UNITED STATES POWER SQUADRONS®
SAIL AND POWER BOATING

P.O. Box 30423, Raleigh, North Carolina 27622
1504 Blue Ridge Road, Raleigh, North Carolina 27607
888-367-8777 • www.usps.org

5 December 2014

Dear Susan Tomczuk:

Enclosed in this package are documents and items to close out Grant 1102.26, *Reducing Recreational Boating Accidents through Advanced Risk Analysis*. United States Power Squadrons is pleased to have conducted and concluded this grant project successfully.

The documents contained in this close-out package are listed below:

- Abstract
- Executive Summary
- Project Dissertation
- SF 425

If any, all products and materials produced by this grant comply with the USCG Grant Product Guidelines and include the appropriate statement/logos indicating that the product was completed under a U.S. Coast Guard grant.

Sincerely,

Mary Catherine Berube,
Executive Director

ABSTRACT

Ernest G. Marshburn. BEYOND HUMAN FACTORS: EXAMINING THE UNDERLYING DETERMINANTS OF RECREATIONAL BOATING ACCIDENTS WITH SPATIAL ANALYSIS AND MODELING (Under the direction of Dr. Ronald L. Mitchelson) Institute for Coastal Science and Policy, Ph.D. Coastal Resources Management, December, 2014.

Recreational boating has grown in popularity in recent decades, accompanied with increased accidents resulting in property damage and personal injury. Some 5,000 recreational boating accidents are reported annually, ranking recreational boating as a leading cause of transportation accidents, second only to automotive.

Recent research suggests that recreational boating accidents stem from multiple factors. In contrast, public perception and public policy overwhelmingly attribute boating accidents to human error, e.g., operator drug or alcohol use or lack of experience. This dissertation offers a comprehensive perspective on recreational boating accidents by exploring human, technological, and environmental factors that most influence these accidents. This level of inclusiveness is absent from previous research.

The conceptual model developed in this dissertation is derived from general accident theory that integrates spatial and temporal qualities of recreational boating (and boating accidents) from satellite imagery, on-the-water boater surveys, and federal boating accident data. Data were assembled for two distinctive research sites, Sandusky, OH and Tampa, FL. Analyses of these data depended, in part, upon various forms of spatial statistics, e.g., hot spot analyses. The boating accident model developed in this study uses the multivariate negative binomial model to analyze accident count data aggregated to 0.25 mi^2 grid cells. The result is a synthetic model with improved parameter estimates and predictive capability compared to previous

boating accident research. Key risk factors contained in the final model clearly represent human (operator experience), technological (boat speed and length), and environmental (boat density and channel character) dimensions.

This research has important societal impact, i.e., to public officials faced with the allocation of limited resources. In particular, this research emphasizes the concentrated nature of boating risk in time (seasonality, day of week, time of day) and in space (shoals, channels, fixed facilities). These features should guide the timing and the placement of mobile law enforcement capacity as well as the location of operation centers near high risk boating sites. Finally, this work emphasizes the need for investigations of additional sites and the importance of including remotely sensed data to complement survey data in studies of recreational boating accidents.

BEYOND HUMAN FACTORS: EXAMINING THE UNDERLYING DETERMINANTS

OF RECREATIONAL BOATING ACCIDENTS WITH SPATIAL ANALYSIS

AND MODELING

A Dissertation

Presented to

the Faculty of the Institute for Coastal Science

and Policy, Coastal Resources Management

East Carolina University

In Partial Fulfillment

of the Requirements for the Degree

Ph.D. Coastal Resources Management

Primary Concentration in Social Science & Coastal Policy

Secondary Concentration in Coastal & Estuarine Ecology

by

Ernest G. Marshburn

December, 2014

BEYOND HUMAN FACTORS: EXAMINING THE UNDERLYING DETERMINANTS

OF RECREATIONAL BOATING ACCIDENTS WITH SPATIAL ANALYSIS

AND MODELING

by

Ernest G. Marshburn

DISSERTATION DIRECTOR: _____
(Ronald L. Mitchelson, Ph.D.)

COMMITTEE MEMBER: _____
(Yong Wang, Ph.D.)

COMMITTEE MEMBER: _____
(Thomas R. Allen, Ph.D.)

COMMITTEE MEMBER: _____
(Claudia L. Jolls, Ph.D.)

COMMITTEE MEMBER: _____
(Lynn L. Harris, Ph.D.)

COMMITTEE MEMBER: _____
(Jason S. Brinkley, Ph.D.)

Director, Ph.D. COASTAL
RESOURCES MANAGEMENT;
INSTITUTE FOR COASTAL
SCIENCE AND POLICY: _____
(Hans Vogelsong, Ph.D.)

DEAN OF THE
GRADUATE SCHOOL: _____
(Paul J. Gemperline, Ph.D.)

DEDICATION

Special thanks go to my family, David Marshburn (brother), Tammy Burroughs (sister), and extended family members (Slater Burroughs, Lorrie Marshburn, Margaret and Baxter Matheson, Matt Burroughs, Michelle Burroughs, Jeremy Marshburn, Eric Marshburn, and Adam Marshburn) without whose love, support, and steadfast belief; I could not have turned this dream into a reality.

Most especially, my continuing gratitude and special thanks go to my wife, Dianne Marshburn. She has been patient, understanding, supportive, and most importantly encouraging during a time when my research focus took priority over family time. This critical and timely support provided the foundation required for the successful completion of an area of recreational boating research that we are both passionate about.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: THE RECREATIONAL BOATING ACCIDENT PARADOX

Recreational boating has grown in popularity in recent decades, accompanied by boating accidents resulting in property damage and personal injury. Some 5,000 recreational boating accidents are reported annually, ranking recreational boating as a leading cause of transportation accidents, second only to automotive transportation. These increases are a consequence of changing industrial, recreational, and residential development along U.S. coastal regions (Burger & Leonard, 2000). Some research suggests that this growth is dependent upon the economy and demographics while others found this increase to be highly dependent upon fuel prices (Bristow & Bennett, 1995). Yet the relative stabilization of fuel prices and a visit to any waterbody during the summer continues to showcase enormous recreational boating use (Bristow & Bennett, 1995). For example, over the past ten years there has been a rapid increase in the use of personal watercraft (PWC), up from 89,710 in 1987 (United States Coast Guard, 1988) to over 960,761 in more recent years (S. Tomczuk, USCG Office of Auxiliary and Boating Safety, pers. comm.). This distinction is important as PWCs can travel faster than conventional boats and operate in shallow coastal (and inland) waters. PWCs account for approximately 10% of the registered boats in many coastal boating areas, yet they are responsible for 35% of all boating accidents (Burger & Leonard, 2000).

Recent research suggests that recreational boating accidents (also referred to as simply boating accidents in this dissertation) stem from multiple factors among several general dimensions. In contrast, public perception and public policy overwhelmingly attribute boating accidents to human error, e.g., operator drug or alcohol use or lack of experience. The research reported in this dissertation offers a more comprehensive

perspective on recreational boating accidents by exploring the human, technological, and environmental factors that most influence these accidents. This level of inclusiveness is absent from previous research.

This research also offers broader societal relevance by illustrating utility to public officials faced with the deployment of limited boating law enforcement resources. In particular, the work emphasizes the concentrated nature of boating risk in time (seasonality, day of week, time of day) and in space (shoals, channels, fixed facilities). These features may help guide the timing and the placement of mobile law enforcement assets as well as the location of operation centers. This research emphasizes the need for expanded study across varied research sites and the importance of including remotely sensed data in these endeavors to complement survey data.

Boating Safety

To reduce the number of recreational boating accidents and injuries, educational measures to improve boat operator safety practice are critically important. While a majority of recreational boaters generally practice safe boating, simply being a member of a boating association or boating in protected waters does not in and of itself imply that a boat operator will exhibit safe boating behaviors (Virk & Pikora, 2011).

One primary issue for water-based resource managers is that increased boating use can create conflicts onsite. Therefore, it is important to understand the factors that most influence boating demand to help marine managers administer resources in more efficient ways (Bristow & Bennett, 1995). Typically, recreation demand models have considered socio-economic demographics and site characteristics to forecast recreation

2

choice. More recently, researchers have defined recreation boating choices as a function of activity (purpose of trip), travel patterns, or as resource distribution and amenities (Bristow & Bennett, 1995).

Boating-related travel is of interest at the regional scale, which has concentrated on boating participation at various park resources. For example, recreational boaters have been thought of as individuals that tend to participate close to home (Bristow & Bennett, 1995). Furthermore, travel behavior research found that recreational boaters exhibited the greatest on-water distance decay function (Bristow & Bennett, 1995; Lentnek & Doren, 1969). Distance decay is defined as a rate of decreased use over a distance (or travel tme). Beyond resource characteristics, nearby facilities such as harbors and marinas may attract recreational boaters (Bristow & Bennett, 1995; Lentnek & Doren, 1969). Recreational boaters have four options (access points) in order to gain access to waterways: marinas, dry storage facilities, private docks, or ramps. The steady increase in the number of these access points and their uses have led to congestion, particularly on the weekends and holidays (Sidman & Flamm, 2001; Sidman, Fik, & Flamm, 2002; Sidman Fik et al., 2005). The potential importance of this congestion on the management of recreational water resources is manifested not only in the field of recreation but also within the field of marine transportation accident research (Schuhmann & Schwabe, 2004).

Under-Reporting

Evidence suggests that non-fatal boating accidents are under-reported for various reasons (e.g., lack of knowledge or deliberate choice), despite the fact that

federal regulations require such reporting (Hoedt, Timmons, & Marmo, 2003). Accident rates are biased downward by under-reporting, which in turn masks the true magnitude of such risks (Hassel, Asbjørnslett, & Hole, 2011; Hoedt, Timmons, & Marmo, 2003; Loeb 1994; Sidman & Fik, 2005; Psarros, Skjong, & Eide, 2010; Yamamoto, Hashiji, & Shankar, 2008;). If boating accident risks are underestimated, "then society may fail to take appropriate actions to control or reduce these risks" (Hoedt, Timmons, & Marmo, 2003). The social and property costs of recreational boating accidents are also an essential cost-benefit consideration involving mitigation strategies that encourage intermodal transportation accident costs comparisons; e.g., the costs of recreational boating accidents as compared with aircraft or automobile accidents (Hoedt, Timmons, & Marmo, 2003; Hassel, Asbjørnslett, & Hole, 2011; Loeb, 1994; Psarros, Skjong, & Eide, 2010; Sidman & Fik, 2005; Yamamoto, Hashiji, & Shankar, 2008).

Under-Reporting Problem

Recreational boating accident under-reporting is not simply a problem for authorities trying to improve maritime transportation safety legislation. Risk management companies, insurers, and other entities use these reported statistics in risk and accident analysis as well (Hassel, Asbjørnslett, & Hole, 2011).   The validity of historical data may be undermined by data uncertainty with upper limits on under-reporting estimated to be as high as 74% (Psarros, Skjong, & Eide, 2010).

Prior studies suggest that unreported boating-related incidents account for nearly 50% of all recreational boating accidents (Psarros, Skjong, & Eide, 2010). Only a few states appear to approach a perfect boat accident reporting level with the best representing about 94% of actual incidents (Hassel, Asbjørnslett, & Hole, 2011).

Regulatory changes in marine transportation are often distilled from past experience, mainly related to accidents. Decisions on how to improve recreational boating safety are thus frequently carried out on an ad-hoc basis, influenced by public pressure, governmental interest, or both. For these reasons, improvements in recreational boating accident risk assessment and mitigation have largely been based on the products of a reactive regulatory approach. This process involves regulatory changes that are imposed to prevent the reoccurance of specific types of boating accidents (Hassel, Asbjørnslett, & Hole, 2011; Psarros, Skjong, & Eide, 2010). However, the common belief is that large numbers of recreational boating accidents go unreported is frequently confirmed by first-hand reports from commercial operators (Devanney, 2008).

By comparison, land and air safety have long held public attention and have been the subject of continuous improvement and reporting for the last few decades. These improvements have been achieved, at least in part, through research on data quality. However, the recreational boating sector has observed few similar initiatives in this area (Alsop & Langley, 2001; Amoros, Martin, & Laumon, 2006; Harris, 1990; Sciortino et al., 2005; Yamamoto, Hashiji, & Shankar, 2008). Based on data from varying marine transportation sectors, it must be assumed that boating accident under-reporting is a continuing problem (Psarros, Skjong, & Eide, 2010).

Traffic Safety and Policy

Today, as a result of evolving transportation models, we are developing a better understanding of recreational boating accidents while reducing injuries and loss of life. Boating traffic strategies and safety are a critical issue for many state and federal

5

agencies'. The tremendous social and economic costs associated with boating accidents have led many governmental authorities and researchers to establish safety management programs aimed at improving safety performance (Sawalha & Sayed, 2006). This trend aligns with the identification of high risk areas on highways, which leads to intervention measures and mitigation strategies by highway traffic officials (Erdogan et al., 2008). Geographical information system (GIS) technology is an increasingly popular research tool for visualizing and analyzing this highway accident data. The range of benefits resulting from the application of GIS in highway accident risk analysis has created a broad literature that forms a basis for better understanding recreational boating accident risk (Erdogan et al., 2008). GIS is a spatial data analyses and visualization platform that is essential to exploring spatial and non-spatial relationships in many fields (Erdogan et al., 2008).

Historical Perspectives

Before the development of land-based transportation systems, marine waterways served as the "super highways" of developing nations enabling the efficient movement of people, goods, and services (Black, 2003). While transportation (marine and land-based) is an essential component of modern civilization, accidents are a functional element of the effectiveness of those systems (Black, 2003). For example, as the number of recreational boats within a marine transportation system increase, the risk of boating accidents increases correspondlingly. This relationship and governmental recognition of it can be evidenced by the formative stages of United States (U.S.) waterway development, i.e., through the establishment of laws, policies, navigation aids intended to guide vessel operators, and the like (Black, 2003).

The first U.S. maritime laws were enacted by Congress in 1789 (Rizzo & Hoedt, 2010). However, in response to growing pressure to respond to boating accidents, Congress enacted the Motorboat Act of 1910 (Motorboat Act of 1910: Public Law 61-201 36 Stat.462., 1910). The Motorboat Act of 1910 established the first federal carriage requirements for marine vessels, e.g., lights, whistles, life preservers, and a way to extinguish fires aboard motorboats up to 65 feet in length (Motorboat Act of 1910: Public Law 61-201 36 Stat.462., 1910). By 1940, continuing recreational boating growth coupled with rapidly increasing numbers of recreational marine accidents, Congress amended the Motorboat Act of 1910 and initiated stronger boating safety requirements (Perry & Timmons, 2003; Hoedt, Timmons, and& Marmo, 2003; House Report 110-899 - HONORING THE HERITAGE OF THE COAST GUARD, 2008). The Federal Boating Act of 1958 standardized the capture of recreational boating accident data. It also tasked the U.S. Coast Guard (USCG) with collecting information on numbered boats as well as analyzing boating accident reports (Federal Boating Act of 1958: Pub. L. 85-911, Sept. 2, 1958, 72 Stat. 1754, 1958). During the 1960s, Congress encouraged increased use of USCG statistics to study additional boating regulations leading to the Federal Boat Safety Act of 1971. By 1990, the USCG as well as many states were enacting boating-under-the-influence (BUI) laws with some states exploring mandatory boat operator education requirements (Perry & Timmons, 2003; Hoedt, Timmons, & Marmo, 2003; House Report 110-899 - HONORING THE HERITAGE OF THE COAST GUARD, 2008; Rizzo & Hoedt, 2010). Today, government authorities believe that current recreational boating laws and programs have reduced boating accident risk to a level where further reduction is not possible (Rizzo & Hoedt, 2010).

However, as many of these boating laws and programs were based on descriptive statistical techniques, the doorway to futher boating accident risk reduction and understanding may spring from deeper analytical research designs and quantative boat accident models. This is the foundation established for this dissertation.

Significance of the Study and Broader Impact

History suggests that the primary focus of boating accident risk mitigation is focused on human factors. Isolated consideration of just human factors is simply too restrictive to capture the complexity of the actual recreational boating accident domain. Conjoint consideration of human, technological, and environmental factors is needed to better understand the highly integrated and complex influences of these factors on boating accidents.

The literature surrounding recreational boating acidents has historically focused on specific accident risk factors that must be present before a recreational boating accident can occur. Previous research supports this theory when suggesting that boating accidents are largely a product of opportunity, access to a recreational boat, presence of a navigable body of water, and varying levels of law enforcement to monitor boating behaviors. This argument is supported by evidence suggesting that most recreational boating accidents occur in close proximity to the boat operator's home port, i.e., neighborhood effects. On the other hand, most previous research has tended to focus on risk factors occuring during specific boating trips. This suggests that underlying boating accident risk behaviors could be represented as patterns in time and space. Boating accidents may not be the result of deterministic causality but rather the

combination of a set of integrated and complex risk factors that collectively increase the probability that a recreational boating accident will result.

Limitations and Assumptions

While the literature has advanced our knowledge about the relationship between boating accident risk factors and boating accidents, prior research is limited in several respects. One limitation is that previous research almost exclusively focuses on individual boat operator characteristics and behavior. These studies may be further weakened by conventional linear regression modeling approaches analyzing relatively rare and discrete phenomena within their spatial and temporal contexts. This inappropriateness stems from violations of conventional regression assumptions relating to error independence and variance homogeneity that require alternative techniques to estimate the desired relationships. In addition, most studies use datasets that are missing significant factors. For example, prior research may suffer from under-reported accident data or be based upon the analysis of boating accidents alone without consideration of representative non-accident data.

Boating Accidents through an Economic, Political, and Social Lens

During the past thirty years, recreational boating has emerged as a major source of economic prosperity influencing the social landscape and expanding 53% between 1980 and 2012 according to the U.S. Coast Guard (United States Coast Guard, 2013). The outcome of unprecedented coastal population growth is rapidly transforming U.S. coastlines and waterways to meet public demands for water-based recreation access (Jaakson, 1989). This coastal effect is compounded by non-coastal (metropolitan) recreational boaters who demonstrate an increased willingness to travel farther to

access coastal waterways (McCarthy & Talley, 1999; Sidman & Fik, 2005; Miller & Pikora, 2008; Swett et al., 2009). Prior research suggests that this one-way travel margin is up to 80 miles further extending the coastal boating access zone (Doll & Stiehl, 1979; Lentnek, 1970; Sidman & Fik, 2005; Swett et al., 2009). This coastal influence is contributing to a rapid rise in recreational waterway use as well as the associated rise in boating accidents (McKnight et al., 2007). It also serves to illustrate why growing public interest in better understanding and estimating boating accident risks are benefical from both a public and political perspective.

Research Overview

This research will assess the applicability of a boating accident risk model designed to estimate local boating accident risks. Additionally, it may offer decision support tools to boating law administrators interested in positioning limited assets in order to minimize boating accidents. From a broader perspective, this research seeks to fill some prior research gaps and advance recreational boating accident risk understanding by considering the complete recreational boating accident domain.

To guide this process, the following questions are asked:

1. What effects do environmental factors play in boating accidents and do those effects reveal specific spatial and temporal patterns?
2. When environmental conditions are properly controlled, (a) which human factors most significantly influence the distribution of boating accidents across time and space and (b) which technological factors most significantly influence the spatio-temporal distribution of boating accidents?

3. Can discrete categories of temporal variation in recreational boating accidents (e.g., annual, monthly, weekly, hourly) be statistically captured, and if so, what are the effects associated with those discrete categories?

CHAPTER 2: A SUMMARY OF EXISTING RESEARCH ON RECREATIONAL BOATING ACCIDENTS

The purpose of this chapter is to summarize the state of recreational boating accident research; a body of knowledge that is young in a temporal sense and relatively weak in concept and method. Nationally, the number of boating accidents is higher than rail and aviation accidents combined so there is clear need for improved intelligence on this phenomenon (O'Connor & O'Connor, 2005; Virk & Pikora, 2011). Despite this signficance, research attempting to explain recreational boating accidents have developed slowly with the most important gains dating back only to the 1980s (Jin, Kite-Powell, & Talley, 2001). This relatively underdeveloped body of knowledge stems in part from government and public emphasis on other modes of transportation including an overwhelming interest in highway transportation research. However, recent public and government attention on recreational boating accidents has generated new research interest. This interest has the potential to enhance our understanding of recreational boating accident risks (Hovden, Størseth, & Tinmannsvik, 2011).

There are close parallels between the dimensions that most influence highway (also called automotive) transportation accidents and those that influence recreational boating accidents. The literature evidences highway accidents in ways that can be readily applied to recreational boating.

To achieve this integrative goal, the information in this chapter is organized into a series of building blocks framed around a conceptual model representing recreational boating accident risk research. This process begins with a research context and theoretical framework that serves as the foundation for this research. That conceptual

framework permits the specification of three key concepts (dimensionals) critical to a better understanding of the nature of recreational boating accidents. Each key dimension will be explored individually and frequently, although not exclusively, through comparisons with automotive transportation. Next, gaps in the literature will be identified and weaknesses will be explored. In part, due to the scarce nature of prior recreational boating accident research, selected references to the automotive transportation literature will be used as a guide to identify deficiencies within the recreational boating accident literature. This is an important part of this comprehensive literature review that also is summarized by a gap analysis of existing recreational boating accident knowledge. This gap analysis is presented as a summary table at the end of this chapter.

Research Context and Theoretical Framework

To frame recreational boating accidents, contemporary accident theory draws heavily from epidemiology (Gabe & Hite, 2003). These conceptual designs express boat accident relationships as a function of system design and the nature of the risks imbedded within them (Anselin 2002; Sidman Fik et al., 2005). These designs borrow from two principal foundations, i.e., Normal Accident Theory (NAT) and High Reliability Organizations (HRO), to understand the nature of accidents, risks, and safety (Leveson et al., 2009). These comprehensive approaches to accident analysis create a "repertoire of analytic tools and intervention strategies" applicable to risk management involving complex systems (Leveson et al., 2009). Of these two approaches, Normal Accident Theory (NAT) is most closely related to the specific recreational boating accident factors and thus becomes the conceptual framework used in this research.

Systems accident literature credits Charles Perrow as first advancing the field of research described as Normal Accident Theory (Perrow, 1984). In the aftermath of the Three Mile Island nuclear power plant accident in 1979, this theory introduced the concept that accidents, i.e., specifically those involving technologically complex systems, are inevitable or in other words "normal" (Perrow, 1984, 1999). The key dimensions describing this inevitability were "*interactive complexity*" and "*loose/tight coupling*." Together, these dimensions determine a system's susceptibility to accident risk. According to NAT, accident risk characterized by interactive complexity and tight coupling cannot be foreseen or prevented. Perrow labelled this type of accident risk as a "system accident" (Marais, Dulac, & Leveson, 2007). Such accident risks are compounded when otherwise trivial components within the system fail in unpredictable ways then cascade with (often) severe consequences (Marais, Dulac, & Leveson, 2007). *Interactive complexity* refers to invisible, incomprehensible, and unplanned, system event sequences. Tightly coupled systems refer to highly interdependent, closely interlinked components when issues in one component can rapidly affect other parts of the system (Marais, Dulac, & Leveson, 2007). In contrast, loosely coupled systems respond more quickly to intervention and often absorb failures without destabilization (Marais, Dulac, & Leveson, 2007). A fundamental flaw in the NAT argument is that only engineering failures are considered although Perrow correctly argues that system "redundancy introduces additional complexity and risk taking" (Marais, Dulac, & Leveson, 2007). This engineering focus yeilds the omission of other key dimensions like human factors and operating environments in systems accidents.

14

To address the overemphasis on highly engineered systems ascribed to Normal Accident Theory, Erik Hollnagel proposed an important framework modification in 2008 that includes the human dimension (Hollnagel, 2008). While both approaches can be used to "locate" accident systems, Hollnagle's perspective employs a three dimensional array abstracted from three primary attributes. These three essential attributes are: (a) the level of *coupling* or system component interdependency, (b) the level of *interactiveness* or the intensity of operator/technology interaction and system linearity, and (c) level of *manageability* or the ability to control a system within its environment (Doll & Stiehl, 1979; Gabe & Hite, 2003; Hollnagel, 2008; Lentnek, 1970; Perrow, 1984).

Although Perrow and Hollnagle both reference system interdependency (coupling), Perrow (1984) focuses on system interactiveness while Hollnagel (2004) focuses on system manageability. To fully represent the complexity of recreational boating accident systems, this research offers a synthetic model of the three dimensional domain considered in this dissertation (see Figure 1). This modified perspective illustrates that recreational boating can be bracketed within a narrow frame of reference characterized as loosely coupled (relatively independent components), easily managed by one or more operators (simple to control), and linear in terms of component interactivity. Furthermore, it allows system attribute levels to vary within changing boat accident risk factors as examined in this research. For example, rapidly changing weather conditions can reduce the effectiveness of boating systems (manageability); the lack of precision in navigation equipment can reduce the effectiveness of boat movement between locations (coupling); and, harbor traffic congestion can increase system complexity (interactiveness).
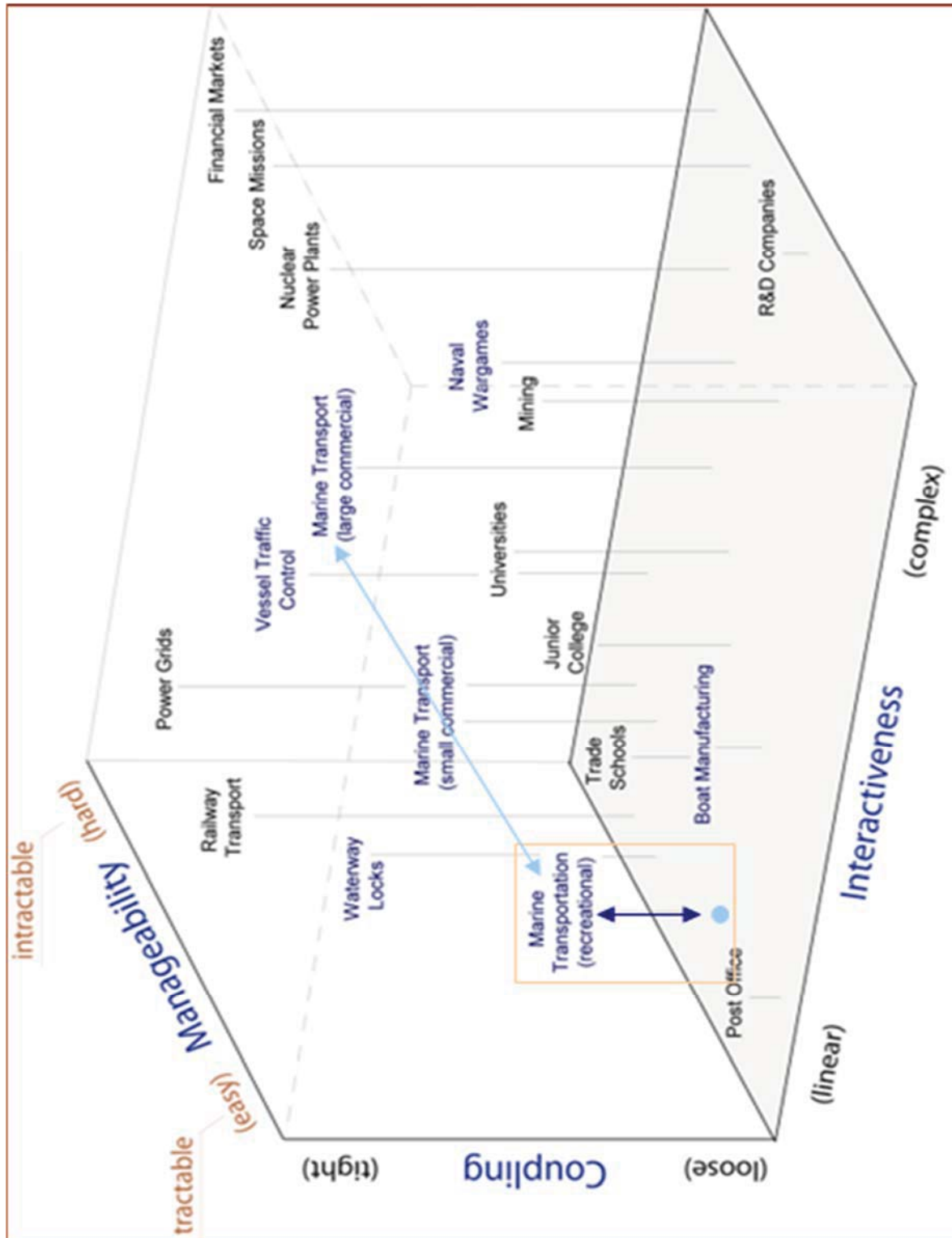
Note. (Synthesized after Perrow and Hollnagel).

Figure 1. 3D Coupling-Interactiveness-Manageability Conceptual Framework.

Recognition of these boat accident risk factors and their effects provide a conceptual as well as analytical focus for this research. Furthermore, this synthetic framework suggests that systems' similarities are directly related to the nature of the accident domains characterizing them (Sidman, Grant, & et al., 2005). It also suggests that certain accident risk factors can be characterized as antecedent conditions. Examples of such conditions include boat operator training, local boating rules and regulations, local cultural influences, and environmental conditions.

Key Concepts

In order to better understand boating accident risk factors, several key concepts must first be introduced. This involves developing an appreciation that the underlying effects can stem from complex interactions; i.e., those that occur with combinations of different human, technological, and environmental factors (Rasmussen & Svedung, 2000). For this reason and to be accurate, good recreational boating accident models must include all risk factors (Hovden, Størseth, & Tinmannsvik, 2011) including:

1. *human factors*: non-use of life jackets (Bell et al., 2000; Treser, Trusty, & Yang, 1997), experience, education, and collective behavior between operators (Zaidel 1992);

2. *technological factors*: inadequate stability and buoyancy (Cassell & Congiu, 2007), engines too large for the vessel, hull and machinery failure (O'Connor, 2008); and navigation system information overload (Hänninen, 2008; Harati-Mokhtari et al., 2007; McKnight et al., 2007); and

3. *environmental factors*: restricted visibility, floating or submerged objects, and bathymetry conditions (O'Connor, 2008); wind speed and sea conditions

17

(Ashby, Cassell, & Congiu, 2008); vessel congestion, variations in vessel directions and velocities, and temporal influences such as seasonality or weekday vs weekend boating operation.

Different aspects of these factors are examined in the subsections that follow. However, it is important to note that prior studies do not properly include important aspects from each of the three dimensions and in most cases do not attempt to model the risk of accidents happening. These inadequacies will form the basis of the added value contributed by the research as reported within this dissertation.

The Human Dimension

Human factors, which in this context relates to recreational boating operators, have been broadly viewed as making the largest contribution to recreational boating accident risk (Cassell & Congiu, 2007; O'Connor, 2008). A few studies have examined the pathways through which operators acquire their skill such as observing immediate family members, formal training, or exposure to boating as a part of life style in general (Factor, Mahalel, & Yair, 2007; Factor, Mahalel, & Yair, 2008). Other studies suggest that operator age is not a significant factor in explaining boating accidents because older operators are at least partially aware of their age-related limitations and adjust accordingly (Sivak, 2002; Borowsky, Shinar, & Oron-Gilad, 2010). This is an important empirical finding because recreational boats on the water and their boat owners are aging (Mahony & Stynes, 1995). A number of studies have examined boat operator age and gender differences (Borowsky, Shinar, & Oron-Gilad, 2010; Factor, Mahalel, & Yair, 2007; Factor, Mahalel, & Yair, 2008; Li et al., 2012; Yagil, 1998). These studies illustrate that younger less experienced boat

operators are more likely to suffer from deficiencies in hazard perception (Borowsky, Shinar, & Oron-Gilad, 2010; Factor, Mahalel, & Yair, 2007; Factor, Mahalel, & Yair, 2008; Li et al., 2012; Yagil, 1998). These authors suggest that younger boat operators possess an exaggerated sense of ability leading to a lower appreciation of risk and correspondingly greater risk-taking behavior (Borowsky, Shinar, & Oron-Gilad, 2010; Factor, Mahalel, & Yair, 2007; Factor, Mahalel, & Yair, 2008; Li et al., 2012; Yagil, 1998). Gender also appears to play a significant role in predicting accident risk. The literature suggests that the frequency of male operators in accidents is at least double that of females (Factor, Mahalel, & Yair, 2008).

The central argument that emerges from this literature focuses on boat operator (human dimension) accident factor as being embedded within a larger social context. This implies that differences in boat operation can be viewed as stemming from cultural (e.g., social norms, parental training) and physical differences (e.g., age, gender) between populations (Factor, Mahalel, &Yair, 2007). Collectively, these characteristics suggest that different social groups can also exhibit unique characteristics (Factor et al., 2007) and that these characteristics influence accident risk. Although still inadequate, this is the dimension that receives the greatest level of attention from researchers interested in better understanding recreational boating accident risks.

The Technological Dimension

Technology factors, as related to boating accident risk, represent an area of the boating literature that is largely underrepresented. Interest in this area stems from modern technology and the increasing complexity of boat operating systems. Such systems significantly impact the accident risk model but it also necessitates new

explanatory mechanisms to understand and assess those risks (Perrow, 1984; Qureshi, 2007). Traditionally, accident related technology failures have been viewed as resulting from cascading risk chains (Qureshi 2007). This perspective restricts accident modeling and ignores factors that arise from interrelated, integrated, and interconnected systems, e.g., interactions between technolgy, operator, and environment.

The central argument is that recreational boating accident risk is embedded within a technology-complexity context. This suggests that differences in boat operation can be viewed as stemming from a boat operator's inability to appropriately understand complex vessel designs, operation, and instrumentation (Qureshi, 2007). In general, this suggests that different types of technology do exist within the boating domain (size, power, navigation) and that these differences can uniquely influence the technological dimension and; therefore, recreational boating accident risk.

The Environmental Dimension

Recent studies have suggested that even when boat operators are "at their best" environmental and situational factors influence accident probability (Andrey et al., 2013). Such risks reflect errors in operator judgment given prevailing operating conditions (Andrey et al., 2013). Most transport systems are sensitive to environmental conditions such as heavy precipitation, fog, and wind (Andrey et al., 2003; Andrey et al., 2013; Bergel-Hayat et al., 2013; Chen & Cai, 2004; Cools, Moons, & Wets, 2010; Khattak, Kantor, & Council, 1998; Rahman & Lownes, 2012; Sewell, Kates, & Phillips, 1968; Snæbjörnsson, Baker, & Sigbjörnsson, 2007; Xu, Wang, & Liu, 2013). These factors have received more research attention than transportation in general in the past. For example, factors such as traffic patterns, vehicular density and intensity,

constricted maneuvering space and built-infrastructure (e.g., intersections, urban areas, and the like), and day vs night operation are frequently ingnored in the boating accident literature although they are frequently the focus of highway accident research (Arditi, Lee, & Polat, 2007; Chen & Cai, 2004; Johansson, Wanvik, & Elvik, 2009; Kockelman, 1998; Prasannakumar et al., 2011; Schmidt et al., 2009; Shankar, Mannering, & Barfield, 1995; Vogt & Bared, 1998).

The central argument is that recreational boating risk is tightly linked to and embedded within an environmental context. This means that differences in boat operation can be viewed as stemming from an operator's inability to effectively adapt to changing environmental conditions to ensure safe operations (Bristow & Bennett, 1995; Jaakson, 1989; Kujala et al., 2009; Pelot & Plummer, 2008; Perdue, 1987; Qureshi, 2007; Yip, 2008). Contemporary recreational boating literature provides little guidance for the inclusion of the key environmental factors that would logically influence boating accident risk, e.g., water depth, waterway constraints, and local boat traffic conditions. One of the key contributions of this dissertation is to elevate the analytical examination of environmental conditions as related to recreational boating accidents.

Proposed Recreational Boating Accident Model

Based on the previous work of Perrow and Hollnagel, the synthetic conceptual model illustrated in Figure 2 serves to anchor this research. This model, called the "Recreational Boating Accident Model", ensures that all boating accident factors are appropriately represented and permitting valid estimates to be obtained. All previously reported risk models suffer from specification bias resulting from the exclusion of one or more of these accident risk factors, e.g., human, technological, environmental.
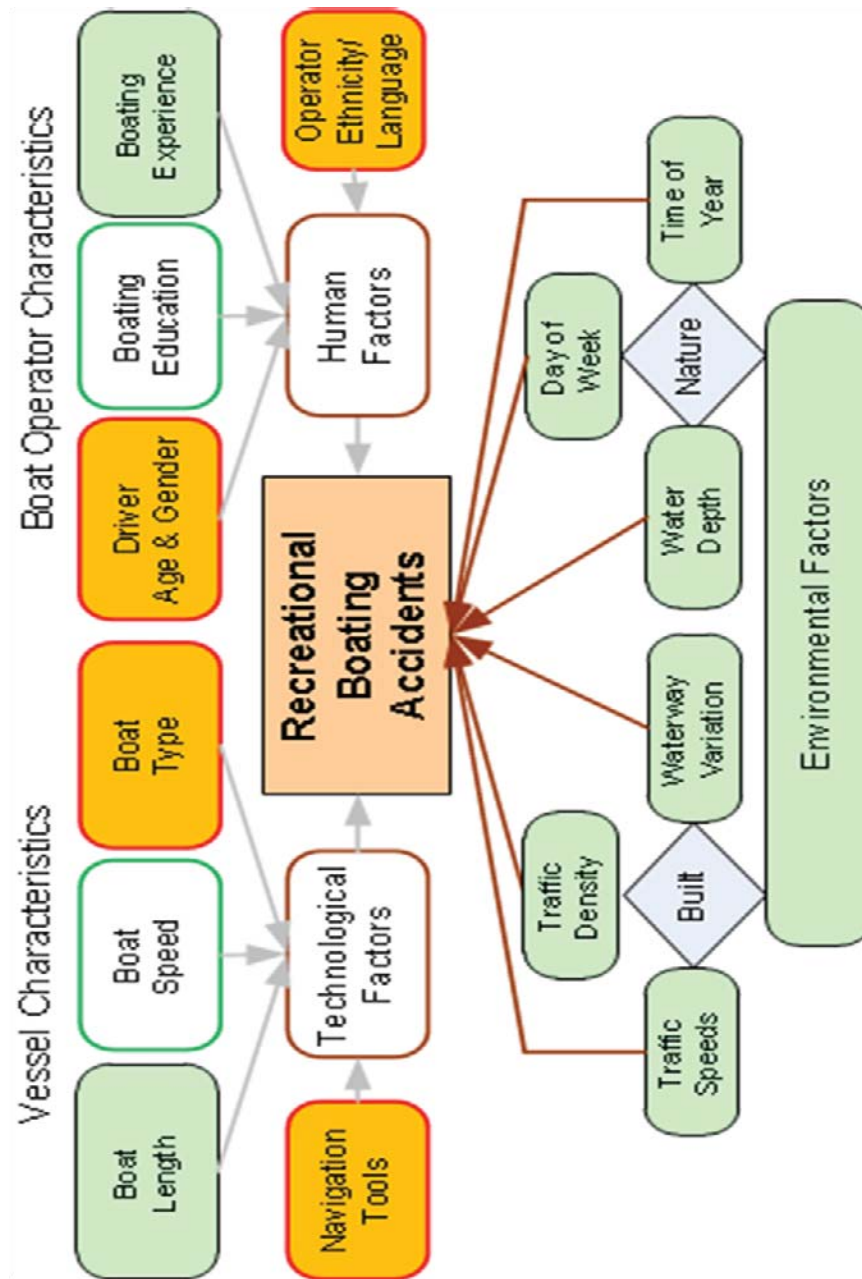
Figure 2. Conceptual Framework for recreational boating accidents.

No existing model includes appropriate measures from each of these factors. The approach in this research also creates a pathway for advanced statistical methodologies permitting technological and human factors to be viewed through an environmental lens. When the evironment is not properly controlled for, the full effect of the recreational boating accident domain on boat operators can not be fully assessed. The existing body of knowledge offers scant evidence of environmental factors influence on recreational boating accidents. For this reason, the use of comparable environmental-based highway accident literature not only serves to illustrate the potential benefit of a comprehensive multivariate modeling approach but the need for a broader perspective when considering recreational boating accidents as well.

Review of Recreational Boating Accident Literature: Prior to 1989

The first evidence of recreational boating accident risk research in the literature is attributed to the work of Doll and Stiehl in 1979, which was sponsored by the United States Coast Guard. Rather than focusing on the more general topic of recreational boating *accident risk*, Doll and Stiehl focused upon recreational boating *fatalities* and linkages to life jacket wear, boater behavior, and very limited environmental conditions. This investigation serves as an important benchmark in recognition of the need for further recreational boat accident research. However, Doll and Stiehl (1979) acknowledged significant flaws in their study, i.e., ineffective system for acquiring recreational boating accident data and significant data gaps in state provided data. In their view, these limitations hampered their work. It also resulted in accident risk models that were too theoretical for practical application (Doll & Stiehl, 1979). The U.S.

Coast Guard commented on this inadequacy attributing the poor quality to underreporting (Doll & Stiehl, 1979).

During the early 1980s, the attention of the general public and government agency priorities relating to recreational boating regulations and managment precipitated increased research. While subsequent research focused on boating accident risk (instead of fatalities), these studies typically examined a single causal variable, e.g., alcohol consumption, legal drinking age, or operating speeds (human factors), and the like (Loeb, 1984; Loeb, 1987). While the resulting models do follow conventional wisdom, the Loeb research provided strong indications of specification bias due to the omission of important explanatory variables (Loeb, 1987) and the absence of statistical control.

Almost all of these studies employ data that is aggregated at the state-level that limits the utility of the data in addressing local recreational boating sites. These inadequacies led Loeb to propose more inclusive multivariate models intended to capture additional independent variables such as personal income, population density, education levels related to vehicle operation, distances traveled, and boat operator age. However, the observational unit in these studies remained at the state level and this type of analysis yields or no little value when attempting to estimate local recreational boating accidents risk. On the other hand, a significant contribution of Loeb's work is the realization that boating accidents in a marine transportation environment are best characterized by multiple explanatory variables drawn from all boating accident dimensions.

Review of Recreational Boating Accident Literature: 1990-1999

Nearly a decade would pass before McCarthy and Talley (1999) advanced the work of Loeb and others by applying similar risk-related explanatory variables to recreational boating accident models. McCarthy and Talley hypothesized that increasing safety regulations should affect the frequency of boating safety behavior (human factors) and thus accident risk. More importantly, they argued that the effects associated with: (a) stricter state and federal regulations relating to boat operator training, experience, and education, (b) BUI (boating under the influence of drugs or alcohol), (c) personal floatation device wear (human factors), and d) weather (an environmental factor) are significant contributors. After considering a range of explanatory boating safety variables, McCarthy and Talley focused their research on the cumulative recreational boating experience and the formal training of boat operators (human factors).

This analysis was couched within the hypothesis that boat operators with minimal training and experience are more likely to engage in risky boating behaviors than those with more training and experience (McCarthy & Talley, 1999). Specifically, their rationale included tests of the effects of previous boat operating experience and formal boat operator training on, e.g., the increased frequency of situationally-appropriate safety behaviors (negative risk compensation) or decreased frequency of situationally-appropriate safety behaviors (positive risk compensation) (McCarthy & Talley, 1999). The United States Coast Guard, Boating Accident Report Database (BARD) served as the foundation for this analysis.

Correlation and t-statistics were extensively analyzed, but no multivariate predictive model was developed. This was due in part to under-reported or missing BARD data. The authors noted that their findings had implications for improving the understanding of risk compensation behavior (human factors) more than safety behaviors (McCarthy & Talley, 1999). For example, McCarthy and Talley suggest that minimally skilled boating operators are more likely to engage in risky boating behavior than those with greater boating skill levels (McCarthy & Talley, 1999). This research firmly establishes human factors as an important element of recreational boating accidents. On the other hand, McCarthy and Talley are illustrative of a basic flaw in almost all of this genre as they only recognized human factors as boating accident risks, i.e., technological and environmental dimensions are ignored. Again, as in previous boating accident research, all data were aggregated to the state level.

There are many reasons why BARD data gaps exist. Some fields are not required by law and some are unknown due to the nature of the accident (no witnesses, deceased sole person onboard, vessel destroyed, etc). Occasionally the public refuses to report information and follow-up is insufficient. While elements of data collection are mandated under regulation (e.g., accident causes), some fields (improper lookout, excessive speed, etc) are not. Thus, BARD data collection is neither uniformly nor consistent, which is a problem for authorities seeking to improve boating safety through legislation. Risk management companies and other entities also employ these data for accident risk assessment (Hassel, Asbjørnslett, & Hole, 2011). While data gaps and accident under-reporting still exist, the quality of BARD data since 1995 has improved to the point where advanced statistical approaches are now feasible.

Review of Recreational Boating Accident Literature: 2000-present

In 2000, Wang advanced the work of McCarthy and Talley (1999) with focus on constructing a multivariate-predictive approach to modeling recreational boating accidents. Wang's (2000) study improved upon McCarthy and Talley's isolation of state regulation effects on recreational boating fatalities by introducing a reduced boat accident model. As was the case in previous studies, the focus remains on explaining accident *fatalities* and not accidents in general (Wang, 2000). While limited, this approach permitted the application of more advanced methods for estimating boating accident risks as attributed to state regulations (Wang, 2000). Similar to Loeb and Giliad (1984) and McCarthy and Talley (1999), Wang chooses to focus exclusively on human factors such as enrollment in boating safety education programs, use of personal floatation devices, and alcohol use by recreational boat operators. He analyzed two dependent variables culminating in a reduced form boat accident/fatality rate model presented as: Rate=f(I,L) where Rate is the boating accident or fatality rate; I is a vector of demographic variables of the states; and, L is a vector describing the state laws, regulations, and boat operator education mandates, i.e., education and alcohol (Wang, 2000). To enable logistic regression, the rate was transformed using the formula: ln[R/(I- R)], where the natural logarithm with rate variables are constrained between zero and one (Wang, 2000). A weighted least squares regression is used to obtain efficient estimates.

Wang's (2000) results suggest that the estimation of two weighted logistic regression equations are comparable to Loeb (1987). Therefore, Wang (2000) advances recreational boating accident risk research through an illustration that boat

operator age, boating education, and personal floatation device wear have significant effects on boating accident risk/fatality rates. Conversely, Wang (2000) fails to entertain the full range of recreational boating risk factors due to a limited focus on human factors. This limited approach completely disregards risk factors from the technological and environmental dimensions. Furthermore and as indicated earlier, the state-level unit of observation is problematic in making any statements about the probability of local recreational boating accidents.

Gabe and Hite (2003) attempted to advance recreational boating accident research again by leveraging the work of McCarthy and Talley (1999) and the earlier work of Loeb et al. (1987). In this study, Gabe and Hite (2003) focused on the impacts of boat operator education programs and the number of enforcement officers in a particular area. This approach has merit because the presence of law enforcement officers in boating areas normally has a moderating effect on operator behavior (Yagil, 1998). Gabe and Hite (2003) advanced this model by investigating the effectiveness of boater education programs, patrol monitoring, and rules and regulations on safe boating behavior. The foundation created by this research is important as it combines the influence of human factors, and to a very limited degree, environmental factors i.e., law enforcement. The authors also used a reduced-form Poisson regression (a model similar to those used extensively in automotive studies) to predict accident risk. Explanatory variables included the number of patrol officers, total hours of boater education, number of registered boats, per capita income, latitude, and regional location (Gabe & Hite, 2003). However, models are limited by the unit of observation and estimates at a state-level of aggregation.

As mentioned above, Gabe and Hite (2003) employ boating accident data by state as the dependent variable and the number of enforcement officers (represented as full-time equivalent agents/1,000 registered boats), the hours of instruction required as part of a state's mandatory boater education program, and other non-policy factors expected as independent variables to explain the variation in the number of boating accidents across states (Gabe & Hite, 2003). Probably most important, the results of this study suggest a negative relationship between the number of enforcement officers available to patrol and the number of boating accidents. In contrast to previous work, Gabe and Hite (2003) also suggest that operator education does not have a statistically significant effect on recreational boating accidents (at the state level of aggregation).

While Gabe and Hite (2003) significantly advance the use of statistical modeling on the study of recreational boating accidents, we still have a limited view of the nature of boating accidents because of the aggregate character of the data (state-level) employed. As in previous studies, the overwhelming emphasis in boating accident risk assessment is attributed to human factors with minimal regard for the technological and environmental risk dimensions. Disregard for these critical boating accident risk dimensions compromises the quality of the resulting boating accident research.

O'Connor and O'Connor (2005) were the first researchers to investigate the full recreational boating accident domain (although only as related to boating *fatalities*). This was accomplished by subdividing boating accident risk factors into the three boating accident dimensions. This research was based on an examination of the cause and prevention of recreational boating fatalities in Australia. The wide-ranging data in this study included a collection of coroners' reports, witness statements, police reports,

autopsy findings, search and rescue reports, weather maps and reports, analysis of forensic and scientific data, assessment of photographic evidence, and review of other related information (O'Connor & O'Connor, 2005).

The focal variables in the O'Connor and O'Connor study are selected to understand the limited number of boating accidents resulting in fatalities. Primary explanatory variables selected are boat operator alcohol use and personal floatation device (PFD) wear, although others are examined. Each explanatory factor was considered in isolation in a more or less binary analysis. As a result, the authors do not offer a recreational boating fatality (multivariate) model but rather provide basic descriptive evidence of fatality differences between explanatory cateogies. Not surprisingly, this study suggests that human factors have the greatest effect on the number of boating fatalities, particularly as related to the use of drugs and alcohol by the boat operator. On the other hand, the study serves to strengthen the rationale for increased boating safety legislation, an increase in the number of boating law enforcement officers, and the increased emphasis on boat operator education and experience (O'Connor & O'Connor, 2005).

Collectively, the information provided by these recent studies broadens our understanding of recreational boating accidents due to the inclusion of a more comprehensive set of factors taken from the three dimensional boating accident domain. However, this prior work does not examine the root causes of recreational boating accident risks. In other words, the authors maintained a limited focus upon boating fatalities as opposed to the broader perspective derived from examination of recreational boating accidents in general. Finally, the lack of statistical sophistication

resulting from the examination of individual factors in a one-at-a-time approach provides insufficient statistical control within an admittedly multivariate process involving the all recreational boating accident dimensions. However, the examination of individual incidents (in contrast to aggregation at the state-level) represents a step in the right direction in understanding the underlying causes of boating, albeit, fatal, accidents.

McKnight et al. (2007) attempted to address these limitations by exploring the comprehensive nature of recreational boating accidents in contrast to boating accidents involving fatalities; a very small subset of the former. McKnight et al. (2007) use a human factor approach to understand recreational boating accidents by attempting to create categorical human (boating) error groups by boat type. The data used in this study were derived from 1996-1998 BARD reports obtained from the U.S. Coast Guard. McKnight, et al. found that because recreational boaters *own* their boats, effective accident prevention programs can be geared toward specific (owned) boat types (Glover, Lane, & Wang, 1995; Miller & Pikora, 2008).

With an overemphasis on human factors, public, state, and federal law enforcement perceptions have elevated drug and alcohol use as the primary risk factors in recreational boat accidents. While the involvement of drugs in some accidents is undeniable, the prominence of this single factor also derives, at least in part, from a body of knowledge that is methodologically immature, i.e., bivariate correlation analysis. For example, Miller and Pikora (2008) paralleled the research of (McKnight et al., 2007) and added alcohol consumption as a potential causal factor in recreational boating accidents. This study illustrated that alcohol consumption could be (correlation) an important recreational boating accident risk factor (Miller & Pikora, 2008).

Subsequent to the work of Miller and Pikora, there has been little comprehensive recreational boating accident research and only a few in the commercial boating accident domain (Mullai & Paulsson, 2011). Instead, recreational boating accident investigations have once again diverged into studies involving just one of the three recreational boating accident risk dimensions. For example, a recent recreational boating accident study explores human factors but with a specific focus on reducing the number of injuries rather than understanding the nature of the accidents (Virk & Pikora, 2011). This specific research design focuses on collecting educational measures important to boat operator safety practices using a Boating Safety Scale survey tool (designed by Virk & Pikora) to measure safe boating practices. Other investigations are pursuing recreational boating accident risks from a technological perspective. For example, Kujala et al. (2009) provide a recent analysis of marine traffic safety in the Gulf of Finland using automatic identification system (AIS) technologies to track and predict collision probabilities. Other investigators have placed focus upon determinants of passenger vessel accidents from environmental causes (Yip, 2008), or simply have reviewed the existing body of knowledge that concerns recreational boat use injury prevention (Cassell & Congiu, 2007).

As indicated earlier, environmental factors have largely been ignored throughout the brief history of this literature. However, a recent study (Niclasen, Simonsen, & Magnusson, 2010) gives special attention to the potential for producing better accident forcasts by including better weather forcasts with information involving sea-state. Specifically, this study examines wave height and steepness as related to small craft advisories (Niclasen, Simonsen, & Magnusson, 2010). While this study offers limited

32

attention to the spatial aspects of recreational boating (in general) and none when it comes to recreational boating accidents, it does offer encouragement for further study. For example, the primary focus of these recent environmental studies has been placed upon recreation and leisure demand rather than on recreational boat accidents (Balaguer et al., 2011; Gray et al., 2011; Sidman & Flamm, 2001; Sidman, Fik, & Flamm, 2002; Sidman & Fik, 2005; Sidman, Grant, & et al., 2005).

Information about spatial patterns of recreational boating is important for describing the environmental risks associated with recreational boating. Such distributions (of vessels, shoals, infrastructure, etc.) are essential to capturing the traffic patterns and other hazards that operators must cope with. In particular, congestion (just as on a highway) is a key component of boating risk that has been ignored in the literature. However, at least to date, traditional spatial mapping techniques involving remote sensing are frequently unable to provide key aspects of boating directions, speeds, and densities (Gray et al., 2011). This limitation will become a key contribution of this dissertation. Managing safe recreational boating waterways requires an understanding of the spatial distribution of recreational boating traffic (Gray et al., 2011). Moreover, the ability to fuse data from varied sources (remote sensing, surveys, etc.) to understand a phenomenon, like recreational boating accidents, is an essential methodological advance that has not yet found its way into the recreational boating accident literature.

Recent European investigations suggest that marine vessel groundings are a dominant accident type with 71% of marine accidents in European waters (2010); less than 45% of these accidents involved collisions (Chauvin et al., 2013; Kujala et al.,

2009). While these findings are interesting, they did not consider relevant

environmental factors nor provide a systematic approach that might be transferred to

recreational boating accident risk analysis (Chauvin et al., 2013).

Collectively, the existing body of recreational boating and other transportation

accident risk research suggests that there are other predictor variables influencing

recreational boating accidents. Unfortunately, the transportation literature is dominated

by work that is highway safety centered. For this reason, the recreational boating

accident literature can be summarized as: (a) studies focused on the human-dimension

while omitting others, (b) studies limited by highly aggregated data of questionable

quality, and (c) studies comparing means or frequencies rather than using more

sophisticated multivariate approaches.

Related Research from the Automotive Literature

Even a partial examination of the highway accident literature provides

transferable lessons to the recreational boating accident domain. For example, the

highway literature consistently asserts the important role of the operating environment

within which accidents take place. In turn, the highway literature reveals a dominant

position for the distribution (density) of vehicles (highway congestion) when explaining

accidents. Because the boating accident literature is so immature when compared to

the highway literature, this section provides transferable lessons that will inform the

methodological design of the current study.

Automotive transportation accident risks often mirror recreational boating

accident risk characteristics. For example, recent highway accident studies of

construction zone accidents clearly reveal the hazards of nighttime construction

operations (Arditi, Lee, & Polat, 2007; Johansson, Wanvik, & Elvik, 2009). Temporal

aspects are just as importrant in boating. In addition, most automotive transportation

systems are sensitive to environmental changes, but especially to heavy precipitation,

fog, and wind (Andrey et al., 2013; Bergel-Hayat et al., 2013; Sewell, Kates, & Phillips,

1968; Xu, Wang, & Liu, 2013). These environmental conditions are just as relevant

within the recreational boating domain as they are within the automotive domain.

In addition to the primary effects of weather and density, the highway literature

also indicates the existence of interactive effects. As a result, interactive effects should

gain attention as we examine recreational boating accidents. For example, wind speed,

temperature, and seasonal factors influence traffic density and correspondingly have

disporportinate effects on automotive accident levels (Al-Harbi, Yassin, & Bin Shams,

2012; Cools, Moons, & Wets, 2010). When traffic density is coupled with adverse

weather and topographic conditions, highway studies clearly suggest that automotive

accidents increase (Chen, Cai, & Wolshon, 2009). The same level of awareness of

interactive effects is absent from the recreationl boating literature.

Another key area of automotive traffic concern represented in the literature are

intersections. Traffic density related concerns regarding automotive intersections are as

applicable to navigation waterway channels and constraints as they are to their highway

counterparts. For example, highway merging/diverging zones force a competition for

vehicle space as compared to other sections of the highway (Mergia et al., 2013). Such

areas of intense competition for space are prone to higher incidence of automotive

accidents than other sections of the highway (Mergia et al., 2013). Although

intersections are notably significant in the highway accident literature, minimal if any

recreational boating research interest has been exhibited in this area. Once again, this environmental neglect is unfortunate from a modeling perspective.

The research undertaken here is intended to *move recreational boating accident research to a more mature platform*, i.e., one that: (a) includes all important boating accident dimensions, (b) pays close attention to appropriate observational units, and (c) employs advanced statistical methodologies to better understand when, where, and why boating accidents occur.

The Potential for Improved Recreational Boat Accident Models

Although prior recreational boating accident models appear to follow conventional wisdom, specification bias created by the omission of important explanatory variables continue to limit this area of research (Gabe & Hite, 2003; Loeb & Gilad, 1984; Loeb, 1987; Loeb, 1994; Wang, 2000). Moreover, too little attention is given to the observational unit (micro-level as opposed to the macro-state-level), that further limits our understanding of recreational boating accidents. Findings in the literature have been mixed at least in part because important risk related explanatory variables have not yet been considered simultaneously with an appropriate observation level. Furthermore, the limited literature reveals that research to date has been highly confined with little attention given to geographical aspects critical to understanding recreational boating safety. This suggests that closer examination of the spatial distributions associated with recreational boating accidents would advance the body of knowledge in marine transportation by including the space-time patterns associated with recreational boating accidents.

As illustrated in this chapter, the existing body of knowledge is too narrowly focused on human influences with insufficient consideration given to the environmental influences on recreational boating accidents. On the other hand, the literature review suggests that the inclusion of such influences could help close large gaps in our understanding of the recreational boating accident domain (Jaakson, 1989; Lentnek & Doren, 1969; Sidman & Fik, 2005).

Several fundamental questions related to recreational boating accidents are thus unaddressed, e.g., where and when do recreational boating accidents occur and why do they occur in those locations? Hence, key factors influencing the volume, spacing, and timing of boat accidents could (and should) include attributes such as: (a) traffic distribution and congestion, (b) variations in speed and direction, (c) boat/operator characteristics, (d) navigation channel characteristics, (e) visibility, and (f) sea state. These factors remain essentially disregarded and statistically uncontrolled for within existing research. At a minimum, it is safe to suggest that previous attempts to model the boating accident record suffer from the severest form of specification error, i.e., left out variables, primarily because key factors and interactions have largely been ignored. Again, isolated consideration of boating accident factors is simply too restrictive to capture the complexity of the full recreational boating accident domain.

The proposed research design utilized in this investigation will permit: (a) the direct and measurable influence of environmental factors on the accident domain and (b) provide statistical control of technological factors and human factors (including possible interactive effects). Since all accidents and influences are tagged in space and time, the desired spatio-temporal results (e.g., some influences are more important in

37

some places than others and some influences are more important during certain seasons than others) can be readily estimated.

Literature Gap Analysis and Summary

A summary table of the literature in the areas of automotive and marine transportation is illustrated in Figure 3. Four areas of interest emerge from this gap analysis. First, the increasing use of analytical techniques to model observations. Second, only four prior recreational boating accident studies partially explore all recreational boating accident dimensions. Third, only two of the studies include a temporal component. This dissertation is the only research incorporating boating accident data, satellite imagery, on-the-water surveys, spatial and temporal analysis, and advanced modeling (negative binominal regression) to better understand recreational boating accident domain.

#	Author	Year	Concept	Auto	Boat	Human	Techno	Enviro	Temporal	OTW Data	Rec Boating Leisure Studies	Rec Boating Accident Risk Study	Rec Boating Fatality Risk Study	Poisson or NB2 Regression
1	Sewell, Kates, and Phillips 1968	1968	x	x				x						
2	Lentnek and Doren 1969	1969			x	x		x				x		
3	Lentnek 1970	1970			x	x						x		
4	Doll and Stiehl 1979	1979	x		x	x	x						x	
5	Loeb 1984, 1987	1984	x	x	x	x	x	x						
6	Loeb and Gilad 1984	1984			x		x							
7	Perow 1984, 1999	1984	x											
8	Loeb 1987	1987		x		x								
9	Perdue 1987	1987				x		x				x		
10	Jaakson 1989	1989	x			x		x				x		
11	Zaidel 1992	1992		x		x								
12	Bristow and Bennett 1995	1995			x	x		x				x		
13	Glover, Lang, and Wang 1995	1995			x	x						x		
14	Mahony and Stynes 1995	1995			x			x				x		
15	Shankar, Mannering, and Barfield 1995	1995	x	x			x	x						
16	Treser, Trusty, and Yang 1997	1997			x	x						x		
17	Harrald et a. 1998	1998			x	x	x	x						
18	Khattak, Kantor, and Council 1998	1998		x		x		x						x
19	Kockelman 1998	1998		x		x	x	x	x					
20	Vogt and Bared 1998	1998	x	x			x	x						x
21	Yagil 1998	1998		x		x								
22	Loeb et al. 1999	1999		x		x	x							
23	Logan et al. 1999	1999												
24	McCarthy and Talley 1999	1999			x	x	x	x	x			x		
25	Bell et al. 2000	2000			x	x	x						x	
26	Rasmussen and Svedung 2000	2000	x											
27	Wang 2000	2000	x		x	x						x		
28	Jin, Kite-Powell, and Talley 2001	2001			x	x	x							x
29	Sidman and Flamm 2001	2001			x	x		x			x			
30	Anselin 2002	2002	x											
31	Sidman, Fik, and Flamm 2002	2002			x	x		x			x			
32	Sivak 2002	2002	x	x		x								
33	Andrey et al. 2003, 2013	2003		x		x		x						
34	Gabe and Hite 2003	2003			x	x	x					x		x
35	Chen and Cai 2004	2004		x		x		x	x					
36	O'Connor and O'Connor 2005	2005			x	x	x	x				x		
37	Sidman and Fik 2005	2005			x	x		x			x			
38	Sidman et al. 2005	2005			x	x	x				x			
39	Keay and Simmonds 2006	2006		x		x		x						
40	Arditi, Lee, and Polat 2007	2007		x		x			x					
41	Cassell and Congiu 2007	2007			x	x	x						x	
42	Factor, Mahjalel, and Yair 2007, 2008	2007	x											
43	Harati-Mokhtari et al. 2007	2007			x	x	x					x		
44	Kilpelainen and Summala 2007	2007		x		x		x						
45	Loeb and Clark 2007	2007		x		x	x							
46	Marais, Dulac, and Leeson 2007	2007	x											
47	McKnight et al. 2007	2007			x	x	x	x				x		
48	Qureshi 2007	2007	x			x	x							
49	Snaebjornsson, Baker, and Sigbjornsson 2007	2007	x	x				x						
50	Ashby, Cassel, and Congiu 2008	2008			x	x	x					x		
51	Hanninen 2008	2008			x	x						x		
52	Hollnagel 2008	2008	x											
53	Miller and Pikora 2008	2008			x	x						x		
54	O'Conner 2008	2008			x	x	x	x					x	
55	Pelot and Plummer 2008	2008	x		x	x	x	x						
56	Yip 2008	2008	x		x	x	x	x	x					x
57	Chen, Cai, and Wolshon 2009	2009		x		x		x						
58	Johansson, Wanvik, and Elvik 2009	2009		x		x			x					
59	Kujala et al. 2009	2009			x	x	x	x	x					
60	Leveson et al 2009	2009	x											
61	Schmidt et al. 2009	2009		x		x			x					
62	Wang, Quddus, and Ison 2009	2009		x		x		x						
63	Borowsky, Shinar, and Oron-Gilad 2010	2010	x	x		x								
64	Cools, Moons, and Wets 2010	2010		x		x		x						
65	Cools, Moons, and Wets 2010	2010		x		x		x						
66	Niclasen, Simonsen, and Magnusson 2010	2010			x	x	x	x				x		
67	Balaguer et al. 2011	2011	x		x	x		x						
68	Gray et al. 2011	2011			x	x		x				x		
69	Hassel, Asbjornslett, and Hole 2011	2011	x		x	x						x		
70	Hovden, Storseth, and Tinmannsvik 2011	2011	x			x	x	x						
71	Mullai and Paulsson 2011	2011	x		x	x	x	x						
72	Prasannakumar et al. 2011	2011		x		x			x					
73	Virk and Pikora 2011	2011	x		x	x						x		
74	Al-Harbi, Yassin, abd Bin Shams 2012	2012	x	x				x						x
75	Li et al. 2012	2012	x			x								
76	Rahman and Lownes 2012	2012		x		x		x						
77	Bergel-Hayat et al. 2013	2013		x		x		x						
78	Chauvin et al. 2013	2013			x	x	x	x						
79	Mergia et al. 2013	2013		x		x		x				x		
80	Xu, Wang, and Liu 2013	2013		x		x		x						x
81	Marshburn 2014	2014	x		x	x	x	x	x	x		x		x

Figure 3. Literature Review Gap analysis.

CHAPTER 3: RECREATIONAL BOATING ACCIDENT RISK -

THE METHODOLOGICAL APPROACH

The purpose of this chapter is to provide a detailed description of the methodology used in connection with this research. It describes the basic approaches being used (e.g., inductive detective work, pattern seeking, data mining, multivariate modeling), the data sources, and data analytic techniques. It further enables a clear understanding of the processes and procedures undertaken. Chapter 3 is organized to include: research questions; the research design; site selection; research data; data fusion, and the techniques for data analysis. These methods provide special attention to: (a) description of the unit of analysis, (b) normalizing the data, (c) data preparation and post-processing, (d) measures and measurement error, and e) research reliability and validity.

Research Questions

As noted earlier, recreational boating accidents are influenced by three boating accident risk factors, i.e., human, technological, and environmental (dimensions) that may include interactive effects varying across space and time. To capture the effects of these risk factors in a working model, the following hypotheses are presented to guide dissertation research.

Environmental: Complex operating environments yield higher recreational boating accident risk probabilities. Thus, operating environments can be attributed with varying traffic densities, direction and patterns, weather, sea-state, etc. Rapid environmental changes can also impact boat maneuverability and on-water boat control.

So:

H1:   The frequency of recreational boating accidents increase as the number of
       boaters in a navigable waterway-space increase.

H2:   Recreational boat accidents are more frequent in areas where the
       environment being navigated is more complex (i.e., channel orientation,
       channel width, water depth, and environmental complexity such as
       intersections, bridges, and related obstructions).

H3:   Recreational boating accidents vary as a function of traffic speed.

H4:   Recreational boating accidents are more frequent during peak boating
       months (e.g., May-August) and less frequent during non-peak boating
       months.

H5:   Recreational boating accidents are more frequent during weekend periods
       (Friday-Sunday) than during weekday periods (Monday-Thursday).

Technological: Boat size, boat speed, boat type, and available navigation
resources influence recreational boating accident risk probability. Boats operating at
faster speeds represent a greater accident risk given that as their momentum increases
their maneuverability decreases and stopping distances increase. So:

H6:   Recreational boating accidents are more frequent in areas where the
       average boat length is smaller and less in areas where the average boat
       length is greater.

H7:   Recreational boating accidents are more frequent in areas where the boat
       type is dominated by powerboat and/or jetski watercraft.

Human: Recreational boat operators with higher levels of boating experience and education exhibit less risky recreational boating behaviors than boat operators with minimal recreational boating experience and education. So:

H8:   Recreational boating accidents are more frequent in areas where average boat operator experience level is lower and less frequent in areas where average boat operator experience level is higher.

H9:   Recreational boating accidents are more frequent in areas where average boat operator education level is lower and less frequent where average boat operator education level is higher.

H10:   Recreational boating accidents are less in areas where average boat operator age is relatively young or relatively old.

Research Design: Site Selection

Consideration of potential coastal research sites for this investigation began with a careful examination of the annual USCG recreational boating accident reports with particular emphasis on 2012 Recreational Boating Statistics (United States Coast Guard, 2013). Florida consistently ranks first nationally with respect to the number of recreational boating accidents (13.7% of the nation's total between 2008 and 2012). For this reason, Florida was high on the list of research sites being considered. The ranks of the top twenty states in terms of the overall number of recreational boating accidents are shown in Table 1. Fifteen of the twenty; i.e., FL, CA, TX, NY, MD, NC, MI, GA, OH, NJ, LA, VA, IL, SC, and WA, have coastal zones, assuming that the Great Lakes are included in the definition of coastal zone. For the purposes of this research

42

Table 1

Recreational Boating Accidents/State (2008-2012)

Total Number of Accidents

	State	2008	2009	2010	2011	2012	Total	%
1	**Florida**	**616**	**610**	**608**	**685**	**662**	**3,181**	**13.7%**
2	California	520	478	412	399	365	2,174	9.4%
3	Texas	218	168	163	197	162	908	3.9%
4	New York	160	148	211	173	197	889	3.8%
5	Maryland	159	174	196	184	145	858	3.7%
6	North Carolina	148	144	148	144	145	729	3.1%
7	Missouri	135	150	161	128	141	715	3.1%
8	Michigan	187	131	132	129	103	682	2.9%
9	Arizona	158	151	113	158	99	679	2.9%
10	Georgia	150	145	135	96	111	637	2.7%
11	**Ohio**	**125**	**105**	**127**	**135**	**136**	**628**	**2.7%**
12	Tennessee	130	117	116	117	147	627	2.7%
13	New Jersey	140	126	116	119	115	616	2.7%
14	Louisiana	110	120	105	112	116	563	2.4%
15	Virginia	95	137	102	121	89	544	2.3%
16	Wisconsin	110	102	104	110	110	536	2.3%
17	Illinois	119	96	97	106	101	519	2.2%
18	South Carolina	107	95	102	93	108	505	2.2%
19	Washington	98	111	72	93	105	479	2.1%
20	Utah	80	87	103	109	99	478	2.1%
	Total Accidents	4,789	4,730	4,604	4,588	4,515	23,226	

study, the Great Lakes are considered coastal waters due to their size and geographic complexity.

In this research, it should be noted that there are two data elements that require the participation of collaborating state partners. These elements include access to the federal BARD data collection system (**B**oating **A**ccident **R**eport **D**atabase; assembled and maintained by the United States Coast Guard) and the collection of non-accident based data through use of an on-the-water survey. On-the-water survey data is officially referred to as vessel safety check or VSC d*ata* by state boating law enforcement officers. To obtain and access BARD data, state sponsors were required before USCG would authorize access. In addition to this project requirement, state collaboration necessitated the need to secure at least one and preferably two states willing to assist in collecting non-accident based recreational boating data (by state boating law enforcement officers). Additional information about this on-the-water survey component can be found in the section On-the-Water (OWS) Survey Data.

Initially, state collaboration was sought from Florida, Texas, North Carolina, Ohio, and Virginia. The rationale for seeking collaboration with these specific states related to their high recreational boating accident ranking and this author's established relationships with the Boating Law Administrators (BLAs) in those states. Following initial contact and subsequent discussions concerning research partnerships, Florida, Texas, and Ohio expressed interest. North Carolina and Virginia indicated an interest in being considered but preferred participation in a future research round. Subsequently, and resulting from state leadership changes, Texas also requested reassignment to a future round of this research. This chain of events lead to signed contracts and a strong

collaborative with the BLAs from both Florida (Major Richard Moore) and Ohio (Chief, Pam Dillon and later Chief Rodger Norcross). In connection with this partnership, the Florida BLA (Major Moore) and the Ohio BLA (Chief Dillon and Chief Norcross) granted access to the Florida and Ohio BARD data. Susan Tomczuk (USCG) authorized access and extracted the initial (and subsequent) BARD datasets for descriptive analysis and analytical use were made available.

Florida ranks first nationally in terms of the number of recreational boating accidents while Ohio ranks eleventh (behind North Carolina ranked sixth). Another reason for the decision to pursue Ohio rather than North Carolina (NC) followed an examination of NC recreational boating accident statistics. These statistics revealed that many NC recreational boating accidents occur on inland lakes rather than in coastal waters. For this reason, the selection of Florida and Ohio as the recreational boating research sites of choice offered strategic advantages.

Once the state partners were selected, the process of identifying a suitable high recreational boating use research site, i.e., a trafficshed, within FL and OH began. A trafficshed is defined as a navigable waterway within which recreational boats operate. Trafficsheds are comprised of common elements such as harbors, harbor channels and associated waterways (e.g., rivers), marinas, and primary - secondary navigation channels that lead to deep open - water access (Swett et al., 2009). Maps of these selected (Tampa, FL) and (Sandusky, OH) trafficsheds (research sites) coupled with an illustration of the satellite containment areas (denoted by a red outline) are illustrated in Figure 4.
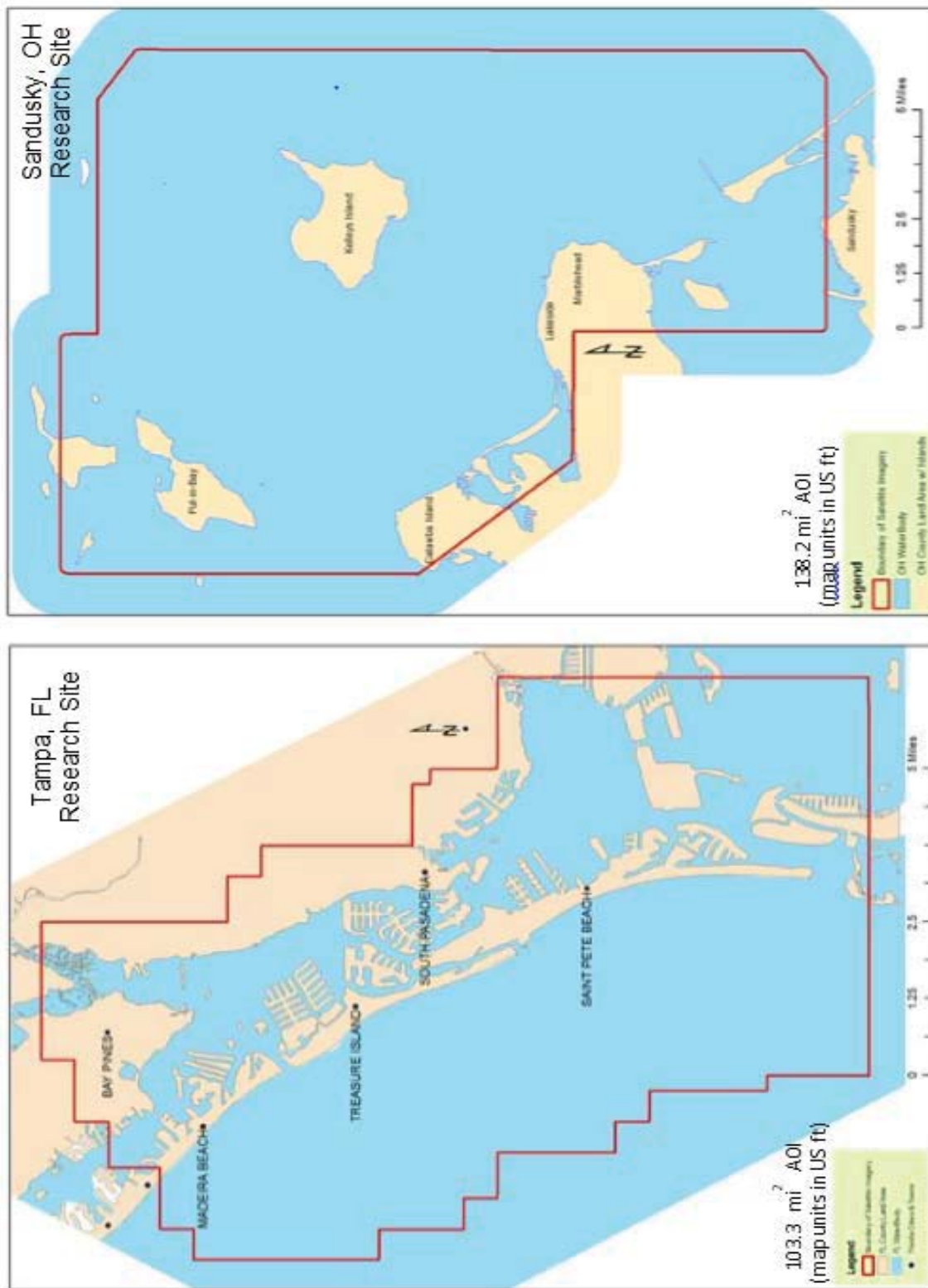
Figure 4. Tampa, FL and Sandusky, OH research sites.

Research Data and Fusion

This section includes definitions that provide a concise explanation of each variable included within the analyses. They are primarily organized by their collection mechanism and further subdivided by the three accident dimensions. In addition, this list is intended to create an organizational structure for understanding each variable, as well as the measures and constructs for which they are operationalized.

Collectively, spatial analysis is accomplished through the collection of accident data, known as the BARD. Additional variables are derived from high-resolution satellite imagery (imaged and provided by DigitalGlobe, Inc., formally GeoEye Inc.), waterway charts provided by the National Oceanic and Atmospheric Administration (NOAA), weather characteristics provided by the National Climatic Data Center (NCDC). The final set of variables is collected via on-water-survey, vessel safety check (VSC) collected by the collaborating state agencies (Florida Fish and Wildlife Conservation and Ohio Department of Natural Resources). These data are inventoried in Table 2: Variable Matrix. With the exception of the BARD data, which is continuously collected, all other data elements were collected during a 16 month data collection period spanning the timeframe between May 1, 2011 and August 31, 2012 (488 days).

USCG BARD Database (DS-1)

As noted above, the variables extracted from the United States Coast Guard BARD system are continuously collected by the Florida Fish and Wildlife Conservation Commission (FWC) and Ohio Department of Natural Resources (ODNR) in connection with recreational boating accident reports. These data were recorded on a BARD accident report (similar to the on-the-water VSC survey form used in this study) initially

Table 2

Variable Matrix

Key Variables	DS-1	DS-2	DS-3	Measurement
Environment variables	-	-	-	
Traffic	-	-	-	
Speed	X	X	X	Satellite
Density	X	X	X	Satellite
Navigation Channel Variation				
Width	X	X	X	Satellite
Depth	X	X	X	Satellite
Orientation	X	X	X	Satellite
Nav Complexity	X	X	X	Satellite
Sea State	-	-	-	
Wind Speed	X	-	X	NOAA/Observation
Waves Height	X	-	X	NOAA/Observation
Atmospheric Visibility	X	-	X	NOAA/Observation
Weather Conditions	X	-	X	NOAA/Observation
Water Conditions	X	-	X	NOAA/Observation
Water Current	X	-	X	NOAA/Observation
Air Temperature	X	-	X	NOAA/Observation
Water Temperature	X	-	X	NOAA/Observation
Technological variables	-	-	-	
Boat Classification	X	X	X	Satellite/Observation
Boat Propulsion	X	-	X	Observation
Boat Length	X	X	X	Satellite/Observation
Boat Age	X	-	X	Observation
Onboard Navigation Tools	X	-	X	Observation
Engine Classification	X	-	X	Observation
Boat Speed	X	X	X	Satellite/Observation
Human variables	-	-	-	
Operator Age	X	-	X	Observation
Operator Gender	X	-	X	Observation
Boater Ethnicity/Language Barrier	X	-	X	Observation
Boat Operator Experience	X	-	X	Observation
Boat Operator Education	X	-	X	Observation
Boat Operator Life Jacket Wear	X	-	X	Observation
Boat Operator Weather Forecast Check	X	-	X	Observation

Table 2 (continued)

Key Variables	DS-1	DS-2	DS-3	Measurement
Temporal variables	-	-	-	
Peak Time of Day	X	X	X	Satellite/Observation
Peak vs Off-Peak Season	X	X	X	Satellite/Observation
Peak Month	X	X	X	Satellite/Observation
Peak Day of Week	X	X	X	Satellite/Observation
Spatial variables	-	-	-	
Location	X	X	X	Satellite/Observation
Accident Occurance (BARD)	X	-	-	Observation

Note. DB-2: BARD Data; DB-2: Satellite/Remote Sensing Imagery; DB-3: On-Water_Survey.

on paper and then transcribed to an electronic report database. The report data are

reviewed and approved by several layers of agency administration who in turn transmit

this data to the USCG for assimilation into the federally maintained BARD system.

However, to protect individual privacy, all BARD reports were redacted of all personally

identifiable information and that de-identification did not compromise the usefulness of

the accident data.

Following extraction of the 1995-2012 Florida and Ohio BARD data (the initial

extraction yielded data for the entire state), all BARD data were filtered into boating

accident report subsets for: (a) Pinellas County, FL and (b) Erie and Ottawa County,

OH. These two subset selections were further reduced to the recreational boating

observations specifically located within the research sites. These satellite containment

observations are identified in Table 3 and serve to establish the boundaries of the study

areas within the identified states/counties.

The Tampa, FL and Sandusky, OH satellite containment areas include

information about recreational boating accidents (e.g., boating accident location, time,

and classifications). ArcGIS was used to plot the latitude and longitude of each

accident observation enabling the precise refinement of these data into a subset. Once

the final subset was obtained, accidents occurring between 1995 and 2004 were

extracted and analyzed for general mapping and descriptive purposes and those

accidents occurring between 2005 and 2012 were extracted and analyzed for more

focused exploratory purposes. Variables associated with these accidents and derived

from the 2005-2012 BARD data source include:

Table 3

Satellite Sampling Distribution

Month	Day		Collection Date	FLA	OH
			2011		
MAY	SUN	1	1-May	X	
	THUR	2	30-May	X	
JUNE	FRI	3	3-Jun		X
	SUN	4	5-Jun	X	
	SAT	5	11-Jun		X
	THUR	6	16-Jun	X	
JULY	WED	7	13-Jul	X	
	FRI	8	22-Jul		X
	SUN	9	24-Jul	X	
	SAT	10	30-Jul		X
SEPT	THUR	11	1-Sep		X
	MON	12	12-Sep	X	
	TUES	13	20-Sep	X	
	FRI	14	23-Sep	X	
OCT	SAT	15	1-Oct	X	
	FRI	16	7-Oct		X
	WED	17	12-Oct	X	
			2012		
MAR	SAT	18	3-Mar	X	
	MON	19	17-Marh		X
	SAT	20	19-Mar	X	
	THUR	21	22-Mar		X
APR	MON	22	2-Apr	X	
	SUN	23	8-Apr		X
	TUE	24	10-Apr	X	
	FRI	25	13-Apr		X
MAY	SAT	26	5-May	X	
	FRI	27	11-May		X
	SAT	28	19-May		X
	SUN	29	27-May	X	
JUN	SAT	30	2-Jun		X
	THUR	31	7-Jun		X
	MON	32	18-Jun	X	

Table 3 (continued)

Month	Day		Collection Date	FLA	OH
JULY	SAT	33	7-Jul		X
	SAT	34	21-Jul		X
	SAT	35	21-JulX	X	
	SUN	36	29-Jul	X	
			2013		
SEPT	FRI	37	24-Sep		X

a. limited environmental characteristics: speed, weather-sea state (e.g., atmospheric visibility, wind speed, and wave height),

b. technological characteristics: boat characteristics (classification, propulsion, length, age), onboard navigation tools, engine classification.

c. human characteristics: operator age, gender, boating education, boating experience, weather forecast pre-check, and life jacket wear,

d. temporal characteristics: time of day and date,

e. spatial characteristics: latitude/longitude of accident locations.

Satellite Imagery and Remote Sensing Data (DS-2)

The satellite imagery and other remotely sensed data (see Table 2) contain the geographic and physically observable elements of each trafficshed under study including: satellite imagery and sea state data from the NOAA National Climatic Data Center (NCDC). Both sets of data were overlaid on electronic nautical charts provided by the National Oceanic and Atmospheric Administration (NOAA Small-Craft Chart 14842, Port Clinton to Sandusky, including the Islands, ed. 9/8/2011; Intercoastal Waterway Chart 11411, St Petersburg, FL, ed. 11/14/2014).

All satellite imagery was collected by GeoEye, Inc., later purchased by DigitalGlobe, Inc. Of the space-based orbital platforms available, both GeoEye and IKONOS satellites were used to collect the high resolution color imagery acquired in connection with this research study. However, in order to simplify satellite referencing, this imagery will simply be referred to as GeoEye from this point forward. All GeoEye imagery collected was rendered with a 0.80m cell size for panchromatic images

and a 3.2 m cell size for all multispectral data. The GeoEye satellite has a sun-synchronous polar orbit with an altitude of approximately 681 km (423 mi) enabling earth orbit every 100 minutes and over-pass of any given location. The satellite generalized collect time of 11:30am and Latitude 27.8264 degrees (Florida) and Latitude 41.6775 degrees (Ohio).

All acceptable images were rendered 95% cloud free with minimal specular reflection. Specular reflection occurs over water when the angle of incidence equals the angle of reflection. The minimization of both cloud and specular reflection is important to ensure that all areas of the trafficshed are unobscured and thus available for observation. For reference purposes, the swath width of each GeoEye image capture is between 16.4 and 18.0 km at nadir. Nadir is the directional image directly below a particular orbit location. Swath size is important because in some instances, the size of the research sites necessitated multiple image collections. Collectively, 36 single frame or multi-frame 11-bit, MSS (multispectral), panchromatic, and near infrared satellite-band images were captured during the 16 month data collection phase. These images were divided between the two peak boating periods (between May-August) during 2011, 2012 and two consecutive non-peak boating periods (between September-April) during 2011-2012 as shown in Table 3.

However, post-processing of all GeoEye satellite imagery was required to generate the high resolution color imagery desired. This post-processing involved the creation of 0.8 m 11-bit pansharpened multispectral images following a BGRN (blue, green, red, and near infrared stacked layer) format. Numerous studies have illustrated the remote sensing benefits associated with pansharpening color imagery; a process

54

otherwise known as image fusion (Alparone et al., 2007; Amro et al., 2011; Choi, 2006; Fonseca, Namikawa, & Castejon, 2009; Helmy, Nasr, & El-Taweel, 2010; Kalpoma & Kudoh, 2007; Nikolakopoulos, 2004; Nikolakopoulos, 2008; Otazu et al., 2005; Tømmervik et al., 2012; Vijayaraj, O'Hara, & Younan, 2004; Wang et al., 2005; Zhang, 2002; Zhang, 2004). Pansharpening is the technique of merging high-resolution panchromatic (gray scale) imagery with the lower resolution multispectral (color) imagery to create a single high-resolution color image (Alparone et al., 2007; Helmy, Nasr, & El-Taweel, 2010; Nikolakopoulos, 2004; Nikolakopoulos, 2008; Vijayaraj, O'Hara, & Younan, 2004; Wang et al., 2005; Zhang, 2002; Zhang, 2004).

Initially, ERDAS Imagine (ver. 2011) was used to post-process the GeoEye satellite imagery into stacked BGRN bands. This involved a sequence, that began with stacking the multispectral BGRN layers and then mosaicing those images (combining two or more georeferenced files into a single output raster) into a final mosaic image representing the entire research site. This technique proved to be tedious and time intensive (estimated to be between 2-3 hrs per image layer). The time intensive nature of this process was partially addressed by building an ERDAS Imagine model that automated the band-stacking process. Following BGRN stacking, each 3.2m multispectral image was individually pansharpened using ERDAS Imagine. Following the stacking and pansharpening process of the Ohio research site, the resulting image quality proved to be lower than expected and a modified sequence was adopted.

A large collection of pansharpening methods has been developed during the past two decades. More importantly, many of those method were developed for specific remote sensing applications (Helmy, Nasr, and El-Taweel 2010). The best known and

most frequently used pansharpening techniques include: the modified intensity–hue–saturation, Brovey transformation, arithmetic/statistical combinations, principal component analysis, multiplicative transformation, wavelet resolution merge, high-pass filtering, and Ehlers fusion (Amro et al., 2011; Helmy, Nasr, & El-Taweel, 2010). Of these techniques, methods based on arithmetic/statistical image combinations proved to be a flexible and accurate way to fuse the multispectral (MSS) and panchromatic images (Amro et al., 2011). Of the various arithmetic/statistical approaches to pansharpening cited in the literature, one of the most successful was developed by Yun Zhang at the University of New Brunswick, Canada. Dr. Zhang's algorithm proved to be so successful that it was subsequently patented and licensed to PCI Geomatics (PCI Geomatics, 2014) and DigitalGlobe, Inc. That algorithm, known as the PanSharp™ module, is built into PCI Geomatics image processing software known as Geomatica. Within PCI Geomatica, PanSharp is an automated tool that incorporates BGRN image layer stacking, mosaicing, orthographic corrections, and rational function processing with pansharpening as an automated two-step automated process. The most significant benefits of this process were a significant reduction in the post-processing time and a reduction in the likelihood of operator error (as compared with a manual multiple-step process). For this reason, PCI Geomatica was used to post-process the entire GeoEye (satellite imagery) collection associated with this research.

Following post-processing, some images were less usable due to weather, clouds, and sun glint. Most of the image degradation noted occurred during 2011 at the Sandusky research site. The Sandusky area suffers from greater cloud cover than the Tampa research site. The issue results from generally poor atmospheric conditions

56

during some collection cycles. For this reason, DigitalGlobe provided additional imagery at no additional cost in 2013 for the Ohio research site.

The independent variables derived from these image sources include: trafficshed characteristics: traffic density, navigation channel orientation and complexity measurements (e.g., intersections, bridges, restricted navigable waterway, and the like), boat length, boat type, compass direction, water depth, estimated boat speed, and some weather and sea-state data. Weather and sea-state data were provided by the National Climatic Data Center (NCDC) and enabled the estimation of weather-sea state (atmospheric visibility, wind speed, wave height, weather conditions, water conditions, air temperature, water temperature, and small craft advisory state) data. These data were obtained from two sensors: NOAA Station KTPW in Florida and NOAA Station KCLE in Sandusky Bay. The NOAA National Climatic Data Center in Ashville, NC (National Environmental Satellite, Data, and Information Service) collect and archive these data while permitting access to educational researchers. Specifically, the 2010 and 2011 FZUS5 data files (Coastal Waters Forecast) for the KCLE and KTPW sensors were extracted (via the access path: http://has.ncdc.noaa.gov). These files were systematically examined for the issuance of "small craft advisory" warnings by the National Weather Service for waters near each of the research sites. The resulting list of dates during which small craft advisory warnings were issued was then cataloged by date in a spreadsheet for later review and analysis.

On-the-Water (OWS) Survey Data

The variables obtained through On-the-Water (OWS) Survey (DS-3) data were collected by the marine law enforcement officers of the Florida Fish and Wildlife

Conservation Commission (FWC) and Ohio Division of Natural Resources (ODNR) patrolling the Tampa and Sandusky research sites respectively. The on-the- water survey (OWS) data was collected during routine "non-accident vessel safety stops" by way of a high precision Trimble Nomad 800L GPS Data Logging Survey instrument (see Figure 5) using a OWS Survey (DS-3) Form (see Figure 6). Grant funding provided each participating FWS and ODNR marine law enforcement officer with individual handheld units. The specialized OWS software was written in collaboration with the Trimble Corporation and implemented on each Trimble device used in this study. That program generated a user-driven, data collection interface. The Trimble Nomad 800L enabled participating boating law enforcement officers to record each stop location with a high degree of precision. The data collected using the Trimble Nomad 800L contributed to the key variables captured as part of DS-3 as shown in Table 2.

Data Fusion

As previously described, all data collected during this investigation were assembled into an ArcGIS database facilitating advanced data fusion techniques. *This aspect of the project is a key feature that separates this research from all of the other studies that have preceeded it*. A generalized model of the data fusion technique used in this investigation is shown in Figure 7. This data fusion approach offers several investigative strengths. First, the collection and representation of all factors operating within the boating accident domain are fully achieved through the location of non-accident based boater characteristics in precise on-the-water locations. Second, the GeoEye satellite imagery collection provides sufficient optical clarity to establish ground-

58

Figure 5. Trimble Nomad 800L.

On-Water: Boating Safety Stop Survey

Officer Name:	E-mail:	Office Telephone ( )

Important Note: Date, Time, and Latitude/Longitude Coordinates are critical elements of this study that must be completed for each report.

Date of Stop	Time	AM/PM	Latitude (from GPS)	Longitude (from GPS)
☐ Day ☐ Night		☐ AM ☐ PM	DDD . MM.MMMM	DDD . MM.MMMM

Weather / Water Conditions / Temperature / Wind / Visibility

Weather
- ☐ Clear
- ☐ Cloudy
- ☐ Fog
- ☐ Rain
- ☐ Snow
- ☐ Hazy

Water Conditions
- ☐ Calm (waves less than 6")
- ☐ Choppy (waves 6" to 2')
- ☐ Rough (waves 2' to 6')
- ☐ Very Rough (greater than 6')
- ☐ No Current
- ☐ Strong/Swift Currents

Temperature
- Air [ ]
- Water [ ]
- (est. degree F)

Wind
- ☐ None
- ☐ Light (0-12mph)
- ☐ Moderate (13-24mph)
- ☐ Strong (25-54mph)
- ☐ Storm (over 55mph)

Visibility
- ☐ Good
- ☐ Fair
- ☐ Poor
- ☐ None

Boat Classification / Hull Material / Engine / Boat Characteristics / Onboard Navigation Tools

Boat Classification
- ☐ Cabin Motorboat
- ☐ Open Motorboat
- ☐ Personal Watercraft
- ☐ Pontoon Boat
- ☐ Airboat
- ☐ Houseboat
- ☐ Sail (with engine)
- ☐ Sail (no engine)
- ☐ Canoe
- ☐ Rowboat
- ☐ Kayak/Race Scull
- ☐ Paddle Craft (not canoe/row/kayak/..)
- Other (Specify) [ ]

Hull Material
- ☐ Fiberglass
- ☐ Aluminum
- ☐ Rubber/Vinyl/Canvas
- ☐ Rigid Hull Inflatable
- ☐ Plastic
- ☐ Wood
- ☐ Steel
- Other (Specify) [ ]

Engine
- ☐ Inboard
- ☐ Outboard
- ☐ Inboard/Outboard
- ☐ Jet Drive
- ☐ Inboard/Outdrive
- Other (Specify) [ ]

Type of Fuel
- ☐ Gasoline
- ☐ Diesel
- ☐ Electric

Propulsion
- No. of Engines [ ]
- Horsepower (total)

Boat Characteristics
- Vessel Length [ ]
- Year Built [ ]
- Boat Manufacturer: [ ]

Operator is the Owner
- ☐ Yes
- ☐ No

Rented Vessel
- ☐ Yes
- ☐ No

Onboard Navigation Tools
Navigation tools and devices onboard and in-use by operator?

	Onboard		In-Use
☐	GPS Handheld		☐
☐	GPS Chart Plotter		☐
☐	Paper Charts		☐
☐	Smart Phone		☐
☐	IPad (similar)		☐
☐	Computer (similar)		☐
☐	Radar		☐

Other Device/s (Specify) [ ] [ ]

Personal Flotation Devices (PFDs)

1. Is boat equipped with the required number of PFD's? ☐ Yes ☐ No
2. Are occupants currently wearing their PFD's? ☐ Yes ☐ No

If occupants are wearing their PFDs, Why? (operator self report)
- ☐ Non swimmer aboard
- ☐ Can swim but feel safer
- ☐ Required by Law
- ☐ Wear It Campaign messages
- Other [ ]

Responses below self reported by operator:

Boat Operations (at time of safety stop)
- ☐ At Anchor
- ☐ Cruising
- ☐ Dock/Undock
- ☐ Drifting
- ☐ Fishing
- ☐ Fueling
- ☐ Hunting
- ☐ Sailing
- ☐ Skiing
- ☐ Skin Diving
- ☐ Swimming
- ☐ Tubing
- ☐ Wake Boarding
- Other (Specify) [ ]

Did operator check weather forecast before going boating?
- ☐ Yes
- ☐ No

Boat speed prior to the safety stop?
- ☐ Not Moving
- ☐ Drifting
- ☐ Idle or Slow
- ☐ Plowing (not planing)
- ☐ Planing
- ☐ Accelerating
- ☐ Decelerating

Boat Information

Boat Information
Number of Persons Aboard [ ]

Boat Operator
- Operator Date of Birth [ ]
- (or alternatively) Operator Age [ ]
- Operator Gender
  - ☐ Male
  - ☐ Female

Boat Operator Ethnicity and Language

Ethnicity:
- ☐ White
- ☐ Hispanic
- ☐ Black or African American
- ☐ Asian
- ☐ Native American (Am. Indian/Native Alaskan)
- ☐ Other race

Language:
- ☐ Speaks Fluent English
- ↳ If not, indicate operator's preferred language [ ]

Operator Experience (with this type boat)
- ☐ Under 10 hours
- ☐ 10 to 100 hours
- ☐ 100 to 500 hours
- ☐ Over 500 hours
- ☐ None

Formal Boating Safety Instruction:

Basic Boating Course
- ☐ State Boating Course
- ☐ USCG Auxiliary
- ☐ United States Power Squadrons
- ☐ Internet
- ☐ None
- ☐ Other [ ]

Advanced Boating Course (beyond basics)
If yes, what course /s ?
- ☐ Yes [ ]
- ☐ No

Figure 6. OWS Survey Form (DS-3).

60
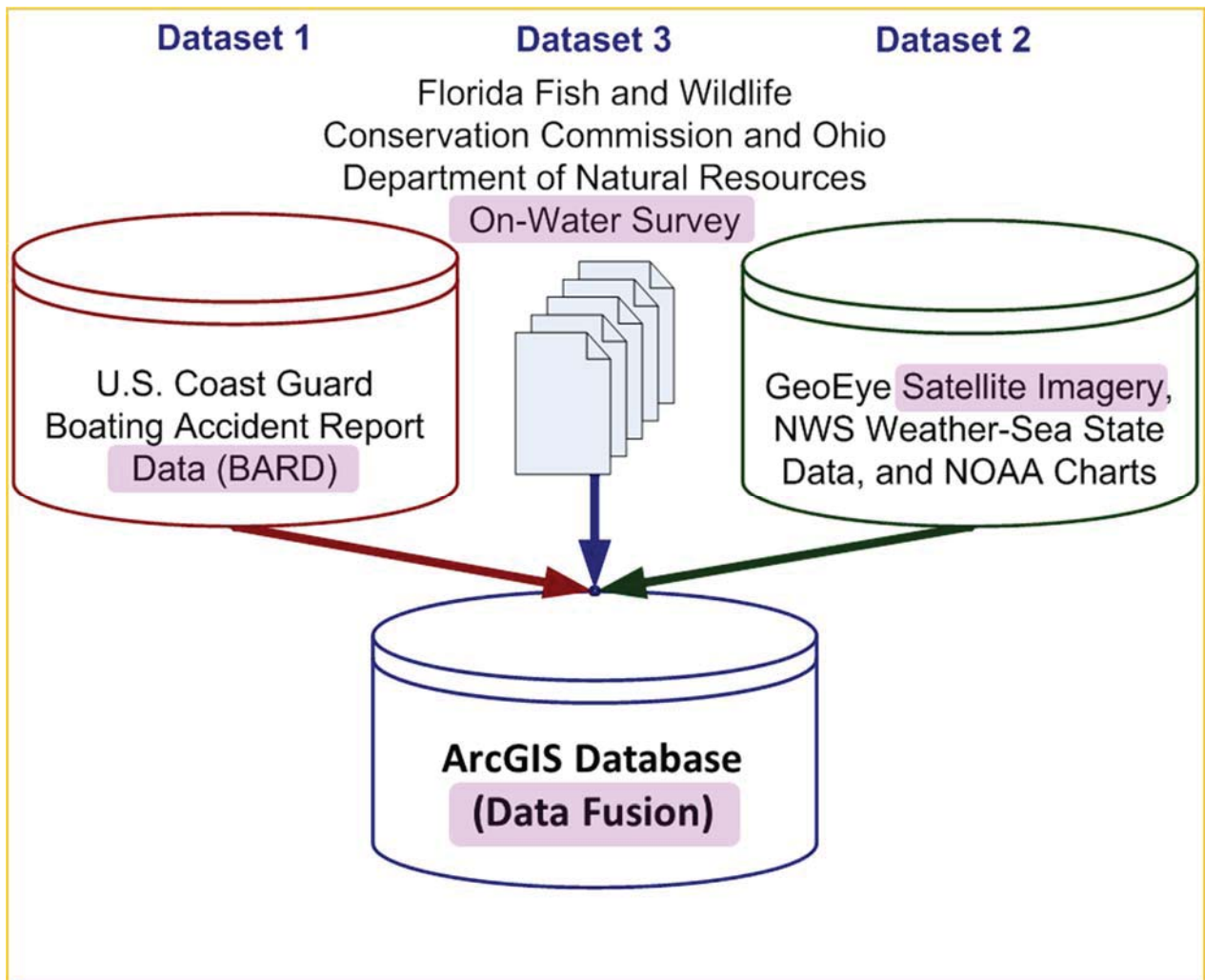
*Figure*7. ArcGIS Data Fusion Model.

truth for key environmental attributes within the accident domain. Lastly, the use of ArcGIS-based data fusion techniques to combine federal, state, and commercial data with other geo-spatial information (GIS layers) assures proper data synthesis (Sidman, Grant, & et al., 2005) and model development.

Methodology

The primary research methodology utilized in this study involved: site selection, data collection, analytic methods, and analysis. The first two of these have already been detailed and we now turn our attention to analytic methods. The dependent variable (number of boating accidents) is measured by the presence or absence of a recreational boating accident(s) within a specific location (geographically located by observations within 1mi radius of a .25 mi^2 grid cell centroid as shown in Figure 8). All independent variables are identified from satellite and on-the-water vessel safety check (VSC) data using the same radius capture technique as used for collecting the dependent variable observations. This methodology permits the capture of non-accident (recreational boating) data concurrently with boating accidents. However, such concurrent data has historically not been collected as the resources required to reconstruct them would be prohibitively expensive.

This research approach is based on a novel, more efficient, equally effective, and cost justifiable alternative utilizing a grid cell based sampling approach. The results, coupled with technological advances in applying data fusion techniques to this type of research, may lead to data collection protocols that can permit the simultaneous collection of both accident and non-accident recreational boating data.

Micro Level Modeling Approach

Primary focus to relate space-time array (ijt) to factors epresenting all boating accident dimensions: human, technological, and environmental:

$$A_{ijt} = f\left(human_{ijt}, technological_{ijt}, environmental_{ijt}\right)$$

A represents the number of accidents within cell (ij) at time (t)

Density measured as the number of boats within ½ **mi** of any cell centroid

Non-Accident Data (Satellite Imagery)

Cell
Unit of Observation (½ mi²)

• **Accident Data** (BARD)
• **Non-Accident Data** (On-the-Water VSCs)
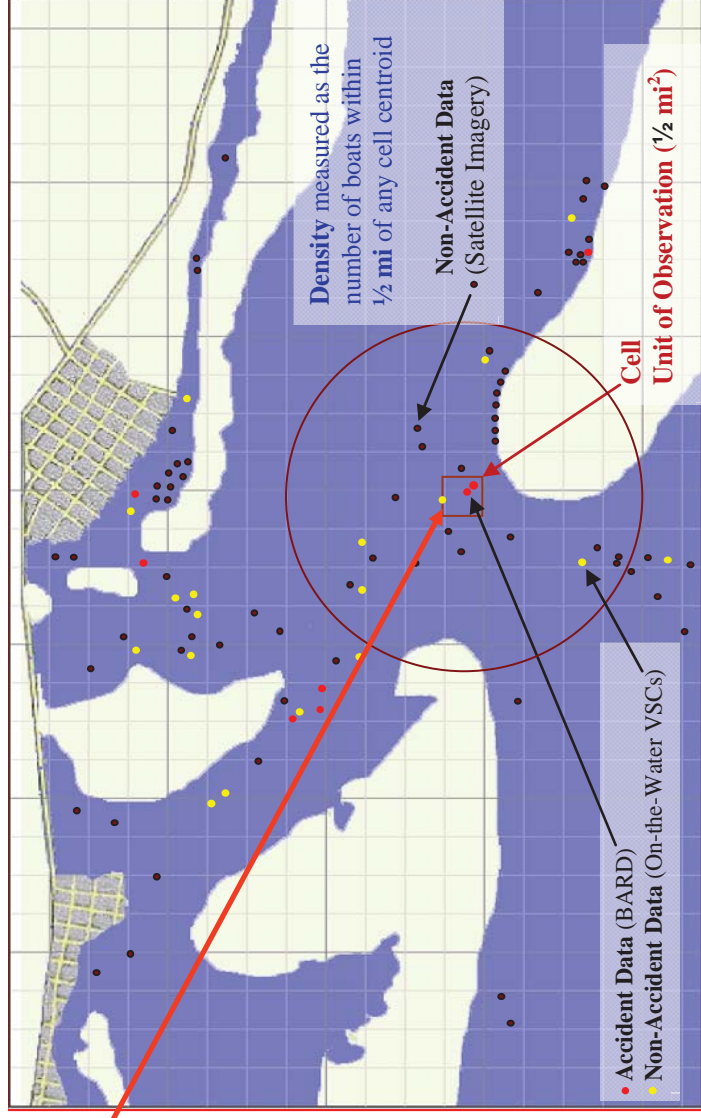
(Observations within 1 mi radius aggregated to the centroid of each cell centroid)

e 8. Unit of analysis, 0.25 mi² grid cell.

Unit of Analysis

A boat can be represented as a point with movement over time in x, y dimensions (Balaguer et al., 2011; Gray et al., 2011; Jaakson, 1989; Kujala et al., 2009; Prasannakumar et al., 2011; Sidman & Fik, 2005; Sidman & Flamm, 2001; Sidman, Grant, & et al., 2005). This representation enables data aggregation based on the use of grid cells as the observational unit. Thus the location of actual boating accidents and all boating accident dimensions are represented within a grid cell and assigned to the grid cell centroid.

The grid cell sampling frame divides the study site into a matrix of uniform cell size. A coarse initial grid cell (ArcGIS Fishnet) matrix was established at 23 x 39 (rows x columns) cells representing a 1 mi^2 area. A more refined grid cell matrix was established at 46 x 78 cells representing an approximate 0.25 mi^2 area. This second and smaller grid cell resolution enables a more refined micro-level spatial analysis. The selection of these grid cell areas is based on prior research suggesting that units of observation smaller than 0.25 mi^2 are too densely packed and that cells larger than 1.25 mi^2 are too sparsely dispursed to effectively capture recreational boating accidents and activities (Jaakson, 1989). This finding is also suggested within contemporary spatial analysis research involving recreational boats (Balaguer et al., 2011; Gray et al., 2011; Jaakson, 1989; Kujala et al., 2009; Prasannakumar et al., 2011; Sidman & Fik, 2005; Sidman & Flamm, 2001; Sidman, Grant, & et al., 2005).

Radius captures at 1.25 mi, 1 mi, and 0.5 mi were tested and the 1 mi distance from the grid cell centroid was found to be optimal in capturing significant relationships (in terms of their influence on boating accidents). This approach, based on tagging all

64

observations in space and time, permits data fusion and varied forms of analyses. It

also permits relationships between variables to be modeled using advanced modeling

approaches and enabling the investigation of potential interactive effects.

The accident value of the dependent variable is derived from individual boating

accidents, contained within the USCG BARD data (see Table 2: Variable Matrix, DS-1).

These observations include reference placement within a grid cell.

Normalized Research Data

In contrast to prior research on boating accidents and risk, the on-the-water

research survey strategy utilized in this investigation provides normalization of the 2011-

2012 BARD data via a non-accident boating population. This is a clear strategic

advantage over prior efforts that rely solely upon land-based surveys to collect

normalizing data (i.e., surveys conducted at marina ramp/dock or mailed surveys to

registered boat owners). While some of the previous land-based approaches did reach

larger samples, the survey responses are based on anecdotal rather than real-time

responses to real-life boating situations, circumstances, and activities. In addition, they

do place the normalizing data within reasonable proximity of boating accident locations.

The advantage of using boating law-enforcement agents to conduct the survey is that

the respondents are more likely to respond accurately and authentically to boating law

enforcement officers than to private citizens, mailed, or phone-based research surveys.

Special protocols were also developed and used during all boating law enforcement

officer on-the-water surveys. Each participant was initially asked if they would be willing

to participate in a study designed to collect boating safety related data as associated

with a recreational boating accident research project. In all but one instance, boaters

stopped voluntarily to participate in this project survey. These survey data provide the

full variation of boaters within the sites and permit this variation to be associated with

cells, e.g., the age of boaters might vary systematically across a research site.

Survey Data Preparation and Post-Processing

The on-the-water data collected by the FWC and ODNR boating law enforcement

officers were received in MS Excel spreadsheet format and only required basic data

clean-up to ensure quality and consistency. All BARD was received in MS Access

format and required export to an MS Excel spreadsheet prior to basic data clean-up.

Processing the BARD data proved to be more complicated than expected as the MS

Access BARD files contained the entire BARD dataset for the state and for the years

between 1995 and 2012. Although the extraction process for the BARD data proceeded

smoothly, it was complicated due to file size. In order to be utilized in this investigation,

an extraction of subsets specific to the counties represented by the two research sites

was required prior to export. Once extracted in MS Excel spreadsheet format, only

basic data clean-up was required to ensure variable response consistency. The most

significant example of this clean-up dealt with consistencies in position entry

(Latitude/Longitude of accidents). A few entries were provided in Latitude and

Longitude. However, a majority of the entries were provided in degrees, decimal

minutes. This format required conversion using the formula: Decimal Degrees =

Degrees + minutes/60 + seconds/3600 to convert all positional data to the appropriate

format.

Subsequent to GeoEye satellite imagery post-processing (described earlier),

each image was imported into ESRI ArcGIS for final image assembly preceding

66

analysis. During this step, various image stretches were applied to improve overall image quality and the minimum-maximum stretch was found to produce the best results. This process also involved the creation of an ArcGIS fishnet grid (0.25mi^2) to enable the subdivision of each image into the desired 0.25 mi^2 unit of observation.

Although other GIS layers such as cities, county boundaries, land area, etc. were added for reference purposes, another primary focus was the incorporation of a geographically referenced NOAA navigation chart (digital version). The NOAA navigation charts (one for the Tampa research site one for the Sandusky research site) created a spatially accurate reference from which selected environmental variables collected in connection with this research could be extracted. To permit viewing of both the satellite image and nautical chart overlays simultaneously, the NOAA chart layer properties were adjusted to a 75% transparency setting.

Data extraction for each satellite image was accomplished by visually inspecting each grid cell and placing a point feature on each observable boat (target). Two scales of image magnification were used to facilitate the search and location of boats in these images. Those magnifications were a 1:4000 scale for rapid and wide area target detection and a 1:500 scale facilitating accurate placement of the point feature at approximately the center-mass of the identified boat. This process was repeated for each grid cell in each GeoEye satellite image with the initial visual scans occurring horizontally from left to right originating from the bottom of the GeoEye satellite image. No reliable automated procedure for this type of extraction was available for this project. To ensure that all boats were effectively captured, i.e., none inadvertently overlooked or mislabeled, this process was repeated for each image but the second time using a

67

visual scan vertically from top to bottom and left to right across each GeoEye satellite image.

Following the location of each observable boat (count = 2539, FL and 3232, OH) within the available GeoEye collection, the following variables were assigned to each point feature listed in the corresponding ArcGIS attribute table:

a. boat length in feet as measured with the ArcGIS measuring tool,

b. boat type as measured by the size, speed, and shape of the target,

c. estimated boat speed as measured by an evaluation of the boat wake and ascribed values of fast, slow, idle, or stationary,

d. compass direction as measured in relation to the north oriented NOAA nautical chart (an eight point compass rose; i.e., N, NE, E, SE, etc. was used),

e. water depth (bathometry) as measured from the NOAA nautical chart overlay,

f. navigational waterway complexity as measured from the NOAA chart overlay and ascribed a value of yes or no depending on the presence of features such as open water, narrow channel, intersecting channels, bridges, etc.,

g. NOAA designated shoals as indicated on the NOAA chart overlay, and

h. whether the area/boats identified were in an observable boat anchorage.

Once these attributes were assigned to each individual point feature (in each respective satellite image), the point feature attribute tables were extracted into an MS Excel spreadsheet for subsequent examination and analysis.

Variable Measurement

For descriptive purposes, the manner in which all variables are extracted from

GeoEye collection is linked to the observational grid cells as described earlier. The

specification used in this research is a 1 mi radius from each grid centroid. Thus, the

influence of each independent variable (on accidents) is assessed within a 1 mi ring

around each cell centroid. However, it is acknowledged that this approach is not the

only alternative (Anselin, 2002). For example, the window of assignment could vary

depending on boat density found during analysis (e.g., 0.5 mi in dense areas to 1.25 mi

in less dense areas) or be based upon the nearest 30 boats to the cell centroid thus

guaranteeing a valid sample of boats/boaters for each cell value. For this reason,

several methods were examined to determine the best assignment approach including

fixed window (0.5 mi), varying window (0.5-1 mi), and nearest neighbor (nearest 30

boats). Based on that examination and experimentation, a fixed window approach was

used with data observations captured using a 1 mi radius from each grid cell centroid.

Environmental variables. Boat density is measured as the number of boats

counted within 1 mi of each cell centroid. These counts were totaled and assigned to

each cell. To measure navigation channel complexity, NOAA nautical charts (see Table

2, DS-3) were superimposed over the GeoEye collection allowing the physical

characteristics of the trafficshed to be measured. Primary channel width and depth

were measured as an aggregated average value for each grid cell. The measurement

of average traffic speeds, MSS satellite imagery was attempted by examination of

frequency shifts between the multispectral bands (red, green, and blue) and the

panchromatic bands as suggested by Zhang & Xiong (2006). However, following

extensive work with GeoEye technicians, it was determined that accurate measurement

69

of this shift involving the GeoEye BGRN bands was not possible at this time. As a consequence, mean traffic speeds were assigned to each grid cell by measuring the generalized wake length of all boats using the speed labels (fast, slow, idle, stationary). These speed descriptors are the same as those used by boating law enforcement officers during on-the-water (VSC, citation, and warning) boating stops.

Visibility, wind, current, and wave influences vary according to the physical characteristics of the trafficshed. However, their effects do influence surface conditions and correspondingly boat navigation. These data were collected as a combination of marine law enforcement officer observations at the time of on-the-water stop or boating accident report coupled with information derived from NOAA National Climatic Data Center's National Environmental Satellite, Data, and Information Service. These conditions were averaged by month then assigned to the corresponding grid cell centroid as associated with each image in the GeoEye collection.

The physical characteristics of a trafficshed determine the nature of the navigation channels through which a boat transits. These channels sometime contain navigation hazards such as multiple and intersecting waterways (similar to highway intersections), waterway narrowing due to geographic constraints, shoals (areas of reduced water depth), sharp bends in the navigation channel, and other obstructions (e.g., pilings and rock jetties) that increase the probability of boating accidents. Each trafficshed grid cell was examined using the NOAA electronic navigation charts to determine if one or more of these navigation hazards existed. This allowed varied hazards present (either individually or in various combinations) to be entertained as a

measure of overall navigation channel complexity (a potential explanatory variable) and then assigned to trafficshed grid cells.

Technological variables. Boat propulsion, length, and type were captured via survey in Table 2, DS-1, (involving accidents) and Table 2, DS-3 (general boat population). Again, each survey observation is tagged with spatial and temporal attributes that will be used to assign each value to a specific grid cell centroid.

Human variables. Table 2, DS-1, and Table 2, DS-3 (accident and survey data), were collected and examined for boat operator age, ethnicity, gender, education, and experience by survey. Each survey observation was also tagged with spatial and temporal attributes used to assign means and variations to specific grid cells. It should be noted that several of these variables were categorical in nature.

Temporal variables. Table 2, DS-1, DS-2 and DS-3 (incident, remote sensing, and survey data) were examined for peak vs. non-peak, weekend vs. weekday, and time of day boating frequencies. Each survey observation was tagged with spatial and temporal attributes that were used to assign each observation to a specific grid cell. Peak periods were defined as the period between May and August while non-peak periods were based all other months. Weekends will be measured as the sum of all incidents occurring on Friday, Saturday, or Sunday. Time of day was aggregated to three hour blocks starting at midnight.

Population Characteristics

The individuals examined in this investigation are all recreational boaters operating different forms of boats in on-the-water environments. That is, all

measurements (GeoEye observations, survey data, and BARD data) are based on *actual on-the-water observations* at each research site. No land-based (ramp, marina, dock, mailing, etc.) recreational boater survey data is included in the research.

Adequacy and Representativeness of Normalizing Groups

The primary weakness or gap associated with previous recreational boat accident research is associated with how the USCG BARD data is normalized. USCG BARD data is a complete record of recreational boating accidents but contains no non-accident boating data that the BARD data can be compared to. In other words, there is no variation in the dependent variable because all of the records represent boating accidents. To address this weakness, this study adopted a novel approach. The unique approach used was to collaborate or partner with two states and thus two state boating law enforcement agencies (i.e., FWC and ODNR) to capture non-accident survey data during routine boating law enforcement officer vessel safety stops (see Figure 9). VSC stops, analogous to automotive driver's license check points, are designed to ensure recreational boat operator compliance with state boat licensing laws and federal boating safety carriage (equipment) laws. Such traffic stops are not routinely recorded. However, the research collaborative established with Florida and Ohio, assigned selected boating law enforcement officers to this project. These officers participated in the capture of VSC observations during 2011-2012. This VSC data mirrors the BARD data extracted for the two research sites. Consequently, the collection of these data enabled non-accident based recreational boating values to be readily compared with the same factors observed in connection with actual boating accidents. The populations of recreational boaters reflect the similarities and

Note. ODNR Officers (top and middle); FWC Officer (bottom); boating law enforcement officers conducting vessel safety stop (VSC). FWC Photo courtesy of Florida Fish and Wildlife Commission

Figure 9. On-the-Water, Vessel Safety Check Stop.

differences between the two research sites and the research design provides a representative sample of these recreational boaters.   Boaters involved in accidents are another subset of the population and they are probably not representative of the larger population. So, the inclusion of VSC data is adequate and appropriate as a normalizing strategy and is representative of the study population in question. That sample permits direct comparison with those involved in recreational boating accidents.

Description of Analytic Techniques

Boating accidents are relatively rare events when confined to a small study area. In this study, OLS and generalized regression models (e.g., Poisson and negative binomial regression) are used to model count response data with acknowledgment that Poisson or negative binomial regression is generally the fundamental modeling method used with count data (Hilbe, 2011). Each technique has the potential to manifest performance issues depending on the explanatory variables specified. These performance issues are most frequently observed through variable omission and multi-collinearity. Generalized regression models frequently, although not exclusively, rely on maximum likelihood techniques to estimate effects (Hilbe, 2011). Where n is large, such as is the cases with the on-the-water and satellite imagery data, those concerns are reduced. However when n is small, as in the case of the accident count related to the more restrictive 2011-2012 model, results may mask model problems. For these reasons and to ensure a sufficient pool of accidents within the research sites, a 1 mi radius capture technique was used to designate BARD accident data for assignment to each respective grid cell centroid.

Building a Case for the Use of Negative Binomial Regression

When modeling data statistically, research investigators frequently begin with an Ordinary Least Squares (OLS) approach. OLS is readily understood and produces an equally well understood R^2 value as a measure of statistical significance. Briefly, R^2 represents a way of demonstrating how the dependent variable is related to the independent variables collected during the course of the study. Regression is useful in illustrating relationships between the dependent and independent variables. On the other hand, OLS is restricted by assumptions involving the Central Limit Theorem (Atkins & Gallop, 2007). One of those restrictions involves sample size. The Central Limit Theorem states that, "as sample size increases the sampling distribution of the mean or regression coefficient becomes normally distributed regardless of the shape of the original distribution in the sample" (Cohen et al., 2013). OLS regression requires a normal distribution on the error terms to generate a probability model.

As in the case of this study, the variables are measured as count data. Because the normal distribution is symmetric, extending potentially from negative to positive, and count data is never negative (even though it may be skewed by low count observations), OLS may not be a good fit. On the other hand, OLS is always a reasonable first step in model building (Atkins & Gallop, 2007). If the resulting model is not stable, two count regression alternatives are Poisson and negative binomial regression. Poisson regression generally offers a better fit for the evaluation of count-based data. However, Poisson regression shares many similarities with OLS regression. With the exception of residual distribution, Poisson regression, like OLS regression, is assumed to follow a Poisson distribution model (Atkins & Gallop, 2007).

The predictors are connected to outcome variables through a natural logarithmic transformation, which in OLS regression transforms the outcome variable in order to normalize residuals (Atkins & Gallop, 2007). In Poisson regression, the log transformation ensures that model predictions will not be negative (Atkins & Gallop, 2007).

These regression limitations are illustrated through the highly restrictive assumptions Poisson regression makes with respect to the relationship between the conditional mean and conditional variance (Atkins & Gallop, 2007). The consequence is that Poisson regression assumptions are almost never met (Atkins & Gallop, 2007; El-Basyouny & Sayed, 2006; Gardner, Mulvey, & Shaw, 1995; Gonzales-Barron et al., 2010; Greene, 1994; Grün & Leisch, 2007; Hilbe, 2011; Hoef & Boveng, 2007).

As noted earlier, a radius capture method for aggregating observational data is used in this research to capture sparsely distributed boating accident and non-accident data. By observation, a majority of the grid cells can be characterized as simply open water, i.e., containing no boating accidents and few, if any, on-the-water surveys or satellite-observed boats. In other words, many grid cells have zero count values (see Figure 10). Excess zero's within data is common. In fact, excess zero's in count-based data is actually a very common phenomenon faced by many researchers (El-Basyouny & Sayed, 2006; Gonzales-Barron et al., 2010; Greene, 1994; Hilbe, 2011; Hoef & Boveng, 2007). The challenge of modeling high frequency zero-valued data is one of determining the most appropriate model to use.

Frequently, statistical models begin with "ordinary least squares" (OLS). OLS is readily understood by those new to statistics (Hilbe, 2011; Shmueli, 2010).
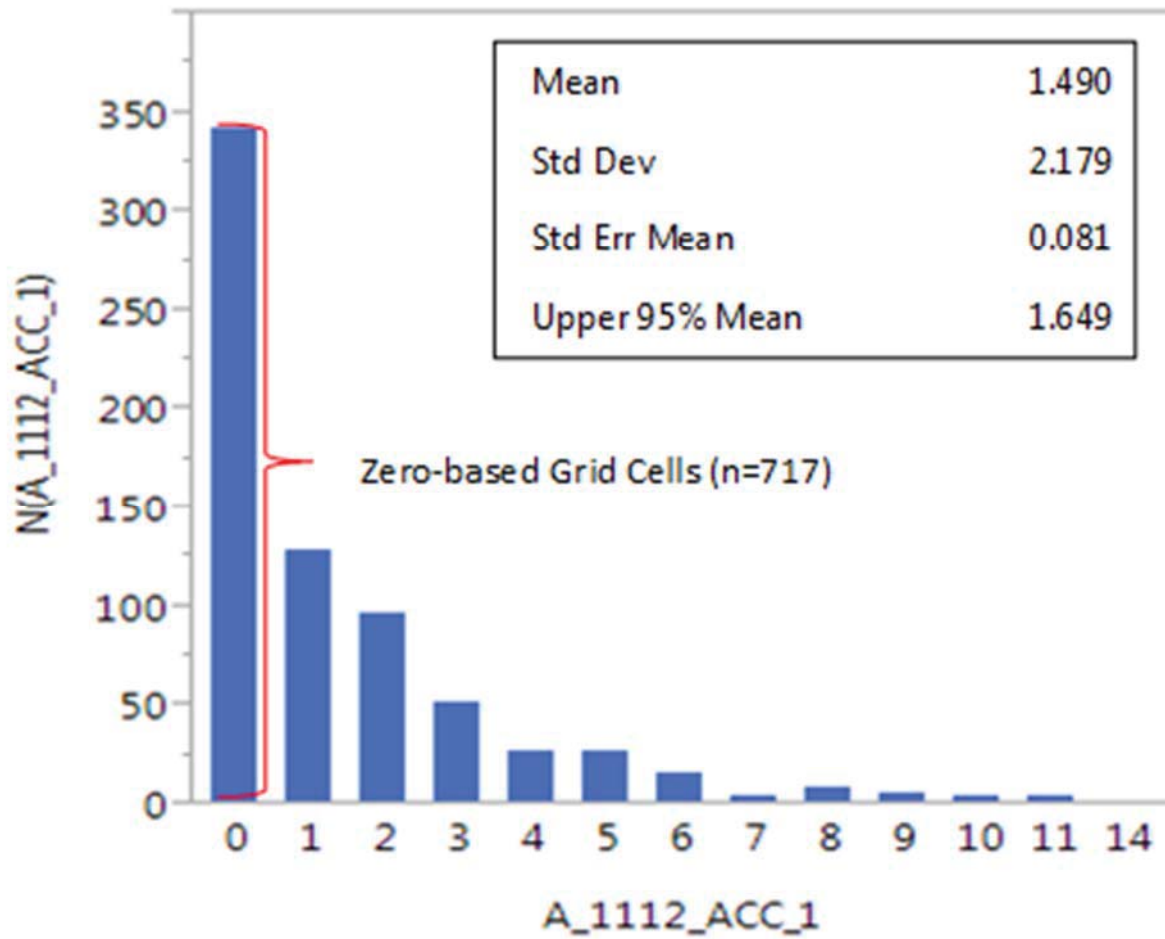
Figure 10. Distribution of grid cell counts showing BARD data.

However, as noted above, neither OLS nor Poisson offer the best methodology for representing count data with a high frequency of zero values. While OLS regression methods can be used to model count variables, linear models may generate negative predicted values and count data can never be negative (Hilbe, 2011; Shmueli, 2010). The other problem associated with OLS regression is that count data can be highly skewed. For example, the extreme difference between a large number of zero-based grid cells with cells containing counts of one or more violate the OLS normality assumption. The limitation of Poisson regression as illustrated in this study relates to the assumption that the variance equals the mean (Hilbe, 2011; Hoef & Boveng, 2007; Morel & Neerchal, 2012). If the variance is greater than the mean, a condition called over-dispersion exists and if the variance is less than the mean, a condition called under-dispersion exists (Hilbe, 2011; Hoef & Boveng, 2007; Morel & Neerchal, 2012). These conditions result in inefficient parameter estimation when employing OLS or Poisson approaches.

A more common approach used to model count data is negative binomial (NB2) regression (Hilbe, 2011; Shmueli, 2010). For purposes of this research, the negative binomial approach is superior because it accommodates the high frequency of zero values and the heavy tail of the distribution, i.e., a few accident hot spots. The negative binomial approach begins with a Poisson regression model then "adds a multiplicative random effect to represent unobserved heterogeneity", i.e., over- and under-dispersion (Atkins & Gallop, 2007; Hilbe, 2011; Hoef & Boveng, 2007). This NB2 characteristic coupled with a relaxation of OLS and Poisson conditional assumptions highlights why NB2 regression is used more frequently when examining distributions with frequent zero

values (Fang, 2013; Hilbe, 2011; Hoef & Boveng, 2007). Thus a Negative Binomial (NB2) regression approach is most applicable to modeling the type of count data in this research as it permits models that account for under- and over-dispersion (Atkins & Gallop, 2007). Due to OLS and Poisson regression limitations, a Negative-Binomial regression model is proposed and developed in this dissertation.

Screening Predictors

In this study, a pre-modeling decision tree analysis involves the use of the SAS JMP bootstrap forest routine, an extension of Leo Breiman's 'random forest' concept (Breiman, 2001). Bootstrap forest is a method of assessing the statistical accuracy of a potential model using random variable selection to identify those variables that most significantly influence the dependent variable. This technique is sometimes referred to as a form of data mining (Klimberg & McCullough, 2013). It is an iterative resampling process that creates new datasets from the original data (Efron & Tibshirani, 1994). It then averages the predicted values of each resample (tree) to obtain a final predicted value for the predictor variables constellation (Stine, 2007). This process does not work in isolation but rather as one used to refine the number of variables ultimately included in a regression model (Efron & Tibshirani, 1994; Grün & Leisch, 2007). Statistically, bootstrapping assigns a measure of accuracy to observation variables by estimating variables properties (e.g., variance) and their influence (Klimberg, 2013, #2473; Stine, 2007, #2480). The result of bootstrapping is a dataset that can be mined to build effective predictive models.

In this study, the initial variable constellations selected for bootstrapping analysis contained a majority of the BARD data and On-the-Water, and Satellite observations.

79

However, initial bootstrap forest observations suggested that the inclusion of any accident (BARD) variable tended to dominate the resulting model. This was found to be the case even if only a small subset of accident variables was included (see Figure 11). As discussed in Chapter 4, a majority of all boating accidents can be attributed to one of two BType variables (powerboats and jetskis). For this reason, no BARD variables were included as explanatory variables in bootstrap forest analyses or final modeling.

Approximately 150 bootstrap forest combinations (various iterations of different explanatory variable combinations) were tested to narrow the range of explanatory variables (from approximately 350 variables in the raw dataset) to a more manageable constellation of 30 candidate variables (see Figure 12). These 30 candidate variables are characterized as 18 satellite and 12 on-the-water (VSC) variables with each exhibiting an above average influence on the outcome variable. Within this constellation, 7 are classified as human factor-based, 3 are classified as technology factor-base, 14 are classified as environmental factor-based, and 6 are temporal in nature (see Table 4). This subset illustrates a balanced distribution between the non-accident-based datasets further illustrating the importance of integrated data in this research.

The Negative Binominal Regression (NB2) Model

The modeling sequence employed in this study began with an examination of Ordinary Least Squares (OLS) regression, a commonly used regression analysis that is employed to generate predictions. This is done by regressing observed values to a line of best fit that results in a set of predicted values.

Response	A_1112_ACC_1			
Distribution	OLS Fit			
RSqure	0.942			
RSqure Adj	0.942			
Mean of Response	0.410			
Observations	717			
AICc	309.113			

Term	aEstimate	Std Error	tRatio	Prob>|t|
Intercept	0.002	0.012	0.150	0.879
A_OpEdu0	0.292	0.029	9.960	<.0001
A_BType1	0.693	0.026	27.000	<.0001
A_BType3	0.807	0.024	34.010	<.0001

Figure 11. Dominance of accident variables in OLS Model

Bootstrap Forest for A_1112_ACC_05r

Specifications

Target Column:	A_1112_ACC_05r	Training rows:	717
		Validation rows:	0
Number of trees in the forest:	100	Test rows:	0
Number of terms sampled per split:	7	Number of terms:	30
		Bootstrap samples:	717
		Minimum Splits Per Tree:	10
		Minimum Size Split:	5

Overall Statistics

Individual Trees	RMSE
In Bag	0.497828
Out of Bag	1.007227

RSquare	RMSE	N
0.633	0.6107947	717

Column Contributions

Term	Number of Splits	SS	Portion
V_WDepth2	131	22.980672	0.0819
S_Intersect1	188	21.0639619	0.0751
S_BLength2	250	19.5318443	0.0696
S_BSpeed3	190	17.8258428	0.0636
S_BType3	168	16.281093	0.0580
V_Channel1	108	15.5280563	0.0554
V_BSpeed3	180	12.2743962	0.0438
V_OpAge2	164	10.8105167	0.0385
S_Constrain_1	122	10.6889448	0.0381
S_WDepth2	184	10.4369115	0.0372
S_Shoal1	172	9.65357263	0.0344
V_Constrain_1	93	9.42091254	0.0336
S_WDepth3	167	8.66686423	0.0309
V_WDepth1	77	7.51913544	0.0268
S_BType1	165	7.06717006	0.0252
S_JunJul	163	7.00594946	0.0250
S_Density1r	172	6.99119504	0.0249
S_MayJuly	147	6.89253124	0.0246
S_SCA_0	158	6.81728347	0.0243
S_WDepth1	148	6.60497537	0.0235
V_PM	145	6.19714563	0.0221
V_OpEdu0	146	6.15481734	0.0219
V_OpAge3	108	5.94416261	0.0212
V_JuneJuly	142	5.729385	0.0204
S_Channel1	113	5.36182292	0.0191
S_Weekend	128	4.82853558	0.0172
S_Weekday	150	4.7577738	0.0170
V_OpExp_N0	43	3.14744751	0.0112
Site	121	2.43116176	0.0087
V_OpAge1	71	1.88293501	0.0067

Figure 12. Bootstrap Forest 30 candidate variables.

Table 4

Baseline Variable Constellation

	Human	Technology	Environment	Temporal
1	S_BSpeed3	S_BLength2	S_Channel1	S_JunJul
2	V_BSpeed3	S_BType1	S_Constrain_1	S_MayJuly
3	V_OpAge1	S_BType3	S_Density1r	S_Weekday
4	V_OpAge2		S_Intersect1	S_Weekend
5	V_OpAge3		S_SCA_0	V_JuneJuly
6	V_OpEdu0		S_Shoal1	V_PM
7	V_OpExp_N0		S_WDepth1	
8			S_WDepth2	
9			S_WDepth3	
10			Site[1]	
11			V_Channel1	
12			V_Constrain_1	
13			V_WDepth1	
14			V_WDepth2	
15	(7)	(3)	(14)	(6)

Note. S=Satellite Observation; V=On-the-Water Observation.

OLS follows the formula:

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...\beta_p X_p + \text{e}$$

(1)

where:

E(Y) is the predicted value of the dependent variable;

β_0 is a additive constant, the value of Y axis when all X=0;

β_1 ... β_p are the coefficients representing the impact of X variables,

X_1 ... X_p are the observed values of the independent variables for each observation,

and e represents the error term (in this case estimators) that define E(Y).

From a general perspective, the error term (e) represents the effect of variables omitted

from the model and more importantly a source of unexplained variation or alternatively,

random variation.

The general linear model used in this study is a negative binomial (NB2) model.

NB2 regression analysis is very similar to OLS regression but without the error term.

This is due to expectations expressed as probabilities of the expected counts. The NB2

formula is expressed as:

$$ln\left\{E\left(Y\right)\right\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...\beta_p X_p$$

(2)

where:

ln{E(Y)} is the natural log of the expected value of the dependent variable,

β_0 is a constant, the point at which the line crosses the Y axis when X=0;

β_1 ... β_p is a coefficient representing the slope of the line, and

X_1 ... X_p is the observed value of the independent variable for each observation.

The case for an NB2 model was outlined in the section *Building a Case for the Use of Negative Binomial Regression* where it will be used to analyze count data (observed over a 2 year period) explain the estimation of data where dispersion is suspected (Fekedulegn et al., 2010; Morel & Neerchal, 2012). Unlike the OLS and Poisson regression techniques, NB2 models contain an additional parameter that permits an examination of "model data heterogeneity while accounting for dispersion" (Fekedulegn et al., 2010). This relationship between the variance and mean can be expressed in the form:

$$V(Y) = E(\mu) + [E(\mu)]^2 \tag{3}$$

where μ is the overall adjusted mean and $\mu=0$ yields the Poisson distribution (Fekedulegn et al., 2010). This model assumes the same form as shown in equation 2 except for the additional dispersion parameter (μ) that allows accounting for variation due to other factors not included in the model. If (μ) represents the mean, then variance can be expressed as:

$$V = \mu(1 + k\mu), \tag{4}$$

where *k* represents the dispersion parameter. If *k* is less than 1, the NB2 model is characterized as being under-dispersed and thus NB2 modeling approaches would be more appropriate than OLS or Poisson models. Of note, Poisson regression is actually a special case of the negative binomial regression where *k*=0. From a modeling perspective, knowing variance (V) is important for its use in generating p-values and inferences. Because the data in this investigation is under-dispersed,

inferences drawn from the NB2 model are more accurate due to the increased accuracy of this parameterization (Fekedulegn et al., 2010; Morel & Neerchal, 2012). It is equally important to understand that k is not calculated to support the model but rather to permit proper adjustments for standard errors and p-values. It is also possible to express variable estimators in terms of the natural log of the mean counts over a two year period as shown in equation 5 where B_i represents one of the variable estimates:

$$B_i = \ln(\mu_2) - \ln(\mu_1) \tag{5}$$

$$B_i = \ln\left(\frac{\mu_2}{\mu_1}\right) \tag{6}$$

The natural log is removed on the right side of equation 6 though exponentiation (e) of both sides of the equation as shown in equation 7:

$$e^{Bi} = \frac{\mu_2}{\mu_1} \tag{7}$$

Based on the formula in equation 7, the variable estimates can best be expressed in terms of an incidence rate (IRR) parameterization (Hilbe, 2011). This is due to the fact that log counts are difficult to explain in practical situations; therefore, "most statisticians prefer working with rate ratios" (Hilbe, 2011). Subsequent interpretation of the IRR is based on its potential value range from 0 through infinity with: (a) an IRR < 1 indicating under-dispersion, and (b) an IRR > 1 indicating over-dispersion. To make this

relationship even simpler to explain, the IRR is frequently converted into a percentage as shown in the following formula:

$$\%Change = 100*(e^{Bi} - 1.00)$$ (8)

For example, if the exponentiation of the parameter estimate shown for S_Density1r (1.080) is examined and converted to a percentage, then a 1 unit increase in x would correspond to a 8.0% (.080*100) increase in the accident rate (i.e., during the two year period of this research). The influence of the IRRs will be explained in greater detail in Chapter 5.

To examine colinearity as a quality of the final explanatory variables selected, the general statistical techniques used in OLS can be applicable to NB2 models (Klimberg & McCullough, 2013). One such OLS colinearity measure is variance inflation factor (VIF). As an additional way to validate the stability of the OLS model, the VIF value of each model variable will be examined. VIF is an index measurement reporting the degree to which the variance (estimate's standard-deviation^2) associated with the regression coefficient is due to collinearity (Klimberg & McCullough, 2013). While a small degree of correlation between explanatory variables does not significantly degrade model performance, too much correlation biases model fit. A VIF of 1 suggests that there is no multi-colinearity and values in excess of 5 but less than 10 suggest that correlation may be present but not problematic (Martz, 2013). VIF values exceeding 10 are heavily biased by multi-colinearity (Martz, 2013).

In the next chapter of this dissertation, some key space-time patterns associated with recreational boating accidents will be illustrated statistically and graphically. This

descriptive treatment of the data is a necessary prelude to final variable selection before

attention is given to the development of the desired model.

CHAPTER 4: KEY SPACE-TIME PATTERNS OF

RECREATIONAL BOATING ACCIDENTS

All events occur within the temporal and spatial domains. Patterns of events within these two key domains (i.e., time and geographic space) are useful in that they can reveal actionable intelligence, e.g., optimal boating law enforcement resource (asset) allocation. Thus the purpose of this chapter is to employ basic graphical and statistical devices that have the potential to reveal recreational boating accidents patterns at both the Tampa and Sandusky research sites. While this treatment is entirely descriptive, it is a necessary prelude to asking the important question - where do boating accidents occur? The "why" question is explicitly posed within a multivariate modeling framework that is presented in Chapter 5.

The data used in this descriptive treatment emanates entirely from the Federal Boating Accident Report Database (BARD) maintained by the United States Coast Guard. The foundation for this descriptive analysis is the BARD accident record datset specific to the 2005-2012 time series. Attempts to model recreational boating accidents will be reserved for Chapter 5 and employ a more limited BARD time series (2011-2012). This limited series treatment during the modeling phase is necessary so that BARD data can be specifically matched with the GeoEye imagery and on-the-water survey data assembled for this dissertation.

Chapter 4 is organized in three primary sections. The first section details the timing of boat accidents, i.e., when they occur including annual variation, monthly variation, daily variation, and time of day variations. Spatial distributions are then

disaggregated by key temporal categories to investigate the spatial dynamics of recreational boating accidents at the two research sites. Maps and spatial analyses will be produced to facilitate comparison of key temporal categories: annual cycles, seasonal, weekday, and time of day. Both research sites are subsequently compared and contrasted across these temporal resolutions. The second section focuses on the geography of recreational boating accidents, i.e., where they occur and associated environmental characteristics (correlates) at those locations. All accidents occurring between 2005 and 2012 are mapped. The boating accident spatial patterns revealed are described and analyzed using point pattern and polygon density maps to support spatial statistical analyses. Key clusters are identified and the levels of clustering are compared and contrasted across both research sites. These treatments permit higher order spatial clustering and hotspot analyses using Moran's I and Getis-Ord (Gi*) statistics. The chapter concludes with a statement on the nature of recreational boating accidents, including the number of boats involved, as a foreshadowing of the multivariate model that will be developed in the next chapter.

Temporal Variation of Boating Accidents within the Research Sites

Table 5 and 6 highlight the 2005-2012 time series distribution of boating accidents at the Tampa and Sandusky research sites that are then illustrated spatially in Figure 13. These data are intended to offer a broad boating accident perspective while permiting appropriate comparison of the research sites. As shown in Table 6, the Tampa research site experienced 172 boating accidents while the Sandusky research site experienced 155 boating accidents. The totals are similar and the resulting means (accidents per year) are 21.5 (Tampa Bay) and 19.4 (Sandusky) respectively.

Table 5

Boating Accidents 2005-2012

	Tampa Research Site		Sandusky Research Site	
Year	Frequency	Percent	Frequency	Percent
2005	10	5.8	25	16.1
2006	20	11.6	10	6.5
2007	16	9.3	15	9.7
2008	36	20.9	25	16.1
2009	29	16.9	12	7.7
2010	19	11.0	12	7.7
2011	17	9.9	23	14.8
2012	25	14.5	33	21.3
Total	172	100.0	155	100.0

Table 6

Distribution of Boating by Accidents and Time

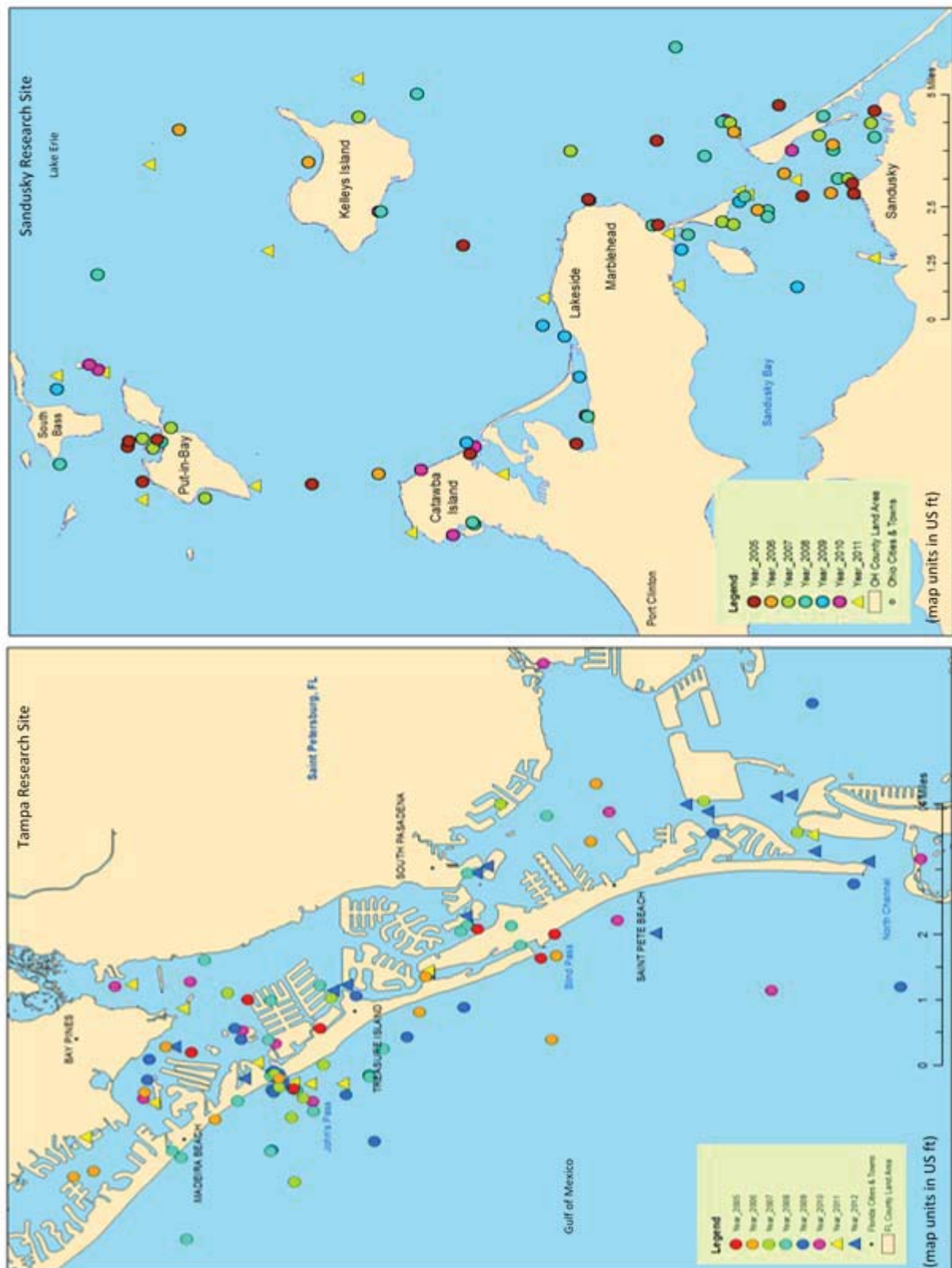| Year	Tampa	by group	Sandusky	by group
2005	10		25	
2006	20		10	
		30		35
2007	16		15	
2008	36		25	
		52		40
2009	29		12	
2010	19		12	
		48		24
2011	17		23	
2011	25		33	
		42		56
Totals		172		155

Note. From the 2005-2013 USCG BARD data.

Figure 13. Spatial distribution of boating accidents by year (2005-2012).

The frequency of recreational boating accidents varies with time at both sites and illustrates a max-min range of 26 at the Tampa site and 23 at the Sandusky site. In terms of frequency, the Tampa site experienced an approximate 70% increase in boating accident frequency during the 2005-2006 period as compared with the 2007-2008 and 2009-2010 periods. Tampa's boating accident frequency declined slightly during the 2011-2012 period. By contrast, recreational boating accidents at the Sandusky site remained relatively constant during the 2005-2008 period, then decreased by approximately 60% during the 2009-2010 period. This was followed by a dramatic increase of more than 100% in 2011-2012.

While the accident distributions between the research sites is similar in terms of the mean number of accidents (21.5 for Tampa vs. 19.4 for Sandusky) and the annual variation (measured by standard deviation) about the mean (8.2 in Tampa and 8.3 in Sandusky), the arrangement of accident values across time is quite dissimilar. For example, Tampa has peaks in 2008 and 2009 while Sandusky has peaks in 2005, 2008, and 2012 (see Table 7 and Figure 14). In fact, the correlation between the two series is only 0.10 (i.e., not significantly different from zero). This strongly suggests that the factors driving annual boat accidents vary in a substantial way (e.g., regional economy, business cycles, weather, reporting requirements, and the like) between regions. However, given the focus and purpose of this dissertation, such annual features will be reserved for future study. Such a regional analysis (albeit with the inclusion of many additional regions, i.e., national level, and many years) could form the basis of a worthy follow-up research project. In addition, it could conceivably support more informed

94

Table 7

Boating Accident Distribution by Month

| Month		Tampa Research Site				Sandusky Research Site		
	N	Cum.N	%	Cum.%	N	Cum.N	%	Cum.%
January	8	8	4.7	4.7	0			
February	3	11	1.7	6.4	0			
March	16	27	9.3	15.7	0			
April	26	53	15.1	30.8	6	6	3.9	3.9
May	20	73	11.6	42.4	14	20	9.0	12.9
June	19	92	11.0	53.5	22	42	14.2	27.1
July	29	121	16.9	70.3	46	88	29.7	56.8
August	21	142	12.2	82.6	42	130	27.1	83.9
September	15	157	8.7	91.3	14	144	9.0	92.9
October	4	161	2.3	93.6	10	154	6.5	99.4
November	6	167	3.5	97.1	1	155	0.6	100.0
December	5	172	2.9	100.0	0			
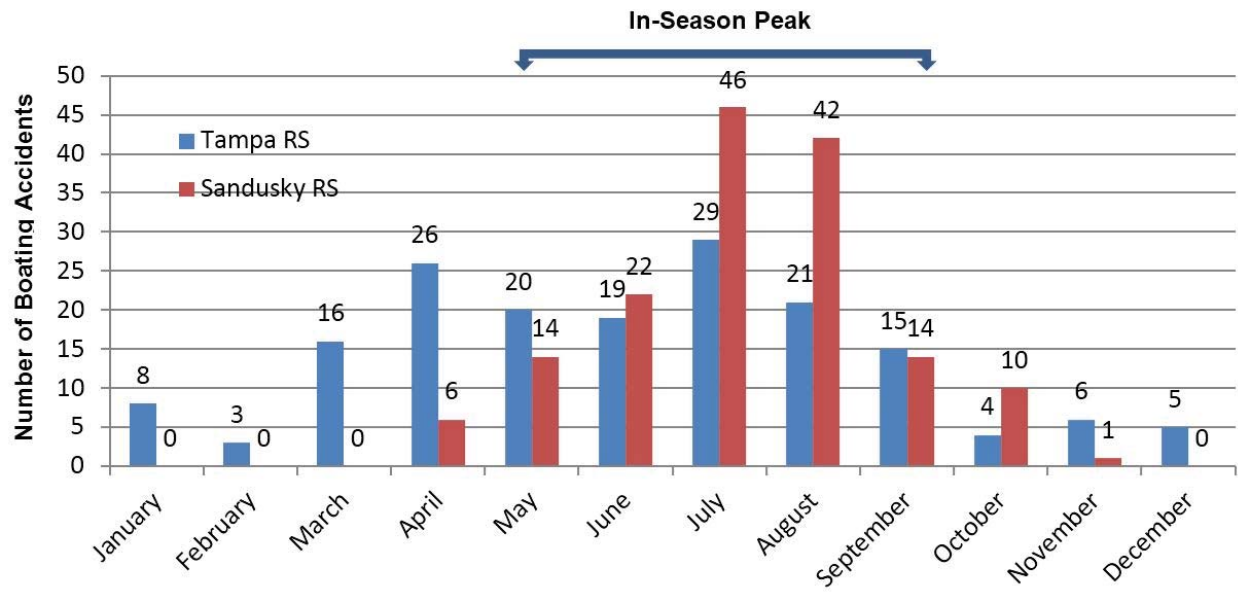
Figure 14. Boating accident distribution by month.

decisions about boating law enforcement asset deployment at the national as opposed to regional level.

In contrast to their similarity in total accident frequency, *monthly* boating accident variation suggests a very strong distinction between the research sites (see Table 7 and Figures 14-16). As expected, given weather and climate differences, the Tampa research site is characterized by a more uniform twelve month boating season (activity distribution) while the Sandusky research site is characterized by a more peaked seven to eight month boating season. From a general perspective, these monthly frequencies suggest that both research sites experience a greater number of recreational boating accidents during the in-season period (May to September) time frame than the off-season period.

This pattern of monthly boating accidents is referred to simply as recreational boating "seasonality". Table 7 and Figures 14-16 illustrate the key seasonal (monthly) differences between the two research sites. The Sandusky research site exhibitis a much stronger seasonality than the Tampa research site. Nearly 60% of the Sandusky boating accidents occur within a two month window, i.e., July and August. By contrast, most boating accidents occurring at the Tampa research site occur between March and September timeframe and represent a much broader seasonal time scale. While the mean monthly accident rate for the two sites is similar (14.3 in Tampa and 12.9 in Sandusky) the level of variation across months, as measured by the standard deviation, is decidedly different (16.2 in Sandusky as contrasted with 8.9 in Tampa). These statistics capture the increased level of recreational boating intensity illustrating the seasonal differences (in terms of monthly variation) as witnessed at the Sandusky
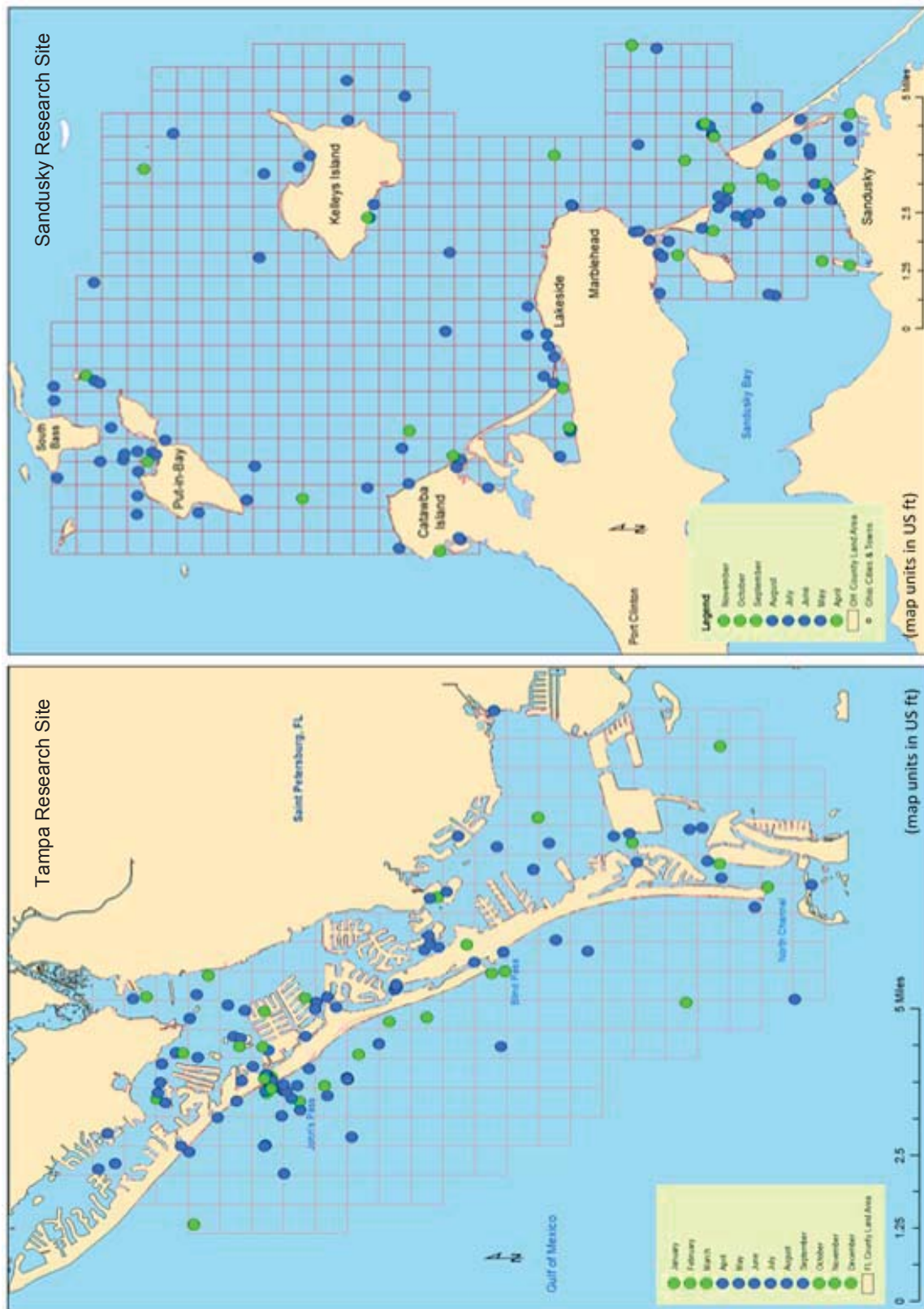
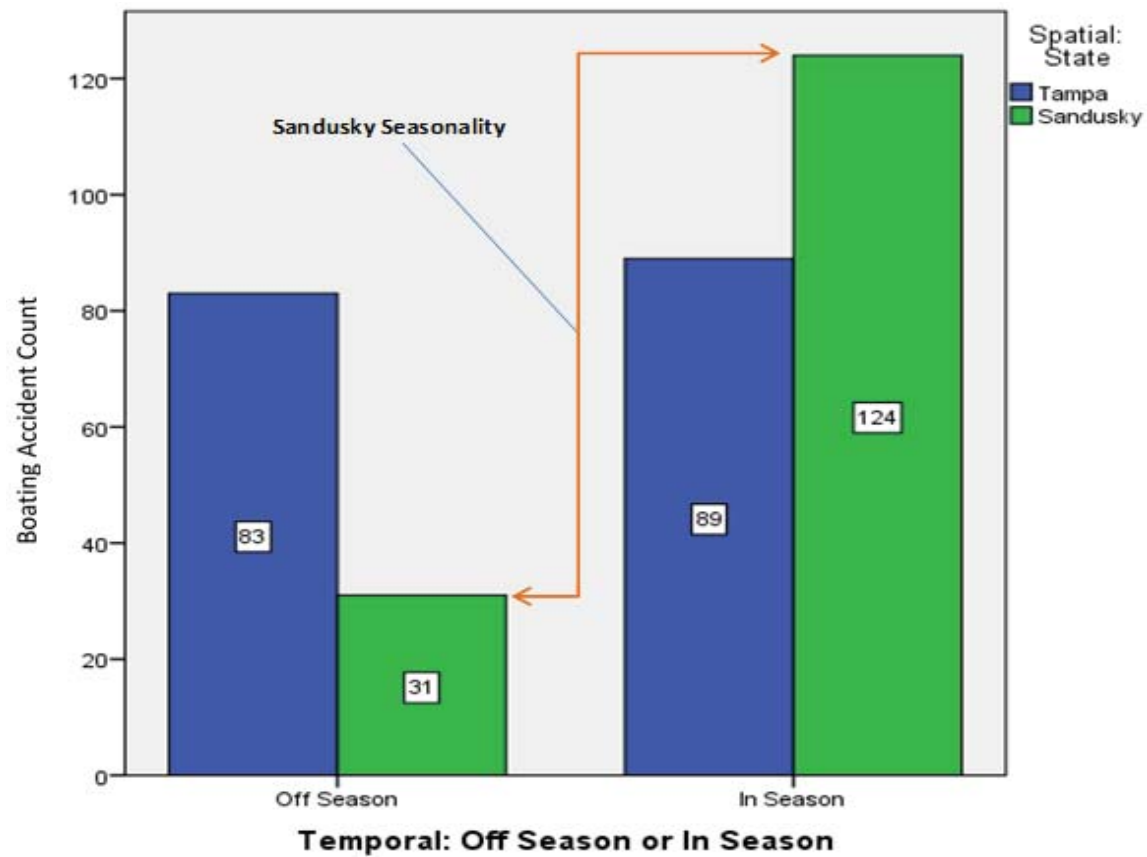*Figure 15*. Spatial distribution of boating accidents by month.

Figure 16. Research site temporal seasonality.

site. This seasonal characteristic is clearly illustrated from a spatial perspective in

Figures 15 and Figure 17.

Figure 18 illustrates daily boating accident variation at both research sites

highlighting observable regional boating accident differences. From a general

perspective, the regional (site specific) boating accident frequency strongly suggests an

increased accident rate during weekend periods, i.e., those days of the week defined as

Saturday and Sunday, as opposed to the traditional weekday period, i.e., Monday

through Friday. The only deviation from this generalized "weekend" accident pattern

pertains to the Sandusky site where boating accident frequency is substantially higher

on Friday's than other weekdays. This empirical observation suggests an extended

weekend recreational boating accident pattern that includes Friday, in the case of

Sandusky.

Similar to seasonality and using this extended weekend definition, the Sandusky

research site is characterized by a more concentrated weekend (Friday-Sunday)

accident peak pattern. Seventy one percent of Sandusky's boating accidents occur

during the extended weekend period. In the case of the Tampa research site, the

extended weekends capture approximately 58% of the total number of recreational

boating accidents. Another interesting characteristic of this weekend boating accident

pattern is highlighted by Tampa's peak on Sunday (as contrasted with Sandusky's peak

on Saturday). During weekdays (Monday-Thursday), the Sandusky research site

averages 11.2 boating accidents while the Tampa research site is much more active,

averaging 18.3 boating accidents as illustrated spatially in Figure 19.
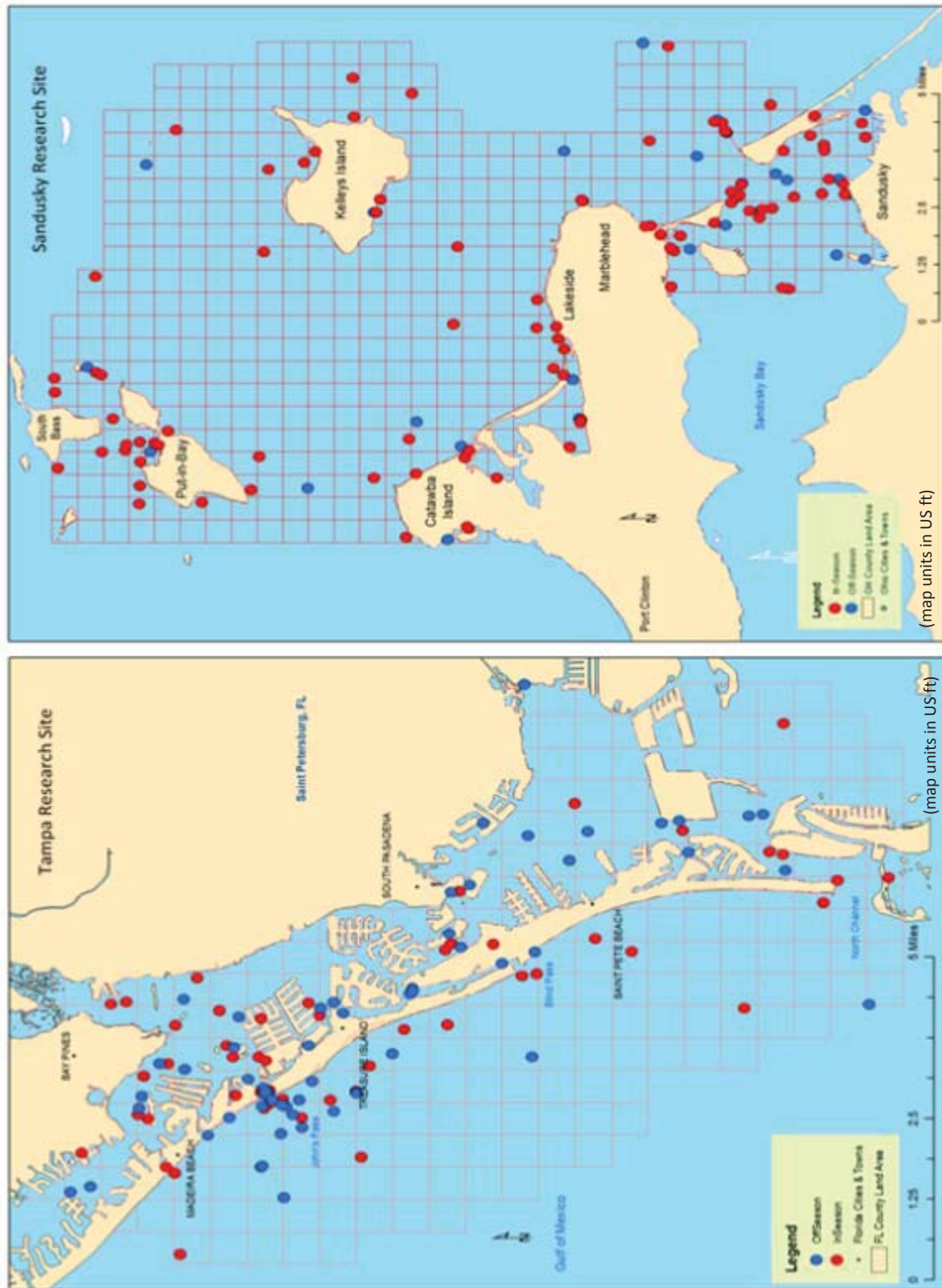
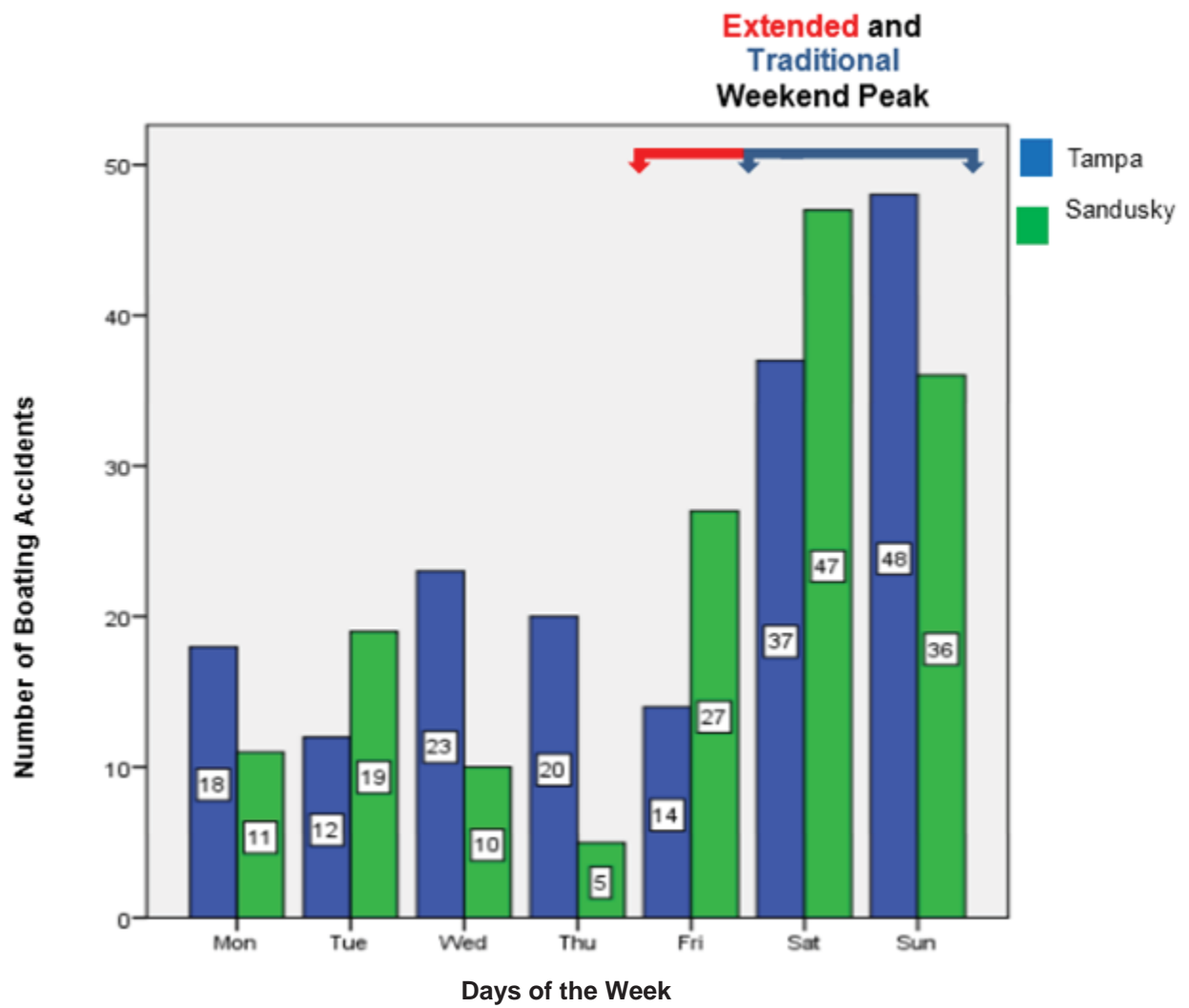*Figure 17*. Distribution of boating accidents by season.
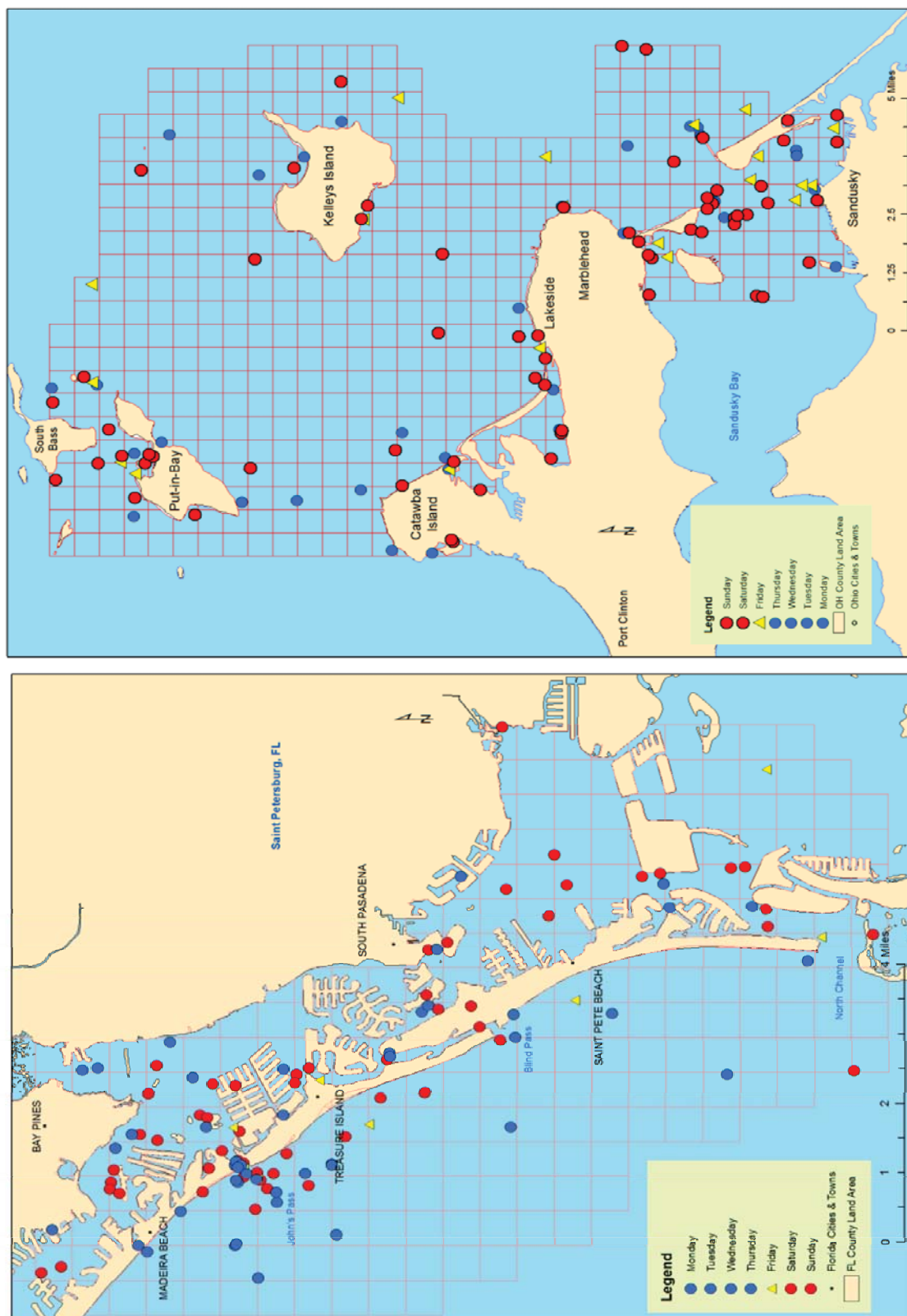
Figure 18. Boating Accidents by day of week.

Figure 19. Spatial distribution of boating accidents by day of the week.

As previously mentioned, these temporal patterns have significant implications for boating law enforcement resource (asset) allocations not to mention the deployment of those assets. Patterns and differences related to accidents' time of day are understandably more challenging and complex due to the range of hours per day (0-24).

To better represent time of day variation in accidents, eight 3-hour time segments are used to aggregate observed accident data into more convenient time periods (see Table 8). For example, 0000-0259 (representing midnight to 2:59 AM) is assigned to Time Period 1. Subsequent time periods range in 3-hour increments to 2100-2359 (9:00 PM to midnight) and representing Time Period 8.

Table 8 illustrates the hourly distribution shown spatially in Figure 20 and represents boating accident distributions occurring at the Tampa and Sandusky sites. The Tampa site experiences an hourly peak between noon and 6:00 PM (periods 4, 5, and 6) when 59% of the Tampa recreational boating accidents occur. Sandusky data suggests that this hourly peak, somewhat surprisingly, occurs during time period 1 (basically midnight) and 3:00 AM with a secondary temporal peak occurring during Time Periods 3, 4, and 6 (basically between 6:00 AM and 6:00 PM). Sandusky is particularly distinctive in its level of night time accidents (periods 8 and 1, basically 9 PM to 3 AM) with nearly 30% of accidents occurring at night. That same time value represents only 14% of the recreational boating accidents in the case of Tampa.

Despite Sandusky's surprising nocturnal boating character, these hourly data suggest that most boating accidents do occur during daylight hours, i.e. between 6:00 AM though 6:00 PM as illustrated in Table 8. Figure 20 further illustrates this accident pattern by illustrating that 65% of boating accidents occur during daylight hours at the

Table 8

Distribution of Boating Accidents by Time of Day

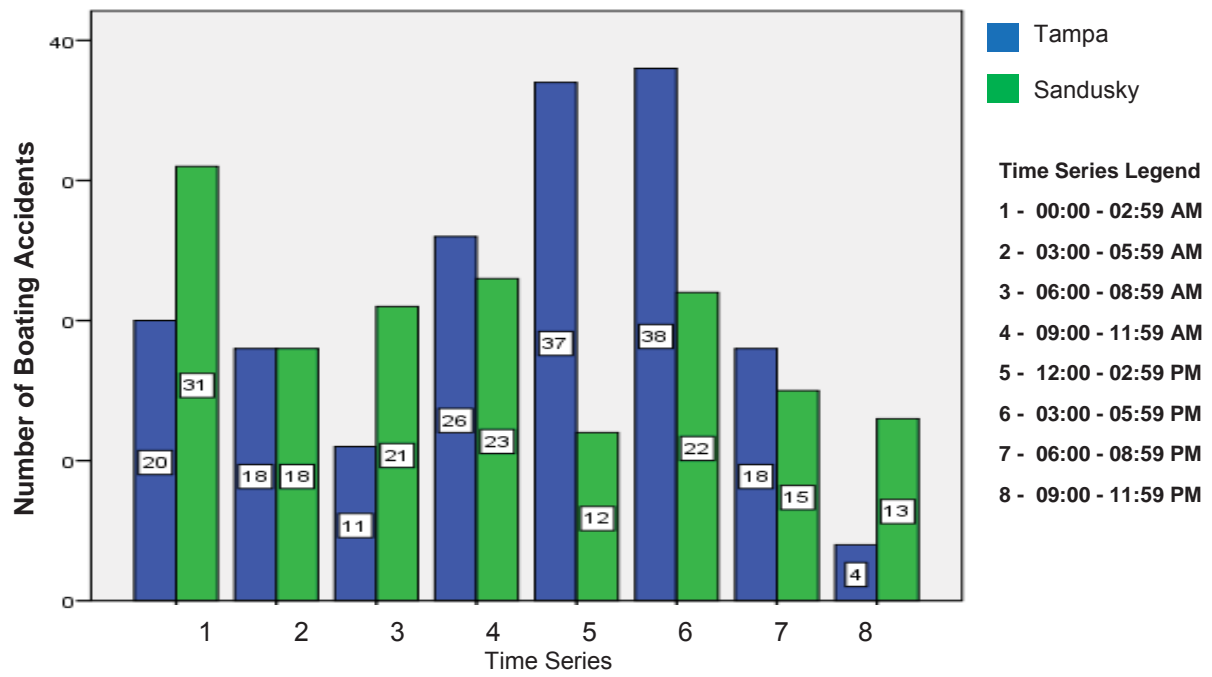| Time Period	Tampa RS		Sandusky RS	
1	20	11.63%	31	20.00%
2	18	10.47%	18	11.61%
3	11	6.40%	21	13.55%
4	26	15.12%	23	14.84%
5	37	21.51%	12	7.74%
6	38	22.09%	22	14.19%
7	18	10.47%	15	9.68%
8	4	2.33%	13	8.39%
	172	100.00%	155	100.00%

Figure 20. Boating accident distribution by hour.

Tampa site as contrasted with 50% of boating accidents at the Sandusky site. This evidence illustrates that boating accident clusters can be categorized as morning or evening and this is further evidenced in Figure 21. Figure 21 also shows observable accident clustering near shore, harbor, or navigation inlet areas.

Spatial Disaggregation

Figure 13 illustrates the spatial distribution of all boating accidents occurring within the Tampa and Sandusky sites between 2005 and 2012. The intent of that illustration is to offer a high-level perspective as to where boating accidents are occurring. While clustering is evident in this illustration, the strategy in this section focuses on disaggregation of the data to reveal potential spatial patterns in greater detail.

The first analytical data disaggregation involves point pattern analyses and tabular data using polygon-based quadrat (grid) analysis. This technique permits key clusters to be identified and levels of clustering to be compared and contrasted across research sites. Density patterns, frequency distribution and grid square counts serve as the measures for assessing boating accident distribution. The second analytical approach considers the potential of spatial autocorrelation, i.e., Tobler's First Law of Geography, "all things are related, but near things are more related than distant things". Spatial clustering and autocorrelation are analyzed and described using spatial statistics such as Getis-Ord (Gi*) and Moran's I.

Examination of clustering is appropriate when investigating spatial distributions of values while looking for unexpected spatial spikes of high values. However, when both
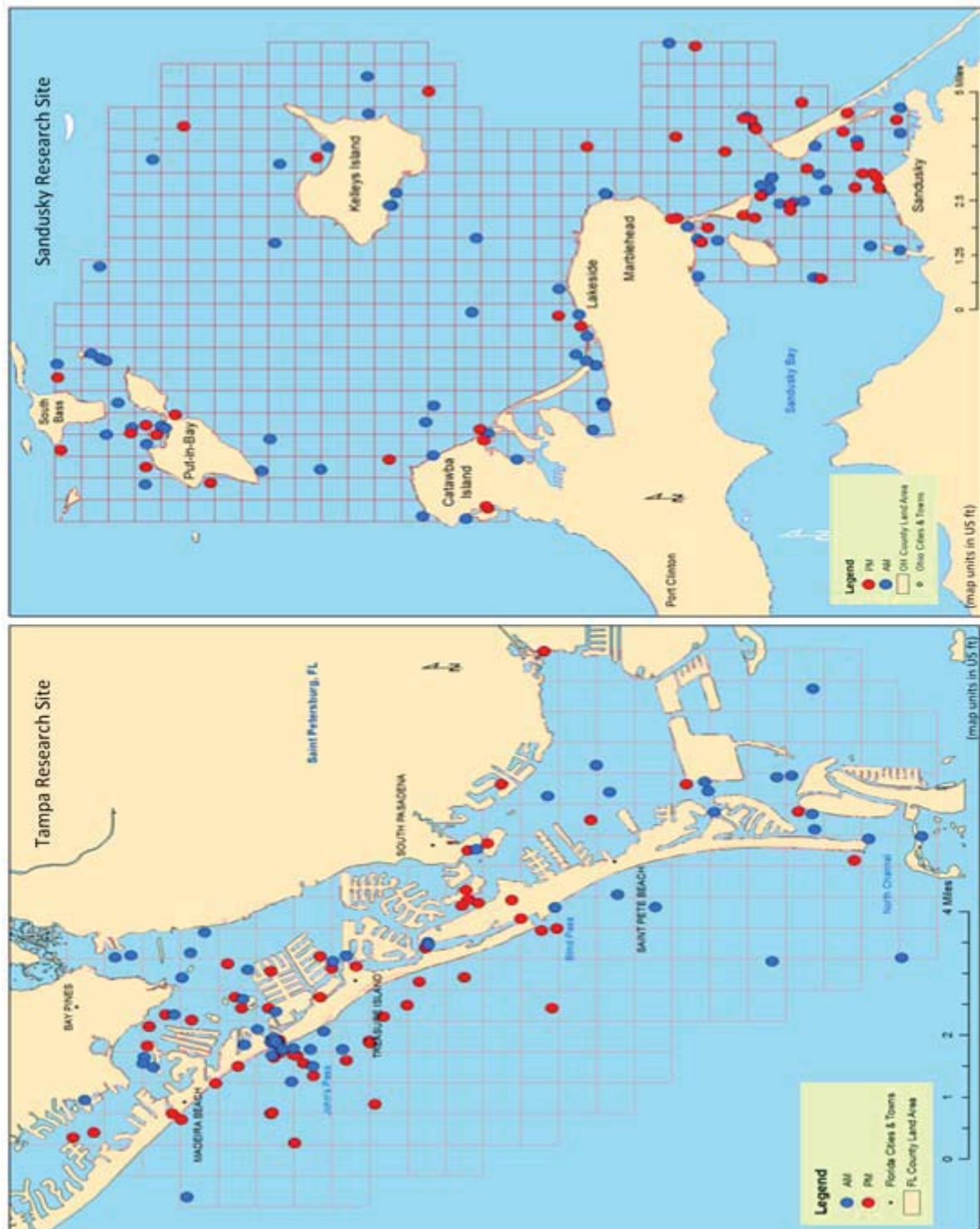
Figue 21. Boating accident spatial distribution by morning vs evening.

high and low valued clusters are present, they tend to cancel one another. For this reason, it is important to measure localized spatial clustering (when both the high valued (hot spots) and the low valued (cold spots) clusters are present), using a local measure of spatial autocorrelation.

Spatial Disaggregation of Boating Accidents by Key Temporal Factors

In general, the boating accident data suggests higher accident frequency in shallow water. However, the most significant clusters in shallow water areas are more apparent in areas of constrained navigation. Areas of navigational constraint are exemplified by John's Pass, within the Tampa research site (see Figure 22). This area illustrates the constrained nature of navigable inlets positioned between open water and the inner-harbor zone. The yellow lines within the inset (see Figure 22) denote the approximate boundaries of safe-passage and the resulting platoon effect that this environmental constraint imposes on boating traffic. The high concentration of boating accidents is readily evident within this zone and this empirical outcome provides further evidence that navigation constraints and shallow water depths are key factors that influence boating accidents. John's Pass is also distinctive as containing a grid cell with the largest number of boating accidents (by count) at either research site.

Despite this generality, other differences can be observed between the research sites. In Sandusky, accident frequencies suggest a relatively uniform accident distribution regardless of season. An examination of the Sandusky data reveals the presence of three significant accident clusters as highlighted by circles and arrows in Figure 23. Data from these areas illustrate a pronounced level of clustering with 46.5%
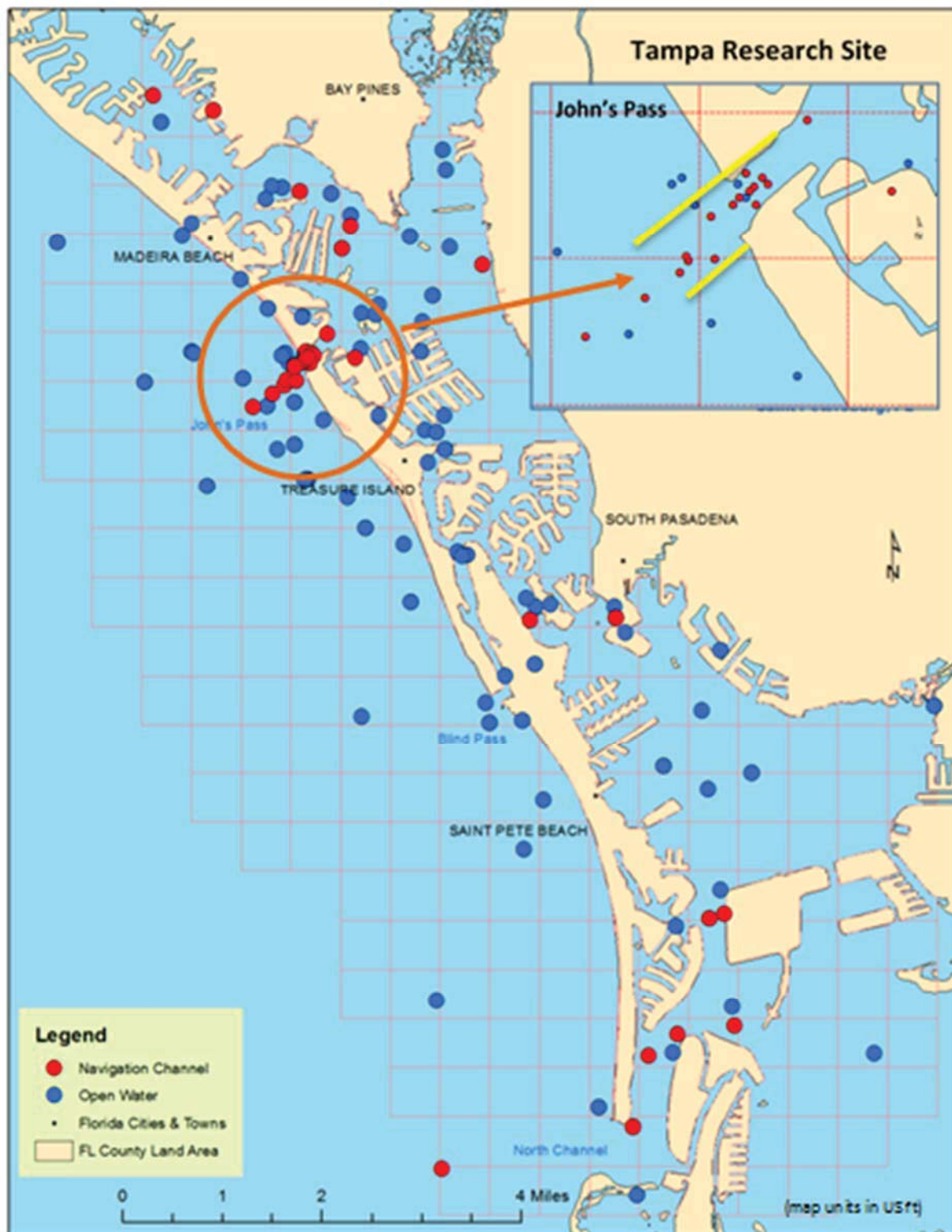
Figure 22. Intensity of boating accidents at John's Pass (FL).

Figure 23. Distribution of boating accidents by in-season vs. off-season characteristics.

of the accidents occurring in-season as contrasted with only 10.3% of the accidents occurring in the same locations during the off-season. Similar to seasonality, accident clusters (indicated by circle and arrow) suggest that certain areas are subject to more intense clustering during the weekend period (see Figure 24). On the other hand, seasonality exhibits less influence on the frequency of accidents at John's Pass with 15.6% of the boating accidents occurring during the off-season period and 12.8% of the boating accidents occurring in-season. At the Tampa research site, 11.6% of the all weekday accidents occur within the identified cluster (excluding Friday) and 16.9% of the all extended weekend accidents.

Similar to seasonality, the weekend accident pattern at both research sites suggests clustering near navigation inlets and more specifically within boating waterways constrained by shallow water depth. As highlighted by the circle and arrow in Figure 25, the patterns suggest that the number of recreational boat accidents are both higher and more concentrated in constrained waterways but more specifically in those areas characterized by low water depths. For example, the area described as "John's Pass" mentioned above illustrates the critical nature of navigational constraints, water depth, and to a degree the temporal factors that accompany both. These concentrations reminds us that there are direct effects such as (shallow water groundings and collisions with rigid or submerged structures) and indirect effects (increased density of boats yielding increases in multi-vessel accidents).

Location of Boating Accidents

As indicated by earlier examples, water depth has frequently been observed as a correlate with boating accident patterns. As observed in *Location of Boating Accidents,*
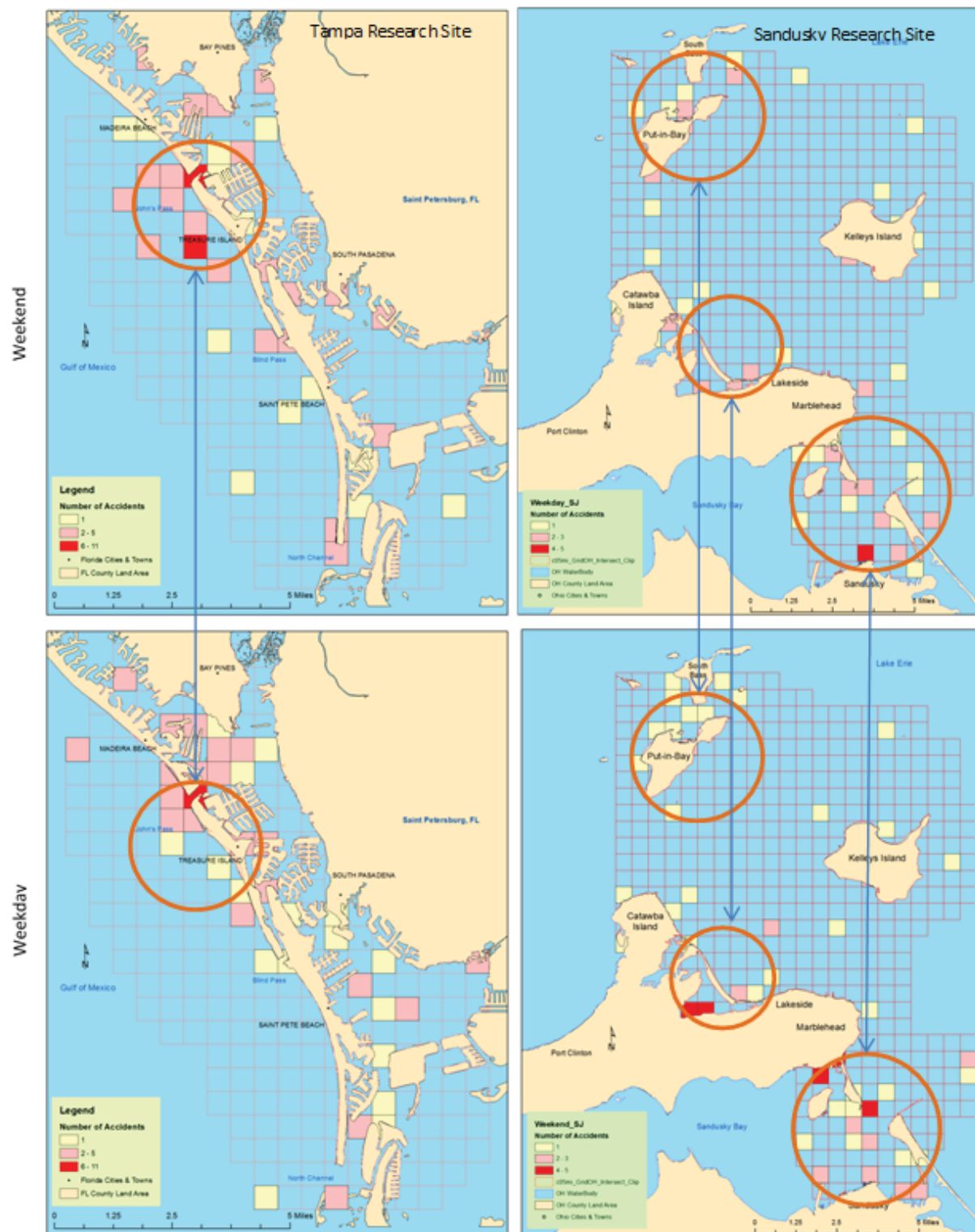
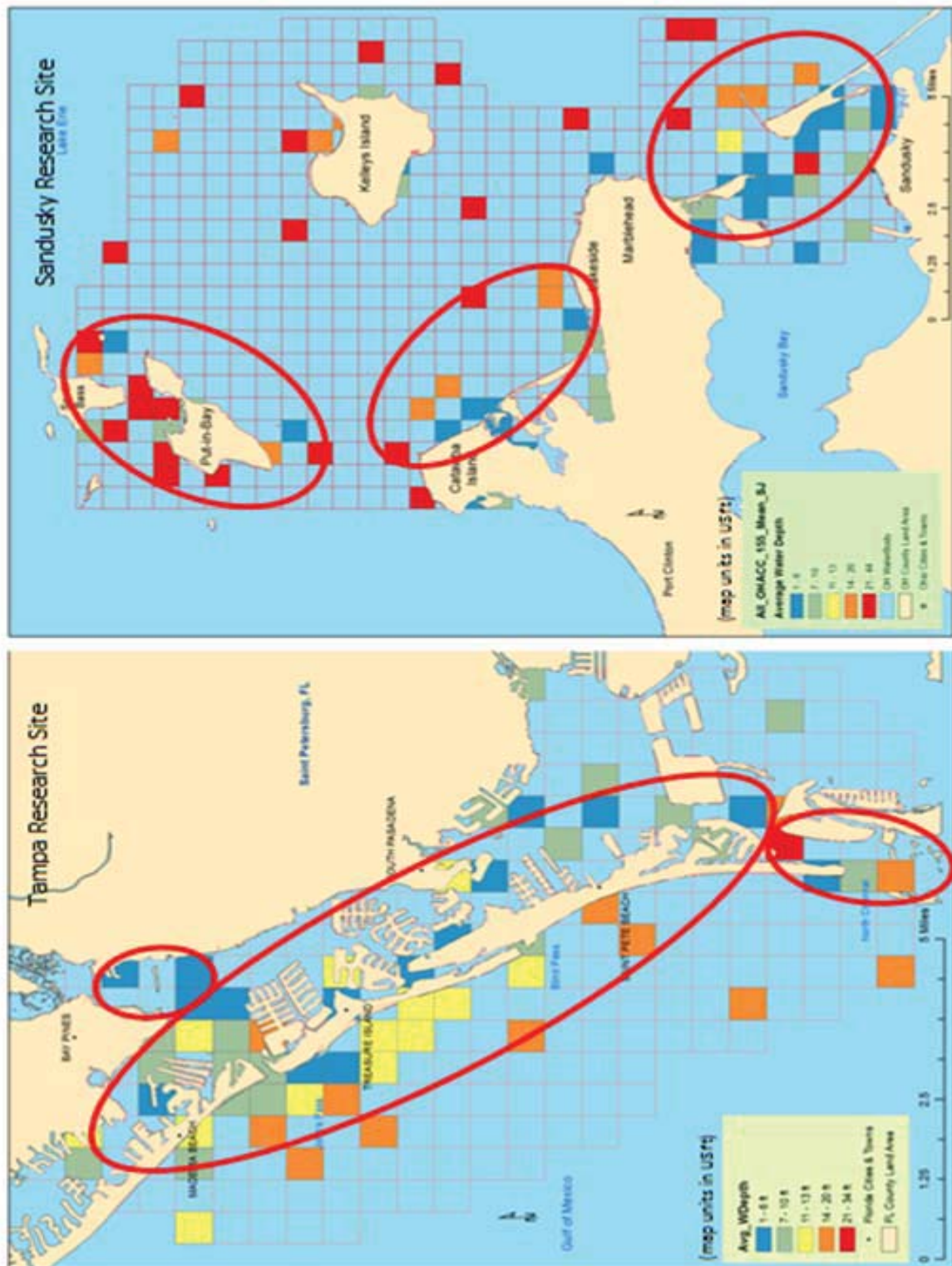*Figure 24*. Distribution of boating accidents by weekday vs. weekend characteristics.

Figure 25. Distribution of boating accidents by water depth characteristics.

accidents tend to occur in areas that are in close proximity to constraining environments, especially where geographic constraints to navigation (e.g., navigation channels, bridges, and waterway intersections) and natural features (e.g., channel width and water depth) exert strong influences. An examination of polygon accident densities in comparison to shallow water areas (illustrated by circles in Figure 26) further reinforce the strong relationship between boating accidents and water depth. These areas (highlighted in red) suggest that a majority of boating accidents at both research sites occur in areas where water depth is less than 15 feet. Figure 26 further illustrates the spatial distribution of boating accidents by water depth. This figure represents a boating accident distribution using a choropleth map overlay atop shallow water ranges. The red, orange, and yellow coded grids characterize water depths averaging 15 feet or less. This shallow water zone encompasses the majority of the boating accident clusters (defined by circle and arrow). This set of elliptic illustrations represent 91.9% of the recreational boating accidents within the Tampa site and 71.6% of the boating accidents within the Sandusky site.

There are other environmental factors such as wind speed (see Figure 27) and wave height (see Figure 27) that one might expect to have a strong influence on boating accidents. However, these factors seem unimportant within this investigation. Frequently, boaters self-select to be off the water during times of bad weather, e.g., after small-craft advisories have been issued. This self-selection suggests that the statistical power of these environmental features would be limited in explaining such accidents. In fact, the patterns observed in Figure 27 can otherwise be explained by navigational constraints and/or water depth as detailed earlier in this chapter.
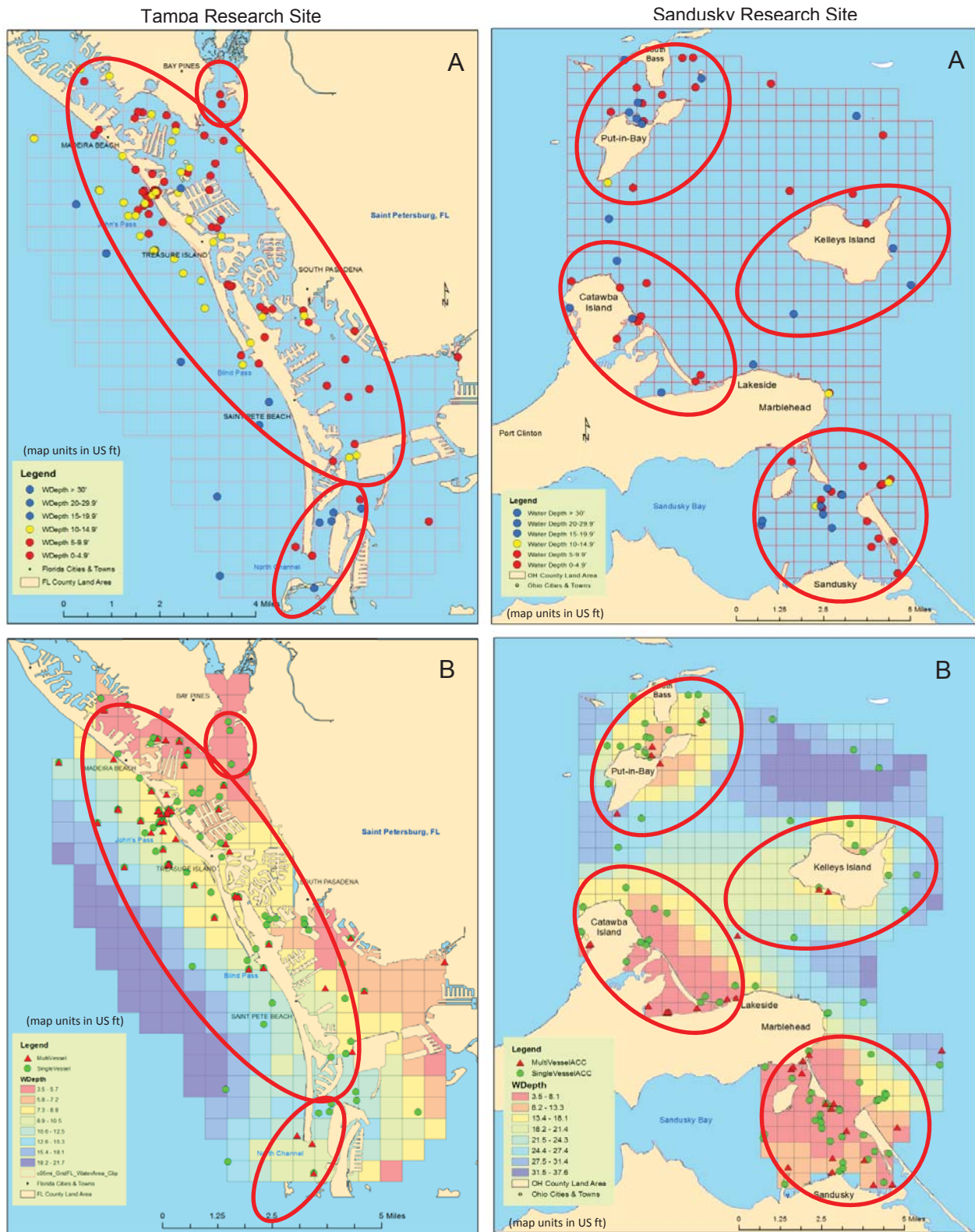
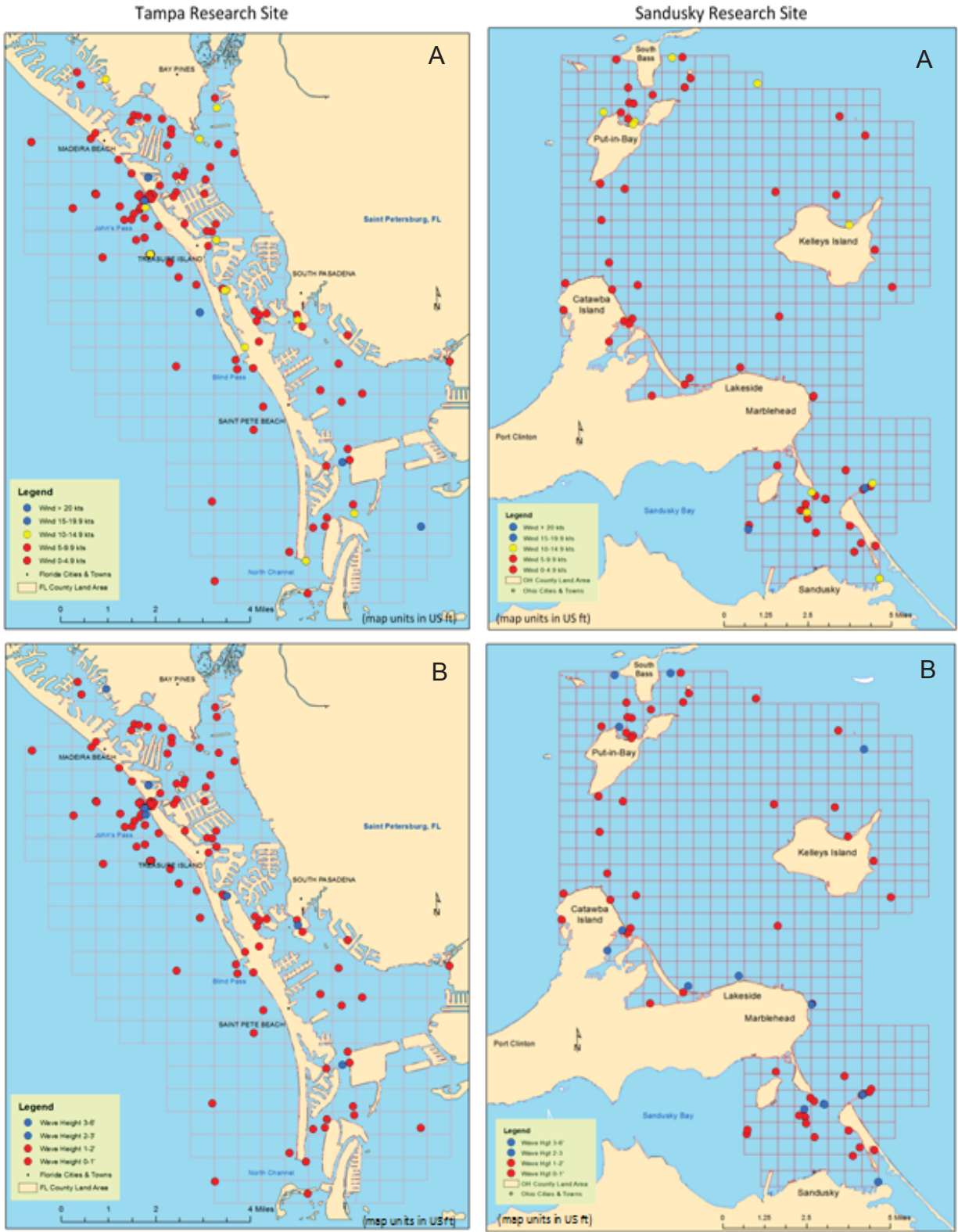*Figure 26*. Spatial distribution of boating accidents by water depth.

Figure 27. Spatial distribution of boating accidents by wind speed (A) wave heights (B).

Another environmental factor that merits special attention is evidenced by the number of boats operating within any specified area, i.e., boat density. Higher boat densities increase the probability of accident risk. Figure 28 illustrates the spatial distribution of boating accidents by grid cell. This figure indicates the areas where higher than average boating accident counts (circled in red) exist when compared with neighboring cells. They correlate with boat density in general.

Spatial Statistics and Boating Accidents

The previous quadrat (grid) analysis of boating accident density is extended by a Hot Spot Analysis (Getis-Ord Gi* & Moran's I statistics) as illustrated in Figures 29-30. The resultant z-scores and p-values (Figure 29, section I and Figure 30, section II) describe whether the boating accidents at each site have a large or small spatial clustering value. While grid cell/accident counts are interesting, the more important observation is that high count grid cells are generally surrounded by other high value cells, i.e., positive spatial autocorrelation. This feature is usually an indication of hot spots. The local sum of a feature and its neighbors is compared proportionally to the sum of all its features. If the grid cell sum is significantly different from the expected sum and that difference is too large to be the result of random chance, a statistically significant z-score results. Most statistical tests begin with a null hypothesis assuming no study area spatial patterns among features or values associated with those features, i.e., the pattern is one of the many possible versions of spatial randomness. The larger positive z-scores illustrated (see Figure 29, section I) indicate boat accident hot spot intensity. The most intense of these hotspot areas are identified by circle and arrows referencing comparable boat accident density (accident counts) in the related figures.
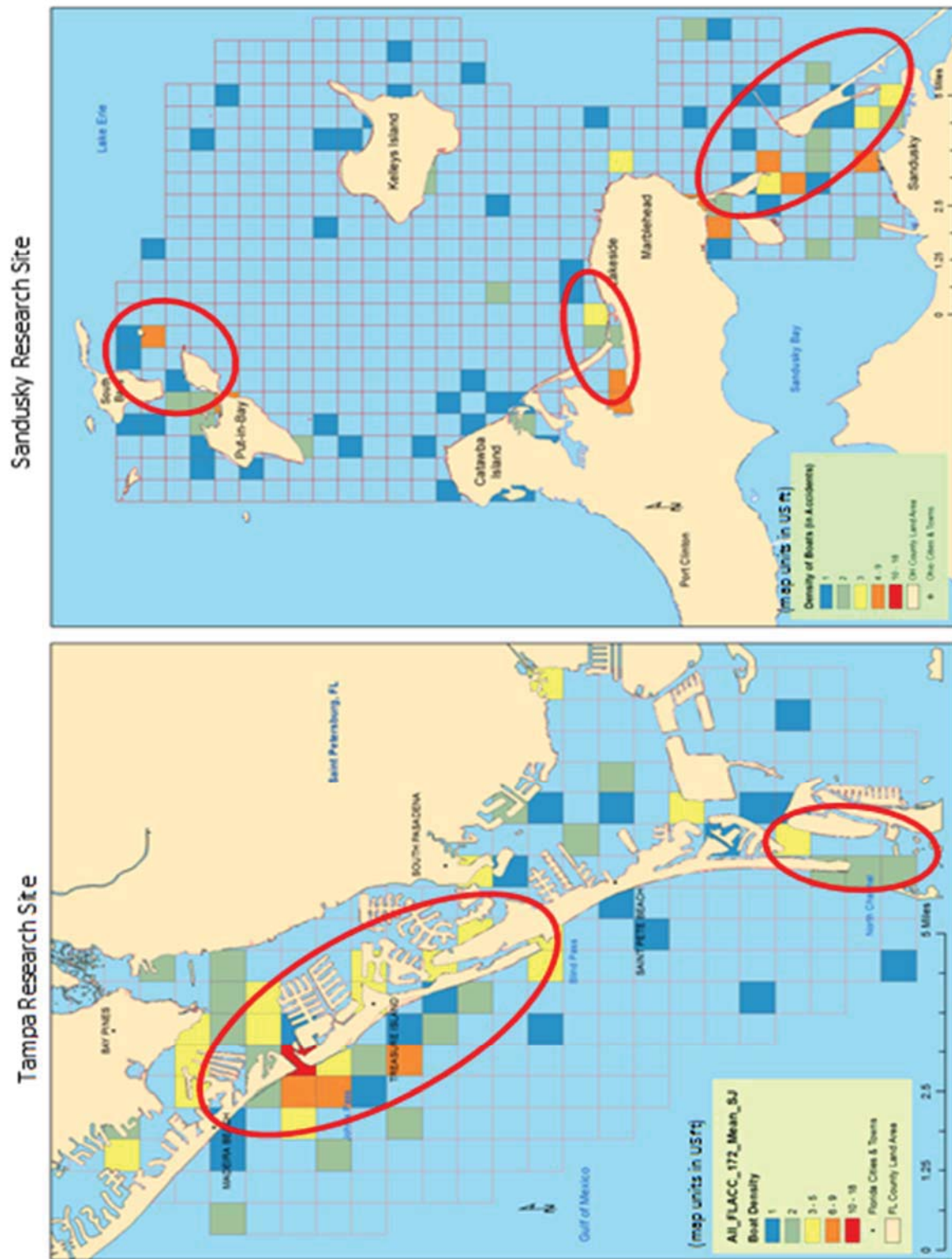
118

Figure 28. Spatial distribution of boating accidents by density (satellite).
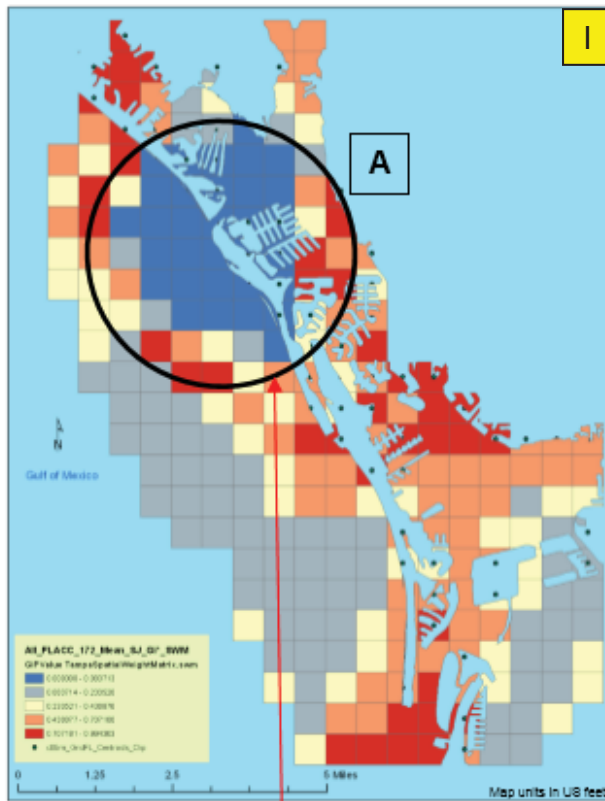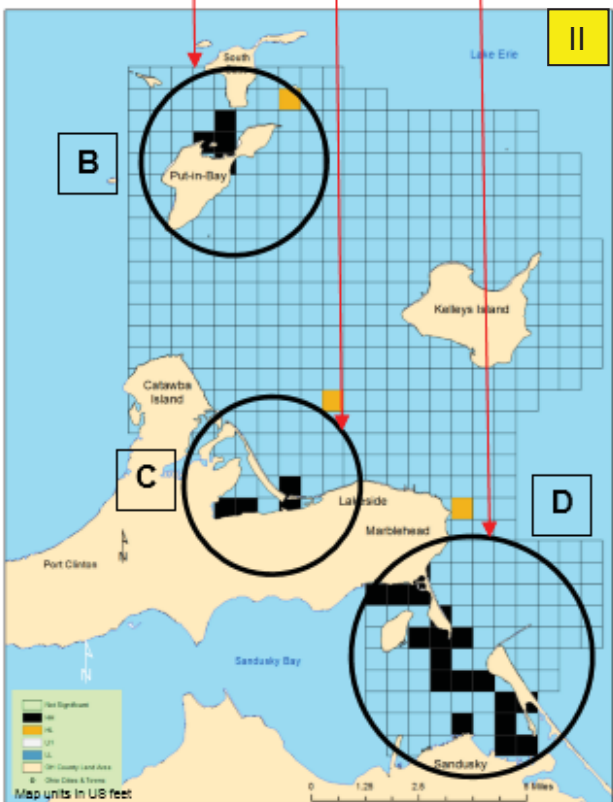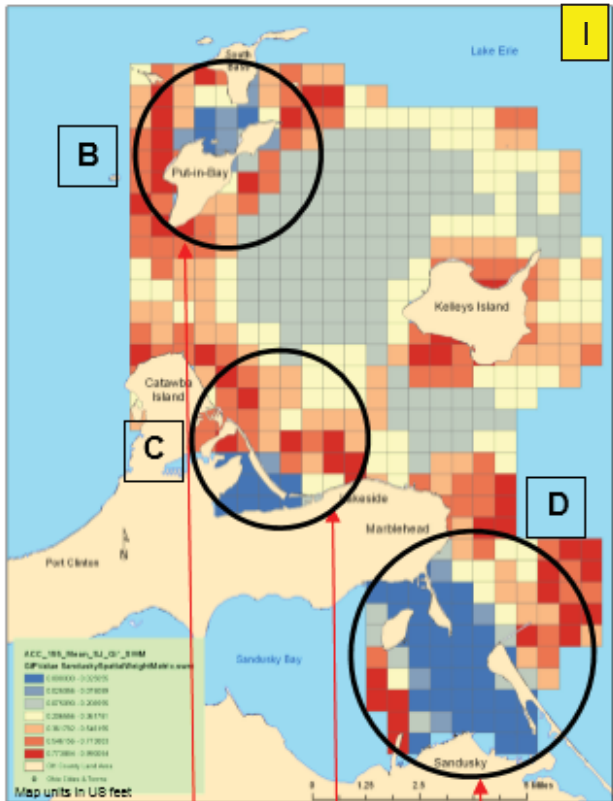
Figure 29. Hot Spot (Getis-Ord Gi) & Moran's I Analysis.*
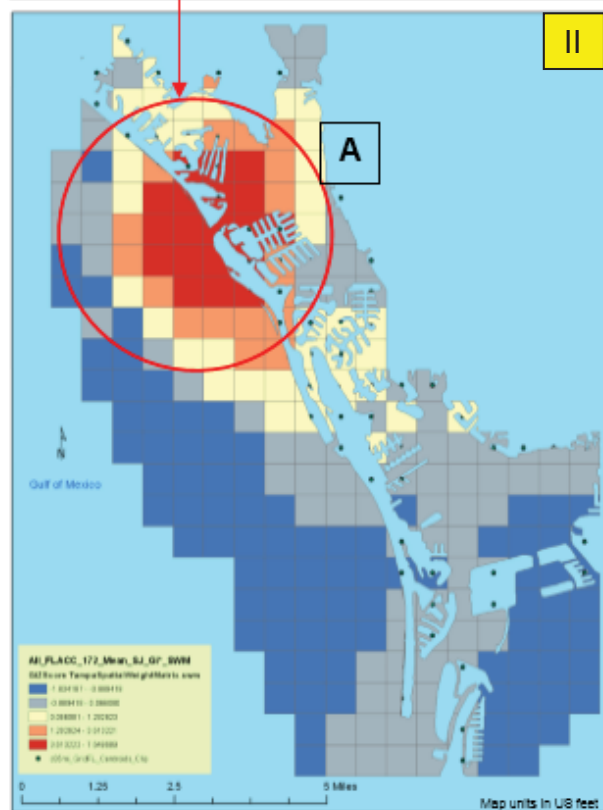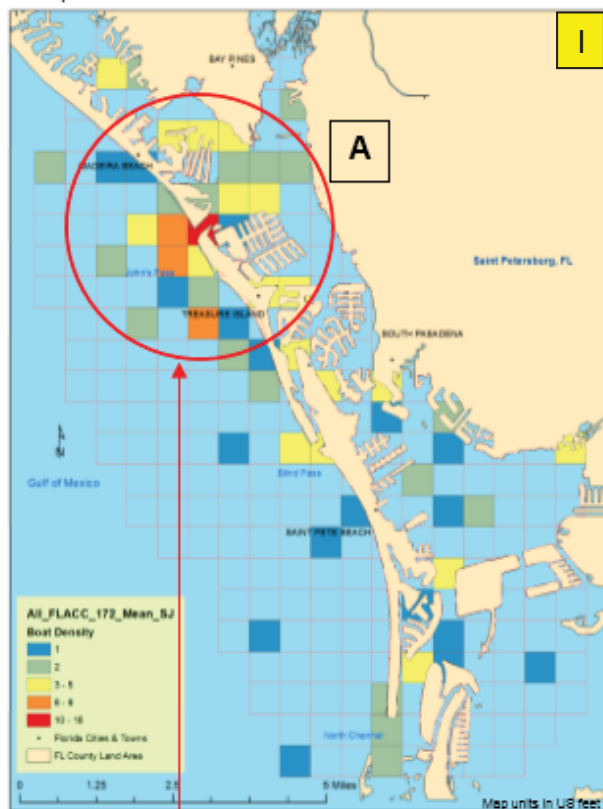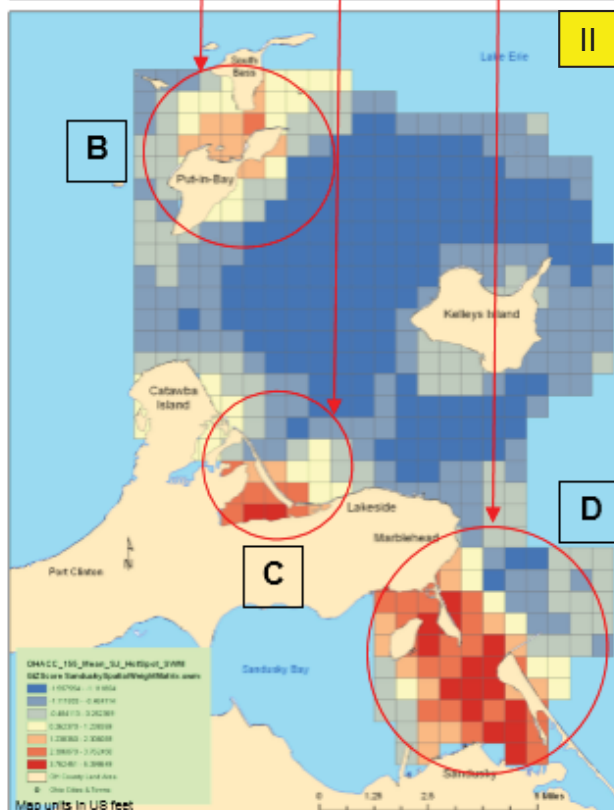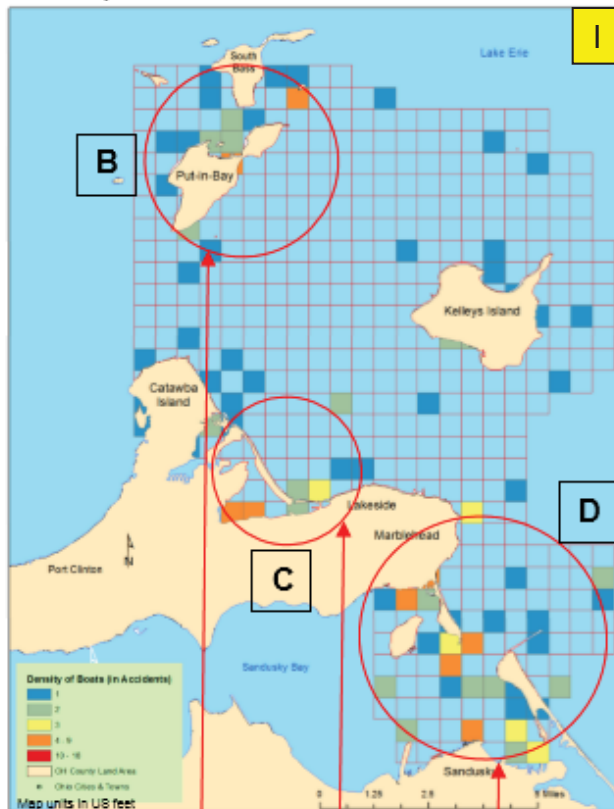
Figure 30. Boat Density and Moran's I Summary.

As z-scores are measures of standard deviation, the Tampa research site hot spots range from 1.2 – 7.6 standard deviations away from the mean while the Sandusky research site ranges from 1.2 – 5.4 standard deviations away from the mean. By contrast, the negative z-score values illustrated in both research sites (from gray to blue in color) are an indication of areas representing relative accident voids (cold spots). Figure 29 also extends this Getis-Ord Gi* approach with a Moran's I hotspot analysis.

Moran's I p-values serve as a statistical significance index of cluster values (as compared to a random spatial distribution). This distribution allows significance and confidence levels to be attached to each Z score. Figure 30, section II illustrates the range of p-values by grid cell. The blue grid cell areas (identified by red circles) indicate those grid cells with p-values (less than 0.05) suggesting statistical significance. It should be noted that these areas of statistical significance closely match the hot spots identified in Figure 29, section II (Moran's I) as well as the boat accident density grid analysis shown in Figure 30, section I. To complete this analysis, a spatial weights matrix based on Queen's Contiguity and a fixed Euclidean distance threshold of 2011.68 meters (1.25 mi) were adopted (see Table 9). Table 9 serves to summarize the hot spot analysis when applied to the boating accident spatial data. Both illustrations reinforce identified hot spot statistical significance (see Figure 29) through z-scores of 8.31 for Tampa and 7.23 for Sandusky with significant p-values of 0.00.

Of the two sites, the Tampa research site highlights the largest cluster (see Figure 31 A) in a single area (John's Pass) representing 80 of the 172 boating accidents within the research site (46.5%) including the most intense cluster in either research site. By contrast, the Sandusky research site is represented by one primary

122

Table 9

Spatial Weights Matrix Summary and Global Moran's I Summary

Location	Number of Features	Percentage of Spatial Connectiivty	Average Number of Neighbors	Minimum Number of Neighbors	Maximum Number of Neighbors	Moran's Index	Expected Index	Variance	Z-score	p-value
Tampa	254	6.58	16.72	5	20	0.1584	-0.0040	0.0004	8.3148	0.0000
Sandusky	463	3.64	16.85	2	21	0.11235	-0.0022	0.0003	7.2341	0.0000
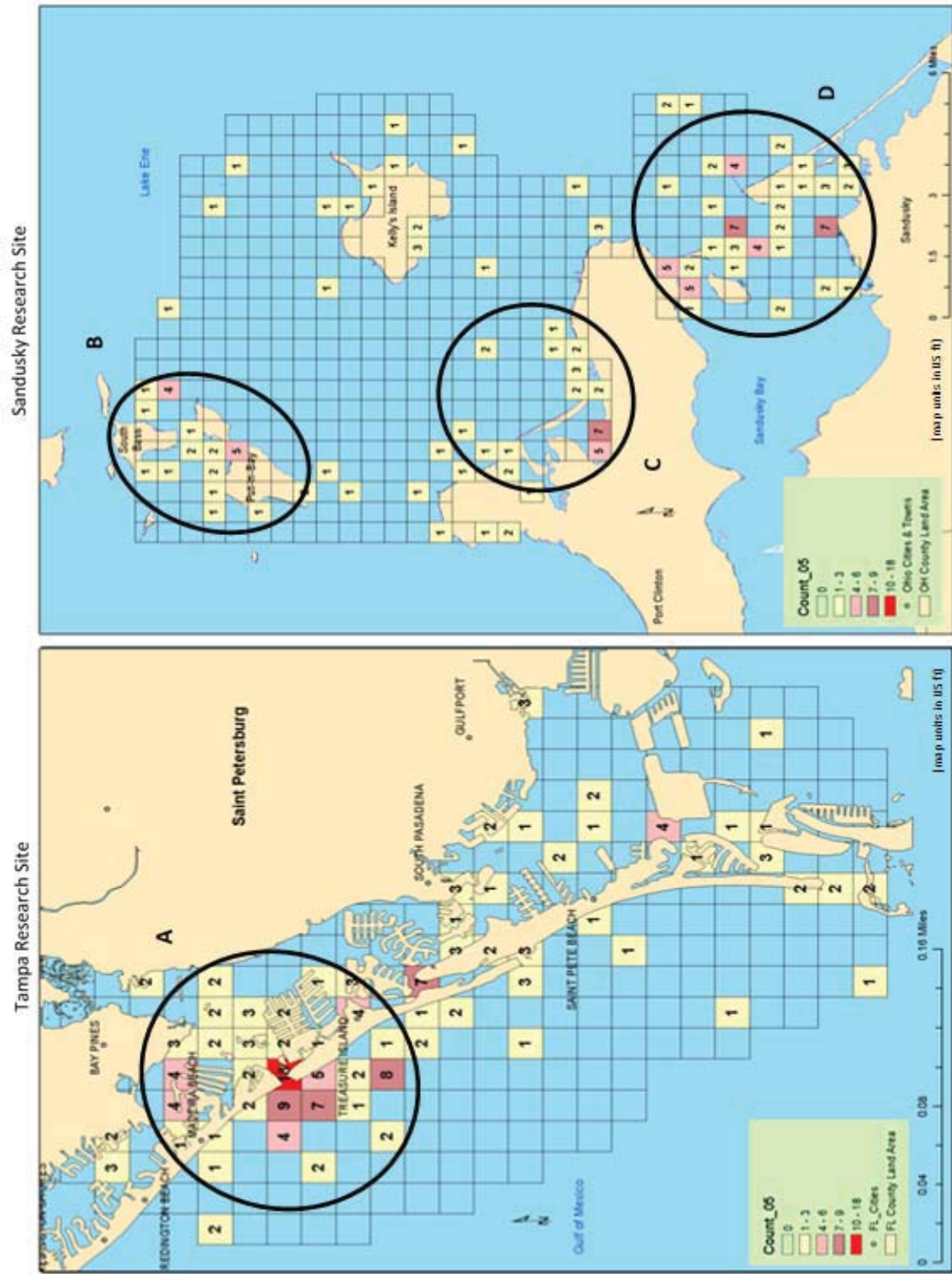
Figure 31. Cluster analysis by boat accident counts.

(see Figure 31 B) and two secondary clusters (see Figures 31 C and D). The primary

cluster is defined by the Sandusky Bay harbor entrance and represents 67 of the 155

boating accidents (43.2%) within the research site. The two secondary areas are

defined by the bay nearest to East Harbor State Park and Catawba Island representing

30 of the 155 boating accidents (19.4%) and the Kelly's Island area representing 25 of

the 155 boating accidents (16.1%). It is observable that the two main harbor entrances

(John's Pass at the Tampa Research site and the Sandusky Bay harbor entrance at the

Sandusky research site) are closely comparable representing 46.5% and 43.2% of the

boating accidents respectively.

Also observable is the cumulative magnitude of these hotspot areas. The Tampa

hot spot accounts for 46.5% (boat accident count, 80) of the accidents within that

research site while the Sandusky's three hot spots cumulatively account for 78.7% of all

accidents within that research site (see Figure 31).

Although Florida (represented by the Tampa research site) is ranked first

nationally in terms of boating accidents, the Sandusky research site is illustrative of a

stronger regional effect. This is further illustrated by (including the Tampa and

Sandusky sites) are similarly characterized by constrained waters including channelized

areas, shallow water depths, and resulting increased boat density (as measured by the

GeoEye satellite imagery collection).

A closer examination of Figure 31 as compared with Figure 29, section I (Hotspot

Analysis standardized by z-score) indicates that 46.5% of Tampa's are represented by

z-scores: a) one deviation above the mean (1.30-3.60; identified by the orange range)

and b) two deviations above the mean (3.61-7.65; identified by the red range).

At the Tampa research site, the above positive deviations are represented by 35 grid cells:

- A: 16 grid cells at 1 deviation above the mean, 19 grid cells at 2 deviations above the mean

By contrast, the positive deviations at the Sandusky research site include:

66 grid cells:

- B: 8 grid cells at 1 deviation above the mean, 2 grid cells at 2 deviations above the mean

- C: 10 grid cells at 1 deviation above the mean and 2 grid cells at 2 deviations above the mean

- D: 25 grid cells at 1 deviation above the mean and 15 grid cells at 2 deviations above the mean

These observations suggest that the boating accident rate is more intense in the John's Pass (A) area of the Tampa research site. However, what the Sandusky research site lacks in terms of a single intense cluster it possesses a broader field of clustering, i.e., 44 higher risk grid cells with an elevated boating accident rate at the Tampa research site vs. 66 higher risk grid cells with an elevated boating accident rate at the Sandusky research site (see Figure 31).

Although the above discussion suggests a close parallel between the Tampa and Sandusky research site locations (A), the spatial extents of these areas are calculated differently. Tampa's single hot spot is 13.5 mi^2 while Sandusky's trio of hot spots is 32.5 m^2. These spatial extents illustrate the differentiation between the designated hotspots but more importantly suggests that the Tampa hotspot (A) is far more compact

than are Sandusky's hot spots, which has a spatial extent that is 2.4 times larger than the Tampa hotspot.

Boating Accident Distribution by Key Human and Technological Factors

The factors considered in this section are specific to characteristics of boating operators involved in recreational boating accidents. The two most frequently observed human boat operator characteristics in terms of recreational boating accidents are age and gender. Boat operators between 30-49 (age) represent 75% of boating accidents at the Tampa research site and 76.7% of the boat operators at the Sandusky research site, if the age range is expanded to 30-59 years of age (see Table 10). Figure 32 illustrates the spatial distribution of boat operators by the peak age ranges observed to dominate the maps. It can also be observed that boat operators with an age greater than 50 or less than 20 clustered in near shore shallow waters and protected harbors (see Figure 32).

Boat operator gender are also observable as a (human) recreational boating accident indicator at each research site. Table 10 illustrates that males represent 67.4% of the boat operators involved in a boat accident at the Tampa research site and 59.4% of the boat operators involved in a boating accident at the Sandusky research site. By contrast, the number of female boat operators (aggregated with those data where gender was not specified by the recording boating law officer) represented a much lower frequency of the boat operators at the Tampa research site (32.6%) although a more balanced distribution at the Sandusky research site (40.6%). Figure 32 illustrates the apparent random spatial distribution of these two boat operator classifications.

Table 10

Distribution of Boating Accidents by Age and Gender

	Tampa		Sandusky	
	Freq	Pct	Freq	Pct
Boat Operator Age				
0-19	34	19.8	12	7.7
20-29	26	15.1	16	10.3
30-39	23	13.4	17	11
40-49	46	26.7	31	20
50-59	19	11	32	20.6
60-69	8	4.7	18	11.6
70-79	4	2.3	6	3.9
Over 80	12	7	6	3.9
Total	172	100	155	100
Boat Operator Gender				
Female/n.a.	56	32.6	63	40.6
Male	116	67.4	92	59.4
Total	172	100	155	100
Boat Type				
Houseboat	1	0.6	-	-
Commercial	-	-	2	1.3
Powerboat	73	42.4	119	76.8
Sailboat	10	5.8	8	5.2
Jet Ski	88	51.2	26	16.8
Total	172	100	155	100
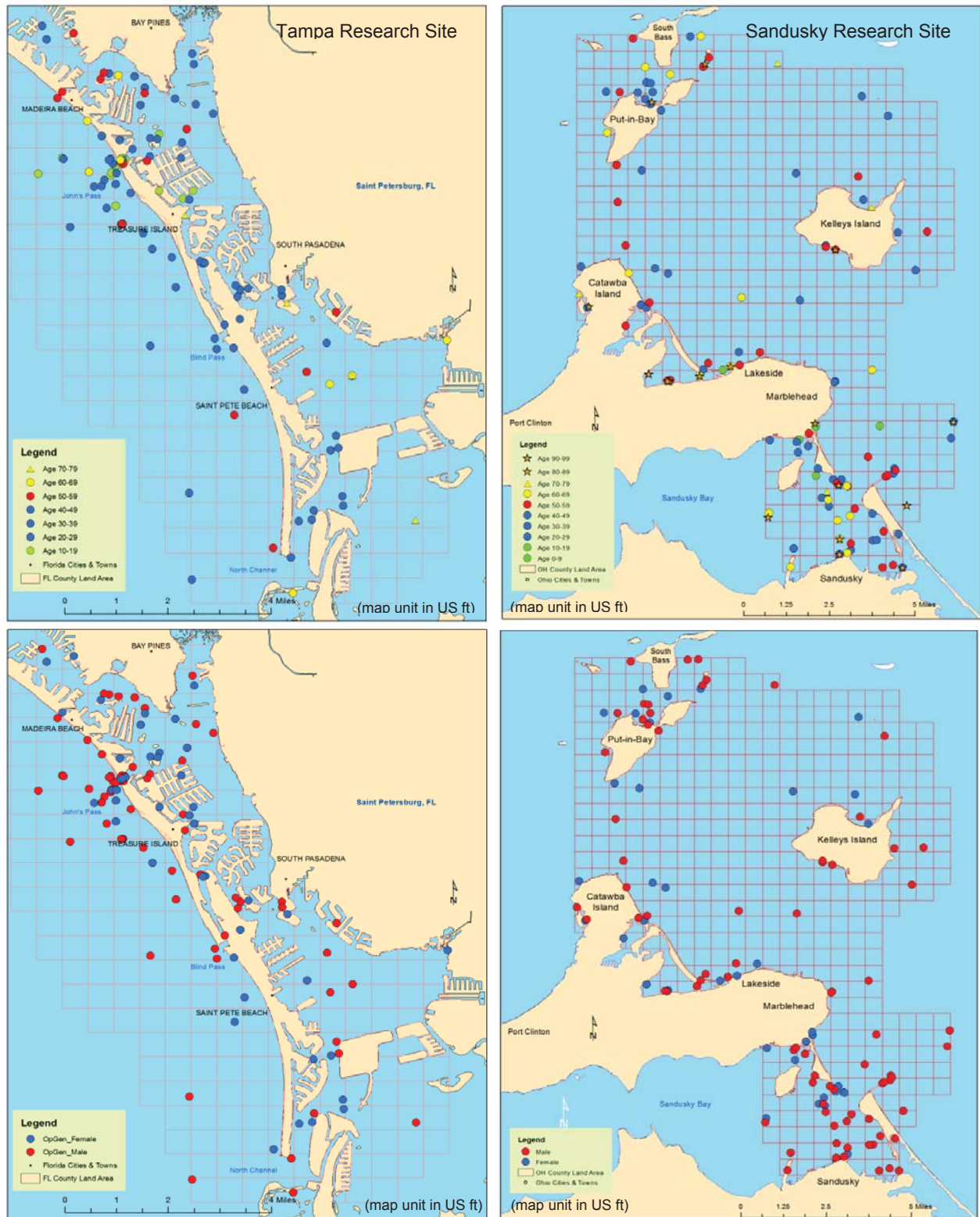
Figure 32. Distribution of Boat accidents by operator age and operator gender.

However, neither factor suggests an observable patten that can not be otherwise explained in terms of environmental factors detailed in this chapter and in Chapter 5.

Technologically, boat type is most suggestive of a dominate boat accident pattern. At the Tampa site, 51.2% of the accidents are attributed to jetskis while 42.4% of boating accidents are attributed to powerboats (see Table 10). By contrast, powerboats dominate the Sandusky site representing 76.8% of all boating accidents with a secondary peak representing jetskis observed at 16.8%. Figure 33 illustrates the spatial distribution of boating accidents by boat type. However, as was the case with respect to the human factors mentioned above, the descriptive strength of boat type as an explanatory technological factor is not suggestive of an observable patterns that can not be otherwise explained in terms of aforementioned environmental factors.

The Nature of Boating Accidents: Foreshadowing of a Statistical Model

Now that we have examined the temporal and spatial aspects of boating accidents within and between each respective research site, an appropriate next step is to consider the causal mechanisms that potentially yield these recreational boating accident patterns. Some of these accident risk influences have been alluded to within this chapter as well as a few boating accident correlates, e.g., constrained by navigation channels, water depth, and boat density. A more inclusive modeling approach permitting the simultaneous consideration of all influences is presented in Chapter 5. The research within this dissertation uses grid cells as the primary unit of analysis. The focus is to capture sparsely distributed boating accident (and non-accident) data into grid cells represented by the two research sites. Within each site, a majority of grid cells can be characterized as simply open water, i.e., those containing no boating
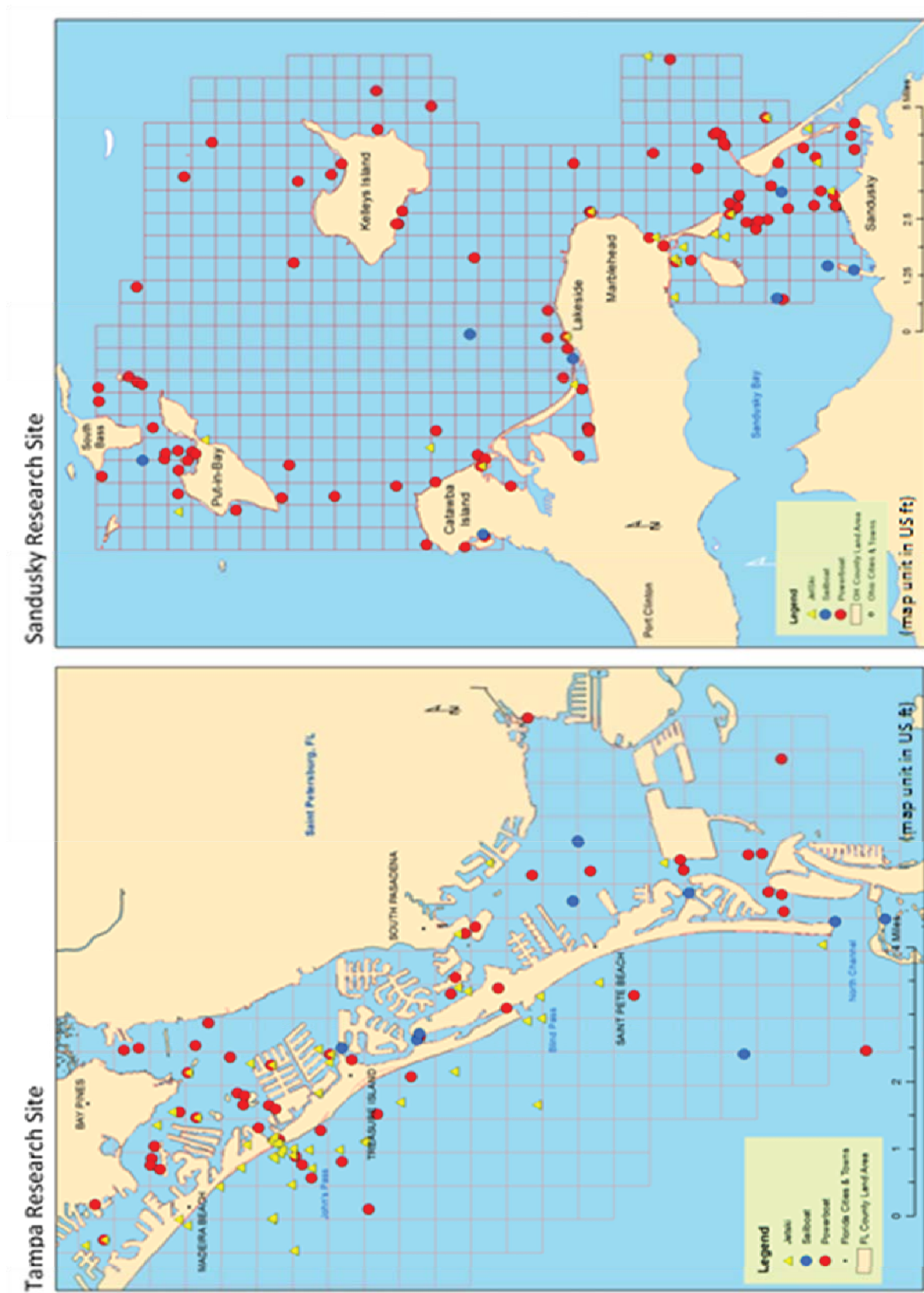
Figure 33. Distribution of boat accidents by boat type (classification).

accidents, no on-the-water observations, or no satellite observations related to the variables examined. In other words, grid cells with zero count values. As discussed earlier, excess zero's in research data is not an uncommon nor infrequent problem encountered in count-based research data (Hilbe, 2011; Hoef & Boveng, 2007). However, when modeling such data statistically, investigators often begin with an ordinary least squares (OLS) approach although it may not represent the best modeling approach. As illustrated in Chapter 3, if OLS is not optimal, two other regression-based modeling approaches are available, both of which are more typically used with count data; i.e., Poisson regression and negative binomial models (Hilbe, 2011; Shmueli, 2010). The main difference between the two approaches is that Poisson makes very restrictive assumptions about the relationship between the conditional mean and conditional variance while the negative binomial approach relaxes those assumptions (Hilbe 2011; Ismail & Zamani, 2013; Shmueli & Koppius, 2011). Furthermore, negative binomial theory suggests that excess zero counts can be separated from the desired variable count values (Fang, 2013; Hilbe, 2011; Ismail & Zamani, 2013; Shmueli & Koppius, 2011). For these reasons and in this study, OLS and NB2 regression models will be considered. Poisson will not be considered due is restrictive nature.

CHAPTER 5: DEVELOPMENT OF RECREATIONAL BOATING ACCIDENT MODEL

All accidental events exist within two primary domains (i.e., time and geographic space). Studies involving the intersection of these domains reveal important boating accident characteristics associated with the descriptive "when" and "where" described in Chapter 4. This chapter extends those questions by asking an equally important question. Why do boating accidents occur then and there? This why question is explicitly posed within a quantitative modeling framework. It begins with an Ordinary Least Squares (OLS) modeling approach that is comparatively used to illustrate the improved performance of the Negative Binomial (NB2) regression model that will ultimately be adopted and presented in response to the "why" question posed above.

The data used in this quantitative causal modeling framework originates from a combination of Federal Boating Accident Report Database (BARD), On-the-Water Vessel Safety Check (VSC) observations, and high resolution optical (0.8 m) GeoEye satellite imagery. BARD data represents the actual boating accident record whereas the on-the-water surveys and GeoEye satellite imagery collection represent boating activities and operator attributes activities in non-accident locations. All accident, on-the-water, and satellite imagery data used in the models and developed in this section are specific to the 2011-2012 time period. The use of the same time frame for all research data ensures that the response (accidents) and explanatory factors (explanatory variables) will be appropriately and properly matched.

The development of a Negative Binomial (NB2) model (hereafter called the Recreational Boating Accident Model) for this study represents a method of analyzing count data that explicitly accommodates excessive (high frequency) zero values. In this

research study and related to the units of analysis (0.25 mi^2 grid cells), count outcomes include a large number of observational units characterized as open water, i.e., typically containing few boats and no boating accidents (a value of 0). This high zero count phenomenon is known as over- or under-dispersion. Statistically this is adjusted through the inclusion of an over-dispersion parameter in the model (Fang, 2013). The NB2 model proposed in this investigation models non-zero observations (counts by grid cell) and minimizes dispersion that could otherwise bias the resulting parameter estimates and understate the standard errors associated with them (Fang, 2013).

The remainder of Chapter 5 is organized in four primary sections. The first section details and builds a case for using a negative binomial model to model boating accidents. The second section offers a descriptive analysis of the NB2 model explanatory variables included in the approach, e.g., boat density, seasonality, boat operator experience, boat speed, boat length, waterway intersections and channels, boat operator age, water depth, day of the week, and the possible effect of research site itself (Tampa and Sandusky) on overall results. This descriptive treatment is essential to laying a proper foundation for understanding the rationale behind explanatory variable selection as well as those included in the final model. Section three details the hypothesized model and final variable constellation proposed for modeling recreational boating accidents. This section also offers an in-depth analysis of each explanatory variable included and the influence of the variables selected for inclusion in the final model. The chapter concludes with a general statement about the nature of boating accidents. This will serve as a foundation for the results, conclusions, and future research directions presented in Chapter 6.

Construction of the Negative Binomial Regression Model

Before a statistical model can be constructed, a constellation of potential explanatory variables must be identified and tested. As detailed in Chapter 3, this pre-modeling analysis involved the use of the SAS JMP bootstrap forest routine. The variables selected and analyzed include on-the-water and satellite observations only. From an approximately 350 variable candidates (see Table 11), a subset of 30 final variables was selected including 18 satellite and 12 on-the-water explanatory variables (see Table 4) with above average influences on the outcome variable (boating accidents). Within this constellation, 7 of these variables are classified as human factor-based, 3 variables are classified as technology factor-base, 14 variables are classified as environmental factor-based, and 6 are classified as temporal factor-based. This 30 variable subset yields balance between the two primary components (surveys and satellite imagery) of the fused dataset further illustrating the importance of data fusion.

A Descriptive Analysis of the NB2 Model Variables

As developed and described in Chapter 2, recreational boating accidents stem from complex interactions, i.e., those that occur with combinations of boat accident dimensions (Rasmussen & Svedung, 2000). For this reason, boating accident risk models must include all variables in the recreational boat accident domain (Hovden, Størseth, & Tinmannsvik, 2011). To better illustrate the impact of these dimensions, the variables identified within the Recreational Boating Accident Model will be described in terms of dimensions they represent.

135

Table 11

Final Variable Candidates from Bootstrap Forest Constellation

	Term	Fusion Type	Dimension	Estimate	Std Error	Wald Chi²	p-value	Lower 95%	Upper 95%
	Intercept	satellite	environmental	-0.062	0.045	1.932	0.165	-0.150	0.026
1	S_Density1r	satellite	environmental	0.896	0.706	1.613	0.204	-0.487	2.279
2	S_BSpeed3	satellite	human	14.730	3.636	16.409	<.0001	7.603	21.857
3	S_BLength2	satellite	technological	12.836	4.188	9.393	0.002	4.627	21.046
4	S_BType3	satellite	technological	3.124	2.790	1.254	0.263	-2.344	8.593
5	S_Intersect1	satellite	environmental	12.756	3.145	16.450	<.0001	6.592	18.920
6	S_Constrain_1	satellite	environmental	-4.075	4.955	0.676	0.411	-13.787	5.636
7	V_WDepth2	on-water	environmental	10.864	3.241	11.236	0.001	4.511	17.216
8	S_Channel1	satellite	environmental	-6.096	2.311	6.956	0.008	-10.626	-1.566
9	V_OpAge2	on-water	human	-3.511	2.837	1.531	0.216	-9.071	2.050
10	V_OpEdu0	on-water	human	9.999	9.016	1.230	0.267	-7.672	27.669
11	S_JunJul	satellite	temporal	2.521	8.498	0.088	0.767	-14.135	19.177
12	V_JuneJuly	on-water	temporal	-0.205	1.954	0.011	0.916	-4.036	3.625
13	S_WDepth2	satellite	environmental	-15.497	5.850	7.017	0.008	-26.962	-4.031
14	V_PM	on-water	temporal	-20.541	4.240	23.466	<.001	-28.852	-12.230
15	S_Weekend	satellite	temporal	-28.215	14.163	3.969	0.046	-55.974	-0.456
16	S_MayJuly	satellite	temporal	-0.369	14.340	0.001	0.980	-28.474	27.737
17	V_OpAge3	on-water	human	5.858	3.584	2.671	0.102	-1.167	12.883
18	V_BSpeed3	on-water	human	5.494	2.030	7.321	0.007	1.514	9.474
19	S_SCA_0	satellite	environmental	4.383	11.766	0.139	0.710	-18.678	27.445
20	S_BType1	satellite	technological	3.490	8.402	0.173	0.678	-12.977	19.957
21	V_OpExp_N0	on-water	human	7.867	3.516	5.006	0.025	0.975	14.758
22	V_WDepth1	on-water	environmental	-7.722	6.267	1.518	0.218	-20.005	4.562
23	S_WDepth1	satellite	environmental	-1.090	6.132	0.032	0.859	-13.108	10.929
24	V_OpAge1	on-water	human	4.432	3.054	2.106	0.147	-1.553	10.418

136

Table 11 (continued)

	Term	Fusion Type	Dimension	Estimate	Std Error	Wald Chi^2	p-value	Lower 95%	Upper 95%
25	S_WDepth3	satellite	environmental	-2.342	3.327	0.495	0.482	-8.863	4.180
26	S_Shoal1	satellite	environmental	22.606	9.044	6.248	0.012	4.880	40.332
27	S_Weekday	satellite	temporal	-3.309	5.857	0.319	0.572	-14.789	8.170
28	V_Channel1	on-water	environmental	6.422	2.652	5.864	0.016	1.224	11.621
29	V_Constrain_1	on-water	environmental	-6.498	5.538	1.377	0.241	-17.351	4.356
30	Site	regional	regional	11.120	2.457	20.481	<.0001	6.304	15.935
	Dispersion			0.064	0.039	2.656	0.103	-0.013	0.140

Note. Response - A_1112_ACC_1; Distribution - Negative Binomial; Estimation Method - Maximum Likelihood; Number of rows - 717; AIC - 1922.2265; Generalized RSqure - 0.517.

There are three NB2 Recreational Boat Accident Model variables that are characterized within the human dimension. These include boat speed (S_BSpeed3, p-value <0.0001) and boat operator experience (V_OpExpNo, p-value 0.0069). S_BSpeed3 (satellite observations involving fast boat speeds) is organized under the human domain because boat speed is a matter of choice that is only limited by the specifications of the boat and engine. As illustrated in Table 12, 47.1% of the boating accidents within the Tampa research site involved boats characterized as moving at a fast rate of speed. By contrast, 40.6% of the boating accidents within the Sandusky research site involved boats characterized as moving at a fast rate of speed. If this category is expanded to include those boats moving at the second fastest boat speed (slow), the observed frequencies increase to 70.4% of the boating accidents at the Tampa research site and 71.6% of the boating accidents at the Sandusky research site. Figure 34 illustrates the spatial distribution of boating accidents as categorized by boat speed. It can be observed fast boat speeds seem to be clustered in areas associated with channels and shallow waters as opposed to open (deeper) water depths. This observation will be further examined in connection with an interactive variable included with the NB2 Recreational Boat Accident Model.

The two most discussed boating operator characteristics at the state and federal level pertaining to boating accidents are operator experience and the level of boating education. Table 12 illustrates these specific human dimension characteristics. Although boat operator education is not specifically included as part of the NB2

Table 12

Classification by Boat Length, Speed, & Operator Experience/Education

	Tampa		Sandusky	
	Freq	Pct	Freq	Pct
Classification by Boat Length				
<16'	90	52.3	34	21.9
16-26'	72	41.9	104	67.1
26-39'	10	5.8	17	11
Total	172	100	15	100
Classification by Boat Speed				
Stationary	28	16.3	17	11
Idle	18	10.5	19	12.3
Slow	40	23.3	48	31
Fast	81	47.1	63	40.6
Unknown	5	2.9	8	5.2
Total	172	100	155	100
Boat Operator Experience				
<10 hrs	35	20.3	12	7.7
11-100 hrs	35	20.3	17	11
101-500 hrs	57	33.1	45	29
>500 hrs	-	-	55	35.5
Unspecified	45	26.2	26	16.8
Total	172	100	155	100
Boat Operator Education				
none	135	78.5	94	60.6
Basic	27	15.7	35	22.6
Advanced	10	5.8	26	16.8
Total	172	100	155	100
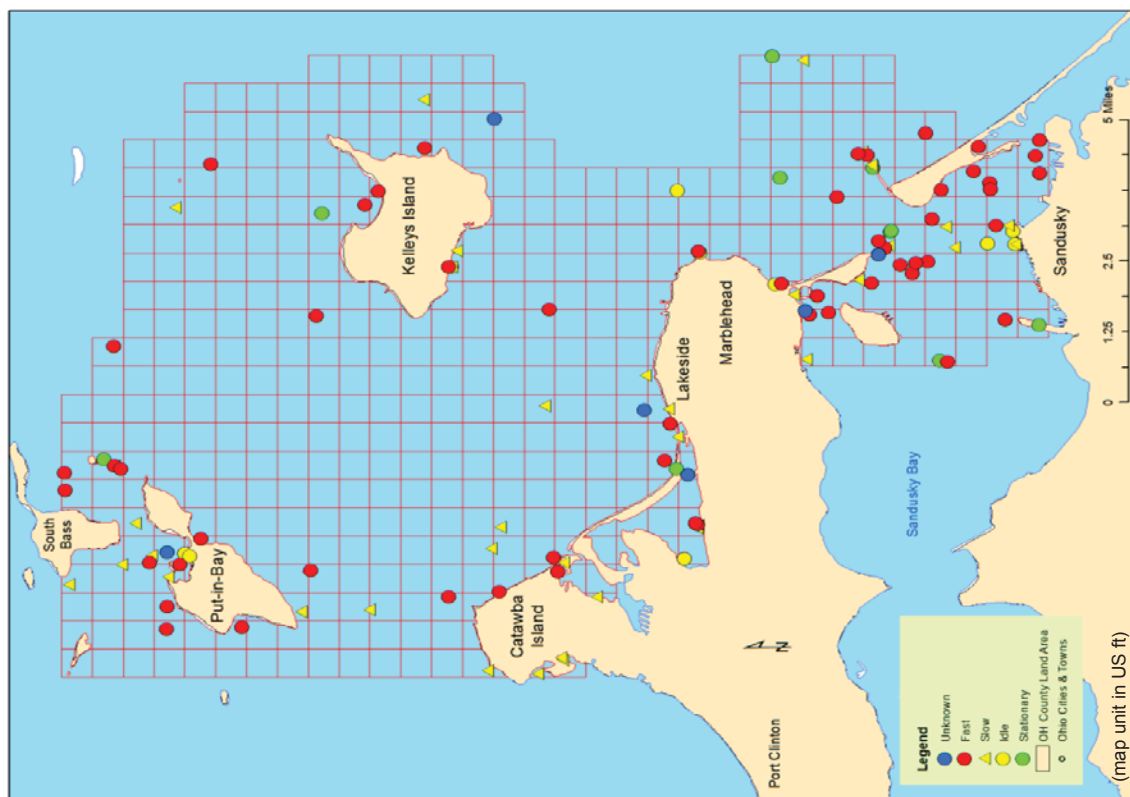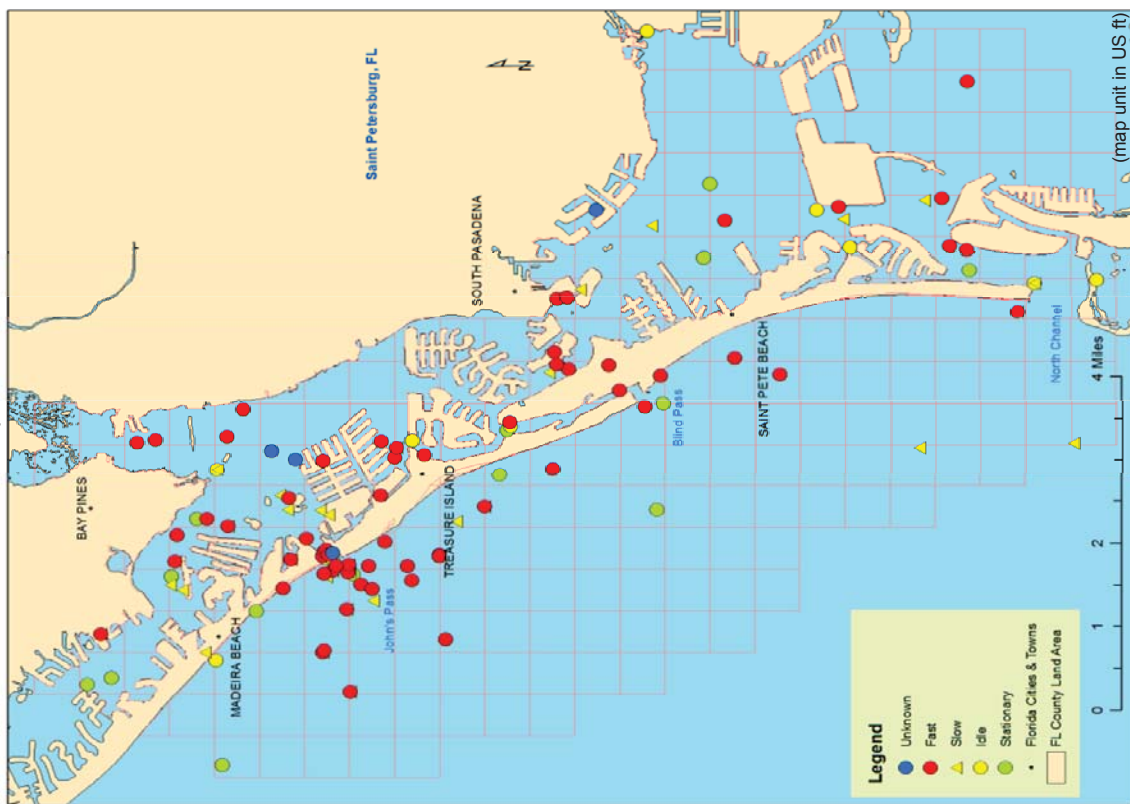
Figure 34. Boating accident distribution by boat speed.

140

Recreational Boat Accident Model, boat operator experience might serve as a proxy for

boat operator education. The rationale for this linkage is that although it is not

impossible to gain on-the-water boating experience in a self-taught manner, such

experience is frequently gained through formal boat operator education (Doll & Stiehl,

1979; Loeb, 1994; McKnight et al., 2007; O'Connor & O'Connor, 2005; Virk & Pikora,

2011). With respect to boat operator experience, operators within the Tampa research

site exhibit similar accident profiles at each experience level, i.e., less than 10 hours of

on-water-experience (20.3%), between 10 and 100 hours of experience (20.3%), and

more than 100 hours of experience (33.1%) totaling 73.7% of all accidents (see Table

12). By contrast, operators within the Sandusky research site suggest a slight

predominance of accidents with boat operator experience levels of 100 hours or more

(64.5%). This feature is clearly illustrated by the 100-500 hours of experience category

(33.1% in Tampa and 29.0% in Sandusky). These data suggest that most of the

operators involved in a boating accident have moderate to high levels of boating

experience on-the-water (see Table 12). This suggest a level of boat operator

confidence that permits more experienced boat operators to spend more time in open,

deeper water while less experienced boat operators spend more time closer to shore.

Closely coupled with the state and federally monitored operator experience

characteristic, is the level of boat operator education associated with boating accidents.

Of the boating operators involved in a boating accident in the Tampa site, 78.5%

reported that they had received no formal boat operator education (classroom or on the

water) and an additional 15.7% reported that they had received only the most basic boat

operations instruction (see Table 12). Collectively, these less-educated Tampa boat

operators represent 94.2% of all boating accidents within the research site. By contrast, 60.6% of the boating operators involved an accident in the Sandusky site reported that they had received no formal boat operator education and an additional 22.6% reported that they had received only the most basic boat operations instruction (see Table 12).

Collectively, these less-educated Sandusky boat operators represent 83.2% of all accidents within the research site. These boat operators can also be mapped spatially as illustrated in Figure 35. Both the Tampa and Sandusky sites illustrate that primary navigation channel and protected harbor areas show a strong mixture of experienced boaters with other boaters who have little or no experience in boat operation. It also suggests that the areas where this experienced vs. inexperienced boater intermixing is most frequent, there are higher concentrations of recreational boaters in general, e.g., John's Pass, navigation inlets, and protected harbor areas. If one broadens the boater experience perspective to include boat operator education, a strong intermixing of boat operators involved in an accident in locations with less boating education it can be observed (see Figure 35).

There is only one NB2 Recreational Boat Accident Model variable characterized within the technology domain; i.e. boat length (S_BLength2, p-value = 0.0016). S_BLength2 (class 2 boat length) is derived from satellite observations involving boat lengths (see Figure 36). Class 2 vessels are in the range of 16-25 feet in overall boat length. As illustrated in Table 12 and Figure 36, boats in this length category comprise 41.9% of all accidents at the Tampa site and 67.1% of the accidents at the Sandusky site. This range is expanded slightly to include smaller class 2 vessels, i.e., S_BLength_1 (class 1) ranging from less than 16 feet. As this range is further narrowed

Tampa Research Site

Sandusky Research Site
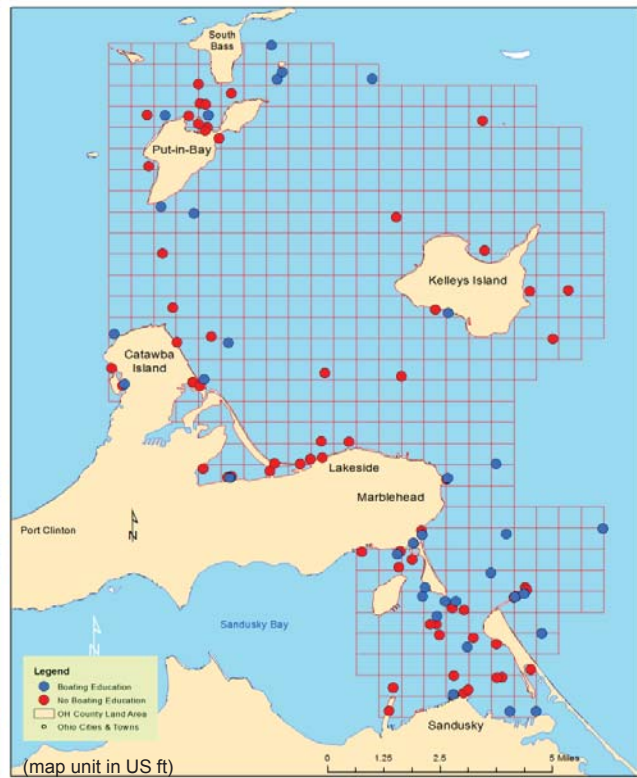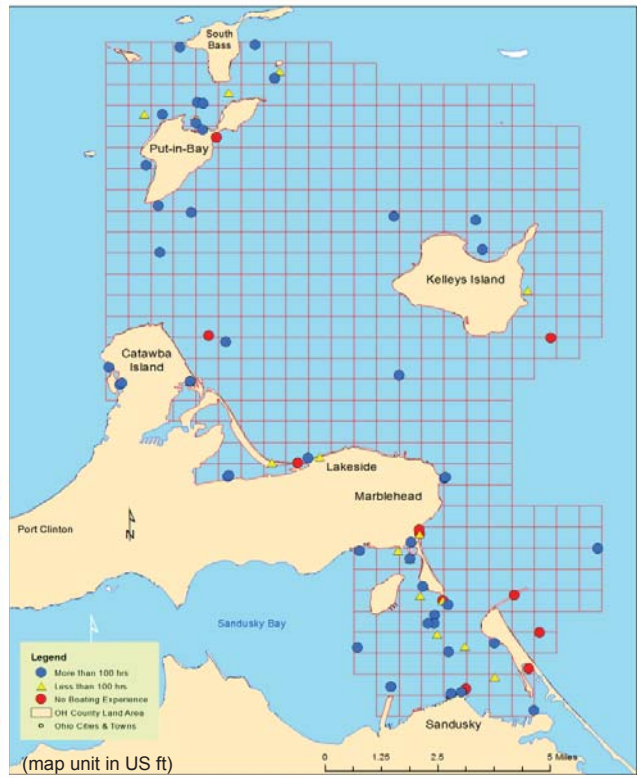
Figure 35. Distribution by boat operator experience and education.

143
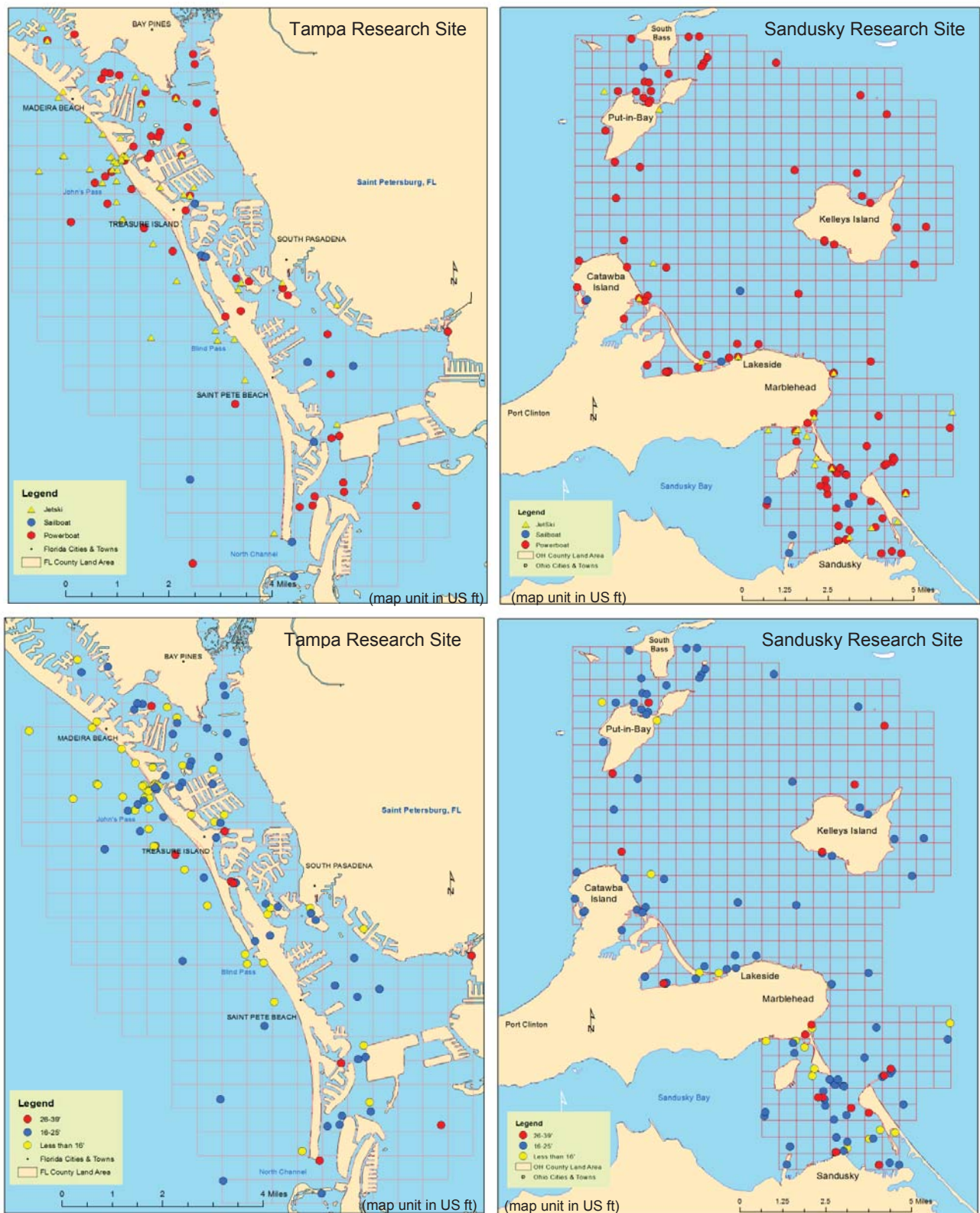
Figure 36. Boating accident distribution by boat type and length.

by boats that typically have motors (8 feet or greater), the combined influence of boats in the class 1 and 2 range represents 94.2% of the boating accidents at the Tampa site and 89.0% of the boating accidents at the Sandusky site.

There are four NB2 Recreational Boat Accident Model variables characterized within the environmental domain. These are: boat density (S_BDensity1r), the research site (Site-Tampa or Sandusky), a description as to whether the boat being observed is operating in a channel or not (V_Channel1), and a description of the water depth in which boat is operating (V_WDepth2).

Examination of boat accident density (S_BDensity1r, p-value 0.0153) reveals clues as to the importance of this variable in the NB2 Recreational Boating Accident Model. Figure 37, section I contrasts the distribution of boating accidents from 2005-2012 against the mean density of non-accident boats (as observed by satellite imagery). These data provide evidence that boating accidents are positively linked to the average number of boats operating within any selected area. The spatial distributions as each site in general suggest that as the number of boats operating within an area increase, there is a corresponding increase in boating accident frequency. On the other hand, Figure 37 suggests that this is a relatively loose coupling with some higher density boating areas having few, if any, boating accidents. This is particularly true in areas characterized by open water and non-channelized areas (see Figure 37, section II).

The next significant variables in the Recreational Boating Accident NB2 model are V_Channel1, p-value <0.0001 and V_WDepth2, p-value <0.0001. Both of these observations are collected by boating law enforcement officers in their respective sites. Figures 37-38 illustrate the distribution of boating accidents with respect to whether the
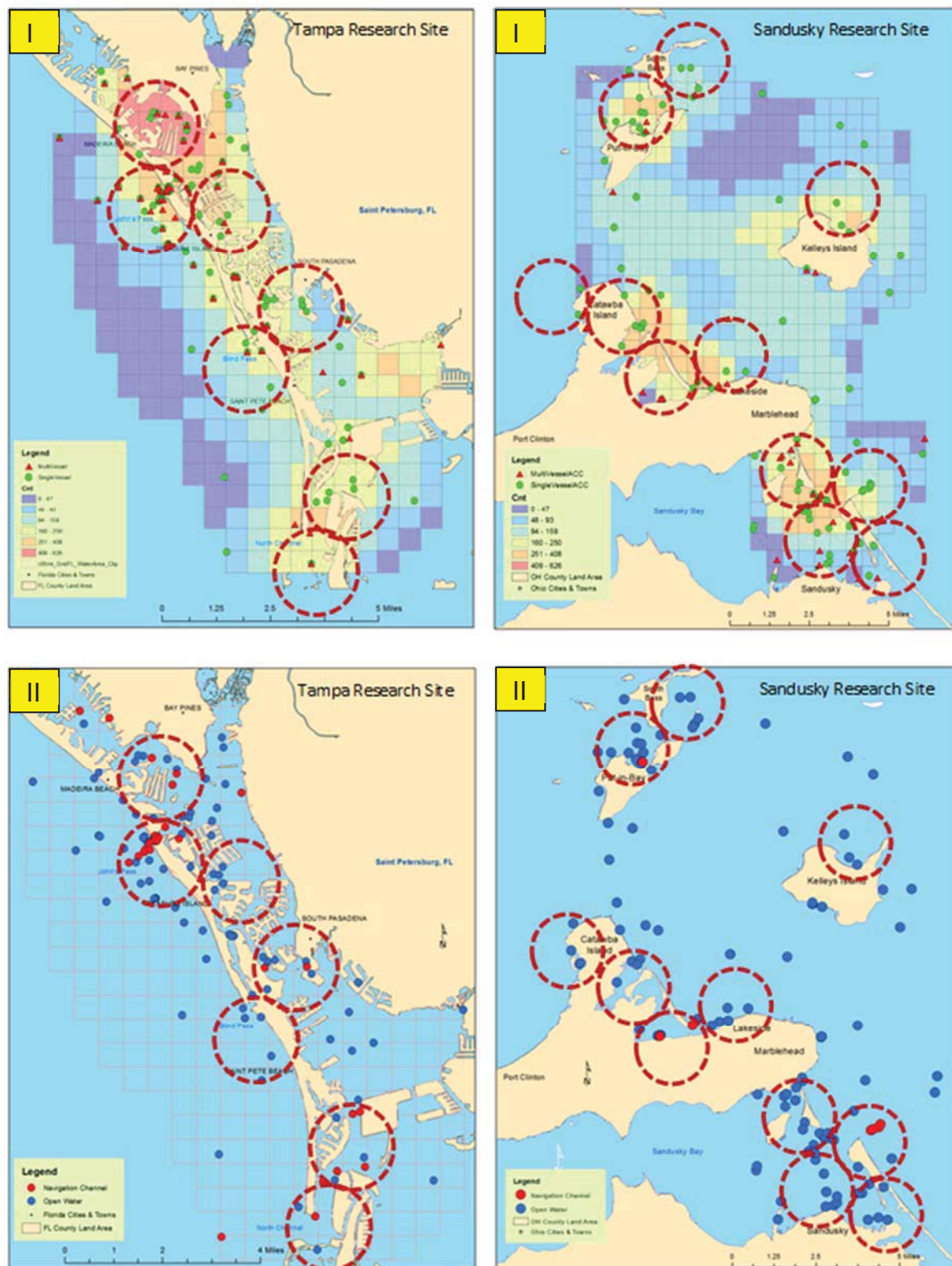
Figure 37. Boat accident density / satellite-based boat density overlay showing distribution in-proximity to NOAA Designated Channels vs Open Water.
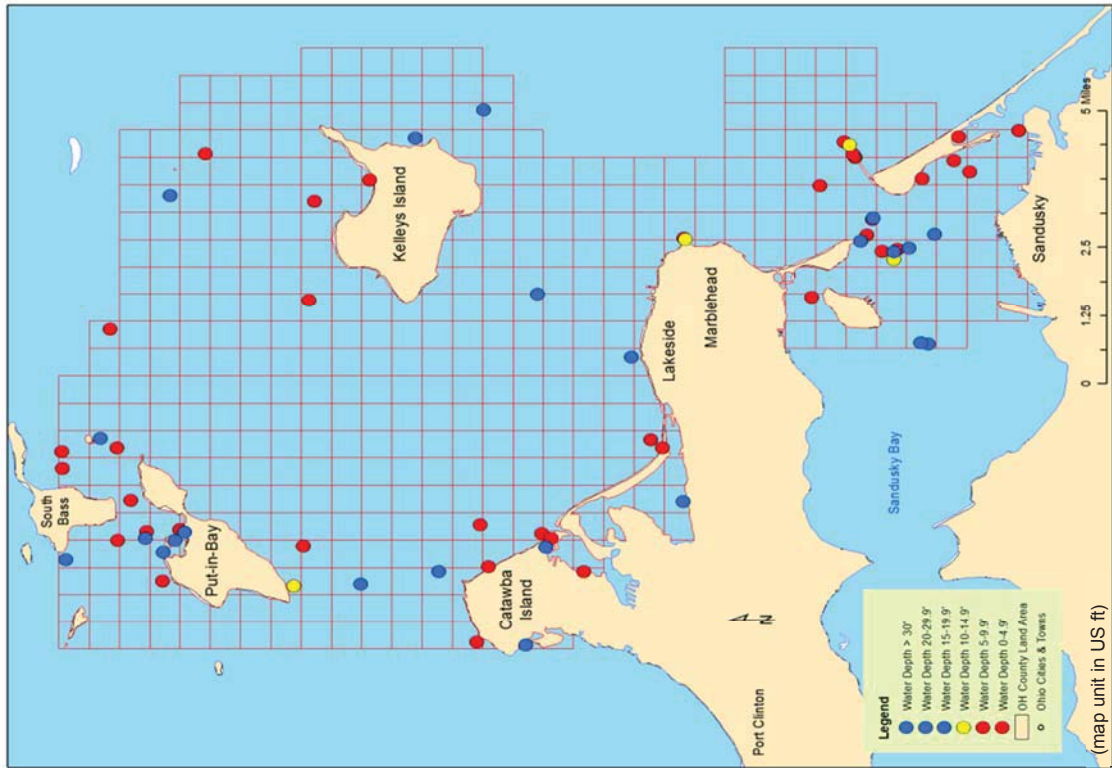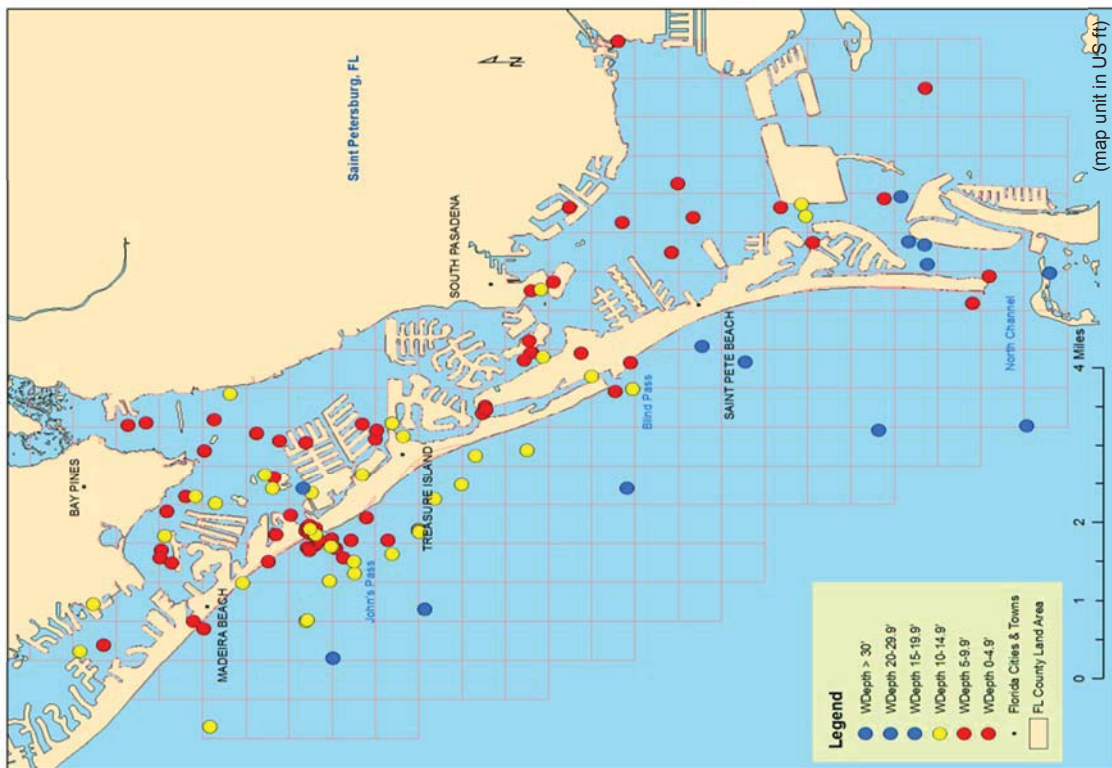
Figure 38. Spatial distribution of boating accidents by water depth.

147

boating accident occurred in a NOAA designated channel or not and the depth of the water where the accident occured. Clusters of accidents within areas where channels exist is evident from the maps. However, if this specifically defined area is expanded to 1 mile radius around the grid cell centroids of these designated channel areas, a substantial increase in the frequency in boating accidents can be observed (see Figure 37). For this reason, the influence of NOAA designated channels on accidents may compliment or interact with other influences; e.g., the joint influence of channel and water depth and will be tested as an interactive variable included with the NB2 Recreational Boat Accident Model.

The frequency of boating accidents that occurs in water depths between 5-9.9 feet represents 40.1% of all boating accidents at the Tampa site and 41.3% at the Sandusky site (see Table 13). A slight expansion of this range to include water depths between 1-14.9 feet accounts for 90.7% of the boating accidents at the Tampa site and 69.7% of the boating accidents at the Sandusky site. Collectively, these observations suggest that shallow water depths, i.e., those below 15 feet influence a majority of the boating accidents. This influence can be observed spatially in Figure 38. These two areas are significantly different and those differences are captured by the SITE variable.

There are two NB2 Recreational Boat Accident Model variables characterized within the temporal domain; i.e. satellite-based boat observations that occur during the June-July timeframe (S_JunJul, p-value <0.0473) and whether boating observations occur on a weekday or during the weekend boating period (S_Weekday, p-value <0.0001). Figure 39 illustrates monthly boating accident trends referred to as boating accident seasonality. A seasonal peak is clearly evident in the June through

Table 13

Distribution of Boating Accidents by Site and Water Depth

	Tampa		Sandusky	
	Freq	Pct	Freq	Pct
Water Depth 0-4.9ft	28	16.3	36	23.2
Water Depth 5-9.9ft	69	40.1	64	41.3
Water Depth 10-14.9ft	59	34.3	8	5.2
Water Depth 15-19.9ft	12	7	18	11.6
Water Depth 20-29.9ft	2	1.2	15	9.7
Water Depth >30ft	2	1.2	14	9
Total	172	100	155	100
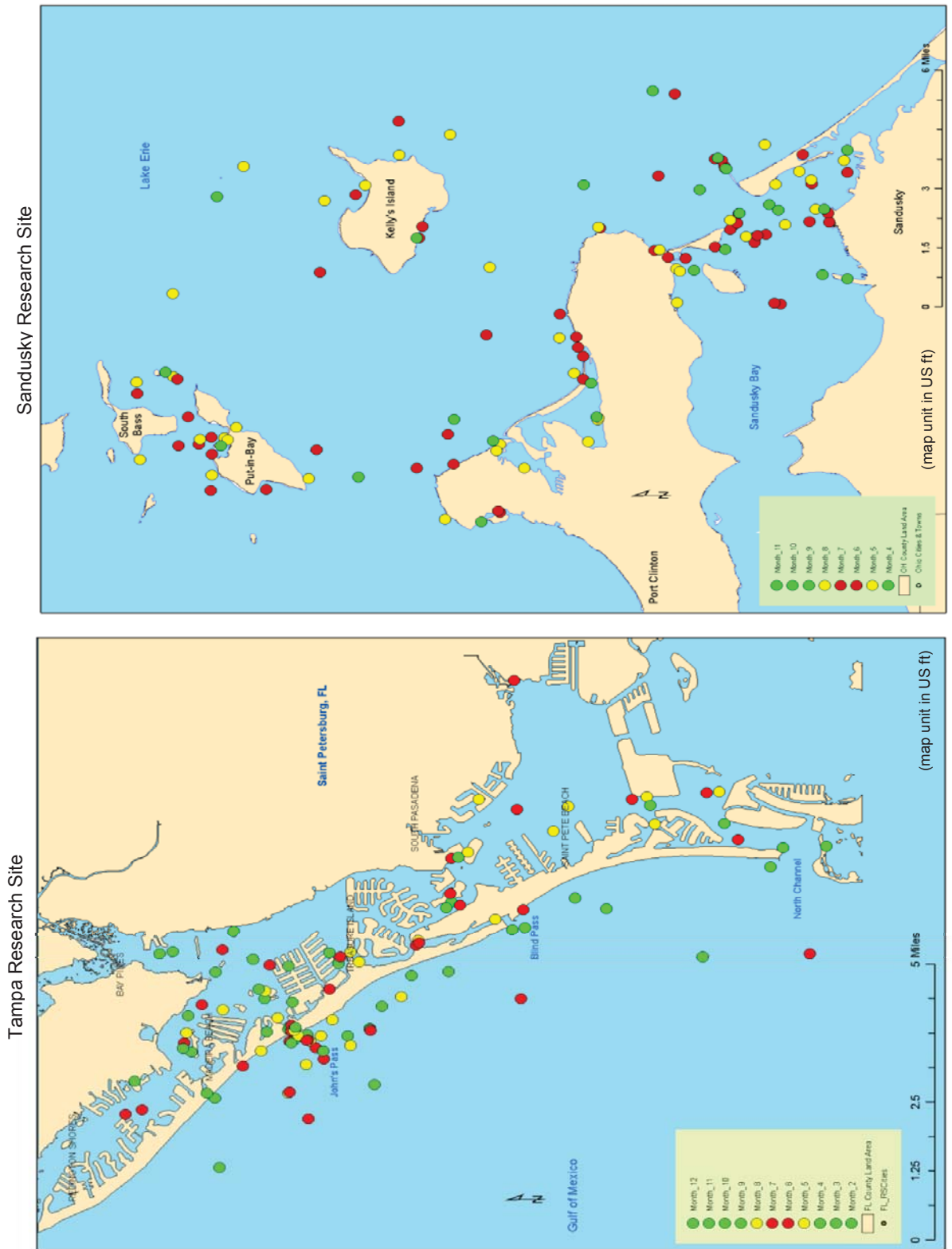
Figure 39. Seasonality in boating accident distribution by month.

August period. This peak is more observable and pronounced in the Sandusky site (80%) than in the case of Tampa (52%). Satellite observations during the June-July period serve as a proxy for this mid-seasonal temporal period from which seasonal impacts can be estimated. Daily variation at both research sites is illustrated by accident trends as shown in Figure 40. The frequency of boating accidents at both the Tampa and Sandusky sites provides evidence of an increased accident frequency during weekends (defined as Saturday and Sunday). The only deviation from this generalization is evidenced on Friday's especially observable at the Sandusky site. In general, monthly variation suggests a stronger spatial relationship with respect to the distribution of boating accidents than daily variation (see Figure 41). The Tampa site illustrates twelve months of relatively continuous boating activity, the Sandusky research site exhibits seven to eight months of boating activity with a strong winter lull. This fraction of the time series can be described as "in season" while the remainder of the year is described as "off season". On the other hand, closer observation reveals that the period of higher boating accident frequency at the Sandusky site actually extends into October (early fall).

The Negative Binomial Based Boating Accident Model

To better understand the parameter estimates shown in Equation 1, the estimates were converted to an incidence rate ratio (IRR) parameterization (Hilbe, 2011) and can be interpreted as "risk" ratios. IRRs are adopted because log counts are difficult to handle and explain in practical situations (Hilbe, 2011). The influences of the resulting IRRs and the associate contribution percentages are discussed in depth later in this section.
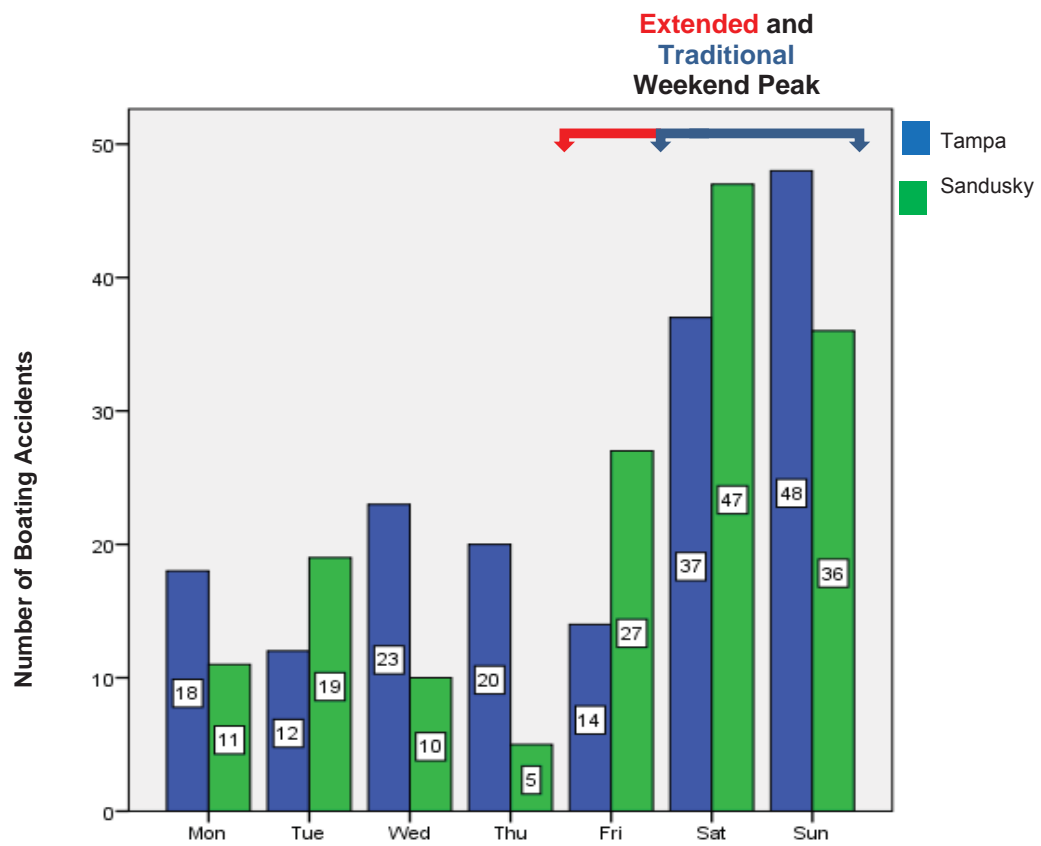
Figure 40. Boating Accidents by day of week.
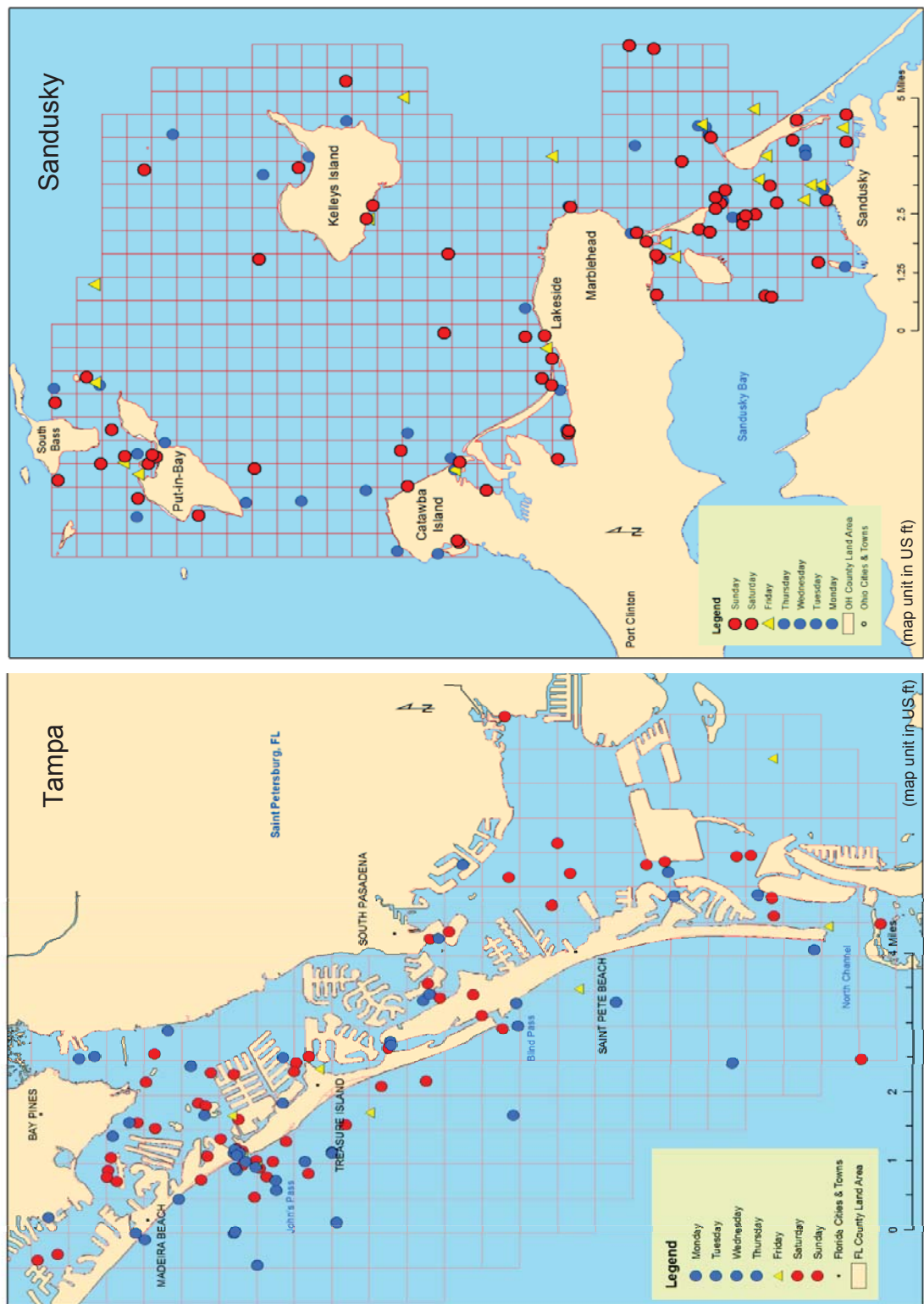
Figure 41. Seasonality in boating accident distribution by day of week.

As documented in this chapter, the most significant explanatory variables (candidates for a final model) included: S_BDensity1r, S_JunJul, V_OpExp_No, S_BSpeed3, S_BLength2, V_Channel1, V_WDepth2, S_Weekday, and the interactive variable V_Channel1 * V_WDepth2 (see Figure 42). The NB2 model resulting from the inclusion of these variables is illustrated in the following equation:

$\ln(E(\text{Boating Accidents}))$ = -1.790 + 0.077 (S_BDensity1r) + 0.003 (S_JunJul) +

0.078 (V_OpExp_No) + 0.017 (S_BSpeed3) + 0.013 (S_BLength2)

- 0.009 (S_Weekday) + 0.996 (Site)

+ 0.035 (V_Channel1) + 0.040 (V_WDepth2)

- 0.002 (V_Channel1 - 7.63) * (V_WDepth2 - 4.36)   (9)

where each coefficient represents the exponentiation of the NB2 parameter estimates for original predictors. For example, the exponentiation of S_BDensity1r ($e^{0.077}$) equates to an IRR value of 1.080. ln (E(Boating Accidents)) expresses the expected (mean) number of accidents as a linear function of explanatory variables. The NB2 over-dispersion parameter was estimated to be 0.167 (with a 95% Confidence Interval between 0.069 and 0.265). This estimate suggests that the model is slightly over-dispersed.

As shown in Figure 42, the explanatory variables make a statistically significant contribution to the prediction of boating accidents. For each of the predictor variables, the regression coefficient was tested under the null hypothesis that the count parameter is equal to zero, controlling for all the other variables in the model. The larger the absolute value of the regression coefficient, ignoring any negative signs, the stronger

Response	A_1112_ACC_1					
Distribution	Negative Binomial					
Estimation Method	Maximum Likelihood					
Number of rows	717					
AIC	1957.259					
Generalized RSquare	0.463					
Term	Estimate	Std Error	Wald Chi^2	p-value	Lower 95%	Upper 95%
Intercept	-1.790	0.177	102.685	<.0001	-2.136	-1.444
S_Density1r	0.077	0.032	5.887	0.015	0.015	0.139
S_JunJul	0.003	0.001	3.934	0.047	0.000	0.005
V_OpExp_N0	0.078	0.029	7.291	0.007	0.021	0.135
S_BSpeed3	0.017	0.004	20.537	<.0001	0.009	0.024
S_BLength2	0.013	0.004	9.974	0.002	0.005	0.021
V_Channel1	0.035	0.005	43.310	<.0001	0.024	0.045
V_WDepth2	0.040	0.006	41.674	<.0001	0.028	0.053
S_Weekday	-0.009	0.002	17.840	<.0001	-0.014	-0.005
Site	0.996	0.180	30.480	<.0001	0.643	1.350
(V_Channel1-7.62762)*(V_WDepth2-4.35983)	-0.002	0.000	64.613	<.0001	-0.003	-0.002
Dispersion	0.167	0.050	11.226	0.001	0.069	0.265

Figure 42. NB2 recreational boating accident model.

the unique contribution of that variable in explaining the dependent variable, i.e., boating accidents

The influence of the variable **S_BDensity1r** has a p-value of 0.015. The incident rate ratio (IRR) of **1.080** suggests that increasing boat density as measured by satellite imagery strongly influences whether a boat will be involved in a boating accident or not. This indicates that as the number of boats within a specific area increases by one unit (boat density), the accident rate increases by 8.0%. Thus, boat density is observed to have a strong influence on recreational boating accidents. Refer to Table 14 for a summary of these IRR values and % change probabilities for this and other variables listed in equation 1 (above).

The influence of the variable **S_JunJul has** a significant p-value of 0.047. The IRR of **1.003** suggests that boat operating on the water during the June-July period are 0.3% more likely to be involved in a boating accident than those operating a boat during other times of the year. As noted in section 5.4, the importance of this parameter estimate is one of illustrating the impact of seasonality in recreational boating. However, the weakness (small size) of the parameter estimate suggests that seasonality has only a small influence on the dependent variable when compared to other predictors.

The influence of the variable **V_OpExp_No** is one of the strongest predictors in this model. In addition to being significant with a p-value of 0.007, the IRR of **1.081** suggests that boat operator experience is a strong indicator as to whether a recreational boater will be involved in a boating accident. This predictor variable suggests that with each additional on-the-water boater (within a cell) with little or no experience, the

156

Table 14

Incident Rate Ratio (IRR) Parameterization based on Figure 42 Parameter Estimates

Term	Estimate	IRR	% Chg	effect
Intercept	-1.790			
S_Density1r	0.077	1.080	8.0%	very strong effect
S_JunJul	0.003	1.003	0.3%	effectively 0%; weak influence
V_OpExp_N0	0.078	1.081	8.1%	very strong effect
S_BSpeed3	0.017	1.017	1.7%	moderate effect
S_BLength2	0.013	1.013	1.3%	moderate effect
V_Channel1	0.035	1.035	3.5%	very strong effect
V_WDepth2	0.040	1.041	4.1%	very strong effect
S_Weekday	-0.009	0.991	-0.9%	effectively 1%; weak influence
Site	0.996	2.708	170.8%	strong regional effect
(V_Channel1-7.62762)*(V_WDepth2-4.35983)	-0.002	0.998		
Dispersion	0.167			

Note. IRR represents the exponentiation of the estimates.

157

accident rate increases by 8.1%. Clearly, experience is a good teacher when it comes to recreational boating and consequently the prevention of boating accidents.

The influence of the variable **S_BSpeed3** is a moderately strong predictor with a p-value of <0.0001. The IRR of **1.017** suggests that fast boat speeds influence whether a recreational boater will be involved in a boating accident or not. This indicates that with each additional fast-moving boat (in a cell) the accident rate increases by 1.7%.

The variable **S_BLength2** is a moderately strong predictor with a p-value of 0.016. The IRR of **1.013** suggests that boat lengths between 16-25.9 feet influence whether a recreational boat will be involved in a boating accident or not.   This indicates that with each additional on-the-water boat (in a cell) with a length in the 16 to 25 feet range, the accident rate increases by 1.3%.

The variable **V-Weekday** is a weak predictor with a p-value of <0.0001. The IRR of **0.991** suggests that weekend boat traffic (as measured by on-the-water vessel safety check stops) more strongly influences whether a recreational boat will be involved in a boating accident or not as opposed to weekday boat operations. More specifically, this variable indicates that weekday recreational boat will experience a 0.9% accident rate decrease as compared with recreational boats during a weekend. This variable estimate suggests a weak negative influence on boating accidents during the weekday.

The variable **Site** is the strongest predictor with a p-value of <0.0001. The IRR of **2.708** suggests that regional effects strongly influence the probability of boating accidents, when other explanatory variables are controlled for (see Figure 42 and Figure 43). This IRR suggests that the boating accident rate at the Sandusky site is 170.8% greater than at the Tampa research site, all else being equal. On the surface,

Response	A_1112_ACC_1					
Distribution	Negative Binomial					
Estimation Method	Maximum Likelihood					
Number of rows	717					
AIC	1989.519					
Generalized RSquare	0.437					
Term	Estimate	Std Error	Wald Chi^2	p-value	Lower 95%	Upper 95%
Intercept	-0.982	0.082	142.840	<.0001	-1.143	-0.821
S_Density1r	0.075	0.032	5.536	0.019	0.012	0.137
S_JunJul	0.006	0.001	25.026	<.0001	0.004	0.008
V_OpExp_N0	0.033	0.028	1.392	0.238	-0.022	0.087
S_BSpeed3	0.016	0.004	18.398	<.0001	0.009	0.023
S_BLength2	0.009	0.004	5.510	0.019	0.002	0.017
V_Channel1	0.022	0.005	20.088	<.0001	0.012	0.031
V_WDepth2	0.033	0.006	29.377	<.0001	0.021	0.045
S_Weekday	-0.012	0.002	27.794	<.0001	-0.016	-0.007
(V_Channel1-7.62762)*(V_WDepth2-4.35983)	-0.002	0.000	37.841	<.0001	-0.002	-0.001
Dispersion	0.172	0.053	10.541	0.001	0.068	0.275

Figure 43. NB2 recreational boating accident model excluding the site variable.

159

this seems contrary to a national boating accident ranking maintained by the U.S. Coast Guard within which Florida traditionally ranks 1st in boating accidents and Ohio typically ranks about eleventh. The Site variable suggests a strong regional effect. This implies that if other regions in different areas with the *exact* same conditions (as defined by the model) were compared with the Sandusky research site, the accident rate would be higher in Sandusky than in other areas. Chapter 6 will include a discussion regarding the probability of identifying other areas within the two research sites with the same conditions as described in the model; e.g., areas with channels, low water depth, less experienced boat operators, higher boat density, and the like. This indicates that with each additional vessel located within navigation channel cells, the accident rate increases by 3.5%.

The variable **V_WDepth2** is a strong predictor with a p-value of <0.0001. The IRR of **1.041** suggests that water depth (shallow water) between 0-4.9 ft. significantly influences whether a recreational boater will be involved in a boating accident or not. This estimator suggests that with each additional boat within a cell in a shallow water area, the accident rate increases by 4.1%.

Collectively, the IRRs (see Table 14) presented above represent percent incident rate ratios from the mean. Since the number of boating accidents in any specific observational unit is typically small, the percentage increase is reflective of a correspondingly small change as well. For example, assume that a specific observation unit is represented by 1.5 boating accidents and the %IRR suggests that a specific influence is manifested as a 5% increase in the number of accidents, then 1.5 accidents multiplied by a 1.05 %IRR yields an increase of 0.075 accidents. Since the analysis is

based on actual accident counts and the number of accidents in any one unit of observation is typically small, changes in the mean counts are small in terms of their impact on the other explanatory variables in the model. Another way to conceptualize the significance of the %IRRs is that although the NB2 process is a significantly improved method for analyzing count data, the number of counts within the observational unit is so small that the referenced changes are comparable to relative risks and odds ratios. Thus, while a small percent change will have a limited (sometimes negligible) effect on the model, the percent change for an accident variable increases or the number of accidents within the observational unit increases, or both, the effect of that influence would increase significantly.

The interactive variable is a more complex explanation than the explanatory variables in this equation as it represents a *nonlinear* interaction that is centered at the mean of **V_Channel1** and **V_WDepth2.**

$$e^{0.035} = \text{IRR } \textbf{(1.035)} \tag{10}$$

where a 1 unit change in **V_Channel1** where **V_WDepth2** is held constant at the average value (4.36)

$$e^{0.040} = \text{IRR } \textbf{(1.041)} \tag{11}$$

where a 1 unit change in **V_WDepth2** where **V_Channel1** is held constant at the average value (7.63) **(12)**

Collectively, **V_Channel1 * V_WDepth2** form an interactive explanatory variable with a significant p-value of <0.0001 where the:

$$\text{Interactive term} = -0.002 \; (\textbf{V\_Channel1 - 7.63}) * (\textbf{V\_WDepth2 - 4.36}) \tag{13}$$

As V_Channel1 increases beyond the mean of 7.63 and V_WDepth2 falls below the

average, the accident incident rate increases at a mild nonlinear rate. For example, if

V_Channel1 = 10 and V_WDepth2 = 2, the:

Interaction = -0.002 (2.37) (-2.36) = 0.0112 **(14)**

As V_Channel1 decreases below the mean of 7.63, then the opposite affect occurs.

V_WDepth2 is below the average and the accident incident rate decreases at a mild

nonlinear rate. For example, if V_Channel1 = 5 and V_WDepth2 = 2, the:

Interaction = -0.002 (2.63) (-2.36) = -0.0124 **(15)**

It is important to recognize that the complicated effect of these influences represents a

nonlinear change on a logarithmic scale. Further consideration of this aspect of the

interpretation will be reserved for discussion in Chapter 6.

The estimate of the dispersion parameter also has a significant p-value of 0.0008

with an estimate of 0.17. The estimate is roughly centered within the 95% confidence

range indicating that the actual estimate ranges from a value of 0.069 to 0.263. This

further suggests that the data are under-dispersed and that there are more count values

below the mean than should be expected under a standard Poisson framework (where

the mean and variance are equal). As such the other slope values and standard errors

are adjusted to offset this under-dispersion. Further consideration of this aspect of the

interpretation will be reserved for discussion in Chapter 6.

A final point of emphasis pertains to the question about whether the model will

remain stable if subjected to scrutiny involving data disaggregation by state (Site). The

consolidated model represents both research sites collectively. Figure 42 and 43

illustrate the effects associated with disaggregating the research sites while including

the remaining explanatory variables. As observed, there are no significant changes in the p-values or parameter estimates between the two models. However, there is an observable degradation in AIC performance (model fit) from 1957.26 to 1989.52 (1.6% reduction). Although this degradation is slight, it does represent reduced model performance except with respect to the variable V_OpExp_No that has a much less significant p-value of 0.238. This specific variable degradation is observable as an artifact of the difference between boat operator experience levels at the Tampa vs. Sandusky sites illustrating one of the key regional differences between the two sites. The general trend associated with boat operator experience at the Tampa site is toward the lower end of the experience range while the trend at the Sandusky site is toward the upper end of the experience range (see Table 12 and Figure 35). In each of the model iterations, variables surfaced that are more appropriate to capture regional differences between the two research sites. While every effort was made to minimize these regional effects, limitations in the number of research sites and data collected make it difficult to fully remove these regional effects. The inclusion of the Site variable tends to capture and smooth out these regional effects permitting results that are more generalizable as pertaining to the performance of the overall NB2 Recreational Boating Accident model.

Overall, all parameter estimates in the model are significantly different from 0 and have the correct sign. In addition, the overall model fit is good (given the limited number of explanatory variables).

163

The General Nature of Boating Accidents

As indicated earlier, this study uses 0.25 mi^2 grid cells as the unit of observation (analysis) to capture sparsely distributed recreational boating accident and non-accident data within the Tampa and Sandusky research sites. The focus and goal of this examination is to better understand the nature of recreational boating accidents, e.g., the when and where and more importantly to open the door to the question of why they occur. These questions were explored in depth in Chapters 4 and 5 of this dissertation. Chapter 6 will suggest study conclusions beginning with a high level review of what has been learned, what can be estimated, and what opportunities await continuation of research into recreational boating accidents.

CHAPTER 6: FINDINGS, DISCUSSION, RECOMMENDATIONS, AND FUTURE STUDY

The primary objective of this investigation is to better understand the impact of generally ignored environmental influences on recreational boating accidents. These factors are considered in combination with human and technological influences that have been the focus of previous research. An improved understanding has been achieved through the collection and analysis of key factors using appropriate data fusion techniques. In this investigation, data fusion was achieved through the integration of BARD accident reports, on-the-water vessel surveys (VSC), satellite imagery data, and a variety of agency data that were integrated into a common GIS database. The highly integrative nature of this approach permitted the thoughtful examination of key variables tagged in both geographic space and time. This approach has led to findings about the human, technological, and environmental influences that affect the rate of recreational boating accidents.

This research illustrates that while the contemporary boating accident literature, and much of the current government mindset, are concentrated on human and technological risk factors, the inclusion of environmental considerations is essential to a comprehensive understanding of recreational boating accidents. Previous research has focused on human and technological risk factors at the expense of ignoring environmental factors. As a result, the environmental focus in this research yields significance within the field of marine transportation and more specifically boating accidents.

The purpose of this chapter is to provide a summary and synthesis of the findings. Section one creates a position for this project and its findings within the contemporary body of literature. Section two briefly restates the research questions then couples them with the findings presented in Chapters 4 and 5. Section three offers a discussion of the implications associated with these empirical findings including methodological limitations.   Section four presents policy and regulatory recommendations based on the findings as well as a discussion about the implications for boat operator training and behavior.   The final section of this dissertation indicates social significance and suggests opportunities for future research.

Position of Research within Literature

As detailed in Chapter 2, boating accident theory draws upon epidemiology theory (Gabe & Hite, 2003) and the key dimensions within which boating accident risks are embedded (Anselin, 2002; Sidman, Grant, & et al., 2005). Specifically, this conceptual design leverages Normal Accident Theory (Perrow, 1984, 1999). Normal Accident Theory also serves to bracket the primary boating accident variables defined in this study. It permits primary and loosely coupled independent variables (e.g., boat density, boat length, seasonality, weekend operation, boat speed, boater experience and navigation channel operations). This conceptual framework characterizes latent conditions, e.g., controlling boat operator experience/education or boat speed through local rules, regulations, and cultural influences. Latent conditions are significant because they can combine in unexpected ways (Marais, Dulac, & Leveson, 2007) and trigger interactive effects that influence accident risk. The works of McCarthy and Talley (1999) and Wang (2000) are suggestive of this foundation as related to interactive

multivariate approaches to boating accident models. The cumulative advancement of these approaches creates a pathway for future analyses using statistical methodologies that consider the influences of interactive effects across the space-time dimension.

As described in Chapter 2, the existing body of knowledge offers limited evidence of significant predictor variables from the environmental dimension and technological dimensions while concentrating on the human dimension. For example, while Loeb and Giliad (1984) and McCarthy and Talley (1999) focused on the human dimension, Wang (2000) and Gabe and Hite (2003) added support for the importance of boating safety experience and education as significant boating accident rate influences. Principally, this investigation builds upon the work of O'Connor and O'Connor (2005) who were the first to investigate the full (i.e., human, technological, and environmental) recreational boating accident domain.

Within the methodology developed in this study, 350 observational categories (variables) representing human, technological, and environmental characteristics were explored. The rationale is not just to produce a more effective and efficient model but to develop one that captures those variables that most influence recreational boating accidents. Additional Negative Bionomial (NB2) modeling was carefully evaluated to reach a final subset of influences that considered the exclusion of overlapping variables (i.e., those basically measuring the same human, technological, and environmental characteristics). These NB2 iterations culminated in a final model. The variables in this model included:

1.   boat density, representing observations derived from satellite imagery,

2. peak month observations, representing satelliteobservations during the June-July timeframe,

3. Boat Operator Experience, representing on-the-water boat surveys that reported the boat operator's "boating operations experience level",

4. Boat Speed, representing observations derived from satellite imagery (stationary, idle, slow, fast)

5. Boat Length, represents observations derived from satellite imagery based upon the visible boat characteristic of overall boat length,

6. Waterway Channel characteristics, represents observations derived from on-the-water surveys as being in a NOAA designated navigation channel,

7. Water Depth, represents observations derived from on-the-water surveys that characterize each boat observation according to the NOAA designated water depth of the area surrounding the observation.

8. Day of the Week characteristics, represents a satellite observation of as occurring during a weekday or a weekend.

9. Research site designation, represents a satellite observation from one of the two designated research sites: (1) Tampa, Florida and (2) Sandusky, Ohio.

10. Waterway Variation; represented by an interactive variable illustrating the combined effect navigation channel complexity and water depth.

Research Questions and Findings

As developed and described in Chapters 2, recreational boating accidents stem from complex interactions; i.e., those that occur with combinations of different human, technological, and environmental factors (Rasmussen & Svedung, 2000). For this

reason, superior boating accident risk models must include variables in the human, technological, and environmental domain (Hovden, Størseth, & Tinmannsvik, 2011). In addition, since all boating accidents occur within space and time, boating accident risk models must additionally include variables in the temporal domain. To better illustrate the impact of these space and time dimensions, the variables identified within the NB2 Recreational Boat Accident Model have been described collectively as a part of the domain which they represent.

To guide the data collection process that underlies this empirical dissertation, the following research questions were developed. Those research questions and hypotheses are presented as follows.

Research Question 1: Environmental Factors

There are four NB2 Boat Accident Model variables characterized within the environmental domain: boat density, the research site, channel location or not, and a description of the water depth.

Initially, an examination of boat accident density revealed clues as to the importance of this variable in the NB2 Recreational Boating Accident Model. These data provide evidence that boating accidents are closely linked to the average number of boats operating within any selected area. The observable pattern suggests in general that as the number of boats operating within an area increases, there is a corresponding increase in the number of boating accidents.

The next significant variables in the NB2 Recreational Boat Accident Model site (Tampa vs. Sandusky) are channel location and channel depth. Both variables are collected by boating law enforcement officers in their respective states. Clusters of

accidents exist in and around (one mile radius) areas where channels exist.

Collectively, these observations suggest that shallow water depths; i.e., those below 15

feet, influence a majority of the boating accidents.

It is clear from the results of this research that complex operating environments

yield higher recreational boating accident probabilities. Thus, the environment within

which boats operate can be attributed with varying environmental characteristics such

as traffic densities and weather states. Such environmental factors impact boat

maneuverability and boat control.

H1:   The frequency of recreational boating accident risks increase as boat

density in a navigable waterway-space increase.

Evidence supporting this hypothesis is found throughout Chapters 4 and 5. It can be

noted that there is a very close relationship between those grid cells with higher

numbers of boats located by satellite imagery and boat accident locations. This is

illustrative of a direct relationship between density and boating accidents. Findings

suggest that with each percentage point increase in boat density that the accident rate

increases by 8%.

H2:   Recreational boat accident risks are more frequent in areas where the

environment being navigated is more complex (i.e., channel orientation,

single vs. multi-channel waterways, water depth, and environmental

complexity such as intersections, bridges, and related obstructions).

Evidence supporting a positive relationship between boating accidents and

navigational complexity is also strongly indicated in this research. Although channel

orientation, intersections, bridges, and related obstructions were not suggested as

170

significant, the presence of multi-channel waterways and water depth are. This finding

suggests that complex operating environments do complicate the decision making

process in boating operations. However, differences in regional geography probably

play a role in the specific environmental characteristics that define complexity. The

areas established as having the greatest boat density were associated with the near

shore areas that are characterized as shallow in the presence of one or more navigation

channels. Moreover, other than boat density and operator experience, water depth and

the presence (or absence) of a navigation channel were the next most significant

predictors. Both environmental characteristics suggest that with each percentage point

increase in water depth and in the presence of a navigation channel, that the accident

rate increases by 4.1% as related to water depth and 3.5% as related to navigation

complexity. In part, the inference is that increasing water depth leads to increased

traffic and potentially boat speed while channelized waterways tend to reduce boat

maneuverability and increase overall boat density within an area.

The interactive variable included (channel and water depth) in the recreational

boating accident model is a significant contributor to boating accidents. It illustrates a

more complex explanation than attributed to the preceding explanatory variables and it

represents a *nonlinear* influence**.** This suggests that for each additional channel, there

is a corresponding 3.5% increase in the number of boating accidents but only when

accompanied by an average value of water depth. This represents one reason why

neither the effects of the number of channels nor water depth in isolation are fully

accounted for. It is important to recognize that the complicated effect of these

influences represents a nonlinear change on a logarithmic scale. However, the

generalizable observation is that as the number of channels increases and water depth

decreases or in situations where you have fewer channels but larger water depths,

boating accident risks increase in a nonlinear manner.

The implication of this finding is generalized into a mild interaction as the result of

"waterway complexity." In areas of decreasing water depth such as marinas, local

cruising (party) destinations such as beaches or sandbars, typically are also

characterized by multiple access pathways or channels. This increasing complexity

suggests an increased opportunity for boat operator confusion or inattention to detail

that is based on a geographic characteristic. Alternatively, fewer channels with deeper

water depth suggest the potential for greater boat operator confidence and faster boat

speeds, but again an increased accident risk associated with the local geographic

characteristics. Thus, when the interactive variables are the same; i.e., high channels

and high water depth or low channels and low water depth, the effects tend to cancel

out the overall effect of the interactive variable. This interaction has the potential to

greatly influence boating accident risk if these variables are significantly out of balance.

> H3:   *Recreational boating accident risks vary as a function of mean boating*
>
>   *traffic speed and variability.*

Almost half of the boats observed in the satellite imagery were moving at fast (as

measured by wake size) rates of speed (47.1% at the Tampa research site and 40.6%

at the Sandusky research site). More importantly, it was observed that most boaters

(approximately 70%) were operating in one of two modes, i.e., fast or slow. The model

suggests that with each percentage increase in boat speeds, the accident rate

increases by 1.7%.

There are two NB2 Recreational Boat Accident Model variables within the temporal domain, i.e., satellite-based boating accident observations that occur during the June-July peak and whether boating observations occur on a weekday or during the weekend.   The seasonality peak is seen in the June through August period. This trend is much more pronounced in the Sandusky research site. Satellite observations during the June-July period serve as a proxy for this peak seasonal period

    Daily variation at both research sites is illustrated by observable frequency of boating accidents (at both the Tampa and Sandusky) increased during weekend periods (defined as Saturday and Sunday). The only deviation from this trend is evidenced by a larger frequency on Friday especially at the Sandusky site.

    While temporal variation is inherent in most studies, environmental factors tend to dominate temporal aggregations. For example, boating accident "hot spots" can be more seasonal in nature than other recreational boating accident risk factors.

H13:   Recreational boating accident risks are more frequent during peak boating months (May-August).

    The empirical evidence simultaneously supports but slightly modifies the hypothesis. This hypothesis defines the peak season as May-August while the data suggest that September should be included within the peak season. A closer examination of the September data reveals that the first half of September time frame is more active than the second half. This suggests that the modifying influence during September is the Labor Day holiday period. However, satellite imagery data only weakly supported the inclusion of this temporal variable as an indicator leading to the decision not to include this variable in the model at this time.

173

H14: Recreational boating accident risks are more frequent during weekend periods (Friday-Sunday) than weekdays (Monday-Thursday).

From a general perspective, the boating accident records at both research sites strongly suggest increased accident frequency during weekend periods, i.e., those days of the week defined as Saturday and Sunday, as opposed to the weekday period defined as Monday through Friday. The only deviation from this generalized "weekend" boating accident pattern occurs at the Sandusky research site where the frequency of boating accidents is substantially higher on Friday's in comparison with other weekdays. This empirical feature suggests an extended weekend accident frequency pattern, which includes Friday, in the case of Sandusky. Similar to seasonality and using this extended weekend definition, the Sandusky research site is characterized by a much more concentrated weekend (Friday-Sunday) accident peak. Another interesting characteristic of the weekend boating accident pattern is highlighted by Tampa's peak on Sunday (as contrasted with Sandusky's peak on Saturday). As previously mentioned, these temporal patterns have significant implications for governmental boating safety resource allocations not to mention the deployment of those safety resources and assets.

Research Question 2: Human Factors

There are three significant NB2 Recreational Boat Accident Model variables that are from the human domain. These include boat speed and boat operator experience. S_ Boat speed is organized under the human domain unlike other factors such as boat length, boat speed is largely a matter of choice.It has been observed that fast boat speeds seem to be clustered in areas associated with channels and shallower water as

opposed to open (deeper) water depths. This observation was further explored in connection with the interactive variable included with the NB2 Recreational Boat Accident Model.

The two most discussed boating operator characteristics, are operator experience and the level of boating education. Although boat operator education is not specifically included as part of the NB2 Recreational Boat Accident Model, boat operator experience serves as a proxy for boat operator education. With respect to boat operator experience, operators within the Florida research site show similar distributions at each experience level. These data suggest that most of the boating operators involved in an accident have moderate to high levels of boating operations experience on-the-water. This may further suggest a level of boat operator confidence that permits more experienced boat operators to spend more time in open, deeper water while less experienced boat operators spend more time closer to shore.

The level of boat operator education of those involved in a boating accident is closely coupled with the level of boat operator experience of those operators. Of the boating operators involved in a boating accident at the Florida research site, 78.5% reported that they had received no formal boat operator education and an additional 15.7% reported that they had received only the most basic boat operations instruction. By contrast, 60.6% of the boating operators involved in a boating accident at the Ohio research site reported that they had received no formal boat operator education and an additional 22.6% reported that they had received only the most basic boat operations instruction.

These boat operators, i.e., those involved in a boating accident, were also mapped spatially. Both the Tampa and Sandusky sites illustrate that primary navigation channel and protected harbor areas show a strong mixture of experienced boaters with other boaters who have little or no experience in on-the-water boat operation. Evidence also suggests that this experienced vs. inexperienced boater intermixing is most frequent in areas where there are higher concentrations of recreational boaters in general, e.g., Johns Pass at the Tampa research site.

The rationale for including these closely coupled boat operator characteristics is that in general, boat operator experience is gained through boat operator education. There are regional differences in both boat operator experience and education between the two sites. For this reason, high frequencies of low levels of boat operator education reinforce findings that boat operator experience is one of the key explanatory variables in the NB2 Recreational Boat Accident Model.

H9:   *Recreational boating accident risks are more frequent in areas where average boat operator experience level is lower.*

H10:   *Recreational boating accident risks are more frequent in areas where average boat operator education level is lower.*

As noted in Chapter 5, the two most discussed boating operator (human factor) characteristics at the state or federal level pertain to boat operator experience and boat operator education. Chapter 5, Table 12 and Figure 35 illustrate these characteristics as existing within the Tampa and Sandusky research sites. At the Tampa research site, findings illustrated that nearly half of the boat operators involved in an accident have fewer than 100 hours of on-water-experience. By contrast, results from the Sandusky

site suggest a smaller probability of accidents associated with low levels of experience. This evidence points to a strong regional effect with respect to boat operator experience. The data also suggest that in general boating operators with moderate to high levels of boating experience have accidents in the same places as less experienced operators, i.e., areas near navigation inlets and in water that is shallow. The most significant boat experience evidence can be found in the NB2 Recreational Boating Accident model developed in this study. That model indicates that boat operator experience is a significant predictor of boating accidents suggesting that with each increase of one percentage point in the number of on-the-water boat operators with no experience that the boating accident rate increases by 8.1%.

In addition, the more boat operator education received when coupled with on-the-water practical experience, the more proficient the boat operator. If one examines boat operator education as related to boating accidents within the respective research sites, it can be observed that 94.2% of the boat operators involved in a boating accident reported that they had basic to no formal boat operator education. When coupled with evidence provided by the NB2 Recreational Boating Accident model, indicating that boat operator education was not a significant predictor of boating accidents, this suggests that there is a stronger relationship between boat operator experience and education than is yet to be discovered. It also tends to disprove on of the findings of the Gabe and Hite (2008) study, which indicated that boat operator education does not have a statistically significant effect on recreational boating accidents.

H11: Recreational boating accident risks are more frequent in areas where average boat operator age is below 20 or greater than 70.

With respect to boat operator age, 75% of the boaters involved in an accident at the Tampa research site were between 30-49 years of age while 76.7% of the boat operators at the Sandusky research site who were involved in a boating accident were between 30-59 years of age. This evidence is coupled with that provided by the NB2 Recreational Boating Accident model, indicating that boat operator age is not a significant predictor of boating accidents.

Research Question 3: Technological Factors

There is one NB2 Recreational Boat Accident Model variable characterized within the technology domain, i.e., boat length. Boat length (class 2 boat length) is based on satellite observations of boats in the range of 16-25 feet. As illustrated in Chapter 4, boats in this range comprise 41.9% of all accidents at the Tampa research site and 67.1% of the accidents at the Sandusky research site. The NB2 model developed clearly indicates that while technological factors are significant with respect to recreational boating accidents, boat length and boat type are far more important when estimating accidents than is the size of the power plant (i.e., engine size). Additionally, although some anecdotal evidence provided by FWC and ODNR boating law enforcement officers pointed toward to potential importance of the navigational equipment (e.g., GPS) as used by the boat operators in determining accident risk, the evidence collected in this investigation was inconclusive due to insufficient data in this area.

> *H5:   Recreational boating accident risks are more frequent in areas where the boat type is dominated by powerboat and jetski watercraft.*

No evidence, either bivariate or multivariate, was found in support of a change in recreational boating accident risk related to average engine size.

> *H6:*   *Recreational boating accident risks are more frequent in areas where the average boat length is greater and less in areas where the average boat length is smaller.*

> *H7:*   *Recreational boating accident risks are higher in areas with greater boat type diversity and lower in areas where there is minimal boat type diversity.*

The empirical evidence from this research supports both the H6 and H7 hypotheses. From a technology perspective, both boat type and boat length were found to be strong indicators with respect to boating accidents. From Chapter 5, bivariate relations strongly indicate that powerboats and jetskis dominate the boat categories involved in an accident. At the Tampa and Sandusky sites, powerboats represent 42.4% and 76.8% respectively of all boating accidents observed. By contrast, jetskis represent 52.1% (Tampa) and 16.8% (Sandusky) respectively of all boating accidents observed. Collectively and generally, these two boat types constitute 94.5% and 93.6% of all boating accidents at the Tampa and Sandusky research sites, respectively. This dominance partially explains why the insertion of this variable within the NB2 Recreational Boating Accident model so dominated the model that other relevant indicators were rendered insignificant. So while the evidence collected in this research study supports the inclusion of boat type as a predictor, it does not contribute to effective model construction. Tthe evidence clearly suggests that boat length is a significant predictor of boating accidents with an estimator that suggests that with each

increase in the number of boats within the 16-26 foot range operating on-the-water that the rate of boating accidents will increase by 1.3%.

> *H8:* *Recreational boating accident risks are more frequent in areas where a majority of the boaters do not practice safe navigation practices, e.g., relying on basic* GPS.

Although a concerted effort was made to collect data appropriate to this question, it was found that the data collected was unreliable and thus it was not used as part of this research.

Implications for Regulation and Policy

As observed through both bivariate and multivariate statistics, several common underlying themes seem to dominate recreational boating accidents. The first and foremost is a human-technological theme. This theme is expressed in terms of boat type, operator gender, boat operator experience/training, boat speed, and boat length. Characterized in terms of an increase in accident risk, that risk is increased by 16-25 foot power boats with male operators who have little to no boat operations experience/education that are operating at relatively fast speeds. This is not a surprising outcome and is echoed in governmental accident profiles. However, if one examines Investigating Officer Observed Accident Causality, it is easy to see that while dominant, the human-technological theme explains only part of the spatial variation in boating accidents. This is reflected in typical accident causes such as collision with a fixed object, striking a submerged object, collision with a floating object, grounding, and flooding/swamping. These categories represent approximately 75% of the human-technological accident count. These characteristics are represented in the NB2

Recreational Boat Accident Model as surrogates like boat density, waterway channel

characteristics, water depth, and the interaction of waterway channels and water depth.

In terms of increased accident risk, these characteristics would be manifested as

higher boat densities operating in shallow depth and highly channelized waterways (or

alternatively deeper water in a single channel). This environmental theme is thus

important in terms of its contribution to the literature because it is a recreational boating

accident characteristic that has largely been ignored. Lastly, the importance of the

temporal terms should not be overlooked. Both bivariate and multivariate statistics

indicate that recreational boating accidents are more probable during peak or in-

season months and on weekends at particular times of the day. Given the need to

design systems for peak activity, these traits are essential to planning for efficiency.

With respect to government policy, the findings related to the human-technology

theme highlight two potential areas of governmental regulatory or policy control. These

areas relate to boat operator experience/training and boat operating speeds. Unlike the

automotive highway system, there are few controls for safe boating operations beyond

USCG carriage requirements, age restrictions mostly coupled with lifejacket wear, and

boating under the influence (BUI). However, this study illustrates that there are parallels

between highway and marine transportation accidents. The strength of the boat

operator's level of experience and training suggest one area for policy and regulatory

consideration. As illustrated by the findings in this study, the accident risk associated

with boat operators who have little to no boating experience/education is significantly

higher than those boat operators who do have higher boating experience/education

levels. As discussed earlier, this is an important characteristic for consideration as the

impact is not simply limited to the subset of boaters who have comparably low boating experience/education levels but impacts all boaters within their area of operation including those with higher levels of experience/education. One means of normalizing boat operators with respect to minimal competencies, would be to phase in a basic skills requirement similar to those required to operate automotive vehicles on public roads. Another area of policy and regulatory consideration related to the findings in this study relate to boat speeds. Again, boat speed was another strong indicator related to boating accidents. However, beyond the minimal use of "no wake zones," there are few regulations for boat speed. Again, using the analogy of automotive transportation, there should/could be posted speed limits in congested marine environments, waterways with multiple intersecting channels, and areas with low water depth. In contrast, broader open water areas could be unregulated speed zones. The automotive analogy would be consistent with higher speeds permitted on freeways but regulated lower speeds in urbanized areas. This research suggests that boating accidents could be further reduced by studying the impact of boat speed controls in selected environments using policy or regulatory mechanisms.

The three significant environmental variables related to the increased risk of boating accidents are: boat density, the presence of waterway channels, and water depth. Boat density and waterway channels both suggest environmental mechanisms that concentrate the number of boats operating within an area. The findings in this study suggest that as the number of boats operating within an area increases, the risk of boating accidents rises as well. Again, using the highway vs urban automotive analogy, speed controls may be the best available regulatory or policy control.

182

Similarly, the findings that shallow water depth is a common indicator of boating accidents suggests that fewer collisions with a fixed, submerged, or floating object, grounding, and flooding/swamping could be reduced by lower boat operating speeds or improved waterway markings/guides/signage.

In summary, these observations yield eight regulatory and policy suggestions for future consideration by state and federal government authorities:

1. Deployment of regulatory resources should account for significant spatial and temporal variation in boating accidents

2. The location of regulatory facilities should reflect the location of boating accident hot spots

3. Consideration for Federal Boat Operator Licensing legislation

4. Consideration of speed zones in high density, high risk boating areas

5. Improved signage in areas of restricted navigation & low water depth

6. Significant rationale that "cookie cutter policies" might be inappropriate given the magnitude of the Site variable (regionality)

7. Boating education should stress the need for greater boat separation in high density areas

8. Real time surveillance of high risk areas with communications capacity

Limitations of Current Research

There are regional differences between the two sites (Tampa and Sandusky) that can only be explained by the characteristics unique to those areas and how those regional characteristics bear upon the technologic, human, and environmental factors. For example, ethnicity, gender and education levels from the on-the-water survey were

not found to be significant in the final model sets. These demographic data were broadly available but probably not in a way that could be matched with the final count regression models. These unique interactions remain a course for future study. A second limitation is represented by the 2-year study period used to define a typical boating environment, i.e., where observed boating accidents occurred. This temporal limitation may not provide an adequate representation. Within this project, this limitation on accident data is also manifested as 2 years of on-the-water observations and satellite imagery (in order to match the accident data). Correspondingly, 7 to 10 years of BARD (boating accident) data with a relatively high accuracy are available. So, expanding the survey and imagery data is a possible next research step. Expanded data sets could enable consideration of additional multi-level interactive effects between the human, technological, and environmental influences. Lastly, other than errors introduced due to significant human data collection methods as well as analysis, satellite imagery degradation in some images due to cloud cover and sun glint issues could have damaged boat density and related on-the-water counts. Although it is believed that any such degradation is minimal, the potential for such degradation is not zero, thus it should be considered as a limit to study accuracy.

Potential for Improved Recreational Boat Accident Models

Although the simple regression models constructed by some recreational boating accident researchers appear to illustrate conventional wisdom, probable specification bias created by the omission of important explanatory variables continues to plague this field of study. In adition, the default observational unit has typically been the state-level, which is inappropriate for understanding such incidents. Findings in the literature have

184

been mixed at least in part because important risk-related explanatory variables have not yet been considered simultaneously and at the appropriate level of observation. Furthermore, this limited literature reveals that the research to date has confined its focus to temporal aspects with significantly less attention given to geographical (spatially explicit) variables critical to recreational boating safety. This suggests that close examination of the spatial distributions associated with recreational boating accidents would continue to advance the body of knowledge and discovery associated with the space/time patterns of recreational boating accidents. In that respect, this dissertation, while innovative, is just a beginning. Another area that needs to be more fully explored is the influence of spatial autocorrelation within the boating accident model presented in this investigation.

Several fundamental questions related to recreational boating accidents remain unaddressed; e.g., the volume, spacing, and timing of boat accidents, the impact of traffic congestion or density in spatially constrained environments, variations in speed and direction, the influence of navigation tools on boat/operator characteristics, and the like. However, similar to previous research and its focus on human and technological factors, isolated consideration of environmental influences is simply too restrictive to capture the complexity of the recreational boating accident domain. Conjoint consideration of environmental factors (e.g., waterway traffic and navigation channel characteristics, visibility, wind, wave current, boat location, and day of week/time of year) coupled with human factors (e.g., age, gender, ethnicity, boat operator education, and boat operator education experience, life jacket wear) and technology factors (e.g., boat propulsion, operating speed, length, type, onboard navigation tools) is needed to

185

better understand the potential relationships and interactive effects. Hierarchical modeling that can accommodate interactive effects is a clear path for further model development and refinement.

The increasing availability of high quality, high resolution satellite imagery and availability of accurate representative on-the-water survey data captured through the use of boating law enforcement citation and warning data would minimize the need for data captured through vessel safety check (VSC) observations. The combination of increased satellite imagery covering a broader temporal period coupled with matching on-the-water citation/warning and BARD data would enable the potential for observations for decades rather than a two-year temporal series. Like many fields, this is a call and opportunity for "big data" in order to reveal key patterns and correlations.

In summary, these observations yield seven areas of potential future research that are illustrated as follows:

1. Increase the number of study sites and data capture

2. Improve on the on-the-water sampling methods

3. Increase spatial/temporal coverage of satellite imagery

4. Increase efficiency and effectiveness in capturing all dimension (human, technological, and environmental) conditions at the time of the boating accident

5. More thoroughly and rigorously examine of interactive effects

6. Investigate the potential of hierarchical (multi-level interactive) modeling approaches permitting more complex/integrated models

7. Improve control of spatial and temporal autocorrelation

Conclusions

The research reported in this dissertation is intended to offer a first glimpse into the complex nature of recreational boating accidents. This goal is made all the more challenging to achieve with relatively small samples (two research sites with spatial constraints of approximately 100 km^2 each), depicting relatively rare accident events. Since the analysis is based on actual accident counts and the number of accidents in any one unit of observation (a grid cell in this dissertation) is typically small, changes in the mean counts are small in terms of the impact of explanatory variables in the model. On the other hand, small percent changes have a very limited effect on the probability of an accident within a grid cell. For these reasons, the accident influences and their impact need to be carefully considered with respect to policy and regulatory application. In this regard, it would be easy to over-react to the statistical significance of the factors discovered.

Secondly, in some statistical circles, modeling might have stopped with the Bootstrap Forest (30 variables and a generalizable R^2 of 0.52). The Bootstrap Forest technique is a predictive modeling output just as is the case with ordinary least squares (OLS) or negative binominal (NB2) regression. However, relying simply on an automated bootstrap technique yields interest in the predictive end related to the model rather than thoughtful consideration of the individual causal elements within a final model. So, variable selection and reduction using the NB2 technique were not only preferred but more logical (including representation from each of the key dimensions of causality) even though its use slightly degraded the resulting generalizable R^2 to 0.46.

The debate surrounding the best sources of on-the-water data, the most influential boating accident variables, the best way to control for the influence of low frequency regional effects (noise), and the most appropriate analytical technique (whether Bootstrap Forest, NB2, or another regression method) remain unclear. However, aspects of this research do offer the potential for continuing review and new observational approaches, some of which are already being contemplated by this author in advancing this research to the next level. What this investigation does show is that boating accidents are a function of a comprehensive domain controlled by human, technological, and environmental factors that must be collectively considered to specify a model that appropriately and statistically explains the space-time distributional qualities of recreational boating accident risks. Furthermore, this boating (marine transportation) accident domain in many respects has been found to parallel the automotive (highway transportation) accident domain with the primary difference only being the number of accidents within a given area, i.e., boating accidents are relatively rare. This same analogy follows with respect to boating accidents, i.e., density is a key driver, with the primary difference being the large number of automobiles on U.S. highways when compared to the number of boats on U.S. waterways.

The broader contribution of this research derives from the specification of a space-time statistical model that likely avoids the large specification errors of prior research when key variables are omitted, improves our understanding of boating accidents, and enables the design of more effective management strategies to reduce the frequency of recreational boating accidents. The level of methodological sophistication embedded in this quest, using data fusion techniques and negative

binomial regression, is much higher than has been previously attempted within this body of literature. More importantly, this investigation has advanced the science of boating accident research by clearly linking boat density (not previously explored within the literature), boat operator experience (including the associated influence of boat operator education), boat speed, boat length, and waterway variation (especially as pertaining to channelized waterways and water depth) as key predictors of boating accidents. This field of transportation accident research, which largely began during the 1970s, has received only limited attention within the literature since that time. This investigation offers an opportunity for not only a reinvigoration within this area of research but a new foundation from which future research can be based.

So what have we learned? First, this study definitely demonstrates that like other forms of transportation, e.g., automotive transportation, marine (recreational boat), transportation accidents are a function of a highly integrated space-time domain that includes human, technological, and environmental, influences. While there can be no doubt that the human dimension is the predominant recreational boating accident influence (i.e., if there were no humans in boats on the water, there would be no boating accidents), technological and environmental influences are shown to be similarly influential. With this in mind and with respect to the findings from this study, some general observations suggest topics that are appropriate for continued research and exploration.

The first pattern involves the human dimension. In contrast to public and government suspicions, drugs and alcohol were not the major drivers of boating accidents within the two research sites. This is not to suggest that drug and alcohol use

are unimportant nor that boating accidents resulting from drugs or alcohol are not horrific. However, within the constraints of this research, neither were found to be significant contributors (to the location of accidents).

On the other hand, there have been boating studies where the perceived influence of drugs and alcohol was in fact greater than what was observed in their data. In that light, drug and/or alcohol use by boat operators (BUIs: boating under the influence) continues to be a primary focus of boating law enforcement officers whose interest is in ensuring public safety. The findings in this study that these specific variables are not significant in the boating accident model may be a function of increased law enforcement, local policies, and effective public awareness campaigns. It may not be that accidents are not caused by drugs and alcohol, but instead their impact has been significantly lessened due to the policies and vigilance of government agencies and the boating law enforcement officers in these specific regions.

From a governmental policy perspective with specific consideration to the constraints of this investigation, these findings suggest that a more effective boating law enforcement distribution might be based on other human factors. Specifically, those factors are directly related to boat operation speeds (especially in confined spaces such as in-shore areas and harbor areas) and boat operator experience/education. Of interest and as noted above, both of these elevated boating accident risk factors could be reduced through the implementation of policy and regulatory guidelines similar to the requirements for automotive transportation, i.e., the implementation of generalized boat speed regulations in shallow water and navigationally constrained environments and

mandatory boat operator education requirements implying boat operator certification rather than boat operator licensing.

With respect to technologic factors, the most significant efficiencies with respect to the allocation and distribution of boating law enforcement officers, this study suggests the greatest return on investment occurs when more closely monitoring the activities of smaller boats in the 15-26 foot range. Although profiling boat operators may not be politically correct, an increased awareness of and attention within this size group is suggested. More specifically, when the influence of technological factors on boating accident risk is coupled with environmental factors, i.e., boat density, navigational waterway complexity (channels), and water depth, this return on investment in boating officer deployment is further magnified. Again, this research illustrates that the aforementioned combination of technological-environmental boating accident factors are collectively as important as human factors. Therefore, one take-away from this research is the observation that the public and government focus on human factors alone should be set aside in favor of a more broadly defined concern that incorporates human, technological, and environmental boating accident factors more effectively.

Lastly, while temporal (subset of environmental) factors were shown to be significant from a quantitative modeling perspective, the evidence suggested by this research indicates that boating law enforcement resource deployments currently based on seasonality with an increased emphasis on weekend monitoring, continues to have the greatest return on investment. If in no other way, this finding reinforces the effectiveness of current governmental operational procedures pertaining to increased deployment of resources during these more narrowly defined temporal periods. In

191

addition, given the highly clustered spatial distributions that characterize boating accidents, regulatory agencies should pay very careful attention to the location of shore facilities and customize their location to the activity patterns inventoried.

It is acknowledged that regional variation (Site) is significant with respect to the areas under study and that these local settings, due to their inherent uniqueness, will differentially influence boating accident risk. However, if effectively captured, these regional effects can also be used to support increased modeling efficiency as a supplement to the proposed boating accident risk model developed in this dissertation.

In addition to the identification of boat density, boat speed, and operator experience/education as significant boating accident risk contributors, this study also reveals the benefit of multivariate, negative binomial statistical modeling approach that is based on count data as opposed to simpler approaches. In fact, the use of the technology (handheld Trimble GPS observation recorders) used to capture on-the-water observations as well as satellite imagery ultimately fused into a single ArcGIS database with accident data proved to be an effective way to capture the needed count data. An external evaluation provided by Ms. Tammy Terry, Chair of the National Association of State Boating Law Administrators, ERAC (Engineering, Reporting & Analysis Committee) reported that the boating law enforcement officers who participated in this study believe that the availability of advanced technology enabled the effectiveness and accuracy of their data collection activities.

Currently, one of the greatest constraints on boating accident risk reduction as well as one of the greatest government (federal and state) concerns is related to the accuracy and quality of data capture. This dissertation supports this governmental

192

concern in two ways. First, it illustrates how the application of relatively low cost technology can significantly improve data (capture) quality and accuracy. The future application of electronic data capture not only eliminates the redundancy and increased error potential of converting paper based data into an electronic form, it also permits that data to be immediately beamed via an appropriate communications medium (cell phone, Wi-Fi, or VHF radio) to a shore-based data collection point. The potential for real-time monitoring yields a variety of new applications. Additional innovation is currently under design with both a mobile application and technology interface that would permit the application of the Trimble-based data capture tools to be ported to an IOS or Android platform permitting further data capture technology acquisition, cost reductions, and increased device screen sizes. The increased device screen size is especially important as reported by the boating law enforcement officers participating in this study who frequently indicated difficulty in reading the small Trimble screens in bright sunlight while wearing sunglasses. The current prototypes under development involve an armored Apple iPad effectively minimize this data capture challenge. In addition, this investigation found that there is an increasing public resistance to on-the-water vessel safety stops, making this form of on-the-water data capture less effective for broader implementation. However, this understanding and creative thinking offer the potential for an alternative on-the-water data capture strategy. That strategy would involve the use of a barcode tag affixed to each registered vessel that contains basic demographic information both about the registered boat operator and the specifications of the boat being observed. This barcode tag could readily be scanned by the observing officer using a mobile device in much the same way that retail items can be

193

scanned for later analysis. Most of the essential temporal and geographic information could be readily obtained via GPS using the same mobile device. Such a technique would significantly increase the accuracy and speed with which data capture of the boating stop could be obtained leaving the boating law enforcement officer with few required supplemental questions to ask. This would permit an increased focus on the boat operator and occupants by boating law enforcement officers. Secondly, the findings from boating accident model suggest that an equally effective way to obtain the needed on-the-water data would be to capture and analyze citation and warning data as issued during official boating stops for perceived rules or regulatory reasons. Based on the distribution of VSC (vessel safety check) based boating safety stops and given increasing public resistance to this data collection method, use of citation and warning data would not only create data that is more randomly distributed but less likely to result in public resistance. The findings in this research study suggests that the most significant variables that should be further explored in future iterations of this research could be effectively captured using citation and warning data as opposed to the VSC data.

The findings from this research illustrate that the field of recreational boating accident risk research is wide open with a history that only spans approximately forty years. More importantly, it illustrates that a broader perspective to better understand boating accidents is needed. The direct benefit of this research is its illustration that this understanding is both accessible, practical, and cost effective through the use of electronic accident data, on-the-water surveys, and satellite data fused with geospatial tools.

REFERENCES

Al-Harbi, M., Yassin, M., & Shams, M. (2012). Stochastic modeling of the impact of meteorological conditions on road traffic accidents. *Stochastic Environmental Research and Risk Assessment, 26*(5), 739-750.

Alparone, L., Wald, L., Chanussot, J., Thomas, C., Gamba, P., & Mann Bruce, L. (2007). Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest. *Geoscience and Remote Sensing, IEEE Transactions, 45*(10), 3012-3021.

Alsop, J., & Langley, J. (2001). Under-reporting of motor vehicle traffic crash victims in New Zealand. *Accident Analysis & Prevention, 33*(3), 353-359.

Amoros, E., Martin, J., & Laumon, B. (2006). Under-reporting of road crash casualties in France. *Accident Analysis & Prevention, 38*(4), 627-635.

Amro, I., Mateos, J., Vega, M., Molina, R., & Katsaggelos, A. (2011). A survey of classical methods and new trends in pansharpening of multispectral images. *EURASIP Journal on Advances in Signal Processing,* (1), 1-22.

Andrey, J., Hambly, D., Mills, B., & Afrin, S. (2013). Insights into driver adaptation to inclement weather in Canada. *Journal of Transport Geography, 28*(0), 192-203.

Andrey, J., Mills, B., Leahy, M., & Suggett, J. (2003). Weather as a chronic hazard for road transportation in Canadian cities. *Natural Hazards, 28*(2-3), 319-343.

Anselin, L. (2002). Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics, 27*(3), 247-267.

Arditi, D., Lee, D., & Polat, G. (2007). Fatal accidents in nighttime vs. daytime highway construction work zones. *Journal of Safety Research, 38*(4), 399-405.

Ashby, K., Cassell, E., & Congiu, M. (2013). *Follow up study of hospital treated recreational boating injury*. Monash University Accident Research Centre 2008 [cited 10/25/2013 2013]. Available from http://www.transportsafety.vic.gov.au/__data/assets/pdf_file/0008/41975/Follow-up-study-of-hospital-treated-recreational-boating-injury--the-longer-term-effects-of-injury-May-2008.pdf

Atkins, D. C., & Gallop, R. J. (2007). Rethinking how family researchers model infrequent outcomes: A tutorial on count regression and zero-inflated models. *Journal of Family Psychology, 21*(4), 726-735.

Balaguer, P., Diedrich, A., Sardá, R., Fuster, M., Cañellas, B., & Tintoré, J. (2011). Spatial analysis of recreational boating as a first key step for marine spatial planning in Mallorca (Balearic Islands, Spain). *Ocean & Coastal Management, 54*(3), 241-249.

Bell, N. S., Howland, J., Mangione, T. W. , & Senier, L. (2000). Boater training, drinking and boating, and other unsafe boating practices. *Journal of Drug Education, 30*(4), 467-482.

Bergel-Hayat, R., Debbarh, M., Antoniou, C., & Yannis, G. (2013). Explaining the road accident risk: Weather effects. *Accident Analysis & Prevention,* (0).

Black, W. R. (2003). *Transportation: A geographical analysis*: Guilford Press.

Borowsky, A., Shinar, D., & Oron-Gilad, T. (2010). Age, skill, and hazard perception in driving. *Accident Analysis & Prevention, 42*(4), 1240-1249.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

196

Bristow, R. S., & Bennett, W. D. (1995). Boating opportunities: A geographical analysis of travel patterns and motivations. *Notes*.

Burger, J., & Leonard, J. (2000). Conflict resolution in coastal waters: The case of personal watercraft. *Marine Policy, 24*(1), 61-67.

Carl, P., & Timmons, J. (2003). A regulatory history. Paper read at Marine Safety Council, at Washington, DC.

Cassell, E., & Congiu, M. (2007). Review of the literature on the epidemiology and prevention of unintentional submersion and non-submersion injury in recreational boating: Melbourne, Australia: Marine Safety Victoria. Retrieved from: http://www.marinesafety.vic.gov.au/doi/doielectnsf/2a6bd98dee287482ca256915 001cff0c/2cc27bcc69173ee7 ca2572d40020a451/$ FILE/Submersion% 20injury

Chauvin, C., Lardjane, S., Morel, G., Clostermann, J., & Langard, B. (2013). Human and organisational factors in maritime accidents: Analysis of collisions at sea using the HFACS. *Accident Analysis & Prevention, 59*(0), 26-37.

Chen, S., Cai, C., & Wolshon, B. (2009). From normal operation to evacuation: Single-vehicle safety under adverse weather, topographic, and operational conditions. *Natural Hazards Review, 10*(2), 68-76.

Chen, S. R., & Cai, C.S. (2004). Accident assessment of vehicles on long-span bridges in windy environments. *Journal of Wind Engineering and Industrial Aerodynamics, 92*(12), 991-1024.

Choi, M. (2006). A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter. *Geoscience and Remote Sensing, IEEE Transactions, 44*(6), 1672-1682.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences.* Routledge.

Cools, M., Moons, E., & Wets, G. (2010). Assessing the impact of weather on traffic intensity. *Weather, Climate, and Society, 2*(1), 60-68.

Devanney, J. (2008). Uses and abuses of ship casualty data. *The Center for Tankship Excellence.* Retrieved from: http://www.c4tx.org

Doll, T. J., & Stiehl, C. C. (1979). Accident recovery model for estimating benefits of recreational boating safety programs. *Journal of Safety Research, 11*(2), 72-90.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Vol. 57: CRC press.

El-Basyouny, K., & Sayed, T. (2006). Comparison of two negative binomial regression techniques in developing accident prediction models. *Transportation Research Record: Journal of the Transportation Research Board, 1950*(-1), 9-16.

Erdogan, S., Yilmaz, I., Baybura, T., & Gullu, M. (2008). Geographical information systems aided traffic accident analysis system case study: City of Afyonkarahisar. *Accident Analysis & Prevention, 40*(1), 174-181.

Factor, R., Mahalel, D., & Yair, G. (2008). Inter-group differences in road-traffic crash involvement. *Accident Analysis & Prevention, 40*(6), 2000-2007.

Factor, R., Mahalel, D., & Yair, G. (2007). The social accident: A theoretical model and a research agenda for studying the influence of social and cultural characteristics on motor vehicle accidents. *Accident Analysis & Prevention, 39*(5), 914-921.

Fang, R. (2013). Zero-Inflated Negative Binomial (Zinb) Regression Model for over-
dispersed count data with excess zeros and repeated measures, an application
to human microbiota sequence data, University of Colorado.

FEDERAL BOATING ACT OF 1958: Pub. L. 85-911, Sept. 2, 1958, 72 Stat. 1754.
1958. edited by U. S. C. Guard. Washington DC: Congress.

Fekedulegn, D., Andrew, M., Violanti, J., Hartley, T., Charles, L., & Burchfiel, C. (2010).
Comparison of statistical approaches to evaluate factors associated with
Metabolic Syndrome. *The Journal of Clinical Hypertension, 12*(5), 365-373.

Fonseca, L., Garcia, M., Namikawa, L. M., & Castejon, E. F. (2009). Digital image
processing in remote sensing. Paper read at Computer Graphics and Image
Processing (SIBGRAPI TUTORIALS), 2009 Tutorials of the XXII Brazilian
Symposium.

Gabe, T. M., & Hite, D. (2003). The effects of boating safety regulations. *Coastal
Management, 31*(3), 247-254.

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and
rates: Poisson, overdispersed Poisson, and negative binomial models.
Psychological Bulletin, 118(3), 392-404.

Glover, E. D., Lane, S., & Wang, M. Q. (1995). Relationship of alcohol consumption and
recreational boating in Beaufort County, North Carolina. *Journal of Drug
Education, 25*(2), 149.

Gonzales-Barron, U., Kerr, M., Sheridan, J. J., & Butler, F. (2010). Count data distributions and their zero-modified equivalents as a framework for modelling microbial data with a relatively high occurrence of zero counts. *International Journal of Food Microbiology, 136*(3), 268-277.

Gray, D. L., Canessa, R. R., Keller, C. P., Dearden, P., & Rollins, R. B. (2011). Spatial characterization of marine recreational boating: Exploring the use of an on-the-water questionnaire for a case study in the Pacific Northwest. *Marine Policy, 35*(3), 286-298.

Greene, W. H. (1994). *Accounting for excess zeros and sample selection in Poisson and negative binomial regression models*.

Grün, B., & Leisch, F. (2007). Fitting finite mixtures of generalized linear regressions in R. *Computational Statistics & Data Analysis, 51*(11), 5247-5252.

Hänninen, M. (2008). Analysis of human and organizational factors in Marine Traffic Risk Modeling. 2010 (2/24/2010): This report is a literature review that describes the existing techniques for modeling human factors as a cause in marine traffic accidents. The report serves as a background for marine traffic accident risk assessment in the Gulf of Finland.

Harati-Mokhtari, A., Wall, A., Brooks, P., & Wang, J. (2007). Automatic Identification System (AIS): Data Reliability and Human Error Implications. *The Journal of Navigation, 60*(03), 373-389.

Harris, S. (1990). The real number of road traffic accident casualties in the Netherlands: a year-long survey. *Accident Analysis & Prevention, 22*(4), 371-378.

Hassel, M., Asbjørnslett, B. E., & Hole, L. P. (2011). Underreporting of maritime accidents to vessel accident databases. *Accident Analysis & Prevention, 43*(6), 2053-2063.

Helmy, A., Nasr, A., & El-Taweel, G. (2010). Assessment and evaluation of different data fusion techniques. *International Journal of Computers, 4*(4), 107-115.

Hilbe, J. M. (2011). *Negative binomial regression* (2nd ed.). Cambridge University Press.

Hoedt, J., Timmons, J., & Marmo, A. (2003). IN THE BEGINNING: THE FEDERAL BOAT SAFETY ACT OF 1971. Paper read at Marine Safety Council, at Washington, DC.

Hoef, J., Ver, M., & Boveng, P. L. (2007). Quasi-Poisson vs. Negative Binomial Regression: How should we model overdispersed count data? *Ecology, 88*(11), 2766-2772.

Hollnagel, E. (2004). *Barriers and accident prevention*: Ashgate Pub Ltd.

Hollnagel, E. (2008). The changing nature of risks. *Ecole des Mines de Paris, Sophia Antipolis, France*.

Hovden, J., Størseth, F., & Tinmannsvik, R. K. (2011). Multilevel learning from accidents – Case studies in transport. *Safety Science, 49*(1), 98-105.

Ismail, N., & Zamani, H. (2013). Estimation of Claim Count Data using Negative Binomial, Generalized Poisson, Zero-Inflated Negative Binomial and Zero-Inflated Generalized Poisson Regression Models. Paper read at Casualty Actuarial Society E-Forum, Spring 2013.

Jaakson, R. (1989). Recreation boating spatial patterns: Theory and management. *Leisure Sciences, 11*(2), 85-98.

Jin, D., Kite-Powell, H., & Talley, W. (2001). The safety of commercial fishing: Determinants of vessel total losses and injuries. *Journal of Safety Research, 32*(2), 209-228.

Johansson, Ö., Wanvik, P. O., & Elvik, R. (2009). A new method for assessing the risk of accident associated with darkness. *Accident Analysis & Prevention, 41*(4), 809-815.

Kalpoma, K. A., & Kudoh, J. I. (2007). Image fusion processing for IKONOS 1-m color imagery. *Geoscience and Remote Sensing, IEEE Transactions, 45*(10), 3075-3086.

Khattak, A. J., Kantor, P., & Council, F. M. (1998). Role of adverse weather in key crash types on limited-access: Roadways implications for advanced weather systems. *Transportation Research Record: Journal of the Transportation Research Board, 1621*(1), 10-19.

Klimberg, R., & McCullough, B. D. (2013). *Fundamentals of predictive analytics with JMP*. SAS Institute.

Kockelman, K. M. (1998). Changes in flow-density relationship due to environmental, vehicle, and driver characteristics. *Transportation Research Record: Journal of the Transportation Research Board, 1644*(1), 47-56.

Kujala, P., Hänninen, M., Arola, T., & Ylitalo, J. (2009). Analysis of the marine traffic safety in the Gulf of Finland. *Reliability Engineering & System Safety, 94*(8), 1349-1357.

Lentnek, B., & Doren, V. A. N. (1969). Spatial behavior in recreational boating. *Journal of Leisure Research, 1*(2), 103-124.

Lentnek, B., Van Doren, C., S., Tiall, J. R. (1970). Spatial behavior in recreational boating. *Journal of Leisure Research, 1*(2), 103-124.

Leveson, N., Dulac, N., Marais, K., & Carroll, J. (2009). Moving beyond normal accidents and high reliability organizations: A systems approach to safety in complex systems. *Organization Studies, 30*(2-3), 227-249.

Li, H., Raeside, R., Chen, T., & McQuaid, R. W. (2012). Population ageing, gender and the transportation system. *Research in Transportation Economics, 34*(1), 39-47.

Loeb, P., Talley, W., & Zlatoper, T. (1994). *Causes and determinants of transportation accidents: An analysis by mode*. Westport, CT: Quorum Books.

Loeb, P. D., & Gilad, B. (1984). The efficacy and cost-effectiveness of vehicle inspection - A state specific analysis using time-series data. *Journal of Transport Economics and Policy, 18*(2), 145-164.

Loeb, P. D., & Gilad, B. (1984). The efficacy and cost-effectiveness of vehicle inspection: a state specific analysis using time series data. *Journal of Transport Economics and Policy,* 145-164.

Loeb, P. D. (1987). The determinants of automobile fatalities: With special consideration to policy variables. *Journal of Transport Economics and Policy, 21*(3), 279-287.

Mahony, E., & Stynes, D. (1995). Recreational boating carrying capacity: A framework for managing inland lakes. *Michigan State University, Department of Park, Recreation, and Tourism Resources, East Landing, Michigan*.

Marais, K., Dulac, N., & Leveson, N. (2007). Beyond normal accidents and high reliability organizations: the need for an alternative approach to safety in complex systems.

Martz, E. (2014). *Enough Is enough! Handling multicollinearity in regression analysis*. Minitab, INc 2013 [cited 3 October 2014]. Available from http://blog.minitab.com/blog/understanding-statistics/handling-multicollinearity-in-regression-analysis

McCarthy, P., & Talley, W. K. (1999). Evidence on risk compensation and safety behaviour. *Economics Letters, 62*(1), 91-96.

McKnight, A. J., Becker, W. W., Pettit, A. J., & McKnight, A. S. (2007). Human error in recreational boating. *Accident Analysis & Prevention, 39*(2), 398-405.

Mergia, W. Y., Eustace, D., Chimba, D., & Qumsiyeh, M. (2013). Exploring factors contributing to injury severity at freeway merging and diverging locations in Ohio. *Accident Analysis & Prevention,* 55(0), 202-210.

Miller, J. R., & Pikora, T. J. (2008). Alcohol consumption among recreational boaters: Factors for intervention. *Accident Analysis & Prevention, 40*(2), 496-501.

Morel, J. G., & Neerchal, N. (2012). *Overdispersion models in SAS.* SAS Institute.

Motorboat Act of 1910: Public Law 61-201 36 Stat.462. 1910. edited by U. S. C. Guard.

Mullai, A., & Paulsson, U. (2011). A grounded theory model for analysis of marine accidents. *Accident Analysis & Prevention, 43*(4), 1590-1603.

Niclasen, B. A., Simonsen, K., & Magnusson, A. K. (2010). Wave forecasts and small-vessel safety: A review of operational warning parameters. *Marine Structures, 23*(1), 1-21.

Nikolakopoulos, K. G. (2004). Comparison of four different fusion techniques for IKONOS data. Paper read at Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International.

Nikolakopoulos, K. G. (2008). Comparison of nine fusion techniques for very high resolution data. *Photogrammetric Engineering and Remote Sensing, 74*(5), 647.

O'Connor, P. J., O'Connor, N. (2005). Causes and prevention of boating fatalities. *Accident Analysis & Prevention, 37*(4), 689-698.

O'Connor, P. (2013). *National assessment of boating fatalities in Australia 1999–2004.* Northern Territory Government of Australia 2008 [cited 10/25/2013 2013]. Available from http://libraries.nt.gov.au:8080/lib/item;jsessionid=44BC1E1D90A04A3EF4CD4B6 23CCE82EC?id=chamo:721988

Otazu, X., Gonzalez-Audicana, M., Fors, O., & Nunez, J. (2005). Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods. *Geoscience and Remote Sensing, IEEE Transactions, 43*(10), 2376-2385.

PCI Geomatics. (2014). *Pansharpening Data in Geomatica.* PCI Geomatics 2014 [cited 1/18/2014 2014]. Available from http://www.pcigeomatics.com/pdf/Pansharp.pdf

Pelot, R., & Plummer, L. (2008). Spatial analysis of traffic and risks in the coastal zone. *Journal of Coastal Conservation, 11*(4), 201-207.

Perdue, R. R. (1987). The influence of awareness on spatial behavior in recreational boating. *Leisure Sciences, 9*(1), 15-25.

Perrow, C. (1984). N*ormal accidents.* New York.

Perrow, C. (1999). *Normal accidents: Living with high risk technologies.* Princeton, NJ: Princeton University Press.

Perrow, C. (1984). *Normal accidents : Living with high-risk technologies*. New York: Basic Books.

Prasannakumar, V., Vijith, H., Charutha, R., & Geetha, N. (2011). Spatio-temporal clustering of road accidents: GIS based analysis and assessment. *Procedia - Social and Behavioral Sciences, 21*(0), 317-325.

Psarros, G., Skjong, R., & Eide, M. S. (2010). Under-reporting of maritime accidents. *Accident Analysis & Prevention, 42*(2), 619-625.

Qureshi, Z. H. (2007). A review of accident modelling approaches for complex socio-technical systems. In *Proceedings of the twelfth Australian workshop on Safety critical systems and software and safety-related programmable systems*. Adelaide, Australia: Australian Computer Society, Inc.

Rahman, A., & Lownes, N. E. (2012). Analysis of rainfall impacts on platooned vehicle spacing and speed. *Transportation Research Part F: Traffic Psychology and Behaviour, 15*(4), 395-403.

Rasmussen, J., & Svedung, I. (2000). *Proactive risk management in a dynamic society.* Swedish Rescue Services Agency Karlstad.

Rizzo, Capt M., & Hoedt, J. (2010). U.S. Coast Guard Office of Auxiliary and Boating Safety. *Journal of Safety and Security at Sea, 67*(2), 3.

Sawalha, Z., & Sayed, T. (2006). Traffic accident modeling: Some statistical issues. *Canadian journal of civil engineering, 33*(9), 1115-1124.

Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., & Kincses, W. E. (2009). Drivers' misjudgement of vigilance state during prolonged monotonous daytime driving. *Accident Analysis & Prevention, 41*(5), 1087-1093.

Schuhmann, P. W., & Schwabe, K. A. (2004). An analysis of congestion measures and heterogeneous angler preferences in a random utility model of recreational fishing. *Environmental and Resource Economics, 27*(4), 429-450.

Sciortino, S., Vassar, M., Radetsky, M., & Knudson, M. M. (2005). San Francisco pedestrian injury surveillance: Mapping, under-reporting, and injury severity in police and hospital records. *Accident Analysis & Prevention, 37*(6), 1102-1113.

Sewell, W. R. D., Kates, R. W., & Phillips, L. E. (1968). Human response to weather and climate: Geographical contributions. *Geographical Review, 58*(2), 262-280.

Shankar, V., Mannering, F., & Barfield, W. (1995). Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis & Prevention, 27*(3), 371-389.

Shmueli, G. (2010). To explain or to predict? *Statistical Science,* 289-310.

Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *Mis Quarterly, 35*(3), 553-572.

Sidman, C., Grant, F., Swett, R., & Fik, R. (2005). A recreational boating GIS for sustainable Florida waterways. Paper read at Proceedings of the 14th Biennial Coastal Zone Conference, July 17 to 21, 2005, at New Orleans, Louisiana.

Sidman, C. F., & Fik, T. J. (2005). Modeling spatial patterns of recreational boaters: vessel, behavioral, and geographic considerations. *Leisure Sciences, 27*(2), 175-189.

Sidman, C., Fik, T. J., & Flamm, R. (2002). Blueways recreational boating characterization for Charlotte Harbor, Florida. *Final Report to the Florida Marine Research Institute*.

Sidman, C., Fik, T., Swett, R., Sargent, B., & Fann, S. (2005). Estimating land and

    water-side service areas and use potential for boat ramps: A case study of

    Tampa and Sarasota Bays. Florida Sea Grant Publication TP-142. University of

    Florida, Gainesville, Florida.

Sidman, C., & Flamm, R. (2001). A survey of methods for characterizing recreational

    boating in Charlotte Harbor, Florida. *FLSGP-S-01-002, 2001, 48*.

Sivak, M. (2002). How common sense fails us on the road: Contribution of bounded

    rationality to the annual worldwide toll of one million traffic fatalities.

    Transportation Research Part F: Traffic Psychology and Behaviour, 5(4), 259-

    269.

Snæbjörnsson, J. T., Baker, C. J., & Sigbjörnsson, R. (2007). Probabilistic assessment

    of road vehicle safety in windy environments. *Journal of Wind Engineering and*

    Industrial Aerodynamics, 95(9–11), 1445-1462.

Stine, B. (2007, October 3). *Classification and Regression Trees*. University of

    Pennsylvania [cited 2014 October 3]. Available from http://www-

    stat.wharton.upenn.edu/~stine/mich/DM_07.pdf

Swett, R. A., Sidman, C., Fik, T., & Sargent, B. (2009). Developing a spatially enabled

    inventory of recreational boats using vessel registration data. *Coastal*

    Management, 37(5), 405-420.

Tømmervik, H., Johansen, B., Høgda, K. A., & Strann, K. B. (2012). High-resolution

    satellite imagery for detection of tracks and vegetation damage caused by all-

    terrain vehicles (ATVs) in Northern Norway. *Land Degradation & Development,*

    23(1), 43-52.

Treser, C. D., Trusty, M. N., & Yang, P. P. (1997). Personal floatation device usage: Do educational efforts have an impact? *Journal of Public Health Policy, 18*(3), 346-356.

United States Coast Guard. (1988, May 16). Boating Statistics 1987. United States Coast Guard, Office of Boating Safety. Document: COMDTINST P16754.1.

United States Coast Guard. (2013, June 17). *Recreational Boating Statistics* (United States Coast Guard). United States Coast Guard, Office of Boating Safety 2013 [cited cited 2013 June 17]. Available from http://www.uscgboating.org/assets/1/News/2012ReportR2.pdf

USCG Boating Safety Division. (2010). *National RBS Strategic Plan* (2007-2011). United States Coast Guard 2010 [cited 23 Dec 2010]. Available from http://www.uscgboating.org/about/strategic_plan.aspx

U.S. House of Representatives. 2008. Committee on Transportation and Infrastructure. *Honoring the Heritage of the Coast Guard*. Report 110-889. U.S. Government Printing Office. Washington, D.C. 11 p. Available at: http://thomas.loc.gov/cgi-bin/cpquery/?&dbname=cp110&sid=cp1102R81m&refer=&r_n=hr899.110&item=&&&sel=TOC_790&. Date retrieved: 11/20/2014.

Vijayaraj, V., O'Hara, C. G., & Younan, N. H. (2004). Quality analysis of pansharpened images. Paper read at Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International.

Virk, A., & Pikora, T. J. (2011). Developing a tool to measure safe recreational boating practice. *Accident Analysis & Amp; Prevention, 43*(1), 447-450.

Vogt, A., & Bared, J. (1998). Accident models for two-lane rural segments and

    intersections. *Transportation Research Record: Journal of the Transportation*

    Research Board, 1635(1), 18-29.

Wang, W. R. (2000). The effects of state regulations on boating accidents and fatalities.

    Applied Economics Letters, 7(6), 373-378.

Wang, Z., Ziou, D., Armenakis, C., Li, D., & Li, Q. (2005). A comparative analysis of

    image fusion methods. *Geoscience and Remote Sensing, IEEE Transactions,*

    43(6), 1391-1402.

Xu, C., Wang, W., & Liu, P. (2013). Identifying crash-prone traffic conditions under

    different weather on freeways. *Journal of Safety Research, 46*(0), 135-144.

Yagil, D. (1998). Gender and age-related differences in attitudes toward traffic laws and

    traffic violations. *Transportation Research Part F: Traffic Psychology and*

    Behaviour, 1(2), 123-135.

Yamamoto, T., Hashiji, J., & Shankar, V. N. (2008). Underreporting in traffic accident

    data, bias in parameters and the structure of injury severity models. *Accident*

    Analysis & Prevention, 40(4), 1320-1329.

Yip, T. L. (2008). Port traffic risks – A study of accidents in Hong Kong waters.

    Transportation Research Part E: Logistics and Transportation Review, 44(5),

    921-931.

Zaidel, D. M. (1992). A modeling perspective on the culture of driving. *Accident Analysis*

    & Prevention, 24(6), 585-597.

Zhang, Y., & Xiong, Z. (2006). Moving vehicle detection using a single set of Quickbird

    Imagery—An initial study.

Zhang, Y. (2004). Understanding image fusion. *Photogrammetric Engineering and Remote Sensing, 70*(6), 657-661.

Zhang, Y. (2002). A new automatic approach for effectively fusing Landsat 7 as well as IKONOS images. Paper read at Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International.

APPENDIX A: INSTITUTIONAL REVIEW BOARD APPROVAL LETTER

EAST CAROLINA UNIVERSITY
University & Medical Center Institutional Review Board Office
1L-09 Brody Medical Sciences Building• 600 Moye Boulevard • Greenville, NC 27834
Office 252-744-2914 • Fax 252-744-2284 • www.ecu.edu/irb

TO: Ronald Mitchelson, PhD, Dept. of Geography & Planning, ECU—Mailstop 157

FROM: UMCIRB $\mathcal{K}\mathcal{K}$

DATE: October 4, 2010

RE: Expedited Category Research Study

TITLE: "Doctoral Dissertation Research: Spatial Perspectives on Recreational Boating Accidents"

UMCIRB #10-0536

This research study has undergone review and approval using expedited review on 10.1.10. This research study is eligible for review under an expedited category number 7. The Chairperson (or designee) deemed this **grant funded** study **no more than minimal risk** requiring a continuing review in **12 months**. Changes to this approved research may not be initiated without UMCIRB review except when necessary to eliminate an apparent immediate hazard to the participant. All unanticipated problems involving risks to participants and others must be promptly reported to the UMCIRB. The investigator must submit a continuing review/closure application to the UMCIRB prior to the date of study expiration. The investigator must adhere to all reporting requirements for this study.

The above referenced research study has been given approval for the period of **10.1.10** to **9.30.11**. The approval includes the following items:
- Internal Processing Form (dated 9.20.10)
- Grant Application
- Request for Waiver of Informed Consent (9.28.10)
- Boating Safety Stop Report

The Chairperson (or designee) does not have a potential for conflict of interest on this study.

The UMCIRB applies 45 CFR 46, Subparts A-D, to all research reviewed by the UMCIRB regardless of the funding source. 21 CFR 50 and 21 CFR 56 are applied to all research studies under the Food and Drug Administration regulation. The UMCIRB follows applicable International Conference on Harmonisation Good Clinical Practice guidelines.

IRB00000705 East Carolina U IRB #1 (Biomedical) IORG0000418
IRB00003781 East Carolina U IRB #2 (Behavioral/SS) IORG0000418
IRB00004973 East Carolina U IRB #4 (Behavioral/SS Summer) IORG0000418
Version 3-5-07

UMCIRB #10-0536
Page 1 of 1