

THE FLORIDA STATE UNIVERSITY  
COLLEGE OF SOCIAL SCIENCE AND PUBLIC POLICY

QUANTIFYING EXTREME HURRICANE RISK IN THE NORTH ATLANTIC  
AND GULF OF MEXICO

By

JILL TREPANIER

A Dissertation submitted to the  
Department of Geography  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

Degree Awarded:  
Spring Semester, 2012

Jill Trepanier defended this dissertation on March 15, 2012.

The members of the supervisory committee were:

James Elsner  
Professor Directing Dissertation

Joseph Donoghue  
University Representative

Victor Mesev  
Committee Member

Tingting Zhao  
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with the university requirements.

For my husband.

“Two roads diverged in a wood, and I...I took the one less traveled by, and that has  
made all the difference.” –Robert Frost  
Thank you for traveling this road with me.

# ACKNOWLEDGMENTS

Dr. James Elsner, I will never be able to fully extend my gratitude for all you have done for me during my tenure at Florida State. It is impressive how quickly the time passes, but even more impressive how much of your wisdom has been imparted upon me during that time. My only hope to repay you is to offer an inexperienced graduate student the same courtesy you bestowed upon me five years ago, the courtesy of an opportunity. Thank you for giving me that chance.

Dr. Victor Mesev, thank you for believing that I would achieve as highly as I have. I appreciate all of the advice you have given me, as well as all of the laughs. Dr. Tingting Zhao, your insight into the public domain and the importance of understanding societal risk is something that has helped to make this research more well-rounded and applicable to the public. Thank you. Dr. Thomas Jagger, thank you for being my statistical R guru. Your guidance has helped shape me into the researcher I am today. Dr. Joseph Donoghue, thank you for teaching me the importance of collaborative work, as well as the importance of getting your hands dirty. Hal Needham, to you and your coauthors, thank you for the unlimited use of your beautiful surge data set.

Dr. Kelsey Scheitlin and Rob Hodges, my colleagues in the Hurricane Task Force, you have helped me grow so much as a person and as an expert in my field over the last five years. Sharing my graduate experience with you both has provided me a unique opportunity to grow intellectually and have a very fun time doing so. Thank you.

Finally, for all of the members of the Geography Department at Florida State, I will never forget my time in graduate school and will look back on it fondly. The experiences I have had and the friends I have made shaped me into the person I am and will guide me to the person I become. Thank you all for your support and friendship.

# TABLE OF CONTENTS

List of Tables . . . . .	vii
List of Figures . . . . .	ix
Abbreviations . . . . .	xiii
Abstract . . . . .	xiv
<b>1 HURRICANE RISK</b>	<b>1</b>
1.1 Hurricanes . . . . .	3
1.2 Extreme Value Theory . . . . .	4
1.3 Copulas . . . . .	9
1.4 Tessellated Space . . . . .	11
<b>2 DATA SOURCES</b>	<b>14</b>
2.1 Hurricane Data . . . . .	14
2.2 Storm Surge Data . . . . .	17
2.3 Sea Surface Temperature Data . . . . .	19
<b>3 RISK OF EXTREME HURRICANE WINDS IN FLORIDA</b>	<b>21</b>
3.1 Estimation of Florida Wind Risk . . . . .	22
3.2 Return Period Estimates . . . . .	25
3.3 Changes in Frequency and Intensity . . . . .	35
3.4 Florida Wind Risk in Summary . . . . .	37
<b>4 COMBINED WIND AND SURGE RISKS FROM HURRICANES ALONG THE U. S. GULF COAST</b>	<b>41</b>
4.1 Collating Wind and Surge Events . . . . .	42
4.2 Storm Surge Risk . . . . .	43
4.3 Relationship between Wind and Surge . . . . .	48
4.3.1 Bivariate Dependence . . . . .	51
4.4 Combined Risk in Summary . . . . .	56
<b>5 SPATIAL RISK OF HURRICANE WINDS OVER THE GULF OF MEXICO AND NORTH ATLANTIC</b>	<b>59</b>
5.1 Data Collation . . . . .	61
5.2 Estimation of Extreme Value Model . . . . .	62

5.2.1	Visualization of Latent Variables . . . . .	63
5.2.2	Hexagon Size Comparison . . . . .	66
5.3	Spatial Autocorrelation . . . . .	67
5.4	Local Parameter Relationship with SST . . . . .	75
5.5	Spatial Risk Summary . . . . .	79
<b>6</b>	<b>CONCLUDING REMARKS</b>	<b>81</b>
6.1	Study Summaries . . . . .	81
6.2	Societal Risk . . . . .	83
<b>A</b>	<b>APPENDIX</b>	<b>86</b>
A.1	Chapter 4 Code . . . . .	86
A.2	Chapter 5 Code . . . . .	89
	References . . . . .	93
	Biographical Sketch . . . . .	100

# LIST OF TABLES

3.1	Number of hurricanes by wind speed intervals known to have passed within a 100 km radius of the specified city center from 1851–2008. Positions and wind speeds interpolated to 1-hourly values following the procedure outlined in Jagger and Elsner (2006). . . . .	29
3.2	Return period for selected wind speed exceedance values. The return levels are based on the HRC presented above and include data over the period 1851–2008. The return period refers to the expected time between winds of this level blowing somewhere within a 100 km radius of specified cities. The confidence interval is the 90% interval. . . . .	31
3.3	Return period for selected wind speed exceedance values for Miami and Pensacola. The return levels include data over the period 1899–2008. The confidence interval is the 90% interval. The results using the full data set are presented alongside for comparison. . . . .	32
3.4	Estimated maximum possible wind speeds. Estimates are for the strongest winds ( $\text{ms}^{-1}$ ) within 100 km of the city center. . . . .	33
3.5	Return period comparisons. Return periods (year) from HURISK and HRC are shown for selected Florida cities. Hurricanes are denoted by H and major hurricanes are denoted by MH. St. Lucie County is included to show a comparison between the return periods for Fort Pierce (HURISK) and Port St. Lucie (HRC). Lee County is included to show a comparison between Fort Myers (HURISK) and Cape Coral (HRC). “Differ” is the return period from HURISK minus the return period from HRC. . . .	34
4.1	Sample of data used for analysis. Includes five hurricanes that affected KAFB, including the name, year, surge height, surge location, wind speed, and wind location. . . . .	43
4.2	Surge height return level estimates for specified return periods at EAFB. The CI listed is the 95%. . . . .	48
4.3	Surge height return level estimates for specified return periods at MAFB. The CI listed is the 95%. . . . .	48

4.4	Surge height return level estimates for specified return periods at KAFB. The CI listed is the 95%. . . . .	49
4.5	Wind speed and surge level return periods for EAFB. Return period year information for Saffir-Simpson Category 1–4 wind speeds are shown along with 3, 4, and 5 meter surges. The quartile pointwise CI is shown in parentheses. . . . .	56
4.6	Wind speed and surge level return periods for MAFB. Return period year information for Saffir-Simpson Category 1–4 wind speeds are shown along with 3, 4, and 5 meter surges. The quartile pointwise CI is shown in parentheses. . . . .	57
4.7	Wind speed and surge level return periods for KAFB. Return period year information for Saffir-Simpson Category 1–4 wind speeds are shown along with 3, 4, and 5 meter surges. The quartile pointwise CI is shown in parentheses. . . . .	57

# LIST OF FIGURES

2.1	Hurricane Andrew (1992) intensities along an hourly plot. Dashed line represents the observed values, and the solid line represents the spline interpolated values. . . . .	15
2.2	North Atlantic hurricanes from 1851–2009. Tracks shown are the 1-hourly spline interpolated location values. The shading represents Saffir-Simpson wind speed categories. The darkest gray represents Category 1 (at least $33 \text{ ms}^{-1}$ ) and the lightest gray represents Category 5 ( $\geq 69 \text{ ms}^{-1}$ ). . . . .	16
2.3	Location points for all U. S. Gulf coast hurricane storm surge events from 1886–2010 as taken from SURGEDAT. . . . .	18
2.4	Location grid of NOAA reconstructed SST values. The locations are taken at $5^\circ$ latitude-longitude grid boxes from the equator north to $70^\circ\text{N}$ latitude and spanning the North Atlantic Ocean. . . . .	20
3.1	Florida cities. The twelve Florida cities used in this study are shown with the city centroid location denoted with a square. The 100 km radial distance from the city center is shown with a circle. This distance is used in the HRC and encompasses each city. County shading represents the population density per square kilometer. . . . .	23
3.2	HRC parameters. The values for the HRC parameters ( $\lambda$ , $\sigma$ , and $\xi$ ) as a function of radial distance are shown. Parameter values are obtained from a maximum likelihood procedure using the fastest wind speeds from hurricanes occurring within a radial distance of the geographic center of Miami. The bold line indicates the regression line of the parameter on distance from city center. . . . .	26
3.3	Threshold comparison. Miami’s $\xi$ parameter as a function of radial distance is shown. The dotted line shows the parameter using a $40 \text{ ms}^{-1}$ threshold. The solid line shows the parameter using a $35 \text{ ms}^{-1}$ threshold. . . . .	27

3.4	Miami’s return level. The return level as a function of return period for hurricanes passing within 100 km of the geographic center of Miami. The solid line shows the return level estimates from the HRC using data over the period 1851–2008. The 90% confidence intervals are shown as vertical lines. The dotted lines shows the return-level estimates from the HRC using data over the period 1899–2008. . . . .	28
3.5	Florida cities’ return level curves. Return level curves are shown as a function of return periods for hurricanes passing within 100 km of the geographic center of the specified city using the HRC. . . . .	30
3.6	Florida threat area. The threat area around Florida is shown with a black circle. This area is described by Hamid et al. (2009) as a 1000 km radius of a location (26°N latitude and 82°W longitude) off the southwest coast of Florida. . . . .	36
3.7	Hurricane frequency and intensity. Frequency and intensity of hurricanes in the vicinity of Florida are shown. (a) Frequency of all hurricanes as a function of time. (b) Frequency of major hurricanes (Category 3 and above) as a function of time. (c) Intensity (maximum sustained wind speed in $\text{ms}^{-1}$ ) as a function of time with trend lines shown for the 50 <sup>th</sup> and 90 <sup>th</sup> percentiles. . . . .	38
3.8	Trends in the maximum intensification rate of hurricanes as a function of quantile. The maximum intensification rate is calculated as the rate of change between hourly observations. Trends are estimated coefficients from quantile regression in units of $\text{ms}^{-1}/\text{hr}$ per year. The point-wise 90% confidence band is shown in grey. The solid line is the trend from a least-squares regression of the rate of intensification as a function of year, and the dashed lines delineate the 90% confidence band about this trend. . . . .	39
4.1	Conceptual schematic depicting data collation. Shows a hypothetical situation. Shapes of the variables are arbitrary. Wind and surge are point observations. The arrow represents a hypothetical hurricane track.	44
4.2	AFB study areas. For EAFB, the base is denoted with a red polygon. For MAFB and KAFB, the base is denoted with a red point. The surge events are shown with squares. The inset map includes a bounding box in the southeastern United States showing where the locations are globally.	45
4.3	The annual exceedance probability for storm surge heights (m). Return levels are shown on the $x$ -axis and the probabilities of annual exceedance are shown on the $y$ -axis. The 95% CI is shown with red lines. (a) EAFB, (b) MAFB, and (c) KAFB. . . . .	47

4.4	Maximum sustained wind speed ( $\text{ms}^{-1}$ ) regressed on storm surge height (m) at EAFB, MAFB, and KAFB. The red line shows the linear model. The adjacent axes show the marginal distributions of the respective variables. . . . .	50
4.5	Density plot and wind speed/surge level return periods for EAFB. (a) Density plot for the EAFB copula. The points shown are the empirical data. Maximum sustained wind speeds ( $\text{ms}^{-1}$ ) are shown on the $x$ -axis and maximum surge heights (m) are shown on the $y$ -axis. (b) Return period information for wind and surge levels at EAFB. Each line represents a specific return period year. The years shown are 20, 30, 50, 75, 100, 150, 200, 300, and 500. . . . .	53
4.6	Density plot and wind speed/surge level return periods for MAFB. (a) Density plot for the MAFB copula. The points shown are the empirical data. Maximum sustained wind speeds ( $\text{ms}^{-1}$ ) are shown on the $x$ -axis and maximum surge heights (m) are shown on the $y$ -axis. (b) Return period information for wind and surge levels at MAFB. The years shown are 20, 30, 50, 75, 100, 150, 200, 300, and 500. . . . .	54
4.7	Density plot and wind speed/surge level return periods for KAFB. (a) Density plot for the KAFB copula. The points shown are the empirical data. Maximum sustained wind speeds ( $\text{ms}^{-1}$ ) are shown on the $x$ -axis and maximum surge heights (m) are shown on the $y$ -axis. (b) Return period information for wind and surge levels at KAFB. The years shown are 20, 30, 50, 75, 100, 150, 200, 300, and 500. . . . .	55
5.1	Each variable plotted per hexagon. (a) $\lambda$ parameter, or rate of hurricanes exceeding $33 \text{ ms}^{-1}$ per year. (b) $\sigma$ parameter, or scale (dispersion) of wind speeds in $\text{ms}^{-1}$ . (c) $\xi$ parameter, or shape of the distribution tails. This is a dimensionless quantity. (d) The thirty year return level in $\text{ms}^{-1}$ . The maps are projected in a Lambert conformal conic planar projection with parallels at $23^\circ$ and $38^\circ$ N and a center longitude of $60^\circ$ W.	65
5.2	Comparison of empirical maximum wind and theoretical maximum wind. (a) Empirical wind speed ( $\text{ms}^{-1}$ ). (b) Theoretical wind speed ( $\text{ms}^{-1}$ ). . . . .	66
5.3	Comparison of rate parameter at three different hexagon sizes. $n = 125$ is the size used in the chapter analysis. . . . .	68
5.4	Comparison of $\sigma$ parameter at three different hexagon sizes. $n = 125$ is the size used in the chapter analysis. . . . .	69
5.5	Comparison of $\xi$ parameter at three different hexagon sizes. $n = 125$ is the size used in the chapter analysis. . . . .	70

5.6	Comparison of the thirty year return level at three different hexagon sizes. $n = 125$ is the size used in the chapter analysis. . . . .	71
5.7	Graph of hexagon connectivity defined by first-order contiguity . . . . .	72
5.8	Moran's scatter plot. (a) Scatter plot of the per hexagon annual probability versus the neighborhood average annual probability. (b) Scatter plot of the per hexagon scale parameter, $\sigma$ , versus the neighborhood average $\sigma$ . (c) Scatter plot of the per hexagon shape parameter, $\xi$ , versus the neighborhood average $\xi$ . (d) Scatter plot of the per hexagon thirty year return level versus the neighborhood average thirty year return level. Neighbors are defined by spatial contiguity. The slope of the best-fit line is Moran's I as a measure of spatial autocorrelation. . . . .	74
5.9	Marginal effect of SST on the number of hurricanes per year (rate). . . . .	76
5.10	Marginal effect of SST on $\sigma$ ( $\text{ms}^{-1}$ ). . . . .	77
5.11	Marginal effect of SST on $\xi$ [Dimensionless]. . . . .	78
5.12	Marginal effect of SST on the thirty year return level ( $\text{ms}^{-1}$ ). . . . .	79

# ABBREVIATIONS

The following short list of abbreviations are used throughout the document.

AIC - Akaike Information Criterion  
AFB - Air Force Base  
ASO - August, September, October  
CDF - Cumulative Distribution Function  
CI - Confidence Interval  
EAFB - Eglin Air Force Base  
EVT - Extreme Value Theory  
GEV - Generalized Extreme Value  
GoM - Gulf of Mexico  
GPD - Generalized Pareto Distribution  
HRC - Hurricane Risk Calculator  
HURDAT - Hurricane Database  
KAFB - Keesler Air Force Base  
MAFB - MacDill Air Force Base  
MAUP - Modifiable Areal Unit Problem  
NAtl - North Atlantic  
NCAR - National Center for Atmospheric Research  
NetCDF - Network Common Data Frame  
NHC - National Hurricane Center  
NOAA - National Oceanic and Atmospheric Administration  
NWS - National Weather Service  
POT - Peaks Over Threshold  
SST - Sea Surface Temperature  
SURGEDAT - Surge Database

# ABSTRACT

Hurricanes threaten the United States every year. It is important to quantify the risk of these events for emergency managers. Extreme value statistics are used to model hurricane characteristics at different locations. Using wind speeds over a specified threshold, the risk of extreme winds are estimated in twelve Florida cities. The risk estimates are provided as statistical return periods, or the expected frequency of specific hurricane magnitudes. Results show that the city of Miami can expect to see hurricane winds blowing at  $50 \text{ ms}^{-1}$  (45.5–54.5) [90% CI] or stronger, on average, once every 12 years. In comparison, the city of Pensacola can expect to see hurricane winds of  $50 \text{ ms}^{-1}$  (46.9–53.1) [90% CI] or stronger once every 24 years. Hurricanes in the vicinity of Florida are found to be increasing in intensity over time as a product of higher offshore intensification rates. The risk of hurricane strikes can be further understood by including an additional variable, storm surge. Using observational data, the joint probability distribution of hurricane wind speeds and storm surge is found at three Air Force Bases along the U. S. Gulf coast. The quartile pointwise uncertainty is quantified using a Monte Carlo procedure. Eglin Air Force Base can expect wind speeds blowing at  $50 \text{ ms}^{-1}$  and surge heights of 3 m, on average, once every 28 years (23–36). MacDill Air Force Base can expect wind speeds blowing at  $50 \text{ ms}^{-1}$  and surge heights of 3 m, on average, once every 27 years (22–34). Keesler Air Force Base can expect wind speeds blowing at  $50 \text{ ms}^{-1}$  and surge heights of 3 m, on average, once every 15 years (13–18). Utilizing a spatial tessellation across the North Atlantic and Gulf of Mexico provides additional insight into the risk of hurricane strikes. Parameters from the extreme value model are mapped across space to visualize patterns. Sea surface temperature is included as a covariate in a geographically weighted regression model with each parameter. It is found that as sea surface temperatures increase, the expected hurricane wind speed for a given return period also increases.

# CHAPTER 1

## HURRICANE RISK

Nowhere in the world is safe from natural disasters. Earthquakes, forest fires, hurricanes, and countless other hazards are responsible for the loss of life and the massive monetary costs that areas around the globe feel annually. Thankfully, changes in warning systems and broadcast technology over the last few decades has allowed for people to be more aware of these hazards and the types of risks an occurrence might pose. This awareness, however, does not mitigate against the potentially staggering economic losses experienced after a major disaster. In the 20<sup>th</sup> century alone, monetary losses that have come from tropical cyclones have reached over \$450 billion (Pielke Jr. et al., 2008; Malmstadt et al., 2009). Much of this damage is caused by the high wind speeds and deep storm surges from the most intense hurricanes. Storm surge is considered the most destructive component and is defined as the difference between the storm tide and the normal tide (Elsner and Kara, 1999). As a hurricane makes landfall, the larger storm tide will pour over the landscape and inundate much in its path, potentially costing large sums of money.

Considering the damage from these types of events reach such a high monetary cost, it is necessary that the scientific community pursues a deeper understanding of statistical risk. Risk, here, refers to the probability of an area experiencing a natural disaster in a given time period. It is separate from loss of life or livelihood. The likelihood of a hurricane occurrence is different than the likelihood of other natural hazards. The location of an earthquake's epicenter, for example, cannot be predicted or anticipated more than a few days before the actual occurrence. Once the event occurs, the spatial extent of the actual quake can be very localized or spread out in a much larger domain. Most importantly, it is difficult to accurately estimate the risk of earthquake occurrences (Masala, 2012). A forest fire, on the other hand, is fairly easy to anticipate if the conditions in an area have been conducive to fire formation. If the area has had drought-like conditions for an extended time and the air temperature has been higher than average, the likelihood of a forest fire disaster increases (Finney et al., 2011). The spatial extent of a forest fire rarely stays localized. These events usually encompass very large areas and can devastate communities hundreds of miles away from where the first ignition point occurred. Understanding these two very different types of hazards allows for a differentiation to be made between these types

and a tropical cyclone, or hurricane, hazard.

Hurricanes pose a particularly interesting threat. Monitoring global and local weather using satellites since the 1960s has allowed researchers to capture the formation of hurricanes in the North Atlantic Ocean, Caribbean Sea, and Gulf of Mexico. There is a heightened sense of awareness and anticipation of these events from the moment of genesis until they dissipate or make landfall. The environmental conditions conducive to formation have been researched to the extent that scientists can actively anticipate a hurricane type disturbance before it actually manifests. The spatial extent of this type of hazard varies from localized areas to state sized areas. In comparison to the other hazards mentioned, the ability to anticipate and model hurricane strikes makes the quantification of this risk different. The likelihood of hurricane occurrence can be estimated in many ways and it can be localized for specific areas.

The purpose of this dissertation is to provide innovative methodologies for understanding the statistical risk of hurricanes using various spatial extents. Citizens all around the world suffer from the detrimental effects of these events making landfall on their communities, so it is the responsibility of scientists to drive our knowledge further for the benefit of the public. There are three major questions being asked. The first: what is the risk of extreme hurricane winds in various cities around the state of Florida? The second: what is the risk of hurricane storm surge along the Gulf of Mexico, and what is the combined risk of hurricane storm surge and maximum wind speeds at the locations of interest? The third: what is the spatial distribution of hurricane winds across the North Atlantic, and how does this spatial distribution compare with the spatial distribution of climatological sea surface temperature? Sea surface temperature is included because the relationship between the warmth of the ocean surface and the location of extreme hurricane winds might provide insight into the effects of a warming climate. The methods presented can be utilized in other locations not analyzed here, and further developed in future research so the discipline continues to grow. The results presented are useful to those living in the areas of interest, as well as those creating government policies or insurance rates in the areas studied.

All three research studies include wind speed as a particular hazardous characteristic from hurricanes. Study 2 also includes hurricane storm surge. A major difference in the approaches presented lies in the spatial aspect of each. The first study's spatial scale is state sized as it focuses on the probability of hurricane occurrence in local cities around Florida. The second study is localized and looks at three individual locations along the U. S. Gulf coast. The third study has the largest spatial scale and can be considered a more regional approach to the understanding of hurricane occurrence. Each of the studies considers virtually the same length of record, and the wind speed and storm surge variables can be considered instantaneous characteristics. That being said, the unifying factor through each of these is the temporal element. The differentiating factor is space.

Six chapters describe this work. The first chapter discusses the statistical risk of

hurricanes, and in the remaining sections relevant terms and literature are defined and discussed to provide a proper background for the studies performed. Chapter 2 describes the data that were collected, refined, and utilized in these studies. Chapter 3, Chapter 4, and Chapter 5 each present a research study including the methodologies used and the results that were collected. Chapter 6 summarizes the approaches and results, and also draws a connection between the results and societal vulnerability. The Appendix provides some of the statistical code used to calculate various results in Chapter 4 and Chapter 5.

## 1.1 Hurricanes

A tropical cyclone is a closed circulation of air that develops over the warm waters of the tropical latitudes. In the Northern Hemisphere, these spin in a counterclockwise direction around a center of low pressure and are called hurricanes. In order for a hurricane to develop, certain conditions need to exist. The temperature of the ocean surface needs to be at least  $26.5^{\circ}\text{C}$ , there must be a pre-existing atmospheric disturbance, it must be further north of  $8^{\circ}$  latitude, and the horizontal winds in the atmosphere must be relatively weak ( $< 10 \text{ ms}^{-1}$ ) (Elsner and Kara, 1999).

The difference between the intense low pressure and surrounding higher pressure causes extreme winds to develop over time. A tropical cyclone is not automatically a hurricane, but rather passes through a life cycle from the development of an atmospheric disturbance to a major hurricane. First, a tropical disturbance, or tropical wave, will develop when there is a small pressure drop that causes enhanced cloudiness and convection. In some cases, this stage is essentially absent and the event becomes a tropical depression immediately. A tropical depression has a maximum sustained wind speed (taken as a one-minute average at 10 meters above the surface) less than  $18 \text{ ms}^{-1}$ . The depression becomes a tropical storm when the one-minute average wind speed is between 18 and  $32 \text{ ms}^{-1}$ . It becomes a full blown hurricane when it reaches  $33 \text{ ms}^{-1}$ , and is considered a major hurricane when it reaches  $50 \text{ ms}^{-1}$ . There is a scale that is used to estimate the intensities of hurricanes once they reach  $33 \text{ ms}^{-1}$ , and it is known as the Saffir-Simpson Scale, named after Herbert Saffir and Robert Simpson. The categories of hurricanes span from 1–5, with 5 being the most catastrophic, experiencing wind speeds greater than  $69 \text{ ms}^{-1}$ .

Hurricanes affect the North Atlantic during an “official” season that runs from June 1 through November, designated by the U. S. National Weather Service. It is, of course, possible for a hurricane to form outside of this season if the conditions are met, but this is rare. The National Oceanic and Atmospheric Administration’s (NOAA) National Hurricane Center (NHC) keeps records of the events that occur each season and have created a record that extends back to 1851. These data will be described in further detail in Chapter 2.

The most destructive characteristics of a hurricane are its wind speeds and its storm surge. One of the second study’s goals is to describe the statistical relationship

between wind speed and storm surge because, physically, they are connected. Storm surge is created through the momentum flux between the energy exchange of the winds at the hurricane's base and the waves at the top of the water column under the sea surface below the hurricane. As the hurricane nears the continental shelf, the water column shortens and pushes more water toward the surface (Blain et al., 1994). By direct interaction with the air flow, this change in swell modifies the wind stress (in this case, defined by air density and the turbulent fluctuations in vertical and horizontal velocities) and will, therefore, affect the exchange rate of the momentum flux between the ocean and atmosphere (Donelan et al., 1997). Considering these characteristics interact with one another on a physical basis, finding the statistical relationship of these variables may provide additional insight into the understanding of their combined probability of occurrence.

There is no shortage of literature on climatological characteristics of particular basins or regions (Landsea, 1993; Elsner and Kara, 1999; Emanuel, 2003; Scheitlin et al., 2011), simulated models that show their interaction with other environmental characteristics (Tsutsui and Kasahara, 1996; Walsh and Katzfey, 2000; Knutson et al., 2007; Caron et al., 2011), and risk estimations of particular characteristics (Darling, 1991; Coles and Casson, 1998; Jagger and Elsner, 2006; Malmstadt et al., 2010). This dissertation focuses on an area of risk estimation that involves the use of extreme value theory to analyze a sample of extreme data. These are data that fall on the tail ends of a distribution, and by theory, do not occur often in records. When interested in extreme hurricanes, the desire is to better understand the upper limits of the known wind speed and storm surge distributions so that society can better prepare for the disasters that cause the most damage.

## 1.2 Extreme Value Theory

The First Law of Geography states that “all things are related, but nearby things are more related than distant things” (Tobler, 1970). This law is what makes spatial data unique because it allows us to use information collected in one place at a nearby location when those data are not available. In particular, extreme data occur at the limits of a distribution and are not as common as those nearest the mean. At any specific location, the record of hurricane events is short because the likelihood of an event making a landfall is not high. Considering this, the likelihood of a major hurricane, or extreme hurricane, at a specific location is even less. For the studies presented here, extreme hurricane wind data and surge data are used to represent the likelihood of occurrence at a specific location using some of the recorded data near the locations of interest, similar to what has been done in Elsner et al. (2008a).

According to Kotz and Nadarajah (2000), probabilistic extreme value theory (EVT) blends an enormous variety of applications involving natural phenomena such as rainfall, floods, air pollution, and wind gusts. Recently, this theory has been applied to tropical cyclone observations to try and estimate the occurrence of hurricanes

affecting the United States. Embrechts et al. (1999) suggest that, because the reinsurance industries have experienced higher losses in recent years from unprecedented catastrophes, extreme value modeling can be an exceptionally useful tool in emergency management. Before discussing applications of extreme value distributions, it is important to understand the theory behind it.

Probabilistic EVT deals with the stochastic behavior of the maximum and the minimum of independent and identically distributed random variables. The distributional properties of extremes, as well as of exceedances over (below) high (low) thresholds are determined by the upper and lower tails of the underlying distribution (Kotz and Nadarajah, 2000). In particular, extreme value analysis usually requires estimation of the probability of events that are more extreme than have ever been observed (Coles, 2001). EVT provides a framework that enables this type of extrapolation.

Historically, work on extreme value problems can be traced back to as early as 1709 when Nicolas Bernoulli discussed the mean largest distance from the origin given  $n$  points lying at random on a straight line of a fixed length  $t$ , see Gumbel (1958). The need for EVT originated mainly from the needs of astronomers in utilizing or rejecting outlying observations. A systematic development of the general theory can be regarded as starting from von Bortkiewicz (1922) that dealt with the distribution of range in random samples. It was in this paper that the concept of “distribution of largest value” was introduced for the first time. Frèchet (1927) followed with a paper in which asymptotic distributions of largest values were considered. Fisher and Tippett (1928) provided a study that identified the only three possible extreme limit distributions that exist. In the 1930s and 1940s, numerous papers dealing with practical applications of extreme value statistics were conducted, including on radioactive emissions (Gumbel, 1937), strength of materials (Weibull, 1939), flood analysis (Gumbel, 1941), seismic analysis (Nordquist, 1945), and rainfall analysis Potter (1949). Gumbel was the first to call the attention of engineers and statisticians to certain distributions which had previously been dealt with empirically. Gumbel (1941) was the first research study using EVT in the U. S. dealing with meteorological phenomena (Kotz and Nadarajah, 2000). This research has spurred decades of subsequent meteorological analysis.

According to Coles (2001), the extreme value paradigm refers to an approach that comprises a principle for model extrapolation based on the implementation of mathematical limits as finite-level approximations. This suggests an implicit assumption that the underlying stochastic mechanism of the process being modeled is sufficiently smooth to enable extrapolation to unobserved levels. This may be flawed, but no credible alternative has been proposed. It is important to understand the limitations of extreme value modeling from the outset. First, models are developed using asymptotic arguments, so care is needed in treating the models as exact results for finite samples. Second, the models themselves are derived under idealized circumstances, and third, the models may lead to information loss when implemented in practice. This final limitation comes into play when taking block maxima per unit time, for

example, and is less important in the research presented in the upcoming chapters where a different approach is taken to obtaining maximum data (Coles, 2001).

As stated above, only three possible extreme limit distributions exist (Fisher and Tippett, 1928). Extreme value distributions are usually considered to comprise the following three families: Type 1 (Gumbel-type distribution) whose upper tail is infinite (though shorter than the Frèchet distribution), Type 2 (Frèchet-type distribution) with an infinite upper tail, and Type 3 (Weibull-type distribution) with a finite upper tail (Kotz and Nadarajah, 2000). The Gumbel type is shown by

$$\Pr[X \leq x] = \exp \left[ -e^{\frac{(x-\mu)}{\sigma}} \right], \quad (1.1)$$

the Frèchet-type is shown by

$$\Pr[X \leq x] = \begin{cases} 0 & x < \mu \\ \exp \left\{ - \left( \frac{x-\mu}{\sigma} \right)^{-\xi} \right\} & x \geq \mu \end{cases} \quad (1.2)$$

and the Weibull-type is shown by

$$\Pr[X \leq x] = \begin{cases} \exp \left\{ - \left( \frac{\mu-x}{\sigma} \right)^{\xi} \right\} & x \leq \mu \\ 0 & x > \mu \end{cases} \quad (1.3)$$

where  $\mu$ ,  $\sigma$ , and  $\xi$  are parameters. According to Coles (2001), these three families can be combined into a single family of models having distribution functions of the form

$$G(x) = \exp - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{\frac{-1}{\xi}}, \quad (1.4)$$

where the parameters satisfy  $-\infty < \mu < \infty$ ,  $\sigma > 0$ , and  $-\infty < \xi < \infty$ . This is the generalized extreme value (GEV) family of distributions. The model has three parameters, a location parameter  $\mu$ , a scale parameter  $\sigma$ , and a shape parameter  $\xi$ . This unification into a single family greatly simplifies statistical implementation. Through inference on  $\xi$ , the data themselves determine the most appropriate type of tail behavior, and there is no necessity to make subjective a priori judgements about which individual extreme value family to adopt.

However, according to Coles and Powell (1996), there is one main drawback to the GEV family of distributions as an inference model. It has restriction to annual maximum data when in fact the series,  $X_1, X_2, \dots, X_n$  may contain many other data which are informative about the tail of the resultant distribution. Thus, models based on the threshold approach were developed. This leads to the second general characterization based on exceedances over a high threshold. If an entire time series of data are available, a better use is made of the data by avoiding altogether the procedure of blocking (Coles, 2001). Many times, in practical applications, the parent distribution  $F$  of a given set of observations is not known. Approximations that

are broadly applicable for high values of the threshold are thus sought after. This parallels the use of the GEV as an approximation to the distribution of maxima of long sequences when the parent population is unknown. An asymptotic model characterization is known as the generalized Pareto distribution (GPD).

The following theorem discusses the GPD. Let  $X_1, X_2, \dots$  be a sequence of independent random variables with a common distribution function  $F$  and let

$$M_n = \max\{X_1, \dots, X_n\}. \quad (1.5)$$

Denote an arbitrary term in the  $X_i$  sequence by  $X$ , and suppose that  $F$  satisfies, so that for large  $n$ ,  $\Pr\{M_n \leq x\} \approx G(x)$ , Equation 1.4 applies for some  $\mu, \sigma > 0$ , and  $\xi$ . Then, for a large enough threshold,  $u$ , the distribution function of  $(X - u)$ , conditional on  $X > u$ , is approximately

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} \quad (1.6)$$

defined on  $\{y : y > 0 \text{ and } (1 + \frac{\xi y}{\tilde{\sigma}}) > 0\}$ , where  $\tilde{\sigma} = \sigma + \xi(u - \mu)$  and  $\mu, \sigma$ , and  $\xi$  are the GEV parameters of Equation 1.4. Taking into account also the crossing rate of the threshold  $u$  gives a three-parameter model equivalent to the three-parameter GEV model (Coles and Powell, 1996). Inference based on this GPD approach is generally superior because it relies on more data (Coles and Powell, 1996). It is for this reason that the GPD threshold approach is adopted for the studies in this dissertation.

There are numerous studies that have been conducted utilizing the various distributions mentioned above. Neumann (1987) employs a Weibull distribution for tropical cyclone wind speeds and a Poisson distribution for hurricane frequency and estimates return periods at the county level. This model is called HURISK, and is utilized by NOAA for estimating tropical cyclone return periods at site-specific locations. The latest output is presented in Blake et al. (2007).

Heckert et al. (1998) use the peaks-over-threshold model (POT) and a reverse Weibull distribution to obtain the mean recurrence intervals for extreme wind speeds at various locations along the U. S. coastline. The GPD is utilized as the threshold modeling approach. It is found that the results are consistent with the assumption that the reverse Weibull distribution is an appropriate probabilistic description of simulated extreme wind speeds. The results tend to be lower, however, than what are offered in other literature, which is due to the inclusion of the POT approach, which, the author's argue should be adopted in future studies.

Chu and Wang (1998) employ the use of the Gumbel, the Weibull, and the log-normal distribution to model return periods for the relative intensities of tropical cyclone wind speeds in the vicinity of Hawaii. Relative intensity is a dimensionless quantity that is efficient in its use in extreme value distributions because of it represents the characteristics of tropical cyclones in a compact fashion. After model adequacy checks, the authors find that the lognormal distribution is best for the overall goodness of fit, so it is this distribution that is used in the Monte Carlo method to simulate the relative intensities in the vicinity of Hawaii.

Jagger et al. (2001) use a maximum likelihood estimator to determine a linear regression for the parameters of the Weibull distribution for tropical cyclone wind speeds in coastal counties of the United States. The results from this approach compare favorably with the HURISK approach (Neumann, 1987), but the inclusion of covariate information shows discrepancies from the normally used HURISK. This paper suggests that the inclusion of covariate information can provide a better understanding when assigning probabilities in a given, say El Niño, year. The model provided in Jagger et al. (2001) generates exceedance probabilities that are conditioned on climate factors by modeling the parameters with linear regression. It is suggested that at the highest hurricane intensities, the results offered by the dynamic model are likely too high as a result of the model fitting the data to the distribution (a heavy tail). They suggest that an approach utilizing a thresholding model would alleviate that problem.

Coles and Casson (1998) use a GPD, with a slight re-parameterization, that enables models fitted at different thresholds or different locations to be easily compared. They also employ the maximum likelihood as the method of inference. They only model wind speeds above a certain threshold  $u$ , and additionally specify the exceedance rate of  $u$ , with  $\lambda$ . The slight re-parameterization mentioned above accounts for the spatial differences in the simulated hurricane data. The results in this paper identify that the spatial variation in return levels is not solely due to variation in hurricane rates, but also due to different patterns of hurricane behavior in different locations.

Jagger and Elsner (2006) produce estimates for extreme hurricane winds near the United States using a GPD. The approach includes covariate information to adjust the return levels depending on climate factors, like El Niño and the North Atlantic Oscillation. A maximum likelihood approach is used to model the maximum wind speeds, and the GPD describes the maximum wind distribution for each hurricane with winds exceeding  $u$ . It incorporates the frequency, or rate, of hurricane occurrence as an exceedance rate modeled with a Poisson distribution. This is represented by  $\lambda$ . The approach is modified in Elsner et al. (2008a) to estimate hurricane return levels for Lake Shelby, Alabama as a way to compare geological proxies of past hurricanes with historical records. The return period estimates using only the known historical data are compared with the estimates that utilize the additional geologic proxy data. Parisi and Lund (2008) use a similar approach (the GPD) but add a component for the day of hurricane occurrence. Malmstadt et al. (2010) automate the approach offered in Elsner et al. (2008a) and apply the GPD, with the exceedance rate, to various cities around the state of Florida.

Within the field of tropical cyclone analysis, extreme value modeling has not only been applied to extreme wind speeds, but has also begun to be applied to storm surge analyses. Walton Jr. (2000) assesses a variety of potential probabilistic models for storm surge in New Jersey. The various extreme value distributions applied are compared by determining the quality of fit in the upper tail of the data. The authors conclude that many different probability distributions may be adequate for models in hydrology, although underprediction may occur in the largest extremes of the data,

so the choice of distribution should be centered on how well the distribution models the upper tail of the data. Butler et al. (2007) use both observational data and model reanalysis data of surge to study temporal trends. In this paper, the authors utilize the  $r$ -largest value model (as an extension of GEV discussed above) to allow for the inclusion of additional data. They argue that the GEV model alone makes highly inefficient use of the available data since it only allows us to use one observation per year.

The vast majority of these analyses are univariate models. In Chapter 3 hurricane wind speed is used to estimate the univariate risk to various locations in Florida. However, it is interesting to consider the benefit of including an additional variable in a model. In the second study presented in this dissertation, the relationship between two hurricane characteristics, wind *and* surge, is explored using a bivariate copula approach.

### 1.3 Copulas

Extreme wind speeds and deep storm surges from hurricanes work in conjunction to cause severe widespread damage when major hurricanes make landfall. According to Liu et al. (2009), the overall risk from a tropical cyclone event cannot be predicted by the traditional univariate extrapolation from annual maximum data sampling alone. This suggests that a method must be employed that allows for multiple cyclone characteristics to be incorporated into the estimation scheme. The traditional way to estimate a combined risk is to use copulas. These discuss the dependency structure of one random variable given the occurrence of another. Coles et al. (1999) discuss that dependence between random variables occurs when different processes have a stochastic behavior that is linked to common meteorological conditions. It is also possible for variables to be dependent when a single process is studied at multiple locations or in terms of its temporal evolution.

The second study of this dissertation, presented in Chapter 4, estimates the dependency structure of extreme hurricane winds and storm surges at three Air Force Bases (AFB) along the U. S. Gulf coast. Hurricane wind speeds and hurricane storm surges are linked because of the meteorological phenomenon of the hurricane itself. The broad objective of this study is to develop a measure of extremal dependence for wind and surge. Probabilities of specific surge and wind levels occurring at the three locations are estimated.

The following is the statistical theory behind copula models. For a given pair of  $(X, Y)$  values, if it is assumed that the marginal distributions of these variables are the same, a natural measure of dependence is

$$\chi = \lim_{z \rightarrow z^*} \Pr(Y > z | X > z), \quad (1.7)$$

where  $z^*$  is the upper limit of the support of the common marginal distribution (Coles et al., 1999). Stated more loosely,  $\chi$  is the probability of one variable being extreme

given that the other is extreme. It is well understood that marginal distributions of random variables are often not identical, which leads us to the copula function that allows for the removal of marginal influences. Tawn (1988) and Coles et al. (1999) describe the theory and dependence measures for extreme values analysis such as this. The copula function for a measure of extremal dependence is described as follows:

For any random vector,  $(X, Y)$ , the distribution function

$$F(x, y) = \Pr(X \leq x, Y \leq y) \quad (1.8)$$

provides a complete description of dependence between variables  $X$  and  $Y$ . Marginal aspects can play an influence on the variables and can be removed by observing that, subject to continuity conditions, there is a unique function  $C(\cdot, \cdot)$  with domain  $\mathcal{A} = [0, 1] \times [0, 1]$  such that

$$F(x, y) = C\{F_X(x), F_Y(y)\}, \quad (1.9)$$

where  $F_X$  and  $F_Y$  are the marginal distribution functions given by

$$F_X(x) = F(x, \infty) \text{ and } F_Y(y) = F(\infty, y). \quad (1.10)$$

The function  $C$  is the copula and it contains complete information about the joint distribution of  $X$  and  $Y$  apart from the marginal distributions. In this way,  $C$  describes the association between the two variables in a form that is unaffected by marginal transformation. Stated another way,  $C$  is the joint distribution function of the two variables after transformation to variables  $U$  and  $V$ , with Uniform  $[0, 1]$  margins, via  $(U, V) = \{F_X(X), F_Y(Y)\}$ .

For a more complete summary of dependence measures, see Coles et al. (1999); Nelsen (1998) and Joe (1997).

Copulas are a popular multivariate modeling tool in various research avenues. They were first introduced in 1959, when Sklar (1959) showed that any multivariate distribution can be represented using a copula. He also showed that if the marginal distributions are continuous, then there is a unique copula representation. In recent years, copulas have been applied to actuarial science, where they are used to model dependent mortality and losses (Frees and Valdez, 1998). Financial research use copulas to allocate assets, score credit, and manage risk (Embrechts et al., 2003). Also, in engineering, copulas are used for hydrological modeling (Genest and Favre, 2007).

Smith et al. (1990) provide a thorough discussion of the statistical aspects of multivariate extreme value distributions. Notably, the authors discuss the relationship between univariate extremes and the multivariate extreme copula. They also include a kernel function to alleviate any smoothing problems. They find that the smoothing method of incorporating a kernel function in the calculation of the dependence parameter provides smoother estimates. They suggest this method is the most satisfactory of the nonparametric approaches, but it cannot be expected to work well in small samples. It will provide some indication of the dependence structure.

Coles et al. (1999) provide a thorough description of dependence measures for extreme value analyses and suggest three examples of data sets that have potential dependence. The third, which has the most relevance here, are wave and surge height data at a coastal location in England. In the third example, the motivation is that both the surge and wave processes are driven by similar meteorological conditions, so extreme behavior in both processes is likely to occur simultaneously. A similar relationship between wind speeds and surge heights is presented using this copula approach for estimation in Chapter 4.

Chapters 3 and 4 provide probability estimates for more localized areas. Specifically, Chapter 3 focuses on the likelihood of hurricane occurrence for a U. S. state location analyzing individual cities around Florida. Chapter 4 analyzes the combined probability of wind *and* surge at three local AFBs in the southeastern United States. Chapter 5 visualizes the univariate risk of hurricane wind speed across the Gulf of Mexico and North Atlantic by using a hexagonal tessellation representing larger spatial regions than the first two studies.

## 1.4 Tessellated Space

Chapter 5 utilizes a spatially oriented approach to visualizing statistical hurricane wind risk. The parameters of the univariate GPD model described previously are mapped out using hexagonal tessellations across the Gulf of Mexico and North Atlantic. Brettschneider (2008) states that the only geometric shapes that completely cover the Atlantic basin without any overlap are squares, triangles, and hexagons. The square is primarily limited by the corner representation. A network of triangles magnifies the corner issue, and also magnifies the orientation issue; that is, the direction the triangle points affects the number of storm tracks observed intersecting the triangle. A hexagon represents the best compromise between overlap and uniformity. Hexagons can be tessellated where no area is left uncovered. All of the interior angles of a hexagon are equal, and the sides are of equal length, so the hexagon best approximates the idealized shape of a circle. Hexagons are used to visualize the GPD model parameters to see if the information changes over space. In particular, the  $\lambda$ , or rate, the  $\sigma$ , or statistical scale, and the  $\xi$ , or shape, parameters are visualized.

As weather services monitor hurricanes, values of location, wind speed, pressure, size, and speed are gathered along the existing path of the event. Hurricanes rely on particular atmospheric and environmental conditions to continue to grow or sustain intensity. These events move through space and are thus potentially correlated through space. If multiple hurricanes exist near one another temporally and spatially, the values may have a dependency as well. In Chapter 5, spatial autocorrelation, or the propensity for similar values to be found together, is tested to see the dependence of particular parameters near one another spatially. Spatial dependence is an inherent property of an attribute in geographic space because of the underlying continuity of space (Haining, 2003), and, in this case, due to the operation of hurricane mechanics.

Christaller (1933) in his famous dissertation theorizes Central Place Theory (CPT), which is a spatial theory in urban geography that attempts to explain the distribution patterns, size, and number of cities and towns around the world. Christaller (1933) came to the conclusion that people gather together in cities to share goods and services. There is a rank order of central places where each smaller subdivision fits into the order above it. When mapped out visually, the central places are located at the points of equilateral triangles, and as they connect, a series of hexagons is formed. CPT attempts to understand the social construction of cities throughout space. It is different than the understanding of hurricane phenomenon throughout space because hurricanes are bounded by the physical elements of the atmosphere, while social behavior has no such bound. However, spatial autocorrelation is a property of attributes occurring throughout geographic space, physical or social. CPT describes similar people locating near one another in space for access to a common good. Hurricane values occurring near one another temporally and in similar environmental conditions are often similar as well. In this way, the geographic theory of central places and the geographic understanding of hurricanes can be similar.

Hexagons are used to map out space and compare differences in two relevant pieces of literature. Brettschneider (2008) uses a tessellated Gulf of Mexico and Atlantic basin to first, determine whether spatial patterns exist in the climatological record, and second, to determine whether probability can indicate the relative likelihood of landfall as a function of storm position. This study utilizes an empirical probability approach to assess the relative likelihood of movement of tropical cyclones based solely on the past movement of cyclones. The methodological approach of using hexagons is adopted for the final portion of this dissertation as an attempt to better understand patterns in the frequency and magnitude of these extreme wind events in the Gulf of Mexico and in the North Atlantic. Elsner et al. (2011) demonstrate a spatial framework for studying hurricane climatology using hexagons. The authors find that hexagons are efficient at covering hurricane tracks, as well as providing a background to combine attribute data from tropical cyclones with spatial climate data.

Other types of spatial hierarchical models exist. Cooley et al. (2007) apply a Bayesian analysis for spatial extremes of rainfall in Colorado to produce a map characterizing how extreme behavior behave across a geographic region. The authors spatially model the exceedance rate parameter assuming that the regional extreme precipitation is driven by a latent spatial process characterized by geographical and climatological covariates. They find that effects that are not fully described by the covariates included are captured by the spatial structure in the hierarchies. This is an interesting and innovative approach to understanding rainfall. That same assumption of the latent spatial process is applied in Chapter 5 when analyzing the parameters of the hurricane threshold model across space.

In Chapter 5, the relationship between each of the GPD parameters and sea surface temperature (SST) is analyzed using a geographically weighted regression (GWR) model. By including the SST covariate, the potential effect of an increasing SST on hurricane characteristics can be further understood. GWR slides a moving window

over the study area and fits the same regression equation to each spatially defined subset (Haining, 2003). A GWR model allows insight into the spatial differences of the relationship that a global regression model cannot provide. Hurricanes depend on particular environmental conditions, so it is beneficial to see where the various aspects of hurricane occurrence might be differently affected by SSTs.

The relevant literature and theories for a complete understanding of the upcoming studies have been discussed. Now, the necessary data are presented.

# CHAPTER 2

## DATA SOURCES

There are three main sources of data utilized in this dissertation to represent hurricane location and wind, hurricane storm surge, and sea surface temperature. The following sections describe each of these sources in detail. The specifics of how the data were further utilized in each study are discussed in the respective chapter.

### 2.1 Hurricane Data

Hurricanes are tropical cyclones whose wind speeds reach at least  $33 \text{ ms}^{-1}$ . The hurricane occurrence, location, and wind speed data used in this analysis are derived from the NHC's Hurricane Database (HURDAT; or best track) (Landsea et al., 2004). This is the official record of tropical cyclones for the Atlantic Ocean, Gulf of Mexico, and Caribbean Sea, including those that have made a U. S. landfall. This record consists of the 6-hourly cyclone location and intensity for individual storms back to 1851. For cyclones prior to 1931, the 6-hour information is interpolated from once-daily (1200 UTC) estimates. For cyclones in the period 1931–1956, the 6-hour information is interpolated from twice-daily observations (0000 and 1200 UTC). Analyses were previously conducted on storms during the second half of the nineteenth and early twentieth century to improve the quality of the data (Fernandez-Partagas and Diaz, 1996; Landsea et al., 2004). These data are further explained and can be downloaded from the NHC site <http://www.nhc.noaa.gov/pastall.shtml#hurdat>.

There remain limitations to these data that are relevant to the work presented in this dissertation. Storm information over the earlier part of the record is less certain than information over the more recent decades (Landsea et al., 2004). This time variation in uncertainty is likely larger in the collection of tropical cyclones occurring over the open ocean, but presents itself to some degree in land falling hurricanes. Unless the area was at least sparsely populated at time of landfall, the hurricane wind speed may not have been recorded. Despite the limitations, these data are frequently used for hurricane risk analysis (Emanuel et al., 2006).

In order to utilize as much information as possible, the 6-hourly hurricane track data were interpolated hourly for geographic position of the hurricane center (cen-

ter fix) and wind speed using splines described in Jagger and Elsner (2006). Using splines, a parsimonious smooth representation of the data is gained that preserves local properties. An example of the difference between the 6-hourly information and the 1-hourly information is shown in Figure 2.1. The hurricane intensity as a function of storm hour using the raw (6-hourly) and 1-hourly spline interpolated values for Hurricane Andrew (1992) is shown in Figure 2.1. The spline smoothing preserves

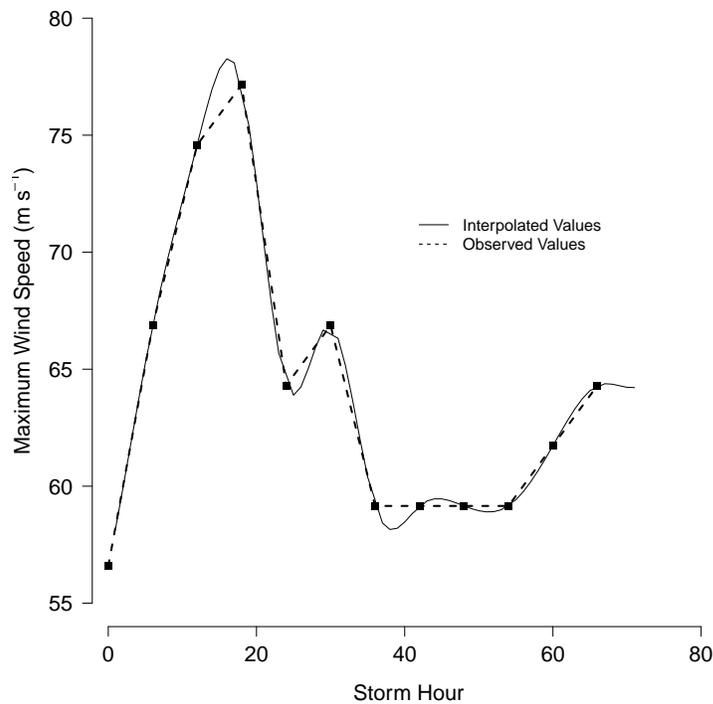


Figure 2.1: Hurricane Andrew (1992) intensities along an hourly plot. Dashed line represents the observed values, and the solid line represents the spline interpolated values.

the 6-hourly wind speed and gives a nonlinear interpolation of values at each hour between these values. The advantage of the hourly interpolated values is that the chance of missing a hurricane passing through any given area is greatly reduced. For example, with the 6-hourly data and an average translational velocity of  $18 \text{ km hr}^{-1}$  one would expect to miss about 10% of all hurricanes crossing a circular region with

a radius of 75 km. In contrast, one would expect to miss only about 0.2% of the hurricanes (2 in 1000 storms) crossing through the same region using hourly data. The missing track proportion is calculated assuming uniform track density and a fixed wind speed.

Figure 2.2 shows the known hurricane tracks with the 1-hourly locations that have affected the North Atlantic from 1851–2009. The color shading of the tracks denotes intensity. The darkest gray shows Category 1 (at least  $33 \text{ ms}^{-1}$ ) hurricanes and the lightest shade shows Category 5 hurricanes ( $\geq 69 \text{ ms}^{-1}$ ). Categories are based on the Saffir-Simpson Scale mentioned in Chapter 1. The weakest hurricanes are the most frequent, while the most extreme occur less often in the record. Each of the studies in

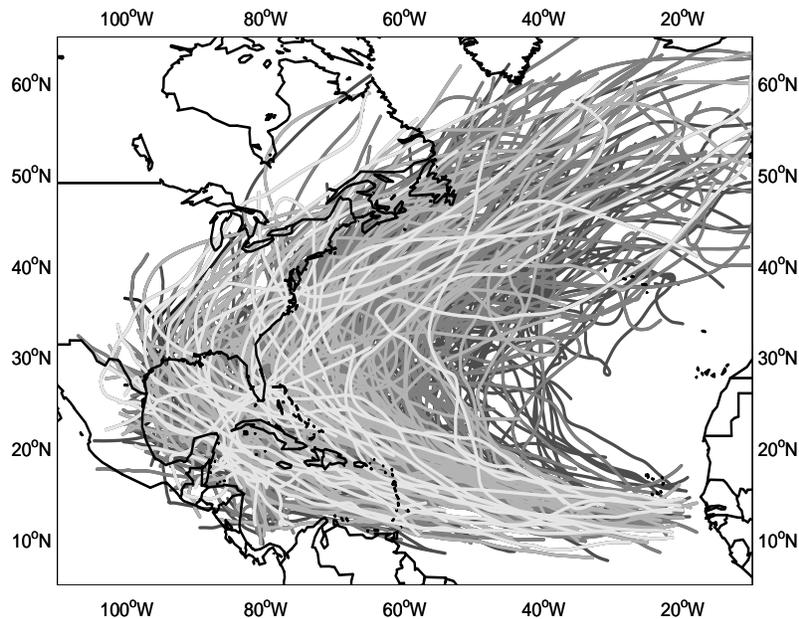


Figure 2.2: North Atlantic hurricanes from 1851–2009. Tracks shown are the 1-hourly spline interpolated location values. The shading represents Saffir-Simpson wind speed categories. The darkest gray represents Category 1 (at least  $33 \text{ ms}^{-1}$ ) and the lightest gray represents Category 5 ( $\geq 69 \text{ ms}^{-1}$ ).

the upcoming chapters utilize the 1-hourly interpolated wind speed data. However, each study utilizes different subsets of these data depending on the question being asked. The specific method of subsetting the data is discussed in the relevant chapter.

## 2.2 Storm Surge Data

The second study utilizes storm surge data along with the wind speed data. The storm surge information used in these analyses come from SURGEDAT, a global storm surge database (Needham and Keim, 2011). This dataset identifies the magnitude and location of peak storm surge for tropical cyclone-generated surge events that have occurred around the world since 1880. The database is available online at <http://surge.srcc.lsu.edu/>.

SURGEDAT utilized 62 sources of data, including 28 Federal government sources, numerous academic publications, and more than 3,000 pages of historical newspaper, to identify more than 200 storm surges along the U. S. Gulf coast. Federal data sources include the NHC, United States Geologic Survey, U. S. Army Corps of Engineers, and National Weather Service weather forecast offices. A given surge event often utilized multiple data sources to estimate the location and height of peak storm surge. For example, SURGEDAT referenced ten data sources, including six Federal sources, three books and one website, to make the most accurate possible storm surge estimate for Hurricane Camille (1969). All available information was documented from every source, producing more than 600 pages of metadata for surge events along the U. S. Gulf coast alone. The 202 storm surge events for the U. S. Gulf coast can be seen in Figure 2.3. Precise surge magnitudes were obtained for events in which tide gauges measured the highest surge levels. As much as possible, SURGEDAT removed tidal influences from the dataset. Where tide gauge data were available, the normal astronomical tide level was subtracted from the maximum observed water level to obtain the magnitude of maximum storm surge. However, when tide gauge data were unavailable, or gauges were not located in the area of peak storm surge, maximum storm surge levels were estimated from scientific and anecdotal documentation. Generally, these estimations provided water levels above normal, sometimes rounded off to the nearest foot, which were then converted to meters.

The influence of waves on maximum water levels was also removed from the dataset as much as possible. The manner in which waves were removed from scientific surge observations depends upon the data type. Waves often appear as noise on tide gauge graphs; in these cases the surge level was visually estimated at the lower end of the observed heights, where Mean Sea Level rose to produce a new baseline level. In regards to post-storm field work, greater weight was given to observations that utilized still water marks inside structures, such as interior mudlines, which minimize the effects of waves. For more details about the methodology utilized to construct SURGEDAT, please see Needham and Keim (2011).

Chaper 4 is the only study to utilize these data. Each location analyzed in the study uses a different subset of these data, and the details of the specific data collection are outlined when necessary.

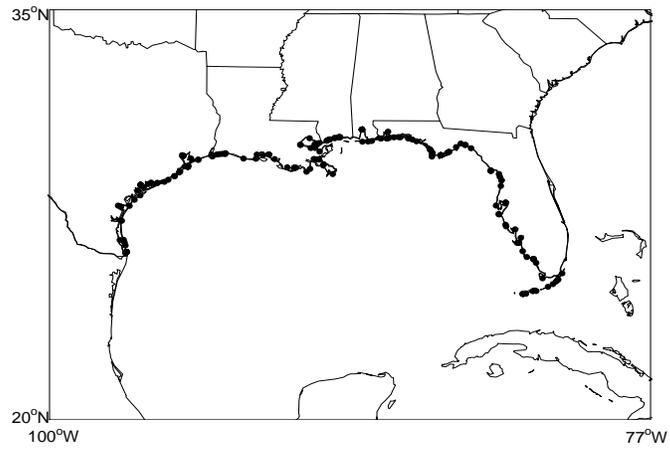


Figure 2.3: Location points for all U. S. Gulf coast hurricane storm surge events from 1886–2010 as taken from SURGEDAT.

## 2.3 Sea Surface Temperature Data

The final study in this dissertation (Chapter 5) looks at the spatial distribution of statistical risk across the North Atlantic using hexagonal tessellations. The relationship between the parameters in the risk model and SST over the North Atlantic is estimated using a GWR model across the hexagons. The data are the NOAA's reconstructed SSTs, version 3, from the Earth System Research Laboratory Physical Science Division and are available in network common data form (netCDF) format. NetCDF is a set of software libraries and data formats from the Unidata community that support the creation, access, and sharing of data arrays. Spatial climate data such as monthly SST grids are organized as arrays and stored in netCDF files by The National Center for Atmospheric Research (NCAR). The SST variable is an area-weighted average ( $^{\circ}\text{C}$ ) using values in  $5^{\circ}$  latitude-longitude grid boxes from the equator north to  $70^{\circ}\text{N}$  latitude and spanning the North Atlantic Ocean (Enfield et al., 2001) for the set of months starting with January 1854 and spanning through November 2009. This grid is shown in Figure 2.4. The data are available at [www.esrl.noaa.gov/psd/data/correlation/amon.us.long.data](http://www.esrl.noaa.gov/psd/data/correlation/amon.us.long.data). Further specifics about the SST data utilized in Chapter 5 are discussed when needed. This is the only study to utilize these data. A solid foundation has now been put in place for the upcoming three chapters. A discussion of the relevant literature has provided the background for this research, and the basic sources of data have been described to provide an initial framework. Each of the following three chapters represents statistical risk in a similar way using various spatial domains. Chapter 3 presents a research study about hurricane wind risk in twelve different Florida cities.

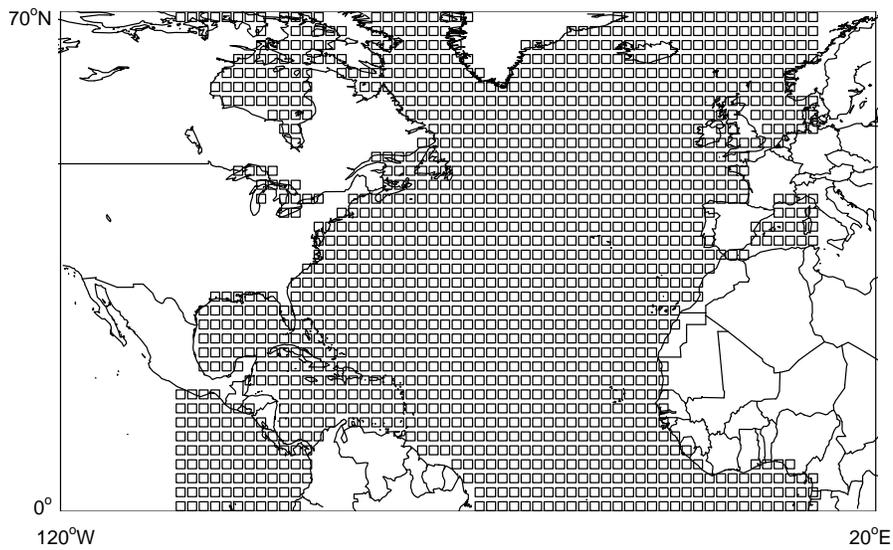


Figure 2.4: Location grid of NOAA reconstructed SST values. The locations are taken at  $5^\circ$  latitude-longitude grid boxes from the equator north to  $70^\circ\text{N}$  latitude and spanning the North Atlantic Ocean.

# CHAPTER 3

## RISK OF EXTREME HURRICANE WINDS IN FLORIDA

The state of Florida has a particularly unique vulnerability to hurricane strikes. Geographically, it sits surrounded by warm seas, with the Atlantic Ocean to the east, the Gulf of Mexico to the west, and both warm bodies connected by the Florida Straits. Other than the small northern boundary, the entire state has a coastline susceptible to hurricanes. According to Blake et al. (2007), 40% of all U. S. hurricanes hit Florida; and 83% of Category 4 or 5 hurricane strikes have hit either Florida or Texas. Due to increasing populations and economic development in the areas prone to devastation, average economic losses are increasing over time (Malmstadt et al., 2009). For society to better cope with and mitigate these disasters, a more precise estimate of the statistical risk of high winds on the local level is needed.

In this chapter, wind risk from a hurricane is estimated as the probability of a region experiencing winds of hurricane force or greater within a specified period. In this context, risk is an attribute of a region that is determined by the size and location of the region and not by the potential for loss of life or property. Risk is an unobserved and changing quantity that is estimated using a statistical procedure. The procedure provides an estimate of the exceedance wind speed (return level) as a function of return period (inverse of the annual probability). The primary goal here is to quantify the risk of hurricane winds of varying magnitudes blowing over a particular region over a given time period. This method can provide useful return level estimates for homeowners on a 20–30 year mortgage as well as provide information at similar time scales as offered for the 100-year flood plain in hydrology literature (Bin et al., 2008).

In particular, this chapter estimates return levels of hurricane winds affecting twelve Florida cities and examines the case for secular trends in the intensity and frequency of hurricanes in the vicinity of the state. This is an application of the Elsner et al. (2008a) model, where information from past cyclones affecting the city and its vicinity are combined through the parameters of an extreme value model to estimate return levels. This chapter goes beyond the initial work by automating the procedure and applying it to cities in Florida. This procedure, called the Hurricane

Risk Calculator (HRC), borrows information from a larger region to estimate the distribution parameters for a smaller region. The methodology implicitly assumes a stationary climate with regard to the occurrence and intensity of the wind events. Some light is shed on this assumption by examining the case for trends in hurricane activity in and around the state of Florida using the method of quantile regression as introduced in the study of hurricane climatology in Elsner et al. (2008b). This chapter can be seen in final publication in Malmstadt et al. (2010).

This chapter continues by discussing the statistical procedure for estimating local wind risk in Section 3.1. Section 3.2 gives the results of the return period calculations for selected Florida cities. Section 3.3 presents the methods and results of the trend analysis of hurricanes in the vicinity of Florida. Section 3.4 summarizes the results and emphasizes the important conclusions for this chapter. The work is important in providing reliable estimates of hurricane wind return periods for the population centers of Florida and for demonstrating a statistically significant trend in hurricane intensification rates in the vicinity of the state.

### 3.1 Estimation of Florida Wind Risk

All analysis and modeling was performed using the open source R Project for Statistical Computing software (R Development Core Team, 2010).

For this study, the 1-hourly wind speed estimates used for the years 1851–2008 are described in Chapter 2. As noted earlier, there are limitations to the data set that are even more prevalent in the earliest portion of the record. To test the effect of the limitations, this chapter compares results from the HRC using all of the available data with the results from a shorter, more reliable set of years, 1899–2008. As noted in Neumann (1987) estimated return levels assume the hurricane climatology over the record will remain unchanged. It is also assumed that the occurrence of a hurricane within the city is independent of future hurricanes within the same city. This latter assumption is reasonable, but as shown in Section 3.3 there is some evidence that Florida’s hurricane climate is changing.

The locations of the twelve cities used in this study were obtained from the United States Geological Survey’s database, available at <http://geonames.usgs.gov>, and describe the geographic center of the city limits as recognized by the federal government. Each city is chosen based on its level of population density, as well as its potential risk for hurricane damage. Twelve cities were chosen to map out the spatial differences for locations from the western Panhandle to the northeast coast of the state (Figure 3.1). The city center is considered the geographic center of the city limits as defined by the federal government. A 100 km radial circle is drawn around each center. This distance is used in the HRC. The 2000 U. S. Census information is provided to show Florida’s county population per square kilometer. Another way to map the Census data would be to show metropolitan statistical area population for the cities of interest.

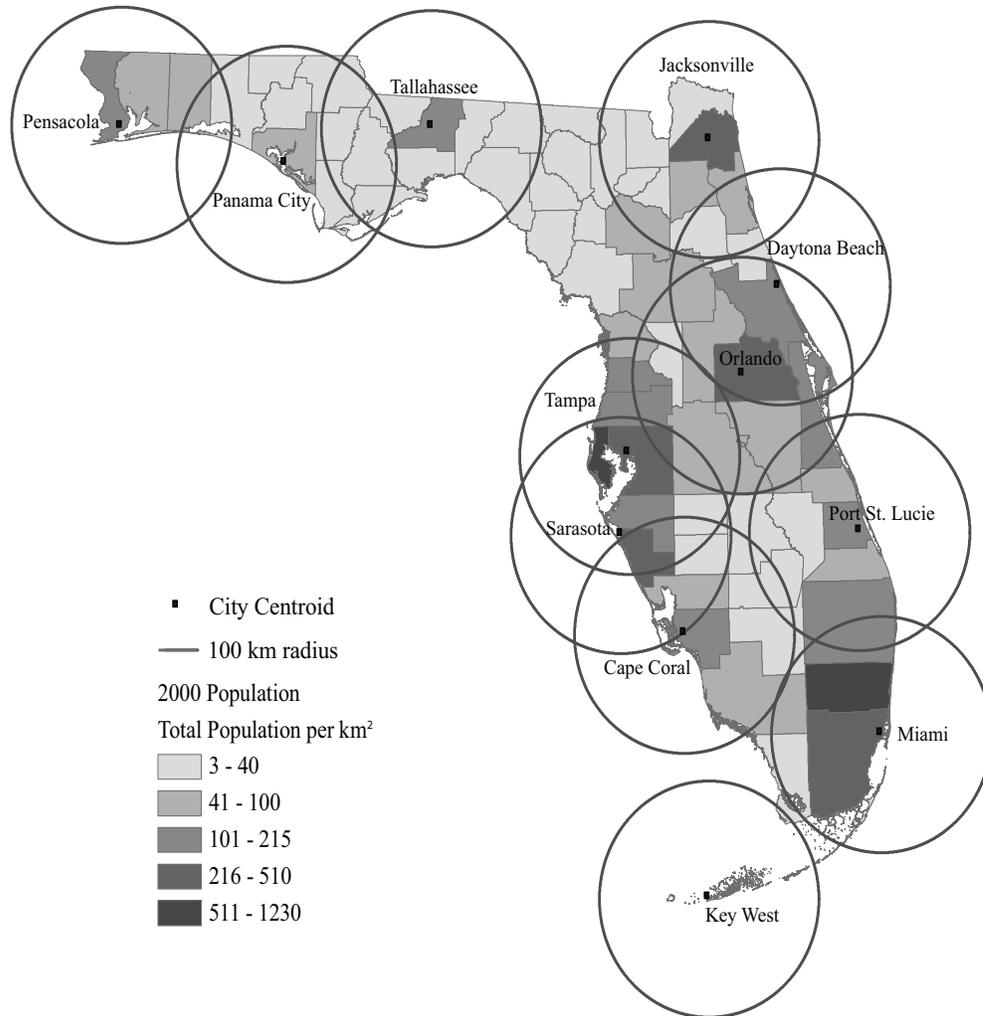


Figure 3.1: Florida cities. The twelve Florida cities used in this study are shown with the city centroid location denoted with a square. The 100 km radial distance from the city center is shown with a circle. This distance is used in the HRC and encompasses each city. County shading represents the population density per square kilometer.

The first portion of this study estimates return levels locally. The climatological probability of hurricane winds can be estimated in various ways, as was shown in Chapter 1. For this research, the approach of Elsner et al. (2008a) is automated to estimate local return levels for twelve cities in Florida. This is the HRC. Specifically, the distribution of the maximum wind above a threshold value  $u$  is assumed to follow a GPD (refer to Chapter 1). The threshold value for the extreme value model is determined using the mean residual life plot (Davison and Smith, 1990). This plot is produced by averaging the difference in the observed wind speeds above a specified level as a function of the level. For example, at a wind speed level of  $30 \text{ ms}^{-1}$ , 30 is subtracted from each observed wind speed and only the positive values (excesses) are averaged. This is repeated for all wind speed levels. The mean excess is the expected value of the amount that the observations exceed the particular level. In other words, if extreme values follow a GPD, then the expected value of the excesses is a linear function of the threshold.

Following Jagger and Elsner (2006) and the theorem discussed in Chapter 1, the exceedences  $W - u$  are modeled as samples from a family of GPD, so that for an individual hurricane with maximum wind  $W$ ,

$$\begin{aligned} \Pr(W > v | W > u) &= \left(1 + \frac{\xi}{\sigma}[v - u]\right)^{-1/\xi} \\ &= \text{GPD}(v - u | \sigma, \xi) \end{aligned} \quad (3.1)$$

where  $\sigma > 0$  and  $\sigma + \xi(v - u) \geq 0$ . For negative values of the shape parameter ( $\xi$ ) the GPD family of distributions has an upper limit of  $W_{\max} = u + \sigma_u/|\xi|$ .

The frequency of storms with intensity of at least  $u$  follows a Poisson distribution with a rate,  $\lambda_u$ , the threshold crossing rate. Thus the number of hurricanes per year with winds exceeding  $v$  is a thinned Poisson process with mean  $\lambda_v = \lambda_u \Pr(W > v | W > u)$ . This is the POT method and the resulting model is completely characterized for a given threshold  $u$  by  $\sigma$ , or scale,  $\xi$ , and  $\lambda_u$ ; the GPD parameters and the threshold crossing rate, respectively. It is important to note that the scale parameter, or  $\sigma$ , is a measure of the dispersion of the distribution, and does not refer to a type of geographic scale.

Since the number of storms exceeding any wind speed  $v$  is a Poisson process, the return period for any  $v$  has an exponential distribution, with mean  $r(v) = 1/\lambda_v$ . By substituting for  $\lambda_v$  in terms of both  $\lambda_u$  and the GPD parameters then solving for  $v$  as a function of  $r$  the corresponding return level for a given return period can be estimated as

$$\text{rl}(r) = u + \frac{\sigma}{\xi} [(r \cdot \lambda_u)^\xi - 1]. \quad (3.2)$$

For a more complete description of the statistical theory supporting this model please refer to Coles (2001), and for an example of the application of POT to maximum wind speeds in hurricanes refer to Jagger and Elsner (2006).

This model estimates how long on average an area can expect to wait to experience a hurricane of a given intensity. Unfortunately for most small areas like cities, there

are too few historical hurricanes to provide a stable estimate of the model parameters. Therefore, the parameters at the smaller radial distance are extrapolated using linear regression from multiple correlated sets of parameters taken at larger radial distances surrounding the location. An example of how this is done in the HRC is provided in the next section.

## 3.2 Return Period Estimates

The HRC procedure is outlined here for the city of Miami. The goal of interest is to find return level estimates for a 100 km radius around each city center. This allows for the city to be included, as well as a substantial portion of the areas outside of the city boundaries. Figure 3.2 shows the GPD model parameters and their 90% confidence intervals for increasing radial distances from the geographic center of the city from 75 to 200 km at increments of 25 km. Each parameter is estimated based on the statistical procedure described in the previous section and plotted as a function of radius from the city center. A confidence interval (CI) about the mean return level is obtained by scaling the CI obtained from using a larger radial distance (see Elsner et al. (2008a)). This is done by first, calculating the return levels for a given return period, and then assuming the estimate is asymptotically normal from a knowledge of the limiting variance of the estimator (delta method). Note that for radii larger than 125 km the change in parameter values is linear. This linear relationship is exploited and a separate bivariate regression for each parameter as a function of search radius between 125 and 200 km is determined. The regression line (solid) is used to extrapolate the parameter value at the radial distance of 100 km for each parameter (point indicated on regression line). Thus parameters at the small radial distance of interest are extrapolated using the parameters at larger radial distances.

The threshold  $u$  used in the model varies for each location depending on the mean residual life plot produced. The choice is a compromise between having wind speeds high enough for a proper GPD fit but enough of them that the parameter estimates of the GPD are reliable. The thresholds range from 20–35  $\text{ms}^{-1}$ . With a proper threshold, the values of  $\xi$  are greater than  $-1$  and the change as a function of radius is small. Figure 3.3 shows the  $\xi$  parameter from Miami’s estimates as an example. The dotted line shows the  $\xi$  parameter using a threshold that is too high for the model (40  $\text{ms}^{-1}$ ) and the solid line shows  $\xi$  using a threshold that is appropriate for the model (35  $\text{ms}^{-1}$ ). Although the choice of the precise threshold is somewhat arbitrary, for each location the choice made is not overly sensitive to the search radius.

The return level curve for Miami using data from 1851 and parameters extrapolated to a radius of 100 km is shown in Figure 3.4. The return period in years is on the horizontal axis with a log scale. The return level indicating wind speeds ( $\text{ms}^{-1}$ ) is on the vertical axis with a linear scale. The 90% uncertainty range on the return levels are shown with vertical lines. The model curve indicates that Miami (within a 100 km radius of the city center) can expect a hurricane of at least 33  $\text{ms}^{-1}$ , with

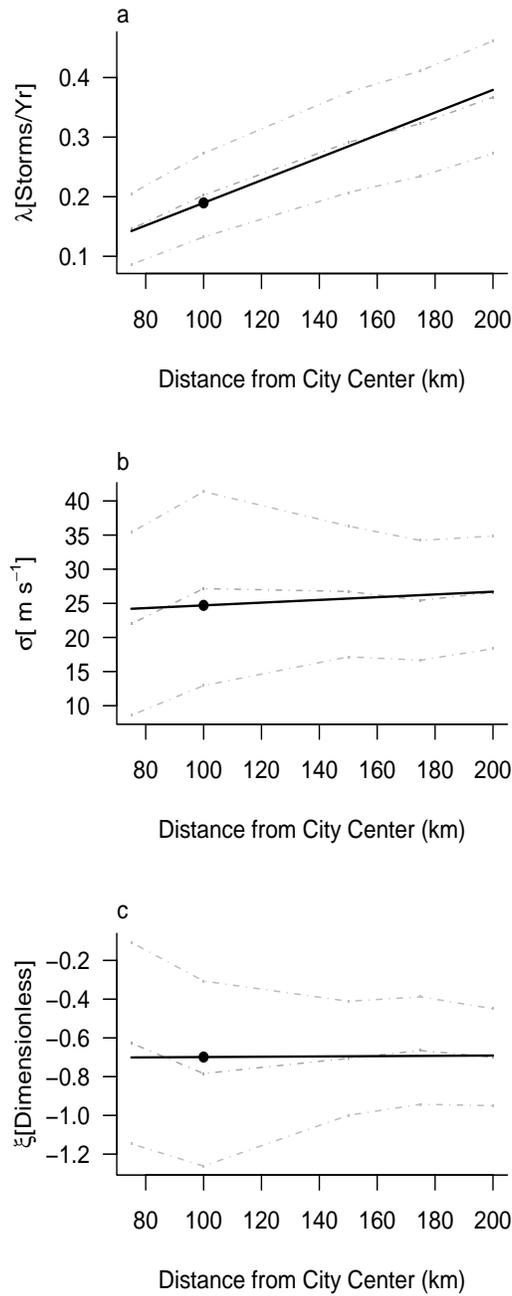


Figure 3.2: HRC parameters. The values for the HRC parameters ( $\lambda$ ,  $\sigma$ , and  $\xi$ ) as a function of radial distance are shown. Parameter values are obtained from a maximum likelihood procedure using the fastest wind speeds from hurricanes occurring within a radial distance of the geographic center of Miami. The bold line indicates the regression line of the parameter on distance from city center.

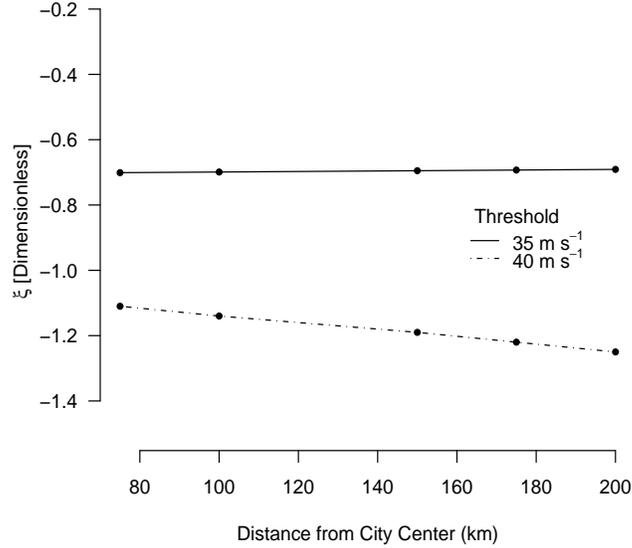


Figure 3.3: Threshold comparison. Miami’s  $\xi$  parameter as a function of radial distance is shown. The dotted line shows the parameter using a  $40 \text{ ms}^{-1}$  threshold. The solid line shows the parameter using a  $35 \text{ ms}^{-1}$  threshold.

a 90% uncertainty of (26.9–39.1) on average once every 5 years and a hurricane of at least  $50 \text{ ms}^{-1}$  (45.5–54.5) on average once every 12 years. Since the estimate is over an area and wind speeds diminish rapidly as the hurricane moves over land, it most accurately represents the portion of the area over (or nearest) the ocean. For comparison, the model is rerun using hurricanes limited to the period 1899–2008. The return-level curve is similar (dashed line). The estimates using data from 1899 produce higher return levels for return periods in the range from about 5 to 100 years. However, the estimates fall within the 90% confidence level produced by using data from 1851. This justifies using the entire record of data.

The HRC is repeated for the other eleven cities. Table 3.1 displays the number of hurricanes passing through each city over the period 1851–2008, inclusive. The pass through is defined by the center fix passing within a 100 km great circle distance of the geographical center of the city. The hurricane wind speeds are broken into the five Saffir-Simpson categories and show variation between locations. Each city is threatened by hurricane winds and all have experienced at least Category 3 winds ( $50\text{--}58 \text{ ms}^{-1}$ ), except Tallahassee. Pensacola has experienced the most Category 1 hurricanes. Overall, Miami has experienced the most hurricanes, with 35 strikes over the 158-year time period, and Key West has the second highest strike occurrence, 33. Miami, Cape Coral, Port St. Lucie, and Key West have experienced Category 4 wind

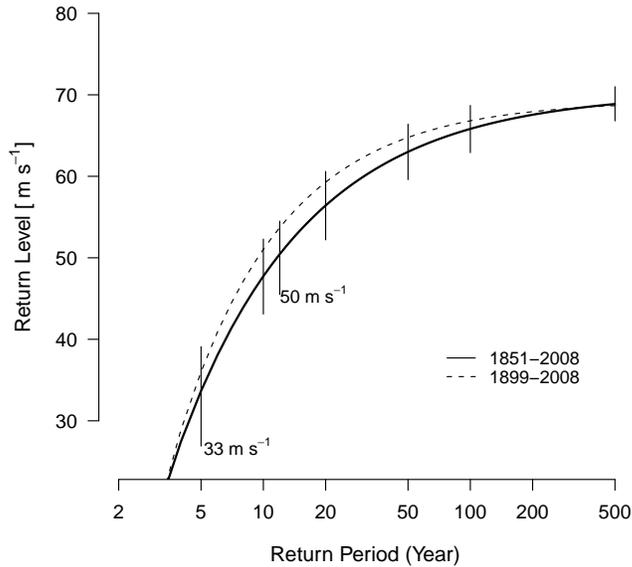


Figure 3.4: Miami’s return level. The return level as a function of return period for hurricanes passing within 100 km of the geographic center of Miami. The solid line shows the return level estimates from the HRC using data over the period 1851–2008. The 90% confidence intervals are shown as vertical lines. The dotted lines shows the return-level estimates from the HRC using data over the period 1899–2008.

speeds. Only Miami has experienced a Category 5 hurricane (Andrew in 1992). The Labor Day hurricane of 1935 reached Category 5 strength, but at a distance greater than 100 km from Key West.

Looking at these results from a geographic perspective, it is seen that the “Big Bend” area of Florida (Tallahassee through Tampa, along the Gulf coast) experiences fewer hurricanes than the western Panhandle and southeastern peninsula. Also, the northeast corner of the state, Jacksonville and Daytona Beach specifically, experience fewer hurricanes than these areas as well. The peninsula of Florida protects the Big Bend region from the strongest hurricanes approaching from the southeast, and the southeast to northwest orientation of the coastline north of Port St. Lucie is parallel to the tracks of hurricanes at this longitude and latitude.

Figure 3.5 shows return-level curves for the 100 km radial distance surrounding each city center. They are plotted based on their geographic locations to help visualize spatial patterns. As expected, areas like the Big Bend and the northeast coastline show weaker wind speeds for a given return period when compared to areas like the Panhandle and the southeastern coastline. The highest return levels for return

Table 3.1: Number of hurricanes by wind speed intervals known to have passed within a 100 km radius of the specified city center from 1851–2008. Positions and wind speeds interpolated to 1-hourly values following the procedure outlined in Jagger and Elsner (2006).

Maximum Wind Speed ( $\text{ms}^{-1}$ )	Number of Hurricanes			
	Miami	Orlando	Jacksonville	Tampa
33–42	12	15	7	10
43–49	7	7	6	6
50–58	11	3	5	4
59–69	4	0	0	0
$\geq 70$	1	0	0	0
	Tallahassee	Cape Coral	Port St. Lucie	Daytona Beach
33–42	8	4	10	13
43–49	6	5	9	4
50–58	0	8	6	5
59–69	0	3	2	0
$\geq 70$	0	0	0	0
	Pensacola	Sarasota	Panama City	Key West
33–42	13	10	17	11
43–49	4	6	4	10
50–58	8	6	6	9
59–69	0	0	0	3
$\geq 70$	0	0	0	0

periods of 500 years are noted in Key West, Miami, Port St. Lucie and Cape Coral. In contrast, Tallahassee has the lowest return levels at the longest return periods. These results are tabulated in Table 3.2 including the 90% CI on the return-level estimates.

Table 3.3 shows the return level estimates for Miami and Pensacola with their confidence intervals using data only back to 1899. The results using the entirety of the data set are presented alongside for comparison. The differences can be attributed to the exclusion of an additional 48 years of data and to data limitations. It should be noted, however, that the differences are not large relative to the range of the CI. In fact, the return level estimates using the shorter data period fall within the CI of the return level estimates from the longer period.

Based on the HRC, a maximum wind speed can be estimated using the equation  $u + (\sigma/|\xi|)$ . Table 3.4 shows the maximum wind speed, the GPD parameters, and the threshold for each city. This wind speed can be understood as the highest possible

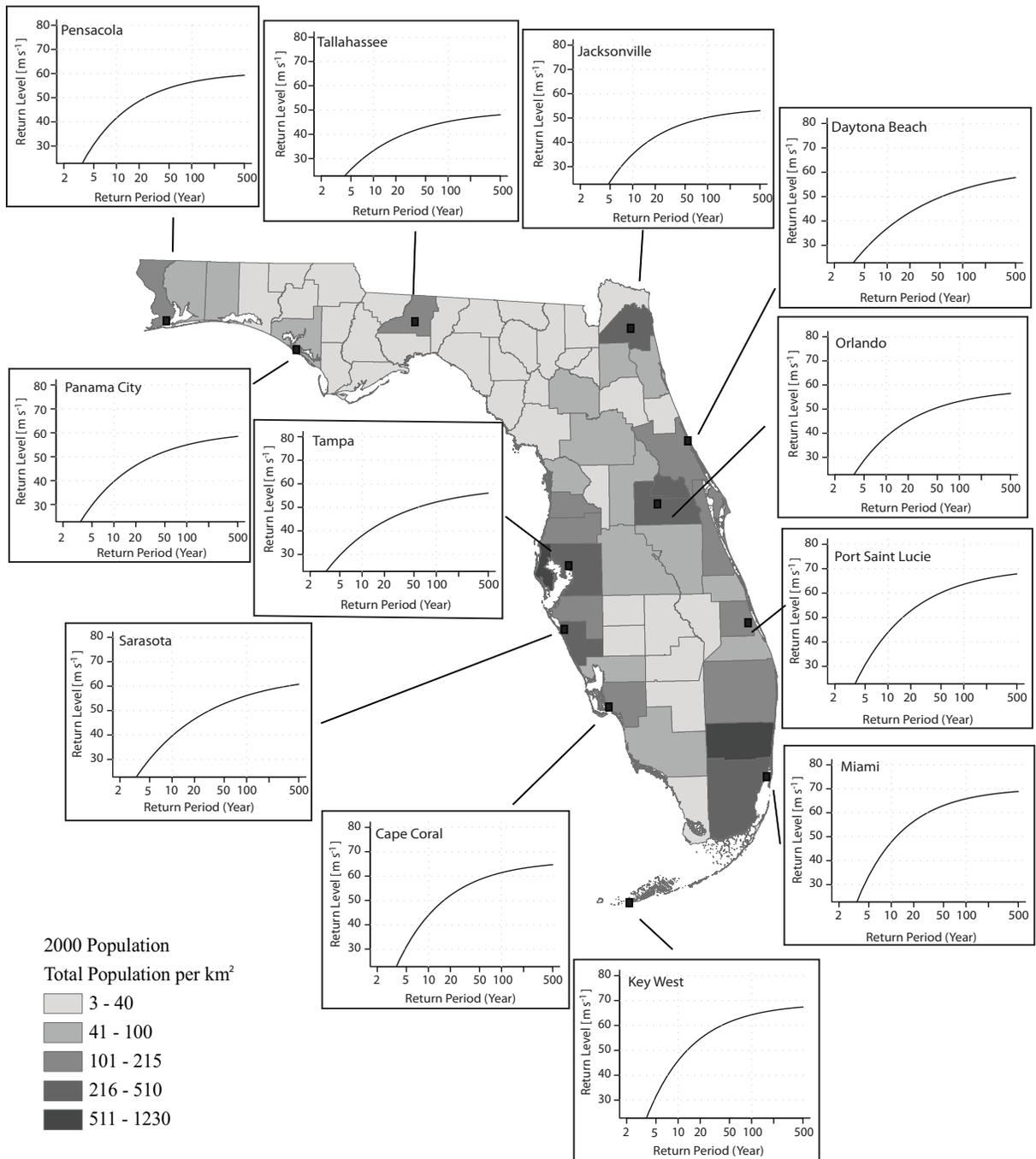


Figure 3.5: Florida cities' return level curves. Return level curves are shown as a function of return periods for hurricanes passing within 100 km of the geographic center of the specified city using the HRC.

Table 3.2: Return period for selected wind speed exceedance values. The return levels are based on the HRC presented above and include data over the period 1851–2008. The return period refers to the expected time between winds of this level blowing somewhere within a 100 km radius of specified cities. The confidence interval is the 90% interval.

Return Period (Year)	<b>Hurricane Wind Speed</b> ( $\text{ms}^{-1}$ )			
	Miami	Orlando	Jacksonville	Tampa
10	47.7 (43.1–52.3)	38.5 (34.5–42.5)	35.1 (32.1–38.1)	38.0 (34.6–41.4)
20	56.4 (52.2–60.6)	45.1 (41.4–48.8)	42.2 (38.9–45.5)	44.0 (40.7–47.3)
50	63.0 (59.6–66.4)	50.6 (46.9–54.3)	47.8 (44.5–51.1)	49.5 (46.3–52.7)
100	65.8 (62.9–68.7)	53.2 (49.7–56.7)	50.3 (47.0–53.6)	52.3 (49.3–55.3)
500	68.9 (66.8–71.0)	56.5 (53.2–59.8)	53.0 (49.7–56.3)	56.1 (53.4–58.8)
	Tallahassee	Cape Coral	Port St. Lucie	Daytona Beach
10	33.2 (30.3–36.1)	43.9 (39.8–48.0)	43.9 (39.9–47.8)	37.0 (33.7–40.3)
20	38.5 (35.5–41.5)	52.1 (48.4–55.8)	52.8 (49.0–56.6)	43.5 (40.4–46.7)
50	43.0 (40.1–45.9)	58.5 (55.5–61.5)	60.1 (56.7–63.5)	49.7 (46.9–52.5)
100	45.2 (42.4–48.1)	61.4 (58.9–63.9)	63.6 (60.5–66.7)	53.0 (50.6–55.4)
500	48.0 (45.2–50.8)	64.7 (62.8–66.7)	67.9 (65.1–70.7)	57.8 (55.8–59.8)
	Pensacola	Sarasota	Panama City	Key West
10	41.6 (38.1–45.2)	39.6 (35.9–43.3)	39.6 (36.2–43.0)	45.8 (41.1–50.5)
20	48.5 (45.3–51.7)	46.5 (43.0–50.0)	46.4 (42.9–49.9)	54.7 (50.5–58.9)
50	54.0 (51.3–56.7)	52.8 (49.7–55.9)	52.1 (48.8–55.4)	61.4 (57.8–65.0)
100	56.5 (54.2–58.8)	56.2 (53.5–58.9)	54.9 (51.1–57.3)	64.3 (61.1–67.5)
500	59.3 (57.4–61.2)	60.8 (58.7–62.9)	58.5 (55.7–61.3)	67.4 (64.9–69.9)

Table 3.3: Return period for selected wind speed exceedance values for Miami and Pensacola. The return levels include data over the period 1899–2008. The confidence interval is the 90% interval. The results using the full data set are presented alongside for comparison.

Return Period (Year)	<b>Hurricane Wind Speed</b>			
	$(\text{ms}^{-1})$			
	1851–2008		1899–2008	
	Miami	Pensacola	Miami	Pensacola
10	47.7 (43.1–52.3)	41.6 (38.1–45.2)	51.0 (45.5–56.5)	42.9 (38.0–47.8)
20	56.4 (52.2–60.6)	48.5 (45.3–51.7)	59.3 (56.0–62.6)	50.4 (46.2–54.6)
50	63.0 (59.6–66.4)	54.0 (51.3–56.7)	64.8 (62.9–66.7)	55.3 (52.0–58.6)
100	65.8 (62.9–68.7)	56.5 (54.2–58.8)	66.8 (65.4–68.2)	57.0 (54.2–59.8)
500	68.9 (66.8–71.0)	59.3 (57.4–61.2)	68.7 (67.8–69.6)	58.6 (56.3–60.9)

Table 3.4: Estimated maximum possible wind speeds. Estimates are for the strongest winds ( $\text{ms}^{-1}$ ) within 100 km of the city center.

City	$\sigma$ ( $\text{ms}^{-1}$ )	$\xi$	Threshold ( $\text{ms}^{-1}$ )	Max Wind Speed ( $\text{ms}^{-1}$ )
Port St. Lucie	26.5	-0.581	25	70.6
Miami	24.7	-0.699	35	70.3
Key West	23.7	-0.698	35	69.0
Cape Coral	27.0	-0.652	25	66.4
Sarasota	18.3	-0.459	25	64.9
Daytona Beach	15.9	-0.422	25	62.7
Panama City	14.2	-0.547	35	61.0
Pensacola	23.2	-0.648	25	60.8
Tampa	16.5	-0.480	25	59.4
Orlando	13.5	-0.569	35	58.7
Jacksonville	16.2	-0.662	30	54.5
Tallahassee	16.3	-0.541	20	50.1

wind speed that can be experienced within a 100 km distance of the given city center. The list is in descending order of maximum wind speeds. At the top are Port St. Lucie and Miami with maximum wind speeds exceeding  $70 \text{ ms}^{-1}$  and at the bottom are Jacksonville and Tallahassee with wind speeds less than  $55 \text{ ms}^{-1}$ .

The NWS also estimates statistical hurricane wind risk for selected Florida cities. They use the HURISK model developed in Neumann (1987), which employs a Weibull distribution for tropical cyclone wind speeds and a Poisson distribution for hurricane frequency and estimates return periods at the county level. The latest output from HURISK is presented in Blake et al. (2007). The HRC estimates a return level for a given return period while the HURISK provides a return period for a given return level. Comparisons are made by converting our HRC return levels at  $33$  and  $50 \text{ ms}^{-1}$  into return periods (Table 3.5). The HRC estimates can be converted to provide a return period for a fixed return level via Equation 3.2.

The two procedures produce similar estimates for hurricane intensities. In four of the ten locations, HURISK and HRC give the same return period and nine out of ten are within one year. At major hurricane intensities the results diverge with the HRC indicating longer return periods (lower probabilities) for the stronger storms by an average of 16 years (an average difference in annual probability of 6%). The other cities (i.e. Tallahassee and Orlando) are not included in this table because the latest output from HURISK does not include a comparison. The differences in the results are attributable to differences in the wind speed distributions used by the two approaches. The Weibull distribution used in HURISK is appropriate for wind

Table 3.5: Return period comparisons. Return periods (year) from HURISK and HRC are shown for selected Florida cities. Hurricanes are denoted by H and major hurricanes are denoted by MH. St. Lucie County is included to show a comparison between the return periods for Fort Pierce (HURISK) and Port St. Lucie (HRC). Lee County is included to show a comparison between Fort Myers (HURISK) and Cape Coral (HRC). “Differ” is the return period from HURISK minus the return period from HRC.

	Return Period (Year)					
	HURISK		HRC		Differ	
	H	MH	H	MH	H	MH
Pensacola	7	17	6	24	1	-7
Panama City	6	17	6	34	0	-17
Tampa	6	23	7	56	-1	-33
Sarasota	6	19	6	32	0	-13
Key West	5	12	5	13	0	-1
Miami	4	9	5	12	-1	-3
Daytona Beach	8	31	7	52	1	-21
Jacksonville	9	28	9	92	0	-64
St. Lucie Cty	7	17	5	16	2	1
Lee Cty	6	15	5	16	1	-1

speeds at all intensities, however by using the more frequent weaker hurricanes the distribution tends to produce a heavy tail (Darling, 1991; Jagger et al., 2001). A heavy tail is indicative of higher probabilities for the most extreme speeds. These higher probabilities translate to shorter return periods as noted in the table for major hurricanes. Thus while HURISK is appropriate for estimating the wind risk of typical hurricanes, it may overestimate the threat of winds in the most extreme hurricanes. For estimating the return period of the strongest hurricane wind speeds it might be better to use a GPD as is done in the HRC.

Having examined hurricane return levels across the state, next it is considered whether the statistical wind risk from hurricanes is changing over time. A stationary climate is assumed in the results presented in this section. This assumption is examined in the following section by considering evidence for changes in the frequency and intensity of hurricanes in the vicinity of Florida.

### 3.3 Changes in Frequency and Intensity

The usefulness of a return level estimate depends to some extent on the assumption that the hurricane climate is stationary. Explicitly, the methodology of estimating the return levels presupposes that the frequency and intensity of hurricanes is not changing over time. Trends in frequency are examined using linear regression and trends in intensity are examined using quantile regression (Koenker and Bassett, 1978) as used recently in Elsner et al. (2008b). Quantile regression is an extension of linear regression applied to quantiles of the response variable. A quantile is a point taken from the inverse cumulative distribution function so that the 0.5 quantile is the value such that 50% of the values are less than the value.

Because of the rarity of hurricanes affecting any small area the uncertainty on a trend estimate will be large. So here the data are examined for trends in the frequency and intensity of hurricanes falling within a 1000 km radius of 26°N latitude and 82°W longitude off the southwest coast (Figure 3.6). This is the area described by Hamid et al. (2009) in an attempt to identify storms that could conceivably impact Florida. Possible effects of climate change could influence hurricanes that pass within this vicinity of Florida, so including a wide search radius around Florida allows for insight into the potential changes that a warmer climate may bring.

The Florida Commission on Hurricane Loss Projection Methodology, created during the 1995 Florida Legislative session to evaluate hurricane risk models, advises modelers to consider hurricanes in this area as a potential threat to Florida. This is done for the purposes of comparing different statistical risk models used by the insurance industry (Hamid et al., 2009; Jagger and Elsner, 2009) using a common set of hurricanes. The area encompasses cyclones that affect the Panhandle, west, and northeast coasts of Florida, as well as cyclones that approach south Florida from the vicinity of Cuba and the Bahamas. This location and data back to 1899 are used because, according to Landsea et al. (2004), the turn of the 20th century is the ap-

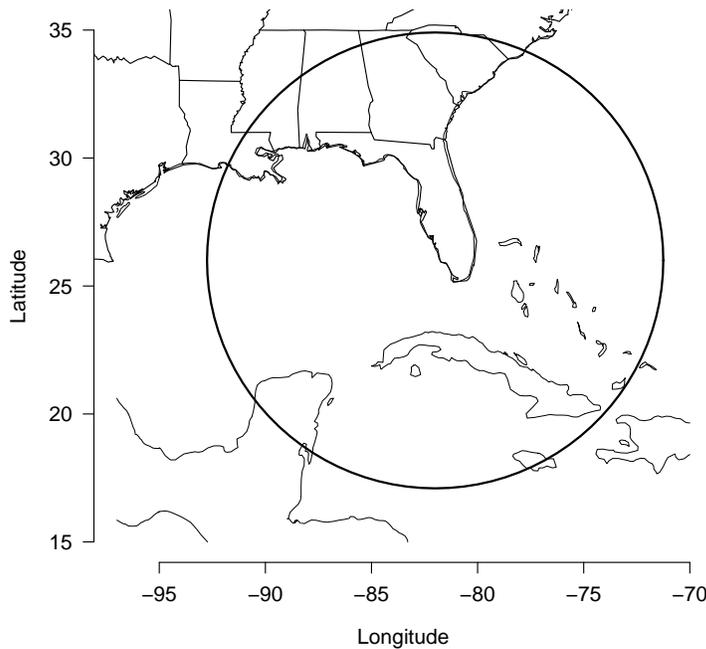


Figure 3.6: Florida threat area. The threat area around Florida is shown with a black circle. This area is described by Hamid et al. (2009) as a 1000 km radius of a location (26°N latitude and 82°W longitude) off the southwest coast of Florida.

appropriate place to start for reliable landfall records for the United States. The data include 162 individual hurricanes over the period 1899–2008, inclusive.

The frequency and intensity of hurricanes as a function of year are shown in Figure 3.7. There is no upward or downward trend in the occurrence rate of hurricanes or major hurricanes in the vicinity of Florida, but there does appear to be a low frequency fluctuation that matches variations in SSTs (Saunders and Harris, 1997; Landsea et al., 1999; Wang et al., 2008). The strongest hurricanes appear to be getting stronger. The trend lines represent the 50<sup>th</sup> and 90<sup>th</sup> percentile regressions of maximum wind speed on year. The maximum wind speed is the highest wind speed for each hurricane within the Florida threat area. The upward trend at the median is  $7 \text{ ms}^{-1}$  per century with a standard error of  $3 \text{ ms}^{-1}$  per century. The upward trend at the 90<sup>th</sup> percentile is  $13 \text{ ms}^{-1}$  per century with a standard error of  $6 \text{ ms}^{-1}$  per century. If only the years from 1950 onward are included, the upward trend increases to  $21 \text{ ms}^{-1}$  per century with a standard error of  $20 \text{ ms}^{-1}$  per century. The trends assume the data are uniformly reliable over the period of record. This assumption may not hold up under closer scrutiny of the influence that changes in measuring techniques

have on the wind estimates. These results suggest that the return levels presented in the previous section might be too low for estimates of future hurricane wind risk.

To better understand the upward trends in hurricane wind speeds, the hourly intensification rates along the hurricane tracks within the Florida threat area are estimated. The intensification is the time change of the maximum hurricane wind speed. The data set is divided into values that are above zero indicating hurricane strengthening and values below zero indicate hurricane weakening. Then the maximum value of strengthening (intensification) along the track of each hurricane within the threat area is taken to obtain one intensification value per hurricane. These values are multiplied by 24 to express the maximum hourly intensification in terms of a daily rate and plot them as a function of year in Figure 3.8. The 10<sup>th</sup>, 50<sup>th</sup>, and 90<sup>th</sup> percentiles are analyzed for trends using quantile regression. Significant upward trends in intensification rates are observed. The greater intensification rates of late do not necessarily imply more intense hurricanes at the point of landfall. These upward trends are consistent with increasing ocean heat content noted over the Gulf of Mexico and the western Caribbean (Carton and Santorelli, 2008). This relationship between hurricanes and SSTs is further explored in Chapter 5.

### 3.4 Florida Wind Risk in Summary

Hurricanes top the list of most destructive and costly natural disasters in the United States. Estimates of return levels for the strongest hurricane winds provide emergency planners and the insurance industry information by which to make decisions. In this chapter, standard hurricane data are used to model wind speed return levels for the strongest hurricanes passing within a 100 km radius of twelve cities in Florida. The model is based on the GPD with parameters estimated locally by borrowing information on the frequency and intensity of hurricanes over larger regions. This procedure is termed the HRC.

Not unexpectedly, it was found that the likelihood of extreme wind occurrence from hurricanes varies across the state. Areas in the northeast, such as Jacksonville, and in the Big Bend, between Tampa and Tallahassee, have longer return periods for a given strong wind speed compared with areas such as Miami in the southeast and Pensacola in the Panhandle. Thus the annual threat of a catastrophic hurricane event is highest in the southeast and the western Panhandle of the state.

The sensitivity of the return level estimates was examined by limiting the input data to the more recent years (1899 and after). The return levels are quite similar indicating that, despite the greater level of uncertainty on the records prior to 1899, the frequency and intensity of the strongest hurricanes are consistent over the longer period of record.

An implicit assumption underlying the results is that of stationarity, which was examined by trend analysis on hurricane frequency and intensity in the vicinity of Florida. It was found that the frequency of hurricanes and major hurricanes is rather

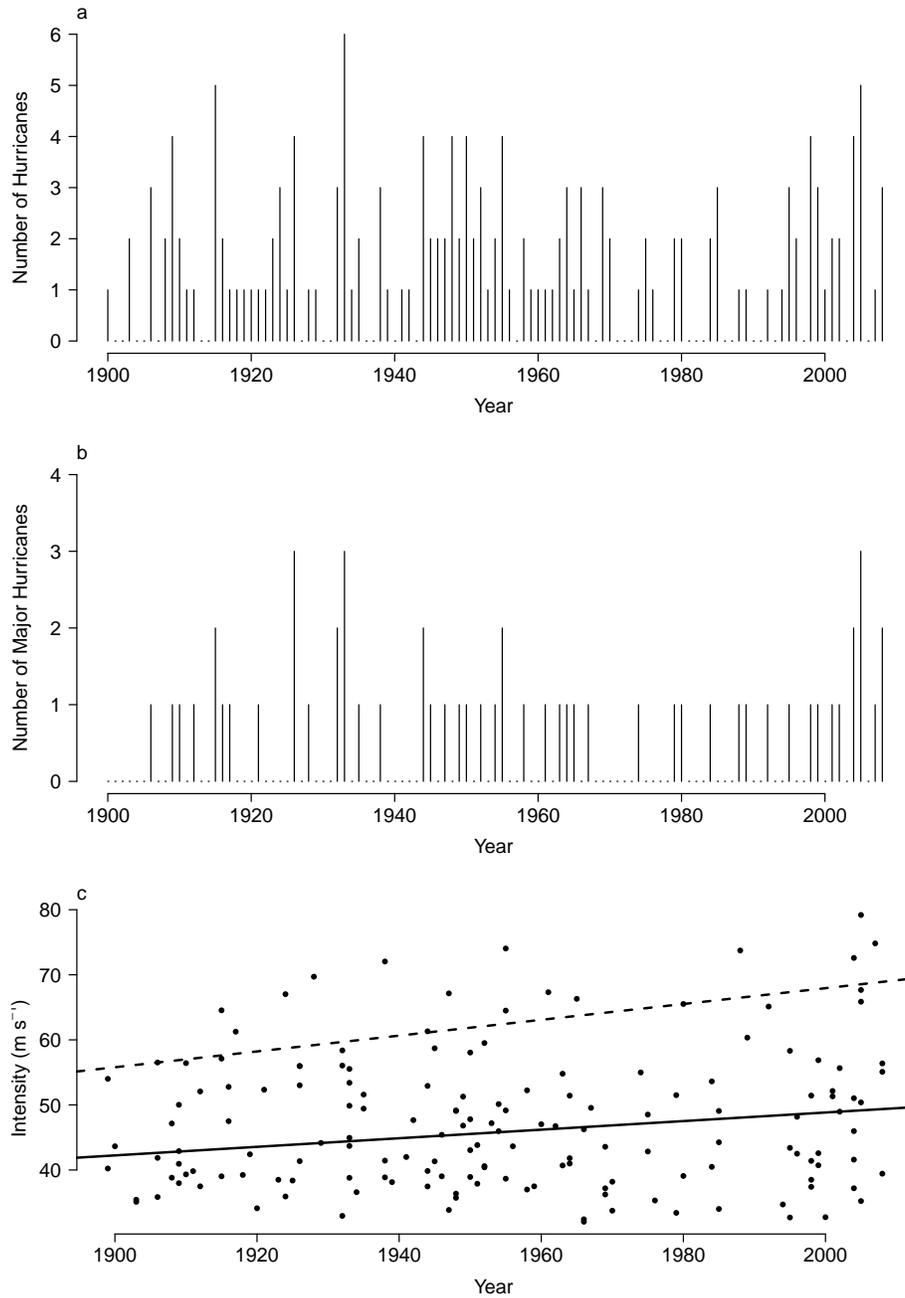


Figure 3.7: Hurricane frequency and intensity. Frequency and intensity of hurricanes in the vicinity of Florida are shown. (a) Frequency of all hurricanes as a function of time. (b) Frequency of major hurricanes (Category 3 and above) as a function of time. (c) Intensity (maximum sustained wind speed in  $\text{m s}^{-1}$ ) as a function of time with trend lines shown for the 50<sup>th</sup> and 90<sup>th</sup> percentiles.

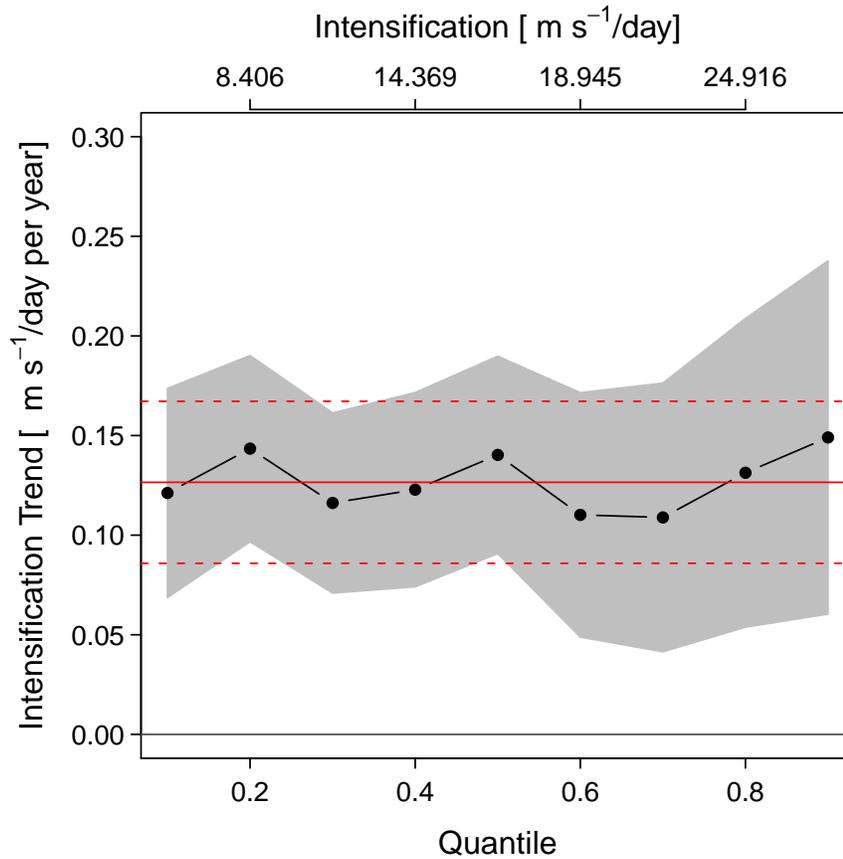


Figure 3.8: Trends in the maximum intensification rate of hurricanes as a function of quantile. The maximum intensification rate is calculated as the rate of change between hourly observations. Trends are estimated coefficients from quantile regression in units of  $\text{ms}^{-1}/\text{hr}$  per year. The point-wise 90% confidence band is shown in grey. The solid line is the trend from a least-squares regression of the rate of intensification as a function of year, and the dashed lines delineate the 90% confidence band about this trend.

constant through time. However, there is evidence of upward trends in the intensity of the strongest hurricanes. In fact at the 90<sup>th</sup> percentile the upward trend is estimated at  $13 \pm 6 \text{ ms}^{-1}$  per century. This is occurring in conjunction with statistically significant upward trends in the intensification rates and is consistent with rising ocean heat content over the Gulf of Mexico.

The study can be improved by determining a method to adjust the return level estimates to reflect the upward trends in hurricane intensity. One possibility is to model hurricane frequency as a function of time or sea temperature and integrate the modeled frequencies over, say 100 years. The ratio of the modeled frequency to the historical frequency could provide a scaling factor for the rate parameter of the Poisson distribution in the HRC. A similar adjustment could be made for the parameters of the GPD.

In the next chapter, the GPD approach is used to estimate return levels for hurricane storm surge at three AFB locations along the U. S. Gulf coast. It furthers the study by exploring the relationship between storm surge *and* wind speed within hurricanes using a bivariate copula model.

# CHAPTER 4

## COMBINED WIND AND SURGE RISKS FROM HURRICANES ALONG THE U. S. GULF COAST

The previous chapter showed that hurricanes create extreme winds in areas around Florida, and that these winds are becoming more extreme over time. Although this is certainly an important characteristic to model, it is also interesting to consider another principal hazard of hurricanes; the storm surge. Storm surge refers to the wall of water that is pushed onshore as the momentum of the hurricane tracks inland. Storm surge is a unique variable to model because it is highly dependent upon the topography and terrain of the surface it passes over. It was not monitored as closely as wind speeds in the earlier portion of the 20<sup>th</sup> century creating relatively short observational records. Up until recently, there was not an available raw data set of sufficient length that allowed for the modeling of storm surge at a given location. Thankfully, a global storm surge dataset has been created that allows for statistical analysis (refer to Chapter 2).

In previous research, the univariate statistical risk of hurricane characteristics has been calculated in various ways. Most notably, researchers have employed EVT to estimate the occurrence of variables beyond those that have been recorded in known data sets, as was shown in Chapter 3. In the case of this study, EVT is used to estimate the occurrence of hurricane storm surge events at three locations along the Gulf coast. In particular, a POT approach and the GPD are used to estimate the likelihood of occurrence. The analysis is focused on three locations; Eglin Air Force Base (EAFB) near Valparaiso, Florida (30.46°N, 86.55°W), MacDill Air Force Base (MAFB) in Tampa, Florida (27.85°N, 82.50°W), and Keesler Air Force Base (KAFB) in Biloxi, Mississippi (30.41°N, 88.93°W).

It is also a goal of this study to understand the relationship between hurricane wind speed and storm surge at these locations. To get an estimate of the overall risk of losses, Liu et al. (2009) use a multivariate compound EVD to combine the risk of typhoon hazards in China. Here another way to combine the likelihood of a wind and surge occurrence from hurricanes is presented using a bivariate statistical model.

In particular, data from hurricane events with high winds and storm surges are used in a copula model. The copula model allows for the estimation of the dependency between high winds and surge. In this chapter, a proof-of-concept for the approach is offered using data in the vicinity of the three AFBs.

In this chapter, Section 4.1 describes how the data described in Chapter 2 are further subset and collated for the AFBs. Section 4.2 describes the methodology and results for the univariate risk model using storm surge. Section 4.3 describes the statistical relationship between wind *and* surge and presents the methodology and results for the bivariate copula risk model. Section 4.4 summarizes the important points of this chapter.

## 4.1 Collating Wind and Surge Events

For each of the three locations mentioned above, the hurricane events ( $\geq 33 \text{ ms}^{-1}$ ) that passed within a 140 km radius of the geographic centers of the AFBs are collected using the entire available HURDAT record (again, using the 1-hourly interpolated values). This radius is used because, according to Keim et al. (2007), the 140 km radius is one-half of the typical width of the swath of hurricane force winds for hurricanes that exceed  $50 \text{ ms}^{-1}$ . This subset of hurricanes is used at each location to find the matching peak surge location and height for events in SURGEDAT.

A 140 km radius is drawn around each surge event to find the corresponding maximum wind speed along the hurricane track that has come within that specified distance. This is the maximum wind speed closest to the recorded surge event. Each location now has a sample of hurricane occurrence, location, wind speed, and surge height data. For EAFB and MAFB, the years of data span from 1896–2004. For KAFB, the years of data span from 1893–2005. No hurricane events are included outside of these ranges due to a lack of observational surge data. Table 4.1 shows an example of five hurricane wind and storm surge pairs that affected one of the locations; KAFB. It includes the name and year of the hurricane, the surge height in meters, the longitude (Slon) and latitude (Slat) of the surge height, the maximum wind speed nearest the surge in  $\text{ms}^{-1}$ , and the longitude (Wlon) and latitude (Wlat) of the wind speed.

As a note, the 140 km radius around EAFB could be changed to an oval due to the rectangular shape of the base. This could be considered in future studies when modeling storms around EAFB. It may allow for additional hurricanes to be considered.

Figure 4.1 is a conceptual model describing the data collation. The light gray rectangle denotes an arbitrary piece of coastline. All hurricanes (arrow) that come within 140 km of the AFB (dark gray rectangle) are collected. All of the surge events (blue square) that are associated with those tracks are collected. A 140 km radius (blue circle) is drawn around the surge event to find the closest maximum wind speed (green triangle) along the track (circle not drawn to scale). Note this maximum wind

Table 4.1: Sample of data used for analysis. Includes five hurricanes that affected KAFB, including the name, year, surge height, surge location, wind speed, and wind location.

Name	Year	Surge (m)	Slon	Slat	Wind ( $\text{ms}^{-1}$ )	Wlon	Wlat
Katrina	2005	8.47	-89.3	30.3	57.3	-89.6	29.3
Ivan	2004	4.57	-87.3	30.3	55.2	-88.1	29.4
Danny	1997	1.99	-87.8	30.2	36.5	-89.1	29.7
Florence	1988	1.83	-89.9	30.3	35.1	-89.3	28.9
Elena	1985	3.05	-85.0	29.7	56.7	-86.4	29.5

speed is different than the maximum wind speed at landfall (red circle). It also may not be the maximum wind speed the hurricane produced over its entire life cycle. Using the data, two sets of distances are calculated. D1 is the distance between the recorded wind event and the recorded surge event (the distance between the green triangle and the blue circle). D2 is the distance between the surge event and the AFB (the distance between the blue circle and the dark gray rectangle). These will be used in various regression models later to test the significance. As a note about the schematic: the shapes of the variables are arbitrary, and the surge and wind locations are point observations. The polygons used here are simply for representation.

An index is added to the data that notes the position of the maximum wind speed relative to the surge location. If the wind speed occurred to the left of the surge location, it was given an index of 1. If the wind speed occurred to the right of the surge location, it was given an index of 0. It is well known that the highest magnitude of wind speeds is in the right, front quadrant of the hurricane. This index is used in later regression analyses as an attempt to test that relationship.

Figure 4.2 shows the study area at EAFB, MAFB, and KAFB. For EAFB, the boundary of the AFB is shown with a red polygon. For MAFB and KAFB, the AFB is very small so they are denoted by single red point at the base's center location. The storm surge event locations are shown with squares. An inset map is included to show the global location of these surge events. The inset map includes a bounding box in the southeastern United States showing where, specifically, the surge events come from. There are 20, 18, and 17 surge events recorded within 140 km of the EAFB, MAFB, and KAFB central location, respectively. At KAFB, a few of the locations are duplicates so the number of squares is less than the number of surge events.

## 4.2 Storm Surge Risk

As before, all analysis and modeling was performed using the open source R Project for Statistical Computing software (R Development Core Team, 2010). Specif-

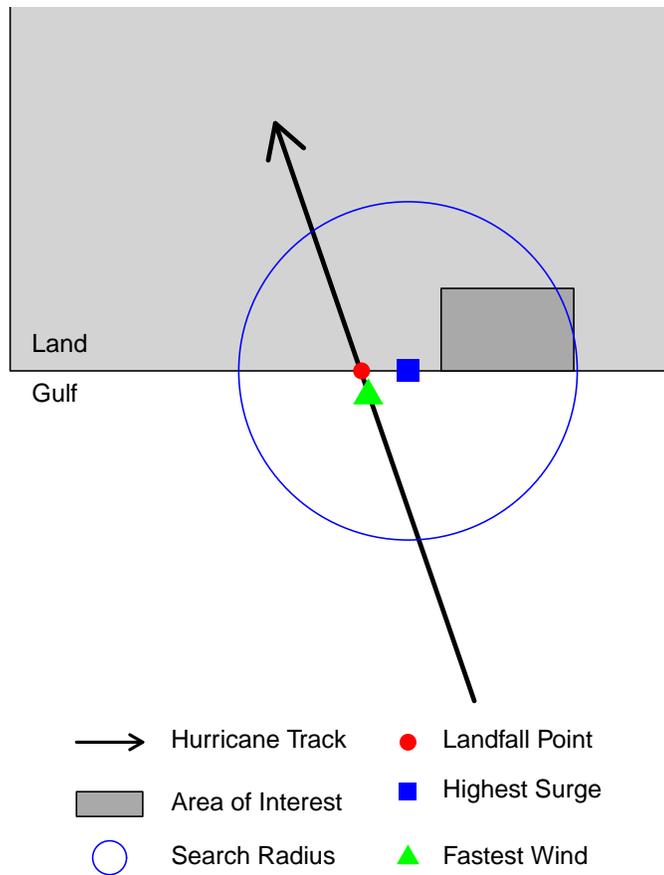


Figure 4.1: Conceptual schematic depicting data collation. Shows a hypothetical situation. Shapes of the variables are arbitrary. Wind and surge are point observations. The arrow represents a hypothetical hurricane track.



Figure 4.2: AFB study areas. For EAFB, the base is denoted with a red polygon. For MAFB and KAFB, the base is denoted with a red point. The surge events are shown with squares. The inset map includes a bounding box in the southeastern United States showing where the locations are globally.

ically, the **ismev** (Introduction for Statistical Modeling of Extreme Values) (Coles and Stephenson, 2010) and **copula** packages (Yan, 2007; Kojadinovic and Yan, 2010) were utilized.

The first portion of this analysis calculates the statistical risk of storm surge at the three AFBs using a POT extreme value model. This model is characterized by the GPD as it only considers surge events that pass a certain threshold  $u$ . The theorem for the GPD is described in detail in Chapter 1, and is shown in an application in Chapter 3. At each AFB, the surge data are collected and modeled using the GPD theorem. The threshold chosen for the model is 1.2 meters. The threshold of 1.2 was chosen because it is a value that can be used at all three AFBs that includes enough data to fit the model while excluding the more common lower surge height values.

Following Jagger and Elsner (2006), Malmstadt et al. (2010), and the GPD theorem, the exceedances  $S - u$  are modeled as samples from a family of GPDs, so that for an individual hurricane with maximum surge  $S$ ,

$$\begin{aligned} \Pr(S > h | S > u) &= \left(1 + \frac{\xi}{\sigma}[h - u]\right)^{-1/\xi} \\ &= \text{GPD}(h - u | \sigma, \xi) \end{aligned} \quad (4.1)$$

where  $\sigma > 0$  and  $\sigma + \xi(h - u) \geq 0$ . For negative values of the shape parameter ( $\xi$ ) the GPD family of distributions has an upper limit of  $S_{\max} = u + \frac{\sigma u}{|\xi|}$ .

The frequency of storms with surge heights of at least  $u$  follows a Poisson distribution with a rate,  $\lambda_u$ , the threshold crossing rate. Thus the number of hurricanes per year with surge levels exceeding  $h$  is a thinned Poisson process with mean  $\lambda_h = \lambda_u \Pr(S > h | S > u)$ . This is the POT method and the resulting model is completely characterized for a given threshold  $u$  by  $\sigma$ ,  $\xi$ , and  $\lambda_u$ ; the GPD parameters and the threshold crossing rate, respectively.

Since the number of storms exceeding any surge height  $h$  is a Poisson process, the return period for any  $h$  has an exponential distribution, with mean  $r(h) = 1/\lambda_h$ . By substituting for  $\lambda_h$  in terms of both  $\lambda_u$  and the GPD parameters then solving for  $h$  as a function of  $r$  we can find the corresponding return level for a given return period as

$$\text{rl}(r) = u + \frac{\sigma}{\xi} \left[ (r \cdot \lambda_u)^\xi - 1 \right]. \quad (4.2)$$

As a note, this is the same theorem as described for Florida winds in the previous chapter. Here it has been adjusted to deal with storm surge values.

Return levels for surge heights at EAFB, MAFB, and KAFB using the GPD theorems described above are shown in Figure 4.3. Surge height exceedance levels in meters are shown on the  $x$ -axis and the probabilities of annual exceedance are shown on the  $y$ -axis. The dots are the empirical points used in the model, and the red lines indicate the 95% CI around the model. The CI values are calculated using the delta method that employs the use of the quantile function for the GEV equation given in Equation 1.4 and the estimated covariance matrix of the GEV parameters ( $\mu$ ,  $\sigma$ , and

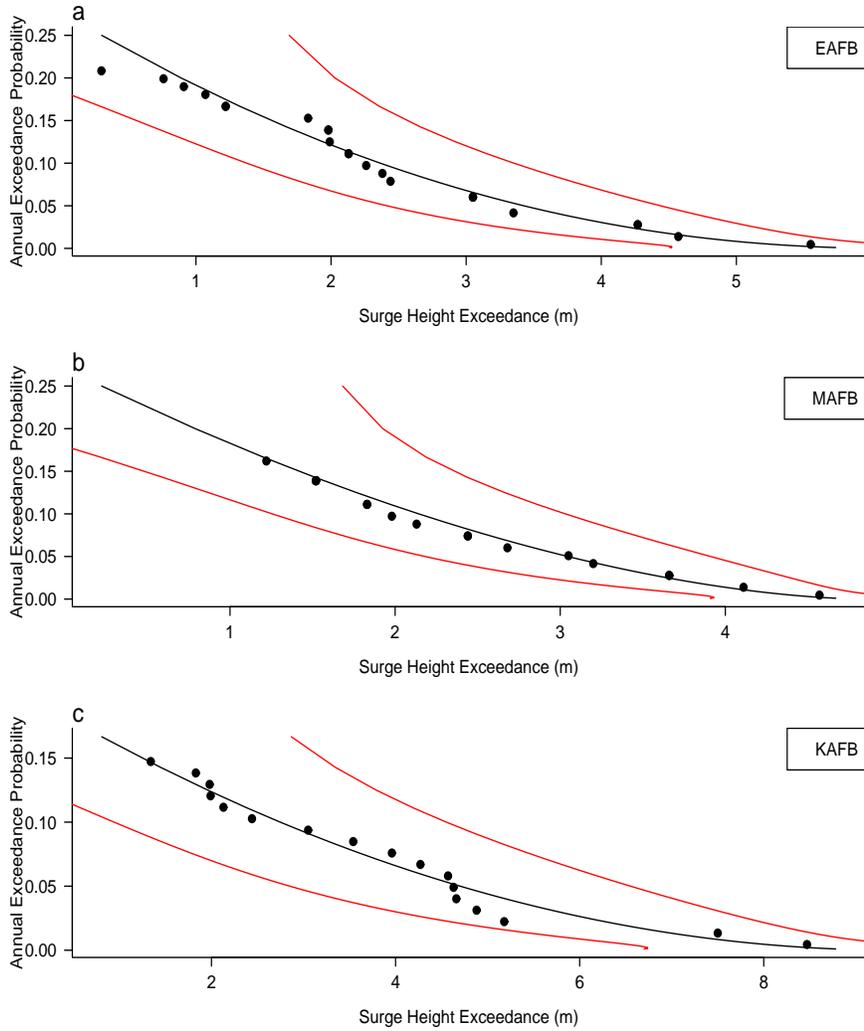


Figure 4.3: The annual exceedance probability for storm surge heights (m). Return levels are shown on the  $x$ -axis and the probabilities of annual exceedance are shown on the  $y$ -axis. The 95% CI is shown with red lines. (a) EAFB, (b) MAFB, and (c) KAFB.

$\xi$ ). The plot is interpreted as follows. EAFB (a) can expect surge height levels of 2.4 m (1.4–3.4) with an annual probability of 10%. MAFB (b) can expect surge height levels of 2.1 m (1.3–3.0) with an annual probability of 10%. KAFB (c) can expect surge height levels of 2.8 m (0.90–4.6) with an annual probability of 10%.

Table 4.2 shows the surge height exceedance estimates for 10, 20, 50, 100, and 500 years, including the 95% CI at EAFB. The 500 year return level for this area is just over 5.5 meters (4.5–6.6). Table 4.3 shows the surge height exceedance estimates for 10, 20, 50, 100, and 500 years, including the 95% CI at MAFB. A major difference

Table 4.2: Surge height return level estimates for specified return periods at EAFB. The CI listed is the 95%.

Year	Surge Height (m)	Confidence
10	2.36	(1.38–3.35)
20	3.43	(2.41–4.44)
50	4.39	(3.48–5.30)
100	4.89	(4.04–5.73)
500	5.57	(4.52–6.62)

between MAFB and EAFB is that the 500 year return level for MAFB is only 4.6 meters (3.9–5.2), almost one full meter below EAFB. Table 4.4 shows the surge height

Table 4.3: Surge height return level estimates for specified return periods at MAFB. The CI listed is the 95%.

Year	Surge Height (m)	Confidence
10	2.14	(1.26–3.03)
20	3.05	(2.18–3.91)
50	3.80	(3.10–4.49)
100	4.15	(3.58–4.72)
500	4.58	(3.93–5.23)

exceedance estimates for 10, 20, 50, 100, and 500 years, including the 95% CI at KAFB. This location has the highest return level for 500 years at 8.5 meters (6.7–10.3).

### 4.3 Relationship between Wind and Surge

The next step of analysis is to analyze the individual surge and wind variables, as well as assess their relationship with one another. Figure 4.4 shows the relationship between maximum sustained wind speed ( $\text{ms}^{-1}$ ) and surge height (m) at EAFB, MAFB, and KAFB. The regression model is shown with a red line. The significant (.001) relationship between these variables at EAFB is as follows: for every  $1 \text{ ms}^{-1}$  increase in wind speed, surge heights are expected to increase by 0.11 m. On the third and fourth axes, the marginal distributions of the respective variables are shown. At EAFB, there are no surge heights above 6 m (max 5.55), nor wind speeds above  $57 \text{ ms}^{-1}$  (max 56.88).

Table 4.4: Surge height return level estimates for specified return periods at KAFB. The CI listed is the 95%.

Year	Surge Height (m)	Confidence
10	2.75	(0.94–4.55)
20	4.70	(2.85–6.55)
50	6.44	(4.79–8.10)
100	7.32	(5.85–8.80)
500	8.50	(6.73–10.27)

The significant (.01) relationship between these variables at MAFB is as follows: for every  $1 \text{ ms}^{-1}$  increase in wind speed, surge heights are expected to increase by 0.06 m. At MAFB, there are no surge heights above 5 m (max 4.57), nor wind speeds above  $63 \text{ ms}^{-1}$  (max 62.73). The significant (.002) relationship between these variables at KAFB is as follows: for every  $1 \text{ ms}^{-1}$  increase in wind speed, surge heights are expected to increase by 0.11 m. At KAFB, there are no surge heights above 9 m (max 8.47), nor wind speeds above  $80 \text{ ms}^{-1}$  (max 79.96). This location experienced the highest surge height out of all three locations with Hurricane Katrina (2005), and the highest wind speed with Hurricane Camille (1969).

At EAFB, the relationship between wind and surge was tested including a third variable, position of the wind speed relative to the surge. Again, if the wind speed occurred to the left of the surge location, it was given an index of 1. If the wind speed occurred to the right of the surge location, it was given an index of 0. If a storm travels to the left of the surge event, the right front quadrant would be nearest the surge. It is tested to see if maximum sustained wind speed occurring to the left of the surge event produces a higher surge than if it occurs to the right. A significant value (0.10) of 0.78 leads us to reject the null hypothesis that there is no relationship and suggest that maximum wind speeds occurring to the left of a surge height location produce a higher surge. Neither distance variable as described above were significant at this location. The Akaike Information Criterion (AIC) is a measure of the goodness of fit of a statistical model. The lower the value, the better the fit. The smaller AIC value of 54.9 versus 56.2 suggests that the model using only wind and surge as variables better represents reality.

At MAFB, it is found that the model is improved by including D1. As a reminder, this represents the distance from the maximum wind location and the location of the surge height. It is found that both the distance and the maximum wind speed have a significant relationship with surge. For every  $1 \text{ ms}^{-1}$  increase in wind speed, surge heights are expected to increase by 0.07 m (significance of 0.001), and for every 1 km increase between the wind and surge height locations, surge is expected to increase by 0.02 m (significance of 0.004). This suggests that the greater the distance between wind and surge, the greater the surge will be. The relationship between wind,

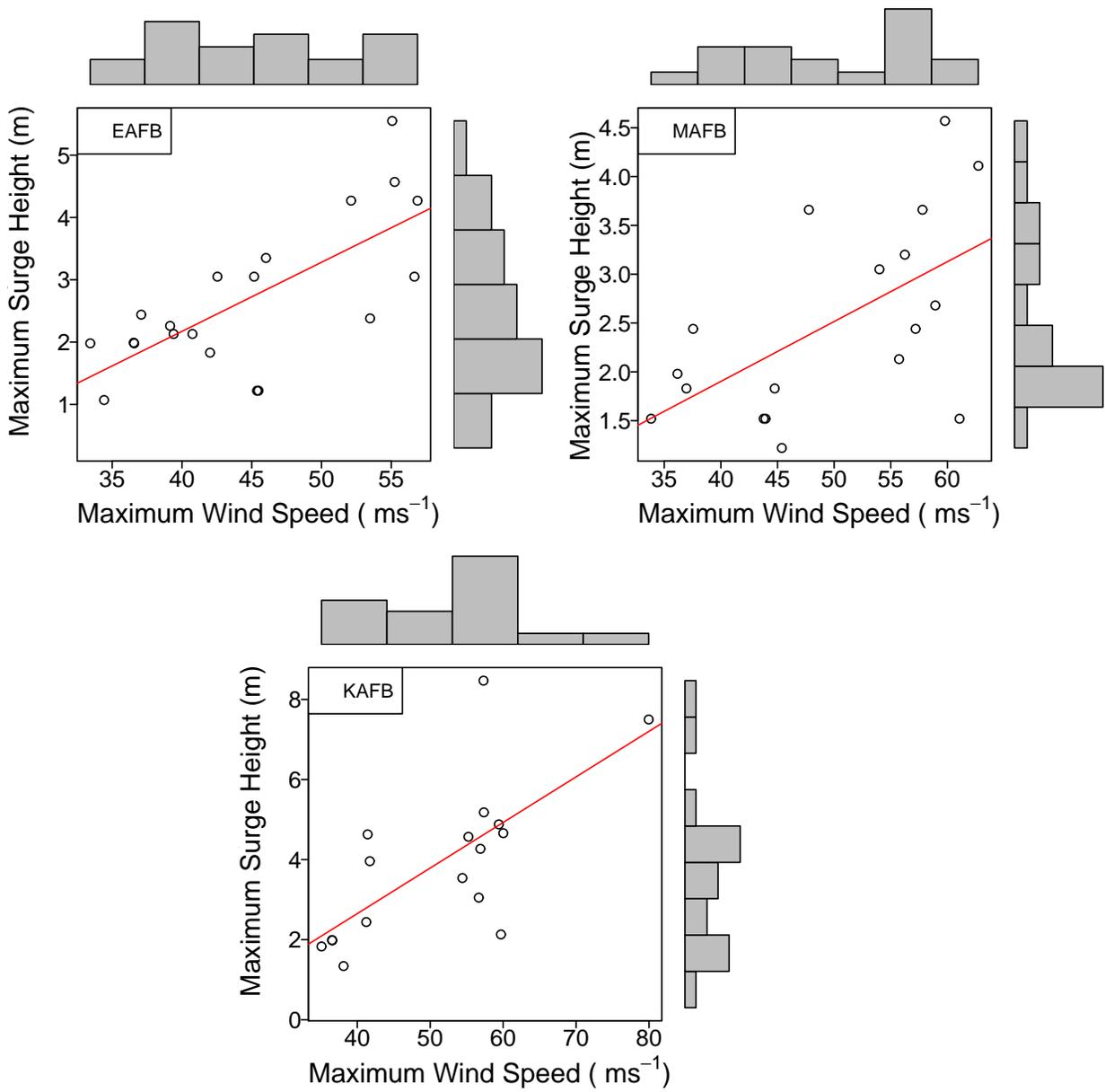


Figure 4.4: Maximum sustained wind speed ( $\text{ms}^{-1}$ ) regressed on storm surge height (m) at EAFB, MAFB, and KAFB. The red line shows the linear model. The adjacent axes show the marginal distributions of the respective variables.

surge, and position of the wind speed relative to surge was also tested at MAFB. A significance value (0.02) of  $-1.18$  leads us to reject the null hypothesis that there is no relationship and suggest that maximum wind speeds occurring to the left of a surge height location produce a lower surge. This result is opposite from EAFB. Surge is a very localized event and the orientation of the east/west versus north/south coastline at EAFB and MAFB might be the reason for this difference. The smaller AIC value of 40.5, compared to 44.5 and 48.7, suggests that the model including surge, wind, and distance between wind and surge is the most appropriate model at MAFB.

At KAFB, it is found that the model can, again, be improved by including D1. It is found that both the distance and the maximum wind speed have a significant relationship with surge. For every  $1 \text{ ms}^{-1}$  increase in wind speed, surge heights are expected to increase by 0.10 m (significance of 0.003). For every 1 km increase between the wind and surge height locations, surge is expected to decrease by 0.05 m (significance of 0.07). This suggests that the shorter the distance between wind and surge, the greater the surge will be. This is opposite then at MAFB, perhaps, again, due to the orientation of the coastlines. The smaller AIC value of 62.9, compared to 65.0, suggests that the model including the three variables better represents reality. The relationship between wind, surge, and position of the wind speed relative to surge was tested. The model did not produce significant results, so there is no need to pursue the inclusion of this variable any further.

### 4.3.1 Bivariate Dependence

Extreme wind speeds and deep storm surges from hurricanes work in conjunction to cause severe widespread damage when major hurricanes make landfall. The traditional way to estimate a combined likelihood of occurrence is to use copulas. Copulas describe the dependency structure between two random variables (Coles et al., 1999). Here the dependency arises from different physical processes (air and water) arising from the same meteorological conditions associated with hurricanes. Specifically, the objective is to develop a measure of dependence for wind and surge ( $W, S$ ). The model developed finds a return period (year),  $Z$ , for a given occurrence level of wind speed and storm surge. Tawn (1988) and Coles et al. (1999) describe the theory and dependence measures for extreme values analysis such as this. The copula function for a measure of extremal dependence is described as follows.

Based off the theorem discussed in Chapter 1, for any random vector,  $(W, S)$ , the distribution function  $F(w, s) = \Pr(W \leq w, S \leq s)$  provides a complete description of dependence between variables  $W$  and  $S$ . Marginal aspects can play an influence on the variables and can be removed by observing that, subject to continuity conditions, there is a unique function  $C(\cdot, \cdot)$  with domain  $\mathcal{A} = [0, 1] \times [0, 1]$  such that

$$F(w, s) = C\{F_W(w), F_S(s)\}, \quad (4.3)$$

where  $F_W$  and  $F_S$  are the marginal distribution functions given by

$$F_W(w) = F(w, \infty) \text{ and } F_S(s) = F(\infty, s). \quad (4.4)$$

The function  $C$  is the copula and it contains complete information about the joint distribution of  $W$  and  $S$  apart from the marginal distributions.  $C$  describes the association between the two variables after transformation to variables  $U$  and  $V$ , with Uniform  $[0, 1]$  margins, via  $(U, V) = \{F_W(W), F_S(S)\}$ . For a more complete summary of dependence measures, see Coles et al. (1999); Nelsen (1998); Joe (1997) and Yan (2007). The code used to create this copula for one location is available in the Appendix of this dissertation. The data necessary to run the code at all locations (as well as another copy of the code) is available at <http://jtrepanier.wordpress.com/data/>.

First the correlation between wind and surge is determined. Next, a two-dimensional normal copula is defined with the parameter equal to the sample correlation value. The dimension is the number of variables, here being wind and surge. An elliptical copula is the copula corresponding to an elliptical distribution by the Sklar's theorem (Fang et al., 1990). Let  $F$  be the multivariate cumulative density function (CDF) of an elliptical distribution,  $F_i$  is the CDF of the  $i^{\text{th}}$  margin, and  $F_i^{-1}$  is its inverse function,  $i = 1, \dots, p$ . The elliptical copula determined by  $F$  is

$$C(u_i, \dots, u_p) = F[F_1^{-1}(u_1), \dots, F_p^{-1}(u_p)] \quad (4.5)$$

Elliptical copulas are popular in finance and risk management because they are easy to implement (Yan, 2007). The model was also tested using a copula created with a  $t$ -distribution. Little difference was found in the results.

The wind and surge data are fitted to separate Weibull distributions and the parameter values are obtained for these marginals using a maximum-likelihood procedure (Venables and Ripley, 2002). The multivariate density is generated from the copula definition and the Weibull parameters (shape  $\xi$  and scale  $\sigma$ ). The copula-based multivariate density is fitted to the data using a maximum-likelihood procedure. The normal copula is redefined using the updated correlation parameter value. The multivariate density is generated from the new copula definition and the updated Weibull parameter values. This provides a fitted copula model using the observational data.

The density is plotted on a two-dimensional grid spanned by the range of wind and surge values. The joint probability of the wind and surge based on the copula density is determined using a function created in the R program (R Development Core Team, 2010). The per event probability,  $[\text{Pr}(W > w_{max}, S > s_{max})]$  times the hurricane frequency (the number of hurricanes divided by the total number of years), provides the annual event frequency. That is, the yearly frequency of events with  $W$  exceeding  $w_{max}$  and  $S$  exceeding  $s_{max}$ . The yearly probability of an event is then  $1 - \exp(-\text{frequency})$ , where  $\exp()$  is the exponent function, and the return period is 1 divided by  $1 - \exp(-\text{frequency})$ , or roughly 1 divided by the frequency plus 0.5 year. For small frequencies, the return period is roughly 1 divided by the frequency. This provides the probability of specific wind *and* surge events.

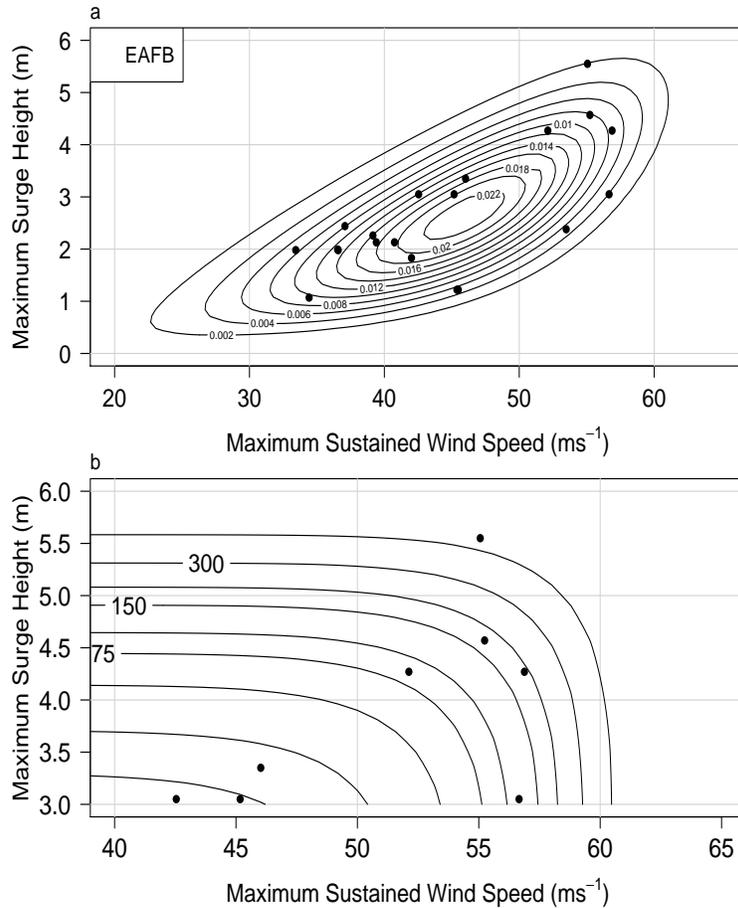


Figure 4.5: Density plot and wind speed/surge level return periods for EAFB. (a) Density plot for the EAFB copula. The points shown are the empirical data. Maximum sustained wind speeds ( $\text{ms}^{-1}$ ) are shown on the  $x$ -axis and maximum surge heights (m) are shown on the  $y$ -axis. (b) Return period information for wind and surge levels at EAFB. Each line represents a specific return period year. The years shown are 20, 30, 50, 75, 100, 150, 200, 300, and 500.

In Figure 4.5(a), the joint density plot for EAFB is shown with the observed data points. Contours, in units of events per bin size, are plotted at intervals of .001 starting at .001. The bin size is in units of meters times  $\text{ms}^{-1}$  and the number of bins is 50 times 50 over a range from 0 to 6 m and 20 to 65  $\text{ms}^{-1}$ . The model represents the data well. In Figure 4.5(b), the return period probability of specific wind and surge events are shown. Contours are given at 20, 30, 50, 75, 100, 150, 200, 300, and 500 years. A 500 year return period has an annual probability of 1/500 or 0.2%.

In Figure 4.6(a), the joint density plot for MAFB is shown with the observed data

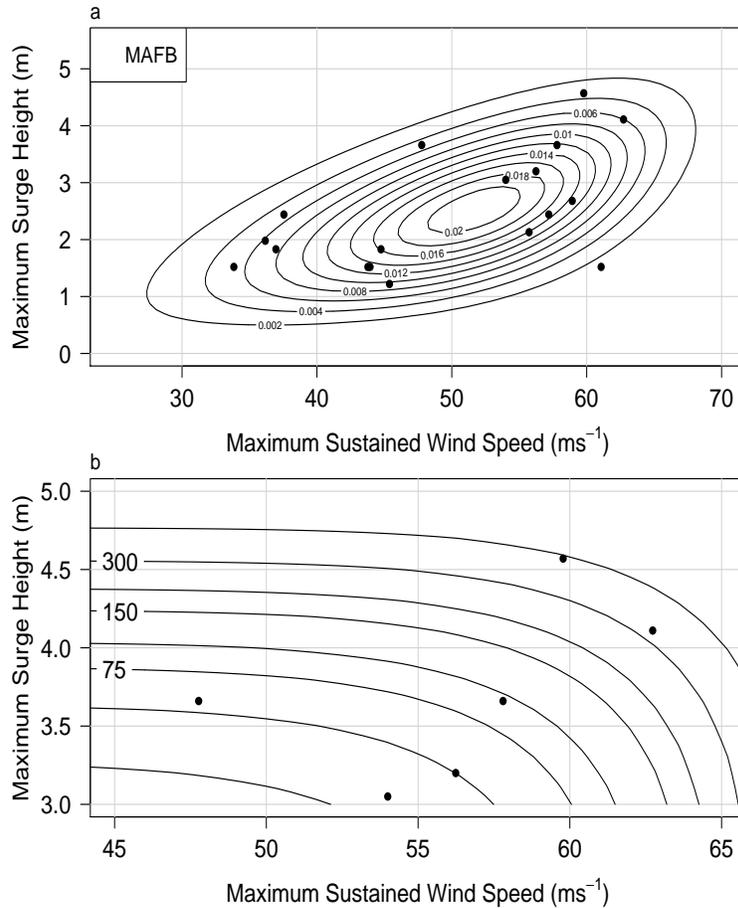


Figure 4.6: Density plot and wind speed/surge level return periods for MAFB. (a) Density plot for the MAFB copula. The points shown are the empirical data. Maximum sustained wind speeds ( $\text{ms}^{-1}$ ) are shown on the  $x$ -axis and maximum surge heights (m) are shown on the  $y$ -axis. (b) Return period information for wind and surge levels at MAFB. The years shown are 20, 30, 50, 75, 100, 150, 200, 300, and 500.

points. The bin size is in units of meters times  $\text{ms}^{-1}$  and the number of bins is 50 times 50 over a range from 0 to 5 m and 25 to 70  $\text{ms}^{-1}$ . The model, again, represents the density distribution of the observed data well. In Figure 4.6(b), the return period probability of specific wind and surge events are shown. The lines represent specific years. Contours are given at 20, 30, 50, 75, 100, 150, 200, 300, and 500 years.

In Figure 4.7(a), the joint density plot for KAFB is shown with the observed data points. The bin size is in units of meters times  $\text{ms}^{-1}$  and the number of bins is 50 times 50 over a range from 0 to 9 m and 15 to 80  $\text{ms}^{-1}$ . The match between the model and data appears to be good although the highest surge does not correspond

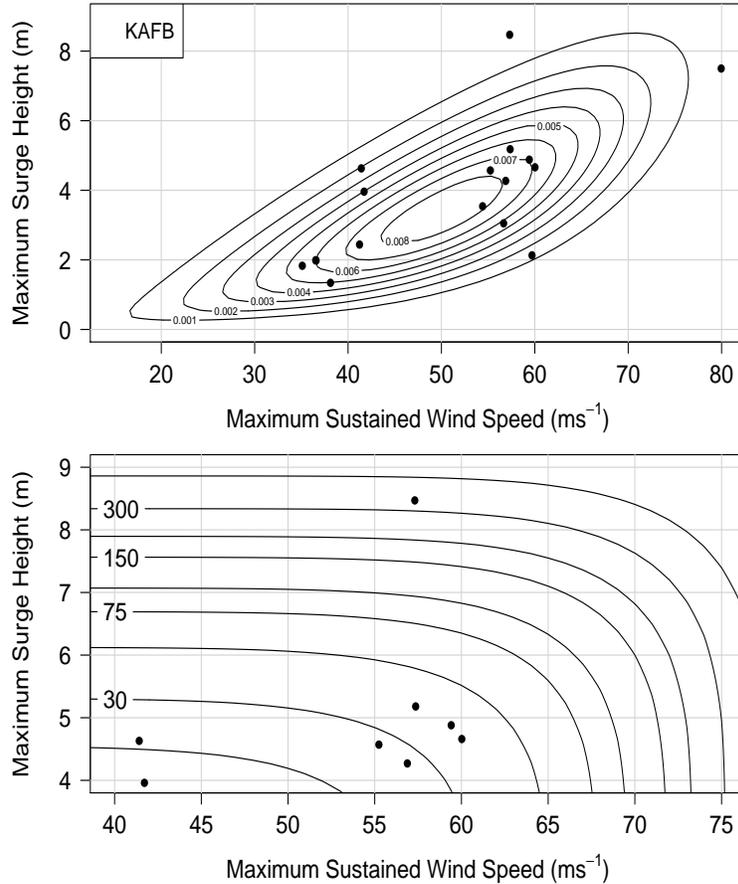


Figure 4.7: Density plot and wind speed/surge level return periods for KAFB. (a) Density plot for the KAFB copula. The points shown are the empirical data. Maximum sustained wind speeds ( $\text{ms}^{-1}$ ) are shown on the  $x$ -axis and maximum surge heights (m) are shown on the  $y$ -axis. (b) Return period information for wind and surge levels at KAFB. The years shown are 20, 30, 50, 75, 100, 150, 200, 300, and 500.

with the fastest wind nor does the fastest wind correspond with the highest surge. In Figure 4.7(b), the return period probability of specific wind and surge events are shown. Contours are given at 20, 30, 50, 75, 100, 150, 200, 300, and 500 years.

KAFB is the most vulnerable for high wind speeds and high storm surges on the shortest time spans when comparing to EAFB or MAFB. MAFB experiences slightly higher wind speeds than EAFB, but the surge values are lower. This could be due to coastline orientation. As it was discussed earlier, the index value denoting the side of landfall was a significant factor at EAFB and MAFB when describing the relationship between wind and surge. It could be considered an important factor with the return

period results. The results can be further analyzed in Tables 4.5, 4.6, and 4.7. Estimates of uncertainty around these point estimates, shown in parenthesis, are based on 1000 Monte Carlo simulations. The Monte Carlo simulations use random draws of the Weibull parameters for the marginals (two parameters for the wind marginal and two for the surge marginal) and a random draw from the copula dependency to generate return periods. The uncertainty is expressed as the interquartile range (25<sup>th</sup> to 75<sup>th</sup> percentiles) of the simulated return periods.

Table 4.5: Wind speed and surge level return periods for EAFB. Return period year information for Saffir-Simpson Category 1–4 wind speeds are shown along with 3, 4, and 5 meter surges. The quartile pointwise CI is shown in parentheses.

Wind Speed (ms <sup>-1</sup> )	Return Period (Year)		
	Surge Height (m)		
	3	4	5
33	15.6 (13.6-18.6)	41.9 (31.6-56.1)	174.1 (100.5-298.3)
43	17.2 (14.7-20.6)	42.9 (32.3-57.6)	174.7 (101.3-300.6)
50	28.3 (22.5-36.4)	54.7 (40.1-77.2)	189.3 (110.9-344.1)
59	244.0 (128.0-496.4)	293.0 (153.7-632.1)	546.0 (274.3-1309.2)

## 4.4 Combined Risk in Summary

Hurricanes are capable of generating significant economic and human loss. Losses result from high winds and overwhelming surge. The purpose of this study was to (1) describe the univariate risk of storm surge at three AFBs along the U. S. Gulf coast, and (2) describe the dependency structure between extreme wind speeds and storm surges at these locations. EVT and a POT model, using a GPD, are used to calculate the univariate risk. It is found that Eglin Air Force Base can expect a 2.3 meter surge, on average, once every 10 years, MacDill Air Force Base can expect a 2.1 meter surge, on average, once every 10 years, and Keesler Air Force Base can expect a 2.8 meter surge, on average, once every 10 years.

This chapter demonstrates a methodology for estimating the return period of the joint hazard based on a bivariate copula model. Wind and surge data from historical

Table 4.6: Wind speed and surge level return periods for MAFB. Return period year information for Saffir-Simpson Category 1–4 wind speeds are shown along with 3, 4, and 5 meter surges. The quartile pointwise CI is shown in parentheses.

Wind Speed ( $\text{ms}^{-1}$ )	Return Period (Year)		
	Surge Height (m)		
	3	4	5
33	21.6 (18.1-26.0)	93.2 (59.3-146.0)	920.4 (336.7-2683.3)
43	22.5 (19.0-27.4)	94.1 (60.5-147.9)	921.6 (339.7-2685.8)
50	26.8 (22.3-33.6)	100.8 (65.1-160.9)	937.9 (353.2-2750.4)
59	60.3 (43.8-86.8)	163.8 (100.9-287.9)	1166.3 (460.0-3701.7)

Table 4.7: Wind speed and surge level return periods for KAFB. Return period year information for Saffir-Simpson Category 1–4 wind speeds are shown along with 3, 4, and 5 meter surges. The quartile pointwise CI is shown in parentheses.

Wind Speed ( $\text{ms}^{-1}$ )	Return Period (Year)		
	Surge Height (m)		
	3	4	5
33	11.0 (10.2-12.5)	15.5 (13.8-18.5)	24.7 (20.5-31.4)
43	12.1 (11.2-14.1)	16.3 (14.5-20.2)	25.3 (21.2-32.7)
50	14.8 (13.3-17.9)	18.6 (16.5-23.5)	27.3 (22.9-36.2)
59	26.3 (21.9-34.0)	29.7 (24.9-40.1)	38.3 (31.0-54.9)

hurricanes are collated and an elliptical copula with Weibull margins is specified. Density plots and probability plots are shown to summarize the results. Eglin Air Force Base can expect a hurricane with at least  $50 \text{ ms}^{-1}$  winds and surge heights of at least 3 m, on average, once every 28 years (23–36). MacDill Air Force Base can expect a hurricane with at least  $50 \text{ ms}^{-1}$  winds and surge heights of at least 3 m, on average, once every 27 years (22–34). Keesler Air Force Base can expect a hurricane with at least  $50 \text{ ms}^{-1}$  winds and surge heights of at least 3 m, on average, once every 15 years (13–18). From these results, it is suggested that Keesler Air Force Base has the highest probability of experiencing extreme wind speeds and high storm surges at the shortest time periods.

The broad impact of this individual study is the hope that locations around the world that experience these natural weather phenomena will adopt this methodology to better understand their statistical risk of hurricanes. Many locations rely simply on the expectation of certain wind levels, when there is another deadly factor they could be considering. This methodology provides a way to assess the frequency of hurricane wind speeds along with hurricane storm surges at any location where surge and wind speed data are available.

The final research chapter of this dissertation extends the univariate hurricane wind risk approach by mapping the GPD parameters as a function of location across the North Atlantic and Gulf of Mexico. Spatial variations in the model parameters may provide new insight into hurricane behavior that is lost in traditional climatological and meteorological approaches.

## CHAPTER 5

# SPATIAL RISK OF HURRICANE WINDS OVER THE GULF OF MEXICO AND NORTH ATLANTIC

The likelihood of a hurricane occurrence, no matter how it is calculated, is a very localized attribute. As previously seen in Chapters 3 and 4, a univariate statistical risk represents the probability of occurrence for a specific variable at a given location. In this chapter, the probability of hurricane wind occurrence is estimated at various locations across the Gulf of Mexico (GoM) and the North Atlantic (NAtl). This statistical risk is calculated using the same univariate approach applied in the previous chapters. The innovative method provided by this study is the way in which the parameters and return level estimates of wind risk are calculated and visualized; using hexagonal tessellations.

Space is used to help understand the particular characteristics that make hurricanes dangerous. In past literature, hurricane occurrence estimates are provided across space using different approaches. One of the first attempts at calculating return periods was a study conducted by Simpson and Lawrence (1971) where the authors analyzed the number of strikes at multiple locations over a particular time period. A more recent example of a similar method is in Keim et al. (2007) where the authors analyze 105 years of hurricane storm occurrence at 45 locations along the U. S. coast. Other approaches are also prevalent in the literature that involve including other variables besides frequency (e.g., intensity, forward speed, size) and more advanced statistical methods to analyze them. For example, Jagger and Elsner (2006) consider how climate variables affect hurricanes, including the El Niño Southern Oscillation, the North Atlantic Multidecadal Oscillation, the North Atlantic Oscillation, and the global temperature.

Geographic research focuses on how phenomena change through space and time. A hurricane is a phenomenon that has the attributes of moving from one location to another over a period of time. Not only does it move, but each individual hurricane has a different shape dimension that causes the space affected to be different for each event. It is the goal of this research to bridge the gap between the mete-

orological/climatological approaches used most frequently in past literature and the geographical research of analyzing space.

As mentioned in Chapter 1, the only geometric shapes that completely cover the Atlantic basin without any overlap are squares, triangles, and hexagons (Brettschneider, 2008). A hexagon represents the best compromise between overlap and uniformity and can be tessellated where no area is left uncovered. All of the interior angles of a hexagon are equal, and the sides are of equal length, so the hexagon best approximates the idealized shape of a circle. This allows for a tessellation with a shape most similar to the object of interest.

In this chapter, the GoM and NAtl are tessellated with hexagonal shapes to see if there is a relationship between location and the various parameters of the univariate risk model. It is assumed that there is a latent spatial process characterized by geographic information that drives hurricane characteristics, much like the study in Cooley et al. (2007). The parameters, or characteristics, of interest include the  $\lambda$ , or rate of occurrence, the  $\sigma$ , or scale of wind speeds, and the  $\xi$ , or shape that dictates the tail ends of the distribution (the extremes). It is very important to note that the scale that  $\sigma$  represents is not the same as the way geographers typically use scale. It does not refer to spatial dimension, but rather to the dispersion of wind values at the particular hexagon (a statistical scale). The thirty year return level is also mapped over space to test the relationship between return period and location. Thirty years is chosen to represent the typical homeowner's mortgage.

Tobler (1970) states that all things are related, but nearby things are more related than distant things. Visualizing the latent model parameters using hexagons allows for insight into whether this is the case over the GoM and NAtl. Latent variables are those that cannot be observed but are instead output from a model. Using the Moran's  $I$  test, the effect that spatial autocorrelation plays in the parameters, if any, is estimated. The final portion of analysis introduces a covariate; SST. The relationship of local SST on the parameters is estimated using a GWR model. This covariate is included to test the relationship of warming SSTs on hurricane frequency, intensity, and return levels. In a warming climate, this insight could be beneficial to policy makers as they prepare for the potential of more extreme weather patterns.

In this chapter, Section 5.1 discusses the details about data collection and describes how the basin is tessellated. Section 5.2 discusses the calculation of the GPD parameters and visualizes the results. In this section, the effect of the modifiable areal unit problem (MAUP) is also visualized by changing the sizes of the hexagons. Section 5.3 shows the effect of spatial correlation on the parameters. Section 5.4 includes the GWR models for each parameter and the thirty year return level using SST as the covariate. Significance is included using the  $t$ -statistic. Section 5.5 summarizes the main points of this chapter.

## 5.1 Data Collation

The 1-hourly hurricane data utilized in this study are the same as presented in Chapter 2. Specifically, all hurricane events (with winds reaching at least  $33 \text{ ms}^{-1}$ ) that occurred from 1854 to 2009 are included. The range of years used corresponds with the available SST data currently on record. The study area domain is dictated by the occurrence of hurricane winds across the NAtl and GoM. Before the data are subset any further, the hurricane tracks are first geographically projected to the World Geodetic System 1984 coordinate frame. This is done using the R Program for Statistical Computing (R Development Core Team, 2010) and the packages **rgdal** (Keitt et al., 2012) and **sp** (Pebesma et al., 2011). They are then reprojected to a Lambert conformal conic (LCC) planar projection using the parallels  $23^\circ$  and  $38^\circ$  N and a center longitude of  $60^\circ$  W. In order to subset the hurricane data to particular locations, the hexagonal tessellations must be created. As a note, the R syntax to create the hexagons is available in the Appendix of this dissertation. The data necessary can be found at [jtrepanier.wordpress.com/data](http://jtrepanier.wordpress.com/data).

The question of size becomes important when creating the hexagonal tessellation. Keim et al. (2007) provides a summary of the size of typical hurricanes, while noting that there is considerable variability in the size of hurricanes and that spatial dimensions are not well correlated with storm intensity. For major hurricanes the hurricane intensity swath is 240 km wide, with tropical storm conditions stretching an additional 80 km to the right and 40 km to the left. This suggests that the area of a storm of this magnitude is over 100,000 square kilometers. 125 hexagons are fit to the area dictated by the wind speed occurrences. The area is covered by 110 equal area hexagons, so the additional are removed. Each has an area of 761,821 square kilometers. This is an area slightly larger than Texas. This was chosen as a compromise between the scale (size) of a hurricane and the representation of the data. Stated another way, in order to have enough data at each hexagon, the size cannot be too small, and, in order to have estimates that are considered regional, the size cannot be too large. Additional estimates are attempted using different sizes in the Subsection 5.2.2 as an attempt to understand the effect of the MAUP on this approach.

The wind speeds are subset so that each hexagon has an individual set of storms. At this point, each projected hexagon has a subset of projected hurricane wind speeds that occurred within that hexagon. Hexagons from the original 110 that do not have wind speed values are removed. 54 hexagons remain. Some hurricanes have more than one value per hexagon, while others have only one. The number depends on how long the hurricane took to travel through the hexagon. For the extreme value model, only one maximum wind value per hurricane per hexagon is needed. For example, if the first hexagon has five hurricane events with a series of values per event, only the maximum per hurricane wind speed is needed, providing five maximum values. Each hexagon's wind speed values are subset to include only the maximum per hurricane value within each hexagon. Now each hexagon has a subset of maximum values for

all hurricanes that passed through the hexagon. As a note, in order to fit the GPD model to the data, there needs to be a sufficient amount. Any hexagons with less than 12 hurricane events are not included. These occur only at the edges of the domain where hurricane occurrences are less frequent in the record. A final 38 hexagons are left for analysis over the GoM and NATl.

The NOAA’s extended reconstructed SST data set for the NATl is described in Chapter 2. The specific SST data used in this chapter are for the domain bounded by 0° and 70° N latitude and 100° W and 10° E longitude for the set of months starting with January 1854 through November 2009. The data are further subset to include only the average combined SST per hexagon from August, September, and October (ASO), as these heat values are during the peak of the hurricane season and are indicative of the amount of available energy for hurricanes to form. See the Appendix for how this is done. The `ncdf` package provides functions for working with netCDF files in R (Pierce, 2011). Each hexagon has the maximum per hurricane wind speed and the average ASO SST.

## 5.2 Estimation of Extreme Value Model

Modeling the return period estimates for extreme hurricane winds at each hexagon is done in the same way that wind risk in Florida was calculated in Chapter 3 and the same way that surge risk was calculated at the three AFBs in Chapter 4. As a reminder, the estimation is explained again.

The statistical risk of extreme winds at all the hexagons is calculated using a POT extreme value model. This model is characterized by the GPD as it only considers wind events that pass a certain threshold  $u$ . The threshold chosen for the model is  $33 \text{ ms}^{-1}$ , or a Category 1 hurricane on the Saffir-Simpson Scale. This particular threshold value was chosen because it excludes any wind values below hurricanes strength, but does not place the threshold so high that there would be too few data points for analysis at each hexagon.

The exceedances  $W - u$  are modeled as samples from a family of GPDs, so that for an individual hurricane with maximum wind  $W$ ,

$$\begin{aligned} \Pr(W > v | W > u) &= \left(1 + \frac{\xi}{\sigma}[v - u]\right)^{-1/\xi} \\ &= \text{GPD}(v - u | \sigma, \xi) \end{aligned} \tag{5.1}$$

where  $\sigma > 0$  and  $\sigma + \xi(v - u) \geq 0$ . For negative values of the shape parameter ( $\xi$ ) the GPD family of distributions has an upper limit of  $W_{\max} = u + \frac{\sigma u}{|\xi|}$ .

The frequency of hurricanes with intensity of at least  $u$  follows a Poisson distribution with a rate,  $\lambda_u$ , the threshold crossing rate. Thus the number of hurricanes per year with wind speed levels exceeding  $v$  is a thinned Poisson process with mean  $\lambda_v = \lambda_u \Pr(W > v | W > u)$ . This is the POT method and the resulting model is completely characterized for a given threshold  $u$  by  $\sigma$ ,  $\xi$ , and  $\lambda_u$ . Since the number

of hurricanes exceeding any wind speed  $v$  is a Poisson process, the return period for any  $v$  has an exponential distribution, with mean  $r(v) = 1/\lambda_v$ . By substituting for  $\lambda_v$  in terms of both  $\lambda_u$  and the GPD parameters then solving for  $v$  as a function of  $r$ , the corresponding return level for a given return period is represented as

$$\text{rl}(r) = u + \frac{\sigma}{\xi} [(r \cdot \lambda_u)^\xi - 1]. \quad (5.2)$$

The GPD model is run at each hexagon on the maximum per hurricane wind speeds above a threshold of  $33 \text{ ms}^{-1}$ . Not only is insight gained about return periods of particular levels of hurricane wind speeds, but there is also access to the parameters estimated from the model. This includes the rate,  $\lambda$ , the scale,  $\sigma$ , and the shape,  $\xi$ , parameters. Each of these parameters, as well as the thirty year return level, are displayed at each hexagon in the next subsection. Before examining the parameters, however, it is important to understand what might be gained from this type of visualization.

By viewing the parameters in this way, a form of latent variable modeling is being done. In statistics, latent variables are those that are not observed but are rather inferred through a mathematical model. In this chapter, the GPD model is providing the latent variables that represent a threshold crossing rate, the overall dispersion of the wind speeds, and the shape of the tail ends of the distribution at each hexagon. These are visualized, as well as modeled further in a GWR model with SST values to provide insight into the way that hurricane characteristics behave over space. In particular, the threshold crossing rate, or  $\lambda$ , will show the area (defined by hexagon) that is most susceptible to the number of maximum hurricane occurrences above the Category 1 level. The scale, or  $\sigma$  parameter describes the stretching or shrinking of the distribution of the wind speed, so it provides insight into the dispersion of the winds above threshold  $u$  at the individual hexagons. The shape, or  $\xi$  parameter dictates the tail end of the distribution, so it is this parameter that represents the most extreme levels within a particular hexagon. The values are  $> -1$  because if  $\xi < -1$ , the likelihood is unbounded and maximum likelihood fails within the model. Finally, the thirty year return level will provide insight into the location that can expect the highest wind speeds on a fixed time period.

### 5.2.1 Visualization of Latent Variables

Figure 5.1 shows each variables plotted over space. Figure 5.1(a) shows the rate of hurricane occurrence in units of hurricanes per year. The hexagons over Florida, Bermuda and the far western Atlantic suggest that these locations experience at least one hurricane with winds reaching  $33 \text{ ms}^{-1}$  annually. Considering that hurricanes drop in intensity as they approach and cross land, these estimates are most accurate out in the ocean. For example, the eastern coast of the U. S. might not receive a landfalling hurricane each year, but a  $33 \text{ ms}^{-1}$  hurricane wind speed will pass through the hexagon somewhere. Near the edges of the map nearest Africa, the annual probability becomes

very low. This makes physical sense since many NATl tropical cyclones originate off the western coast of Africa and are not yet hurricane strength.

Figure 5.1(b) shows the scale parameter in  $\text{ms}^{-1}$  at each hexagon. The brighter values of yellow indicate locations where the scale of the distribution is higher. The map suggests that the peninsula of Florida, the Western Antilles Islands, and the southwestern Caribbean Sea are the most susceptible to wind speeds above the threshold crossing rate having a greater dispersion. Stated another way, these locations experience the greatest range in hurricane wind speeds (i.e., they receive Category 1, 2, 3, 4, and 5 wind speeds). Again, towards the eastern most edge of the map, the values are low as the wind speeds do not vary much in these hexagons. This is also a product of the origination of NATl tropical cyclones. Any hurricane force winds occurring are usually low.

Figure 5.1(c) shows the shape parameter, which is a dimensionless quantity, at each hexagon. This represents the tail ends of the distribution of the wind speeds. The two hexagons in the southeastern most corner of the map have values that are  $< -1$ , which means the model was unable to fit the  $\xi$  parameter to the subset of winds at these hexagons. This is due to the low scale parameter as witnessed in Figure 5.1(b). The overall dispersion of the winds in these hexagons is too low to fit a maximum likelihood estimate. Therefore, these two hexagons have been removed from the map. The darker hexagons over the center of the map, the peninsula of Florida, and the Western Antilles Islands suggest that these locations experience the most extreme wind events. The blue hexagon in the North, and the two blue hexagons near the east central are likely low due to the dispersion of the data and the low rate values. That is, the lack of extreme, or maximum, data in these hexagons causes the model estimates to be slightly skewed at these locations.

Figure 5.1(d) shows the thirty year return level estimate in  $\text{ms}^{-1}$  at each hexagon. This map is different from the previous three because it is a final estimate calculated after considering the three previous parameters. The results are to be expected. The eastern and northern most edges of the map have very low return levels in each hexagon due to the minimal historic occurrences of hurricanes in these locations. The areas over the western Caribbean and the center of the GoM have the highest return level of all locations.

Based on this GPD model, every hexagon has a theoretical maximum wind speed that can be calculated as

$$u + \left( \frac{\sigma}{|\xi|} \right). \quad (5.3)$$

Figure 5.2 shows the empirical maximum wind speed compared to the theoretical maximum wind speed. The theoretical wind speed can be understood as the highest possible wind speed that can be experienced within the hexagon. Figure 5.2(b) suggests that the maximum possible wind speeds can occur over the southeastern U. S. , the GoM, portions of northern Central America, and over the northern tip of South America. Any location with green, orange, or red hexagons can theoretically expect

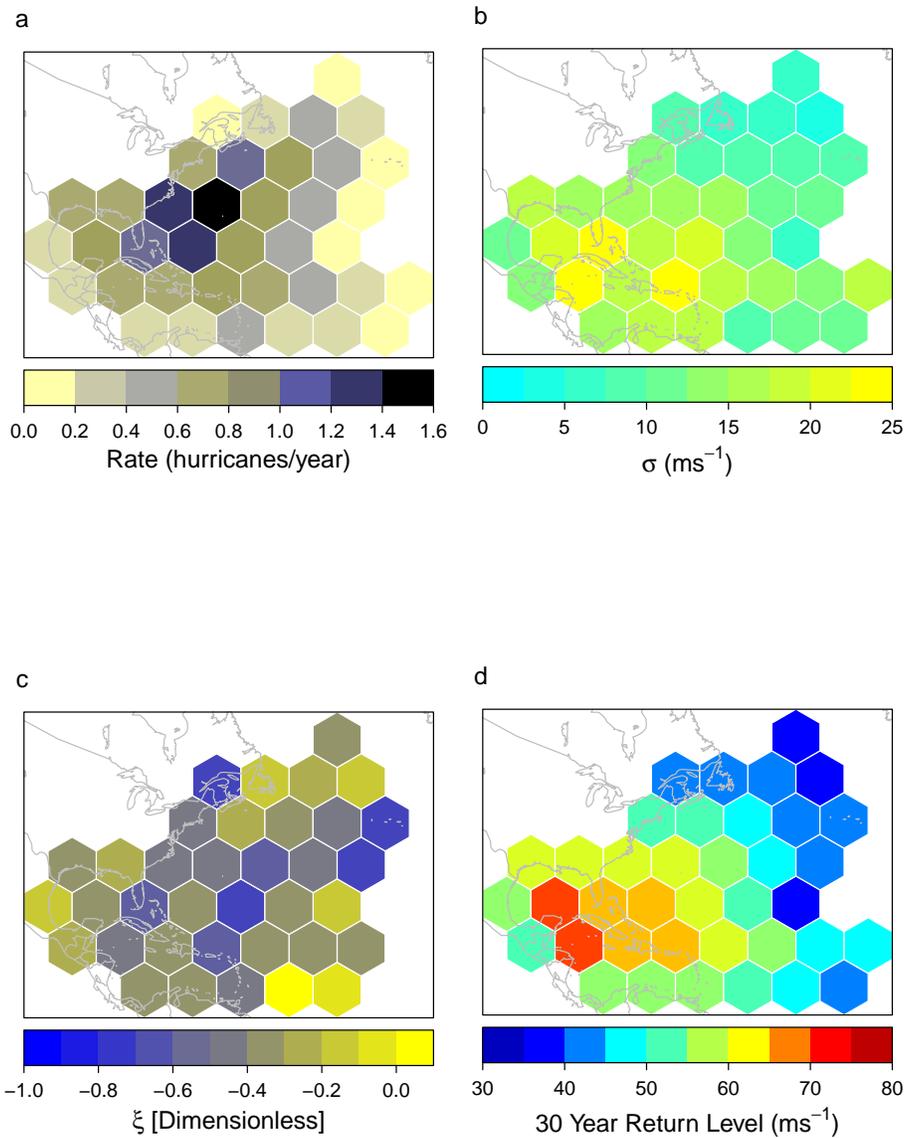


Figure 5.1: Each variable plotted per hexagon. (a)  $\lambda$  parameter, or rate of hurricanes exceeding  $33 \text{ ms}^{-1}$  per year. (b)  $\sigma$  parameter, or scale (dispersion) of wind speeds in  $\text{ms}^{-1}$ . (c)  $\xi$  parameter, or shape of the distribution tails. This is a dimensionless quantity. (d) The thirty year return level in  $\text{ms}^{-1}$ . The maps are projected in a Lambert conformal conic planar projection with parallels at  $23^\circ$  and  $38^\circ$  N and a center longitude of  $60^\circ$  W.

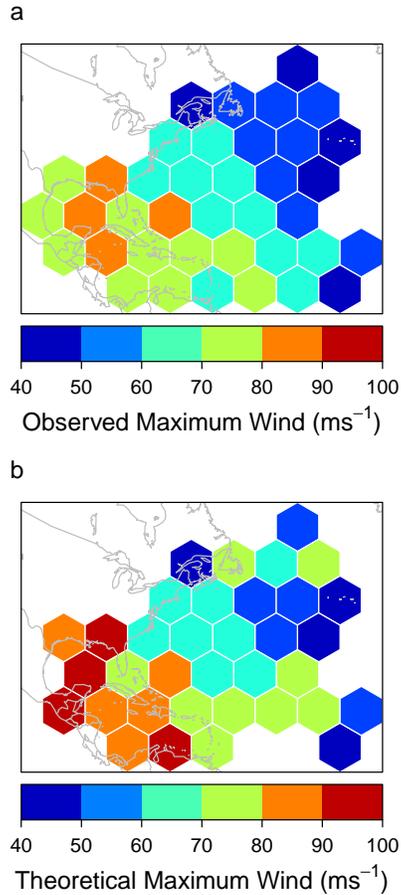


Figure 5.2: Comparison of empirical maximum wind and theoretical maximum wind. (a) Empirical wind speed ( $\text{ms}^{-1}$ ). (b) Theoretical wind speed ( $\text{ms}^{-1}$ ).

Category 5 wind speed levels ( $\geq 69 \text{ ms}^{-1}$ ). Any hexagons that have been removed from the map in Figure 5.2(b) were done so due to model fit problems.

The parameters above are dependent to some level on the size of the hexagon chosen. Next, different sizes are compared to visualize the effect of the MAUP.

## 5.2.2 Hexagon Size Comparison

A good strategy to test the effect of the modifiable areal unit problem is to see how much the parameter estimates change when using different hexagon sizes (see Elsner et al. (2011)) and see if the change influences the conclusions. Figures 5.3, 5.4, 5.5, and 5.6 all show the parameter and model estimates using different hexagon sizes. When  $n = 75$ , the area of each hexagon is 1.27 million  $\text{km}^2$ . When  $n=100$ ,

the area of each hexagon is 952,276 km<sup>2</sup>. When n=125 (as it does in the chapter), the area is equal to 761,821 km<sup>2</sup>. As it is seen in Figures 5.3, 5.4, 5.5, and 5.6, the results do not vary greatly with different hexagon size. The size chosen for the results is done so because it allows for enough values per hexagon for the model to run successfully, while keeping the estimates at a regional size for application purposes. For comparison, all color intervals for the maps are the same per parameter.

Now that the parameters and the thirty year return level estimate have been mapped and discussed, it is interesting to test if the variables are spatially correlated. Hurricanes pass through space over similar environmental conditions so it is possible that the parameters from one location to the next are spatially dependent.

### 5.3 Spatial Autocorrelation

Metrics of spatial autocorrelation can reveal if there is a local dependency between observations that is not accounted for by the independent variables. Stated another way, it tests to see if there is a dependency between observations that occur near one another in space and time. Since the model is used on the maximum per hurricane wind speeds in each hexagon, locations may influence one another. A maximum wind speed in one hexagon will not be that dissimilar from the maximum wind speed from the same hurricane in a nearby hexagon. A common metric for spatial autocorrelation can help provide insight into whether a given parameter is dependent upon nearby values of the same parameter.

This common metric is Moran's  $I$  (Moran, 1950) defined as

$$I = \frac{n y^T W y}{s y^T y} \quad (5.4)$$

where  $n$  is the number of hexagons,  $y$  is the vector of values within each hexagon (e.g., rate parameter) where the values are deviations from the overall mean,  $W$  is a weights matrix,  $s$  is the sum over all the weights, and the subscript  $T$  indicates the transpose operator.

Values of Moran's  $I$  range from  $-1$  to  $+1$  with a value of zero indicating a pattern with no spatial autocorrelation. Although not widely used in climate studies, de Beurs and Henebry (2008) use it to identify spatially coherent eco-regions and biomes related to the North Atlantic oscillation. To compute Moran's  $I$  a weights matrix is needed. The weights matrix is square with the number of rows and columns equal to the number of hexagons. Here the weight in row  $i$ , column  $j$  of the matrix is assigned a non-zero value if hexagon  $i$  is contiguous with hexagon  $j$ . Otherwise it is assigned a zero. The **spdep** package (Bivand, 2011) in R (R Development Core Team, 2010) has functions for creating weights based on contiguity (and distance) neighbors.

A contiguity-based neighborhood list is created from the spatial hexagons. The list is ordered by hexagon number starting with the southwestern most hexagon. This hexagon has three neighbors; hexagon numbers 2, 8, and 9. Hexagon numbers

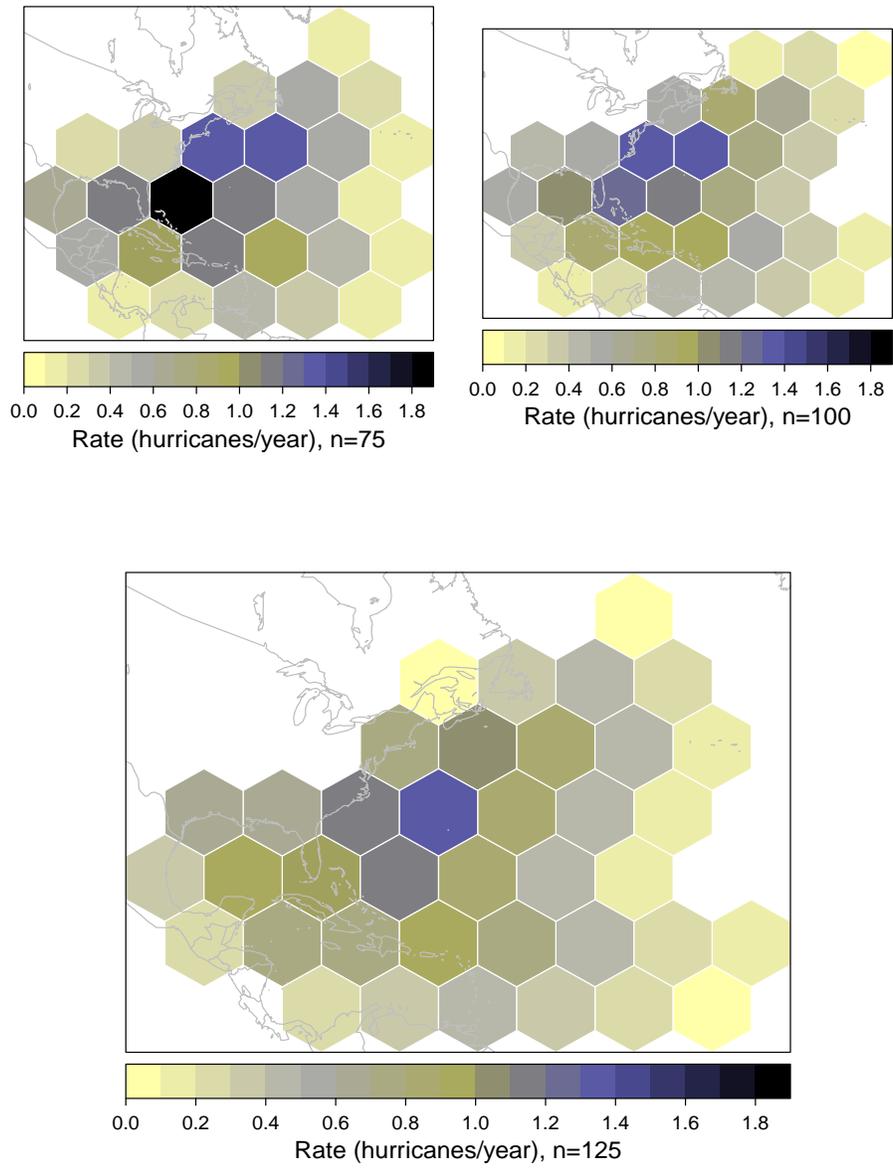


Figure 5.3: Comparison of rate parameter at three different hexagon sizes.  $n = 125$  is the size used in the chapter analysis.

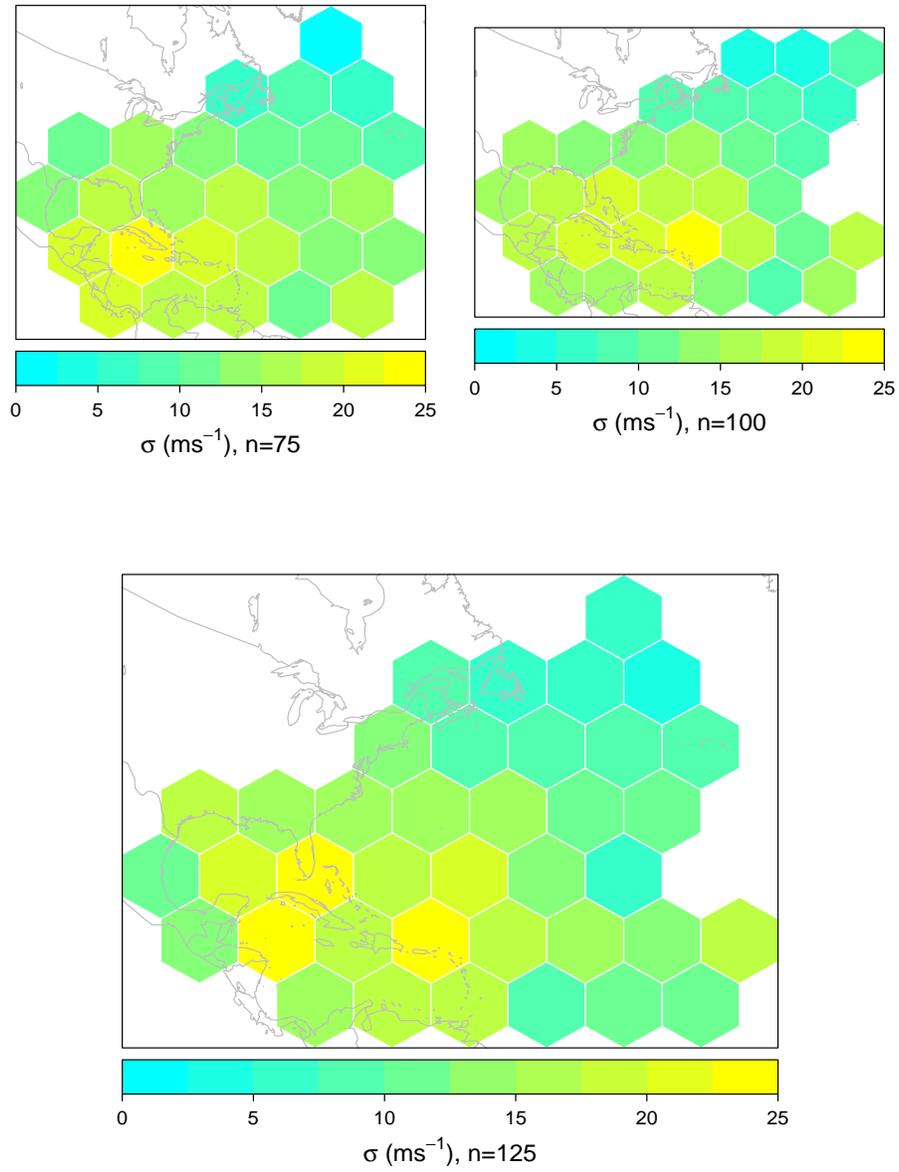


Figure 5.4: Comparison of  $\sigma$  parameter at three different hexagon sizes.  $n = 125$  is the size used in the chapter analysis.

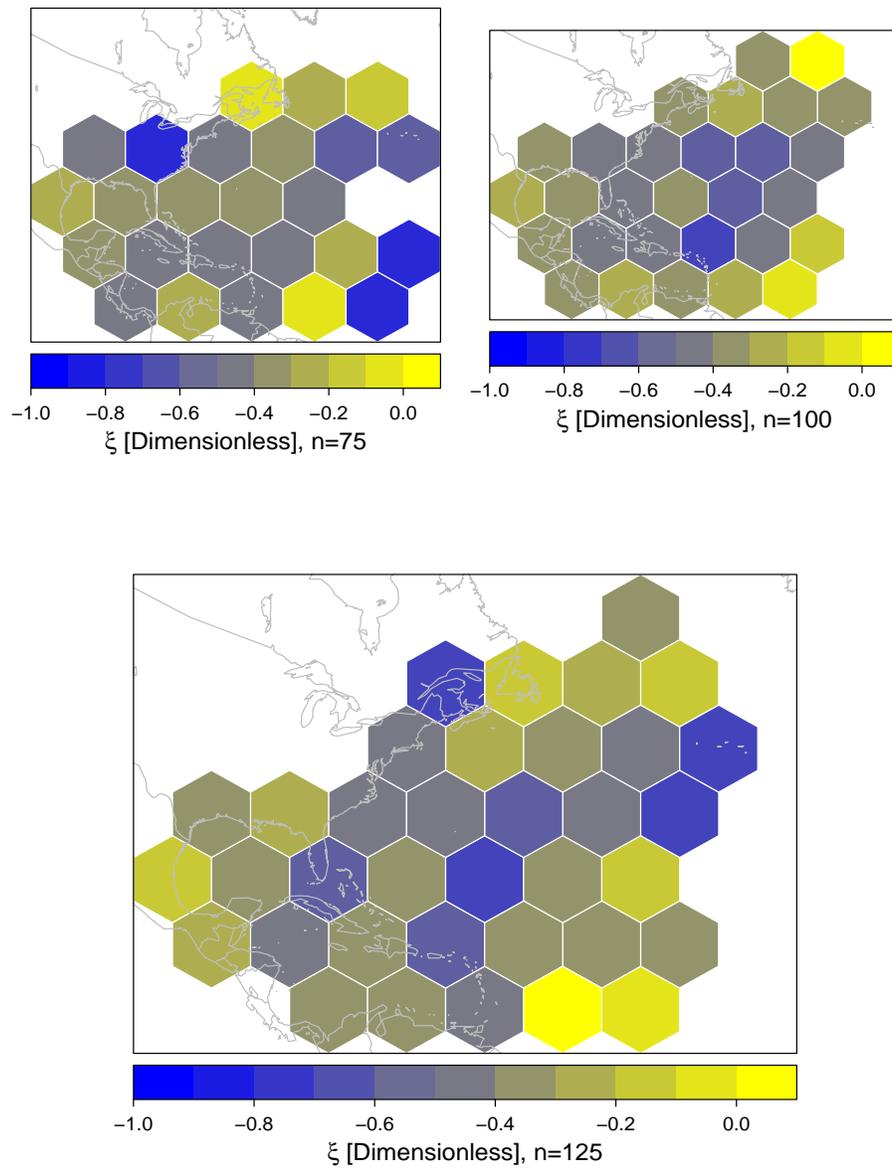


Figure 5.5: Comparison of  $\xi$  parameter at three different hexagon sizes.  $n = 125$  is the size used in the chapter analysis.

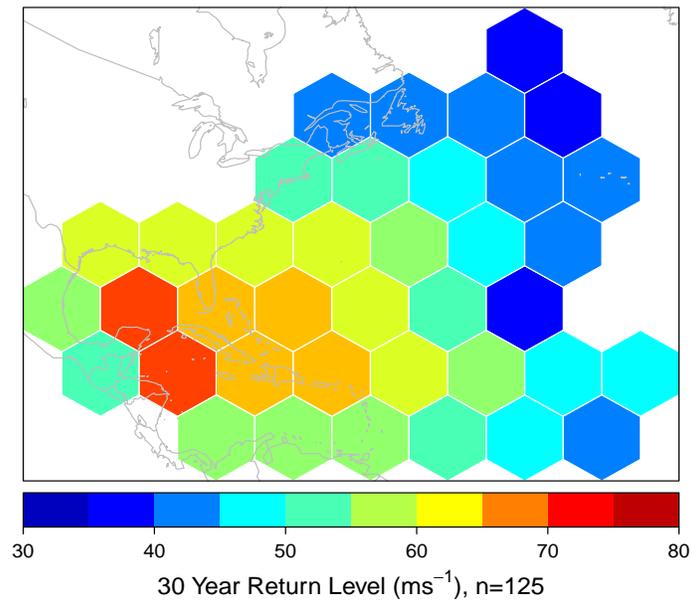
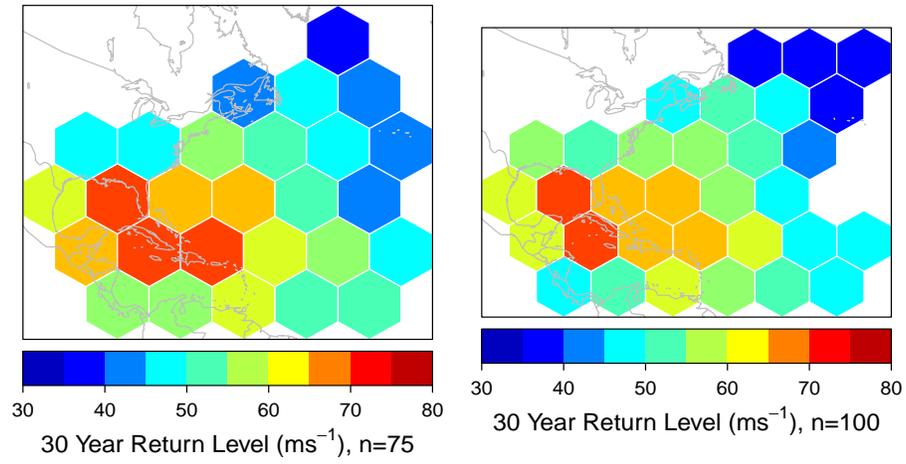


Figure 5.6: Comparison of the thirty year return level at three different hexagon sizes.  $n = 125$  is the size used in the chapter analysis.

increase to the west and north. A hexagon has at most six contiguous neighbors. Hexagons at the borders have fewer neighbors. Figure 5.7 shows a graph of the hexagon connectivity, here defined by first-order contiguity.

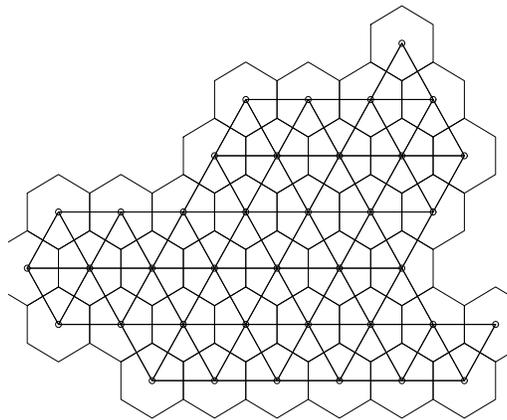


Figure 5.7: Graph of hexagon connectivity defined by first-order contiguity

The weights are the inverse of the number of neighbors. For instance, the six neighbors of a fully connected hexagon each have a weight of  $\frac{1}{6}$ . Now that a neighborhood matrix and a weighting scheme exists, the value of Moran's  $I$  can be computed. The value of 0.53 indicates positive spatial autocorrelation in the frequency, or rate, of occurrence at the hexagons. The value of 0.66 indicates positive spatial autocorrelation in the  $\sigma$ , or scale, parameter. The value of 0.19 suggests there is very little spatial correlation of the dimensionless  $\xi$  parameter. Finally, the value of 0.71 indicates positive spatial autocorrelation for the 30 year return level at the hexagons.

The positive spatial autocorrelation witnessed in these parameters is to be expected for two reasons. First, the data are arbitrarily subset into hexagons for this project. An individual hurricane passed through a series of hexagons and a maximum wind speed value per hexagon was recorded. A maximum occurring in one hexagon from a given hurricane will likely be very similar to another maximum occurring from the same hurricane at a nearby hexagon. The second cause for the positive spatial autocorrelation is the physical mechanics that control hurricane formation. Warm SSTs are necessary for a hurricane to form. If a given hurricane event traveled over warm SSTs, and a different hurricane event took a similar path over the same warm

temperatures at a different time, both of these hurricanes could have similarly high maximum wind speed values.

Insight into the Moran’s  $I$  statistic is obtained by noting that it is equivalent to the slope coefficient in a regression model of  $Wy$  on  $y$ , where  $Wy$  is called the spatial lag variable (see Equation 5.4). Let  $y_i$  be the set of parameters in each hexagon when  $i = \lambda, \sigma, \xi$  and the thirty year return level. A spatial lag variable for each  $i$  can be created using the functionality in R (R Development Core Team, 2010). For each parameter value there is a corresponding value representing the mean parameter over the neighboring hexagons. For a completely connected hexagon the average is taken over the adjoining six neighboring values.

Scatter plots of the neighborhood average versus the latent variables are shown in Figure 5.8. Figure 5.8(a) shows the rate parameter, Figure 5.8(b) shows the  $\sigma$  parameter, Figure 5.8(c) shows the  $\xi$  parameter, and Figure 5.8(d) shows the thirty year return level. In each plot, the slope of the least-squares regression line through the points is the value of Moran’s  $I$ . The scatter plot of the rate,  $\sigma$ ,  $\xi$ , and the thirty year return level suggests the possibility of a significant amount of spatial autocorrelation. Hexagons with high values tend to be surrounded, on average, by other hexagons with high values and vice versa as evidenced by the upward slope of the regression lines. The  $\xi$  value is the least spatially correlated. This is likely due to the sensitivity of  $\xi$  to the most extreme events.

To test the significance of these Moran values, the expected value of Moran’s  $I$  under the hypothesis of no spatial autocorrelation is

$$E(I) = \frac{-1}{n-1} \tag{5.5}$$

where  $n$  is the number of hexagons. This allows the testing of the significance of the sample Moran’s  $I$ . The output shows the standard deviate of Moran’s  $I$  computed as  $I$  minus the expected value of  $I$  divided by the square root of its variance. The  $p$ -value is the chance of observing a standard deviate this large or larger assuming there is no spatial autocorrelation. For the rate and  $\sigma$  parameters, and the thirty year return level, the  $p$ -values are extremely small leading to a conclusion that there is significant autocorrelation of these parameters. For the  $\xi$  parameter, the  $p$ -value of 0.01 provides moderate evidence to reject the null hypothesis of zero autocorrelation. Moran’s  $I$  and the corresponding significance test are sensitive to the definition of neighbors and to the neighborhood weights, so conclusions stated are conditional on the neighborhood definition.

Understanding the dependency of the parameters is important in order to truly understand the conclusions gained from the study. The goal of the final portion of analysis is to test the relationship between the different latent variables and the ocean’s surface temperature. This can provide insight into the effect that a warmer SST value might have on a particular hurricane characteristic.

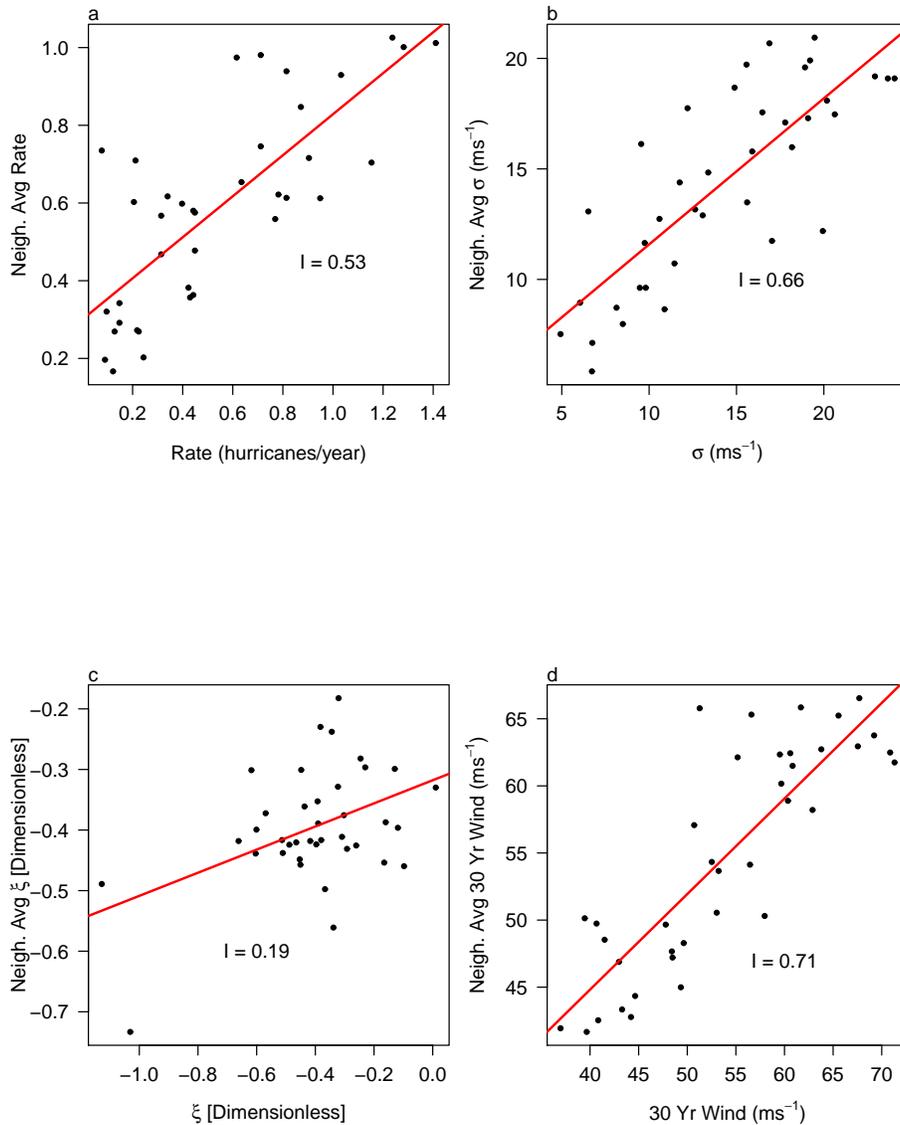


Figure 5.8: Moran's scatter plot. (a) Scatter plot of the per hexagon annual probability versus the neighborhood average annual probability. (b) Scatter plot of the per hexagon scale parameter,  $\sigma$ , versus the neighborhood average  $\sigma$ . (c) Scatter plot of the per hexagon shape parameter,  $\xi$ , versus the neighborhood average  $\xi$ . (d) Scatter plot of the per hexagon thirty year return level versus the neighborhood average thirty year return level. Neighbors are defined by spatial contiguity. The slope of the best-fit line is Moran's  $I$  as a measure of spatial autocorrelation.

## 5.4 Local Parameter Relationship with SST

Models that exploit spatial autocorrelation are called spatial regression models. If significant autocorrelation exists, spatial regression models have parameters that are more stable and statistical tests that are more reliable than non-spatial alternatives. A GWR model allows the relationship between the response and the explanatory variable to vary across the domain (Brunsdon et al., 1998; Fotheringham et al., 2000). GWR allows one to see where an explanatory variable contributes strongly to the relationship and where it contributes weakly. It is similar to a local linear regression.

With GWR the SST parameter is replaced by a vector of parameters, one for each hexagon. The relationship between the response vector and the explanatory variables is expressed mathematically as

$$y = X\beta(g) + \varepsilon \quad (5.6)$$

where  $g$  is a vector of geographic locations, here the set of hexagons with different latent variables and

$$\hat{\beta}(g) = (X^T W X)^{-1} X^T W y \quad (5.7)$$

where  $W$  is a weights matrix given by

$$W = \exp\left(\frac{-D^2}{h^2}\right) \quad (5.8)$$

where  $D$  is a matrix of pairwise distances between the hexagons and  $h$  is the bandwidth. The elements of the weights matrix,  $w_{ij}$ , are proportional to the influence hexagons  $j$  have on hexagons  $i$  in determining the relationship between  $X$  and  $y$ . Weights are determined by an inverse-distance function (kernel) so that values in nearby hexagons have greater influence on the local relationship compared with values in hexagons farther away. The bandwidth controls the amount of smoothing. It is chosen as a trade-off between variance and bias. A bandwidth too narrow (steep gradients on the kernel) results in large variations in the parameter estimates (large variance). A bandwidth too wide leads to a large bias as the parameter estimates are influenced by processes that do not represent the conditions locally. Here, an adaptive bandwidth is chosen that allows the estimates to vary depending on the location of the samples.

It is interesting to see the percent change in the intercept values of each variable across the hexagons. This suggests which latent variable is more sensitive to SST. There is an 18.5% change in the  $\lambda$ , a 4.0% change in the  $\sigma$ , a 62.4% change in the  $\xi$ , and a 5.6% change in the thirty year return level. This suggests that the  $\xi$  parameter is the most sensitive to SSTs. The  $\xi$  parameter represents the most extreme events, so these results suggest that the maximum intensity of hurricanes is most sensitive to a changing SST value.

The results for the GWR model of SST on  $\lambda$  are shown in Figure 5.9. In Figure 5.9 it can be seen the rate of hurricane occurrence is only affected by SSTs near the

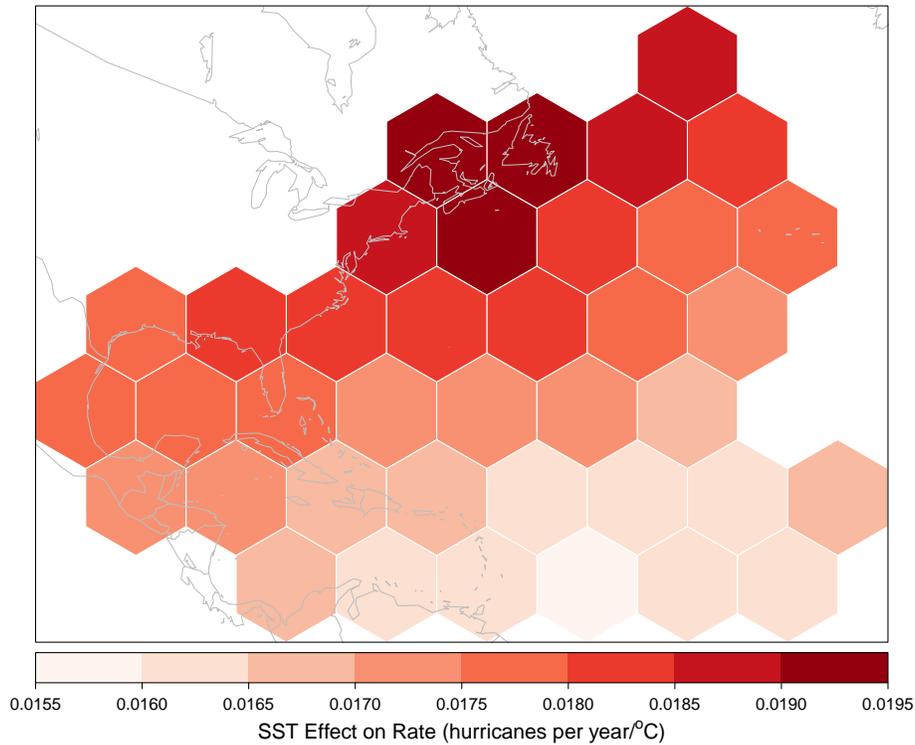


Figure 5.9: Marginal effect of SST on the number of hurricanes per year (rate).

northern perimeter of the domain. This is because hurricanes do not occur very often in these locations now because the required environmental conditions (i.e., the SST values) are not met during most of the season. Any increase in the SST values will alter the frequency of hurricanes in these hexagons.

The results for the GWR model of SST on  $\sigma$  are shown in Figure 5.10. In Figure 5.10 it is shown that the scale of hurricane wind speeds over the Caribbean Sea, GoM, and western Atlantic are influenced by SST values more so than the other hexagons. This suggests that as SSTs increase in warmth, the range of wind speeds occurring in these hexagons will also increase. That is, more hurricanes of differing magnitudes will occur.

The results for the GWR model of SST on  $\xi$  are shown in Figure 5.11. In Figure 5.11 it is shown that the parameter representing the most extreme values is most heavily influenced by SST in the GoM, Caribbean Sea, and the very western portion of the Atlantic. This suggests that as SSTs increase, the most extreme events will become even more intense.

Finally, the results for the GWR model of SST on the thirty year return level are shown in Figure 5.12. In Figure 5.12, it is shown that SST values have the greatest influence, again, over the western Atlantic, GoM, and Caribbean Sea. As

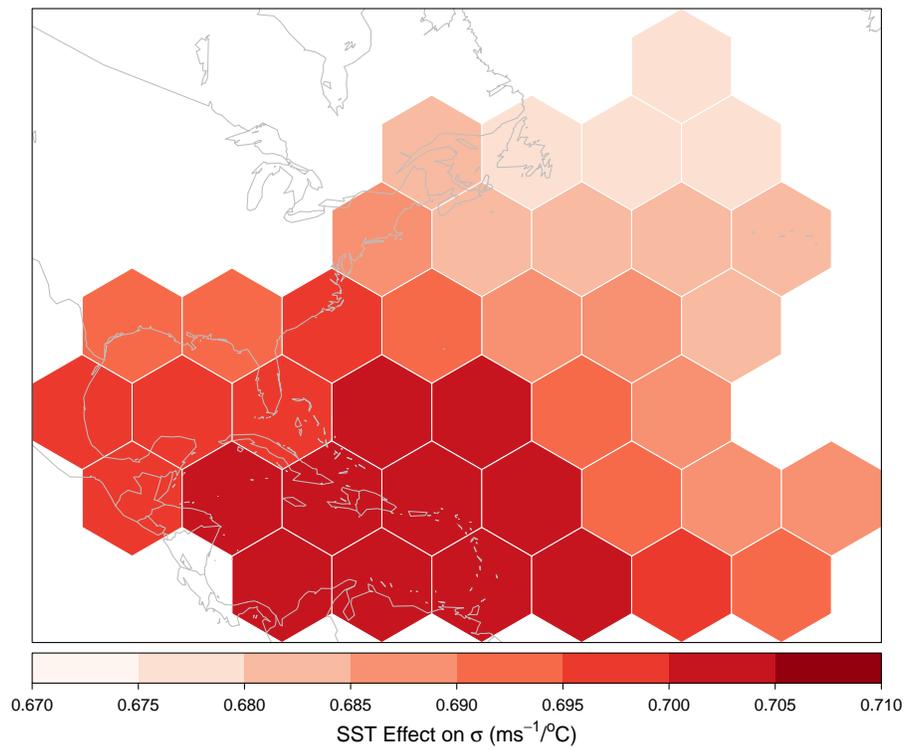


Figure 5.10: Marginal effect of SST on  $\sigma$  ( $\text{ms}^{-1}$ ).

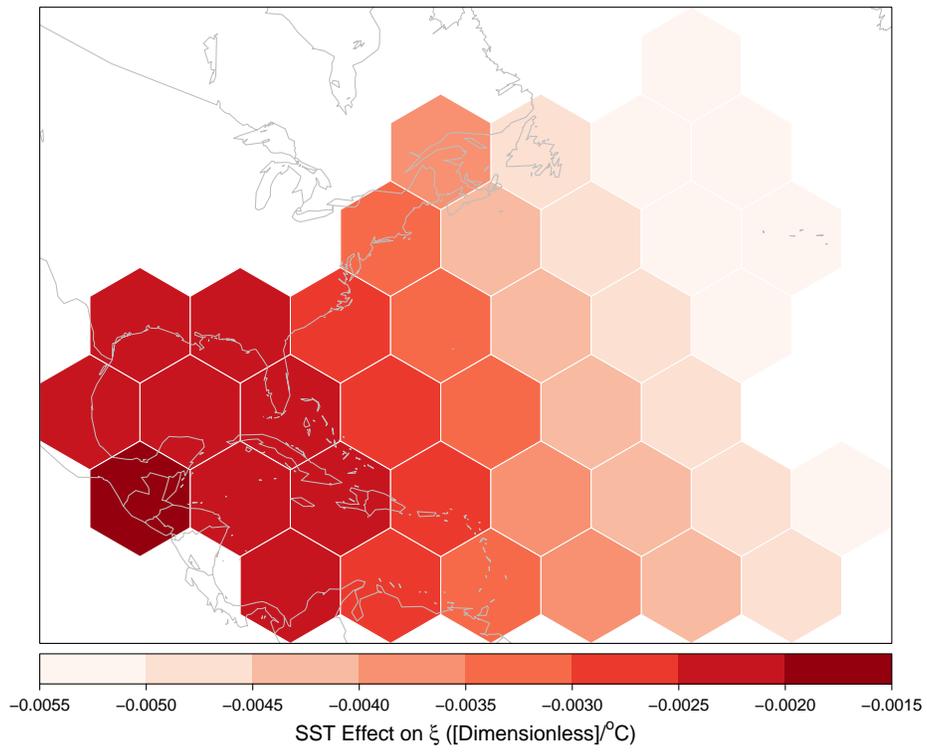


Figure 5.11: Marginal effect of SST on  $\xi$  [Dimensionless].

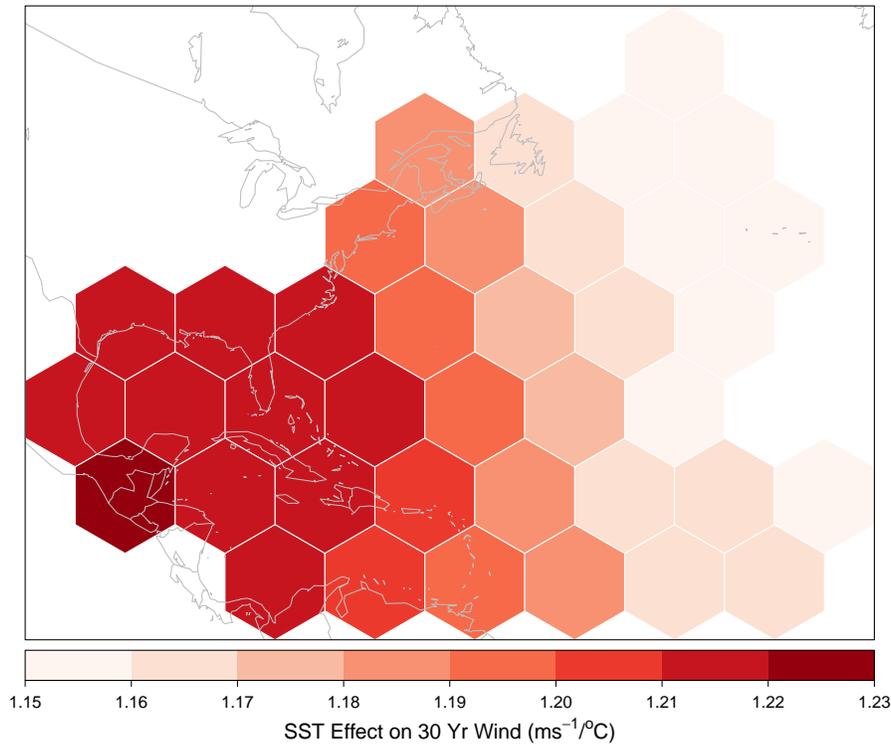


Figure 5.12: Marginal effect of SST on the thirty year return level ( $\text{ms}^{-1}$ ).

SSTs increase, the expected return level for a fixed time period will increase.

Each of these plots suggests the overall influence of the ocean’s surface temperature on individual hurricane characteristics. This provides an insight into hurricane characteristics that only a geographic approach can supply.

## 5.5 Spatial Risk Summary

Hurricane wind speeds are mapped using a hexagonal tessellation over the GoM and NAtl Ocean. This approach provides a unique insight into the way that hurricane characteristics vary over space. A GPD extreme value model is used to calculate parameters of interest using the maximum hurricane wind speed values known to occur in each hexagon. Specifically, the  $\lambda$ , or rate, the  $\sigma$ , or scale, and the  $\xi$ , or shape, parameters are catalogued. The rate represents the number of expected hurricanes per year and is highest over Florida, Bermuda, and the western Atlantic. These locations can expect the highest number of hurricanes exceeding  $33 \text{ ms}^{-1}$  in any given year. The scale represents the dispersion of wind speeds occurring in each hexagon and is highest over the Florida peninsula, the Western Antilles Islands, and the southwestern

Caribbean Sea. These locations experience the widest range of wind speeds, meaning they receive many different levels of hurricanes. The shape parameter represents the tail end of the distribution, where stable model values nearest to  $-1$  suggest the most extreme tails. The locations over the Florida peninsula and Western Antilles Islands experience the most extreme events.

Not only are the latent parameters visualized, but the thirty year return level is as well. Thirty years is chosen to represent the average homeowner's mortgage. The results estimated can provide potential and current homeowners with their overall likelihood of hurricane occurrence during the time they might own a home. The western Caribbean Sea and the Gulf of Mexico experience the highest return level for thirty years. The hexagons surrounding these locations also experience thirty year return levels exceeding  $50 \text{ ms}^{-1}$ , or a Category 3 hurricane.

The effect of the MAUP is tested by visualizing these parameters at different sized hexagons. The size chosen for this study was based on a trade off between having enough data at the locations (not too small), while still having a regional application for each hexagon (not too large). The effect of spatial autocorrelation was calculated using the Moran's  $I$  statistic. It was found that the rate,  $\sigma$ , and thirty year return level are all significantly autocorrelated throughout space. Due to the nature of how hurricanes form, and how these data were subset, this is to be expected. The  $\xi$  parameter had only moderate significance of minimal autocorrelation. This difference is likely due to the particular sensitivity of the  $\xi$  parameter to the most extreme of events.

The final portion of analysis was to test the relationship between the average August through October SST values per hexagon with each of the parameters and the thirty year return level using a GWR model. Based on the rate of change in the intercept value of the GWR models, the  $\xi$  parameter is the most sensitive to a change in SST values. The model suggests that as SSTs increase, the range of wind speeds occurring in the hexagons over the Caribbean Sea, the GoM, and the western Atlantic, will also increase. That is, more hurricanes will occur of differing magnitudes. The model for the thirty year return level suggests, again, that as SSTs increase, the expected return level for a fixed time period will increase over those same locations.

Based on these results, it can be suggested that as the oceans' surfaces increase in temperature, the expected maximum intensity for hurricanes will increase over the GoM, the Caribbean Sea, and the western Atlantic. This is important for any population of people living in these locations because it provides them with a deeper understanding of the expected risk as they move into the future. The unique approach offered by this study allows for a spatial visualization of hurricane characteristics that has never been attempted before. By visualizing the latent variables of the GPD model in this way, the geographic element of individual hurricane characteristics is supplied.

# CHAPTER 6

## CONCLUDING REMARKS

Throughout this dissertation, multiple methodologies have been presented to calculate the statistical risk of hurricane characteristics at a particular location. Specific locales were focused on, but these methodologies have the benefit of being useful wherever a hurricane record can be produced. In this concluding chapter, the methods and results of the three studies are summarized. The implications of these results are discussed as they relate to society and overall vulnerability. The vulnerability of any system (at any scale) is a function of the exposure and sensitivity of that system to any hazardous condition, as well as the ability or capacity of the system to cope with and adapt from the effects of that condition (Parry, 2002). In this case, the hazardous condition refers to the hurricane strike itself with the characteristic high wind speeds and deep storm surges. A discussion is offered relating this particular hazardous condition to societal vulnerability.

### 6.1 Study Summaries

In Chapter 3, extreme value theory was used to create a statistical model that can calculate the probability of hurricane winds in different cities around Florida. In particular, this extreme value model was based on the peaks over threshold approach, where only wind speeds above a certain threshold are considered in the estimation. All values over a specified threshold fit the generalized Pareto distribution. This particular approach is useful when dealing with hurricanes because lower wind speed levels are expected in many areas where hurricanes are common, but the most extreme events are fairly rare. This thresholding approach allows for particular insight into the most extreme events that other extreme value distributions, such as the Weibull, do not. The model created, the Hurricane Risk Calculator, used wind speeds from neighboring locations to calculate return periods (frequency of occurrence) of particular return levels (magnitudes) for a 100 km distance around twelve Florida cities. Uncertainty for these results was provided using the delta method, which assumes that the estimates are asymptotically normal from a knowledge of the limiting variance of the estimator. More broadly stated, the delta method is a fairly general

central limit theorem. It was found that Miami experiences hurricane wind speeds that blow at  $50 \text{ ms}^{-1}$  (45.5–54.5) [90% CI] or stronger, on average, once every twelve years. In comparison, Pensacola experiences hurricane wind speeds that blow at  $50 \text{ ms}^{-1}$  (46.9–53.1) [90% CI] or stronger, on average, once every 24 years. The second portion of this study assessed whether or not hurricane wind speeds and frequency in the vicinity of Florida have been changing over the period of record. A quantile regression is applied to the hurricane wind speeds around Florida and results show that the strongest hurricanes are getting stronger as a consequence of higher offshore intensification rates. This suggests that the return period estimates might be too conservative, or too low, as the intensity of hurricane winds increases into the future.

In Chapter 4, the probability of hurricane storm surge occurrences was estimated using the same peaks over threshold approach as used in the first study. This was conducted at three Air Force Bases along the U. S. Gulf coast to provide the overall statistical risk of surge events at various military infrastructures. It is found that Eglin Air Force Base near Valparaiso, Florida can expect a 2.4 m (1.4–3.4) surge, on average, once every 10 years. MacDill Air Force Base in Tampa, Florida can expect a 2.1 m (1.3–3.0) surge, on average, once every 10 years. Keesler Air Force Base in Biloxi, Mississippi can expect a 2.8 m (0.90–4.6) surge, on average, once every 10 years. These results suggest that the Mississippi coastline is more at risk of higher storm surge events than either the panhandle or western coast of Florida. The second portion of this study was aimed at creating a bivariate statistical approach that could assess the local risk of hurricane storm surge *and* hurricane wind speeds. This bivariate probability was quantified using a copula model on observed data. The copula has marginals with Weibull distributions and a correlation parameter for the dependent relationship between wind and surge. The quartile pointwise uncertainty is quantified using a Monte Carlo procedure. Eglin Air Force Base can expect wind speeds blowing at  $50 \text{ ms}^{-1}$  and surge heights of 3 meters, on average, once every 28 years (23–36). MacDill Air Force Base can expect wind speeds blowing at  $50 \text{ ms}^{-1}$  and surge heights of 3 meters, on average, once every 27 years (22–34). Keesler Air Force Base can expect wind speeds blowing at  $50 \text{ ms}^{-1}$  and surge heights of 3 meters, on average, once every 15 years (13–18). From these results, it is suggested that Keesler Air Force Base has the highest likelihood of extreme wind speeds and high storm surges at the shortest time periods when comparing these three locations.

In the final study of this dissertation, Chapter 5, a unique spatial approach was offered for the assessment of hurricane risk. Hexagonal shapes were used to tessellate the Gulf of Mexico and North Atlantic Ocean and the likelihood of wind speed occurrence was mapped over space. The peaks over threshold model presented in the previous chapters was used to estimate the return periods for hurricane winds at each hexagon over the domain. In calculating this model, three parameters are produced, as well as the specified return level estimates. The first parameter,  $\lambda$ , estimates the rate of hurricane occurrence per year at each hexagon using the available maximum wind speed data. The second parameter,  $\sigma$ , estimates the scale (statistical) of the available wind speed data. In other words, it represents the overall dispersion of the

values. The third parameter,  $\xi$ , is a dimensionless quantity that represents the tail end of the distribution of wind speeds. Stated another way,  $\xi$  represents the most extreme events (those that do not occur very often). Each of these parameters, as well as the thirty year return level are mapped out in each hexagon over the spatial domain. By visualizing the parameters in this way, the locations most affected by the particular characteristics become apparent.  $\lambda$  is highest over the Florida, Bermuda, and the western Atlantic. These locations can expect the highest number of hurricanes exceeding  $33 \text{ ms}^{-1}$  in any given year.  $\sigma$  is highest over the Florida peninsula, the Western Antilles Islands, and the southwestern Caribbean Sea. These locations experience the widest range of wind speeds, meaning they receive many different levels of hurricanes.  $\xi$  is lowest (values closer to  $-1$  represent extremes) over the Florida peninsula and Western Antilles Islands. These locations experience the most extreme events. The thirty year return level is highest over the western Caribbean and the Gulf of Mexico, suggesting that these locations experience the highest wind speeds on a fixed temporal scale.

Using a geographically weighted regression model, the relationship between the average August through October sea surface temperatures per hexagon with each of the parameters and the thirty year return level is tested. This is done to estimate the overall influence of sea surface temperatures on individual hurricane characteristics. The significance of the results was estimated using the  $t$  statistic. The model suggests that as sea surface temperatures increase, the range of wind speeds occurring in the hexagons over the Caribbean Sea, the GoM, and the western Atlantic, will significantly increase. That is, more hurricanes will occur of differing magnitudes. The model for the thirty year return level suggests that as the water temperatures increase, the expected return level for a fixed time period will significantly increase. Both the  $\lambda$  and  $\xi$  parameters did not have a significant relationship with sea surface temperature.

Each of these studies provided a statistical assessment of hurricane risk at various locales. Now that each has been presented and the results are understood, it is interesting to consider what these estimates suggest for the societies living in the locations experiencing the highest risk. The next, and final, section of this dissertation provides a brief discussion of how hurricane hazards relate to societal vulnerability.

## 6.2 Societal Risk

Adaptation, in evolutionary biology, refers to the development of a genetic characteristic that allows the organism to prevail or survive an environmental stressor or change. Adaptation, in the human context, refers to a process, outcome, or action that can better the system's ability to cope with, manage, or adjust to some changing condition (Smit and Wandel, 2006). O'Brien and Holland (1992) believe that the process of adaptation occurs when a group of people add new and improved methods of coping to their environment to their "cultural repertoire". Denevan (1983)

believed this adaptation is a process of change in response to a shift in the physical surroundings or internal processes (like economics). The latter definition broadens the range of the types of stresses/changes that can influence a group of people to adapt. Denevan (1983) believed that any society that can quickly respond/cope to a change are considered to have high “adaptability” or “capacity to adapt”. Though inequality exists throughout the United States, for the most part there are enough advanced warning systems and emergency management that unexpected changes can be responded to and coped with quickly. The unexpected change of interest here; hurricane strikes. Of course, the interaction between the environment and human societies is only one constraint of many that influence the vulnerability of a particular population to various social, economic, political, and environmental risks (Fraser et al., 2003). However, this dissertation provides a more thorough understanding of the environmental risks that play a part in establishing vulnerability. Human societies can use the information presented here to change how they interact with hurricane strikes to lead to less life loss and less monetary loss.

Human societies can do little to defend in the way of the physical mechanisms necessary for creating hurricanes. It is unlikely that hurricane mechanics could ever be stopped or redirected by man. It is possible, however, to advance our understanding of the occurrence of these events so that societies can adapt and be better prepared when an event does happen. The studies in this dissertation provide the knowledge of the statistical risk of events at the various locations of interest, so that they may adapt or change their preparation for a hurricane event. The code provided in the Appendix allows for the combined risk methodology to be repeated at any location that can provide a data set of hurricane winds and surges (or any other co-occurring variables in a phenomenon), suggesting that this methodology can be applied at any location experiencing hurricanes, typhoons, cyclones, etc.

One of the forefront research areas for adaptation deals directly with climate change. One of the main concerns is how our policies might be able to offset any negative consequences of climate change by including some adaptation (or mitigation) to the expected effects (Parry, 2002). The United Nations Framework Convention on Climate Change, a convention that commits countries to mitigate greenhouse gas emissions, considers adaptations from this perspective: to what degree can a particular adaptation lessen the expected negative impacts of climate change? It is not the purpose of this dissertation to supply any particular adaptation methods, but rather provide the risk information necessary so those who create policies have as much data as possible when making decisions. In particular, a few portions of the studies look directly at the influence of climate change on hurricane occurrence. For example, in Chapter 3, the intensity of the most extreme hurricanes was found to be increasing through time. All else being equal, if sea surface temperatures are known to be increasing, it is possible that an increased surface temperature will make hurricanes more extreme. The results presented in Chapters 3 and 5 suggest that as the climate changes and sea surface temperatures increase, hurricane maximum intensity will be affected trending toward more extreme events.

Research that focuses on the relative adaptive capacity of a particular region or group of people considers vulnerability as the beginning points (rather than the outcome) (Parry, 2002). This helps focus attention specifically on the regions that are particularly vulnerable to the effects of climate change. The intent of this type of research is to target where adaptation initiatives should take place. The information presented here provides insight into locations at highest risk of hurricane strikes. Statistical risk estimates are separate from societal vulnerability, but they can be used to understand the expected risk of hurricane strikes at a particular societal level. Policy and decision makers can use the information to adjust their policies to better represent the expected risk. Especially at a time when the threat of climate shifts is high, policy makers can use the information provided in this work to prepare for a more tumultuous hurricane future. It is possible for decision makers to allocate resources in an efficient way by focusing on the areas that have the highest probability of hurricane wind speeds, storm surges, or occurrences. Insurance companies can also use this information to adjust their premiums. Most importantly, individual citizens living in the societies analyzed can use the information to understand the level of hurricane risk they live in during the summer months. The goal is to provide them with enough information to lessen their risk by becoming better prepared.

# APPENDIX A

## APPENDIX

### A.1 Chapter 4 Code

Here are the steps in detail to create the copula presented in Chapter 4. The data can be found at [jtrepanier.wordpress.com/data](http://jtrepanier.wordpress.com/data). The code below is only for Eglin Air Force Base, but can be altered to adjust for the location and variables of interest.

Read in the data and grab the wind and surge variables of interest. Remove any NA values.

```
> dateafb = read.table("EAFBfinal.txt", header = T)
> dateafb = data.frame(W = eafb$WmaxMS, S = eafb$SmaxM)
> dateafb = na.omit(dateafb)
```

Next the correlation between wind and surge is determined.

```
> r = cor(dateafb$W, dateafb$S)
```

Next, acquire the **copula** package in R (Yan, 2007; Kojadinovic and Yan, 2010) and define a two-dimensional normal copula using the *ellipCopula* function with the parameter equal to the sample correlation value. The dimension is the number of variables, here being wind and surge.

```
> library("copula")
> myCopula = ellipCopula(family = "normal", dim = 2,
+   param = r)
```

Next fit the wind and surge data to separate Weibull distributions and obtain the parameter values for these marginals using the maximum-likelihood procedure in the *fitdistr* function (**MASS**) (Venables and Ripley, 2002).

```
> library(MASS)
> parW = fitdistr(dateafb$W, "weibull")
> parS = fitdistr(dateafb$S, "weibull")
```

Then generate the multivariate density from the copula definition and the Weibull parameters (shape  $\xi$  and scale  $\sigma$ ) using the *mvdc* function (**copula**).

```
> myMvd = mvdc(myCopula, margins = c("weibull",
+   "weibull"), paramMargins = list(list(shape = parW$est[1],
+   scale = parW$est[2]), list(shape = parS$est[1],
+   scale = parS$est[2])))
```

Next, fit the copula-based multivariate density to the data using a maximum-likelihood procedure in the *fitMvdc* function (**copula**). The first argument is the data (as a matrix), the second is a multivariate copula with Weibull margins, and the third is a vector containing the original parameter estimates.

```
> start = c(parW$est, parS$est, r)
> model = fitMvdc(as.matrix(dateafb), myMvd, start = start,
+   optim.control = list(trace = TRUE, maxit = 20))
```

Finally, redefine the normal copula using the updated correlation parameter value and generate the multivariate density from the new copula definition and the updated Weibull parameter values. This provides a fitted copula model using the data.

```
> myCopula2 = ellipCopula(family = "normal", dim = 2,
+   param = model@estimate[5])
> myMvd2 = mvdc(myCopula2, margins = c("weibull",
+   "weibull"), paramMargins = list(list(shape = model@estimate[1],
+   scale = model@estimate[2]), list(shape = model@estimate[3],
+   scale = model@estimate[4])))
```

The density is plotted on a two-dimensional grid spanned by the range of wind and surge values. This is done using the *contour* function (**graphics**) and the *smvdc* function (Becker et al., 1988). The *smvdc* function was created to use in conjunction with *contour*, which determines the joint probability of the wind and surge based on the copula density. This can be found for download at [jtrepanier.wordpress.com/data](http://jtrepanier.wordpress.com/data).

```
> source("scopula.R")
> ctr = contour(myMvd2, smvdc, xlim = c(20, 70),
+   ylim = c(0, 8))
```

The per event probability,  $[\Pr(W > w_{max}, S > s_{max})]$  times the hurricane frequency (the number of hurricanes divided by the total number of years), provides us with the annual event frequency. That is, the yearly frequency of events with  $W$  exceeding  $w_{max}$  and  $S$  exceeding  $s_{max}$ . The yearly probability of an event is then  $1 - \exp(-\text{frequency})$ , where  $\exp()$  is the exponent function, and the return period is 1 divided by  $1 - \exp(-\text{frequency})$ , or roughly 1 divided by the frequency plus 0.5 year. For small frequencies, the return period is roughly 1 divided by the frequency. This provides the probability of specific wind *and* surge events.

```

> rate = nrow(dateafb)/(2005 - 1896)
> ctr$z = 1/(ctr$z * (1 - ppois(0, rate)))

```

Then plot out the density of the copula model and include the empirical points.

```

> contour(myMvd2, dmvdc, xlim = c(20, 65), ylim = c(0,
+   6), xlab = expression(paste("Maximum Sustained Wind Speed (m",
+   s^-1, ")")), ylab = "Maximum Surge Height (m)",
+   las = 1, cex.axis = 1.4, cex.lab = 1.4, col = "white",
+   drawlabels = FALSE)
> abline(h = seq(0, 6, 2), v = seq(20, 70, 10),
+   col = "lightgray")
> contour(myMvd2, dmvdc, xlim = c(20, 65), ylim = c(0,
+   6), xlab = expression(paste("Maximum Sustained Wind Speed (m",
+   s^-1, ")")), ylab = "Maximum Surge Height (m)",
+   las = 1, cex.axis = 1.4, cex.lab = 1.4, add = T)
> points(dateafb[, 1], dateafb[, 2], pch = 16)

```

Then plot out the yearly probability of an event and include the empirical points.

```

> contour(ctr, levels = c(1, 2, 5, 10, 20, 30,
+   50, 75, 100, 150, 200, 300, 500), main = "",
+   ylab = "Maximum Surge Height (m)", xlab = expression(paste("Maximum Sustained Wind Speed (m",
+   s^-1, ")")), xlim = c(40, 65), ylim = c(3,
+   6), drawlabels = FALSE, cex.axis = 1.4,
+   cex.lab = 1.5, las = 1, labcex = 1.5, col = "white")
> abline(h = seq(3, 6, 1), v = seq(40, 70, 10),
+   col = "lightgray")
> contour(ctr, levels = c(1, 2, 5, 10, 20, 30,
+   50, 75, 100, 150, 200, 300, 500), main = "",
+   ylab = "Maximum Surge Height (m)", xlab = expression(paste("Maximum Sustained Wind Speed (m",
+   s^-1, ")")), xlim = c(40, 65), ylim = c(3,
+   6), drawlabels = TRUE, cex.axis = 1.4,
+   cex.lab = 1.5, las = 1, labcex = 1.5, add = T)
> points(dateafb[, 1], dateafb[, 2], pch = 16)

```

Calculate the CI. This will take some time. Read a summary of the results.

```

> cs = mvrnorm(1000, mu = model@estimate, Sigma = model@var.est)
> cs[, 5] = ifelse(cs[, 5] > 1, 1, cs[, 5])
> mar = c("weibull", "weibull")
> rp = numeric()
> wt = c(33, 43, 50, 59)
> st = c(3.04, 4, 4.96)

```

```

> ns = 1000
> for (j in 1:ns) {
+   shw = cs[j, 1]
+   scw = cs[j, 2]
+   shs = cs[j, 3]
+   scs = cs[j, 4]
+   cc = cs[j, 5]
+   mC = ellipCopula(family = "normal", dim = 2,
+     param = cc)
+   mM = mvdc(mC, margins = mar, paramMargins = list(list(shape = shw,
+     scale = scw), list(shape = shs, scale = scs)))
+   mysim = contour(mM, smvdc, xlim = c(20, 70),
+     ylim = c(0, 8))
+   mysim$z = 1/(mysim$z * (1 - ppois(0, rate)))
+   for (i in 1:4) {
+     for (k in 1:3) {
+       rp = c(rp, mysim$z[mysim$x == wt[i],
+         mysim$y == st[k]])
+     }
+   }
+ }
> rpSamples = t(array(rp, c(12, ns)))

```

## A.2 Chapter 5 Code

The following is the code showing how to work with spatial data sets in R (R Development Core Team, 2010), as was done in Chapter 5. The data can be found at [jtrepanier.wordpress.com/data](http://jtrepanier.wordpress.com/data). An explanation of the best.use data can be found in (Jagger and Elsner, 2006). The code to calculate the model parameters within Chapter 5 is not provided. This shows you how to deal with spatial data, as well as read in and subset SST data.

First load in the cyclone data set and subset to look at specified years of interest. Change the wind speeds (as knots) to meters per second and only grab hurricanes ( $\geq 33 \text{ ms}^{-1}$ ).

```

> load("best.use.RData")
> begin = 1854
> end = 2009
> years = begin:end
> Wind.df = best.use[best.use$Yr >= begin & best.use$Yr <=
+   end, ]
> Wind.df$WmaxS = Wind.df$WmaxS * 0.5144

```

```

> threshold = 33
> Wind.df = Wind.df[Wind.df$WmaxS >= threshold,
+   ]

```

Load the package for spatial data and assign coordinates and projection information.

```

> require(sp)
> coordinates(Wind.df) = c("lon", "lat")
> ll = "+proj=longlat +datum=WGS84"
> proj4string(Wind.df) = CRS(ll)
> slot(Wind.df, "coords")[1:3, ]

```

Transform the map to Lambert Conformal Conic.

```

> lcc = "+proj=lcc +lat_1=38 +lat_2=23 +lon_0=-60"
> require(rgdal)
> Wind.sdf = spTransform(Wind.df, CRS(lcc))
> slot(Wind.sdf, "coords")[1:3, ]

```

Generate hexagon centers and create hexagon polygons.

```

> hpt125 = spsample(Wind.sdf, type = "hexagonal",
+   n = 125, bb = bbox(Wind.sdf) * 1.2, offset = c(0,
+   -1))
> hpg125 = HexPoints2SpatialPolygons(hpt125)

```

Next overlay hexagons on the hurricane locations.

```

> Wind.hexid125 = overlay(Wind.sdf, hpg125)

```

Then split the data frame by hexagon number.

```

> Wind.split125 = split(Wind.sdf@data, Wind.hexid125)

```

Create a vector of ID names for the data frame.

```

> hexID125 = paste("ID", names(Wind.split125),
+   sep = "")

```

Finally, create a spatial polygons data frame with a dummy slot in the data location. This allows you to fill further data slots with whatever information you choose. For example, with maximum wind speeds or model parameters (as was done in Chapter 5). Now, each hexagon over a location of interest has geographic coordinates associated with it as well as whatever data information entered.

```

> hexID125 = as.integer(names(Wind.split125))
> int125 = rep(3, length(hexID125))
> int125 = as.data.frame(int125)
> names(int125) = "test"
> row.names(int125) = paste("ID", hexID125, sep = "")
> hspdf125 = SpatialPolygonsDataFrame(hpg125[hexID125,
+ ], int125, match.ID = TRUE)
> hspdf125@data

```

The following code shows how to read in SST data and pull out the values of interest. The data can be found at [jtrepanier.wordpress.com/data](http://jtrepanier.wordpress.com/data).

First load the data and grab the August through October monthly SST values per grid location.

```

> load("ncdataframe.RData")
> sehur = paste("Y", years, sep = "")
> sstAug = ncdataframe[paste(sehur, "M", formatC(8,
+ 1, flag = "0"), sep = "")]
> sstSep = ncdataframe[paste(sehur, "M", formatC(9,
+ 1, flag = "0"), sep = "")]
> sstOct = ncdataframe[paste(sehur, "M", formatC(10),
+ sep = "")]

```

Then average the monthly location values.

```

> sstAugmean = rowMeans(sstAug)
> sstSepmean = rowMeans(sstSep)
> sstOctmean = rowMeans(sstOct)

```

Get the average SST value from all three months and assign geographic coordinates.

```

> ASOsst = cbind(sstAugmean, sstSepmean, sstOctmean)
> avgsst = rowMeans(ASOsst)
> avgsst = as.data.frame(avgsst)
> avgsst$lon = ncdataframe$lon
> avgsst$lat = ncdataframe$lat

```

Project the SST values.

```

> coordinates(avgsst) = c("lon", "lat")
> proj4string(avgsst) = CRS(11)
> SST.sdf = spTransform(avgsst, CRS(1cc))

```

Calculate the spatial average of SSTs per hexagon.

```
> ssta = overlay(SST.sdf, hspdf125, fn = mean)
```

Then add the SST data to the spatial data frame created above. This is just one example of the type of data you can put into this slot.

```
> hspdf125$avgsst = ssta$avgsst
```

Should any questions or concerns exist about this code or the data available, please contact the author of this dissertation.

## REFERENCES

- Becker, R. A., J. M. Chamber, and A. R. Wilks, 1988: *The new S language*. Wadsworth and Brooks/Cole.
- Bin, O., J. Kruse, and C. Landry, 2008: Flood hazards, insurance rates, and amenities: Evidence from the coastal housing market. *Journal of Risk and Insurance*, **75**, 63–82.
- Bivand, R., 2011: *spdep: Spatial dependence: weighting schemes, statistics and models*. <http://CRAN.R-project.org/package=spdep>, R Foundation for Statistical Computing, r package version 0.5-40 ed.
- Blain, C., J. Westerink, and R. Luttich, 1994: The influence of domain size on the response characteristics of a hurricane storm surge model. *Journal of Geophysical Research*, **99**, 467–479.
- Blake, E., E. Rappaport, and C. Landsea, 2007: The deadliest, costliest, and most intense United States tropical cyclones from 1851 to 2006 (and other frequently requested hurricane facts). Tech. Memo to the National Weather Service, National Oceanic and Atmospheric Administration, 45 pp.
- Brettschneider, B., 2008: Climatological hurricane landfall probability for the United States. *Journal of Applied Meteorology and Climatology*, **47**, 704–716.
- Brunsdon, C., A. Fotheringham, and M. Charlton, 1998: Geographically weighted regression - modelling spatial non-stationarity. *Journal of the Royal Statistical Society: Series D-The Statistician*, **47**, 431–443.
- Butler, A., J. Heffernan, J. Tawn, R. Flather, and K. Horsburgh, 2007: Extreme value analysis of decadal variations in storm surge elevations. *Journal of Marine Systems*, **67**, 189–200.
- Caron, L., C. Jones, and K. Winger, 2011: Impact of resolution and downscaling technique in simulating recent Atlantic tropical cyclone activity. *Climate Dynamics*, **37**, 869–892.
- Carton, J. and A. Santorelli, 2008: Global upper ocean heat content as viewed in nine analyses. *J. Climate*, **21**, 6015–6035.

- Christaller, W., 1933: Die zentralen orte in süddeutschland. Ph.D. thesis, Gustav Fisher, Jena.
- Chu, P. and J. Wang, 1998: Modeling return periods of tropical cyclone intensities in the vicinity of Hawaii. *Journal of Applied Meteorology and Climatology*, **37**, 951–960.
- Coles, S., 2001: *An introduction to statistical modeling of extreme values*. Springer, 208 pp.
- Coles, S. and E. Casson, 1998: Extreme value modelling of hurricane wind speeds. *Structural Safety*, **20**, 283–296.
- Coles, S., J. Heffernan, and J. Tawn, 1999: Dependence measures for extreme value analyses. *Extremes*, **2**, 339–365.
- Coles, S. and E. Powell, 1996: Bayesian methods in extreme value modelling: a review and new developments. *International Statistical Review*, **64**, 119–136.
- Coles, S. and A. Stephenson, 2010: *ismev: An introduction to statistical modeling of extreme values*. R Foundation for Statistical Computing, URL <http://cran.r-project.org/web/packages/ismev/ismev.pdf>.
- Cooley, D., D. Nychka, and P. Naveau, 2007: Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, **102**, 824–840.
- Darling, R., 1991: Estimating probabilities of hurricane wind speeds using a large-scale empirical model. *Journal of Climate*, **4**, 1035–1046.
- Davison, A. and R. Smith, 1990: Models for exceedences over high thresholds. *Journal of the Royal Statistical Society: Series B*, **52**, 393–442.
- de Beurs, K. M. and G. Henebry, 2008: Northern annular mode effects on the land surface phenologies of northern Eurasia. *Journal of Climate*, **21**, 4257–4279.
- Denevan, W., 1983: Adaptation, variation, and cultural geography. *Professional Geographer*, **35**, 399–406.
- Donelan, M., W. Drennan, and K. Katsoros, 1997: The air-sea momentum flux in conditions of wind sea and swell. *Journal of Physical Oceanography*, **27**, 2087–2099.
- Elsner, J., R. Hodges, and T. Jagger, 2011: Spatial grids for hurricane climate research. *Climate Dynamics*, DOI: **10.1007/s00382-011-1066-5**.
- Elsner, J., T. Jagger, and K. b. Liu, 2008a: Comparison of hurricane return levels using historical and geological records. *Journal of Applied Meteorology and Climatology*, **47**, 368–374.

- Elsner, J. and A. Kara, 1999: *Hurricanes of the North Atlantic: Climate and society*. Oxford University Press, 488 pp.
- Elsner, J., J. Kossin, and T. Jagger, 2008b: The increasing intensity of the strongest tropical cyclones. *Nature*, **455**, 92–95.
- Emanuel, K., 2003: Tropical cyclones. *Annual Review of Earth and Planetary Sciences*, **31**, 75–104.
- Emanuel, K., S. Ravela, E. Vivant, and C. Risi, 2006: A statistical deterministic approach to hurricane risk assessment. *Bulletin of the American Meteorological Society*, **87**, 299–314.
- Embrechts, P., F. Lindskog, and A. McNeil, 2003: Modelling dependence with copulas and applications to risk management. *Handbook of Heavy Tailed Distribution in Finance*, S. Rachev, Ed., Elsevier, 329–384.
- Embrechts, P., S. Resnick, and G. Samorodnitsky, 1999: Extreme value theory as a risk management tool. *North American Actuarial Journal*, **3**, 30–41.
- Enfield, D., A. Mestas-Nunez, and P. Trimble, 2001: The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental U. S. . *Geophysical Research Letters*, **28**, 2077–2080.
- Fang, K., S. Kotz, and K. Ng, 1990: *Symmetric multivariate and related distributions*. Chapman and Hall.
- Fernandez-Partagas, J. and H. Diaz, 1996: Atlantic hurricanes in the second half of the nineteenth century. *Bulletin of the American Meteorological Society*, **77**, 2899–2906.
- Finney, M., C. McHugh, I. Grenfell, K. Riley, and K. Short, 2011: A simulation of probabilistic wildfire risk components for the continental United States. *Stochastic Environmental Research and Risk Assessment*, **25**, 973–1000.
- Fisher, R. and L. Tippett, 1928: Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, **24**, 180–190.
- Fotheringham, A., C. Brunson, and M. Charlton, 2000: *Quantitative Geography: Perspectives on Spatial Data Analysis*. Sage.
- Fraser, E., W. Mabee, and O. Slaymaker, 2003: Mutual vulnerability, mutual dependence The reflexive relation between human society and the environment. *Global Environmental Change*, **13**, 137–144.
- Fréchet, M., 1927: Sur la loi de probabilité de l'écart maximum. *Annales de la Societe Polonaise de Mathematique*, **6**, 93–116.

- Frees, E. and E. Valdez, 1998: Understanding relationships using copulas. *North American Actuarial Journal*, **2**, 1–25.
- Genest, C. and A. Favre, 2007: Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, **12**, 347–368.
- Gumbel, E., 1937: Les intervalles extrêmes entre les émissions radioactives. *J. Phys. Radium*, **8**, 446–452.
- Gumbel, E., 1941: The return period of flood flows. *Annals of Mathematical Statistics*, **12**, 163–190.
- Gumbel, E., 1958: *Statistics of extremes*. Columbia University Press, 400 pp.
- Haining, R., 2003: *Spatial Data Analysis Theory and Practice*. Cambridge University Press, 432 pp.
- Hamid, S., B. Kibria, S. Gulati, M. Powell, B. Annane, S. Cocke, J. Pinelli, and C. Chen, 2009: Predicting losses of residential structures in the state of Florida by the public hurricane loss evaluation models. *Statistical Methodology*, in press.
- Heckert, N., E. Simiu, and T. Whalen, 1998: Estimates of hurricane wind speeds by ‘peaks over threshold’ method. *Journal of Structural Engineering ASCE*, **124**, 445–449.
- Jagger, T. and J. Elsner, 2006: Climatology models for extreme hurricane winds in the United States. *Journal of Climate*, **19**, 3220–3236.
- Jagger, T. and J. Elsner, 2009: Modeling tropical cyclone intensity with quantile regression. *Int. J. Climatol.*, **29**, 1351–1361.
- Jagger, T., J. Elsner, and X. Niu, 2001: A dynamic probability model of hurricane winds in coastal counties of the United States. *Journal of Applied Meteorology*, **40**, 853–863.
- Joe, H., 1997: *Multivariate models and dependence concepts*. Chapman and Hall, London.
- Keim, B., R. Muller, and G. Stone, 2007: Spatiotemporal patterns and return periods of tropical storm and hurricane strikes from Texas to Maine. *Journal of Climate*, **20**, 3498–3509.
- Keitt, T., R. Bivand, E. Pebesma, and B. Rowlingson, 2012: *Bindings for the geospatial data abstraction library*. <http://cran.r-project.org/web/packages/rgdal/rgdal.pdf>, R Foundation for Statistical Computing, 0th ed.

- Knutson, T., J. Sirutis, S. Garner, I. Held, and R. Tuleya, 2007: Simulation of the recent multidecadal increase of Atlantic hurricane activity using an 18-km grid regional model. *Bulletin of the American Meteorological Society*, **88**, 1549–1565.
- Koenker, R. and G. Bassett, 1978: Regression quantiles. *Econometrica*, **46**, 33–50.
- Kojadinovic, I. and J. Yan, 2010: Modeling the multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software*, **34**, 1–20.
- Kotz, S. and S. Nadarajah, 2000: *Extreme value distributions: Theory and applications*. Imperial College Press, 185 pp.
- Landsea, C., 1993: A climatology of intense (or major) Atlantic hurricanes. *Monthly Weather Review*, **121**, 1703–1713.
- Landsea, C., R. Pielke Jr., A. Mestas-Nunez, and J. Knaff, 1999: Atlantic basin hurricanes: Indices of climatic changes. *Climatic Change*, **42**, 89–129.
- Landsea, C., et al., 2004: The Atlantic hurricane database reanalysis project: Documentation for the 1851–1910 alterations and additions to the HURDAT database. *Hurricanes and Typhoons: Past, Present, and Future*, R. Murnane and K. b. Liu, Eds., Columbia University Press, 177–221.
- Liu, D., L. Pang, and B. Xie, 2009: Typhoon disaster in China: prediction, prevention, and mitigation. *Natural Hazards*, **49**, 421–436.
- Malmstadt, J., J. Elsner, and T. Jagger, 2010: Risk of strong hurricane winds to Florida cities. *Journal of Applied Meteorology and Climatology*, **49**, 2121–2132.
- Malmstadt, J., K. Scheitlin, and J. Elsner, 2009: Florida hurricanes and damage costs. *The Southeastern Geographer*, **49**, 108–131.
- Masala, G., 2012: Earthquakes occurrences estimation through a parametric semi-Markov approach. *Journal of Applied Statistics*, **39**, 81–96.
- Moran, P., 1950: Notes on continuous stochastic phenomena. *Biometrika*, **37**, 17–33.
- Needham, H. F. and B. D. Keim, 2011: A storm surge database for the U. S. Gulf Coast. *International Journal of Climatology*, DOI:10.1002/joc.2425.
- Nelsen, R., 1998: *An introduction to copulas*. Springer Verlag, New York.
- Neumann, C., 1987: The National Hurricane Center risk analysis program (HURISK). Tech. Memo to the National Weather Service NHC, National Oceanic and Atmospheric Administration, 59 pp.

- Nordquist, J., 1945: Theory of largest values, applied to earthquake magnitudes. *Transactions of American Geophysical Union*, **26**, 29–31.
- O’Brien, M. and T. Holland, 1992: The role of adaptation in archeological explanation. *American Antiquity*, **57**, 36–69.
- Parisi, F. and R. Lund, 2008: Return periods of continental U.S. hurricanes. *J. Climate*, **21**, 403–410.
- Parry, M., 2002: Scenarios for climate impact and adaptation assessment. *Global Environmental Change*, **12**, 149–153.
- Pebesma, E., R. Bivand, B. Rowlingson, and V. Gomez-Rubio, 2011: *A package providing classes and methods for spatial data: points, lines, polygons and grids*. <http://trac.osgeo.org/proj/>, R Foundation for Statistical Computing, package sp version 0.9-84 ed.
- Pielke Jr., R., J. Gratz, C. Landsea, D. Collins, M. Saunders, and R. Musulin, 2008: Normalized hurricane damage in the United States: 1900–2005. *Natural Hazards Review*, **9**, 29–42.
- Pierce, D., 2011: *ncdf: Interface to Unidata netCDF data files*. <http://CRAN.R-project.org/package=ncdf>, R Foundation for Statistical Computing, r package version 1.6.6 ed.
- Potter, W., 1949: Normalcy tests of precipitation and frequency studies of runoff on small watersheds. Technical Bulletin 985, U.S. Department of Agriculture.
- R Development Core Team, 2010: *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing, URL <http://www.R-project.org>, ISBN 3-900051-07-0.
- Saunders, M. and A. Harris, 1997: Statistical evidence link exceptional 1995 Atlantic hurricane season to record sea warming. *Geophys. Res. Lett.*, **24**, 1255–1258.
- Scheitlin, K., J. Elsner, S. Lewers, J. Malmstadt, and T. Jagger, 2011: Risk assessment of hurricane winds for Eglin Air Force Base in northwestern Florida, USA. *Theoretical and Applied Climatology*, DOI: **10.1007/s00704-010-0386-4**.
- Simpson, R. and M. Lawrence, 1971: Atlantic hurricane frequencies along the U.S. coastline. Technical Memo to the National Weather Service SR-58, National Oceanic and Atmospheric Administration.
- Sklar, A., 1959: Fonctions de répartition à  $n$  dimension et leurs marges. *Publications de l’Institut de Statistique de l’Université de Paris*, **8**, 229–231.
- Smit, B. and J. Wandel, 2006: Adaptation, adaptive capacity and vulnerability. *Global Environmental Change*, **16**, 282–292.

- Smith, R., J. Tawn, and H. Yuen, 1990: Statistics of multivariate extremes. *International Statistical Review*, **58**, 47–58.
- Tawn, J., 1988: Bivariate extreme value theory: models and estimation. *Biometrika*, **75**, 397–415.
- Tobler, W., 1970: A computer movie simulating urban growth in the Detroit region. *Economic Geography*, **46**, 234–240.
- Tsutsui, J. and A. Kasahara, 1996: Simulated tropical cyclones using the National Center for Atmospheric Research community climate model. *Journal of Geophysical Research-Atmospheres*, **101**, 15 013–15 032.
- Venables, W. N. and B. D. Ripley, 2002: *Modern applied statistics with S*. New York: Springer.
- von Bortkiewicz, L., 1922: Variationsbreite und mittlerer fehler. *Sitzungsber. Berli. Math. Ges.*, **21**, 3–11.
- Walsh, K. and J. Katzfey, 2000: The impact of climate change on the poleward movement of tropical cyclone-like vortices in a regional climate model. *Journal of Climate*, **13**, 1116–1132.
- Walton Jr., T., 2000: Distributions for storm surge extremes. *Ocean Engineering*, **27**, 1279–1293.
- Wang, C., S. Lee, and D. Enfield, 2008: Atlantic Warm Pool acting as a link between Atlantic Multidecadal Oscillation and Atlantic tropical cyclone activity. *Geochemistry Geophysics Geosystems*, **9**, Q05V03, doi: 10.1029/2007GC001 809.
- Weibull, W., 1939: A statistical theory of the strength of materials. *Ing. Vet. Akad. Handlingar*, **151**.
- Yan, J., 2007: Enjoy the joy of copulas: with a package copula. *Journal of Statistical Software*, **21**, 1–21.

# BIOGRAPHICAL SKETCH

## Date and Place of Birth

27 January 1985

West Allis, Wisconsin

## Education

Doctorate of Philosophy in Geography: Florida State University, Tallahassee, FL  
May 2012

Master's of Science in Geography: Florida State University, Tallahassee, FL  
May 2009

Bachelor's of Arts in Geography: University Wisconsin-Oshkosh, Oshkosh, Wisconsin  
June 2007

## Relevant Experience

Research Assistant, Florida State University

Teaching Assistant, Florida State University

Online Mentor, Florida State University

Lab Assistant, University Wisconsin-Oshkosh

## Publications

1. Scheitlin, K. N., J. B. Elsner, S. W. Lewers, **J. C. Malmstadt**, and T. H. Jagger, 2011: Risk assessment of hurricane winds for Eglin Air Force Base in north-western Florida, USA, *Theoretical and Applied Climatology*, DOI: 10.1007/s00704-010-0386-4.
2. Elsner, J. B., S. W. Lewers, **J. C. Malmstadt**, and T. H. Jagger, 2011: Estimating contemporary and future wind-damage losses from hurricanes affecting Eglin Air Force Base, Florida, *Journal of Applied Meteorology and Climatology*, **50**, 1514–1526.
3. **Malmstadt, J. C.**, J. B. Elsner, and T. H. Jagger, 2010: Risk of strong hurricane winds to Florida cities, *Journal of Applied Meteorology and Climatology*, **49**, 2121-2132.
4. Scheitlin, K. N., J. B. Elsner, **J. C. Malmstadt**, R. E. Hodges, and T. H. Jagger, 2010: Toward increased utilization of historical hurricane chronologies, *Journal of Geophysical Research*, **115**, DOI:10.1029/2009JD012424.

5. **Malmstadt, J. C.**, J. B. Elsner, and T. H. Jagger, 2010: Frequency and intensity of hurricanes within Florida's threat zone. *Hurricanes and Climate Change, Volume 2*.
6. Elsner, J. B., R. E. Hodges, **J. C. Malmstadt**, and K. N. Scheitlin (eds), 2010: *Hurricanes and Climate Change, Volume 2*.
7. **Malmstadt, J. C.** , K. N. Scheitlin, J. B. Elsner, 2009: Florida hurricanes and damage costs. *Southeastern Geographer*, **49**, 108–131.