio and coverage index differences among the trajectories illustrated in Figure 4-4, it is also apparent that the first canoe trajectory (Figure 4-4(a)) travelled near the shoreline, while the second one (Figure 4-4(b)) distanced itself, effectively crossing the channel. Including the furthest distance from shore as a variable can represent this potentially important feature. First of all, let's look at a point's distance from shore instead of a whole trajectory's furthest distance from shore, since the trajectory consists of points. If we can evaluate the distance from shore of a point, then we can easily work out the furthest distance from shore of a whole trajectory. Supposing the distance from shore of point p in Figure 4-5 is r. To calculate this, a bisection method is applied. Initially we set up a lower bound distance r_l equal to 1 meter, which made the calculation very precise, and an upper bound distance r_2 equal to 10,000 meters, because more than three quarters of respondents of a MARIN recreational boating phone survey stay within 10 kilometers of the shoreline (Pelot et al., 2004b). Then a circle is constructed whose radius r equals $(r_2-r_1)/2+r_1$ to determine whether any point on this circle intersects land. If so, the upper bound would be replaced by r, otherwise the lower bound would be replaced by r. Consequently, a new circle can be calculated using the same function $(r_2-r_1)/2+r_1$ and checked according to the same condition. These procedures can be summarized as follows: calculate the radius for a circle, check for intersection of circle and land, and replace either the upper or lower bound for calculating the new radius, — continues until the distance of (r_2-r_1) is less than a pre-defined error ε, in this case set to 1 meter. Therefore, the distance from shore of that point p is $(r_2-r_1)/2+r_1$. Figure 4-6 shows the flowchart for calculating the distance from shore for a point. Applying this algorithm iteratively to all the points of one GPS trajectory, the largest value represents the furthest distance from shore for that trajectory. Once again, although such information has also been solicited through surveys, it is highly inaccurate compared with the GPS-generated results.

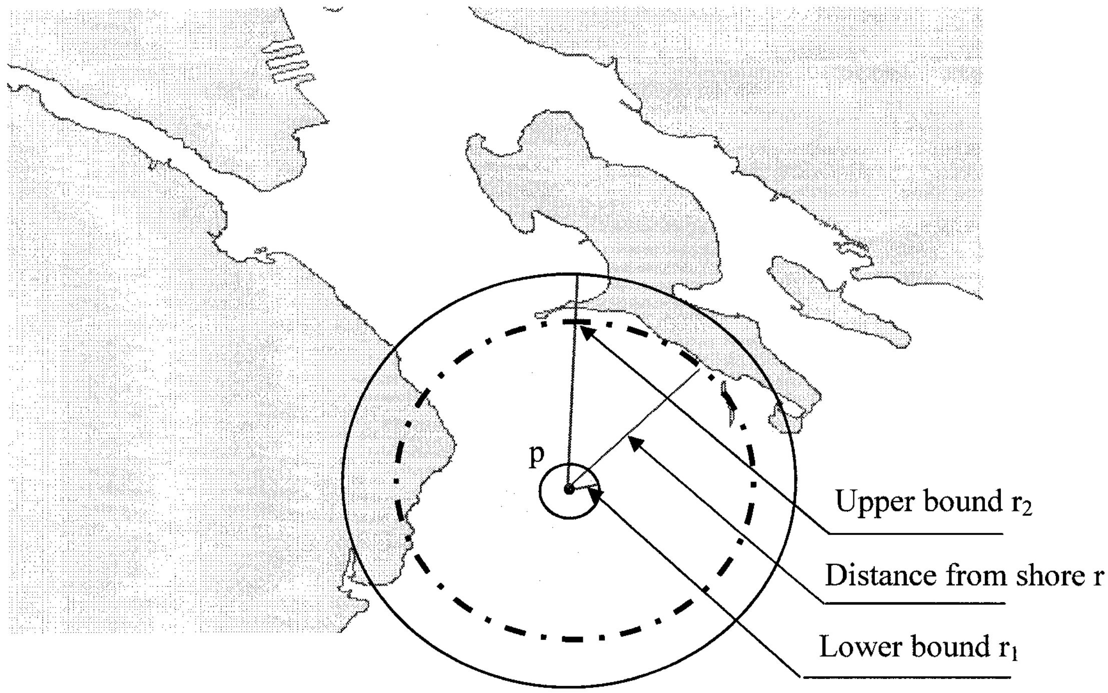


Figure 4-5 Distance from Shore of a Single Point

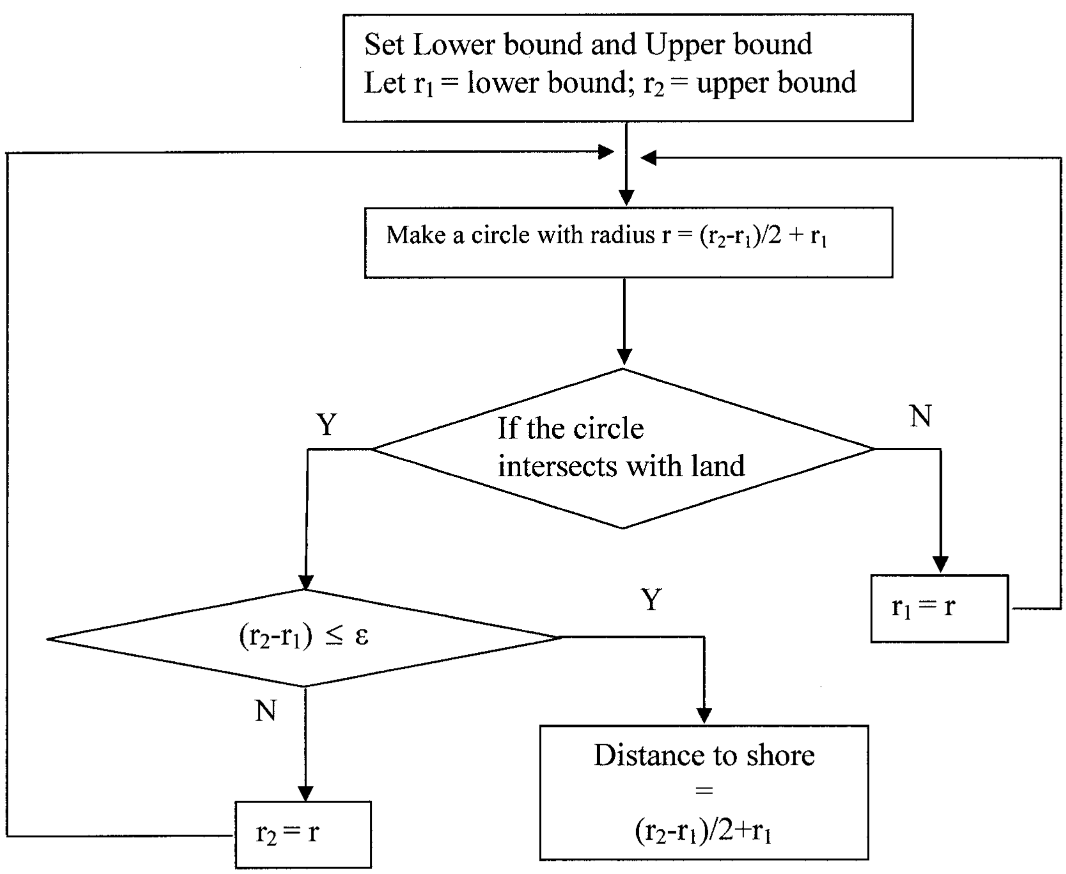


Figure 4-6 Flowchart of Distance from Shore Algorithm

4.5 Turning Angle

Figure 4-7 displays four characteristic boating trajectories for different boat types in Halifax's Northwest Arm. It is easy to observe some differences between these different boat types' trajectories visually. Based on the proposed algorithms to precisely calculate variables such as total distance travelled, coverage index, aspect ratio and furthest distance from shore, statistical tests can be conducted to quantify these pattern distinctions. It can also be observed that there is a big difference in the nature of turning angles among these trajectories. The turning angles of sailboats are unique (Figure 4-7(d)), which is a characteristic of sailboats. Sailboats cannot travel directly into the wind, but employ a sailing technique known as "tacking" to zigzag across a headwind. Tacking allows the boat to use prevailing wind from many angles to travel forward. For example a boat travelled for a time at an angle toward its desired course to the right, and then the boaters swung the boom of the sail to the other side and tacked across the desired course at an angle to the left, which results in a zigzag motion (West, 1994). This behavior can be valuable for discrimination and classification boat types, especially sailboats.

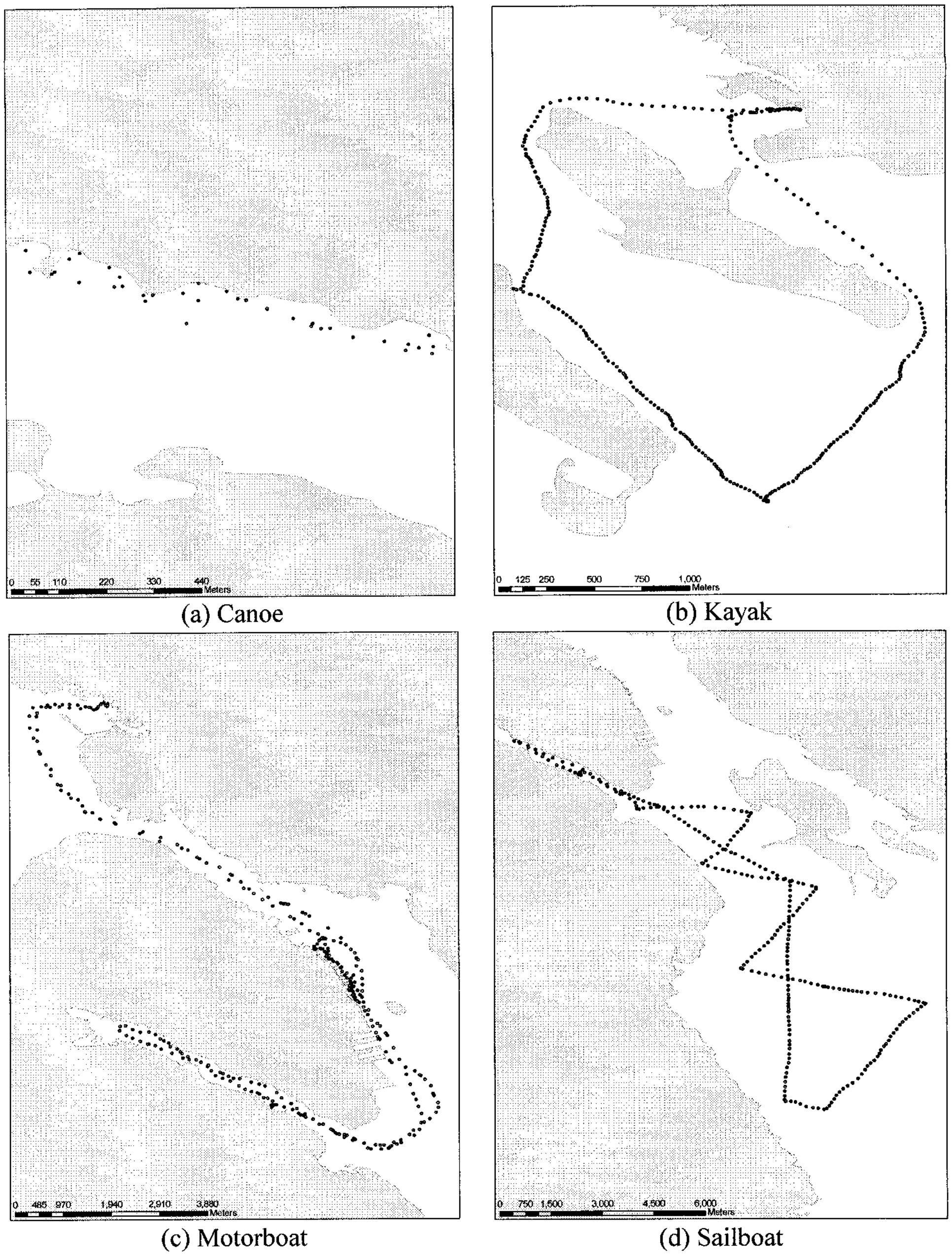


Figure 4-7 Characteristic Boating Trajectories of Different Boat Types

The number of points in a single raw GPS trajectory can be quite extensive, automatically generating a reading every few seconds depending on a combined function of distance, time and turning angle. For instance Figure 4-7(d) consisted of 275 GPS points. Not all of these points are significant for the analysis however. The data associated with the turning angles of the boat is one of the key factors for quantifying movement characteristics. However, since many readings occur when the boat is not turning, retaining these points not only makes calculations tedious, but these records will then skew the results in establishing the fundamental patterns of movements. In other words, including all these instances with "zero" turning angle masks the information from actual turns. For example, in the sailboat trajectory in Figure 4-7(d), we only want to include the turning angle at some zigzag turning locations, not the "turning angle" subtended at every set of three original successive points. The points that are not absolutely necessary in depicting original tracks are referred to as "noise". As shown in Figure 4-8, which is a short segment of the sailboat trajectory in Figure 4-7(d), one might choose the line constructed by only the dark points as a path representing the original trajectory for a specified accuracy. In this case, the other (white) points constitute noise. As one might observe, only 4 out of 14 original points are kept; however the 4 retained points can still correctly represent the principal characteristics of the initial trajectory. Once this "noise" has been removed, the track is "dedensified". This process is referred to as recreational boating trajectory dedensification.

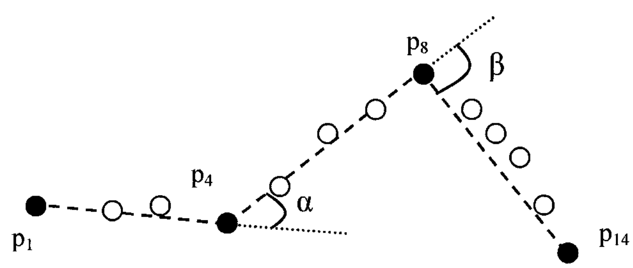
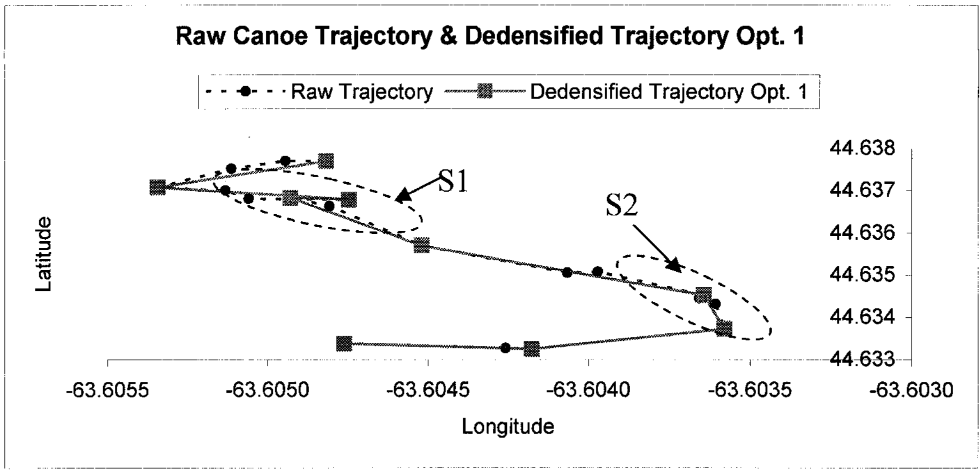


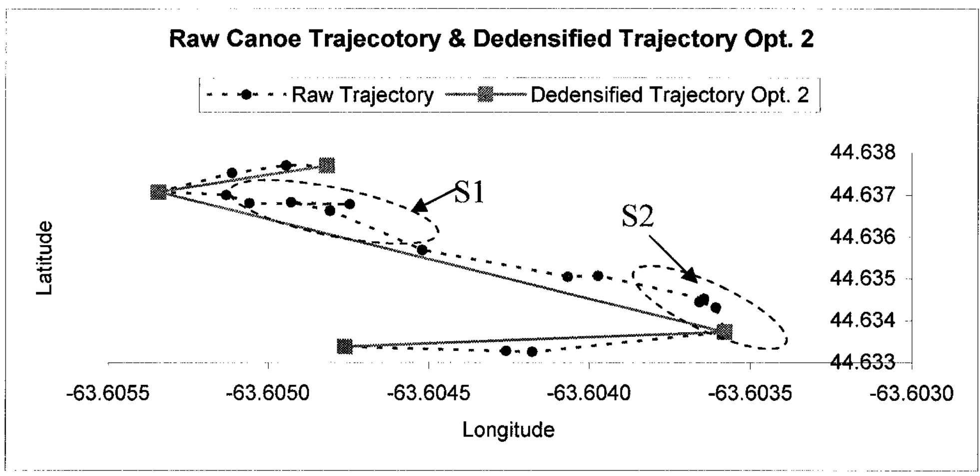
Figure 4-8 Sample of Dedensifying a Trajectory

The distance travelled by the boat before changing direction, called segment length (i.e. line p_1p_4, p_4p_8 and p_8p_{14} in Figure 4-8), and the angle at which it made the turn, relative to its bearing, called the turning angle (i.e. angles α and β in Figure 4-8), are both useful features. In this study, a "feature" is defined as any element that should remain noticeable in the dedensified trajectory due to its importance in representing the characteristics of the trajectory.

Different trajectories exhibit different features, and a simplified trajectory retaining the wrong features will fail to reflect the fundamental features of the vessel's movements. An example of a canoe trajectory is shown in Figure 4-9. The raw GPS trajectory is relatively simple, comprising only 19 points, depicted by the small circles connected by dashed lines in both Figure 4-9(a) and Figure 4-9(b). There are also two dedensified trajectories shown in Figure 4-9(a) and Figure 4-9(b) respectively, depicted by the solid lines and square symbols. Option 1 (Figure 4-9(a)) retains 9 points from the original track, while option 2 (Figure 4-9(b)) retains only 4. The detail dedensification algorithm will be elaborated separately in next chapter (Chapter 5) due to its importance and complexity.



(a) Option 1



(b) Option 2

Figure 4-9 Raw Trajectory and Dedensified Trajectories

It is noticeable that the features extracted from these three trajectories (original plus 2 dedensified) are quite different. The results of statistics conducted on both the segment lengths and turning angles are presented in Table 4-4. The differences are quite noticeable. The mean segment length of the original trajectory is 37 meters, while that of the dedensified trajectory option 1 is 80 meters, and 193 meters calculated from the dedensified trajectory of option 2. Similar large variations appear with respect to the maximum segment length: that of the dedensified trajectory option 1 is nearly one and a half times longer than that of the original trajectory, and that of the dedensified trajectory option 2 is almost 4 times longer than that of the original trajectory. As to the turning angle, the mean turning angle derived from original trajectory is 57°, 75° for the dedensified trajectory option 1, and 70° for the dedensified trajectory option 2. Similar disparities are noted for all the other statistics of segment length, as well as turning angles, summarized in Table 4-4. As would be expected, the segment lengths are longer in the dedensified trajectories compared with the original path. As more points are eliminated, the lengths of some segments increase. The turning angles are also altered during this dedensification process.

Table 4-4 Statistics for Different Trajectories

	Segment Length (meter)			Turning Angle (degree)		
	Raw Trajectory	Dedensified Opt. 1	Dedensified Opt. 2	Raw Trajectory	Dedensified Opt. 1	Dedensified Opt. 2
Mean	36.74	79.71	192.71	56.35	74.56	69.86
Median	24.09	75.60	101.50	49.57	65.68	69.86
Std. Dev.	29.68	42.77	177.19	45.69	54.00	25.20
Min	6.89	15.26	79.71	2.44	13.79	52.04
Max	106.22	146.71	396.93	168.12	163.51	87.69

Enlightened by these examples, two things are realized. One is that dedensification is necessary, and the other is that choosing the dedensified trajectory for a proposed calculation is crucial to the analysis. Because of its importance, the dedensification algorithm and selection approach for recreational boating trajectories will be elucidated separately in the following chapter.

5. Recreational Boating Trajectory Dedensification

Dedensifying recreational boating trajectories is a novel idea, whereas the concept of simplifying coastlines, rivers or roadways is a common practice in cartography. Since the 1970's, a large amount of research has been conducted in this area (Ramer, 1972; Douglas and Peucker, 1973; Marino, 1979; McMaster, 1987; Cromley and Campbell, 1990; Li and Openshaw, 1992). The aims of line simplification underlying those practices are mainly saving computer memory space and expediting computer processing. Little and Gu (2001) used the same concept to smooth trajectories in sports matches such as hockey or tennis. However, their aim differs from the cartographers, hence the algorithms are tailored to get rid of "noise" points except for the corners and high curvature points, to aid with object recognition. Similarly, our aim for recreational boating trajectory dedensification is to keep the main movement features. There have been several approaches and algorithms advanced to reduce the number of points necessary to represent numerically recorded lines. The methods can be placed into broad categories of: (1) elimination of points along the line by one or more criteria, (2) the approximation of the line with a mathematical function, and (3) the deletion of specific (cartographic) features represented by the line (Douglas and Peucker, 1973). Since much research has been conducted in this field as indicated above, the varied nomenclature includes line simplification, smoothing, and line generation. In this study, the term dedensification is introduced, because although some line simplification may result, one of the key aims is to remove intermediate "redundant" points, which in some cases does not result in a simplified line; simply a less dense point representation.

The process of fractal dimension analysis introduced in section 2.3 can be used to attenuate information including reducing the number of points, leading to line generation (Longley and Batty, 1989). However, because of the fixed yardstick length, it cannot guarantee that the retained points are feature points. For example, neither generated line in Figures 2-2(a), (b) or (c) retained feature points. A reasonable dedensification line should be like the line shown in Figure 5-1, which is composed of segment lines with different lengths. The fractal line generation idea is very similar to a simpler idea: retain every n^{th} point. One is a fixed length criterion; the other is a fixed number of points criterion. Despite the simplicity of these ideas, none of them are suitable for this study because only the feature points where the boat actually turned should be retained. In this section, the algorithm applied for dedensifying boating trajectories will be presented, starting with the introduction of the basic Douglas-Peucker algorithm.

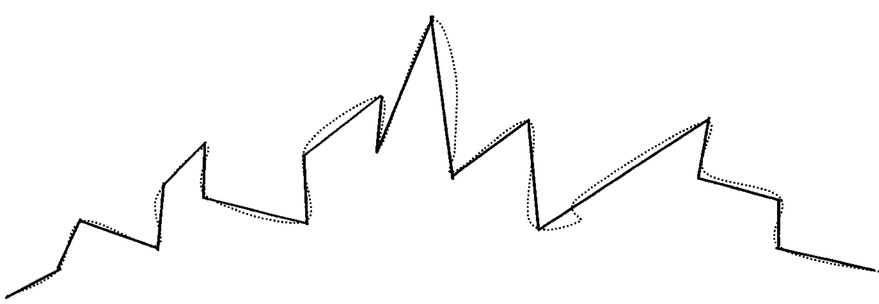


Figure 5-1 A Possible Dedensified Line with Retained Feature Points

5.1 Douglas-Peucker Dedensification Algorithm

The well-known Douglas-Peucker Algorithm (DPA) is associated with numerous early attempts to simplify cartographic feature. This algorithm continues to be widely used as an effective program that focuses on line filtration and simplification (Visvalingam and Whyatt, 1991). The procedure can be illustrated by examining the simple situation comprising a minimum of three points shown in Figures 5-2(a) and (b).

The major difference between Figure 5-2(a) and Figure 5-2(b) is that there is a compulsion to keep the point P_2 in Figure 5-2(a), while there is assumed to be no such need in Figure 5-2(b). The perpendicular distance between point P_2 and the line segment P_1P_3 suggests that a distance criterion could be established to identify feature points (i.e. those whose elimination would substantially change the remaining path), for example the turning angles.

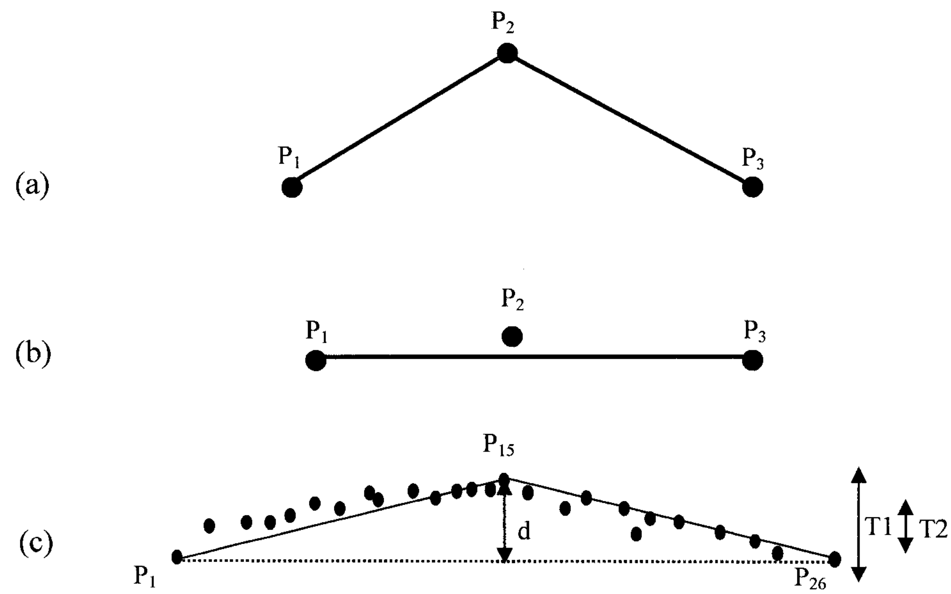


Figure 5-2 Criteria for Point Expulsion along a Trajectory

The first step in DPA is to compare the largest perpendicular distance (d) between any point and the subtended baseline (P_1P_{26}), to a pre-specified distance tolerance (Figure 5-2(c)). In this example, if T_1 is the tolerance, no point on the trajectory is further than the tolerance distance T_1 from the subtended baseline P_1P_{26}, and the baseline (dashed line) is deemed sufficient to represent the original trajectory. If T_2 is used instead for the tolerance, then furthest point (P_{15}) from the baseline exceeds it, and becomes the vertex of new line segments more closely following the original trajectory. For subsequent steps, the same procedure is iteratively carried out with the new segments, and so on until all points are within T_2 of the closest segment of the final path (Douglas and Peucker, 1973). The basic steps of the recursive DPA are listed as follows with the corresponding illustrative diagram in Figure 5-3:

1) Find C furthest point from AB, at distance d from the line

2) If d < deviation tolerance, include AB in the simplified line and eliminate point C (Shown in Figure 5-3 (a))

3) Otherwise split the line at C and evaluate recursively (Shown in Figure 5-3 (b))

Initially A and B are the ends of the original line.

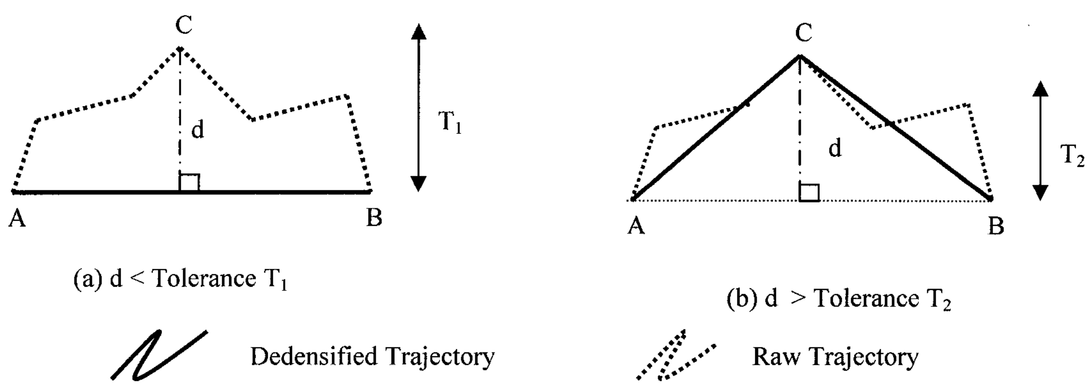


Figure 5-3 Typical DP Algorithm

This algorithm concentrates on choosing the relevant points along the line, preserving these, while removing all other superfluous points. DPA connects the two ends of the original line as the first baseline and chooses the furthest point from the straight segment (Douglas and Peucker, 1973), which makes the algorithm very efficient.

The line generation algorithm (Chang et al., 1991) used in Little and Gu's work (2001) is based on calculating the distance of a point from the straight line connecting its two neighboring points. However, it is not applied in our study, since it is very time-consuming when the number of points is large (i.e. hundreds or more), as in our case. There is another reason why Gu's algorithm could not be employed in our case, as illustrated in Figure 5-4, metaphorically referred to as a myopia problem. A long and flat track of points implies that the distance between a point and its neighbor is very small, leads to discarding all the points. However, when considering the overall trajectory, the furthest distance d from the line connecting the two ends of the track might too large to ignore. Thus at least one point should be retained to represent this general deviation from a straight path. Since many recreational boating trajectories include such elongated track sections, the DPA would be a better choice.

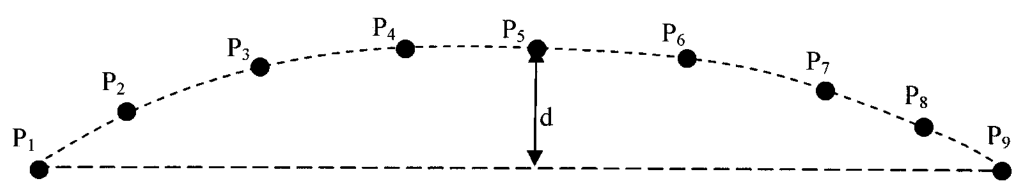


Figure 5-4 Myopia Problem

5.2 MARIN Douglas-Peucker Algorithm

Despite its popularity and simplicity, there are two problems while applying DPA to dedensify recreational boating trajectories. In this section, we will state the problems and the suggested solutions.

5.2.1 The Loss of Important Features

5.2.1.1 Problem Statement

While the DPA process for selecting points using a predefined tolerance provides reasonably good results, the algorithm may also eliminate information that is important for characterizing a recreational boat trajectory. Similar issues were discussed in other work (Vaughn et al., 1991; Visvalingam and Whyatt, 1991). Switchbacks, or movements in which a trajectory demonstrates a series of diagonal zigzag repeats, are quite commonly yet incorrectly eliminated in the dedensification by DPA. Figure 5-5 offers an example of such a situation where DPA results in a trajectory in which important features have been unintentionally eliminated. The trip starting point is A, and the final point is B, going through C and D clockwise. After calculating all perpendicular distances of the other points to the half plane of straight line AB, we find that all are less than the specified tolerance. The output from the DPA dedensification therefore, will be reduced to the straight line AB, omitting the important information consisting of turns that are crucial in characterizing vessel movements in terms of travel/turn combinations. A tighter tolerance may seem like an obvious solution, but this can be problematic as explained in section 5.2.1.2.

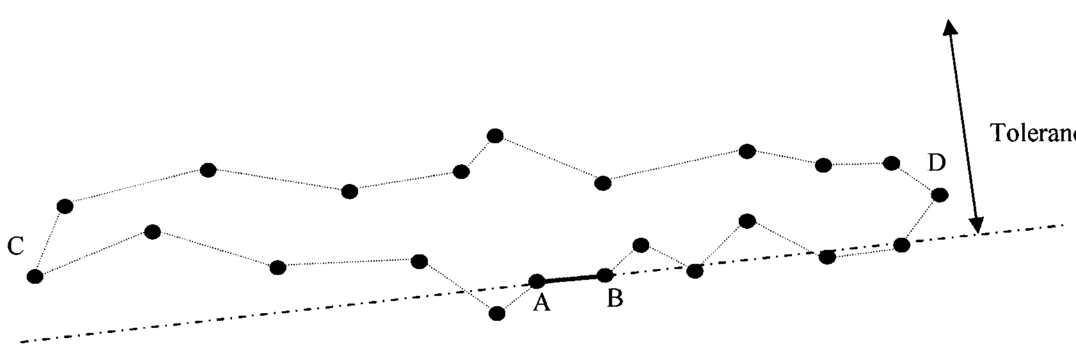


Figure 5-5 The Loss of Important Features in the DP Algorithm

This situation occurs quite often during the dedensification of recreational boat activities, especially those involving kayaks and canoes. Due to the characteristics of these boat types, which are slower and use human power, boaters almost always travel along the coastline. The perpendicular distances (to the straight line joining the start and end points) for near-shore trajectories (such as the left hand path A in Figure 5-6) would be relatively small compared to those for an offshore trajectory (right hand path B in Figure 5-6). Depending on the selected tolerance, the dedensified trajectory for the former case may only keep two points (the start and end points) and in doing so would lose the important features of the trip.

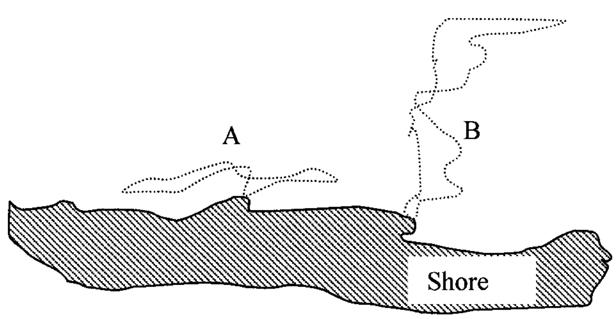


Figure 5-6 Near-Shore and Off-Shore Trajectories

This phenomenon could perhaps be ignored if it rarely arose, but unfortunately this is not the case. Aside from the GPS patterns collected in this study, the broader survey on boaters' activities reinforced that many outings remain near-shore, resulting in the problematic oblong pattern described above. The MARIN phone survey showed that more than three quarters (76%) of respondents stay within 10 kilometers of the shoreline. About 12% of respondents reported travelling a distance of 10 km or more from the coast. Note that the results are a summary of many kinds of recreational boats, including relatively slow boats such as kayaks and canoes.

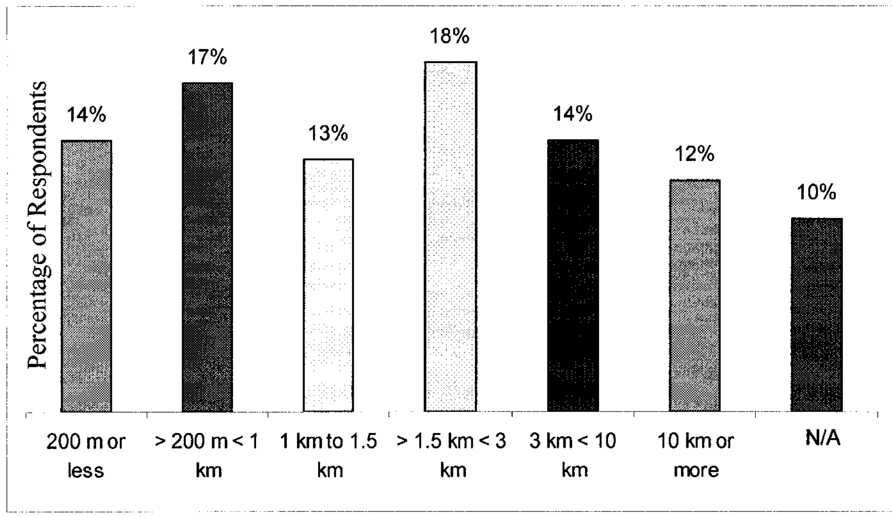


Figure 5-7 The Typical Distance Travelled from Shore

Given that most recreational boating activities occur within 3 km from shore, as depicted in Figure 5-7 (Pelot et al., 2004b), we cannot ignore the loss of important features that will result using DPA in the simplification of the trajectories.

As illustrated in Figure 5-8, Ebisch (2002) found a similar problem when simplifying Alaska's Coastline. If the DPA is programmed to simplify the coastline using a defined tolerance of 105 km, all points along this line, with exception of the first and last, would be eliminated corresponding to the dashed line in Figure 5-8 (Ebisch, 2002). It was calculated that using this tolerance, approximately 475km of coastline would be lost due to the fact that numerous points are within a 105 km distance from the line segment between the start and endpoints (Ebisch, 2002).

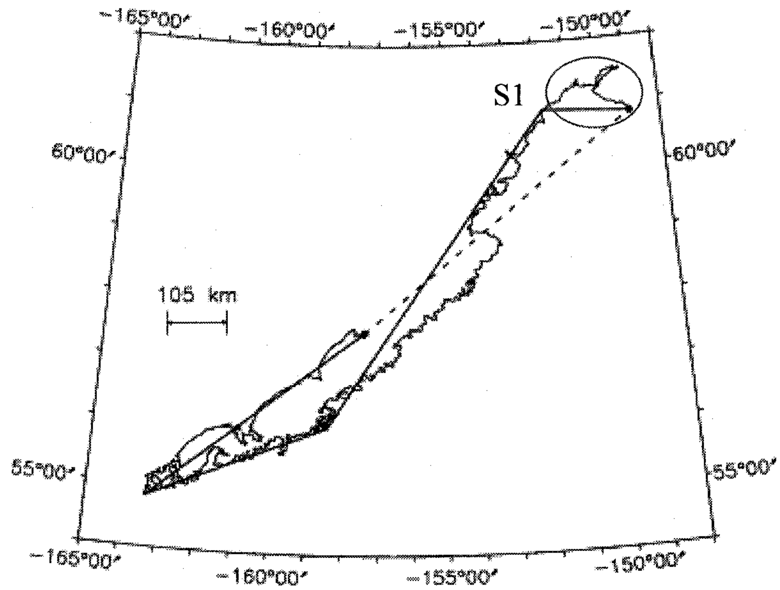


Figure 5-8 An Application of the DP algorithm to the Alaskan Coast Line
(Ebisch, 2002)

5.2.1.2 Modified Approaches

Some might argue that this problem can be avoided by choosing a correspondingly small tolerance. It is true that the important features will be kept only if the threshold is small enough, however there are two problems with this approach. One is the difficulty in choosing the proper threshold; the other is despite using a smaller tolerance, there will usually still be some "noise" points that remain after the dedensification is complete. Noise points prevent a good estimate of turning angles.

Ebisch suggested an improvement concentrating on the definition of perpendicular distances between the points and the straight line. Three different cases are demonstrated in Figure 5-9.

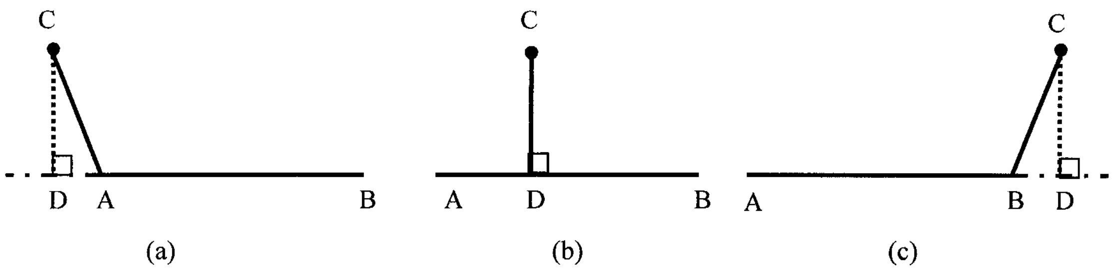


Figure 5-9 Three Perpendicular Cases based on Location Ranges

Given the straight line AB, the perpendicular distance from C to the half-plane AB can be calculated. Point C may be located within the range of AB, as shown in Figure 5-9(b), or may be found outside AB as in Figure 5-9(a) and Figure 5-9(c). Applying standard DPA, regardless of the location of C, the perpendicular distance CD from C to the half-plane AB is always calculated. Ebisch hypothesized that this was the cause of the missing geographic features that commonly resulted from the DPA. He therefore defined different distances according to these three cases. Firstly he determined whether the closest point to C on the straight line segment was A, B, or a point in between. If A or B was the nearest point, he then treated CA or CB as the correct distance. If it was another point in between however, he then used the perpendicular CD as was done in the original DPA (Figure 5-9(b)). The solid line in Figure 5-8 results from his improved algorithm, using the same tolerance as the original DPA.

This modified algorithm mitigates certain problems, but is still not ideal for use in the dedensification of recreational boat trajectories. For example, the simplified coastline in the Figure 5-8 still loses features of segment S1 despite using the improved Ebisch algorithm which, if it were a boating trajectory, would likely not be an acceptable simplification.

Therefore, to achieve our specific aims, we attempted to solve this problem by first dividing the trajectory into segments, then applying the DPA separately to each one. Segment breaks are defined at any point n where the subsequent point n+1 falls closer to the origin than point n. Reworking the example in Figure 5-5 with a start point at A, and an end point at B, in Figure 5-10 the trajectory is divided into three segments AC, CD and DB using this procedure.

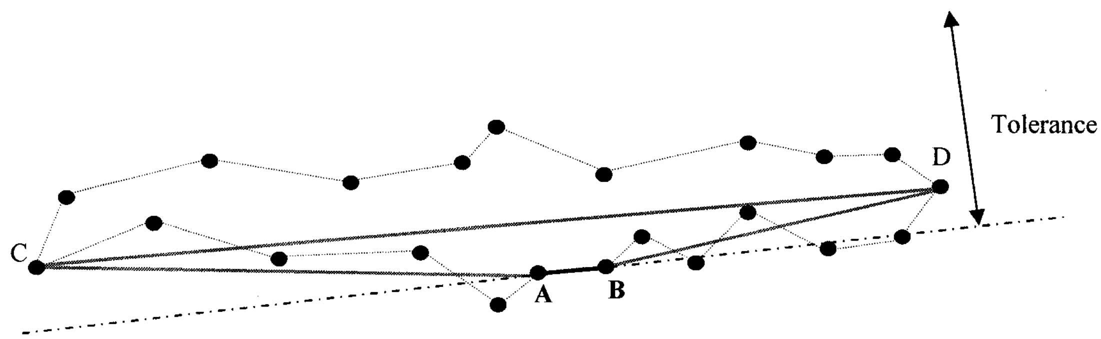


Figure 5-10 MDPA based on Segment Divisions

With this modification algorithm, named MARIN Douglas-Peucker Algorithm (MDPA), none of the points associated with a particular segment extend past the ends of the segment, therefore there is no need to define a different distance calculation as was suggested by Ebisch (2002) to address the issue illustrated in Figures 5-9(a) and Figure 5-9(c). Furthermore, at a minimum, the segment endpoints are retained using MDPA, therefore the dedensified trajectory retains some important features of the original track regardless of the chosen tolerance.

5.2.2 Land-crossing Interference

5.2.2.1 Problem Statement

As illustrated in Figure 5-11, another obvious problem commonly found in the dedensification of boating trajectory points is the issue of land intersection when simplifying the line results in a realistically infeasible path. Since a great percentage of recreational boating is near-shore, this often occurs, and is exacerbated as the dedensification tolerance increases.

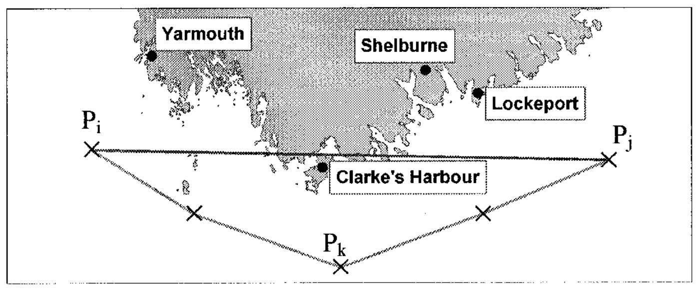


Figure 5-11 An Example of Land Avoidance

5.2.2.2 Solution

During the dedensification process, a sub-procedure is conducted to confirm that any line connecting retained points in the simplified trajectory does not traverse a land mass. This algorithm is applied recursively until an appropriate dedensified trajectory for which all points avoid land is found. As illustrated in Figure 5-11, the simplified segment P_iP_j crosses land and thus it is treated as the baseline with which to locate the furthest point P_k, which is then connected to both P_i and P_j. This sub-procedure is repeated until a feasible trajectory is found, as in Figure 5-11 where neither P_i P_k nor P_k P_j intersects land.

Formally, the basic sub-procedure is as follows:

1) Check if the dedensified line segment P_iP_j crosses land; if so, proceed with steps (2) and (3); otherwise no further operations are required on this segment.

2) Find the furthest point P_k from P_iP_j.

3) Form new lines P_iP_k and P_kP_j, check them both recursively with this procedure.

5.2.3 Results

The MDPA, when applied to the dedensification of recreational boating trajectories, overcomes both the problems of feature loss as well as land interference. To illustrate this success, Figure 5-12 depicts three raw GPS trajectories displayed in the spatial software program ArcGIS (version 9), while the lines represent the dedensified trajectories using this modification.

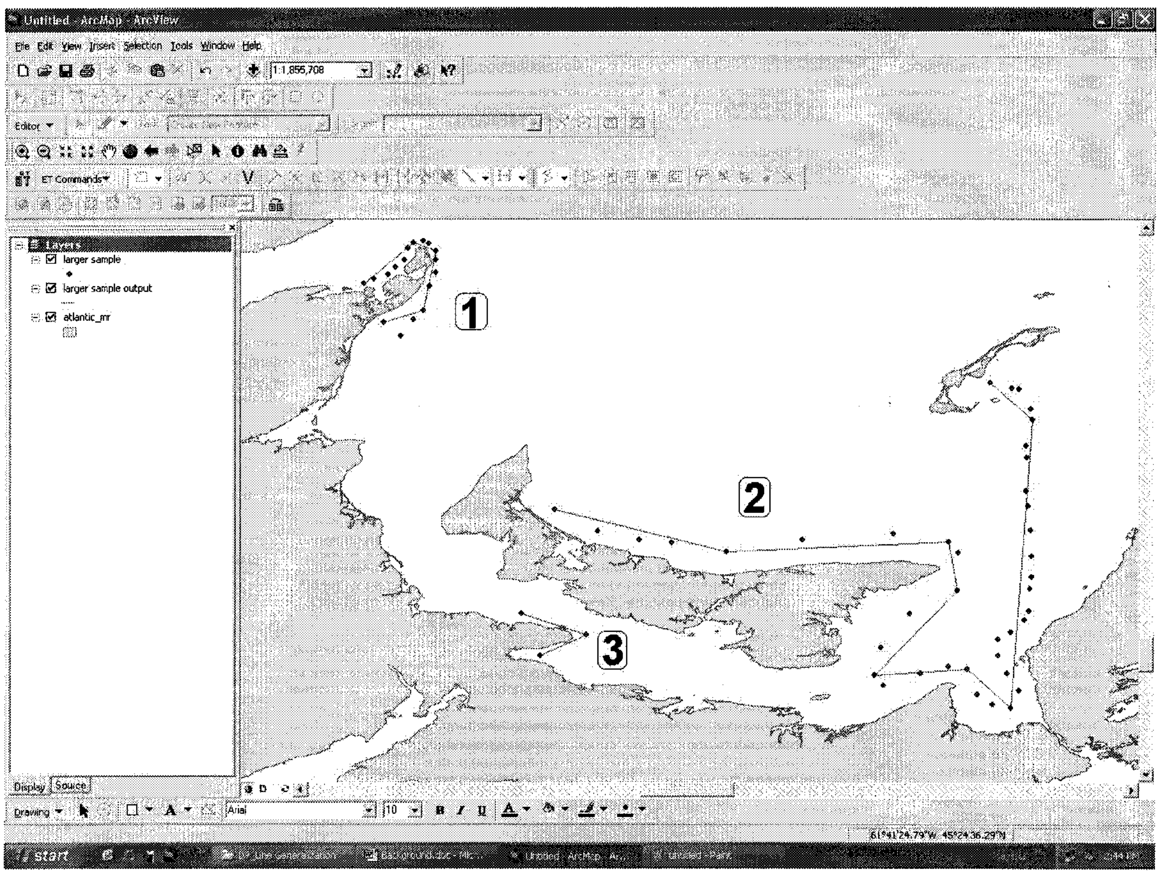


Figure 5-12 Three Raw GPS Trajectories and Dedensified Trajectories

The output (shown in Figure 5-12) represents the dedensified lines using a 0.5 decimal degree tolerance (about 55 km). The original Track 2 contains 40 points, while the completed output from the program uses only nine to approximate the same path. Track 3 is interesting because all of the original points are retained since any dedensification will lead to land intersection for part of the trajectory.

5.3 Two Criteria Dedensification Algorithm

The dedensified trajectory provides two key pieces of information. One is the segment length between two boat turns, and the other is the turning angle between two adjacent segments. In some circumstances, the turning angle is more important than the distance travelled, because the distance is determined by more factors such as the operators' aims and geography; however, the turning angle is influenced mainly by the boat type. So the turning angle is a good attribute of boats' movements. DPA and MDPA only consider one criterion, which is a distance tolerance. The Two Criteria Algorithm considers both the angle and distance tolerances.

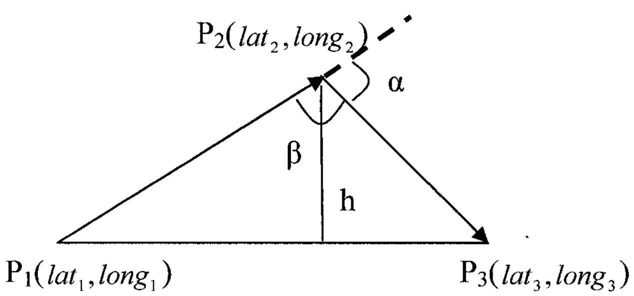


Figure 5-13 Parameters of Two Criteria Algorithm

As shown in Figure 5-13, supposing a recreational boat travelled from point P_1 to point P_3 via point P_2, the two criteria of whether a boat makes a significant turn along its trajectory are:

- Turning angle α is greater than a deviation tolerance of angle, OR
- Perpendicular distance h is great than a deviation tolerance of distance.

Based on the above criteria, if the point P_2 is retained, the boat is considered to have turned. Otherwise this point is deleted, as would be warranted if the boat simply wavered. The turning angle α and the distance h are calculated using Equation [5-1] and Equation [5-2].

$$\alpha = \pi - \arccos\frac{[(lat_1 - lat_2)^2 + (long_1 - long_2)^2] + [(lat_2 - lat_3)^2 + (long_2 - long_3)^2] - [(lat_1 - lat_3)^2 + (long_1 - long_3)^2]}{2\sqrt{(lat_1 - lat_2)^2 + (long_1 - long_2)^2}\sqrt{[(lat_2 - lat_3)^2 + (long_2 - long_3)^2}} \qquad [5\text{-}1]$$

$$h = \frac{|(long_1 - long_3)lat_2 + (lat_3 - lat_1)long_2 + (lat_1 long_3 - lat_3 long_1)|}{\sqrt{(long_1 - long_3)^2 + (lat_1 - lat_3)^2}} \qquad [5\text{-}2]$$

The dedensification algorithm devised by Chang et al. (1991) and applied by Little and Gu (2001) to dedensify sports match players' trajectories also examines three successive points sequentially. For instance in Figure 5-4, the processing order is $P_1P_2P_3$, $P_2P_3P_4$ etc. in the form of $P_{i-1}P_iP_{i+1}$. However, the two criteria algorithm would examine the data as $P_1P_2P_3$, $P_1P_3P_4$ … $P_1P_8P_9$, to overcome the myopia problem. Moreover that algorithm uses distance tolerance as the only criterion. Two criteria algorithm is sensitive to turns too, for example it will keep the sharp turn around S1 in Figure 5-8. Nevertheless, MDPA is better for our specific research aim due to its efficiency and the dedensification results, so it was chosen as the customized algorithm for dedensification to calculate turning angle.

5.4 Context-specific Objective Trajectory Selection Algorithm

Despite the improved performance of MDPA for dedensifying recreational boating trajectories, the arbitrariness of the threshold value selection impedes the extraction of desired attributes such as mean turning angle (MTA) and the length of the straight segments between turns. The analysis of the example in Figure 4-9 illustrates this point. The extreme sensitivity of the outcome variables (turning angle and segment length) to the dedensification threshold value poses a challenging problem, as it is impossible to set the appropriate tolerance beforehand to get a desired result. An objective criterion must be established to select the best simplified line from many candidates of dedensified trajectories that use different tolerances. This section will present an innovative and objective approach to resolve this issue. Figure 5-14 is the automatically generated dedensified trajectory for the sailboat trajectory shown in Figure 4-7(d). The dedensified trajectory was chosen as the "right" one using the approach explained in this section, yielding only 12 points out of the 275 original points, but keeping all of the principal features. The word "right" does not mean correct, as Buja's witty quotation notes: "There is no true interpretation of anything; interpretation is a vehicle in the service of human comprehension. The value of interpretation is in enabling others to fruitfully think about an idea" (Buja, 2000), so "right" means suitable for the study purpose, which is to discern the spatial pattern from an appropriately dedensified trajectory. Since similar sorts of complications may arise in other domains requiring spatial line analysis, such a method would not only be useful for our study, also could be applied to other applications.

Two interim approaches are introduced (5.4.2 and 5.4.3), culminating in a third, more objective and context-specific approach (5.4.4 and 5.4.5) which can then be applied to select the best dedensified trajectory. Note that only coastal data collected in the summer of 2004 in NS were used for analysis in this section to develop the trajectory selection algorithm. Motorboat trajectories are too few to get reliable results, so this type is excluded from the development process. Note that this algorithm was developed in the fall of 2004 at which time only these data were available, but subsequent data usefully served for validation of the method.

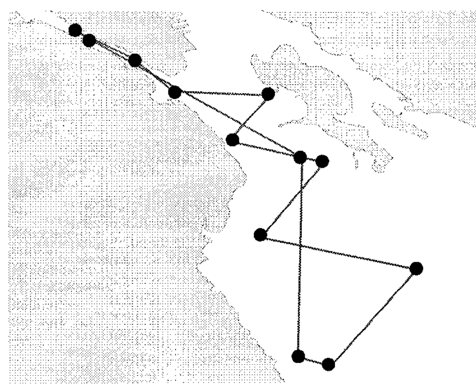


Figure 5-14 Automatically Generated Dedensified Trajectory

5.4.1 Normalized Absolute Deviation (NAD)

There have been many dedensification algorithms developed and studied, but few of them mention an appropriate approach to select the correct dedensified line. One of the criteria used is to pick according to the percentage of original points retained, which can be quite misleading, especially for boating trajectories. Taking the sailboat trajectory in Figure 5-14 for example, 4.4% (12/275) points were kept, while the canoe trajectory in Figure 4-9(a) kept 47.7% (9/19) points. It is impossible to set up a percentage-based standard to arrive at the best dedensified track because usually a more complicated trajectory has more feature points which should be kept to represent the movements. For instance, Figure 5-15(b) is more complicated than the trajectory shown in Figure 5-15(a) although they have the same number of original points. Figure 5-15(a) kept 20% of the original points, but if we use 20% as the criterion for Figure 5-15(b), it definitely results in a very poor representation. Figure 5-15(b) needed 40% of the number of original points to preserve the movement features, and at the same time to delete the noise ones.

Figure 5-15 Different Complexity Trajectories

In order to evaluate the efficiency of MDPA, the normalized absolute deviation (NAD) is introduced. Consider the example in Figure 5-16, where the original trajectory has 10 points. Assuming that points p1, p2,...p8 are not feature points, but that they arise from the boat merely wavering rather than turning, then only points A and B should be retained for the dedensified path. The measure of this simplification begins with the summation ($\sum_i d_i$) of the distances between deleted points and the remaining straight line using latitude and longitude for the coordinate system:

$$\text{Where } d_i = \frac{(long_A - long_B)lat_{P_i} + (lat_B - lat_A)long_{P_i} + (lat_A long_B - lat_B long_A)}{\sqrt{(long_A - long_B)^2 + (lat_A - lat_B)^2}}$$

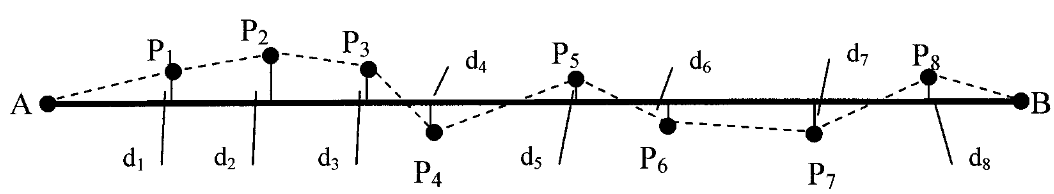
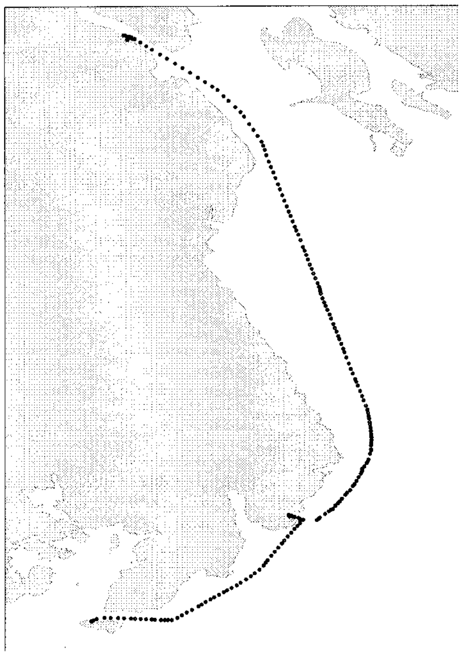


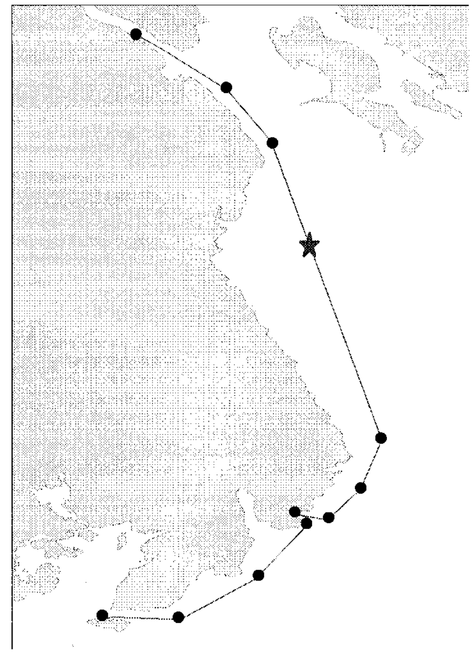
Figure 5-16 The Concept of Normalized Absolute Deviations (NAD)

Since the total distance varies among different trajectories, usually the longer the path, the more points will be deleted, and the larger this summation will be. Therefore, we normalize the error measure by dividing the summation of deviations by the total distance travelled, yielding the NAD used to set the cutoff.

The turning angle attribute should be derived from a dedensified trajectory, because as described earlier using raw data points to calculate turns is erroneous. However, all the other attributes should be calculated based on the cleaned trajectory (i.e. before dedensifying). For example calculating furthest distance from shore for the trajectory shown in Figure 5-17, the cleaned trajectory has 216 points while the dedensified trajectory only keeps 11 points as the dark circles displayed in Figure 5-17(b). Obviously, if calculated based on dedensified trajectory, the computation time would be much less compared to the estimation based on cleaned trajectory. However, it is also evident that the furthest location from shore corresponds to none of the retained points in the dedensified trajectory. It should be the point somewhere near the star in Figure 5-17(b), whereas no point was kept between the two turnings. So the dedensified trajectory is prepared solely for the estimation of turning angle. Conversely, speed, total distance travelled, aspect ratio, coverage index and furthest distance from shore are determined accurately and objectively from the cleaned original trajectory at the expense of computing time.



(a) Cleaned Trajectory



(b) Dedensified Trajectory

Figure 5-17 Cleaned Trajectory and Dedensified Trajectory

5.4.2 Fixed Criteria Approach

On the basis of the coverage index and total distance travelled variables, in general the boat types differ markedly (Table 5-1). Boating trajectories can now be differentiated according to their coverage index and total distance travelled, allowing trajectory-sensitive flexibility in the setting of dedensification tolerances, resulting in the best simplified lines.

Table 5-1 Summary of Total Distance and Complexity of Different Boat Types

	Total Distance (km)			Coverage Index (non-dimensional)		
	Canoe	Kayak	Sailboat	Canoe	Kayak	Sailboat
Min	0.36	1.60	4.80	0.58	0.97	0.20
Mean	1.36	5.50	18.30	1.39	1.46	0.76
Max	2.51	8.70	42.00	2.14	2.19	1.25

Sailboats travel relatively far and exhibit complex movements, whereas canoes and kayaks typically follow the shoreline. In addition, compared with the other two boat types, a sailboat GPS trajectory usually has more points, many of which do not contribute to the definition of important features of the path. Hence, if we want to keep only the significant points for a sailboat, we should allow a relatively large NAD, which will accommodate the deletion of many intermediate points during the dedensification process. However, applying the same NAD cutoff to canoe and kayak trajectories, most of their movement characteristics will be lost because their trajectories are relatively simple and short. Due to this phenomenon, distinct NAD cutoff criteria are set for the different boat types based on trial and error testing: 0.05 for canoes; 0.1 for kayaks; and 1.0 for sailboat. For each boat trajectory, the procedure is to run the MDPA multiple times for a sequence of tolerance values, and then choose the simplified path whose NAD is nearest to the set cutoff criteria for that boat type. If there are two dedensified trajectories having nearest NAD, the one with smaller NAD will be chosen.

5.4.3 Conditional Criteria Approach

The fixed criteria approach described above to select the best dedensified trajectory assumes *a priori* knowledge of the boat type. However, this requirement does not fulfill the main aim of this study, which is to distinguish the boat types based solely on the GPS trajectory data. Therefore the conundrum is how to establish dedensification selection criteria in advance without relying on knowledge of the boat type. From the observation that longer distances demand a higher NAD cutoff, while increasing complexity of the trajectory leads to choices of a smaller NAD cutoff, two objective characteristics of the boating trajectories are used to select the best cutoff: total distance travelled and coverage index. Each of these variables is broken into 10 categories as shown in Table 5-2, by equally divided the range into 10 groups.

Table 5-2 Ten Categories of the Two Variables for Coastal Activity

Total Distance (meter)				Coverage Index (non-dimensional)			
Category	Lower Bound	Upper Bound	Proportion	Category	Lower Bound	Upper Bound	Proportion
1	350	5,590	0.39	1	0.200	0.405	0.06
2	5,590	10,830	0.22	2	0.405	0.610	0.08
3	10,830	16,070	0.13	3	0.610	0.815	0.07
4	16,070	21,310	0.06	4	0.815	1.020	0.20
5	21,310	26,550	0.06	5	1.020	1.225	0.15
6	26,550	31,790	0.03	6	1.225	1.430	0.13
7	31,790	37,030	0.04	7	1.430	1.635	0.03
8	37,030	42,270	0.02	8	1.635	1.840	0.03
9	42,270	47,510	0.02	9	1.840	2.045	0.12
10	47,510	52,750	0.02	10	2.045	2.250	0.13

Since short trips produce relatively few GPS data points, it is important to set the cutoff such that many of the points are retained. Conversely, for longer trips, the shape of the boat trajectory must play a larger role in selecting the best cutoff. These guiding principles lead to the following set of conditional rules for picking the cutoff:

If category of total distance = 1,
Then cutoff $\leq$ 0.05
Else if category of coverage index = 1,
    Then cutoff $\leq$ 0.5,
    Else cutoff $\leq$ 1.0

Therefore, if the total distance is very short (i.e. falls into category 1), we set the cutoff equal to 0.05. Otherwise checking the coverage index category further, if it falls in its category 1 as well, we set the cutoff to 0.5. All the other situations will have the cutoff as 1.0.

5.4.4 Functional Criteria Approach

The conditional method described above is limited to only three cutoff values: 0.05, 0.5 and 1.0 corresponding respectively to extreme cases. An extremely short trajectory results in an extremely small cutoff of 0.05. Conversely an extreme large cutoff equal to 1.0 is associated with very long trips, unless it is extremely complex, in which case the cutoff is moderated to a lower value of 0.5. But this gives little latitude to accommodate other combinations of total distance and coverage index, limiting its usefulness for this study. Therefore, a refinement to the method is required to better reflect the characteristics of individual trajectories.

Let $f(x_1, x_2)$ be a function of total distance category x_1 and coverage index category x_2. This function allows a different cutoff for different distance and coverage index variables, not only three cutoffs as in the preceding fixed and conditional methods. To define a suitable function f for this purpose, two requirements are considered. It should assume a fairly simple form, and it should discriminate across different boat types. The following ten diverse functions, including linear and nonlinear, were tried to check their effectiveness for differentiating the boat types:

$$\eta_1 = x_1 + x_2$$

$$\eta_2 = x_1 + \frac{1}{2}x_2$$

$$\eta_3 = x_1 + \frac{1}{4}x_2$$

$$\eta_4 = x_1 + \frac{1}{8}x_2$$

$$\eta_5 = x_1 \times x_2$$

$$\eta_6 = x_1^2 + x_2$$

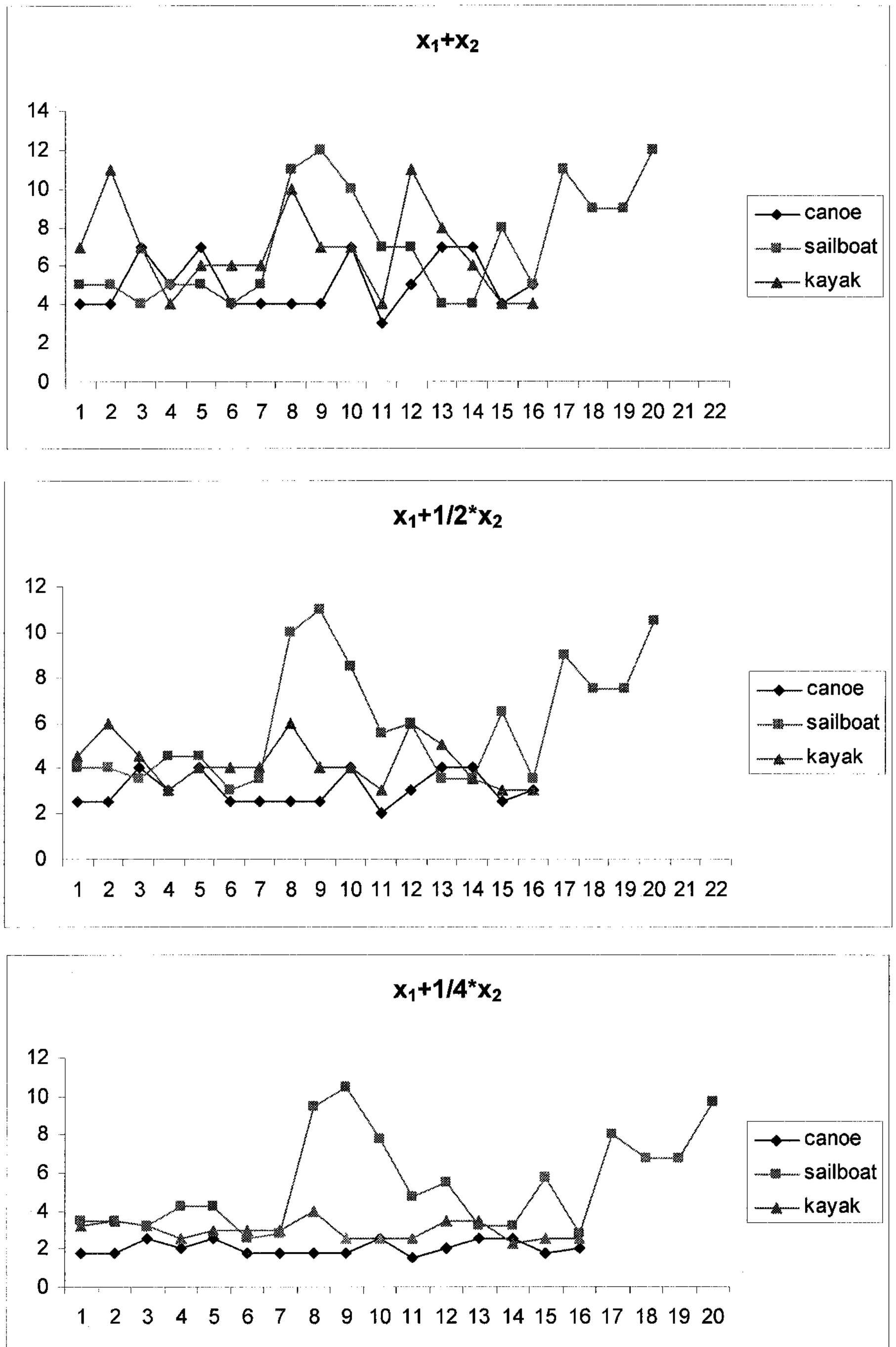
$$\eta_7 = x_1^3 + x_2$$

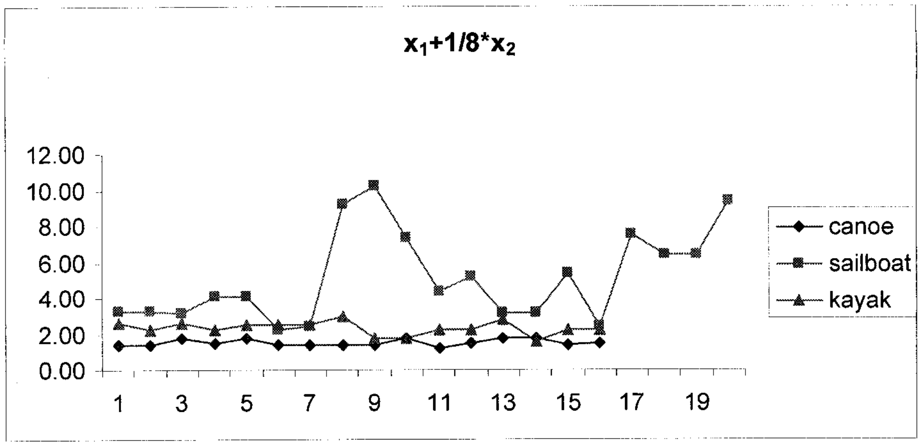
$\eta_8 = x_1 + \ln(x_2)$

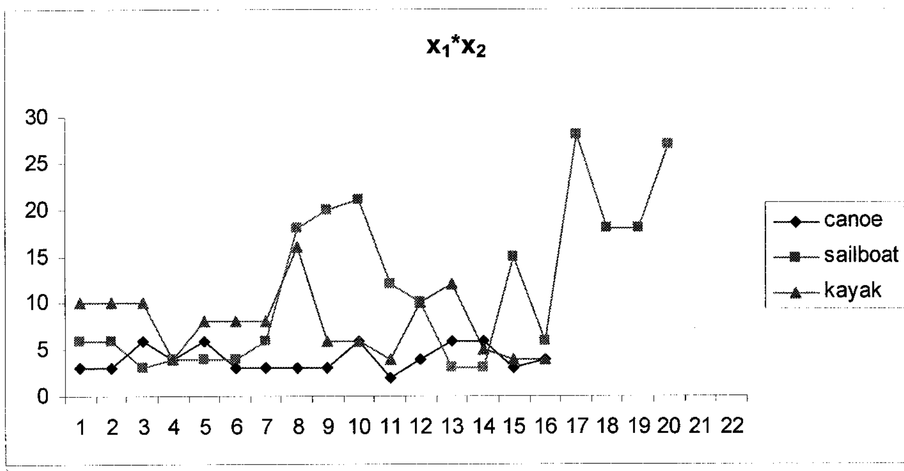
$\eta_9 = \ln(x_1) + x_2$

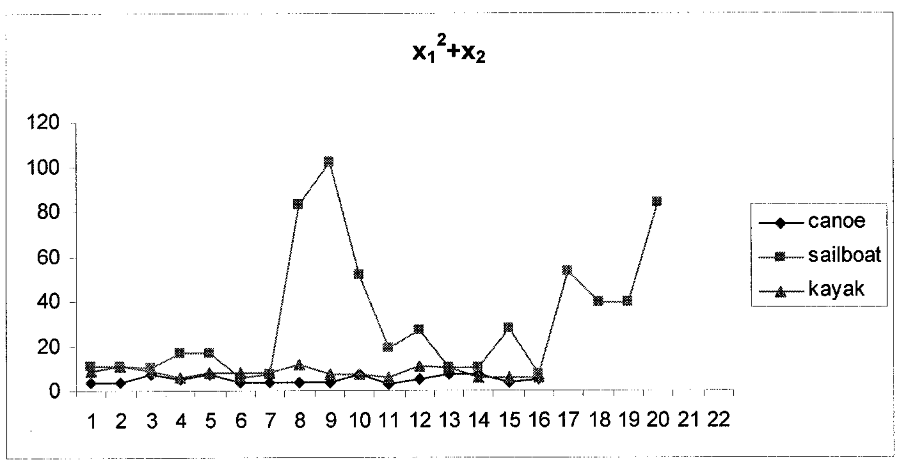
$\eta_{10} = x_1^2 + \ln(x_2)$

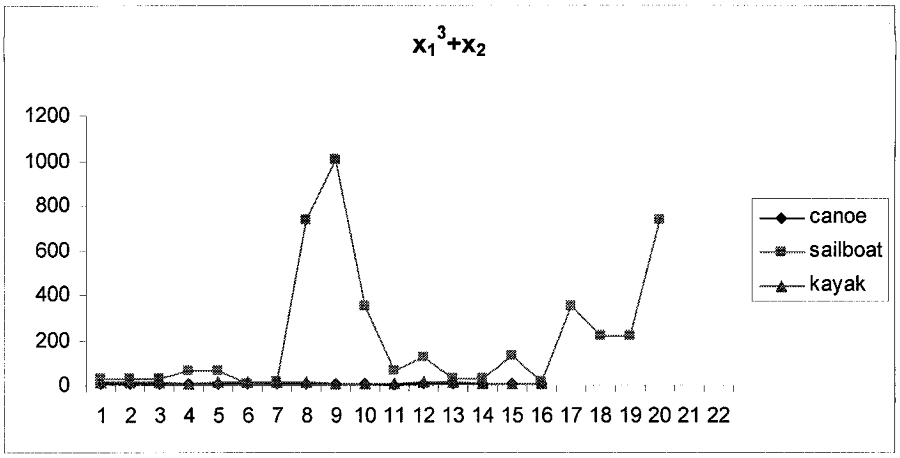
Figure 5-18 Various Functions

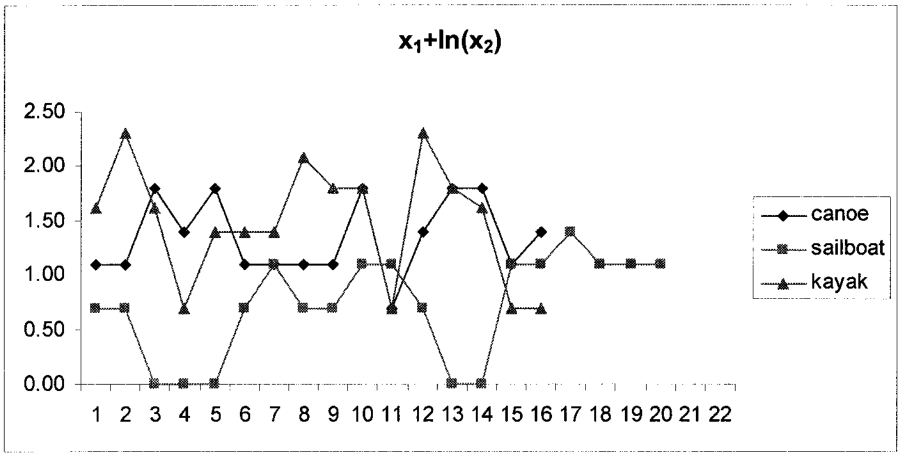


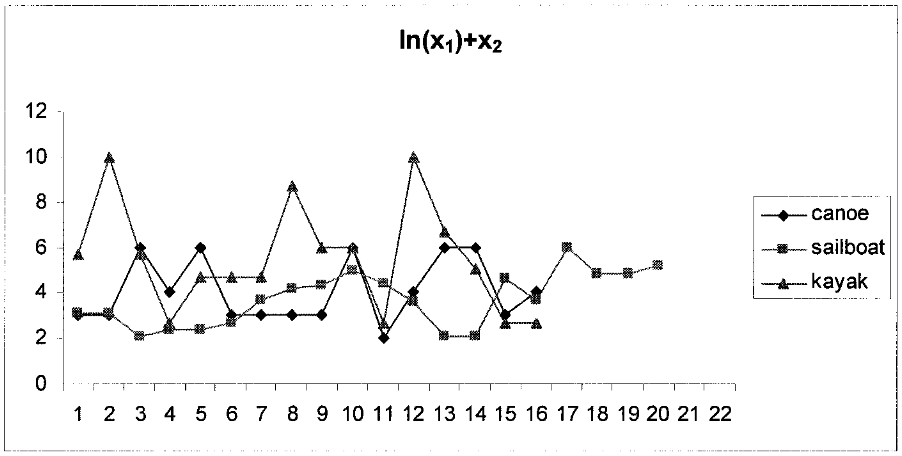

$x_1+1/8*x_2$
12.00
10.00
8.00
6.00
4.00
2.00
0.00
1 3 5 7 9 11 13 15 17 19
canoe
sailboat
kayak

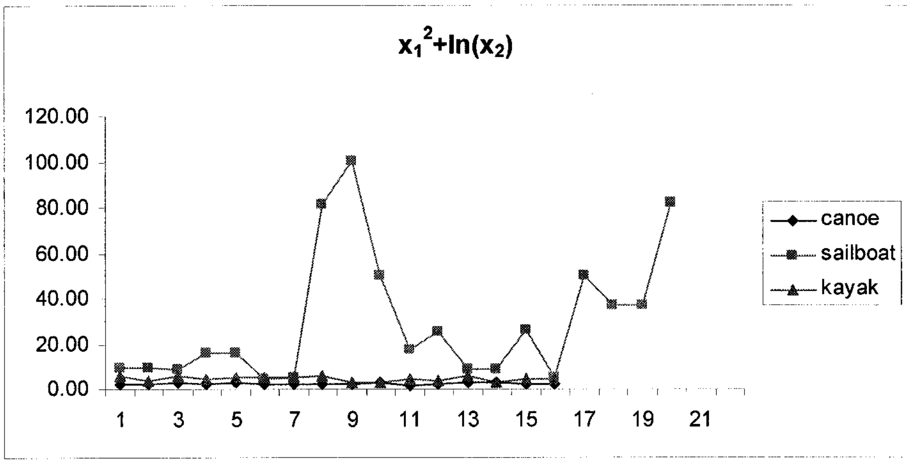

x_1*x_2
30
25
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
canoe
sailboat
kayak


$x_1^2+x_2$
120
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
canoe
sailboat
kayak


$x_1^3+x_2$
1200
1000
800
600
400
200
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
canoe
sailboat
kayak


$x_1+\ln(x_2)$
2.50
2.00
1.50
1.00
0.50
0.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
canoe
sailboat
kayak


$\ln(x_1)+x_2$
12
10
8
6
4
2
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
canoe
sailboat
kayak



The function $\eta_3 = x_1+1/4*x_2$ separates the different boat types better than other functions. The domains of each of the two variable categories are [1, 10], therefore the η_3 range is [1.25, 12.5]. Using linear transformation to map [1.25, 12.5] to the cutoff domain [0.05, 1], the function is built as Equation [5-3].

$$f = 0.085x_1 + 0.021\, x_2 - 0.056 \qquad [5\text{-}3]$$

Placing a greater emphasis on total distance x_1, with a modest weight accorded to the coverage index x_2 variable, is coherent with preceding observations, and yielded good results. Figure 5-19 shows a relatively good separation of the function across the three boat types, whereby different trajectories are assigned their own cutoff values. The imperfect performance of the separation capabilities of this function, for example some sailboats' functions fall into other boat types' range, implies that not every trajectory displays the typical characteristics of its class. For example, not all of the sailboats travelled in a zigzag manner because of the boaters' experience or wind conditions, so they may have turning angle features as well as a coverage index similar to kayak movements.

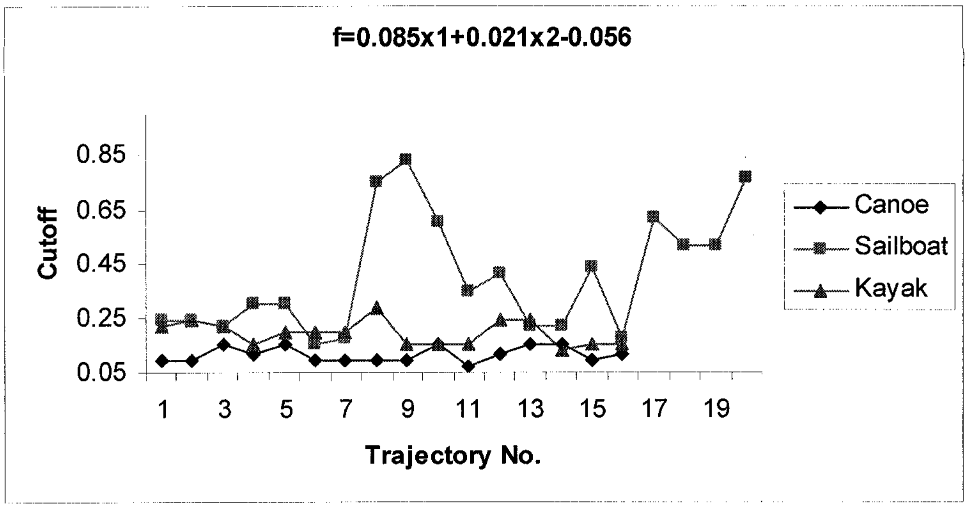
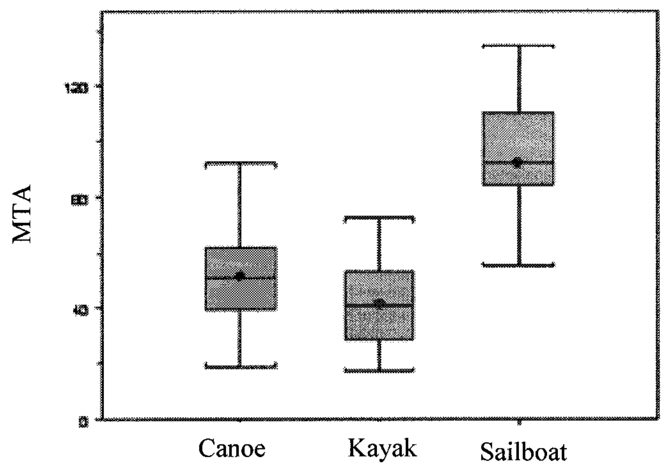


Figure 5-19 Functional criteria applied to diverse boat types

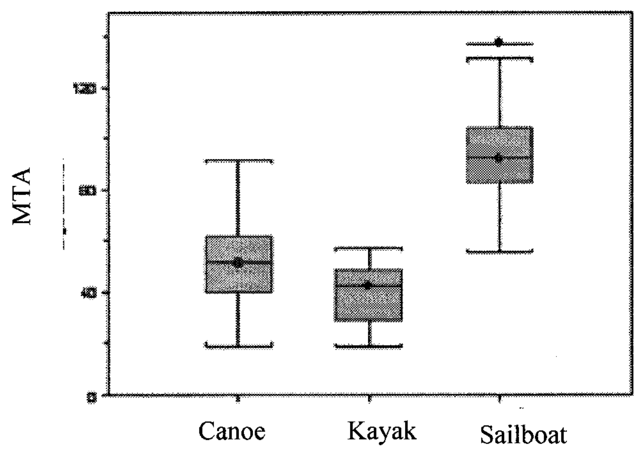
5.4.5 Results and Discussion of the Three Approaches

The data from the NS coastal area were used for developing the parameters for the three approaches described above. Those results will be examined in this section. The NB river area data were used to validate the methods based on ability to distinguish boat types according to turning angles, as presented in the subsequent section.

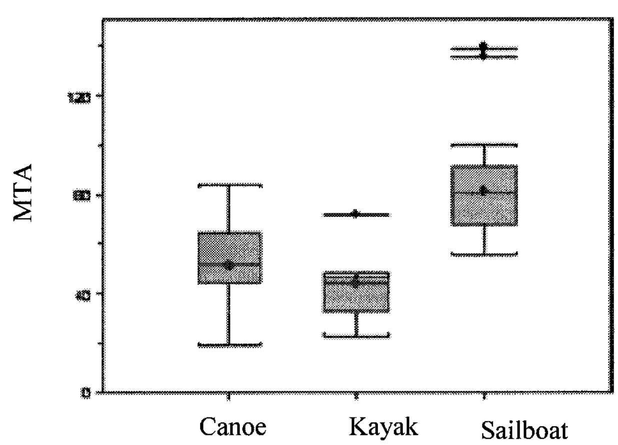
The box plots of mean turning angle (MTA) for coastal data, derived respectively from the fixed, conditional, and function criteria approaches, are shown in Figures 5-20(a), (b) and (c) respectively. An ANOVA analysis is applied to the MTA across boat types for each of the three methods, resulting in *p*-values of $1.0\text{x}10^{-11}$, $2.7\text{x}10^{-12}$ and $9.6\text{x}10^{-9}$ respectively, which are all negligible. In each situation therefore, the null hypothesis H_0 that the MTAs are not significantly different is rejected. Furthermore, applying the multiple comparison Tukey *post hoc* analysis to each case, shows that sailboat mean turning angles are significantly different from those of the other two boat types. However, there are no significant differences between canoe and kayak MTAs.



(a) Fixed Criteria



(b) Conditional Criteria



(c) Function Criteria

Figure 5-20 Box plot of Mean Turning Angle (MTA)

Table 5-3 compares the MTAs for each of the boat types resulting from each of the selection methods. Considering first the implication of going from the fixed approach to the conditional approach, the conditional method yields the same result as the fixed criteria approach for canoes, since the total distances of canoe trajectories all belong to category 1, and therefore the selected criteria of 0.05 from the fixed criteria method does not change. As none of the sailboat trajectories fell into category 1 since the distances are not short, the cutoff choice is dictated by the coverage index attribute. Therefore, under the conditional criteria approach this sometimes results in 0.5 for sailboats, yielding slightly different outcomes for mean turning angle compared with the fixed method. Finally for kayaks, rather than the uniform cutoff selected under the fixed approach, the conditional method allows for the distinction between relatively short trips, and those which are longer but not unduly complex. Hence all of the kayaks were assigned extreme cutoff values of 0.05 or 1.0, slightly altering the MTA compared with the fixed approach.

Table 5-3 Mean Turning Angle derived from Different Criteria Methods

Approach	Canoe	Kayak	Sailboat
Fixed Criteria	51.06°	42.14°	96.08°
Conditional Criteria	51.06°	40.04°	95.59°
Function Criteria	52.68°	41.28°	83.33°

This demonstrates that the conditional approach is worthwhile because it segregates the boat types, as does the fixed approach, but with the benefit of not having to know the type in advance. It can tell the difference among different boat types according to objective variables: total distance and coverage index.

The function criteria approach has relatively little impact on the mean turning angles of canoes and kayaks, but a greater change occurs with the sailboat MTA (Table

5-3). Although the angle got smaller for sailboats, it did not influence the discrimination. As shown in Figure 5-21, there is still a significant difference between the MTA for sailboats and those of canoes and kayaks.

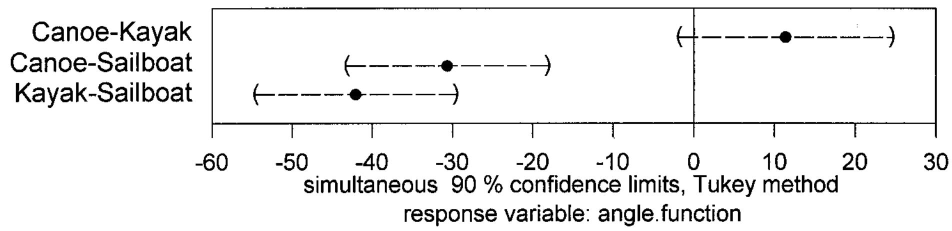


Figure 5-21 Confidence Intervals of MTA according to Function Criteria Method

It is not surprising that on average sailboat turns are significantly larger, because a common goal of sailors is to catch the wind for effective propulsion, which generally involves a lot of turns. In fact, the turning angle of a sailboat is often around 90 degrees, which represents a characteristic of sailboats. A sailboat cannot sail directly into the wind. So when sailing upwind, the angle between the vessel's heading and the wind origin must be at least about 45 degrees to one side of the wind. Then the crew can tack, which means quickly steering the boat so that it is on the other side of the wind, approximately a 90 degree turn (West, 1994). This characteristic is valuable for discrimination and classification boat types, especially sailboats, which is why we have an interest in the turning angle.

5.4.6 Validation and Category Adjustment

Based on the results from these three methods for setting the NAD cutoff, it is evident that all of them are suitable for distinguishing sailboat patterns. However, the function method offers a major advantage over the other two methods. Since it does not rely on *a priori* knowledge of the boat type, it is more objective and the function form is more sensitive to each vessel's trajectory characteristics, selecting a cutoff criterion accordingly.

The function method applied above to the coastal traffic is then validated using the GPS data from the Saint John River. There is a significant difference between the canoes and kayaks' group and the sailboats as shown in Figure 5-22, reaching the same conclusion obtained from the NS coastal activity data.

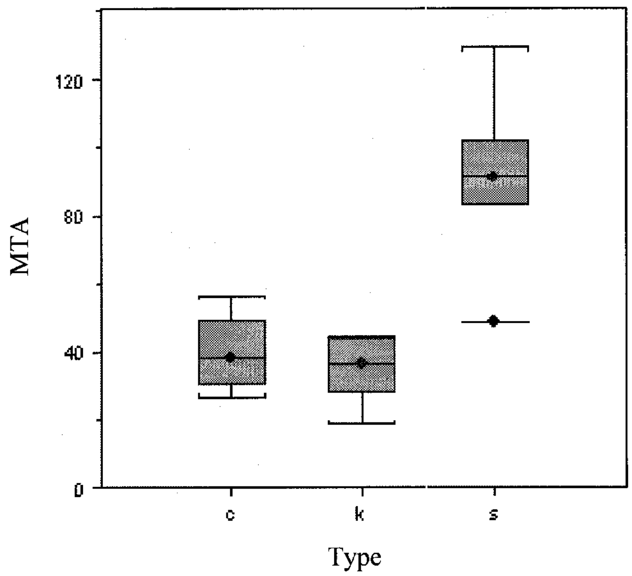


Figure 5-22 Box Plot of Turning Angle of Data from Saint John River

Using the 32 samples from the river area, the ten equally divided categories of total distance travelled are listed in Table 5-4. However, most of the proportions are 0 because of the extremely longest distance of 87,648 meters, which made 71% of the trajectories fall into the first category. Thus, this categorization scheme does not well represent the variability of the differences in total distance travelled. In order to remedy this, the two extremely longest trajectories were deleted, and then the remaining 30 samples formed the basis for equally dividing into 10 distance categories. The resulting distance and coverage index categories for the river data are shown in Table 5-5. Pooling all of the data from the coastal and river trajectories, it is found that the total distance category from the combined data (Table 5-6) are the same as the one from coastal data (Table 5-2), which inspired us to simplify the procedure by using rounded-off categories as shown in Table 5-7. By setting these fixed categories, the approach can be standardized and applied more widely.

Table 5-4 Ten Categories of the Total Distance Travelled for River Activity

Total Distance			
Category	Lower Bound	Upper Bound	Proportion
1	860	9,539	0.71
2	9,539	18,218	0.19
3	18,218	26,897	0.03
4	26,897	35,576	0.00
5	35,576	44,255	0.00
6	44,255	52,934	0.00
7	52,934	61,613	0.00
8	61,613	70,292	0.03
9	70,292	78,971	0.00
10	78,971	87,650	0.03

Table 5-5 Ten Categories of the Two Attributes for River Activity

Total Distance				Coverage Index			
Category	Lower Bound	Upper Bound	Proportion	Category	Lower Bound	Upper Bound	Proportion
1	860	3,035	0.33	1	0.40	0.59	0.06
2	3,035	5,210	0.20	2	0.59	0.78	0.00
3	5,210	7,385	0.13	3	0.78	0.97	0.13
4	7,385	9,560	0.07	4	0.97	1.16	0.41
5	9,560	11,735	0.07	5	1.16	1.35	0.13
6	11,735	13,910	0.10	6	1.35	1.54	0.00
7	13,910	16,085	0.03	7	1.54	1.73	0.00
8	16,085	18,260	0.03	8	1.73	1.92	0.06
9	18,260	20,435	0.00	9	1.92	2.11	0.13
10	20,435	22,610	0.03	10	2.11	2.3	0.09

Table 5-6 Ten Categories of the Two Attributes in both Areas Combined

Total Distance				Coverage Index			
Category	Lower Bound	Upper Bound	Proportion	Category	Lower Bound	Upper Bound	Proportion
1	350	5,590	0.424	1	0.20	0.41	0.06
2	5,590	10,830	0.232	2	0.41	0.62	0.07
3	10,830	16,070	0.128	3	0.62	0.83	0.06
4	16,070	21,310	0.056	4	0.83	1.04	0.25
5	21,310	26,550	0.056	5	1.04	1.25	0.17
6	26,550	31,790	0.024	6	1.25	1.46	0.09
7	31,790	37,030	0.032	7	1.46	1.67	0.02
8	37,030	42,270	0.016	8	1.67	1.88	0.04
9	42,270	47,510	0.016	9	1.88	2.09	0.15
10	47,510	52,750	0.016	10	2.09	2.3	0.09

Table 5-7 Rounded-off Fixed Categories

Total Distance				Coverage Index			
Category	Lower Bound	Upper Bound	Proportion	Category	Lower Bound	Upper Bound	Proportion
1	0	5,000	0.362	1	0	0.25	0.02
2	5,000	10,000	0.260	2	0.25	0.50	0.06
3	10,000	15,000	0.118	3	0.50	0.75	0.09
4	15,000	20,000	0.071	4	0.75	1.00	0.17
5	20,000	25,000	0.063	5	1.00	1.25	0.28
6	25,000	30,000	0.032	6	1.25	1.50	0.09
7	30,000	35,000	0.016	7	1.50	1.75	0.02
8	35,000	40,000	0.024	8	1.75	2.00	0.12
9	40,000	45,000	0.024	9	2.00	2.25	0.14
10	45,000	infinity	0.032	10	2.25	infinity	0.01

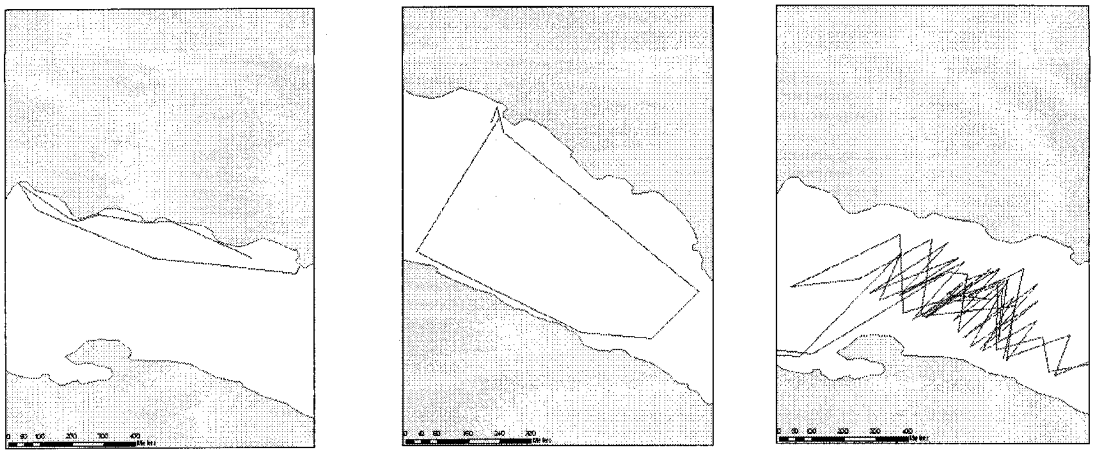
To see if the preceding three category groupings (i.e. derived from respective coastal/river areas, or from the combined data from the two areas, or the rounded-off categories) result in different outcomes, an ANOVA on their resulting MTAs was run. The MTAs of each boat type in each area are not significantly different regardless of which method for establishing categories was applied, since the p-values are very high (Table 5-7). So using the rounded categories listed in Table 5-7 is the most suitable approach for balancing the complexity of the data and the desired simplicity of the method.

Table 5-8 *P*-value of One-Way ANOVA for MTA by different Category Methods

	Canoe	Kayak	Sailboat
Coast	0.99	0.99	0.97
River	1.00	0.82	0.83

5.5 Recreational Boating Trajectory Dedensification Results

All the results and analyses indicate the effectiveness of the MDPA and the context-specific objective selection approach with rounded-off fixed categories to select the best dedensified trajectories for the proposed pattern recognition. For example, Figure 5-23 is the dedensified trajectories automatically chosen using these methodologies (i.e. MDPA and context-specific selection approach) for the original boating trajectories in Figure 4-4. The numbers of points are drastically decreased, only keeping the points when the vehicle turned. Employing these methods, the turning angles would be much more accurate than using the original GPS trajectories, and can discriminate between different boat types. Moreover, this algorithm works very well, even on trajectories which did not apparently adopt their typical boat type movement characteristics. Also this approach may provide suggestions for other applications, which need to choose appropriate dedensified lines for specific analyses.



(a) c-0707-08-03 (b) c-0721-11-04 (c) s-0722-14-10

Figure 5-23 Dedensified Boating Trajectories

6. Recreational Boating Trajectory Pattern Analysis

The study of recreational boating is a very important aspect of maritime risk analysis because of the frequency of accidents, many of which incur serious consequences to the boaters. Despite several studies on the risks associated with recreational boating, unfortunately little is known about detailed recreational boating patterns, although this could provide insight into recreational boat movement characteristics and exposure levels to advance the research on risk analysis associated with this activity. In order to glean some knowledge about recreational boating patterns, in this study, GPS points were collected for a sample of recreational boating trajectories on four types of boats (canoes, kayaks, motorboats and sailboats) in two environments (coastal and river), and then the GPS data were examined to find spatial patterns. Assiduous cleaning and customized algorithms for line dedensification are critical steps in preparation for the pattern analyses, as described earlier. This chapter will concentrate on presenting the analyses relating to the main objectives of this research, in the order of boating pattern discrimination and classification.

6.1 Spatial Boating Trajectory Patterns of Different Boat Types

Six attributes were derived from the GPS data to provide objective information for the pattern analysis. They are segment speed, total distance, bounding box, and furthest distance from shore (which are calculated from the cleaned trajectories), and turning angle and segment length derived from the dedensified trajectories. With these attributes, 9 variants could be obtained: mean speed, maximum speed, $\text{maximum}_{1/20}$ speed, total distance travelled, aspect ratio, coverage index, furthest distance from shore, mean turning angle and mean segment length. Because the variability of segment lengths is so large, it is excluded from the discrimination and classification analyses, however it is necessary for simulating trajectories, as shown in next chapter. Maximum speed was also excluded due to its unreliability compared to $\text{maximum}_{1/20}$ speed. In order to distinguish patterns of different boat types, the data are examined for each study area separately, and then with the areas combined to explore a potential location influence.

This section attempts to answer the first research question: Do different boat types travel differently? For each of the two locations, a univariate ANOVA test was performed for each variable to test for significant differences across the boat types, where the null hypothesis is that they are not different. The *p*-value results are listed in the Table 6-1 for both of the study areas. For consistency, the significance level was set at 0.10. For the coastal area, the result is that only the coverage index was not significantly different across different boat types, and all the other attributes were significantly different. In the river area, total distance travelled was the only attribute having no significant difference across boat types.

Table 6-1 *P*-value of One-Way ANOVA

	River (NB)	**Coastal (NS)**
Total Distance Travelled	4.8×10^{-1}	7.2×10^{-9}
Mean Speed	2.4×10^{-5}	0
Max 1/20 Speed	6.5×10^{-5}	0
Mean Turning Angle	2.8×10^{-2}	1.2×10^{-3}
Coverage Index	4.8×10^{-2}	2.3×10^{-1}
Aspect Ratio	7.2×10^{-2}	7.8×10^{-2}
Furthest Distance from Shore	2.2×10^{-6}	3.3×10^{-7}

Applying Tukey's *post hoc* multiple comparison method, we can determine which boat type(s) are different from others with respect to the different movement attributes. The results are summarized in Table 6-2. The boat types in each circle are significantly different from those in the other circle(s) in the same cell.

Table 6-2 Spatial Boating Trajectory Patterns of Different Boat Types

Attribute	River (NB)		Coastal (NS)		
Total Distance Travelled			C, K	S, M	
Mean Speed	C,K,S	M	C,K	M	S
$Max_{1/20}$ Speed	C,K,S	M	C,K	M	S
Mean Turning Angle	C,K	S	K	S	
Coverage Index	M	S			
Furthest Distance from Shore	C,K,M	S	C,K	S	
Aspect Ratio	K	S	C	S	

As expected, it can be claimed definitively that the mean speed for motorboats is significantly different than that of canoes and kayaks, but one cannot distinguish between canoes and kayaks by speed. The same result applies to $maximum_{1/20}$ speed, since motorboats can attain higher speeds than canoes and kayaks, but it is not possible to differentiate between canoes and kayaks by this attribute.

As seen in Table 6-2, sailboat patterns can be distinguished from other vessel types utilizing four variables in the river, and all variables but one in the coastal area. Sailboats not only travel longer distance than canoes and kayaks, but also further from shore. This is perhaps a consequence of heading to relatively open water to take advantage of the wind, as well as their relatively deeper draught requiring shallow water avoidance. Furthermore, sailboats zigzag when tacking in the wind, which accounts for their mean turning angle being significantly larger than those for other boat types, as well as their higher aspect ratio and lower coverage index.

Being unpowered, it is not surprising that canoes and kayaks stay relatively close to shore and do not cover long distances. This univariate approach is apparently not effective at discriminating between canoe and kayak movements, but subtle differences can be used to full advantage through forthcoming multivariate analysis.

6.2 Spatial Boating Trajectory Patterns at Different Geographic Locations

Since the classification groupings are somewhat different in the two geographic locations as shown in Table 6-2, this leads to the second objective to explicitly test whether the boating trajectories for each type of vessel vary between the coastal and river areas. For each vessel type and attribute, we formed a null hypothesis that the mean of the attribute does not differ significantly between the two locations, coast and river. The asterisks in Table 6-3 indicate that the difference is significant for that attribute and boat type. Most of the attributes are not significantly different between the coast and river, except for the speed of canoes. The Saint John River's current or calmness relative to the ocean might be responsible for that as canoes may achieve noticeably higher speeds than in the coastal waters. In general, geography is not an important influencing factor based on these data. Notably, the sailboats' behavior was unaffected by the location in our results.

Table 6-3 *P*-value of Standard Two-Sample t-Test for Each Boat Type at Two Study Areas

Attribute	Canoe	Kayak	Motorboat	Sailboat
Total Distance	0.0480*	0.5400	0.7300	0.4500
Mean Speed	0.0032*	0.3600	0.7900	0.6100
$Max_{1/20}$ Speed	0.0073*	0.6500	0.0590*	0.9800
Mean Turning Angle	0.1300	0.0740*	0.9600	0.1400
Coverage Index	0.3200	0.2900	0.1900	0.4200
Furthest Distance from Shore	0.3800	0.4800	0.0024*	0.1500
Aspect Ratio	0.9000	0.0230*	0.5200	0.9400

* p-value ≤ 0.10, H_0 is rejected

6.3 Classification of Recreational Boating Trajectory Patterns

It is found that sailboats are significantly different from the other boat types in terms of speed, total distance travelled, bounding box and turning angle. The trajectories of sailboats are long, fast, wide and complex. But canoes and kayaks are similar to each other according to these attributes. Both have slow, narrow, short and simple trajectories. Motorboats are significantly faster, as one may expect. Furthermore, geographic location does not much influence the boating spatial patterns, given the data available. In this section, multivariate analysis is applied to classify the recreational boating trajectory patterns as proposed in this study. The analysis follows the order of coastal area, river area, and then combined areas as well, in order to furnish evidence to compare different patterns. But first of all, the classification methodology: multivariate discriminant analysis is introduced.

6.3.1 Multivariate Discriminant Analysis

A neural network approach was considered for this work but was inadvisable due to the difficulty choosing the number of layers and nodes, and the subjective intervention to do so. The Classification And Regression Trees (CART) method was also avoided because of the difficulty associated with its rules for selecting the best splits and establishing a criterion for choosing the extent of the tree. Furthermore, CART results are not particularly informative with respect to the nature of the relationships between the variables. Given the relatively limited trajectory sample size, and the preference to use a statistical method which does not rely on manual intervention, Multivariate Discriminant Analysis (MDA) was selected to address the objectives.

Multivariate discriminant analysis is a statistical multivariate approach invoked when the dependent variable is categorical and the independent variables are metric. Its primary objective is to identify the group to which an object belongs. The general form is shown in Equation [6-1].

$$Y \quad = X_1 + X_2 + X_3 + \ldots + X_n \tag{6-1}$$

(nonmetric) (metric)

Boat Type $\Longleftrightarrow$

- Max Speed
- Max $_{1/20}$ Speed
- Mean Speed
- Total Distance Travelled
- Coverage Index
- Aspect Ratio
- Mean Turning Angle
- Furthest Distance from Shore

When there are only two groups involved, it is known as logistic regression. This methodology is widely used in many applications such as acoustics for speech recognition, taxonomy in biology and medicine, and market research (Fisher, 1936; Blanc et al., 2001; Jensen et al., 2005; Kelly et al., 2005). Hence, this method has been commercialized in many statistical software packages such as SAS, SPSS, MiniTab and

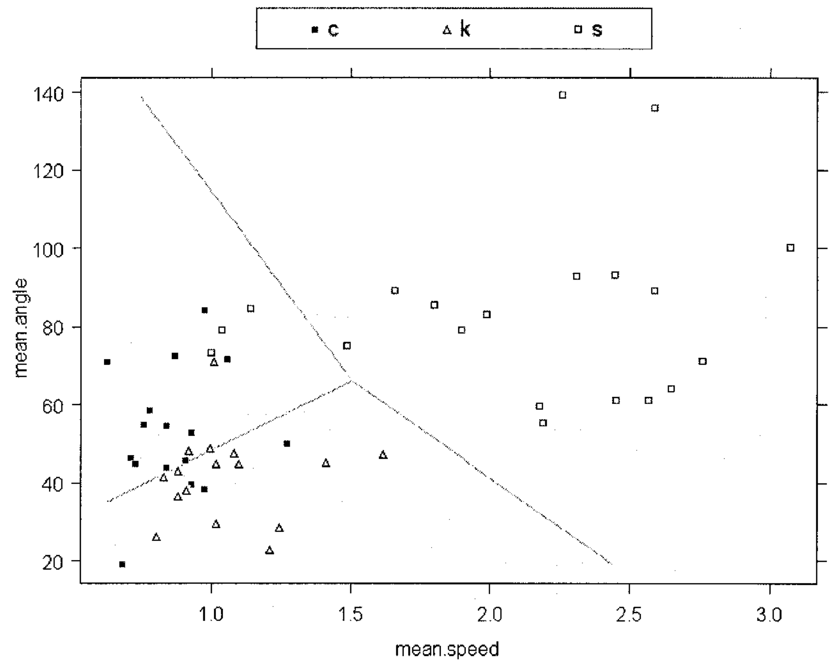
S-Plus. Because this study must handle four boat types, a non-metric variable, and the independent variables are all metric attributes, discriminant analysis is the right choice.

The main idea of discriminant analysis is maximizing between-group variation and minimizing within-group variation by deriving a linear combination: $l(X) = \beta_{i0} + \beta_{i1}X$ or a non-linear combination: $d(X) = \beta_{i0} + \beta_{i1}X + X^T\beta_{i2}X$ of some independent variables that will discriminate best across *a priori* defined types. The covariance matrices determine which model to choose: linear or quadratic. If all the groups have equal covariance matrices, a homoscedastic model (linear) would be constructed, otherwise, a heteroscedastic model (quadratic) would be derived (S-Plus User Manual, 2001).

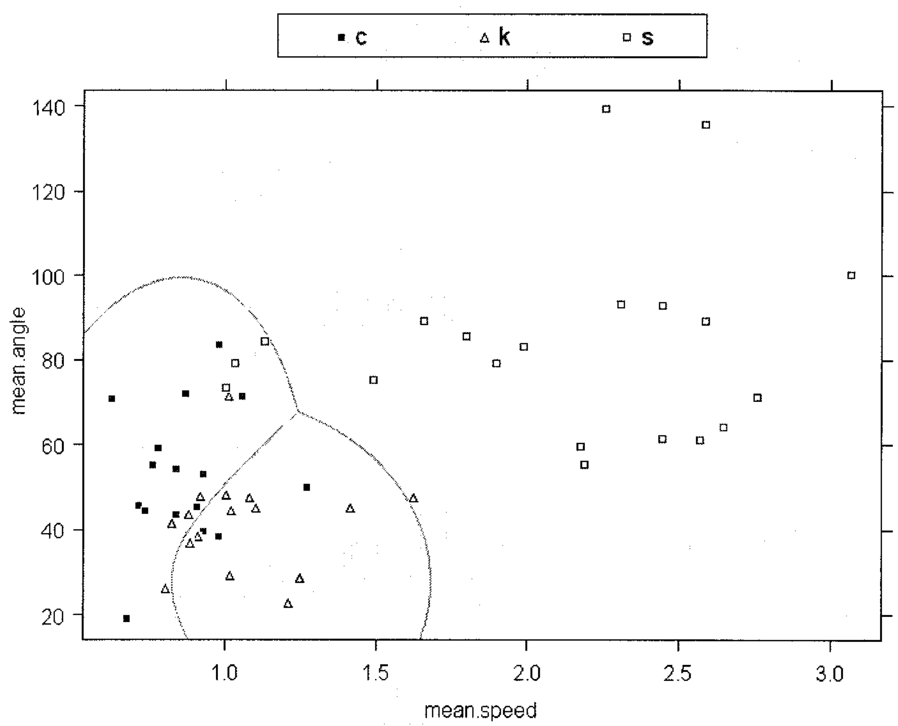
However, for small sample size, even if the various groups have unequal covariance matrices, linear discrimination generally outperforms quadratic discrimination models (Marks and Dunn, 1974; O'Neil, 1992). Many studies suggest that a minimum group size should be at least 20, or five observations per independent variables, which refers to all variables considered in the analysis, even if all of the variables do not enter into the discriminant model (such as stepwise estimation). The sample size (summarized in Table 3-4) is small according to these practical guidelines, so linear discriminant analysis is suitable for this study. Moreover a subset of the 2004 NS data was used to experiment with this method at the initial stages of this project. Mean speed and mean turning angle, and only these two attributes, are employed as independent variables. One reason is that a two-dimensional model is easy for observation, and the other is that both of the attributes can differentiate boat types well. The *p*-value of Box's M test for homogeneity of covariances is negligible, so theoretically a quadric model should be established. A linear model was also constructed for comparison. The results are shown in Figure 6-1, and the cross-validation outcomes are listed in Table 6-4. The overall error of the linear discriminant model is 0.22, which is less than that of the quadratic model. The conclusion is consistent with former studies and supports the decision of applying a linear model. Therefore, later on, when we mention discriminant analysis, it means linear multivariate discriminant analysis (MDA).

Table 6-4 Cross-Validation Table

	Linear Model					**Nonlinear Model**			
	Canoe	Kayak	Sailboat	Error		Canoe	Kayak	Sailboat	Error
Canoe	12	4	0	0.25	Canoe	12	4	0	0.25
Kayak	4	12	0	0.25	Kayak	5	10	1	0.38
Sailboat	3	0	17	0.15	Sailboat	3	0	17	0.15
Overall				0.22	Overall				0.26



(a) Linear Model



(b) Nonlinear Model

Figure 6-1 MDA Linear Model vs. Quadric Model for Three Boat Types

The linear combination for discriminating is termed the discriminant function in MDA, and the standard form is shown in the following Equation [6-2].

$$z_{jk} = a + w_1 x_{1k} + w_2 x_{2k} + ... + w_n x_{nk} \qquad [6\text{-}2]$$

where

z_{jk} = discriminant Z score of discriminant function j for object k

a = intercept

w_i = discriminant weight for independent variable i

x_{ik} = independent variable i for object k

From the above equation, it is seen that discriminant analysis multiplies each independent variable by its corresponding weight, where the rule is that the larger the coefficient is, the larger the discriminatory power is, and then a single discriminant Z score is calculated by summing these products. The group mean, referred to as the centroid, is determined by averaging the Z scores for each individual within a particular group. The centroid indicates the most typical location of any individual from a particular group, and can be tested for the hypothesis that the group means of a set of independent variables for two or more groups are equal by measuring the distance between these group centroids. Combined with the distribution of Z scores and cutting score, it can be verified whether the discriminant functions are good at predicting group membership. The cutting score is the criterion against which each object's discriminant score is compared to determine into which group the object should be classified. If the overlap is small as in Figure 6-2(a), the discriminant function separates the groups well; if the overlap is large like Figure 6-2(b), the discriminant function is poor.

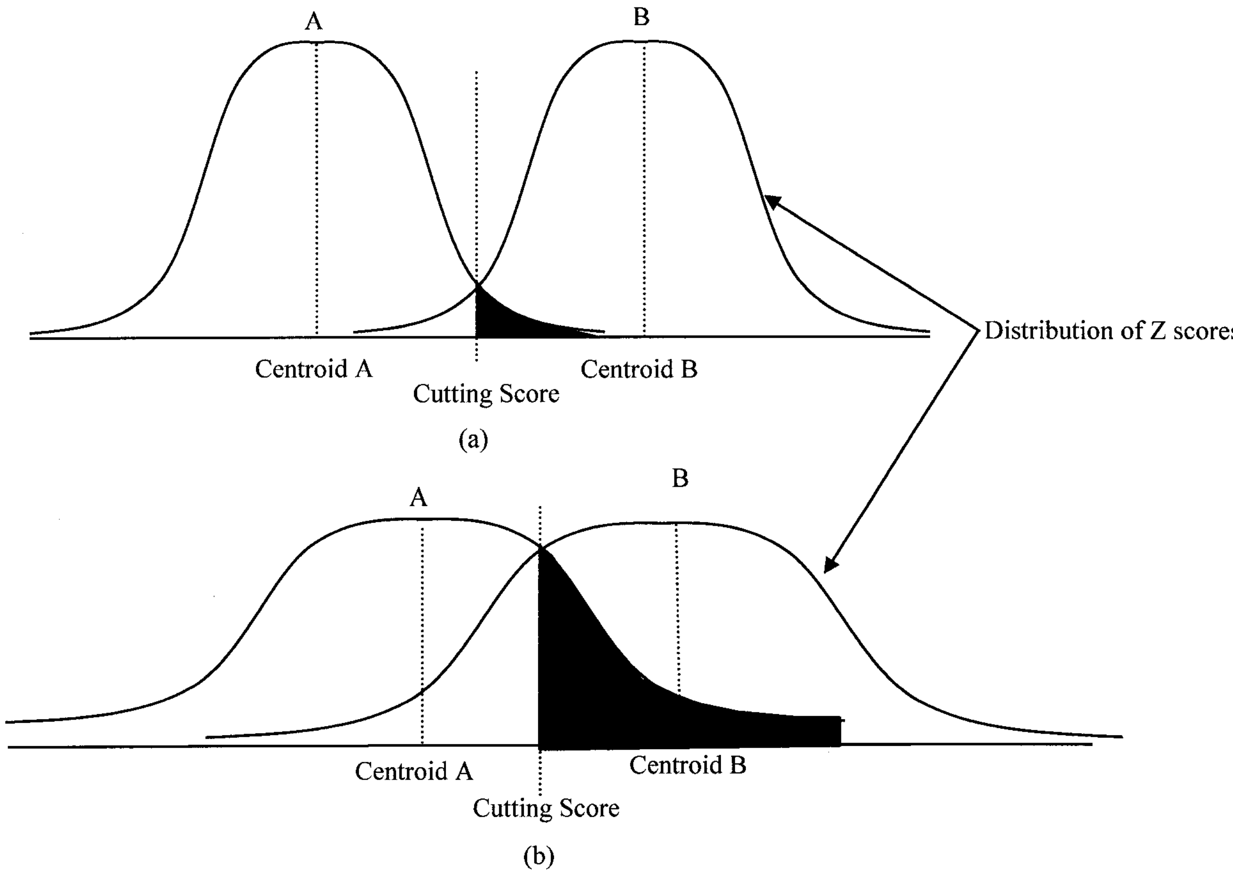


Figure 6-2 Good and Bad Discriminant Functions

The cutting score in Figure 6-2 is halfway between the two groups' centroids because the groups are of equal size. It can be calculated by the following Equation [6-3].

$$Z_{CE} = \frac{Z_A + Z_B}{2} \qquad [6\text{-}3]$$

where

Z_{CE} = critical cutting score value for equal group sizes

Z_A = centroid for group A

Z_B = centroid for group B

However, if the groups are not of equal size and Equation [6-3] for equal sample sizes is nevertheless applied, then the cutting score would be the left vertical straight line called the unweighted cutting score shown in Figure 6-3, which leads to big prediction error.

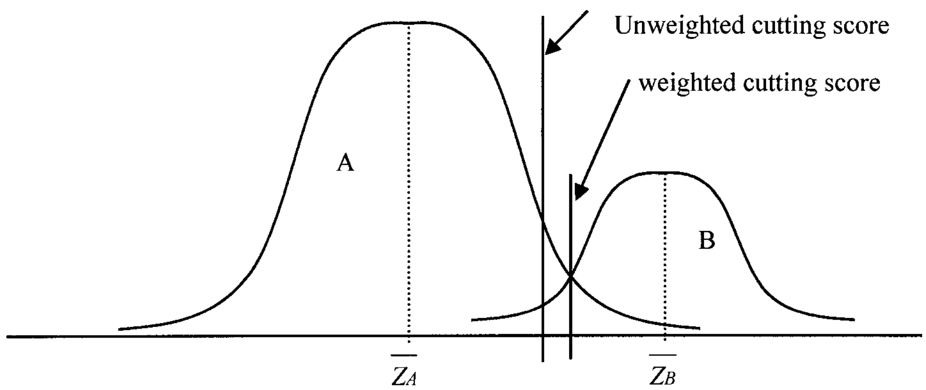


Figure 6-3 Cutting Scores with Unequal Sample Sizes

Assuming the different sample sizes represent the population proportions, an optimal cutting score should be calculated with the weighted average of the group centroids. Equation [6-4] would be applied in such a case, and the results are shown in Figure 6-3 as the weighted cutting score.

$$Z_{CU} = \frac{N_A Z_B + N_B Z_A}{N_A + N_B} \qquad [6\text{-}4]$$

where

Z_{CU} = critical cutting score value for unequal group sizes

N_A = number in group A

N_B = number in group B

Z_A = centroid for group A

Z_B = centroid for group B

The examples illustrated above all comprise two groups, which have been presented for ease of clear explanation and drawing. However, there are usually more than two groups in the analysis. It is obvious that only one discriminant function is needed for discriminating between two groups. If the Z score of an individual is less than the cutting score, it belongs to one type, otherwise it belongs to the other type. So to discriminate between M groups where the groups' sample size is N_i respectively, M-1 discriminant functions are needed. Figure 6-4 shows how to analyze among three groups using two discriminant functions. More dimensions are very difficult to illustrate on a diagram.

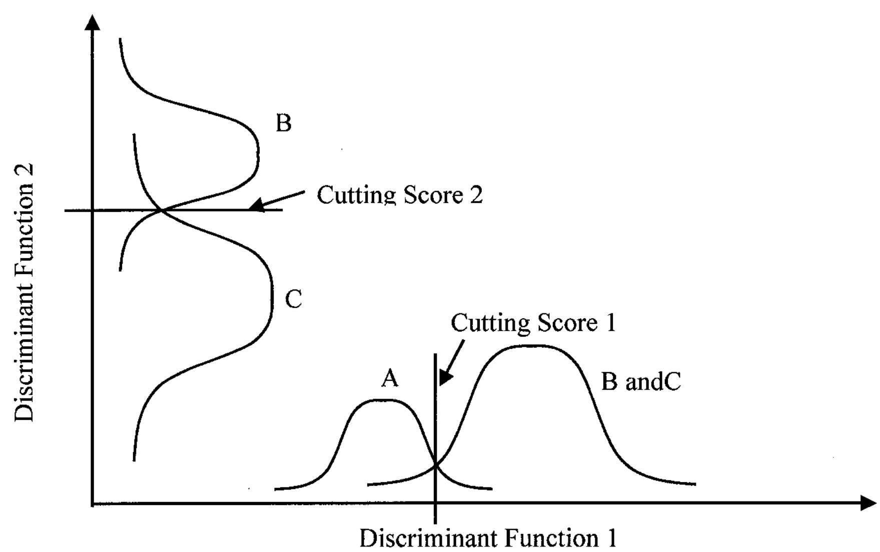


Figure 6-4 Discriminant Analysis among three Groups

Discriminant analysis not only yields the discriminant functions to distinguish between different groups, but also classification functions, which can be used in classifying observations into one of the groups. Equation [6-5] shows the classification functions for M groups. Each individual observation's independent variables are input into every function in Equation [6-5], leading to the conclusion that this individual could be classified in the group with the largest numerical value.

$$\begin{aligned}
&Group\ A = \alpha_1 + \beta_{11}x_1 + \beta_{12}x_2 + \ldots + \beta_{1n}x_n \\
&Group\ B = \alpha_2 + \beta_{21}x_1 + \beta_{22}x_2 + \ldots + \beta_{2n}x_n \\
&\vdots \\
&Group\ M = \alpha_2 + \beta_{m1}x_1 + \beta_{m2}x_2 + \ldots + \beta_{mn}x_n
\end{aligned} \qquad [6\text{-}5]$$

6.3.2 Research Design and Assumptions for Multivariate Discriminant Analysis

As this study aims to classify boat types only using GPS trajectories, the boat type is the dependent variable. The independent variables derive from seven possible independent variables: mean speed (MS), $\max_{1/20}$ speed, mean turning angle (MTA), total distance travelled (TD), aspect ratio (AR), coverage index (CI) and furthest distance from shore (DFS).

A key assumption underlying discriminant analysis is that of multivariate normality. It is very difficult to test this directly. However, by testing the univariate normality for all the variables, multivariate normality can be concluded, although this is not guaranteed (Hair et al., 1998). The Kolmogorov-Smirnov goodness-of-fit test is used to determine whether the empirical distribution of a set of observations is consistent with a random sample drawn from a theoretical normal distribution (Kutner et al., 2005). It is generally more powerful than the chi-square goodness-of-fit test for continuous variables. Applying the Kolmogorov-Smirnov test to the seven variables, the resulting p-values listed in Table 6-5 show that the normality requirement is met.

Table 6-5 *P*-values from Normality Tests

	Coastal	River	Combined
Total Distance Travelled (TD)	0.21	0.26	0.22
Mean Speed (MS)	0.24	0.24	0.13
$Max_{1/20}$ Speed	0.85	0.47	0.72
Mean Turning Angle (MTA)	0.97	0.81	0.40
Coverage Index (CI)	0.95	0.91	0.14
Aspect Ratio (AR)	0.20	0.98	0.33
Furthest Distance from Shore (DFS)	0.97	0.89	0.40

The second crucial assumption is that the predictor variables are not highly correlated with each other. Figure 6-5 is a scatter plot matrix of all seven aforementioned attributes plus maximum speed, demonstrating that mean speed, max speed and $max_{1/20}$ speed have a strong linear relationship ($|\rho| > 0.8$). Table 6-6 shows the correlation values and indicates that all the other variables are not strongly linear related. Of the three speed variables, the mean speed is kept as it is deemed likely to be the most reliable speed measure. Consequently, six independent variables are retained for MDA.

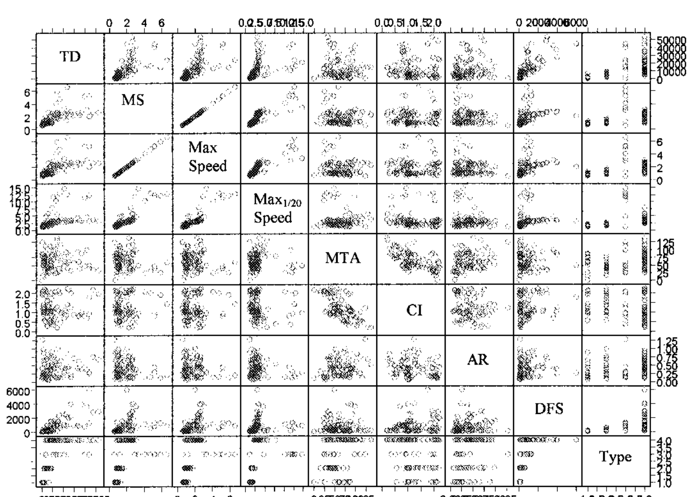


Figure 6-5 Scatter plot Matrix of the Coastal Data

Table 6-6 Correlations for the Coastal Data

	TD	**MS**	**Max1/20 Speed**	**Max Speed**	**MTA**	**CI**	**AR**	**DFS**
TD	1.00	0.63	0.63	0.40	-0.05	0.00	0.11	0.74
MS	0.63	1.00	1.00*	0.86*	-0.10	0.07	-0.17	0.38
Max1/20 Speed	0.63	1.00	1.00	0.86*	-0.10	0.07	-0.17	0.38
Max Speed	0.40	0.86	0.86	1.00	-0.05	-0.09	-0.18	0.13
MTA	-0.05	-0.10	-0.10	-0.05	1.00	-0.64	0.26	-0.02
CI	0.00	0.07	0.07	-0.09	-0.64	1.00	-0.20	0.20
AR	0.11	-0.17	-0.17	-0.18	0.26	-0.20	1.00	0.04
DFS	0.74	0.38	0.38	0.13	-0.02	0.20	0.04	1.00

* strong linear relationship

The third assumption for linear discriminant analysis is homoscedasticity. However, employing Box's M test shows that heteroscedasticity holds in the current situation, but a linear model is still adhered to in this case due to the relatively small sample size, as commented on earlier.

6.3.3 Multivariate Discriminant Analysis for the Coastal Area Data

To begin the assessment of the four-group discriminant analysis in the coastal area, Table 6-7 shows the group means for each of the independent variables, based on the associated 95 GPS recreational boating trajectories. Univariate ANOVA is used to assess the significance between means of the independent variables for the four groups. Using a significance criterion of 0.05, these tests indicate that four of the six independent variables show significant univariate differences among these four groups. Only Coverage Index (CI) and Aspect Ratio (AR) are not significantly different across the boat types.

Table 6-7 Group Mean Equality Test for the Coastal Data

Group Means for the Independent Variables: Boat Types						
Dependent Variable	**TD**	**MS**	**MTA**	**CI**	**AR**	**DFS**
Canoe	1 701	0.88	60.6	1.35	0.30	139
Kayak	5 269	1.09	45.3	1.40	0.41	257
Motorboat	18 699	3.98	60.7	1.21	0.34	718
Sailboat	18 3 63	2.06	71.7	1.12	0.47	1 409
Total	**12 523**	**1.83**	**62.7**	**1.23**	**0.41**	**8 54**
Significance level	**0.000***	**0.000***	**0.003***	**0.227**	**0.078**	**0.000***

* p-value ≤ 0.05

In summation, TD, MS, MTA and DFS are retained as significant and independent variables. Usually, different variables have different discriminant power.

Since one of the objectives is to determine which variables are the most efficient in discriminating across these four boat types, a stepwise computation procedure should be applied. Independent variables enter the discriminant function one at a time on the basis of their discriminant power. This can be done by maximizing Mahalanobis D^2 measure[2] between groups (Hair et al., 1998), which develops the best one-variable model, followed by the best two-variable model, and so forth until no other variables meet the desired selection rule. In this study, a minimum significance value of 0.05 is required for entry of a variable, and 0.10 for removal.

The mean speed (MS) was the first entering variable, as the single best discriminating variable. This initial variable, MS, was then paired with each of the other independent variables one at a time, and the variable best able to improve the discriminating power of the function in combination with the first variable was chosen, in this case the mean turning angle (MTA). Total distance travelled (TD) was included at the next stage followed by furthest distance from shore (DFS). But as the additional variable DFS entered, the previously selected variable TD was removed since the information it contained about group differences was available in some combination of the other variables. Although there are four independent variables which are significant and satisfy multivariate normality, only three of them – mean speed, mean turning angle and furthest distance from shore – are the most efficient variables in discriminating boat types. The following model [6-6] shows the discriminant functions, where Z_i is the

---

[2] Let $X=\{x_1,x_2,\ldots,x_n\}$ be a p-dimensional data set with size n, $MD=\sqrt{(X_i-\overline{X})^{T}V^{-1}(X_i-\overline{X})}$, $i=1,2,\ldots,n$. $\overline{X}$ *and* V are the usual sample mean and sample covariance matrix of the data set X.

discriminant score. There are three functions to differentiate four boat types, because given n groups to be discriminated, (n-1) functions are sufficient.

$$
\begin{aligned}
Z_1 &= 1.027MS + 0.582MTA + 0.114DFS \\
Z_2 &= -0.437MS + 0.496MTA + 0.958DFS \\
Z_3 &= 0.010MS + 0.779MTA - 0.465DFS
\end{aligned}
\qquad [6\text{-}6]
$$

Table 6-8 Summary of Discriminant Functions for Coastal Area

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.964	78.2	78.2	.814
2	.521	20.7	98.9	.585
3	.028	1.1	100.0	.165
Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 3	.216	138.8	9	.000
2 through 3	.640	40.4	4	.000
3	.973	2.5	1	.113

Structure Matrix			
	Function		
	1	2	3
MS	.846*	-.311	-.434
DFS	.299	.688*	-.662
MTA	.167	.414	.895*
TD[a]	.443	.250	-.528*

* Largest absolute correlation between each variable and any discriminant function

[a] This variable not used in the analysis

Table 6-8 shows the summary results from the discriminant functions for the coastal area. The eigenvalues provide information about the relative efficacy of each discriminant function. Nearly all of the variance explained by the model is due to the first two discriminant functions. Wilks' lambda is a measure of how well each function separates the cases into groups; smaller values indicate greater discriminatory ability of the function. The associated chi-square statistic tests the hypothesis that the means of the functions listed are equal across groups. A small significance value indicates that the discriminant function does better than chance at separating the groups. The test of Function 3 has a significance value of 0.113. Since this is greater than the typical threshold of 0.10 for Wilks' test, this function contributes little to the model.

The structure matrix in Table 6-8 shows the linear correlation between each independent variable and the discriminant Z score for each discriminant function. An asterisk (*) marks each variable's largest absolute correlation with one of the functions. Mean speed is most strongly correlated with the first function, and it is the only variable most strongly correlated with this function. Mean turning angle and total distance travelled are most strongly correlated with the third discriminant function, but this function is insignificant. These test results are consistent with the coefficients in the discriminant functions. Mean speed has the absolute largest coefficient in Function 1, equal to 1.027. The same observations can be found for the other two functions.

Discriminant analysis serves two roles. One is to quantify observations to create a model that explains the grouping of the given individuals, by forming discriminant functions as illustrated above. The other is to assign observations to the correct group, generating classification statistics. A linear function is defined for each group according to the following model [6-7]:

$$\begin{aligned} &canoe &&= -6.878+3.066\,MS+0.125MTA+2.739\times10^{-4}\,DFS \\ &kayak &&= -5.624+3.183\,MS+0.104MTA+2.293\times10^{-4}\,DFS \\ &motorboat &&= -27.071+9.611\,MS+0.192MTA+2.997\times10^{-4}\,DFS \\ &sailboat &&= -13.720+5.293\,MS+0.178MTA+1.983\times10^{-3}\,DFS \end{aligned} \tag{6-7}$$

Classification is performed by calculating a score for a given observation using each group's classification function, and then assigning the observation to the group with the highest score. As shown in Table 6-9, 16 out of 17 canoe trajectories are correctly classified, and 1 of them is misclassified as a kayak. 13 out of 21 kayak trajectories are properly recognized, except for 7 misclassified as canoes. Some research studies have combined these two boat types for particular analyses, but even though they represent similar boating behaviors superficially, they can still be discriminated with some certainty using three attributes extracted from their GPS trajectories. However, the lowest correct classification rate (61.9%) is for kayaks, which reflects the difficulty of distinguishing between canoes and kayaks. Sailboats can be distinguished well (91.5%), likely due to the particular nature of their movements. Furthermore, overall 84.2% of the original cases of boat trajectories are correctly classified.

Table 6-9 Classification Results for Coastal Data [a]

		Predicted Group Membership				Total
	TYPE	canoe	kayak	motorboat	sailboat	
Count	canoe	16	1	0	0	17
	kayak	7	13	0	0	21
	motorboat	2	0	8	0	10
	sailboat	2	1	1	43	47
Percentage	canoe	94.1	5.9	.0	.0	100.0
	kayak	33.3	61.9	.0	4.8	100.0
	motorboat	20.0	.0	80.0	.0	100.0
	sailboat	4.3	2.1	2.1	91.5	100.0

[a] 84.2% of original grouped cases correctly classified.

If it is believed that all boat types are equally utilized, then equal prior probabilities can be specified for all groups. This aspect implicitly affects the formation of the discriminant function by implying equal size populations. However, in this study the observed group sizes are used as the prior probabilities since the usage of various boat types varies quite a bit. For example, sailing is the most popular recreational boating activity around the Halifax coast. If equal prior probabilities for the groups were assumed, the correct classification rate would decrease to 78.9%.

6.3.4 Multivariate Discriminant Analysis for the River Area Data

One objective of this study is to examine whether there are significant differences in boating characteristics between two different geographic locations. Therefore, the GPS data from the river study area were examined, and the correlations show the same linear relationships as in the coastal study area. Maximum speed, maximum$_{1/20}$ speed and mean speed exhibit a strong linear relationships. As in the coastal scenario, only mean speed among the above three was retained for further analysis.

Table 6-10 shows the group means for each of the independent variables based on the 32 GPS recreational boating trajectories comprising the river analysis sample. These statistical tests indicate that variables MS, MTA, AR and DFS are independent and significantly different across boat types.

Table 6-10 Group Mean Equality Test for River data

Group Means for the Independent Variables: Boat Types						
Dependent Variable	**TD**	**MS**	**MTA**	**CI**	**AR**	**DFS**
Canoe	3 407	1.17	39.8	1.09	0.28	174
Kayak	4 688	1.01	35.1	1.19	0.20	210
Motorboat	15 465	3.78	53.2	1.54	0.26	307
Sailboat	14 029	1.92	91.0	0.89	0.47	601
Total	**11 039**	**2.47**	**52.9**	**1.29**	**0.28**	**312**
Significance level	**0.479**	**0.000***	**0.001***	**0.048***	**0.072**	**0.000***

* p-value ≤ 0.05

Applying stepwise discriminant analysis, the independent variables that were incorporated into the model remained the same as those identified in the coastal data analysis. The difference is that DFS has the highest discriminant power in the river study area as it was included in the model at the first step, while MS had the most predictive power in the coastal area. The discriminant functions for the river area are model [6-8]:

$$Z_1 = 0.348MS + 0.571MTA + 0.785DFS$$
$$Z_2 = 0.940MS - 0.100MTA - 0.293DFS \quad [6\text{-}8]$$
$$Z_3 = -0.038MS + 0.819MTA - 0.546DFS$$

Among these three discriminant functions, Functions 1 and 2 are significant for distinguishing between boat types as they cumulatively discriminated 99.7% of the cases (Table 6-11). Furthest distance from shore (DFS) has the strongest linear relationship to discriminant Function 1, mean speed (MS) to Function 2, and mean turning angle (MTA) and coverage index (CI) to Function 3.

Table 6-11 Summary of Discriminant Functions for River Area

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	2.850	70.3	70.3	.860
2	1.195	29.5	99.7	.738
3	.012	.3	100.0	.109
Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 3	.117	59.023	9	.000
2 through 3	.450	21.949	4	.000
3	.988	.331	1	.565
Structure Matrix				
		Function		
		1	2	3
DFS		.763*	-.306	.570
MS		.294	.952*	-.088
MTA		.523	-.160	-.837*
CI[a]		-.236	.051	.459*

* Largest absolute correlation between each variable and any discriminant function

[a] This variable not used in the analysis

Table 6-12 shows the classification results for the river data. Except for the two kayak trajectories misclassified as canoes, all the other trajectories are correctly classified. The overall classification accuracy is quite high at 93.8% which is a positive result, but this must be viewed in the context of the small sample sizes which may lead to less stable and generalizable results.

Table 6-12 Classification Result for River Data [a]

		Predicted Group Membership				Total
	TYPE	canoe	kayak	motorboat	sailboat	
Count	canoe	4	0	0	0	4
	kayak	2	6	0	0	8
	motorboat	0	0	15	0	15
	sailboat	0	0	0	5	5
Percentage	canoe	100.0	.0	.0	.0	100.0
	kayak	25.0	75.0	.0	.0	100.0
	motorboat	.0	.0	100.0	.0	100.0
	sailboat	.0	.0	.0	100.0	100.0

[a] 93.8% of original grouped cases correctly classified.

6.3.5 Multivariate Discriminant Analysis for the Combined Data

Combining the data from both areas was done to see how well the boats could be classified overall. First, comparing the means of the attributes for all four boat types between the coastal and river areas, it was found that geographic location does not make a difference in most cases (Table 6-3). Among these 28 tests, excluding max $_{1/20}$ speed due to its strong linear relationship with mean speed, only total distance travelled and mean speed of canoes, aspect ratio of kayaks, and furthest distance from shore of motorboats are significantly different between the two study areas (i.e. p-value < 0.05). This reinforces the preceding analyses whereby the stepwise discriminant analyses for both the coastal and river study areas retain the same independent variables, indicating that geographic location does not influence boating patterns drastically.

Pooling the data from both locations, a test of equality of group means across the four boat types is conducted (Table 6-13). Mean speed, mean turning angle and furthest distance from shore are significantly different across different boat types. Moreover, they are retained in the classification models for both of the study areas. Thus, applying these three metrics as independent variables for a discriminant analysis on the pooled GPS data from both study areas, 81.1% of the original trajectories were correctly classified. Furthermore, by including an explicit location variable which denotes a coastal versus river trajectory, the accuracy rate increased to 84.3%. The results were improved, but not much as is expected, since the influence of geographic location is not strong.

Table 6-13 Tests of Equality of Group Means

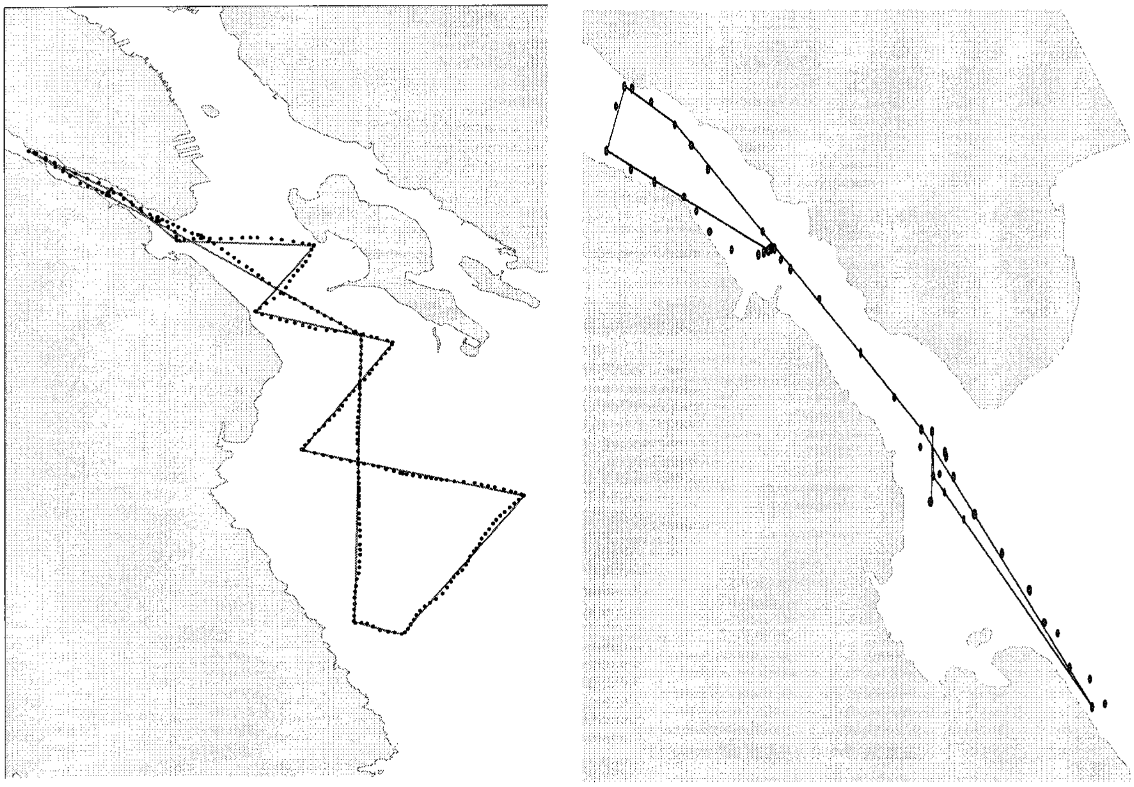
	Wilks' Lambda	Sig.
MS	.401	.000
MTA	.805	.000
DFS	.686	.000

6.3.6 Assessing Multivariate Discriminant Analysis

The predictive accuracy of the discriminant analysis is measured by the rate of correct classification. The value of this analysis can be further enhanced by verifying whether the percentage of correct classifications is significantly larger than that would be expected by chance.

Determining the classification rate that would result purely by chance based on the sample size of the largest group is referred to as the maximum chance criterion. If the rate of correct classification is larger than the maximum chance, then the discriminant analysis results are meaningful. In this study, the maximum chance is 41% based on the largest sample size of the pooled data, comprising the sailboats. A suggested criterion is that the classification accuracy should be at least one-fourth greater than that achieved by chance (Hair, 1998). The percentages of correct classification derived from the discriminant analyses discussed above are significantly larger than 51% (41%*1.25). This demonstrates that the models constructed in this study are useful for discriminating and classifying boat types based only on attributes extracted from GPS boating trajectories.

Finally, some misclassified trajectories are diagnosed to identify any characteristics of these observations that could be incorporated into the discriminant analysis for improving predictive accuracy.



(a) Characteristic Sailboat Trajectory (b) Uncharacteristic Sailboat Trajectory

Figure 6-6 Characteristic Sailboat Trajectory and Uncharacteristic Sailboat Trajectory

Figure 6-6 illustrates two sailboat trajectories. The dots represent the original GPS points. The solid line shows the dedensified trajectory, wherein unnecessary points are deleted, connecting only feature points to show where it turned. Although Figure 6-6(b) is also a sailboat trajectory, the discriminant analysis model misclassified it as a motorboat because it did not have the characteristic movements of a sailboat boating trajectory, especially the zigzag movements of a tacking boat. Hence Figure 6-6(b) movements are more characteristic of a motorboat for example high speed and long distance. Actually, by examining the auxiliary form for the recreational boating GPS trajectory collection, this sailor stated using the auxiliary motor during boating. The classification model evaluated that the probability of being a sailboat is 0.101 while that of being motorboat is as high as 0.899. This may have occurred because of the weather since sailboats cannot zigzag without wind, or because the boater may have been a novice who could not master the sailboat very well. Similar observations apply to other misclassified trajectories, which exhibit characteristics of other boat types, deluding the discriminant model into making a wrong decision. However, the decision is nevertheless reasonable based on the attributes of the boating trajectory presented. Therefore, to some degree, the error does not come from the classification model but from trajectories that did not display characteristic movements of their own boat types.

Moreover, due to the limited sample sizes, it was infeasible to reserve a portion of data for testing the model, and the remaining for training. An inadequate training dataset leads to unreliable models. Therefore, all of the data were included to construct the model to improve its robustness. Then, the model was applied to the same dataset to generate results. The results are compelling, and they have face validity, although the use of a testing dataset would serve to further establish the validity and generalizability of the model.

To sum up, in this chapter we employed the univariate analysis to examine the recreational boating trajectory patterns in terms of different boat types and different geographic locations. However, this approach cannot tell which of the independent variables account most for the differences across the four boat groups. Therefore, multivariate analysis has been applied to add further insight by providing discriminant functions and classification functions. The resulting attributes are essential for trajectory simulation, a component of spatial risk analysis (Pelot, 2005). Moreover the results of this study can help detect boat types through dynamic tracking and characterizing abnormal movements which could prove useful for coastal security, and perhaps engender better response planning in the case of potential maritime casualties. In addition, this study can provide insight on recreational boat exposure, a key determinant of recreational boating risks.

7. Simulation of Recreational Boating Trajectory

After succeeding at recreational boating trajectory pattern analysis, recreational boating traffic can possibly be investigated spatially and temporally. First of all, it is apparent that different patterns of trajectories should be simulated for different boat types because the results of the classification analysis showed that there are significantly discrepancies across the movement features of the four boat types. Secondly, the algorithm developed to calculate the turning angle and segment length for the purpose of discriminating can be applied to generate feasible movement points for simulated trajectories. This chapter attempts to develop an algorithm to simulate recreational boating trajectories in a GIS package, which would fill a gap in this area of application.

7.1 Simulation Purposes

Despite all the challenges of simulating a recreational boating trajectory, it cannot be avoided because of its crucial role in traffic analysis. Traffic patterns and incident patterns could assist decision-makers in locating emergency response resources and developing safety policy. One necessary ingredient for the traffic pattern modelling is trajectories, which requires simulation for recreational boating trajectories, since comprehensive realistic tracking data is generally not available.

Traffic patterns could be quantified as trip-based, as well as area-based, or more specifically grid-based, since gridding is the most common form for spatial delineation. It bears some merits such as simplicity of shape, homogeneity of area, and ease of geocoding[3] and georeferencing[4]. In grid analysis every grid cell represents a geographic

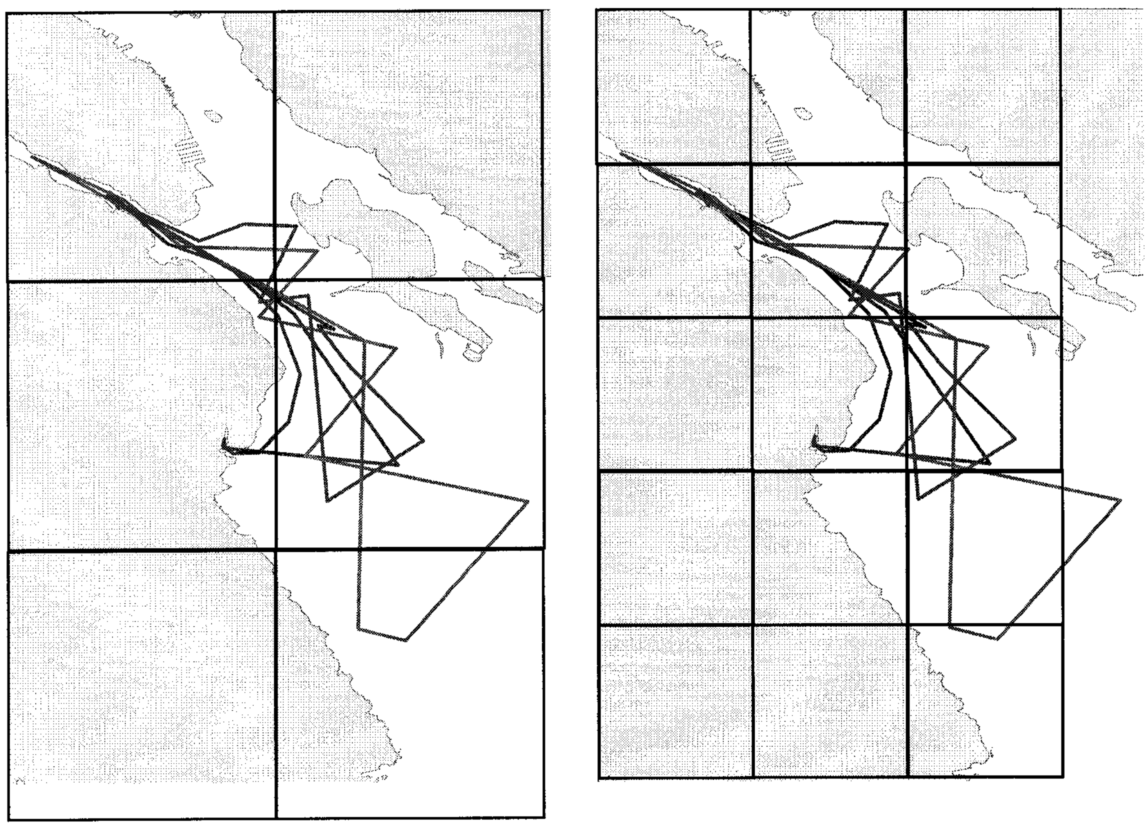
[3] Geocoding is the process of plotting geographic locations (identifiers) for data that are stored in tabular format.

unit for analysis and geographic units are regularly spaced, with the location of each unit referenced by row and column positions. Because gridded geographic units are of equal size and identical shape, area adjustment of geographic units is unnecessary and spatial properties of geographic entities are relatively easy to trace. Shields (2003) and Shahrabi (2003) used grids to estimate maritime vessel exposure density, as introduced in the literature review. In spite of the above advantages, the Modifiable Areal Unit Problem (MAUP) associated with the grid method should not be ignored, because this phenomenon means that different grid sizes could result in substantially different exposure density. This implies that the results may be unstable over the resolution and placement of the grid.

Given some estimate of recreational boating outings from specific departure points (marina, wharf, launch ramp, etc.), the spatial distribution of traffic might be quantified. One way is to predefine a gridding scheme, usually an integer such as 1 decimal degree square would be chosen, and then to estimate the density/frequency in each grid. However, establishing a suitable grid size poses problems. As shown in Figure 7-1, the MAUP can be examined to determine its effects on the analytical results. For example, with Gridding Scheme A (Figure 7-1(a)), five units have some traffic. Conversely, in Gridding Scheme B (Figure 7-1(b)), the same trajectories appear in eight grids. There are few solutions to resolve the MAUP problem, but rather than estimating traffic densities in each cell directly, if we establish trajectories first then we can grid them according to various criteria and evaluate the effect of the MAUP problem. Then sensitivity analysis on grid size and location can be explicitly conducted once the

[4] Georeferencing is the process of scaling, rotating, translating and deskewing an image to match a particular size and position. The word was originally used to describe the process of referencing a map image to a geographic location.

trajectories are simulated. That is why one objective of this study is to define trajectory patterns, then simulate them before performing spatial analyses.



(a) Griding Scheme A (b) Griding Scheme B

Figure 7-1 MAUP in Gridding

Moreover, to simulate recreational boat trajectories is also partly driven by the need for completeness of an existing GIS model of maritime traffic in a system named Maritime Activity and Risk Investigation System (MARIS) developed for the Canadian Coast Guard (Pelot, 2001). One of the analysis capabilities is to generate traffic lines for the overall risk model. It can simulate fishing and merchant shipping paths based on point data reported in existing comprehensive databases. Since such information is not available for recreational boating, by using the results of this study, the simulation of recreational boat paths can render the GIS model more complete and more powerful.

7.2 Simulation Methodology

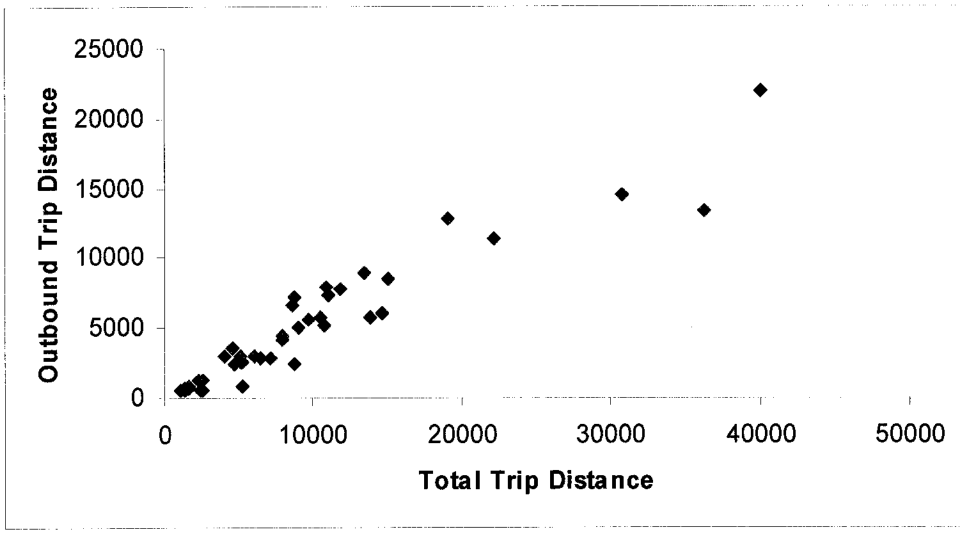
Normally, most recreational boating outings are round trip. Boaters came back to the place they started. For simulation trajectories, a one way trip is much easier to model than a round trip. After choosing the launch point and the destination, a possible trajectory could be simulated according to the geography. Simulating a round trip however must consider more factors. So this study places emphasis on simulating round trip recreational boating trajectories based on the actual data of original GPS-tracked trips in Nova Scotia (NS). Table 7-1 presents the sample size of round trips for every boat type. Because there are only five round-trip motorboat trajectories, the statistical results would not be reliable; therefore, motorboats are excluded from simulation. However, if the method for simulating other boat types' trajectories is effective, motorboat trajectories could be simulated after obtained a large sample.

Table 7-1 Sample Sizes of Round Trips in NS

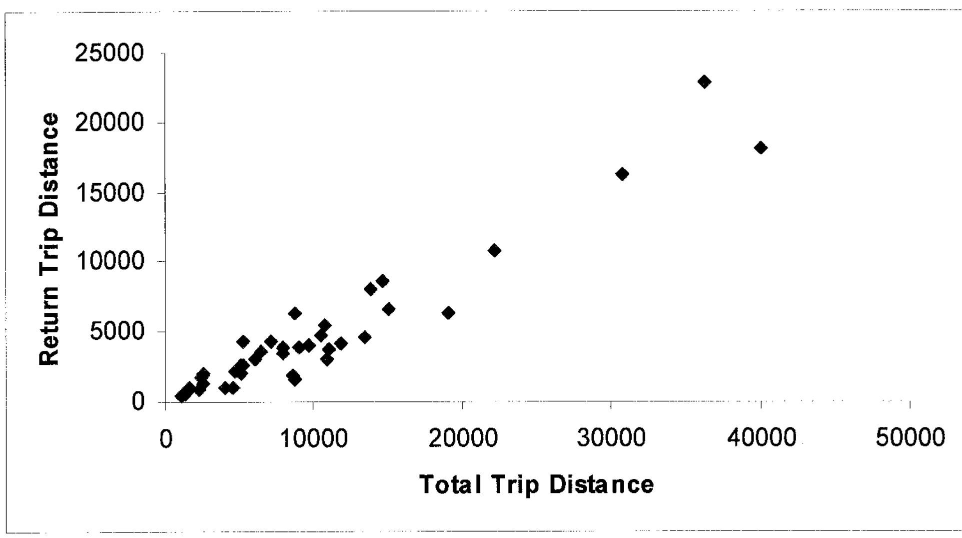
Boat Type	Sample Size
Canoe	11
Kayak	11
Motorboat	5
Sailboat	21

In order to reflect a round trip, outbound trips and return trips are defined correspondingly. An outbound trip refers to the path between the launch point and the furthest point from the beginning, while the remaining path is the return trip. This aspect is intrinsic to the complicated nature of a boating outing. It is a natural idea to generate points one after the other to create a trajectory, and it would seem easy to create a realistic representation based on the information of segment lengths and turning angles measured earlier. However, successfully spawning points does not guarantee that a reasonable recreational boating trajectory is formed. Thus, the attribute of furthest distance from launch point is the key for simulation, although this attribute is not used for discrimination and classification.

There are two versions of distance previously calculated. One is the furthest distance from shore, and the other is the furthest distance covered by the whole trajectory which is defined as the length of the bounding box. Different models (i.e. classification or simulation) demand different information such as which distinct attributes to extract. Moreover, all attributes computed during simulation would be dependent on dedensified trajectories, since it is not necessary to generate all the “noise” points for the simulated trajectory, and the crucial features of segment length and turning angle are also derived from dedensified trajectories.



(a)



(b)

Figure 7-2 Scatter Plots of Distances Travelled

From the scatter plots of Figure 7-2, it seems there are strong linear relationships between total trip distance travelled, outbound trip distance travelled and return trip distance travelled. The travelled distance is the summation of all segments between successive points, not the direct straight distance between the launch point and the furthest point from it. So instinctively, outbound trip distance travelled would not necessarily equal the return trip distance travelled. One may assume that the mood during the outbound trip is more relaxed, which leads to travelling a greater distance, a more complicated trajectory and more sightseeing, while the disposition during the return trip might be tired and rushing to get home so the trajectory could be relatively straight. As the case stands, applying linear regression, let total trip distance be the dependent variable y, and the outbound trip distance and the return trip distance be the independent variables x_1, x_2. The robust linear functions with R-Square equals 0.9085 and 0.9144 respectively are built as follows in meters.

$$y = 237.47 + 1.85x_1 \tag{7-1}$$

$$y = 1349.17 + 1.80x_2 \tag{7-2}$$

It may be concluded that return trip distance is shorter than the outbound trip distance because the intercept of Equation [7-2] is bigger than the one in Equation [7-1], which is consistent with our supposition. However, the supposition could be invalidated due to the slightly bigger coefficient of the independent variables x_1 compared to that of Equation [7-2]. The break-even point corresponds to 22,234 meters.

Considering the measure of the number of points instead of distance travelled, for example the number of total trip points, the number of outbound trip points and the number of return trip points as surrogates for total trip distance, outbound trip distance and return trip distance respectively, Figure 7-3 is the scatter plot revealing a strong linear relationship between these "number of points" variables. These observations provide insight for perhaps generating the total number of points for a certain boat type's trajectory, and determine where the threshold is between the outbound and return portions. Using the relations among the number of points, instead of the relations among the distances travelled, although both of the relations are strongly linear, the concern is that it is hard to generate a trip which is exactly equal to the prescribed distance for the outbound trip and the predetermined distance for the whole trajectory as well. If basing trip extent on the number of points however, this problem disappears. Moreover, the distance travelled could be used as a validation measure for the simulation, because this attribute does not then influence the outcome during the procedure.

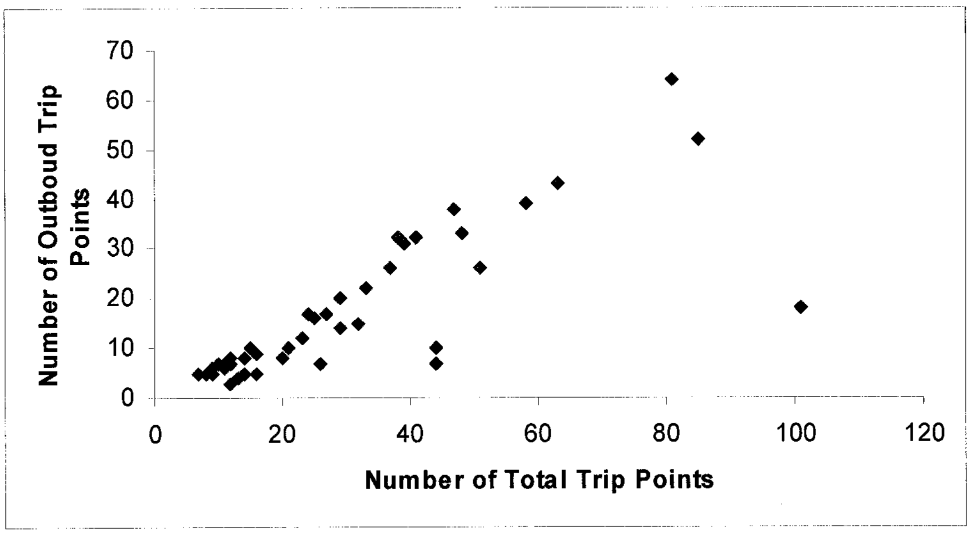


Figure 7-3 Scatter Plot of Number of Trip Points

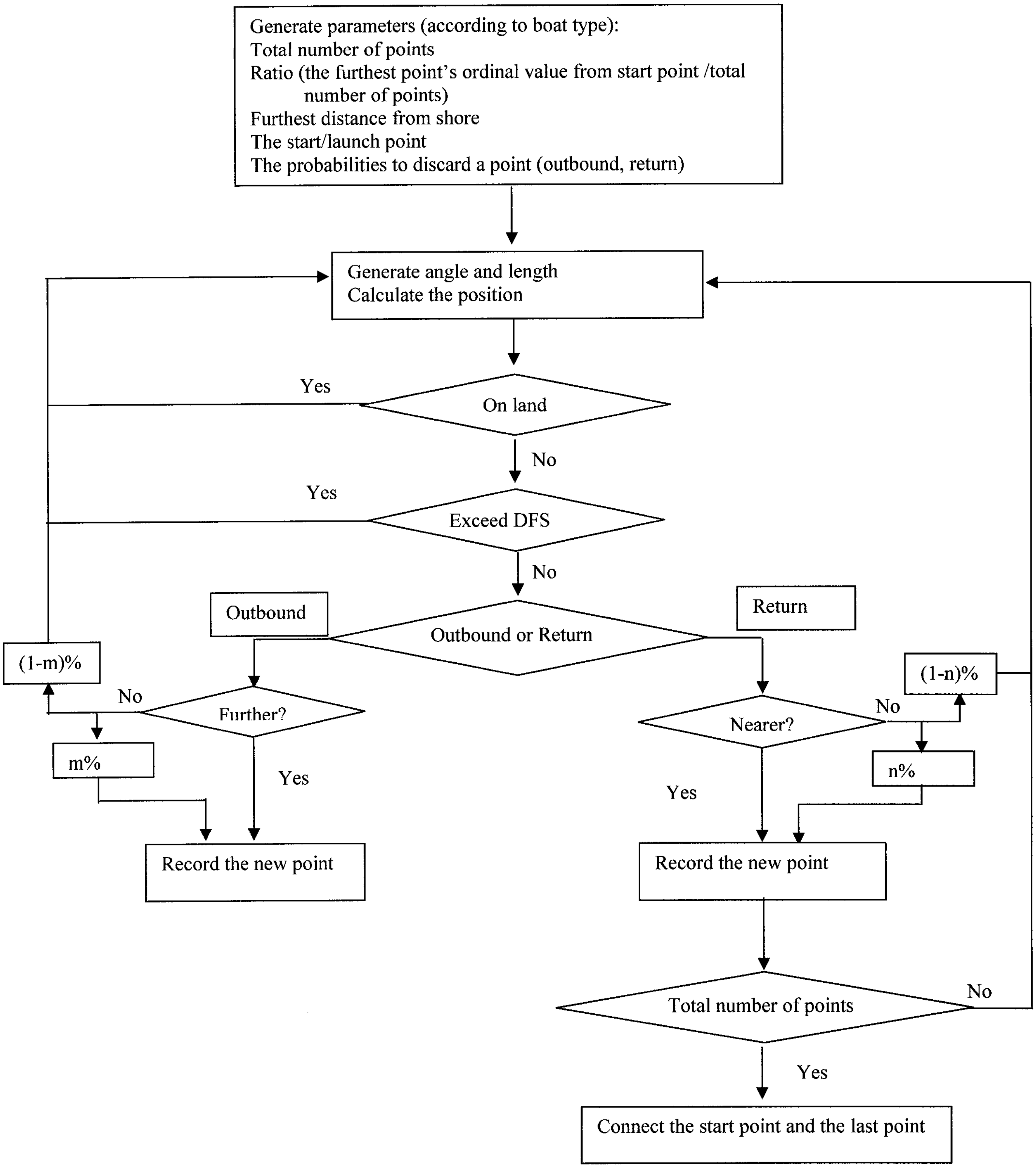


Figure 7-4 Flowchart of Simulation

Therefore, the algorithm proposed now shows that the round-trip difficulty has been solved. The simulation flowchart in Figure 7-4 illustrates the principal steps. There are two operations to accomplish the task of simulating a recreational boating trajectory. One is the generation of parameters and the other is feasibility inspection. Parameters include total number of points, ratio (the furthest point's ordinal value from start point/total number of points), furthest distance from shore, turning angle and segment length. The first three parameters need to be generated only once at the beginning of the simulation process, while the turning angle and segment length must be generated at least as many times as the number of points. Usually, more of these two attributes must be created because some of them might be discarded as unsuitable according to the inspection procedure. The furthest point from the launch position can be calculated based on the total number of points and the ratio, which means that the outbound trip and return trip can be determined. The furthest distance from shore is used for the inspection operation. If a newly generated position is beyond this limitation, the point will be discarded and a new one generated until a feasible position is created.

Aside from this inspection procedure, two other conditions should be tested. One is that a generated position is not on land, otherwise the point is rejected. The other condition-checking depends on whether the generated position is for the outbound trip or the return trip. During an outbound trip, the general tendency of the path should be further away relative to the launch point, until the furthest point from the start position is reached. Conversely, the general path tendency on the return trip is to be nearer and nearer to the launch point until finally it is reached, the characteristic of a round trip. However, not every point's distance from the launch point would be further than the preceding one in the outbound trip, and similarly not every point's distance to the start point would be closer than the preceding one in the return trip. To account for this phenomenon, the probability of keeping a generated outbound trip point if the distance to the launch point decreases is found (m% in Figure 7-4). Likewise, for the probability of keeping a generated point if it is on the return trip and the distance to the launch point increases (n% in Figure 7-4). This probability could be set as 0 under the assumption that sequential points should be further during the outbound trip (m=0) and nearer during the return trip (n=0). This assumption would be the easiest procedure, but the simulated trajectory might be too rigid to represent the complexity of actual recreational boating paths. The alternative is to estimate these probabilities from the sample data, from which the simulated trajectory would adopt more of a back and forth nature, more like a real path.

7.3 Simulation Preparation

Before applying the proposed algorithm, attributes from the dedensified trajectories are extracted including the total number of points, the furthest point's ordinal value from start point, furthest distance from shore, total distance travelled, coverage index and aspect ratio. All the information for the three boat types is listed the Table 7-2, Table 7-4 and Table 7-6 respectively with the mean and standard deviation calculated in the last two rows. Assuming that all the attributes are normally distributed, a probability distribution for each attribute can be created using the sample means and standard deviations (STDEVs) from these tables. For instance, the total number of points for a canoe trajectory can be generated from its corresponding function: ROUND {NORMAL (13.55, 5.7)}. The round function ensures that the generated number is an integer.

Similarly the following two functions are used to generate values for the ratio and the furthest distance from shore respectively NORMAL (0.56, 0.16) and NORMAL (153.59, 44.96). The furthest distance from shore provides a bound on this aspect when generating points for a simulated trajectory.

For the segment lengths and turning angles, Table 7-3, Table 7-5 and Table 7-7 can be referred to. The probability of a length or an angle being generated within an acceptable range is estimated separately for the outbound trip and return trip. The range of turning angles is 10 degrees to 180 degrees, and the range for segment lengths is broken into categories as shown in Table 7-3 and Table 7-5, with the lowest from 0 to 50 meters, and the highest range above 600 meters for canoes and kayaks which travel relatively short distances. For sailboats, the uppermost range for segment lengths starts at 6000 meters (Table 7-7). The probability distribution of generating a value within each range category is assumed to be uniform. If the segment length generated is beyond 600 meters or 6000 meters respectively for different boat types, a threshold has been established based on the sample data. The cutoff for canoes is 850 meters, 1120 for kayaks, and 7000 for sailboat outbound trips and 7500 for sailboat return trips. This data is included in the bottom part of the tables (Tables 7-3, 7-5 and 7-7), which also shows the probability of keeping a generated point whose distance to launch point decreases during the outbound trip and the probability of keeping a generated point whose distance to launch point increases during the return trip.

Table 7-2 Distributions of Once-Generated Parameters for Canoe

Trajectory Name	Number of total points	The furthest point's ordinal value from the start	Ratio	DFS	TD	CI	AR
c-0707-08-03	11	7	0.64	125.39	1348.17	1.19	.18
c-0707-13-03	12	8	0.67	187.82	1349.93	1.03	.17
c-0708-09-03	9	6	0.67	187.82	2243.75	1.29	.54
c-0708-13-08	16	5	0.31	180.02	2532.02	1.05	.44
c-0709-10-08	10	7	0.70	84.41	1129.30	1.09	.26
c-0713-10-08	20	8	0.40	187.82	1631.61	.98	.32
c-0720-09-04	14	8	0.57	88.31	1669.11	1.01	.17
c-0721-07-04	26	7	0.27	187.82	2367.28	.60	.15
c-0721-11-04	8	5	0.63	187.82	1643.47	1.44	.83
c-0727-08-08	7	5	0.71	176.12	1341.28	1.07	.21
c-0810-13-09	16	9	0.56	96.12	6046.85	1.05	.06
Mean	13.55		0.56	153.59	2118.44	1.07	0.30
STDEV	5.70		0.16	44.96	1382.46	0.21	0.22

Table 7-3 Distributions of Multiply-Generated Parameters for Canoe

Turning Angle			**Segment Length**		
Range (Degree)	Probability (Outbound Trip)	Probability (Return Trip)	Range (Meter)	Probability (Outbound Trip)	Probability (Return Trip)
0-10	0.01	0.07	0-50	0.21	0.20
10-20	0.19	0.19	50-100	0.25	0.13
20-30	0.19	0.16	100-150	0.11	0.11
30-40	0.12	0.07	150-200	0.15	0.05
40-50	0.03	0.02	200-250	0.12	0.11
50-60	0.08	0.04	250-300	0.00	0.05
60-70	0.03	0.08	300-350	0.05	0.16
70-80	0.08	0.04	350-400	0.01	0.12
80-90	0.05	0.04	400-450	0.03	0.02
90-100	0.03	0.01	450-500	0.00	0.00
100-110	0.05	0.08	500-550	0.02	0.01
110-120	0.02	0.02	550-600	0.00	0.01
120-130	0	0	600 more	0.04	0.03
130-140	0.04	0.07			
140-150	0	0.02	Longest length (meter)		850
150-160	0.03	0.04	Probability m		0.11
160-170	0	0.01	Probability n		0.12
170-180	0.05	0.03			

Table 7-4 Distributions of Once-Generated Parameters for Kayak

Trajectory Name	Number of total points	The furthest point's ordinal value from the start point	Ratio	DFS	TD	CI	AR
k-0616-11-03	21	10	0.48	622.94	6494.82	1.33	0.79
k-0701-13-11	9	5	0.56	74.66	1547.79	1.39	0.98
k-0708-10-02	25	16	0.64	355.63	4684.64	1.05	0.20
k-0722-11-01	29	14	0.48	185.87	5076.98	1.02	0.34
k-0727-15-06	32	15	0.47	168.31	7887.72	1.10	0.36
k-0807-15-01	15	10	0.67	68.80	5250.84	1.05	0.06
k-0807-18-01	23	12	0.52	158.56	2561.99	1.03	0.22
k-0808-15-03	47	38	0.81	123.44	3991.04	1.02	0.68
k-0808-15-11	9	6	0.67	181.97	5136.67	1.08	0.09
k-0724-14-05	44	7	0.16	304.90	8748.50	0.86	0.73
k-0731-14-04	44	10	0.23	148.80	2554.49	0.59	0.36
Mean	27.09		0.52	217.62	4903.23	1.05	0.44
STDEV	13.62		0.19	159.79	2226.41	0.21	0.31

Table 7-5 Distributions of Multiply-Generated Parameters for Kayak

Turning Angle			**Segment Length**		
Range (Degree)	Probability (Outbound Trip)	Probability (Return Trip)	Range (Meter)	Probability (Outbound Trip)	Probability (Return Trip)
0-10	0.08	0.12	0-50	0.21	0.27
10-20	0.20	0.18	50-100	0.18	0.15
20-30	0.13	0.14	100-150	0.08	0.07
30-40	0.15	0.14	150-200	0.15	0.02
40-50	0.12	0.06	200-250	0.04	0.14
50-60	0.14	0.04	250-300	0.07	0.07
60-70	0.03	0.06	300-350	0.06	0.04
70-80	0.07	0.04	350-400	0.01	0.02
80-90	0.03	0.02	400-450	0.03	0.03
90-100	0.02	0.02	450-500	0.03	0.04
100-110	0.00	0.02	500-550	0.03	0.03
110-120	0.00	0.02	550-600	0.02	0.00
120-130	0.01	0.03	600 more	0.10	0.12
130-140	0.01	0.03			
140-150	0.00	0.02	Longest length (meter)		1120
150-160	0.00	0.04	Probability m		0.09
160-170	0.01	0.01	Probability n		0.10
170-180	0.00	0.01			

Table 7-6 Distributions of Once-Generated Parameters for Sailboat

Trajectory Name	Number of total points	The furthest point's ordinal value from the start point	Ratio	DFS	TD	CI	AR
s-0623-18-09	48	33	0.69	523.43	10481.41	.58	.20
s-0721-18-09	39	31	0.79	1000.25	10940.08	.69	.40
s-0722-14-02	85	52	0.61	193.68	8941.28	.39	.55
s-0722-14-10	63	43	0.68	191.73	11062.43	.26	.50
s-0722-14-11	58	39	0.67	199.53	13411.40	.23	.68
s-0730-14-05	29	20	0.69	3621.80	40100.15	.73	.37
s-0804-16-04	15	10	0.67	1434.73	15092.49	1.12	.36
s-0804-17-05	24	17	0.71	1051.51	19134.50	.73	.33
s-0804-17-08	41	32	0.78	382.94	8536.13	.53	.48
s-0804-18-10	38	32	0.84	365.38	8774.30	.49	.48
s-0823-14-10	12	7	0.58	2494.09	22206.05	1.04	.40
s-0825-15-10	13	4	0.31	3939.12	36241.60	1.02	.38
s-0727-18-07	33	22	0.67	1407.88	9628.22	1.10	.92
s-0727-18-11	37	26	0.70	800.49	11863.48	.78	.62
s-0730-13-11	27	17	0.63	913.66	30753.14	.88	.40
s-0731-17-07	16	9	0.56	1531.15	7994.20	1.04	.37
s-0801-16-07	11	6	0.55	1567.76	10699.88	1.05	.78
s-0802-16-07	14	5	0.36	1566.54	14676.58	1.14	.99
s-0810-10-11	51	26	0.51	498.06	7095.85	.61	.26
s-0810-13-11	101	18	0.18	139.05	5206.36	.30	.84
s-0812-14-11	81	64	0.79	142.95	4630.29	.32	.70
Mean	39.81		0.62	1141.23	14641.42	0.72	0.52
STDEV	25.79		0.17	1080.25	9849.33	0.31	0.22

Table 7-7 Distributions of Multiply-Generated Parameters for Sailboat

Turning Angle			Segment Length		
Range (Degree)	Probability (Outbound Trip)	Probability (Return Trip)	Range (Meter)	Probability (Outbound Trip)	Probability (Return Trip)
0-10	0.01	0.03	0-50	0.10	0.08
10-20	0.05	0.08	50-100	0.13	0.13
20-30	0.12	0.12	100-150	0.10	0.08
30-40	0.05	0.06	150-200	0.08	0.06
40-50	0.03	0.07	200-250	0.06	0.05
50-60	0.06	0.06	250-300	0.06	0.05
60-70	0.03	0.02	300-350	0.05	0.07
70-80	0.09	0.03	350-400	0.03	0.03
80-90	0.04	0.08	400-450	0.02	0.01
90-100	0.07	0.04	450-500	0.03	0.01
100-110	0.07	0.06	500-550	0.02	0.01
110-120	0.05	0.05	550-600	0.00	0.02
120-130	0.04	0.07	600 -1000	0.14	0.14
130-140	0.08	0.07	1000-2000	0.13	0.15
140-150	0.04	0.02	2000-3000	0.01	0.04
150-160	0.06	0.05	3000-4000	0.01	0.03
160-170	0.05	0.05	4000-5000	0.01	0.01
170-180	0.08	0.05	5000-6000	0.02	0.01
			6000 more	0.02	0.01
Probability m		0.28	Longest length of Outbound Trip (meter)		6919.91
Probability n		0.23	Longest length of Return Trip (meter)		7427.95

7.4 Simulation Results and Limitations

Appling the proposed procedure and attribute information extracted from sample trajectories, the simulation of recreational boating trajectory for different boat types can be conducted. Figure 7-5 shows three simulated trajectories for a canoe, kayak and sailboat, randomly chosen from the outcomes. All of them launched from the same location. Different characteristics can be observed across the trajectories, which represent their particular boat types. For example, the canoe travelled the shortest distance, and lingered near the shoreline. Conversely, the sailboat ventured furthest to the open sea and had a remarkably serpentine track.

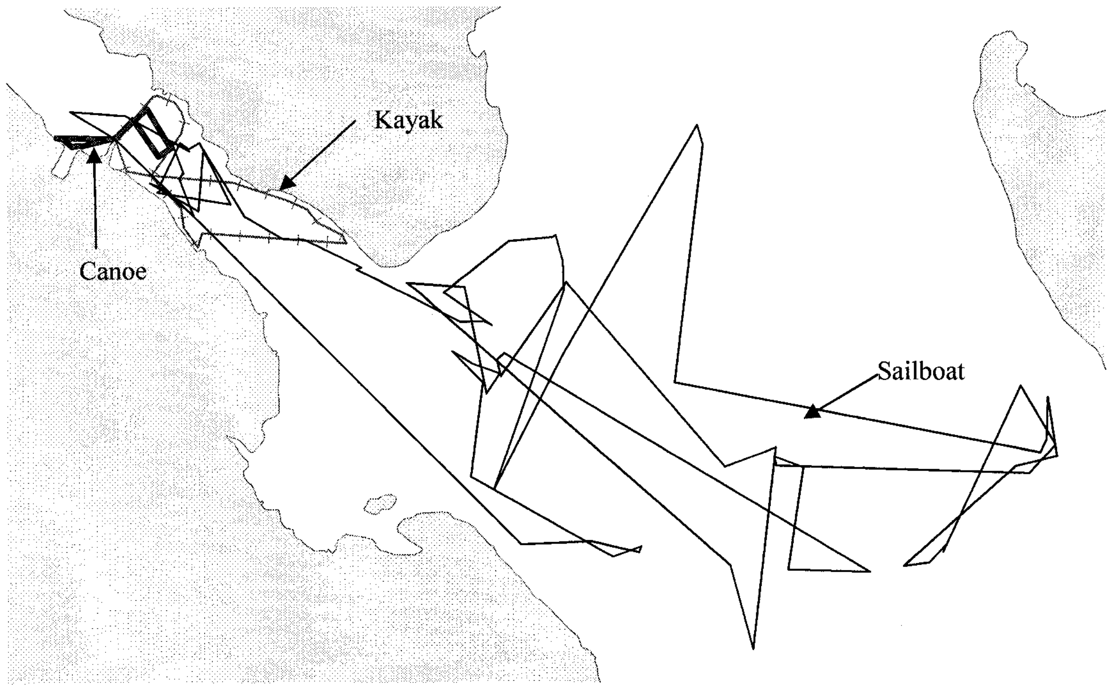


Figure 7-5 Simulated Trajectories for Different Recreational Boat Types

In order to test the effectiveness of the simulation, the classification model is not a suitable validation tool as planned in the objective section in chapter 1, because the mean speed is a variate included in the model for its powerful discriminating ability. However, there is no velocity information included in the simulated trajectory. Therefore, the idea of processing the simulated tracks through the classification model to see if they fall into the correct classes with high probability is not feasible. However, there are three attributes (i.e. TD, CI and AR) listed in the last three columns in Table 7-2, Table 7-4 and Table 7-6, which were not involved in the simulation process. Therefore, they can be used objectively to test certain characteristics of the simulated tracks. Assuming all of them are normally distributed, six sigma ($\mu \pm 3\sigma$) would cover a 99% confidence interval. On this basis, the total distance travelled, coverage index and aspect ratio are extracted from the simulated trajectories, and the performance of the simulation can be rated by the percentage of these attributes falling in the 99% confidence intervals displayed in Table 7-8.

To test the procedure, 500 trajectories are simulated respectively for each recreational boat type. Table 7-9 summarizes how many of the trajectories' attributes fall within acceptable ranges based on the sample data as described above. It is obvious that the results are not satisfactory as a large percentage of track's attributes are unrealistic, which is attributable to several reasons. The major difficulty is the very small sample size (i.e. the round trips at NS in Table 7-1), which leads to unstable means, and large standard deviations. Moreover, another consequence is that it is difficult to confirm a suitable theoretical distribution, therefore, normal distribution is assumed by default, although not a defensible choice. Another deficiency about the samples relates to their quality. Only a small portion of the sample represents the apparent key features for their specific boat types. Therefore, despite the reliance on several accurate measures of actual boat trajectories, this generation procedure is not adequate for realistic reproduction in most respects. One limitation is that the subjects targeted for data collection were not professionals, whose patterns may be more consistent and representative than the laymen's. Therefore, it is not surprising that some trajectories, and some track portions, are not reflective of characteristic movements, which tends to confound the analyses.

In terms of the of simulation algorithm, two improvements could be possibly made. One is the launch point selection. It has been noted that only one position was chosen. Actually, multi-launch points could be included in our application. However, the knowledge from the pattern analysis is that geography is not a significant factor to influence the boating movement features, and it can be foreseen that this change would not improve the simulation performance much. Another improvement would be to determine whether sequential turns are independent or not. If the direction's likelihood of being left or right depends on the direction of the preceding turn, then this could be accounted for in the simulation. Furthermore, more complex relationships could be sought, to see if turns are correlated with preceding turn angles. The application being used now randomly and independently chooses the direction and angle at each turn. These improvements on turn angle generation however would require much larger sample sizes and more characteristic trajectories which represent their own boat type's movement features.

Table 7-8 99% Confidence Intervals

	Canoe		**Kayak**		**Sailboat**	
	Low Interval	High Interval	Low Interval	High Interval	Low Interval	High Interval
TD (meter)	1,043	3,193	3,171	6,635	7,126	22,083
CI	0.91	1.24	0.88	1.21	0.49	1.00
AR	0.13	0.48	0.20	0.68	0.36	0.69

Table 7-9 Simulated Trajectory Attributes Relative to Sample Data Ranges

	Canoe	**Kayak**	**Sailboat**
TD	45.6%	38.6%	46.8%
CI	78.6%	66.6%	79.2%
AR	60%	71.6%	75.4%

There is also another interesting thing reflected in Table 7-9. The percentages of acceptable coverage index and aspect ratio for all the boat types are better than those of total distance travelled. As known, CI and AR have no units, while the unit of TD is meters. The simulation algorithm performs well with respect to the normalized attributes. This may support the argument that if sufficient samples are used to obtain all the parameters, the simulation procedure would work better.

8. Conclusions and Recommendations for Future Work

Indisputably, recreational boating is a popular activity among Canadians. Unfortunately, recreational boating accidents also account for many of the maritime incidents annually. In order to properly target accident prevention programs and improve search and rescue planning, it is necessary to acquire more knowledge about the patterns of recreational boating movements. This study broke through conventional research methodologies, employing GPS units to keep track of recreational boating trajectories. Algorithms were tailored to the aims of research for recreational boating. Univariate and multivariate statistics techniques were applied for discrimination and classification of four different recreational boat types at two different geographic locations. Based on the results of this study, simulation of recreational boating trajectories was attempted. The performance could be enhanced as more characteristic samples are obtained.

This study has set up protocols for a prototype of recreational boating movement pattern identification, from initial data acquisition to intermediate information extraction and final knowledge recognition. In the future, after more samples of more representative trajectories from more different locations are collected, it is predictable that some of the parameters of this prototype model might need to be adjusted. In other words, the specific parameters and results may be confined to the narrow scope of the study areas, but the procedures are believed to be robust enough for general implementation.

Moreover, there is another modification that could be made. At this point, the MDPA and the context-specific selection module are separate computer procedures, so an extension could be to combine them and produce an integral function inserted into the MARIS traffic modelling suite.

Ultimately, the overall challenge for a comprehensive risk representation involves creating a spatial and temporal model of recreational boating in Canada. Surveys such as the nationwide phone survey and the boat owner surveys have been completed, which resulted in information on boating frequency, location, and type of recreational boating in Canada. Since simulating trajectories for different recreational boat types is now within reach based on their movement patterns, the overall recreational boating traffic can be simulated as well. Therefore, incident rates calculated from maritime incidents relative to traffic distribution can be calculated based on location and type of activity. Moreover, prevention measures can be improved through a better understanding of critical factors such as the amount of exposure.

Another maritime risk associated with this research involves coastal security. Although most of the control efforts in that area are focused on large ships, there is also concern about illegal activities associated with small boat traffic, in particular on the Great Lakes couched between the U.S. and Canada. It is possible that such vessels may not travel in usual areas or move in characteristic ways. The results of this study could assist with the detection of abnormal movements, thus helping to mitigate coastal security risks as well.

References

Apps C.D., McLellan B.N., Kinley T.A. and Flaa J.P. (2001) Scale-dependent habitat selection by mountain caribou, Columbia Mountains, British Columbia, *Journal of Wildlife Management* 65: 65-77.

Beavan J., Moore M.C., Pearson C., Henderson M., Parsons B., Bourne S., England P., Walcott R., Blick G., Darby D. and Hodgkinson K. (1999), Crustal deformation during 1994-1998 due to oblique continental collision in the central Southern Alps, New Zealand, and implications for seismic potential of the Alpine Fault, *Journal of Geophysical Research*: 104 (11): 25233-25255.

Blanc L.A.L., Hashemi R.R. and Rucks C.T. (2001), Pattern Development for Vessel Accidents: a Comparison of Statistical and Neural Computing Techniques, *Expert Systems with Application* 20: 163-171.

Borgelt C. and Kruse C., *Graphical Models: Methods for Data Analysis and Mining*, John Wiley and Sons, New York, 2002.

Breen K.C. and Hejzlar Z. (2000), *Report: # 930658 Investigation and analysis of marine accidents*, 57-68.

Brooks C.J. (2001), *Survival in Cold Waters*, Transport Canada, Ottawa.

Buja A. (2000), Discussion of additive logistic regression: A statistical view of boosting, *Annals of Statistics* 28: 387–391.

Canadian Coast Guard: Search and Rescue Fisheries and Oceans Canada, 2003.

Canadian Legal Information Institute (2003) Electronic Source: Small Vessel Regulations: *Minimum Equipment Requirements for Pleasure Craft-Personal Flotation Devices* [online]
Available:
http://www.canlii.org/ca/regu/crc1487/

Chang C., Hwang S. and Buehrer D. (1991), Shape recognition scheme based on relative distances of feature points from the centroid, *Pattern Recognition* 24(11): 1053-1063.

Cromley R.G. and Campbell G.M. (1990), A geometrically efficient bandwidth line simplification algorithm, *Proceeding of the 4th International Symposium On Spatial Data Handling*, Zurich, 1: 77-84.

Dana P. H. (1997), Electronic Source: Global Positioning System Overview, *NCGIA Core Curriculum in GIScience*, [online] Available:
http://www.ncgia.ucsb.edu/giscc/units/u017/u017.html.

Devore J.L., *Probability and Statistics for Engineering and the Sciences*, Brooks/Cole Publish Company, Monterey, California, 1982.

Dicke M. and Burrough P.A. (1988), Using fractal dimensions for characterizing tortuosity of animal trails, *Physiological Entomology* 13: 393-398.

Donnelly M.P., Vaske J.J. and Graefe A.R. (1986), Degree and range of recreation specialization: Toward a typology of boating related activities, *Journal of Leisure Research* 18 (2): 81-95.

Douglas D.H. and Peucker T.K. (1973), Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, *The Canadian Cartographer*, 10(2): 112-122.

Dreyfus-Leon M.J. (1999), Individual-based modelling of fishermen search behaviour with neural networks and reinforcement learning, *Ecological Modelling* 120: 287-297.

Dykes J.A., 1999, An Introduction to Global Positioning Systems, Electronic Source [online] Available:
www.geog.le.ac.uk/jad7/gps/index.html.

Ebisch K. (2002), A correction to the Douglas-Peucker line generalization algorithm, *Computer & Geosciences* 28(8): 995-997.

Environics Research Group ltd, *Boating Safety Study*, Toronto, ON, 1998.

Erol B. and Kossentini F. (2003), Retrieval by Local Motion, *EURASIP Journal on Applied Signal Processing* 1: 41-47.

Fisher R.A. (1936), The use of multiple measurements in taxonomic problems, *Annals of Eugenics* 7: 179-188.

Forrester L. and Hotz S. (1998), Issues concerning the wearing of personal flotation devices: A literature review.

Golden F. and Tipton M. (2002), *Essentials of Sea Survival: Human Kinetics*, Human Kinetics Europe Ltd, UK.

Goodwin E.M. (1975), A statistical study of ship domains, *Journal of Navigation* 28: 328-334.

Great Lakes Commission (2000), Recreational boating and the great lakes-St. Lawrence region, Electronic Source [online] Available:
www.glc.org/advisor/00/recboating.pdf

Groff P. and Ghadiali J. (2003), Will it float? Mandatory PFD wear legislation: a background research paper, Toronto, ON: SMARTRISK.

Haering E.A. (1998), Booming Blackbird: GPS goes supersonic, *GPS World* 9(5): 24-32.

Hair J.F., Anderson R.E., Tatham R.L. and Black W.C., *Multivariate Data Analysis, Fifth Edition*, Prentice Hall, Upper Saddle River, New Jersey, 1998.

Hilliard C. and Pelot R. (2002), *MARIN Report: # 2002-07* Development of a Path Generation Algorithm for Simulating Maritime Traffic, Dalhousie University, Halifax, NS.

Hunter A. (2005), *Animal Movement Analysis: Matching Behaviour with Scale*, GEOIDE Summer School Presentation, Quebec City, Quebec, Canada.

Howarth R. and Buxton H. (1992), Analogical representation of spatial events for understanding traffic behaviour, *Proceedings of 10th European Conference on Artificial Intelligence,* pp.785-789.

Industry Canada: Recreational Boats 2005. Electronic Source [online] Available: http://strategis.ic.gc.ca/epic/internet/inrb-bp.nsf/en/home

International Boat Industry (2003). Electronic Sources: *Key Market Facts* [online] Available:
http://www.ibinews.com/ibinews/key_mkt_facts/index.htm

Jensen B.F., Refsgaard H.H.F., Bro R. and Brockhoff P.B. (2005), Classification of membrane permeability of drug candidates: A methodological investigation, *QSAR & Combinatorial Science* 24 (4): 449-457.

Johnson C.J., Parcker K.L., Heard D.C. and Gillingham M.P. (2002), Movement parameters of ungulates and scale-specific responses to the environment, *Journal of Animal Ecology* 71: 225-235.

Johnson N. and Hogg D. (1996), Learning the distribution of object trajectories for event recognition, *Image and Vision Computing* 14: 609-615.

Kelly D.P., Spillane W.J. and Newell J. (2005), Development of structure-taste relationships for monosubstituted phenylsulfamate sweeteners using classification and regression tree (CART) analysis, *Journal of Agricultural and Food Chemistry* 53 (17): 6750-6758.

Koenig, W.D., Van Duren, D. and Hooge, P.N. (1996), Detectability, philopatry and the distribution of dispersal distances in vertebrates, *Trends in Ecology and Evolution* 11: 514-517.

Kutner M.H., Nachtsheim C.J., Neter J. and Li W., *Applied Linear Statistical Models*, McGraw-Hill/Irwin, New York, NY, 2005.

Lambert M. and Santerre R. (2004), Speed check, performance monitoring with RTK GPS, *GPS World* 15 (2): 14-23.

Lentnek B. Van Donren C.S. and Trail J.R. (1969), Spatial behavior in recreational boating, *Journal of Leisure Research* 1 (2): 103-124.

Li W.J., Wu H.H. and Lin Z.Y. (2005), Fractal attenuation analysis of cartographic object on cartographic generalization, *Science of Surveying and Mapping* 30 (3): 21-24.

Li Z. and Openshaw S. (1992), Algorithms for automated line generalization based on a natural principle of objective generalization, *International Journal of GIS* 6(5): 373-90.

Lisowsi J., Rak A. and Czechowicz W. (2000), Neural network classifier for ship domain assessment, *Mathematics and Computers in Simulation* 51: 399-406.

Little J.J. and Gu Z. (2001), Video retrieval by spatial and temporal structure of trajectories, *Proceedings of SPIE - The International Society for Optical Engineering* 4315: 545-552.

Longley P.A. and Batty M. (1989), Fractal measurement and line generalization, *Computer and Geosciences*1 5 (2): 167-183.

Magee B. and Denys P. (2005), Accurate velocity assessment of a rowing skiff using kinematic GPS, *GEOMATICA* 59(2): 121-129.

Makris D. and Ellis T. (2002), Path detection in video surveillance, *Image and Vision Computing* 20: 895-903.

Mandelbrot B.B., *Fractals, Form, Chance and Dimension*, Freeman, San Francisco, 1977.

Mandelbrot B.B., *The Fractal Geometry of Nature*, Freeman, San Francisco, 1983.

Marino J.S. (1979), Identification of characteristic points along naturally occurring lines: An empirical study, *The Canadian Cartographer* 16: 70-80.

Marks S. and Dunn O.J. (1974), Discriminant function when covariance matrices are unequal, *Journal of the American Statistical Association* 69(346): 555-559.

McAvoy L.H., Lime D.W., Schatz C. and Pitt D.G., *Boating Safety in Minnesota*, University of Minnesota, 1990.

Martin T.P. and Bernfield Y. S. (1980), Effect of stroke rate on velocity of a rowing shell, *Medicine and Science in Sports and Exercise* 12(4): 250-256.

McCarthy P. and Talley W. K. (2001), Safety investments, behaviors and injury severity, *Applied Economics Letters* 33: 701-10.

McMaster R.B. (1987), Automated line generalization, *Cartographica* 24(2): 74-111.

Milne B.T. (1991), Lessons from applying fractal models to landscape patterns. Quantitative Methods in Landscape Ecology (eds M.G. Turner and R.H. Gardner), pp. 199-235. Springer-Verlag, New York.

Ministry of Natural Resources (1990), *Recreational Boating Fatalities in Ontario.*

O'Neill T.J. (1992), Error rates of non-Bayesian classification rules and robustness of Fisher's linear discriminant function, *Biometrika* 79:177-184.

O'Sullivan D. and Unwin D.J., *Geographic Information Analysis*, John Wiley & Sons Canada, Ltd, 2003.

Pasquariello G., Satalino G., Forgia V. and Spilotros F. (1998), Automatic target recognition for naval traffic control using neural networks, *Image and Vision Computing* 16: 67-73.

Pelot R., Shields R. and Sandblom C.L. (1997), *MARIN Report: #1997-01 The Development of a Model of the Demand for Search and Rescue Services Provided by the Canadian Coast Guard: Phase I,* Dalhousie University, Halifax, NS.

Pelot R. (2000a), *Recreational/tourism marine activity assessment in the Bay of Fundy, Nova Scotia,* Dalhousie University, Halifax, NS.

Pelot R. (2000b), *Newfoundland fishing incidents: perspective and analysis,* Dalhousie University, Halifax, NS.

Pelot R., Kendrick P., Shields R. and Deveaux J.P. (2000), *MARIN Report: #2000-05 Recreational/Tourism Marine Activity Assessment in the Bay of Fundy*, Dalhousie University, Halifax, NS.

Pelot R. (2001), *MARIN Report: # 2001-01 Maritime Activity and Risk Investigation System (MARIS): A GIS Tool for Maritime Traffic and Incident Analysis*, Dalhousie University, Halifax, NS.

Pelot R. and Wootton D. (2004), *MARIN Report: # 2004-10 Response Analysis Methodology for SAR Vessel Patrol Area Delineation, Maritime Activity and Risk Investigation Network*, Dalhousie University, Halifax, Canada.

Pelot R., Delbridge C., and Jakobsen M. (2004a), *MARIN Report: #2004-13 Recreational Boating Survey: Phase I*, Dalhousie University, Halifax, NS.

Pelot R., Jakobsen M., Delbridge C., and Wu Y. (2004b), *MARIN Report: #2004-14 Pan-Canadian Phone Survey of Recreational Boating Activity*, Dalhousie University, Halifax, NS.

Pelot R., Delbridge C. and Nickerson C. (2004c), *MARIN Report: #2004-12 Commercial Recreational Boating in Eastern Canadian Navigable Waters: Spatial & Temporal Evaluation*, Dalhousie University, Halifax, NS.

Pelot R. (2005), Spatial Analysis of Traffic and Risks in the Coastal Zone, *Coast GIS*, Aberdeen, Scotland.

Peterson K. (1991), The delineation of recreational boating markets: a multivariate approach, *Journal of Leisure Research* 23(3):209-224.

Poole K.G., Heard D.C. and Mowat G. (2000), Habitat use by woodland caribou near Takla Lake in central British Columbia, *Canadian Journal of Zoology* 78: 1552-1561.

Public Health Data Standards Consortium (2006), Electronic Sources: *External Cause of Injury Codes* [online]
Available:
http://www.quotiki.com/quote.aspx?id=9979

Ramer U. (1972), An iterative procedure for the polygonal approximation of plane curves, *Computer Graphics and Image Processing* 1: 244-256.

Rice D.P., KacKenzie E.J. and Jones A.S. (1989), Cost of Injury in the United States: A Report to Congress. Institute for Health and Aging, University of California and Injury Prevention Center, the John Hopkins University.

Shahrabi J. (2003), *Spatial and Temporal Analyses of Maritime Fishing and Shipping Traffic and Incidents,* Unpublished *Ph.D thesis*, Dalhousie University, Halifax, NS.

Shields Robert C. (2003), *Assessing Fishing Vessel Incident Factors in the Bay of Fundy for Response Planning,* Unpublished *Master thesis*, Dalhousie University, Halifax, NS.

Siderelis C., Brothers G. and Rea P. (1995), A Boating Choice Model for the Valuation of Lake Access, *Journal of Leisure Research* 27 (3): 264-282.

Sidman C, Fik T. and Sargent B. (2004), A Recreational Boating Characterization for Tampa and Sarasota Bays, Sea Grant, University of Florida, TP-130.

Skaloud J., Gyr M., Ladetto Q., Marcacci A., Lüthi P., Merminod B., Vetterli M., and Schutz Y. (2001), Athletic Analysis With Racing Heart, *GPS World* 14-18.

Smith J.N.M. (1974), The food searching behaviour of two European thrushes I, *Behaviour* 48: 276-302.

S-PLUS User's Guide, 6th Edition, Insightful Corporation, Seattle, Washington, 2001. pp. 83-106.

Statistics Canada (1999), Electronic Sources: *Average Age of Death* [online]
Available:
http://www.statcan.ca/Daily/English/020507/d020507b.htm

Statistics Canada (2003), *Census of Population: Immigration, Birthplace and Birthplace of Parents, Citizenship, Ethnic Origin, Visible Minorities and Aboriginal Peoples*, 2003.

Statistics Canada (2005), Electronic Sources: *household of Canada* [online]
Available:
http://www12.statcan.ca/english/census01/products/standard/popdwell/Tabel-PR.cfm

Tarrant M.A. and English D.B.K. (1996), A crowding-based model of social carrying capacity: a application for whitewater boating use, *Journal of Leisure Research* 28(3): 155-168.

The Canadian Red Cross Society (2001), *National Drowning Report, Visual Surveillance Report: 2001 Edition.*

The Canadian Red Cross Society (2003), *What We Have Learned: 10 Years of Pertinent Facts About Drownings and other Water-Related Injuries in Canada, 1991-2000.*

The University Kayak Club (2004), Electronic Sources: *Sea Kayaking Frequently Asked Questions* [online]
Available:
http://students.washington.edu/~ukc/FAQs/SKFAQ.htm#B

Torres C.T. (2001), *A Quantitative Analysis of Fishing Vessel Incidents in the Canadian Atlantic Region*, Unpublished *Master's Thesis*, Dalhousie University, Halifax, NS.

Turchin P. (1991), Translating foraging movements in heterogeneous environments into the spatial distribution of foragers, *Ecology* 72: 1253-1266.

United States Coast Guard (2005), Electronic Source: *Trajectory Analysis Specialist* [online] Available:
http://www.uscg.mil/d14/units/msohono/fog/fog4130.htm

United States Coast Guard DoT, *Boating Statistics: 2000.*

Uremovich C.J. (2002), *Risk Analysis of Merchant Shipping Traffic and Predictive Modelling of Vessel Casualty Rates in the Canadian Atlantic Region*, Unpublished *Master's Thesis*, Dalhousie University, Halifax, NS.

Uremovich C.J. and Pelot R. (2002), *MARIN Report: #2002-04 SISAR Maritime Incident Database Cleaning Methodology in the Canadian Atlantic Region*, Dalhousie University, Halifax, NS.

Vaughn J., Whyatt D., and Brookes G. (1991), A parallel implementation of the Douglas-Peucker line simplification algorithm, *Software—Practice and Experience* 21(3): 331-336.

Visvalingam M. and Whyatt J. D. (1991), The Douglas-Peucker algorithm for line simplification: re-evaluation through visualization, *Computer Graphics Forum* 9 (3): 213-228.

Voss R. (1984), Random fractal forgeries: from mountains to music, *Science and Uncertaint*: 69-86.

Walsh D. (1991), *Marine recreation: a major ocean industry sector,* OCEANS 91 conference: ocean technologies and opportunities in the Pacific for the 90's, Honolulu HI.

Wang W. (2000), The effects of state regulations on boating accidents and fatalities, *Applied Economics Letters*, 7: 373-378.

Water Incident Research Alliance (WIRA) (2005), Electronic Source [online] Available: http://www.waterincident.ca/index.html

West G. (1994), *Basic Cruising Skills*, Canadian Yachting Association, Kingston, Ontario.

Wiens J.A., Crist T.O., With K.A. and Milne B.T. (1995), Fractal Patterns of insect movement in microlandscape mosaics, *Ecology* 76 (2): 663-666.

Wikipedia (2005), Electronic Source: *Global Positioning System* [online]
Available:
http://en.wikipedia.org/wiki/Global_Positioning_System

Appendix 1 Questionnaire of MARIN's pan-Canadian Phone Survey on Recreational Boating Habits

Questionnaire

<u>Introduction (REQUIRED)</u>

'The following 15 questions are a portion of a research study being conducted at Dalhousie University by Dr. Ronald Pelot. These questions are about the recreational boating habits of Canadians. Responses to these questions will be published in research and then used to help make search and rescue programs more efficient. Your participation is completely voluntary and anonymous. Do you agree to participate in this research study?'

Consent: ○ yes ○ no

If the individual gives their consent, the caller will ask our survey questions.

<u>Instructions:</u>

The following are extra statements the caller can use if the respondents ask for clarification:

Q2. If the volunteer asks for clarification of 'Personal Watercraft' the caller can say: "An example of a PWC is a Seadoo".

Q7. If the volunteer ask for clarification of the word 'trip' the caller can say: "A trip is an outing on the water with LESS than a one hour break. This means that if you are out water-skiing and then you dock for 10 minutes to change gear and then go back out again, this would still count as one trip. Only if you stop for an extended period, for example if you stop to have an hour-long picnic and then go back out, could you consider this to be two trips. "

<u>Boating Survey for Phone</u>

1) What is the postal code of your primary residence?

2) What types of the following boats does your household own:
(Read all the following options and check all that apply)

1. Kayak/Canoe
2. Pedal Boat
3. Rowboat
4. Inflatable Raft
5. Zodiac
6. Sailboat without auxiliary power
7. Sailboat with auxiliary power
8. Motorboat with Cabin
9. Motorboat without Cabin
10. PWC (personal watercraft)
11. Pontoon
12. Other (no specify)
13. None

***<u>If none owned then skip the rest of the questionnaire</u>

3) Which ONE of the (following) boats owned by your household is used the most often: (Read all the following options and check ONE)

1. Kayak/Canoe
2. Pedal Boat
3. Rowboat
4. Inflatable Raft
5. Zodiac
6. Sailboat without auxiliary power
7. Sailboat with auxiliary power
8. Motorboat with Cabin
9. Motorboat without Cabin
10. PWC (personal watercraft)
11. Pontoon
12. Other (no specify)
13. None

4) When was the last time this (insert Q3 response) was used by you or a member of your household? (read all the following options in order)

○ less than 1 month ○ 1 month - 1 year ○ 1-5 years ○ more than 5 years

5) Please estimate the length of this boat (insert Q3 response). in feet.____/or metres___.

**take exact response given.

6) Please estimate the TOTAL horsepower of this (insert Q3 response) if that applies.

(Read all the following options in order)

skip this question for non-powered watercraft like canoe, kayak.

Below 10	50 to 99	200 and above
10 to 49	100 to 199	

7) How long are you typically on the water each trip?

(read all the following options in order)

Under 1 hour	3 to 6 hours	over 13 hrs
1 to 2 hours	7 to 12 hours	

8) In the MONTH that your household uses this (insert Q3 response) the most often, estimate how many trips are taken.

1 to 5 6 to 10 11 to 20 21 or more

9) How fast do you normally travel in this (insert Q3 response)?

Under 10 km/hr	25-55 km/hr
10-25 km/hr	above 55 km/hr

10) How far from the shoreline into deeper water do you normally travel in this (insert Q3 response)? (Record value and check appropriate units)

______ Km Miles meters feet

11) How far do you travel from your launch point? (Record value and check appropriate units)

________ Km Miles

12) How many people are typically on board?

1	3-4	more than 10
2	5-10	

13) What is the greatest number of people that have gone out on this boat at one time?

1	3-4	more than 10
2	5-10	

14) What is the name of the location where you boat most often in this (insert Q3 response)? (Could be a lake, river, town etc)

____

15) How far is this boating location, insert Q14 response here, from your primary residence in Kilometers?

____

Appendix 2 Recreational Boat Owners Survey

MARIN | DALHOUSIE University

Recreational Boat Owners Survey

Research Supported By: SEARCH AND RESCUE / RECHERCHE ET SAUVETAGE

In the table, describe each boat owned in your household, from the one YOU use most (A) to the one YOU use least (D
Use the categories below to specify boat type.

#	Type	#	Type
1	PWC	7	Sailboat (with auxiliary power)
2	Kayak/Canoe	8	Sailboat (without auxiliary power)
3	Pedal Boat	9	Motorboat (with cabin)
4	Zodiac	10	Motorboat (without cabin)
5	Rowboat	11	Inflatable Raft
6	Houseboat	12	Other (describe in table under "Type")

Boat	Type # (use key on left)	Length (feet)	Horsepower
A			
B			
C			
D			

The following questions relate to the boat YOU used most frequently in 2003 (Line "A" above).

What is the **usual** purpose of your boating trips? ☐Touring/Cruising ☐Fishing ☐Commuting ☐Partying ☐Hunting ☐Whitewater sports ☐Swimming/Diving/Snorkeling ☐Water-skiing/Tubing/etc. ☐Organized activity (tournament, regatta, etc.) ☐Other: ________________

Estimate the number of trips you take on this boat each month of the year

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec

How long are you **typically** on the water each trip? ☐Under 1hr ☐1 to 6 hrs ☐6 to 12 hrs ☐Over 12 hrs

What is your **maximum** time on the water for a trip? ☐Under 1hr ☐1 to 6 hrs ☐6 to 12 hrs ☐Over 12 hrs

How fast do you normally travel? ☐Under 10 km/hr ☐10-25 km/hr ☐25-40 km/hr ☐40-55 km/hr ☐Above 55 km/hr

How far from shore do you normally travel? _______ ☐kilometres ☐miles ☐metres ☐feet

How far do you travel from your launch point? _______ ☐kilometres ☐miles ☐metres ☐feet

How many people are typically on board? ________ What is the maximum capacity of this vessel? _________

What is the postal code of your primary residence? ___________

Where is your boat normally stored during boating season? (see map for grid number) _________

Is this: ☐your primary residence? ☐your cottage? ☐marina? ☐other?: ________________

Where do you boat most often? (see map for grid number) _______ Name of River/Lake/etc.: ___________________

What percentage of your trips on this boat do you do at this location? ______%

The following section asks about weather conditions under which you would decide **NOT** to go boating on this boat.
Under "Importance", please indicate how much each factor influences your decision **NOT** to go boating by circling a number on the scale **(1 being not important, 5 being very important)**

Weather Factor	I would decide NOT to go boating if:	Importance (Low … High)
Temperature	the temperature was below ______ ☐°C ☐°F **OR** above ______ ☐°C ☐°F	1 2 3 4 5
Wind Speed	wind speed was below _____ ☐km/hr ☐knots **OR** above _____ ☐km/hr ☐knots	1 2 3 4 5
Wave Height	waves were below _____ ☐metres ☐feet **OR** above _____ ☐metres ☐feet	1 2 3 4 5
Sun Exposure	it was ☐Sunny ☐Partly sunny ☐Cloudy	1 2 3 4 5
Fog	there was ☐Light fog ☐Moderate fog ☐Heavy fog	1 2 3 4 5
Precipitation	there was ☐No rain ☐Light rain ☐Moderate Rain ☐Heavy Rain	1 2 3 4 5

What is your: Gender? ☐Male ☐Female

Age? ☐Under 25 ☐25 to 34 ☐35 to 44 ☐45 to 54 ☐55 to 64 ☐65 or over

Annual Household Gross Income? ☐Under $40,000 ☐$40,000 to $59,999 ☐$60,000 to $79,999 ☐$80,000 to $99,999 ☐$100,000 to $119,999 ☐$120,000 and above

Appendix 3 Auxiliary Form for Recreational Boating GPS Trajectory Collection

Recreational Boating Survey Form

Research Supported By: SEARCH AND RESCUE / RECHERCHE ET SAUVETAGE

This survey is to be filled out by the primary boater.

Please turn ON the GPS receiver when you BEGIN your boating trip, but NOT beforehand.

Turn OFF the receiver when you complete your boating trip.

For each boat used, please fill out a separate survey form.

Your gender: ☐ Male ☐ Female

Your age: ☐ Under 25 ☐ 25 to 34 ☐ 35 to 44 ☐ 45 to 54 ☐ 55 to 64 ☐ 65 or over

Education level (highest level completed):

☐ Grade School ☐ High School ☐ Some College ☐ College Graduate ☐ Graduate Degree

What type of boat are you using today?

☐ PWC ☐ Houseboat ☐ Sailboat (with auxiliary power) ☐ Canoe
☐ Zodiac ☐ Inflatable Raft ☐ Sailboat (without auxiliary power) ☐ Kayak
☐ Rowboat ☐ Pedal Boat ☐ Motorboat ☐ Other:__________________

If this boat is powered, what is its horsepower? _______

What is the boat's length? _______ ☐ m ☐ ft

How many years have you been boating on this type of vessel? _____

How often do you boat on this type of vessel during the peak season?

☐ Less than once a month ☐ Once or twice a month ☐ Once a week ☐ More than once a week

How would you describe your experience/comfort level with this type of vessel?

☐ Very Confident ☐ Somewhat confident ☐ Average ☐ Have minimal experience ☐ No experience

When was the last time you took a boating course? ______________________________

Do you have a Pleasure Craft Operator Card? ☐ Yes ☐ No

How fast do you normally travel?

☐ 0-2 km/h ☐ 2-5 km/h ☐ 5-10 km/h ☐ 10-25 km/h ☐ 25-40 km/h ☐ 40-55 km/h ☐ Above 55 km/h

What is the *main* purpose of your trips aboard this vessel? (Please choose only *ONE*)

☐ Cruising ☐ Partying ☐ Water skiing/Tubing/etc. ☐ Lessons/Teaching
☐ Whitewater sports ☐ Hunting ☐ Swimming/Diving/Snorkeling ☐ Guided Tour
☐ Commuting ☐ Fishing ☐ Organized activity (tournament, regatta, etc.)
☐ Other: ______________________________

Use the reverse side of this page to fill out specific information about your trips on this boat using the GPS unit.

Fill out a new trip section for any break in boating activity that is more than 1 hour (either on-land or anchored).

- **For example, if you are on a canoeing trip and stop on land for a 15 minute pit stop and then continue on, this is considered one trip.**
- **If, on your canoeing trip, you stop and camp overnight and continue on in the morning, this would be considered two separate trips.**

Trip 1
What is the number on the back of your GPS receiver? ______
Trip date: ____
Trip *start* time? ________________
Trip *end* time? ________________
Describe the weather today:
Temperature: _____ ☐ °C ☐ °F
Is it foggy? ☐ Yes ☐ No
How would you describe the precipitation? ☐ None ☐ Overcast ☐ Light Drizzle ☐ Light Rain Showers ☐ Heavy Rain Showers
What is the wind speed? ________ ☐ mph ☐ km/h ☐ knots
If you are operating a sailboat, did you use auxiliary power during this trip? ☐ Yes ☐ No
If so, when and for how long? ____

Trip 2
What is the number on the back of your GPS receiver? ______
Trip date: ____
Trip *start* time? ________________
Trip *end* time? ________________
Describe the weather today:
Temperature: _____ ☐ °C ☐ °F
Is it foggy? ☐ Yes ☐ No
How would you describe the precipitation? ☐ None ☐ Overcast ☐ Light Drizzle ☐ Light Rain Showers ☐ Heavy Rain Showers
What is the wind speed? ________ ☐ mph ☐ km/h ☐ knots
If you are operating a sailboat, did you use auxiliary power during this trip? ☐ Yes ☐ No
If so, when and for how long? ____

Trip 3
What is the number on the back of your GPS receiver? ______
Trip date: ____
Trip *start* time? ________________
Trip *end* time? ________________
Describe the weather today:
Temperature: _____ ☐ °C ☐ °F
Is it foggy? ☐ Yes ☐ No
How would you describe the precipitation? ☐ None ☐ Overcast ☐ Light Drizzle ☐ Light Rain Showers ☐ Heavy Rain Showers
What is the wind speed? ________ ☐ mph ☐ km/h ☐ knots
If you are operating a sailboat, did you use auxiliary power during this trip? ☐ Yes ☐ No
If so, when and for how long? ____