2015

Analysis and Prediction of Integrated Kinetic Energy in Atlantic Tropical Cyclones

Michael E. Kozar

FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

ANALYSIS AND PREDICTION OF INTEGRATED KINETIC ENERGY IN ATLANTIC TROPICAL CYCLONES

By

MICHAEL E. KOZAR

A Dissertation submitted to the
Department of Earth, Ocean and Atmospheric Science
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Spring Semester, 2015

Michael E. Kozar defended this dissertation on March 19, 2015.
The members of the supervisory committee were:

Vasubandhu Misra
Professor Directing Dissertation

Ming Ye
University Representative

Robert E. Hart
Committee Member

Philip Sura
Committee Member

Allan J. Clarke
Committee Member

Mark D. Powell
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 university requirements.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my family for their loving support. Their unwavering encouragement has enabled me to chase my dreams of becoming a meteorologist. I am also grateful for the friendships that I have made here at Florida State and in Tallahassee, back at Penn State and at home in New Jersey, and everywhere else in between. All of the fun that we've had through the years has made the difficult task of earning a doctorate degree extremely enjoyable.

In addition, I would like to thank my major professor, Dr. Vasu Misra, for his support and guidance throughout my entire time at Florida State. I am also thankful for the helpful advice given to me by my committee members Dr. Robert Hart, Dr. Philip Sura, Dr. Allan Clarke, Dr. Ming Ye, and Dr. Mark Powell. Furthermore, I owe thanks to Dr. Mark Bourassa for his help during my dissertation defense as well as several of my friends and colleagues at COAPS who gave me significant feedback on various presentations and research projects throughout my studies. Finally, I am indebted to all of my past teachers, advisors, professors, committee members, and colleagues who have helped make me a better meteorologist throughout my young career.

This work was supported by grants from the National Oceanic and Atmospheric Administration (NA12OAR4310078, NA10OAR4310215, NA11OAR4310110) and the United States Geological Survey (G13AC00408).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AM	Angular momentum
AAM	Absolute angular momentum
ANN	Artificial neural network
BIKE	Benchmark of Integrated Kinetic Energy
CFSR	Climate Forecast System Reanalysis
EAM	Earth angular momentum
ESRL/PSD	Earth System Research Laboratory / Physical Sciences Division
ET	Extratropical transition
GEFS	Global Ensemble Forecast System
GFS	Global Forecast System
IKE	Integrated kinetic energy
kts	Knots
LGEM	Logistic Growth Equation Model
MSE	Mean squared error
NCEP	National Centers for Environmental Prediction
NHC	National Hurricane Center
NOAA	National Oceanic and Atmospheric Administration
nm	Nautical miles
PoE	Probability of exceedance
PV	Potential vorticity
RH	Relative humidity
RMW	Radius of maximum sustained winds
RV	Rankine vortex
r_{34}	Radius of 34-knot winds
r_{50}	Radius of 50-knot winds
r_{64}	Radius of 64-knot winds
RAM	Relative angular momentum
RI	Rapid intensification
SHIFOR	Statistical Hurricane Intensity Forecast
SHIPS	Statistical Hurricane Intensity Prediction Scheme
SPIKE	Statistical Prediction of Integrated Kinetic Energy V1 (linear regression model)
SPIKE2	Statistical Prediction of Integrated Kinetic Energy V2 (artificial neural network)
SSHWS	Saffir Simpson Hurricane Wind Scale
SST	Sea surface temperature
TC	Tropical cyclone
TNA	Tropical North Atlantic
TJ	TeraJoules
VMAX	Maximum sustained winds

The abbreviations for the individual model predictors are shown separately in Table 2.1.

ABSTRACT

Integrated kinetic energy (IKE) is a recently developed metric that approximates the destructive potential of a tropical cyclone by assessing the size and strength of its wind field. Despite the potential usefulness of the IKE metric, there are few, if any, operational tools that are specifically designed to forecast IKE in real-time. Therefore, IKE and tropical cyclone structure are analyzed within historical Atlantic tropical cyclones from the past two decades in order to develop an understanding of the environmental and internal storm-driven processes that govern IKE variability. This analysis concurs with past research that IKE growth and decay is influenced by both traditional tropical cyclone development mechanisms and by other features such as extratropical transition and trough interactions.

Using this framework, a series of statistical prediction tools are created in an effort to project IKE in Atlantic tropical cyclones from a series of relevant normalized input parameters. The resulting IKE prediction schemes are titled the “Statistical Prediction of Integrated Kinetic Energy (SPIKE)”. The first version of SPIKE utilizes simple linear regression to project historical IKE quantities in a perfect prognostic mode for all storms between 1990 and 2011. This primitive model acts as a proof of concept, revealing that IKE can be skillfully forecasted relative to persistence out to 72 hours by even the simplest of statistical models if given accurate estimates of various metrics measured throughout the storm and its environment.

The proof-of-concept version of SPIKE is improved upon in its second version, SPIKE2, by incorporating a more sophisticated system of adaptive statistical models. A system of artificial neural networks replaces the linear regression model to better capture the nonlinear relationships in the TC-environment system. In a perfect prognostic approach with analyzed input parameters, the neural networks outperform the linear models in nearly every measurable way. The system of neural networks is also more versatile, as it is capable of producing both

deterministic and probabilistic tools. Overall, the results from these perfect prognostic exercises suggest that SPIKE2 has a high potential skill level relative to persistence and several other benchmarks.

Finally, in an effort to assess its real-time performance, the SPIKE2 forecasting system is run in a mock-operational hindcast mode for the 1990 to 2011 North Atlantic hurricane seasons. Hindcasts of IKE are produced in this manner by running the neural networks with hindcasted input parameters from NOAA's second generation Global Ensemble Forecast System reforecast dataset. Ultimately, the results of the hindcast exercises indicate that the neural network system is capable of skillfully forecasting IKE in an operational setting at a level significantly higher than climatology and persistence. Ultimately, forecasts of IKE from these neural networks could potentially be an asset for operational meteorologists that would complement existing forecast tools in an effort to better assess the damage potential of landfalling tropical cyclones, particularly with regards to storm surge damage.

CHAPTER 1

INTRODUCTION

Over the past few decades, predicting track and intensity fluctuations of North Atlantic tropical cyclones (TCs) has been a major emphasis for operational forecasters and atmospheric researchers alike. Both track and intensity metrics are definitively connected to the risks of a tropical cyclone, as maximum wind speed (VMAX) is related to the maximum potential destructiveness of a storm's wind field (e.g. Emanuel 2005; Bell et al. 2000), and storm position is used to identify the regions at risk of experiencing the storm.

Despite the inherent usefulness of track and intensity metrics, they do not yield information on the size or overall strength of TCs. Over the past few Atlantic hurricane seasons, several notable landfalling TCs have produced more damage than otherwise would have been expected by a storm of its intensity. Hurricanes Ike (2008) and Irene (2011) both produced in excess of $15 billion in damage across the United States as noted by the National Hurricane Center's (NHC) tropical cyclone reports (Berg 2009; Avila and Cangialosi 2011) despite being rated as category two and category one storms respectively at time of landfall on the Saffir Simpson Hurricane Wind Scale ("SSHWS"; Saffir 1975; Simpson 1974). More recently, Sandy (2012) only had maximum sustained winds of 70 knots when it made landfall as an extraordinarily large post-tropical storm, yet it caused in excess of $50 billion in total damage (Blake et al. 2013). As a result of these recent damaging storms and those from the record-setting 2005 Atlantic Hurricane Season (i.e. Hurricane Katrina), some researchers have questioned whether or not intensity metrics, such as maximum sustained winds, are best suited to communicate the risks of TCs (Kantha 2006).

Large hurricanes such as Ike, Irene, and Sandy have proven that VMAX is not the sole meteorological factor connected to the damage potential of landfalling TCs. For instance, Irish et

al. (2008) found that storm size is significantly correlated to storm surge in Atlantic Hurricanes and recommended that, although intensity scales adequately categorize wind damage, storm size must be considered to adequately categorize damage from flooding. Of course, storm surge damage is also significantly influenced by non-meteorological variables such as coastal bathymetry, population density, and property value. Nonetheless, with all else being equal, storm size is thought to have a significant influence over storm surge risks in landfalling TCs. Overall, arguments such as these underscore the importance of assessing TC structure in real time and also help to validate the usefulness of metrics related to TC size such as operational wind radii.

Consequently, new scales and metrics have emerged as a complement to the SSHWS that are more closely tied to the overall kinetic energy of tropical storms (Powell and Reinhold 2007; Maclay et al. 2008). Integrated kinetic energy (IKE) is one such metric that is calculated by accumulating one-half of the 10-meter wind field squared (U^2) times the density (ρ) per unit volume over a one-meter depth volume domain of a tropical cyclone (Powell and Reinhold 2007), or more specifically,

$$(1.1) \quad IKE = \int_v \frac{1}{2}\rho U^2 dV \, .$$

In most cases, IKE is integrated using Equation 1.1 over a portion of the storm volume (V) that contains wind speeds greater than a certain threshold, such as tropical storm force (34kts or 18 m $s^{-}1$) which will be used in the majority of this dissertation to be consistent with past work (Powell and Reinhold 2007; Misra et al. 2013).

Importantly, unlike maximum sustained wind metrics, IKE responds to changes in the overall size, strength, and intensity of a TC. With respect to the physical processes that govern damage potential, IKE is designed to scale with the stress of the wind on the ocean as well as the wind load that acts upon a structure (Powell and Reinhold 2007). Therefore, IKE and other similar

kinetic energy metrics are hypothesized to correspond well to the destructive potential of TCs, particularly with regards to storm surge damage.

For example, Hurricanes Ike, Irene, and Sandy at their respective times of landfall in the United States had extremely high IKE values, despite their low intensities (Figure 1.1). Just prior to landfall, each of these three storms had IKE values that would rank in the top 7.5% of IKE values for all North Atlantic TC fixes between 1990 and 2011. Sandy, in particular, reached a lifetime maximum of over 400 TeraJoules (TJ) of IKE just prior to landfall, giving it the second highest maximum IKE value in any Atlantic TC since 1990. These recent storms and the recent work of Powell and Reinhold (2007), which connected IKE to damage potential, suggest that forecasting IKE, as a complement to existing intensity metrics, could help to better assess the risks of landfalling North Atlantic TCs, particularly for larger and less intense storms such as these three recent examples.

Despite the potential usefulness of real-time IKE updates, there are limited resources currently available to forecasters that are specifically designed to assess the kinetic energy of a tropical cyclone. Therefore, this dissertation will present multiple statistical forecasting models that are specifically designed to assess IKE values within TCs over the Tropical North Atlantic (TNA) basin. The foundation for those statistical models is discussed in Chapter 2 by analyzing the processes that affect storm structure and IKE, with a particular emphasis on relating angular momentum to storm structure. Using the framework from that discussion, a linear regression IKE prediction model is constructed, and its results are discussed in Chapter 3. This initial statistical model serves as a proof of concept by using data from observations and other historical records to determine the plausibility of predicting IKE out to seventy-two hours (Kozar and Misra 2014). In Chapter 4, a more complex artificial neural network system is adapted to model IKE fluctuations in Atlantic TCs in a perfect-prognostic mode. In Chapter 5, the artificial neural network is run through a series of sensitivity tests to further analyze the physical relationships that drive the neural network. The neural networks are adapted for real time use in

Chapter 6 by running the network with hindcast data from the National Oceanic and Atmospheric Administration (NOAA)'s Global Ensemble Forecast System (GEFS) Reforecast Project Version 2. Finally, Chapter 7 contains a discussion on the applications of real-time IKE forecasts, including a simple Rankine Vortex model that can estimate operational wind radii from IKE in real time.

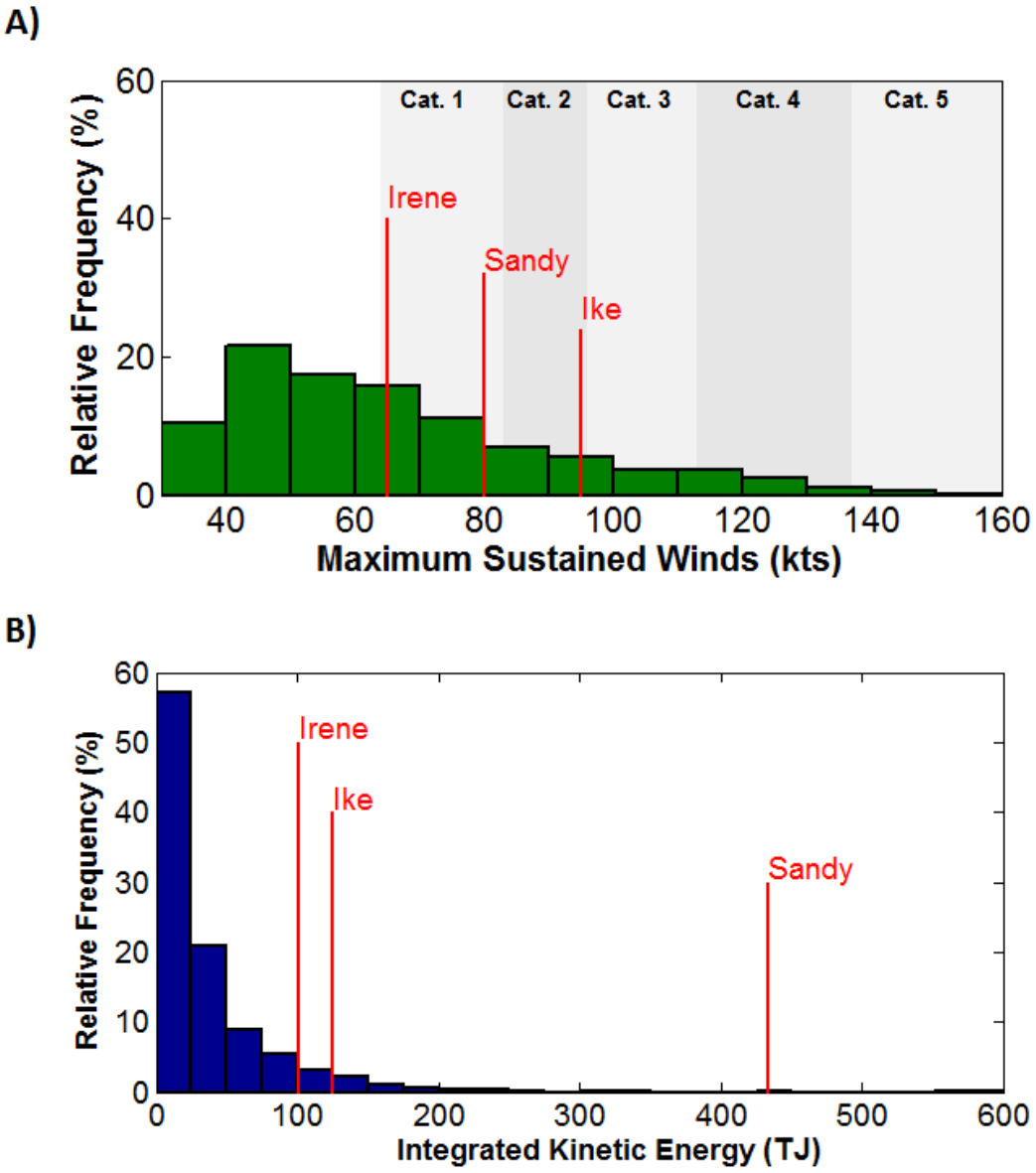

Figure 1.1: Relative frequency histogram of six-hourly VMAX (panel A) and IKE (panel B) measurements in Atlantic TCs between 1990 and 2011. This sample includes 5498 fixes from 291 storms. Vertical lines are shown to indicate VMAX and IKE values for selected hurricanes just prior to a US landfall. The times of these IKE measurements are as follows: Ike 9/13/08 00Z; Irene 8/28/11 06Z; Sandy 10/29/12 18Z. The three storm points would fall in the top 45% of all TCs points in terms of VMAX and the top 7.5% of TCs in terms of IKE from 1990 through 2011.

CHAPTER 2

REVIEW OF TC STRUCTURE AND ANGULAR MOMENTUM

The purpose of this chapter is to analyze some of the basic mechanisms that affect the structure of tropical cyclones, with a particular emphasis on discussing angular momentum. In past works, storm structure has been broken down into three components: *Intensity*, *size*, and *strength (*Merrill 1984). As expected, intensity measures the maximum wind speed and/or minimum pressure extremes found within the inner core of a TC. Storm size relates to the outer areal extent of a TC, and is often defined by the radius of winds above a certain threshold or the radius of the outermost closed isobar. Finally, Merrill (1984) defines storm strength as a conceptual average wind speed measured across the whole cyclone, whereas strength and intensity can potentially change independent of one another as wind speeds in the outer core regions evolve.

Observations indicate that North Atlantic TCs inherently come in many flavors. Some TCs are compact and intense (e.g. Hurricane Charlie, 2004). Other TCs are less intense, but have great storm strength and storm size (e.g. Hurricane Sandy, 2011). Occasionally, some particularly destructive TCs obtain high intensities, large sizes, and great strength all at once (e.g. Hurricane Katrina, 2005). Ultimately, all of these characteristics influence the total amount of IKE within a TC's wind fields. As such, it is critically important to understand the basic physical principles that govern structure before attempting to predict quantities of IKE in North Atlantic TCs.

The first section focuses on the conservation of angular momentum (AM) and how it relates to atmospheric circulations such as those found in a TC. Section 2.2 builds upon the AM discussion by explaining how storm size and intensity interact together with IKE. Finally, the third section in this chapter discusses the role of trough interactions and other features that can

result in a net gain of angular momentum for TCs, ultimately causing them to grow in size, strength, and/or intensity.

2.1 Conservation of Angular Momentum

Angular momentum is a quantity that measures the rotation of a body around a center axis. In a tropical cyclone, the absolute angular momentum (AAM) of the wind circulation is often broken down into to two components: relative angular momentum (RAM) and Earth angular momentum (EAM). Relative angular momentum represents the rotation of the axisymmetric circulation around the center of the storm and EAM represents the contribution from the rotating earth beneath the storm. More specifically these relative momentum quantities are defined per unit mass as

$$(2.1) \quad RAM = v_\theta r \ ,$$

$$(2.2) \quad EAM = \frac{1}{2} f r^2 \ ,$$

$$(2.3) \quad AAM = RAM + EAM = v_\theta r + \frac{1}{2} f r^2 \ ,$$

wherein v_θ is the tangential wind component, r is the radius relative to the storm center, and f is the Coriolis parameter.

If a system is closed and isolated, such that there is no net torque exerted on the system, then AAM will remain constant based on the law of conservation of angular momentum. As per the glossary of the American Meteorological Society (2012), conservation of angular momentum dictates that AAM cannot be destroyed, only transferred from one system to another. Considering that TCs contain a significant amount of cyclonic AM, the conservation of AM is somewhat applicable to these storms. For instance, conservation of AM can be used in part to explain why the fastest surface winds in a hurricane occur near the center of the circulation, after slower winds are imported from much larger radii.

However, if AAM is entirely conserved in a TC environment, a parcel with no tangential velocity originating at a radius of 500 km would contain tangential winds of more than 400 kts if it were brought into the eyewall at a radius of 30 km in a TC located at 20°N latitude. Obviously, wind speeds do not reach those extreme values in even the most intense TCs, and as a result, it can be determined that AAM is not fully conserved within a single parcel. In fact, no more than 25% of AM is conserved as parcels move inward towards the eye. Studies have long shown that the assumption of AM conservation fails because friction and turbulence causes RAM to dissipate as lower-tropospheric air converges towards the center of a TC (e.g. Pfeffer 1958; Holland 1983; Merrill 1984). Ultimately, for a TC to sustain itself, it must overcome AM frictional dissipation by importing additional AM into the system. In this regard, larger systems with more RAM and larger areas of frictional dissipation will require more inward AM transports than will TCs with smaller wind fields (Merrill 1984; Chan and Chan 2014).

Ultimately, the import of AM into the inner core of a TC is often driven by the secondary circulation (e.g. Holland 1983; Merrill 1984). As shown in Figure 2.1, a typical TC will have large areas of cyclonic RAM in its lower levels, and areas of anticyclonic RAM in the outflow layer aloft. Interestingly, some of the largest RAM magnitudes are found away from the center where wind speed is lower but the radius of the circulation is higher. Considering that EAM also increases with increasing radius, it is the responsibility of the secondary circulation's lower inflow branch to draw in higher cyclonic AM into the center from larger radii. Furthermore, the upper-level outflow branch of the secondary circulation is responsible for ventilating lower AM out of the center. Together, the transport of AM by the lower and upper branches of the secondary circulation, in addition to AM transports by various eddy fluxes, cause a net inward advection of AM that opposes the frictional dissipation of AM and allows for a TC to sustain itself (e.g. Holland 1983).

When the net import of AM into the TC outweighs the momentum loss from surface stresses, a TC will grow, strengthen, and/or intensify. For instance, storm growth is often tied to

increased cyclonic AM transport in the lower level of the secondary circulation, and intensification is often related to the transport of low AM out of the core region by the upper branch of the secondary circulation (Holland and Merrill 1984; Chan and Chan 2013). Building upon this result, Holland and Merrill (1984) discovered that the amount of AM transport needed to trigger intensification is much less than the amount of AM transport needed to promote storm size growth. Therefore, internal processes such as redistributing or enhancing convection in a TC can cause intensification, but these processes are typically insufficient for triggering and sustaining storm growth. Instead, low level environmental forcing is often critically important to accumulating of AM and achieving storm growth (Holland and Merrill 1984).

Since RAM in a storm is directly related to the wind speeds found in a TC, features that promote positive cyclonic AM transport into the core regions of the storm would be favorable for increasing IKE. Therefore, one would expect to find a positive relationship between IKE and each of upper-level divergence, low-level convergence, and low-level vorticity. Ultimately, the proposed relationships that arose through this discussion of angular momentum may be useful when constructing the framework for IKE prediction in the next several chapters.

2.2 Relationships Between Intensity, Size, and Kinetic Energy

By definition (Equation 1.1), integrated kinetic energy will increase to some degree, all else being equal, if any one of the storm intensity, size, or strength increases. Simply put, increasing intensity results in higher wind speeds within the core region, increasing size results in larger areas of tropical storm force wind speeds on the outer edges of the storm, and increasing strength results in a general increase in wind speed throughout the storm. An increase in wind speeds for any of these regions would cause IKE to similarly increase to a certain degree. Intensity changes, all else being equal, cause the least significant response to IKE because the fastest winds in a TC are often limited to a confined area within the inner core. Changes to

storm size and/or storm strength often cause a more substantial change to IKE, all else being equal, because they involve changing the wind speed over a larger portion of the TC wind field.

However, the relationships between IKE, intensity, strength, and size are not simple because changes in one aspect of a storm's structure can impact other features within a TC. For example, an increased intensity would result in an increased RAM within the inner core. If AM is conserved as discussed in the previous section, the outer core must lose RAM to offset the RAM gain by the inner core. This hypothetical intensification scenario would ultimately result in the TC contracting. The simultaneous act of intensification and contracting would have a mixed signal with IKE. In this case, increased inner core intensity would cause IKE to increase but decreased outer core size would cause IKE to decrease. As alluded to earlier, changes in storm size often have a greater impact on IKE tendency than do changes in intensity.

Based upon this hypothetical scenario, it is somewhat reasonable to hypothesize that IKE will decrease when intensity increases if RAM is conserved. If this were the case, a negative trend should be evident when plotting VMAX against IKE for all Atlantic TC fixes 1990 and 2011. Instead, Figure 2.2 quickly invalidates the hypothesis, by indicating that IKE and VMAX have a statistically significant positive relationship with one another for TCs that are located equatorward of 30°N. Ultimately, the positive trend raises the question, "How can intensity and IKE both increase if AM is conserved?"

One method for investigating this issue is to examine various sample TC vortices as they undergo intensification while simultaneously conserving AM. To accomplish this, a number of idealistic vortices are generated using the simple Rankine vortex equations (Depperman 1947). These equations produce idealistic axisymmetric TC wind fields, such that the wind speed at any radius is dependent solely upon VMAX, the radius of maximum wind (RMW), and a scaling parameter. More specifically, the equations are given by the following stepwise function, wherein the letter A denotes the scaling parameter:

$$v(r) = \left(\frac{VMAX}{RMW}\right) r \qquad r < RMW$$

(7.1)

$$v(r) = VMAX\left(\frac{RMW}{r}\right)^{A} \qquad r \geq RMW.$$

Ultimately, the inner core of the cyclone assumes a solid body rotation, wherein wind speeds increase linearly from the storms center until reaching RMW within the TC's eyewall. Outward of RMW, winds decrease exponentially, such that the strongest radial wind gradients are located closest to RMAX.

Using the RV equations, a sample modest hurricane wind field with a maximum intensity of 70kts, a RMW of 35 nautical miles (nm), and a scaling parameter of A=0.80 is generated. This hurricane, TC#1, has an IKE of 35 TJ, a tropical storm force wind radius (r_{34}) of 86 nm, and an average of $3 \cdot 10^6$ m^2/s RAM at the surface from a radius 0 to 500 nm (Figure 2.3). If TC#1 was located at 30°N latitude, the storm would have an average of $10 \cdot 10^6$ m^2/s of EAM at the surface from r=0-500nm, resulting in an average surface AAM of $14 \cdot 10^6$ m^2/s from r=0-500nm. If TC#1 was considered to be a closed system, its AAM value would need to remain constant, even if the storm were to intensify, in order to satisfy the conservation of angular momentum.

To simulate the rapid intensification of TC#1, another TC wind field is created using the RV equations. This second storm, TC#2, will remain at 30°N latitude with a RMW at 35nm, but its maximum intensity will increase to 115kts, making it a Category 4 hurricane on the SSHWS. To offset the increased inner core wind speeds and RAM gain, the scaling parameter in the RV equations must increase. An increase of the scaling parameter would cause winds speeds to decrease more quickly away from RMW, thus reducing the outer core RAM. The exact scaling parameter is chosen for TC#2 based on which value gives the second storm a correct value of $14 \cdot 10^6$ m^2/s average r=0-500nm AAM, providing that the new scaling parameter does not make the new wind profile inertially unstable (decreasing AAM with increasing radius).

Ultimately, the second TC vortex can be compared against TC#1 to examine exactly how wind speeds change across an entire idealized vortex during intensification, given the

conservation of AAM. As expected, Figure 2.3 indicates that wind speeds within the inner core of the TC increase drastically as a result of increasing VMAX by 45kts from TC#1 to TC#2. The largest increase unsurprisingly occurs at the RMW, but an increase of wind speed by 10kts or more can be found in a somewhat wide band from a radius of 8 nm out to a radius of 85 nm. On the other hand, the wind speeds in the outer radii do not decrease by the same magnitude. Since RAM is equal to the radius multiplied by the tangential velocity, a small change in velocity at a large radius of 500nm will offset the RAM gain caused by a much larger change in velocity at a smaller radius near the RMW.

Ultimately, the increase in inner core wind speeds pushes the radius of hurricane-force winds and the radius of 50-knot winds out further by more than 50% and 40% respectively. Interestingly, the taller peak in the radial wind speed distribution also pushes r_{34} out further by more than 25%. In fact, the wind speeds in TC#2 do not begin to decrease relative to TC#1 as a result of the intensification process until well outside of the tropical storm force wind field. Therefore, it is of no surprise that intensification in the RV leads to an increase of IKE by 49 TJ from TC#1 to TC#2. This increase in IKE is not unique to this intensification scenario. In fact, every other considered intensification scenario (more modest intensification, more extreme intensification, intensification with simultaneous eyewall contraction) all support that VMAX has a positive relationship with IKE, thus confirming the results found in the historical record.

Once again, it should be noted that the above results (increased intensity corresponds to a slightly weakened outer core, but IKE still increases overall during intensification) are valid given an assumption that TCs can be represented by a closed system with conserved total AM from r=0-500nm. Obviously, this assumption is not necessarily a realistic constraint. Therefore, instead, some may choose to apply the conservation of angular momentum to individual closed air parcels entering the circulation of the hurricane. This alternative approach is not any more realistic, as angular momentum is not truly conserved in individual air parcels either; otherwise slowly moving parcels taken from large radii would have wind speeds that could potentially

exceed the speed of sound when brought to the RMW. Nonetheless, for the sake of completion, it is worthwhile to investigate how the results would change if AM was conserved in individual parcels as opposed to the whole TC system.

Ultimately, in the alternative approach, closed air parcels would follow a constant AM surface as they moved towards the RMW. Given a value for the RMW and VMAX, there is one and only one profile of tangential wind that would be inertially neutral (dm/dr = 0). As a result, closed parcels moving towards the RMW in TC#1 with a VMAX of 70kts would follow a lower constant AM surface than closed parcels moving to the RMW in TC#2 with a VMAX of 115kts. Ultimately, this parcel-based conservation of AM approach would dictate that tangential winds would increase everywhere throughout the storm circulation during intensification, as AM increases uniformly. This differs from the result discussed previously when considering AM conservation across the TC system, wherein wind speeds must decrease over the outer core, in response to the spin up of winds in the inner core. Nonetheless, the two different approaches yield the same conclusion that IKE and VMAX should be positively correlated, agreeing with the observed historical data (Figure 2.2).

Based on these results, intensity parameters such as VMAX and minimum sea level pressure (MSLP) should be useful for projecting the tendency of IKE in Atlantic TCs. However, given the difficultly of operationally predicting storm intensity, it may be more useful to relate IKE with large scale environmental features tied to intensity change. For instance, one can hypothesize that IKE, like intensity, may have a positive relationship with SSTs and a negative relationship with vertical wind shear. All of these relationships are certainly worth investigating when constructing the IKE prediction models in the forthcoming chapters.

2.3 Import of Angular Momentum from Trough Interactions

The findings from the idealistic Rankine Vortex exercise suggest that intensity may be a critically important factor in determining integrated kinetic energy variability. However, in realistic atmospheric conditions, AM is not conserved and TC wind profiles are much more complex than the idealized axisymmetric Rankine Vortex. Therefore, the observed historical relationship between storm structure and intensity is likely to be much weaker than the strong relationship found in the Rankine Vortex exercises. As a result, other environmental and internal factors must also influence IKE variability

This is easily confirmed when looking at a plot of VMAX against IKE for TCs that are located over the northern half of the Atlantic Basin. Figure 2.4 includes a plot of VMAX versus IKE for all storms between 1990 and 2011 that are located north of 35°N in the Atlantic Basin. TCs in the northern part of the basin are typically less intense than TCs found over the warmer waters of the lower latitudes, yet the northern TCs seem to have more IKE on average. Obviously, if VMAX was the controlling factor of IKE, one would expect to find storms with the highest IKE values within the tropics, not the mid-latitudes. Furthermore, in just the higher latitude cases, VMAX does not seem to have any significant effect on IKE. The highest IKE values are collocated with moderate intensities near 65kts, not the higher intensities above 100kts (Figure 2.4). Ultimately, this suggests that intensity is not the major driving factor of IKE growth, especially over the higher latitudes of the TNA basin.

From an AM perspective, storms in the higher latitudes encounter a greater contribution of AM from the rotation of the planet. As such, higher latitude TCs contain higher EAM than lower-latitude TCs. If AAM was conserved, poleward moving TCs would then be required to lose RAM to offset the gain in EAM. However, the high IKE values in northern TCs suggest that they do not have less RAM than those located over the lower-latitudes. If RAM is not decreasing with latitude and EAM is increasing as storms move poleward, then the only possible conclusion is

that TCs in the upper latitude often do not conserve AAM. Instead, they must import substantial amounts of AM from outside environmental sources to grow and gain IKE.

In an effort to more closely examine the evolution of AM and IKE throughout the lifetime of Atlantic TCs, the majority of this section focuses on a case study of a historical TC, namely Hurricane Igor from September 2010. Igor was a classic Cape Verde type hurricane that formed in the eastern TNA. Nearly four days after becoming a tropical depression, Igor reached hurricane intensity, and rapidly intensified into an intense Category 4 hurricane on the SSHWS. Igor began turning to the northwest shortly thereafter, without threatening any of the Caribbean Islands directly. Igor's intensity fell over the central North Atlantic, but the storm grew substantially in size as it moved northward, passing just to the west of Bermuda. The track of Igor along with values of its intensity and IKE are shown in Figure 2.5.

One method for analyzing Igor's evolution is to utilize an intensity-kinetic energy (V-IKE) diagram. As discussed by Musgrave et al. (2012), these diagrams are useful for locating the three distinct phases of development in a TC: *incipient stage*, *deepening stage*, and *mature stage*. In an ideal scenario, the incipient stage occurs after TC genesis as the storm's convection begins to organize, resulting in a slow increase in VMAX and IKE. Following the incipient stage, the hypothetical TC undergoes intensification during its deepening stage. In this stage VMAX rapidly increases while IKE continues to increase at a slower rate. Following this intensification process, the TC will enter its mature phase. The mature phase is identified on a V-IKE diagram by a more pronounced increase of IKE accompanied by maintained and/or slowly decreasing intensities.

Overall, Igor tends to follow this idealistic evolution pattern very well. Initially, Igor struggled to develop, becoming a tropical storm on September 8th, before temporarily weakening back to a tropical depression. On September 11th, Tropical Storm Igor finally showed signs of steady development, reaching hurricane intensity the next day. To this point, Igor was clearly in its incipient stage as defined by Musgrave et al. (2012). Hurricane Igor then entered its

deepening stage on the 12th, rapidly intensifying from a minimal hurricane to a major hurricane in a span of 18 hours. Throughout this deepening stage, Igor's IKE slowly increased to nearly 50 TJ on the 13th. On September 15, Igor reached peak intensity, just short of Category 5 on the SSHWS, with an IKE of approximately 75 TJ, thus completing its deepening phase. Igor's mature phase began with an eyewall replacement cycle, during which VMAX temporarily decreased. By the 17th, Igor began to deintensify, yet its high values of IKE persisted. On the 20th, Igor was merely a Category 1 hurricane on the SSHWS, but its IKE suddenly increased quite rapidly before Igor finally became an extratropical cyclone on the 21st.

Figure 2.7 includes a plot of radially averaged 850mb AAM over the course of Igor's lifetime. These low-level AAM values were calculated from analyses of Igor's wind field within the second generation GEFS reforecast archive (Hamill et al. 2013). Throughout the first several days of Igor's lifetime, AAM was mostly conserved. There is some indication that a net inward flow of AAM exists, but any gain of AAM within the inner core is slow to exist. Therefore, based on the results of the RV experiments, it is not surprising that Igor's rapid intensification on September 12th was accompanied by slower IKE growth. Towards the end of Igor's mature stage, when the storm grew substantially in size, it becomes immediately apparent that the TC did not conserve AAM. Huge amounts of AAM were clearly imported into the storm from September 18th through the completion of Igor's extratropical transition on the 22nd that allowed the storm to increase in size and gain IKE, while still remaining a category 1 hurricane with a VMAX of 65 kts. For AM to be conserved throughout the entire system, the wind field expansion would have been accompanied by a dramatic reduction of inner core wind speeds, and an overall drop in IKE.

The mid-level pattern across the continental United States on September 20, 2010 is plotted in Figure 2.8. This plot, from the National Centers for Environmental Prediction (NCEP)'s daily weather map, coincides with Igor's rapid IKE gain at the end of its lifecycle. Igor's mid-level circulation is clearly visible over the northwestern Atlantic on the far eastern edge of the weather

map. Further upstream, an upper-level low is located over eastern Canada, and an area of higher heights is located over the Gulf Coast states. The upper-level trough associated the low over Canada extends reasonably far to its southeast, such that Igor appears to be embedded within this trough. The trough acted as a significant source of energy and momentum for Igor as it transitioned to an extra tropical cyclone (Fogarty 2010). Physically, AM is imported from upper-level troughs into the TC's circulation, which becomes tilted as it moves into a baroclinic environment, such that the cyclone's outer wind field accelerates at all levels. Ultimately, this process leads to an overall increase in size (Evans and Hart 2008) and subsequently a rapid gain of kinetic energy near the surface (Maclay et al. 2008).. As such, the relationship between trough interactions and increasing IKE is somewhat common in the northern Atlantic. Hanley et al. (2001) found that 78% of TCs superimposed within an upper-level trough and 61% of TCs that interact with troughs at a distance deepen because of the upper-level trough interaction.

In addition to the AM argument, Maclay et al. (2008) also argued that the increased wind shear associated upper-level troughs can be beneficial for storm growth in moderation, wherein the vertical shear acts to spread the tropical convection outside of the storm's inner core. Obviously, if the shear associated with the trough were too great, then the enhanced wind shear would be detrimental to storm growth, implying a Goldilocks principle between wind shear and IKE tendency over the mid-latitudes. Furthermore, another study by Kimball and Evans (2002) used a potential vorticity (PV) argument to relate trough interactions to TC intensification and growth. Essentially, the introduction of positive PV source from an upper-level low outside of the TC's RMW generates radially outward propagating vortex Rossby wave packets are generated. These wave packets transfer energy outward, ultimately causing the RMW to expand, the storm to grow, and therefore a simultaneous gain in IKE.

Aside from trough interactions, extratropical transition (ET) also tends to occur in the northern half of the basin. Historically, ET occurs in just under half of all Atlantic TCs (Hart and Evans 2001). Documented cases of ET have occurred as far south as 24°N, but the majority of

ET cases tends to occur between 35° and 45°N. During ET, TC wind fields often expand as the RMW moves outward and the outer wind field accelerates (Evans and Hart 2008). Thus, in many cases, TCs will gain IKE as ET begins to occur over the northern portion of the TNA basin.

Ultimately, processes such as ET and trough interactions ensure that the underlying relationships between IKE and the near-storm environment is quite different over the higher latitudes compared to the relationships found over the lower latitudes. As previously discussed, warm SSTs, low shear, high humidity, greater upper level divergence, and low-level convergence are traditionally thought as favorable for TC development over the tropics. However, in order to project changes in storm size and IKE over the upper-latitudes, it is plainly obvious that a prediction scheme must also consider the effects of trough interactions and ET to succeed. Therefore, when developing a statistical model to project IKE tendency, it would be wise to consider predictors related to trough interactions and ET. Simple positional variables such as latitude would be a good initial start, given the preference for mid-latitude cyclones and troughs to be located over the higher latitudes and the climatological preference for ET to occur in the upper latitudes. In addition, variables such as shear, SSTs, and/or eddy momentum fluxes could be useful to identify regions where troughs and/or baroclinic effects could lead to storm growth.

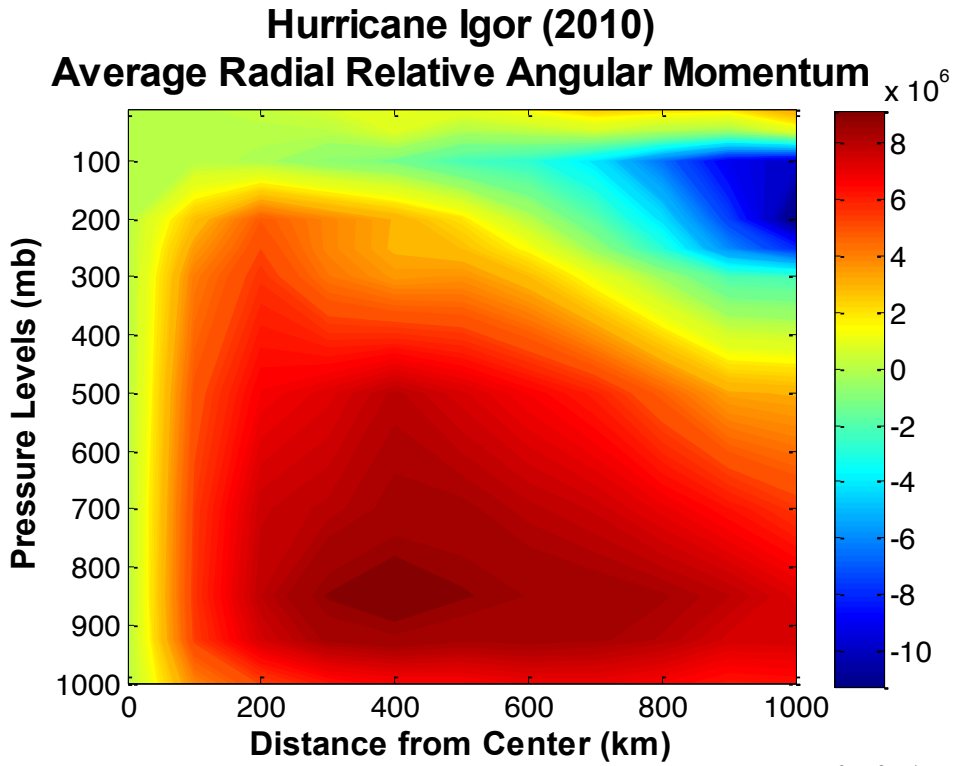

Figure 2.1: Cross Section of radially averaged relative angular momentum (in 10^6 m^2 s^{-1}) for Major Hurricane Igor on September 17, 2010 from the zero hour analysis in the 1 degree resolution control run in the GEFS reforecast database. Note that the lowest layer of RAM plotted within this figure is on a 1000mb surface. Since, surface pressure will increase with increasing radius in a TC, the AM fields at the outer radii are located at a higher altitude than the AM values within the inner core. In reality, if surface or near surface AM were plotted, there would be a much larger gradient present in the data, considering that winds must decrease to exactly surface at ground height.

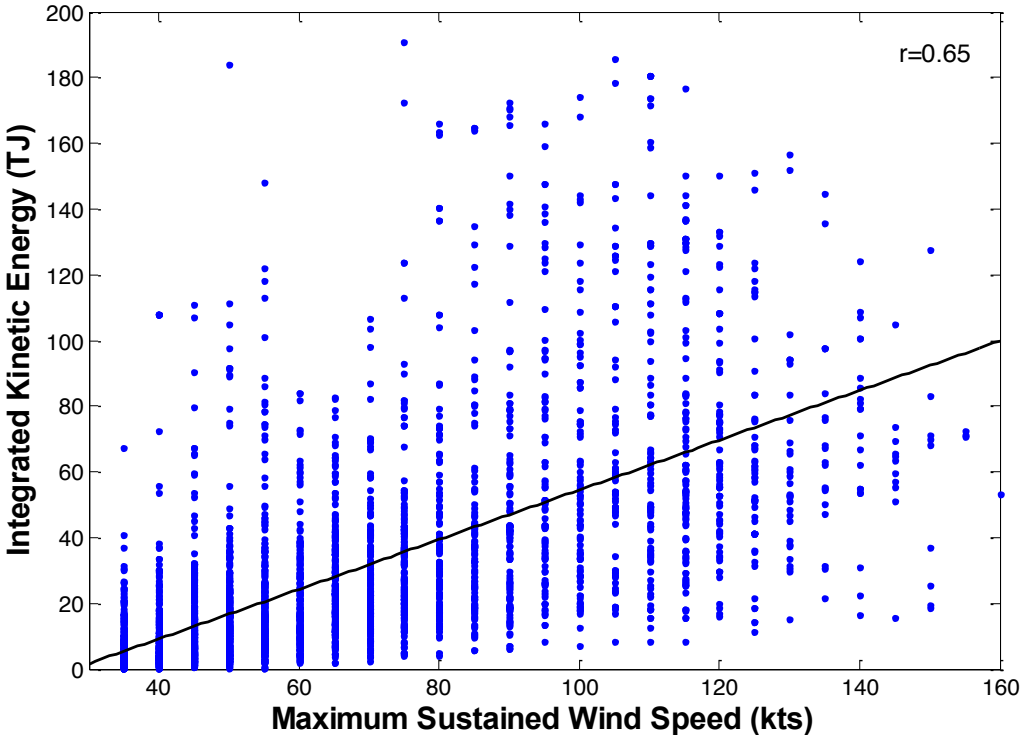

Figure 2.2: Plot of integrated kinetic energy (TJ) versus maximum intensity (kts) for 3896 Atlantic TC fixes located south of 30°N latitude between 1990 and 2011 (blue dots). The black line represents a linear regression fit for the data. Considering the large sample of storm fixes, the linear correlation is easily significant at the p=0.01 level.

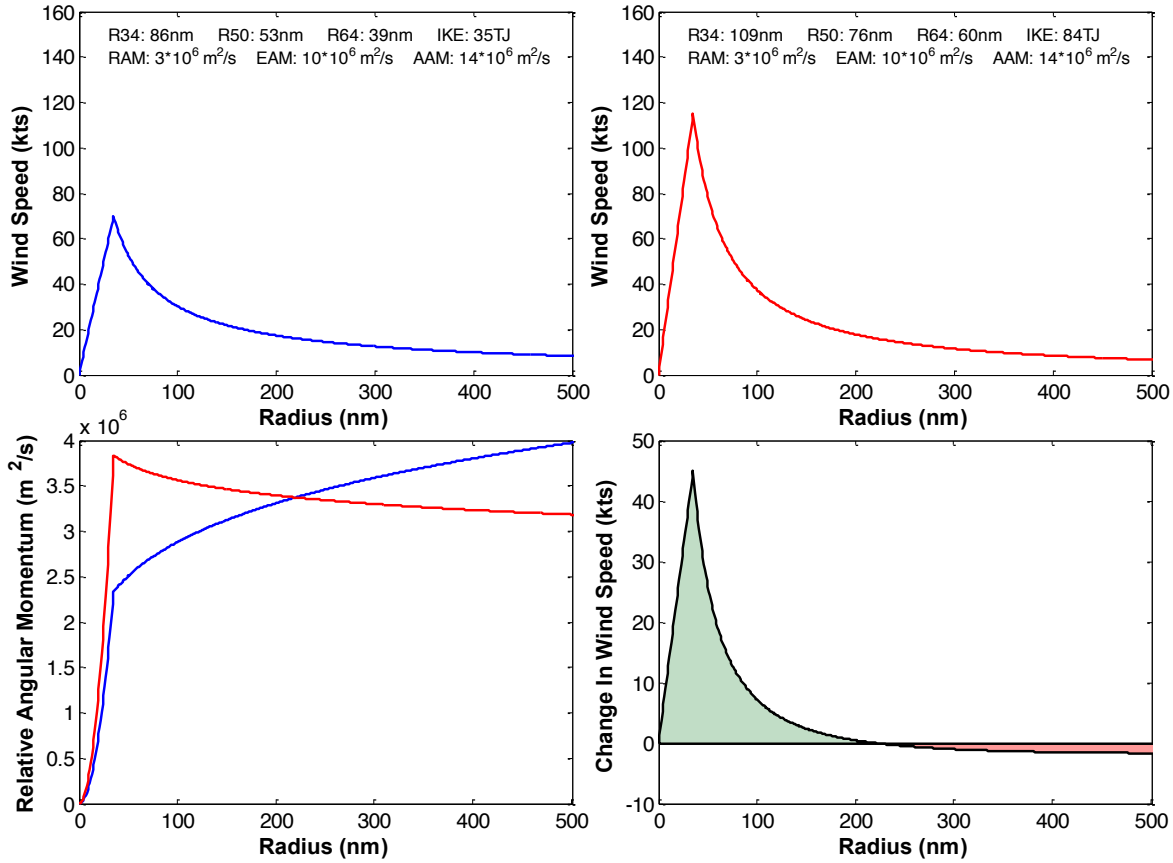

Figure 2.3: Various characteristics of two idealized Rankine Vortex TC circulations. The first TC vortex represents a developing storm that just achieved hurricane intensity, with maximum sustained winds of 70 kts. The second TC represents the same storm at a later time period with a higher maximum intensity, 115 knots. Importantly, the latter TC is restricted to contain the same amount of 0-500 km average absolute angular momentum as the first storm, such that the transition from TC #1 to TC #2 abides by the conservation of angular momentum. Additionally, both storms are located at precisely 30°N, have a radius of maximum winds set to 35 nm, and are stationary. The radial wind speeds of the first idealized TC (in kts) are shown in the top-left plot with a blue curve. The radial wind speeds of the second idealized TC (in kts) are shown in the top-right plot with a red curve. Operational wind radii, integrated kinetic energy values, and various average angular momentum values are written at the top of each of the top two plots for each respective storm. The radial distribution of relative angular momentum (in 10^6 m^2 s^{-1}) is shown in the bottom-left plot, with the blue curve representing TC #1 and the red curve representing TC #2. Although RAM decreases with increasing radius in TC#2, its radial distribution of AAM is inertially stable when considering the contribution from EAM. Finally, the bottom-right panel shows the change in wind speed from the first TC to the second TC (in kts). A green shaded area underneath this difference curve indicates that wind speed increased in TC#2 relative to TC#1 for a given radius, and a red shaded area above the curve indicates that wind speed decreased in TC#2 relative to TC#1 for a given radius.

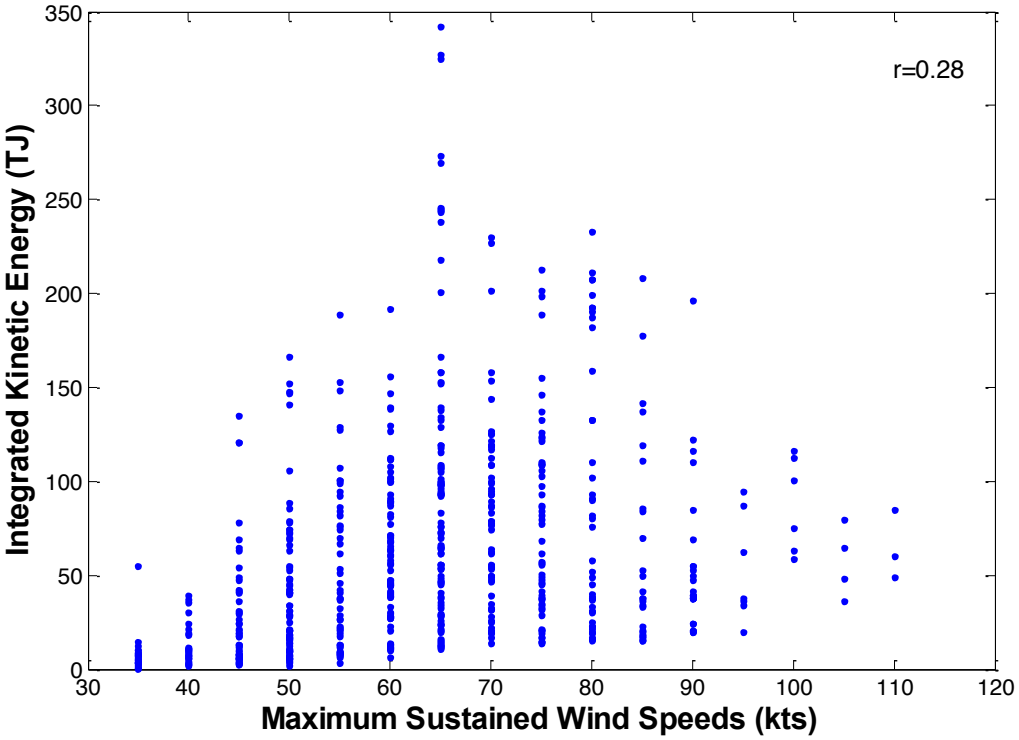

Figure 2.4: Plot of integrated kinetic energy (TJ) versus maximum intensity (kts) for 735 Atlantic TC fixes located north of 35°N latitude between 1990 and 2011. The correlation between VMAX and IKE in this figure is r=0.28.

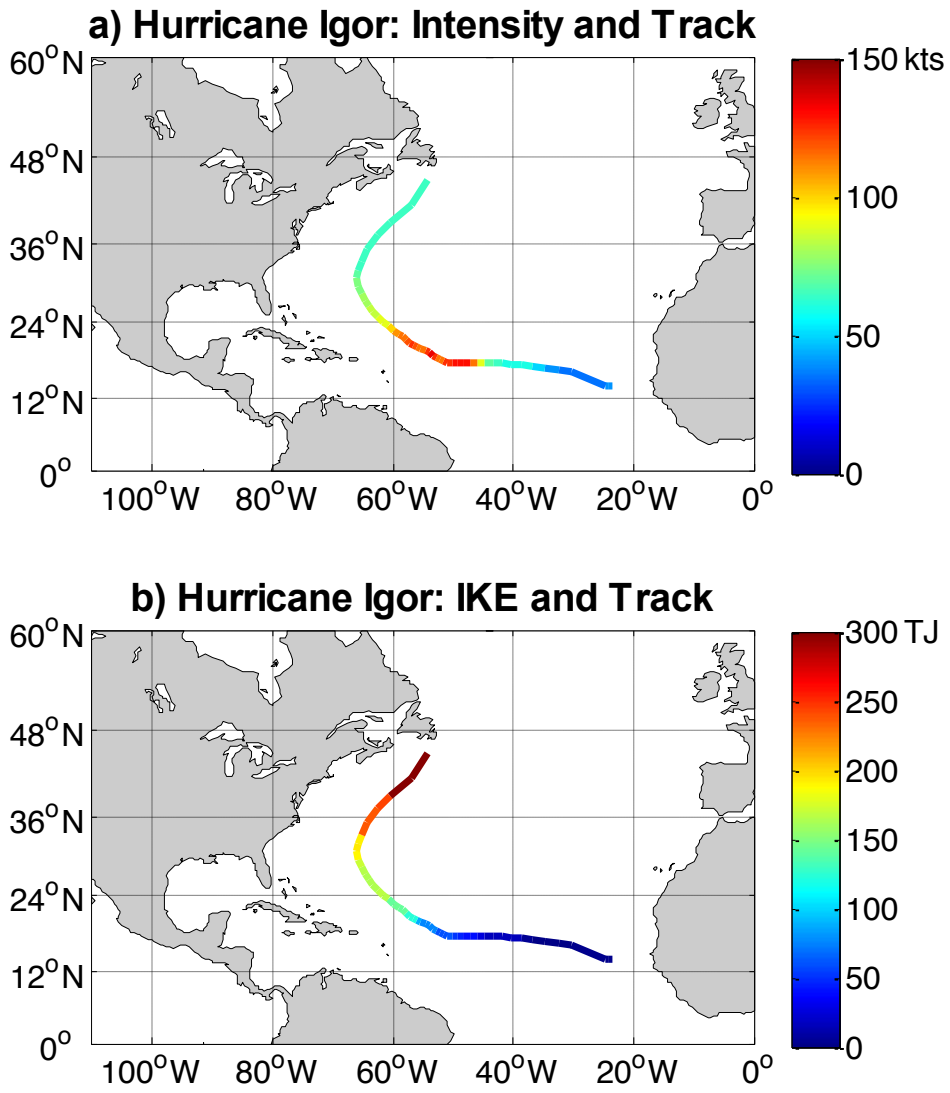

Figure 2.5: Plot of Hurricane Igor's track in September, 2010. In panel A, Igor's track is color coded based on maximum sustained winds. In panel B, Igor's track is color coded based on integrated kinetic energy.

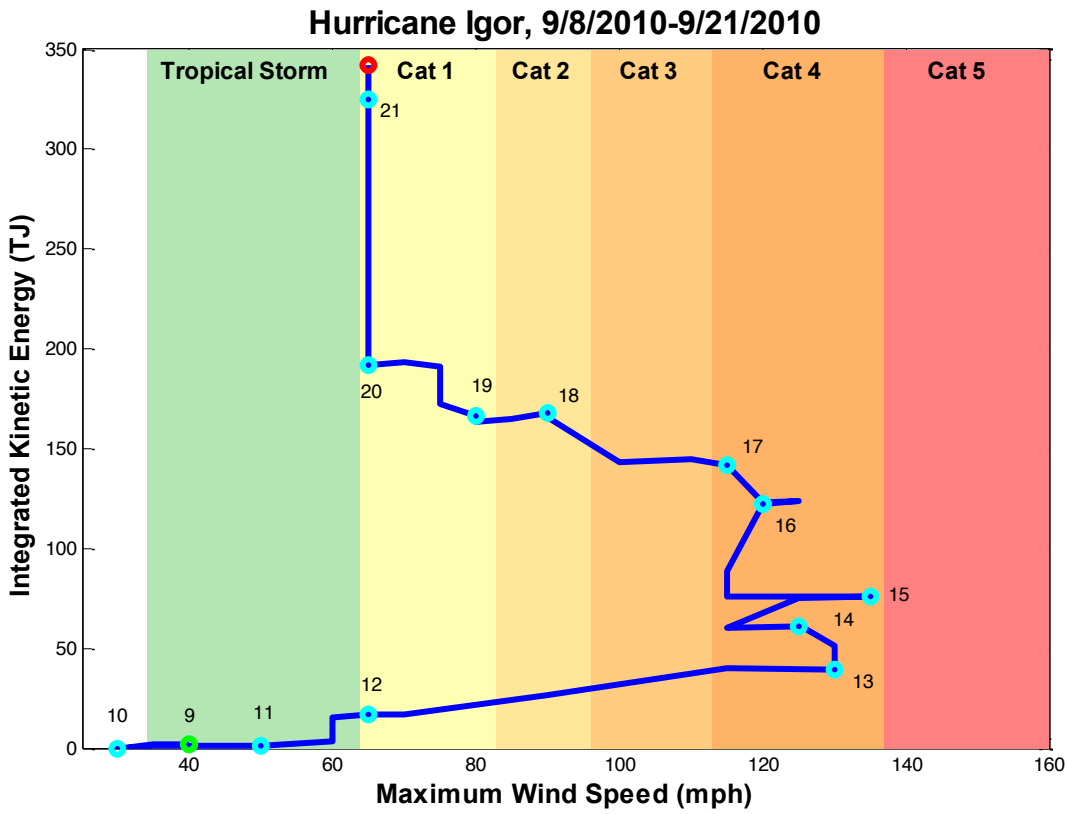

Figure 2.6: Evolution of integrated kinetic energy (TJ) versus maximum intensity (kts) throughout the lifetime of Hurricane Igor in September of 2010, as inspired by the work of Musgrave et al. (2012). Igor first obtained tropical storm intensity on September 8th at 18Z with a VMAX of 40 kts and an IKE of 1.8 TJ. This genesis point is denoted on the plot by the green circle in the bottom left corner. Each of the subsequent cyan circles signals a 00Z storm fix, indicating the passage of a day's time. The date for each 00Z fix is labeled next to each cyan circle. Igor became extratropical on September 21st at 18Z as denoted by the red circle at the top of the plot.

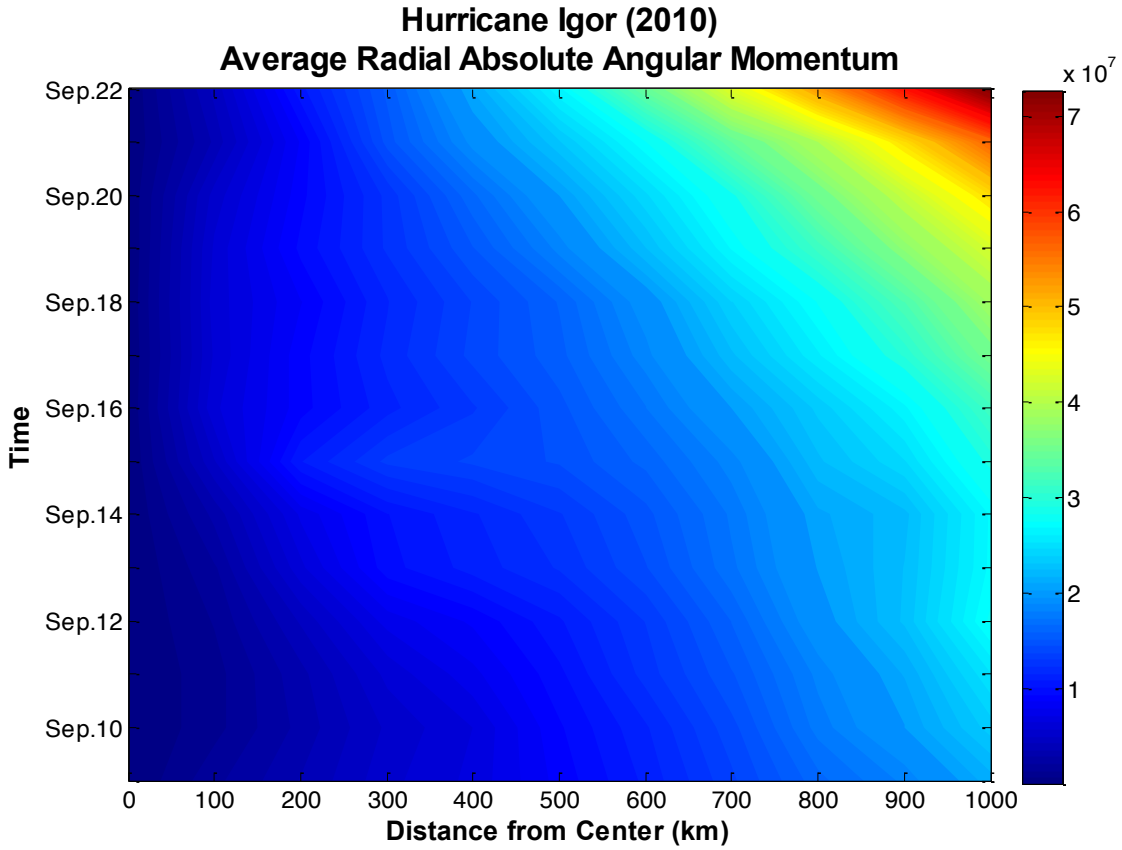

Figure 2.7: Radially averaged 850mb absolute angular momentum (in 10^7 m^2 s^-1) for Hurricane Igor over time from GEFS analyses in the GEFS reforecast database.

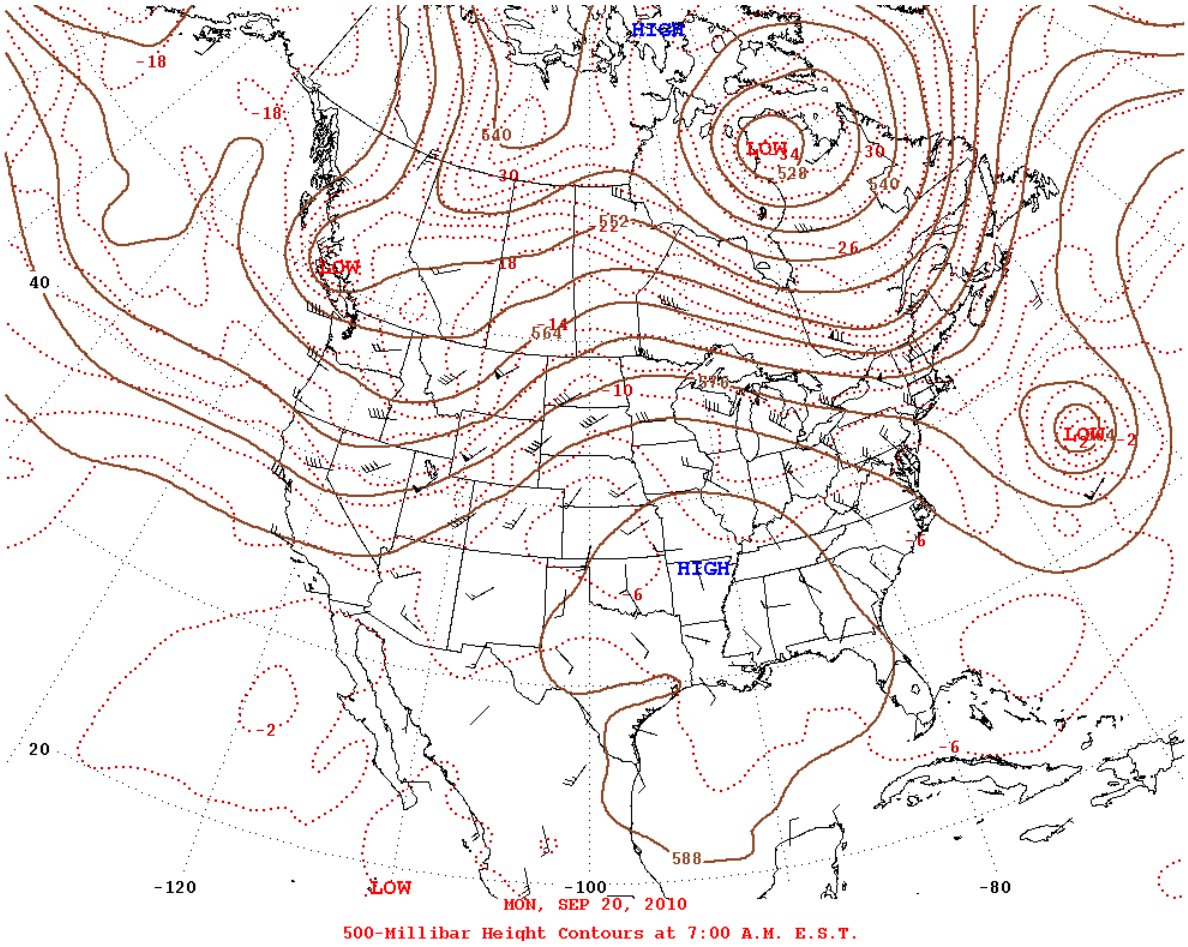

Figure 2.8: Daily weather map at 500 mb on September 20th, 2010 from the National Centers for Environmental Prediction, Hydrometeorological Prediction Center. Archived daily weather plots can be generated online at: http://www.hpc.ncep.noaa.gov/dailywxmap/

CHAPTER 3

PREDICTION OF IKE: A PROOF OF CONCEPT

This chapter discusses a simple multivariate linear regression model, named Statistical Prediction of Integrated Kinetic Energy (SPIKE), which can be used for forecasting IKE fluctuations in North Atlantic tropical cyclones. Much like the operational Statistical Hurricane Intensity Prediction Scheme (SHIPS) that is used operationally for forecasting VMAX (DeMaria and Kaplan 1994; 1999; DeMaria et al. 2005; Jones et al. 2006), SPIKE is trained on a blend of persistence, environmental, and internal storm-driven predictors to predict fluctuations of IKE out to three days in the future. The following sections in this chapter will discuss the data that goes into SPIKE as well as the construction and calibration of the regression model. Afterwards, the model's skill will be evaluated in a 22-year training interval from 1990-2011 and in a single validation season (2012) that notably falls outside of the training interval.

The exercises presented here are performed in a perfect prognostic or "perfect-prog" mode (e.g. Neumann 1987; DeMaria, 2010), wherein the independent predictors and target predictands are taken directly from observations and analyses. Obviously, in an operational setting, a statistical model will require imperfect forecasted predictors to forecast IKE in real time. Therefore, the discussion throughout this chapter is not intended to assess the skill of the SPIKE model in an operational environment. Instead, by performing these exercises in a perfect-prog mode, the maximum potential skill of the linear SPIKE models can be estimated for the ideal scenario during which the forecasted predictors perfectly match the eventual observations. Ultimately, the exercises in this chapter are designed as a proof of concept to determine whether or not it is even plausible to predict IKE in Atlantic TCs.

3.1 Historical Integrated Kinetic Energy Record

In order to calibrate a statistical regression model for IKE, it is first necessary to obtain a historical record of IKE that can be used as the calibration model's dependent variable. Gridded analyses of TC wind fields would be ideal for calculating IKE, but unfortunately datasets like H*Wind analyses (Powell et al. 1998) are discontinuous, as they are not available for every six-hourly storm fix. Alternatively, Powell and Reinhold (2007) developed an approximate set of stepwise equations to calculate estimates of IKE from operational 34-knot, 50-knot, and 64-knot wind radii, the radius of maximum wind (RMW), and VMAX. This methodology for approximating IKE from operational wind radii has been used previously to catalogue historical IKE levels (Misra et al. 2013) and will be used again here to form a large historical record of IKE values to train and test the SPIKE regression models. Specifically, the operational wind radii from the extended best track dataset (Demuth et al. 2006) are utilized to create a six-hourly record of IKE values for all North Atlantic TCs over a twenty-two year training interval between 1990 and 2011. The resulting IKE values in the historical dataset specifically measure the integrated kinetic energy only over the portion of the wind field where the wind speeds are of tropical storm force or greater ($U > 18$ m s^{-1}). This specific IKE quantity is selected to be consistent with past work (Misra et al. 2013) and more importantly because this IKE metric is hypothesized to be closely related to storm surge damage potential by Powell and Reinhold (2007).

In total, the historical IKE record comprises of 5498 six-hourly IKE values from 291 individual TCs included from 1990 to 2011 (one IKE value for each six-hourly TC fix). For storm fixes to be included in this IKE database, the storm must have maximum sustained winds of tropical storm force or greater, and the storm must be classified as a tropical or subtropical cyclone in the official best track database (i.e. extratropical and post-tropical lows are not considered). The vast majority of the fixes that meet these criteria, and are included in the historical IKE database, are located over the open ocean. However, there are a small minority of

points, for which the center of the TC is located over land. These so called “land points” occur mostly over islands or very near the coastlines of mainland North America, as storms typically do not retain tropical characteristics and tropical storm intensity long after landfall. In an effort to gauge SPIKE’s skill at predicting IKE for landfalling storms, these “land points” were not removed from the historical record.

The relative frequency distribution of this large sample of six-hourly historical IKE data points is shown in Figure 1.1. It is immediately evident that Atlantic TCs most often have IKE values that do not exceed 25 TJ. The mean IKE value in the dataset is 34.9 TJ, and the standard deviation is 43.0 TJ. In rare instances, TCs can briefly obtain IKE values greater than 300 TJ. These extreme IKE values typically occur towards the end of a TC’s lifecycle, prior to the completion of extratropical transition. If extratropical cyclones were included in the historical IKE database, it is likely that IKE would continue to increase immediately following extratropical transition over the northern latitudes of the Atlantic, but these storms are not as well documented as TCs and are outside of the main focus of this dissertation.

It should be noted that the historical wind radii dataset used to calculate historical IKE values is subject to year-to-year and storm-to-storm inconsistencies. Landsea and Franklin (2013) estimated that uncertainties for operational 30-, 50-, and 64-knot radii were quite large during the 2010 season. In fact, they estimate that on average the error bars for each radii estimate could exceed 30% of the mean size for each radii. Overall, they find that the accuracy of the historical wind radii is limited to the data available in the operational or post-storm analyses, wherein operational radii measured by aircraft reconnaissance have less uncertainty than do those measured by just satellite observations. Furthermore, Landsea and Franklin (2013) indicate that landfalling storms and more intense storms typically have less uncertainty, as these types of storms are generally observed more frequently and with better equipment, especially landfalling storms which can be observed from the ground. Therefore, the historical

IKE dataset is likely subject to the same inconsistencies of data quality and quantity that is evident within the wind radii data in the extended best track data.

3.2 Linear Regression Methodology

The distribution of historical six-hourly IKE values is decidedly non-Gaussian as shown in Figure 1.1. In fact, the six-hourly total IKE values are approximately log-normally distributed, which is similar to the distribution of storm size as measured by the radii of vanishing winds (Dean et al. 2009). Therefore, it would be inappropriate to use linear regression to model these total IKE values. Instead, SPIKE will seek to predict changes of IKE from initial time out to twelve evenly-spaced time intervals from six hours to seventy-two hours (0-6hrs, 0-12hrs, … , 0-72hrs). Ultimately, these IKE tendencies are more normally distributed (e.g. Figure 3.1), and thus, it becomes more appropriate to use multivariate linear regression. However, in order to create a model for each of these twelve forecast intervals, SPIKE regression models must be calibrated and validated separately for each forecast interval.

Based on the success of the statistical SHIPS model used to predict intensity change relative to other statistical and dynamical models (DeMaria et al. 2005), SPIKE utilizes similar regression methodology to project changes of IKE. As done by DeMaria and Kaplan (1994), both the dependent and independent variables are normalized prior to training the regression model. Ultimately, by normalizing the predictors, it becomes possible to make a comparison between regression coefficients for various predictors and forecast hours (DeMaria and Kaplan 1994).

To avoid over fitting in the SPIKE regression model, predictor screening must be utilized. Once again, SPIKE's methodology is based upon the methodology of DeMaria and Kaplan's SHIPS model (1994; 1999) which uses backwards screening to objectively select the most skillful predictors. Backward predictor screening is done here by training the model upon all of

the predictors for each forecast interval, and then repeatedly removing the single predictor with the least significant regression coefficient one at a time, until all of the remaining regression coefficients are significant at the p=0.01 level. Ultimately, this backward screening methodology retains a smaller subset of predictors that are used in the SPIKE model for each forecast interval. To make the SPIKE prediction model as uniform as possible across the forecast intervals, the same predictors are chosen for all intervals. These predictors are selected if their coefficients are significant at the 99% level for at least half of the forecast intervals. As a result, predictors may be used on intervals when their coefficients are not significant, but as pointed out by DeMaria and Kaplan (1994), when a predictor is not significant, its coefficient becomes small and its influence on the regression model is diminished.

It should also be noted that some of the SHIPS predictors included in Table 3.1 have missing or null data values scattered throughout the training interval of 1990-2011. Most notably, some of the subsurface ocean predictors (i.e. RD26) are unavailable for entire years at the beginning of the record. Therefore, to account for the limited number of missing predictor values, the regression coefficients are calibrated and the predictors are screened using a sample of all storm fixes that contain no missing data in the training interval. In other words, if a certain storm fix contains even one missing predictor, that data point will not be used in the calibration of the model. However, once the predictors are chosen and the coefficients are calculated, IKE is estimated for every storm fix to evaluate model skill. In the case that a predictor is missing, as occurs occasionally in an operational setting, the missing data points are filled with a value equivalent to the sample mean (i.e. exactly zero since the predictors are all normalized).

3.3 Potential Model Predictors

In addition to the historical record of IKE, a pool of potential predictors must also be created to establish the relationships that will govern the statistical model. These predictors are carefully selected based on some of the understood relationships between a storm's environment and the factors that govern the size, strength, intensity, and ultimately kinetic energy of a TC (e.g. Gray 1968; Merrill 1984; McBride 1995; Hill and Lackmann 2009; Maclay 2008; Musgrave et al. 2012).

These environmental, storm-specific, and persistence predictors are gathered from a combination of the NHC best track data set (Jarvinen et al. 1984) and the SHIPS developmental data set (DeMaria and Kaplan 1999). The variables derived from the NHC best track data are storm-specific predictors, such as date, position, duration, intensity, and translational storm motion. The environmental variables, which encompass thermodynamic, dynamic, and moisture related fields known to affect TC behavior, are taken from the aforementioned SHIPS developmental dataset (DeMaria and Kaplan 1999). In all, thirty-one predictors were considered for this regression exercise (Table 3.1).

Once again, it should be noted that the SHIPS developmental dataset, from which the environmental data is obtained, is not a forecast, and instead it utilizes analyses and reanalyses from NCEP to provide estimates for the observed environmental conditions experienced by each storm from genesis to dissipation. Obviously, an operational real-time version of SPIKE will require the predictors to be forecasted by numerical models, since the analyses and reanalyses are not available for future time steps in a real-time operational setting. Therefore, the SPIKE models discussed for the remainder of this chapter, all of which utilize observations and SHIPS developmental data, are not meant to be forecasts or even hindcasts. Instead these regression models will provide an estimate for the maximum potential skill of SPIKE in the idealistic scenario that the forecasted predictors exactly match the future observations.

3.4 Physical Interpretation of Selected Predictors

The remainder of Chapter 3 focuses on the results of the SPIKE regression model after the model is calibrated and the predictors are screened. However, this section first focuses on interpreting the physical relationships that drive the regression model because a statistical relationship is meaningless without an understanding of the underlying physical processes. A list of predictors that are retained through the backward screening exercise discussed in the previous section is shown in the first column of Table 3.2. These predictors encompass a wide array of variables ranging from thermodynamical fields such as the depth of the 26°C isotherm to positional variables such as latitude, dynamical values like upper-level divergence, and persistence variables such as past values of IKE. For the sake of simplicity, the variables are referenced as they are abbreviated in Table 3.1 for the rest of this discussion.

The coefficients for these variables at selected forecast intervals are also shown in Table 3.2. Encouragingly, the sign of most of the predictors' coefficients do not vary with forecast hour. For example, the coefficient for PIKE is negative in all intervals suggesting that storms with higher IKE are more likely to have decreasing IKE over time. As seen in the distribution of historical IKE values, TCs most often have low values of IKE; therefore it would not be terribly surprising for higher IKE storms to regress toward those lower IKE values. The physical reasoning behind this relationship is tied to the timing of maximum IKE during a TC lifecycle. As found by Musgrave et al. (2012), TCs often exhibit storm growth (increasing IKE) through most of their lifecycle. North Atlantic TCs, in particular, maintain or increase their size even after reaching maximum intensity (Knaff et al. 2014). As a result, TCs often have their highest levels of IKE late in their lifecycle, either prior to landfall or during extratropical transition when TCs often undergo wind field expansion (Evans and Hart 2008). Obviously, following landfall or the completion of ET, TCs typically weaken drastically over the hostile environments of land or the cold northern Atlantic Ocean. Therefore, the negative coefficient of PIKE can be attributed to the

negative IKE changes of these large storms, and the fact that weaker storms near genesis typically will gain IKE as they become more mature, provided the environment is not too unfavorable. Similar to PIKE, PDAY has negative coefficients for all forecast intervals. The negative coefficient is tied to the fact that TCs are more likely to have periods of increasing IKE close to the peak of the season (small PDAY), when conditions are typically most favorable for TC development.

The only predictor to have a coefficient that changes sign with forecast hour is dIKE12, wherein the coefficient is positive in the shorter forecast intervals and slightly negative in the longer intervals. The positive dIKE12 coefficients in the first several forecast intervals make sense, as storms typically continue to have increasing IKE in most phases of their lifecycle (Musgrave et al. 2012). Therefore, a growing storm will continue to grow provided the environment remains somewhat favorable. Likewise, if a storm is in an environment unfavorable for IKE growth, the kinetic energy will likely continue to drop, at least in the short term. The slightly negative sign for the longer forecast intervals is more difficult to reason physically, but it should be noted that the coefficient is not significant to begin with, suggesting that past 12hr IKE tendency is only helpful for determining upcoming IKE changes in the immediate future.

In terms of the environmental predictors, some of the underlying physical relationships are immediately apparent. For example, the coefficients for VORT and D200 are positive, suggesting a direct relationship between storm growth and each of these fields. In this case, both low-level vorticity and upper-level divergence are well known conditions that are generally favorable for large scale organized convection and the formation and development of TCs (e.g. McBride 1995). Furthermore, enhanced low-level vorticity, low-level convergence, and upper-level divergence would all lead to an increased influx of AM into the storm's core, which as discussed in the previous chapter, would lead to increased TC size, strength, and/or intensity. Similarly, the negative coefficients for MSLP and PENV are expected, as a more intense storm and/or a storm with a larger area of low pressure will typically have higher wind speeds and

increased IKE with all else being equal. Likewise, the positive coefficients for RD26 are unsurprising, as TCs induce turbulent mixing and upwelling in the upper levels of the ocean which cools sea surface temperatures (SSTs) through the entrainment of cooler subsurface waters (e.g. Price 1981). This SST cooling mechanism plays a significant role of slowing down TC growth and intensification, especially for slower moving storms over shallow oceanic mixed layers (Schade and Emanuel 1999). Thus, an environment with a deeper thermocline, and a higher RD26, is more resistant to the negative SST feedback mechanism, making storm growth more favorable. Finally, the positive coefficients for DTL also make sense, as TCs tend to weaken as they approach and eventually cross landmasses.

In some cases, however, the physical processes that govern the regression coefficients are less apparent. For example, the positive coefficients for SHRD and LAT seem somewhat counterintuitive to conventional TC development theories, wherein TCs favor low shear environments as well as warmer oceans which are typically found in the lower latitudes. However, as discussed in Chapter 2.3, TC growth is also connected to external forcing from trough interactions and baroclinic environments over the higher latitudes (e.g. Maclay 2008; Hanley et al. 2001; Kimball and Evans 2002). The positive coefficients for REFC, for example, reflect the positive influence of trough interactions on storm growth (Maclay et al. 2008; DeMaria et al. 1993). Likewise, the positive coefficients of LAT and SHRD are likely related to ET and trough interactions. As discussed previously, ET and trough interactions climatologically are more likely to occur in the northern part of the TNA basin. Furthermore, a moderate increase of shear can enhance storm growth during trough interactions (e.g. Maclay et al. 2008).

Finally, the negative coefficients for RHLO are particularly counterintuitive because in most cases increased low level humidity is favorable for TC development. This apparent contradiction is likely related to the effect of dry air intrusions on TC structure. First of all, storms undergoing ET are often accompanied by an intrusion of dry air into the storm circulation (Jones et al. 2003), the presence of which would in turn result in a decreased RHLO. Since ET tends to

also be associated with increasing IKE in the mid-latitudes, this could be the cause of the negative RHLO coefficient. Dry air intrusions also tend to occur as TCs enter a landfall environment (Kimball 2006). These dry air intrusions are observed to disrupt the inner core convection of landfalling TCs, often eroding the storm's inner eyewall and triggering an eyewall replacement cycle. Several recent storms in the Gulf of Mexico (e.g. Katrina, Ike, Gustav, etc) have grown substantially because of inner eyewall erosion and the subsequent RMW expansions (Knabb et al. 2005; Berg 2009). This observed process aligns with the work of Musgrave et al. (2012), which found that IKE increases despite the drop in intensity during eyewall replacement cycles. Overall, the negative relationship between IKE and RHLO is likely tied to these two dry air intrusion mechanisms, but additional studies on the physical relationship between IKE and RHLO is clearly warranted to ensure that this negative coefficient is not an simply artifact of the statistical methods that govern SPIKE.

3.5 Linear Model Skill Evaluation from 1990 to 2011

The shared variance between the SPIKE model and changes of IKE is shown for selected forecast intervals during the training period in Table 3.2. As is the case with SHIPS (DeMaria and Kaplan 1994), the explained variance increases with increasing forecast hour. At first, this appears counterintuitive as forecast error typically increases with lead time. However, the average magnitude of IKE change from 1990 to 2011 is much smaller in the shorter forecast intervals than in the larger forecast intervals (9 TJ for 12hr; 32 TJ for 72hr). Considering the errors and biases within the historical archive of operational wind radii, the calculations for observed IKE likely contain biases on the same order as, or even greater than, the IKE changes themselves for the shorter forecast intervals. Therefore, the model will perform poorly at explaining these smaller short-term changes that are dominated by observational biases. Furthermore, the predictors used to train SPIKE in this perfect-prog approach are observed and

not forecasted. Therefore, this exercise is not hurt by forecast biases and errors, which would ostensibly increase with forecast hour.

The shared variance scores of SPIKE for predicting IKE changes are all significant considering the large sample sizes from using thousands of storm fixes between 1990 and 2011. The shared variance statistics for SPIKE are particularly impressive at the longer forecast intervals ($r^2 = 0.54$), where they approach and in some cases exceed the shared variance levels for SHIPS and TC intensity (DeMaria and Kaplan 1994; 1999). Admittedly, the model performs quite poorly in the shorter ranges, especially considering observed predictors are used instead of forecasted predictors.

Although SPIKE is designed to predict the normally distributed quantity of integrated kinetic energy change, it can still be adapted to predict total IKE values (Figure 3.2). This is done by adding the estimate of IKE tendency from SPIKE to the known persistence IKE value. As a result of incorporating persistence into the IKE forecast, the shared variance levels between SPIKE's total integrated kinetic energy estimates are significantly higher than its estimates for just IKE fluctuations. At a forecast interval of 12hrs, SPIKE can estimate total IKE with a staggering explained variance of 84%. This shared variance drops off to 70% by 30 hours and a still impressive 60% by a forecast interval of 72 hours (Figure 3.3a).

The use of persistence to obtain total IKE projections allows the SPIKE products to take advantage of the inertial nature of IKE quantities. Whereas a point metric like maximum sustained wind can and does change rapidly somewhat regularly (e.g. Kaplan and DeMaria 2003), kinetic energy integrated across the entire wind field does not change as rapidly. Although drastic intensity changes do impact IKE values, rapid intensification (RI) events typically result in a drastic increase of near surface winds over a small confined area of convection near the center of the storm. Therefore, the impact on IKE during RI is typically small, provided the overall size of the storm's wind field remains somewhat constant. As a result, a persistence forecast of IKE is typically very skillful, especially in a short forecast

interval. However, at longer forecast intervals, persistence does not fare nearly as well (r^2 = 25% at 72hrs; Figure 3.3a). In fact, the SPIKE regression model is more skillful at estimating total IKE values at a 72-hour forecast interval than is persistence at forecasting the same quantity on a much shorter 30-hour interval. The fact that a persistence forecast dramatically worsens over time indicates that environmental and storm-specific data must be utilized for longer-term forecasts of IKE.

Absolute mean errors were also calculated for the total SPIKE model by computing the average magnitude of the differences between the SPIKE forecast and the observed total IKE values for each of the storm fixes in the 1990-2011 training interval. These mean errors for the total SPIKE model and for a persistence forecast are plotted in Figure 3.3b. It should be noted that the mean errors for the persistence forecast are equal to the mean magnitude of IKE change for each forecast interval since a persistence forecast by definition will predict a change of zero TJ. Therefore, the mean magnitude of errors for the SPIKE model's total IKE projections are of the same magnitude as the IKE changes themselves within the shortest forecast intervals (e.g. ~10 TJ for the 12 hour forecast interval; ~15 TJ for the 24 hour forecast interval). This ultimately means that a short-term total IKE forecast is not significantly better than persistence. While unfortunate, this does not come as a surprise, as SPIKE has low shared variance scores in these shorter forecast windows for predicting the more subtle IKE changes (Table 3.2). However, a persistence forecast has significantly more error than does the SPIKE model in forecast intervals greater than 24 hours. In fact, at a forecast window of 48 hours, total IKE projections from SPIKE have a mean error magnitude of 18.6 TJ compared to a much higher 24.2 TJ from a persistence forecast. Therefore, the mean error statistics support the conclusions drawn from the shared variance statistics, wherein persistence is a tough forecast to beat within 24 hours and the SPIKE model exhibits significant skill over persistence beyond that point.

Analysis of mean squared error (MSE) values in the SPIKE model, with respect to those from a persistence forecast, further emphasize that SPIKE offers a significant skill improvement over a persistence forecast during the training interval (Figure 3.4). In a twelve-hour forecast window, SPIKE exhibits a 14% reduction of MSE values when compared to a persistence forecast. Two-sample bootstrapping tests using the squared errors of the SPIKE model and a persistence forecast suggest that SPIKE's improvement over persistence is not quite significant at the one-sided p=0.10 level for this short forecast window. However, at 30 hours, SPIKE exhibits a 36% improvement over a persistence forecast in terms of MSE, which is significant at the p=0.01 level according to two-sample bootstrapping tests. The percent decrease of MSE in the SPIKE model relative to persistence within the training interval continues to improve across all later forecast intervals, culminating in a 60% reduction of MSE in the 72-hour SPIKE model compared to a 72-hour persistence forecast.

In addition to simply calculating statistics over the calibration interval, some validation exercises are also performed using standard bootstrapping techniques. These bootstrapping exercises are done by training the model over a sample that is created by randomly selecting data points from the overall population of IKE and predictor data (repetition allowed). The regression coefficients from the model trained over this sample are then used over the original population to examine how model performance changes. In the case of the SPIKE model for kinetic energy change, there is an average decrease in shared variance of 3.7% across all twelve of the forecast intervals. The decrease of shared variance for the total kinetic energy estimates is less significant, averaging less than 0.5% across all of the forecast intervals. Ultimately, these simple tests indicate that SPIKE should be able to retain predictive skill when using a different sample of data. However, once again, it should be noted that because developmental SHIPS data is used for the predictors, there likely will be a decrease in skill when using forecasted predictors in an operational setting.

3.6 Validation Tests of Linear Model During 2012 Hurricane Season

The 2012 Atlantic Hurricane Season consisted of 19 tropical cyclones, ten of which were hurricanes and two that eventually reached major hurricane status. A total of 395 storm fixes were taken from the extended best track dataset to estimate IKE for each storm. Similar to storms in the training interval from 1990 to 2011, the TCs of the 2012 Atlantic season had IKE values less than 25 TJ for most of their lifetimes. However, four storms (Hurricanes Leslie, Nadine, Rafael, and Sandy) all obtained in excess of 100 TJ of IKE, mostly in the latter stages of their lifecycles over the middle and upper latitudes of the basin.

The same SPIKE models that were calibrated on the 1990-2011 data in the previous sections (coefficients listed in Table 3.2) are utilized to project observed fluctuations of IKE during the 2012 Atlantic Hurricane Season, as means of determining SPIKE's potential predictive skill outside of the training interval. Predictor data for the 2012 season is once again taken from the best track dataset (Jarvinen et al. 1984) and the SHIPS developmental data set (DeMaria and Kaplan 1999). Therefore, this analysis, which does not use dynamically forecasted predictors, is not meant to assess the operational skill of SPIKE. Instead, this exercise serves as an evaluation of SPIKE's skill outside of the training interval when given idealistically accurate predictors.

Overall, the SPIKE model exhibits comparable skill during the 2012 season when compared to the 1990-2011 training interval. The explained variance and mean absolute forecast errors are shown in Figure 3.5. SPIKE explains an even higher percentage of the observed total IKE variance in 2012 when compared to the training interval. At a forecast interval of 36 hours, the 2012 SPIKE model explains an astonishing 83% of the variance of total IKE values, as compared to the 67% explained by SPIKE during the longer training interval. The main reason for this apparent enhanced skill during the 2012 season, when compared to the longer training interval, likely stems from the fact that persistence corresponds extremely well

with future IKE values during the 2012 season. In fact, out to 36 hours, a simple persistence forecast explains more than 80% of the variance during the 2012 season, suggesting that the kinetic energy levels of TCs during this one season were even more inertial than should be expected on average.

In terms of mean absolute errors, SPIKE does not appear to perform as well during the 2012 season when compared to the 1990-2011 training interval (Figure 3.5b). An increase in forecast error should be expected since the 2012 data was not in the training interval that was used to calibrate the regression coefficients in SPIKE. In this case, the error at a forecast window of 72 hours was 29.2 TJ during the entire 2012 season, which is a staggering 36% higher than the mean 72-hour error during the 22-year training interval. The highest SPIKE errors during the 2012 season unsurprisingly occurred during Hurricane Sandy. As previously mentioned, Sandy had near-record levels of IKE near the end of its lifecycle. The SPIKE model, at all forecast intervals, correctly projected that Sandy would have IKE levels above 150 TJ before reaching the Mid-Atlantic coastline. This forecasted level of IKE from SPIKE would have placed Sandy within the top 2.5% of all TC fixes during the training interval. Therefore, in many regards, the SPIKE forecast still would have been more than adequate to categorize the high damage potential of Sandy, which explains why the shared variance score of SPIKE during the 2012 season is not similarly worse than the training interval shared variance levels. Nonetheless, SPIKE was unable to project that Hurricane Sandy would reach near-record IKE values exceeding 400 TJ. As a result, even a very high forecast of 300 TJ for Sandy would contain an enormous error magnitude of more than 100 TJ, contributing to the much higher than expected mean error levels shown in Figure 3.5b.

If the final dozen storm fixes of Hurricane Sandy were neglected, the SPIKE model would have significantly less mean error during the 2012 season (Figure 3.5b). In fact, the mean error during 2012, excluding these Sandy data points, would be strikingly similar to the mean errors during the full 22-year training interval at all forecast intervals. Mean squared error

statistics (Figure 3.6) tell a similar story, wherein the SPIKE model improves upon a persistence forecast during the 2012 season, excluding Sandy, for all forecast hours. Like the MSE reduction analysis done for the training interval in Section 3.5, SPIKE performs increasingly better than persistence with increasing lead time. The percent reduction of MSE in SPIKE relative to persistence becomes significant at the one-sided p=0.05 level at a forecast interval of 48-hours, where SPIKE has 35% less MSE than a persistence forecast. At a forecast interval of 72-hours, SPIKE has a reduction of MSE relative to persistence of nearly 45% during the 2012 season, excluding Sandy.

Overall, this exercise suggests that the SPIKE model performs well over a validation dataset such as the 2012 season. Skill relative to persistence increases with increasing lead time, as expected. Within a twenty-four hour forecast window, the high correspondence between SPIKE projections and historical IKE in 2012 can be attributed to the skill of a short-term IKE persistence forecast. However, outside of this one-day forecast interval, the SPIKE model performs significantly better than a persistence forecast in this validation exercise.

3.7 Estimating Linear Model Skill in an Operational Setting

Since SPIKE was trained in a perfect-prog mode, it is fair to expect the linear regression model to have somewhat reduced skill levels in a truly operational setting because the predictors will be imperfectly forecasted for future time steps in nearly all scenarios. Previous work done by DeMaria (2010) offers some insight with regards to the amount of skill that could be lost by moving from a perfect-prog approach to an operational approach with imperfectly forecasted predictors. DeMaria (2010) found that the skill of a statistical-dynamical model, such as the Logistic Growth Equation Model (LGEM), did not drastically improve when given perfect large-scale environmental predictors. Since most of the predictors in the linear SPIKE model are

indeed large-scale environmental predictors, this bodes well for the potential skill of the SPIKE model in an operational setting.

However, DeMaria (2010) noted that smaller-scale predictors that are more dependent upon track are more likely to affect model performance when moving from a perfect-prog approach to an operational approach with forecasted predictors. This discovery is relevant to SPIKE, because MSLP is used in the SPIKE model. The intensity of a TC obviously is quite dependent on storm track, especially when a storm is nearing a coastal environment. Considering that MSLP is an important predictor in the SPIKE model, it is worthwhile to examine how an imperfect estimation of MSLP will affect the skill of SPIKE.

Therefore, the observed MSLP predictor in the perfect-prog approach is replaced with an MSLP persistence forecast, leaving all other “perfect” predictors the same. In most cases, a persistence forecast of MSLP will be a poor indication of future MSLP values, as TC intensity tends to change somewhat rapidly, particularly when storms near landfall. Nonetheless, the performance of SPIKE decreases by no more than 20% across all forecast intervals, in terms of both explained variance and MSE, when the poor persistence MSLP predictor is used in place of the “perfect” observed MSLP predictor used throughout the previous sections. In fact, a SPIKE model trained with an MSLP persistence predictor is still significantly better at projecting integrated kinetic energy than an IKE persistence forecast at the one-sided p=0.10 level in the training interval for all forecast intervals exceeding 24 hours. This further suggests that the SPIKE model will likely offer an improvement over a persistence forecast of IKE, even when using imperfect operational predictors.

Table 3.1: Variables considered for use in the SPIKE models. Many of the variables (e.g. RHLO, SHRD) originated from SHIPS and are averaged over specific areas as noted. Others were specifically created from observations for the SPIKE model. Not all of these variables are used in the final models, as many of the regression coefficients fail significance tests in the backward screening methodology.

Variable	Definition	Units
PIKE	persistence of IKE	TJ
dIKE12	previous 12hr change of IKE	TJ
VMAX	maximum sustained wind speed	kts
dV12	previous 12hr change of VMAX	kts
VMPI	Difference between maximum potential intensity and VMAX	kts
LAT	latitude of storm's center	°N
LON	longitude of storm's center	-°W
DTL	distance to nearest landmass	km
UMOT	zonal translational storm motion	$m \cdot s^{-1}$
VMOT	meridional translational storm motion	$m \cdot s^{-1}$
MSLP	minimum sea level pressure	hPa
PENV	average surface pressure *(averaged from r=200-800km)*	hPa
VORT	850 hPa vorticity *(r=0-1000km)*	$10^{-7}\ s^{-1}$
D200	200 hPa divergence *(r=0-1000km)*	$10^{-7}\ s^{-1}$
SHRD	850-200 hPa shear magnitude *(r=200-800km)*	kts
SHTD	850-200 hPa shear direction *(r=200-800km)*	°
SHDC	SHRD but relative to 850 hPa center with vortex removed *(r=0-500km)*	kts
RHLO	850-700 hPa relative humidity *(r=200-800km)*	%
RHMD	700-500 hPa relative humidity *(r=200-800km)*	%
T150	150 hPa temperatures *(r=200-800km)*	°C
RD20	ocean depth of 20°C isotherm	m
RD26	ocean depth of 26°C isotherm	m
OHC	ocean heat content	$kJ \bullet cm^{-1}$
SST	sea surface temperatures	°C
EPSS	mean difference between lifted sfc. parcel θ_e and envir. θ_{es} *(r=200-800km)*	°C
TWAC	850 hpa symmetric tangential wind *(r=0-600km)*	$m \cdot s^{-1}$
TADV	low-level temperature advection by thermal wind *(r=0-500km)*	$10^{-6}\ s^{-1}$
REFC	relative eddy momentum flux *(r=100-600km)*	$m \cdot s^{-1} \bullet day^{-1}$
PEFC	planetary eddy momentum flux *(r=100-600km)*	$m \cdot s^{-1} \bullet day^{-1}$
SDAY	time after tropical storm genesis	days
PDAY	time from peak of season (Sept. 10)	days

Table 3.2: Regression coefficients for each predictor in the SPIKE model. The coefficients listed in a blue font are significant at the 99% level for that forecast hour. Sample size for the training interval of 1990 to 2011 is shown below. Finally, shared variance between the SPIKE regression model, and observed kinetic energy changes are listed in the bottom row.

Variable	Forecast Hour				
	12hr	24hr	36hr	48hr	72hr
PIKE	-0.53	-0.70	-0.72	-0.70	-0.70
dIKE12	0.11	0.02	0.02	0.00	-0.01
RHLO	-0.13	-0.19	-0.22	-0.23	-0.25
SHRD	0.06	0.11	0.12	0.13	0.13
DTL	0.14	0.15	0.15	0.14	0.15
D200	0.12	0.13	0.12	0.12	0.11
MSLP	-0.26	-0.37	-0.39	-0.40	-0.41
VORT	0.16	0.19	0.20	0.20	0.20
LAT	0.24	0.30	0.33	0.34	0.36
RD26	0.11	0.11	0.10	0.10	0.09
EPSS	0.06	0.08	0.09	0.09	0.11
PENV	-0.08	-0.10	-0.13	-0.14	-0.18
REFC	0.07	0.08	0.07	0.05	0.04
PDAY	-0.08	-0.10	-0.11	-0.10	-0.08
Sample Size	4814	4239	3756	3341	2657
Shared Variance	13%	25%	39%	43%	54%

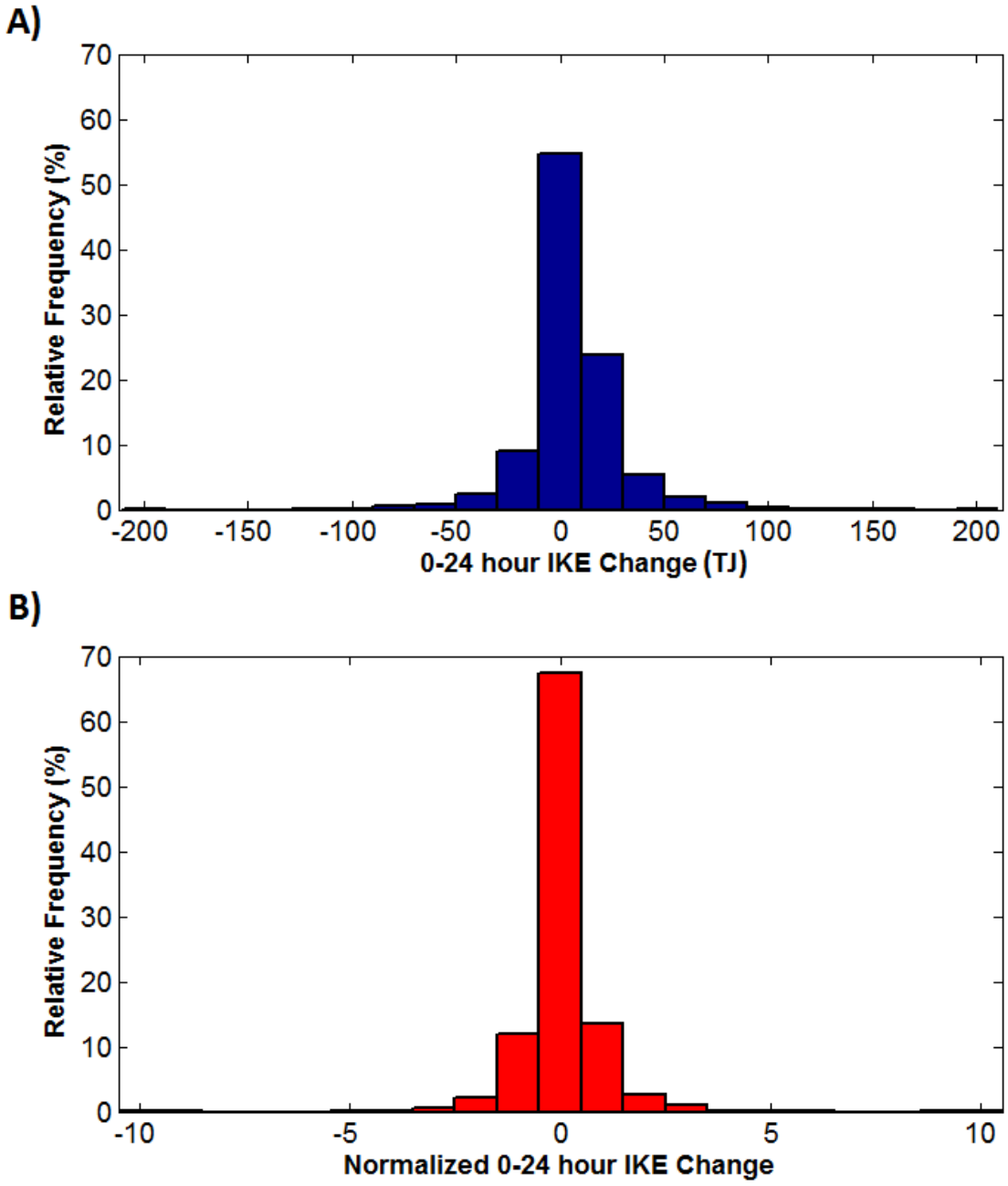

Figure 3.1: Relative frequency distribution of 0-24 hour IKE changes (panel A), and normalized 0-24 hour IKE changes (panel B). The normalized changes of IKE at various forecast times is used as the target predictand for the linear SPIKE regression models.

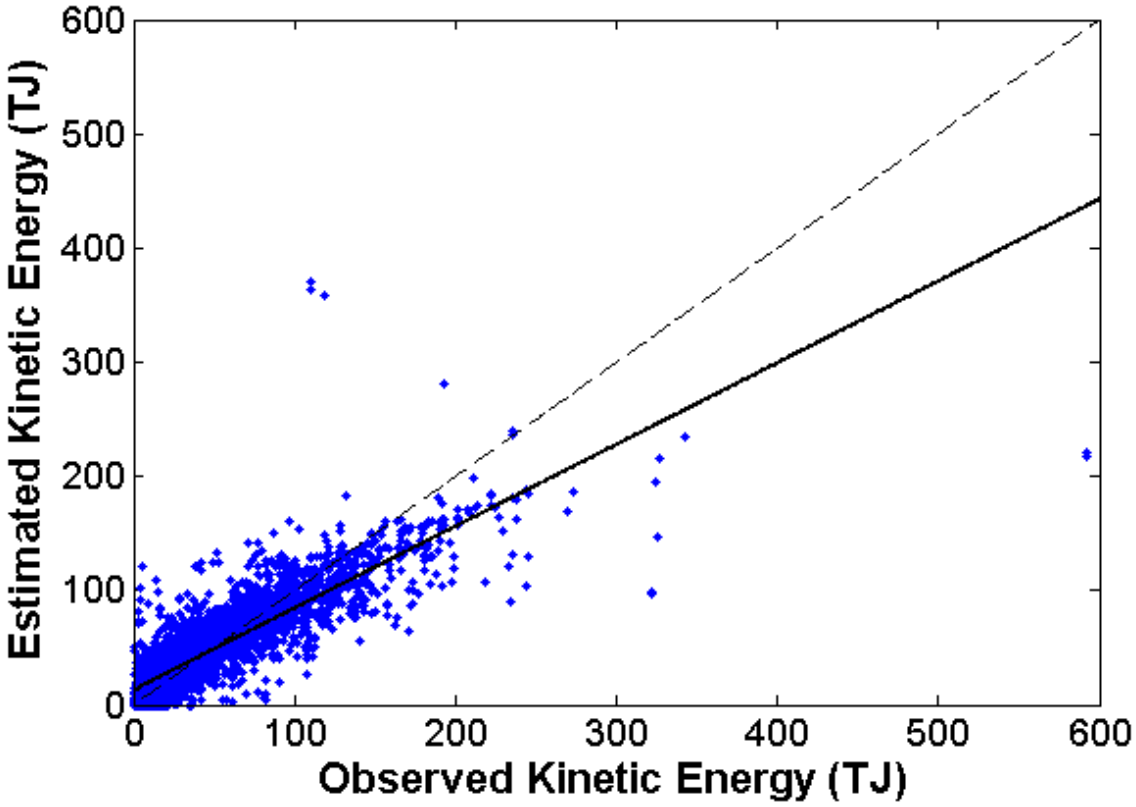

Figure 3.2: Scatter plot of observed total IKE values vs. estimated total IKE values from a 24-hour SPIKE regression model (blue dots). The dark black line represents the best fit line between the 4239 observed and estimated data points. The dashed line represents a perfect forecast (y=x). The correlation between the observed and estimated values is r=0.85 which is significant beyond the p=0.01 level given the high number of samples.

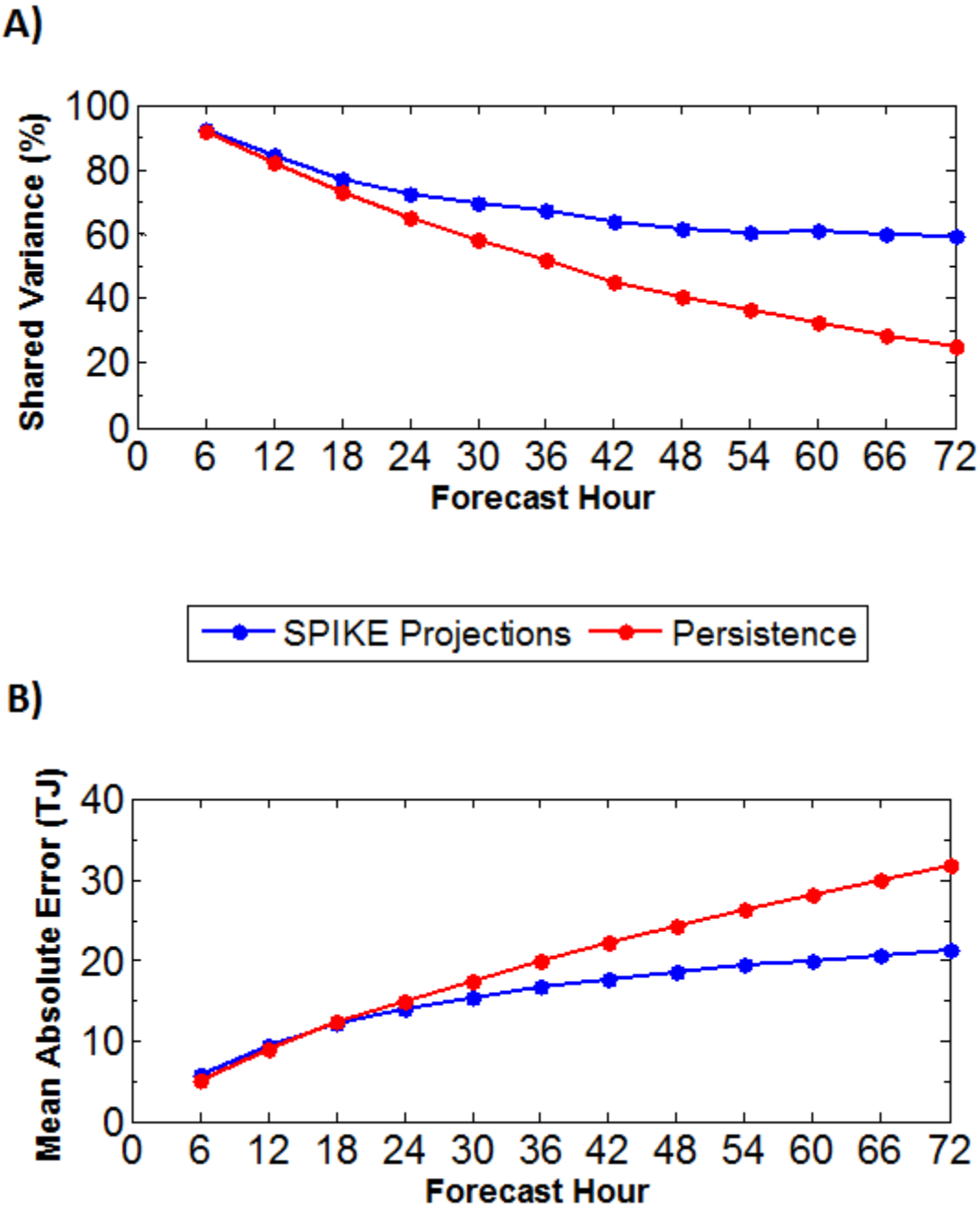

Figure 3.3: Evaluation of SPIKE's performance relative to persistence during the 1990-2011 training interval. Plot of shared variance over forecast hour (panel A) and mean absolute error over forecast hour (panel B) for total IKE. The blue line represents these metrics between the observations and the SPIKE model at each forecast hour. The red line represents these metrics between a persistence forecast and the observed IKE value a certain number of hours later.

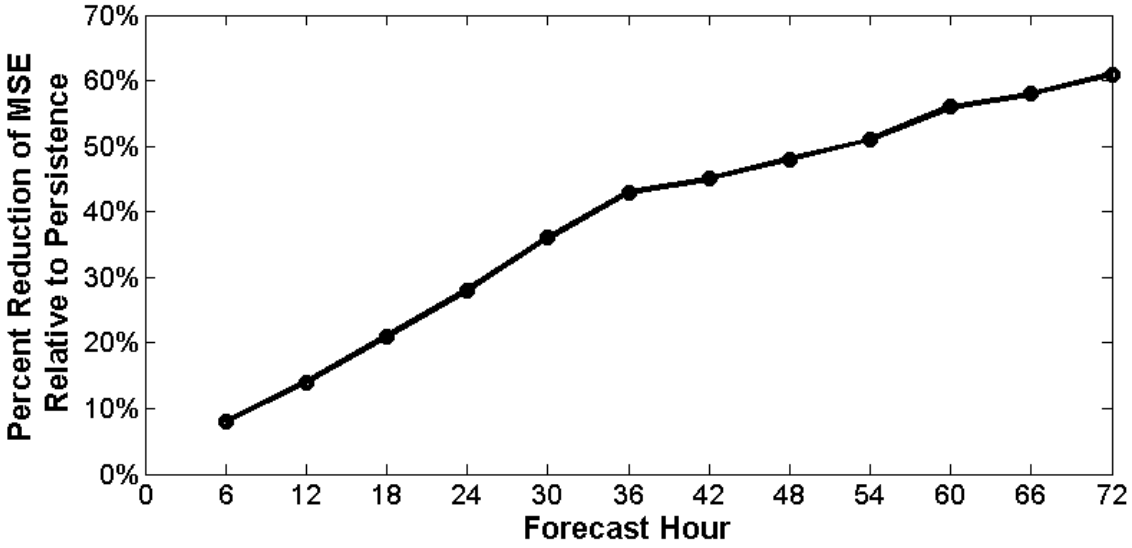

Figure 3.4: Percent reduction of mean squared error (MSE) for the SPIKE model during the 1990-2011 training interval with respect to persistence. A reduction of MSE is plotted as a positive percentage, indicating improved model skill. The model's improvement over persistence is significant at the one-sided p=0.05 level for all forecast hours greater than or equal to 24 hours and at the p=0.01 level for all forecast hours greater than or equal to 30 hours.

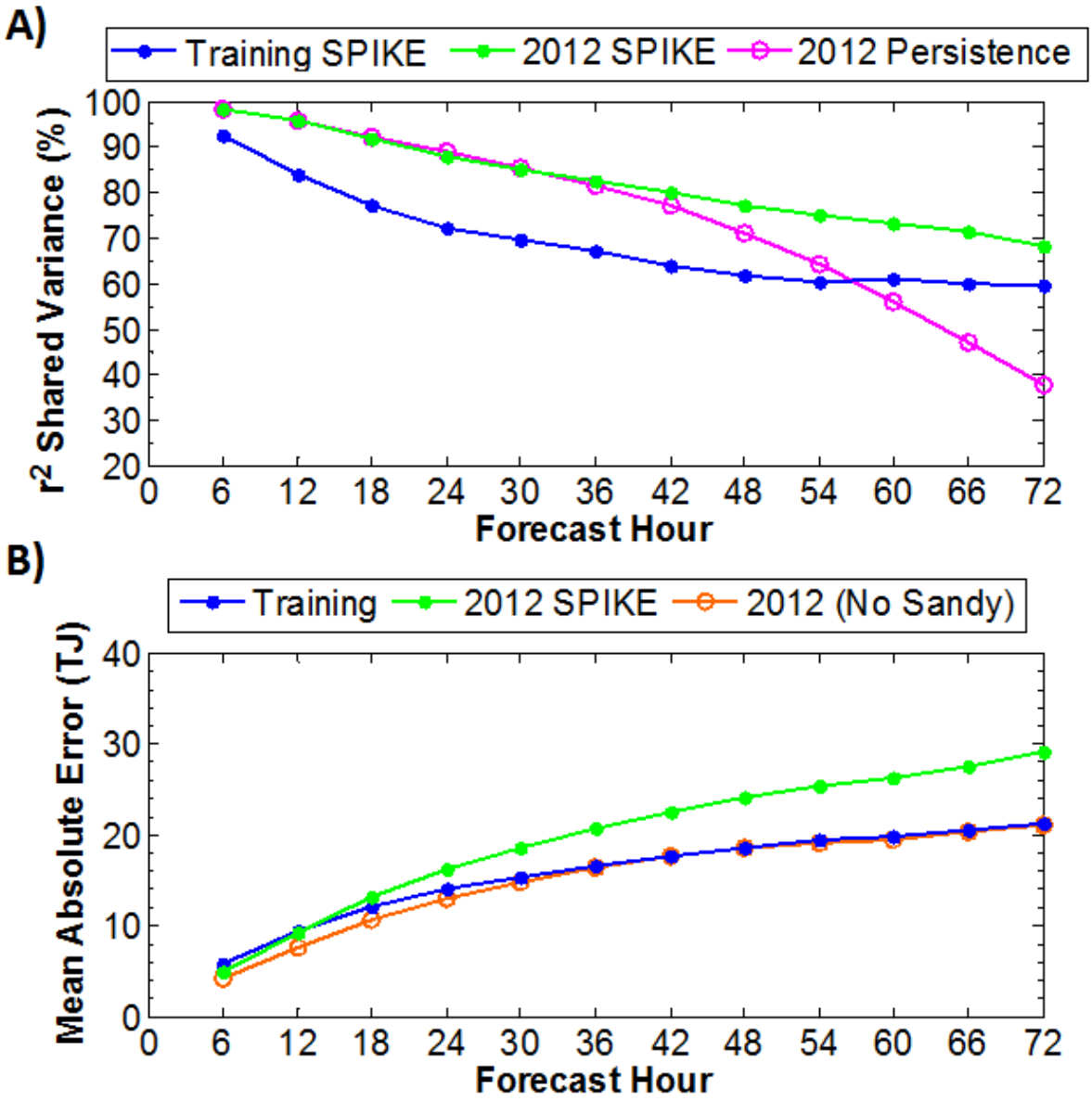

Figure 3.5: Evaluation of SPIKE during the 2012 Atlantic Hurricane season. Panel A depicts a plot of shared variance over forecast hour for total IKE measurements. More specifically, the blue line is the shared variance between the SPIKE model and observed total IKE measurements during the training interval of 1990 and 2011 reproduced from Figure 3.3a, the green line is the shared variance between the validation SPIKE model and observed total IKE values for all storms during the 2012 season, and the magenta line is the shared variance between observations and persistence during the 2012 season. Panel B depicts mean absolute error over forecast hour for total IKE. The blue line represents the mean error between the observations and the SPIKE model during the training period as reproduced from figure 3.3b. The green line represents the mean error of the validation SPIKE model for all storms in the 2012 season. Finally, the orange line represents the mean error of the validation SPIKE model for the 2012 season, excluding the final 3 days of Hurricane Sandy's lifecycle.

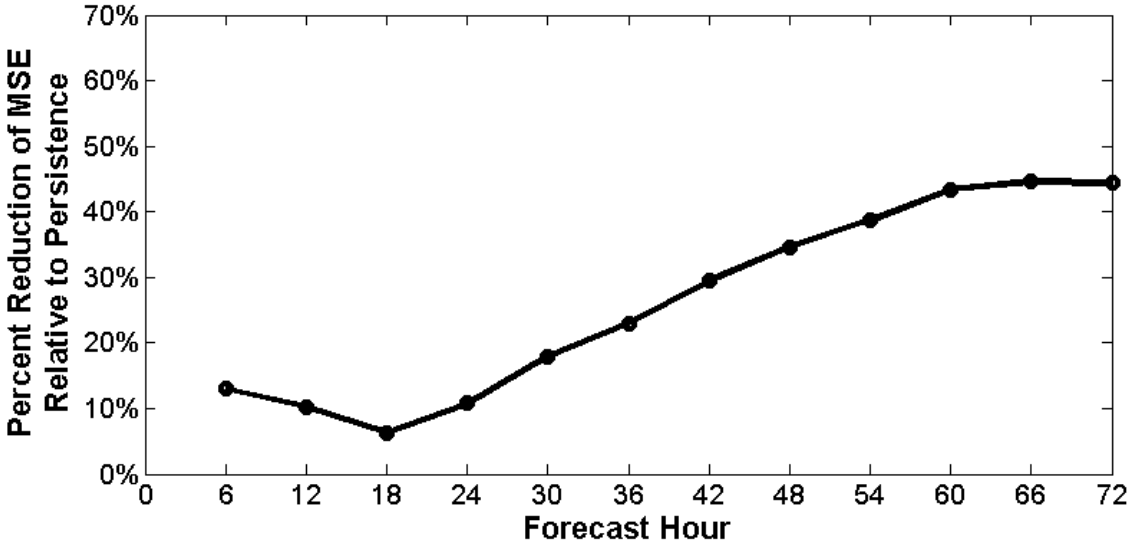

Figure 3.6: Percent reduction of mean squared error (MSE) for the SPIKE model during the 2012 season, excluding the final seventy-two hours of Hurricane Sandy's lifecycle, with respect to a persistence forecast. A reduction of MSE is plotted as a positive percentage, indicating improved skill. The model has significantly lower MSE than persistence at the p=0.05 level for all forecast hours greater than or equal to 48 hours.

CHAPTER 4

PREDICTION OF IKE WITH NEURAL NETWORKS

In the previous chapter, fluctuations of IKE in Atlantic TCs were successfully predicted from a series of environmental predictors using a simple linear regression model. The linear SPIKE model proves to be skillful relative to a simple persistence forecast in a perfect prognostic mode, and as a result, SPIKE serves as a capable proof of concept. However, a linear regression approach is suboptimal for weather prediction because the earth system is clearly nonlinear. As such, the fixed linear regression coefficients in the SPIKE model will never be able to fully process the complex changing relationships between the environment and IKE variability within a TC.

For example, the SPIKE regression model presented in the previous chapter has a positive coefficient for the deep layer wind shear magnitude predictor, SHRD. This positive coefficient captures the influence of extratropical transitions and trough interactions, both of which cause storms to gain IKE as their wind fields expand over the more highly sheared upper latitudes. Since linear regression coefficients are not allowed to change, the SPIKE model assumes higher shear is favorable for IKE growth in every circumstance. The inflexibility of the linear model is unfortunate because the model is incapable of anticipating the many observed scenarios in which decreased wind shear over the lower latitudes of the basin is beneficial to storm intensification, storm growth, and increasing IKE. Ultimately, the conflicting nonlinear signals in the TC-environment system underscore the need for a more complex statistical model that is capable of learning and adapting to a number of different scenarios to better predict observed changes of IKE.

As a result, this chapter focuses on adapting artificial neural networks (ANN) for the prediction of IKE within TCs in the TNA basin. These networks are capable of adapting and

learning, and as a result they are better suited to model a nonlinear system. The first section will serve as a brief introduction to ANNs. The discussion in Sections 4.2 and 4.3 focuses on describing the data and methods that are used to incorporate neural networks into the improved Statistical Prediction of Integrated Kinetic Energy Version 2 models (SPIKE2). Finally, the skill of the SPIKE2 system will be evaluated in a perfect-prog mode from 1990-2011 relative to a persistence forecast and the linear SPIKE model. Throughout the evaluation of SPIKE2, an effort is made to show the distinct advantages of using neural networks over the simpler linear regression models.

4.1 Brief Introduction to Neural Networks and their Applications

In short, artificial neural networks are complex nonlinear systems that have the capability to learn, and as a result are able to generalize and associate data through artificial intelligence (e.g. Kriesel 2007). The motivation for the creation of ANNs stems from the biological processes that govern decision making within the human brain. As a result, neural networks are typically comprised of many layers of connected neurons, through which data is propagated until an appropriate output is generated.

Over time, researchers have found various useful tasks for ANNs including classification or clustering of data, recognition of patterns, and data fitting and prediction. The adaptability of ANNs has made them valuable for researchers across many different fields, including the atmospheric sciences. For example, neural networks have been used in meteorology for processing remotely sensed data (e.g. Atkinson et al. 2010), classifying circulation patterns (e.g. Cawley and Dorling 2005), predicting troposphere ozone levels (e.g. Abdul-Wahab and Al-Alawi 2002), forecasting wind speeds (Cao et al. 2012), forecasting precipitation and flooding (e.g. Hapuarachchi et al. 2011), and predicting the strength of the Indian monsoon on a seasonal scale (Shukla et al. 2011). A more detailed summary of earlier ANN applications in meteorology

can be found in a review by Gardner and Dorling (2002). Ultimately, the goal in this chapter is to build upon these existing meteorological applications by adapting a series of ANNs to predict values of IKE in Atlantic TCs.

Before jumping into neural network based forecasts of IKE, it is appropriate to briefly describe the general components of an ANN. As alluded to earlier, an ANN is comprised of several connected artificial neurons. Each individual neuron is defined as a processing unit of the ANN (Kriesel 2007), and each of these neurons are comprised of several important internal elements. In the most basic sense, a neuron will contain weights, a propagation function, and an activation function (Figure 4.1). The weights in each neuron are used to determine the strength of the connection between the neuron and each of the values that are inputted into the neuron. In most cases, a neuron will be given several input values and a single bias value that is always equal to one. Therefore, a neuron will have a unique weight value for each of these individual inputs. The propagation function within each neuron is often a simple weighted sum, wherein the neuron processes the inputted data by multiplying each input value by its associated weight and then totaling the quantity. This weighted sum is then passed on to the activation function. The activation function may be a stepwise function, a linear function, a sigmoid function, or anything in between. In short, the activation function processes the value passed on by the propagation function, and determines the final output of the neuron.

Figure 4.1 and Table 4.1 can be used to together to create some examples in an effort to better explain how an artificial neuron works. Figure 4.1 depicts a crude representation of a single neuron. The columns of Table 4.1 contain three example sets of hypothetical input values (x) that can be passed through the sample neuron. For each of these examples the weights (w) will be set as follows:

$$(4.1) \quad w_0 = 1;\ w_1 = -1;\ w_2 = 3;\ w_3 = -2$$

Similarly the propagation function will be a weighted sum and the activation function will be a piecewise function, wherein the output (y) will be defined as follows:

$$\sum_{i=0}^{3} x_i w_i < 0 \ \rightarrow y = -1$$

(4.2)

$$\sum_{i=0}^{3} x_i w_i \geq 0 \ \rightarrow y = 1$$

Based upon these functions and weights, the final outputs of the neuron are shown in the bottom row of Table 4.1 for each example.

Neural networks are not comprised of just one neuron, however. In fact, neural networks typically are comprised of multiple layers of several neurons each. These neuron layers can be aligned in many different hierarchies, and data can be passed through them in many different directions. However, in this chapter, the focus will be on a specific type of network called a two layer feed-forward neural network because, ultimately, this type of network is utilized to project IKE in the next several sections. As would be expected in a “feed-forward” network, data is passed on from the input sources through each subsequent layer of neurons until a final output value is reached. In a two layer feed-forward network, the first layer of neurons, or the hidden layer, receives data from each input source. The second layer, or the output layer, receives the outputs from each of the hidden neurons and produces the final output(s) of the network. An example of a small two-layer feed-forward network that produces two outputs is shown in Figure 4.2.

The final component of ANNs that needs to be introduced is the artificial intelligence itself. In short, machine learning within an ANN allows for these powerful tools to adapt and generalize. The neural networks created later in this chapter all utilize a type of machine learning procedure called “supervised learning.” In supervised learning, the neural network is trained upon a series of input patterns and a series of correct output results, such that a precise error vector can be calculated. Supervised learning is advantageous when constructing a

forecast model such as SPIKE when a known set of historical target data is available to train the model.

The most general procedure for supervised learning is to first enter the series of input patterns and propagate them though the network as previously discussed in this section. The output series generated by the network is then compared with the desired target output series. Based on the error vector, the network determines whether or not any corrections are required to improve the predictive skill of the network. In theory, there are many ways to apply corrections (i.e. changing weights, changing activation functions, changing propagation functions, creating new neurons, or deleting existing neurons). However, in nearly all cases, neural networks adapt by making corrections to the weights within each neuron in an effort to converge on a solution with a minimal amount of error. Since the weights within each neuron are corrected throughout the learning process, they are typically initialized as somewhat random or arbitrary values. Therefore, the error values are typically large to start. The learning process will continually correct these weights, reducing the error values in the training dataset until the error values reach a sufficiently low level.

4.2 Data Used to Create Neural Networks

The data used to create the neural networks in this chapter are very similar to the data used to train a linear regression model in the previous chapter. Once again, a perfect prognostic approach will be utilized to establish a proof of concept for using neural networks to predict IKE in Atlantic TCs. As such, historical data from observations and analyses will be used to train the neural network, in an effort to determine the maximum potential skill of the neural networks in SPIKE2 relative to persistence. The target output data for the neural network is based upon historical record of IKE in Atlantic storms from 1990-2011, which was previously used to construct the linear regression version of SPIKE. As previously discussed, this historical record

of IKE is estimated from archived operational wind radii in the extended best track database (Demuth et al. 2006) using an approximate equation set established by Powell and Reinhold (2007). Similar to the linear regression version of SPIKE, the neural networks will not be used to project the exact total IKE values from the historical record. Instead, SPIKE2 will attempt to project changes of IKE over twelve separate forecast intervals ranging from 0-6hrs out to 0-72hrs.

The input variables used by the ANNs in SPIKE2 will be comprised of a pool of environmental and internal storm-specific quantities that are similar to those used in Chapter 3 with the linear version of SPIKE. To reiterate, those predictors are taken from a combination of the NHC best track data set (Jarvinen et al. 1984) and the SHIPS developmental data set (DeMaria and Kaplan 1999). A full list of the eighteen predictors used to construct the neural networks is shown in Table 4.2. Once again, it is important to note that some of the SHIPS predictors have missing or null data values scattered throughout the training interval of 1990-2011. Most notably, some variables are entirely unavailable prior to 1995 at the beginning of the developmental SHIPS record. Therefore, to account for the missing predictor values, the neural networks will be trained on storm fixes that have a complete set of input parameters with no missing variables. As a result of the missing data prior to 1995, the neural networks are trained and calibrated on complete sets of data for storm fixes from the 1995 Atlantic Hurricane Season through the 2011 hurricane season. Ultimately, this results in a sample of more than 4,000 storm fixes in the 6-hour forecast window training sample trailing down to about 2,200 storm fixes for the 72-hour forecast window training sample. Again, the decrease in sample size with increasing forecast hour is attributed to the fact that the longer forecast intervals require long-term persistence variables that may not exist in short-lived TCs.

Each of these predictors is normalized over the 1995-2011 sample before being passed into SPIKE2's neural networks, where in a predictor value of zero is equal to the 1995-2011 sample mean. Normalizing input predictors is not necessary in ANNs because unlike linear

regression models, ANNs are nonlinear by nature. Nonetheless, the predictors for SPIKE2 are normalized with operational forecasting in mind. To use SPIKE2's neural networks in an operational forecast setting, predictors must be imported in real time from a dynamical weather prediction model. Many of these dynamical forecast models obviously contain significant biases that will hurt the adaptability of SPIKE2 in a forecast environment. For example, coarser global models will often underestimate the intensity predictors such as VMAX and MSLP because these models lack the resolution needed to resolve the eyewall of a TCs. By normalizing each predictor over a representative sample of data from archived forecast model data, the biases in the mean and variance of the dynamical model's predictors can be somewhat minimized, allowing SPIKE2 to be adapted to various different dynamical forecast models without much inconvenience.

In addition, it should be noted that the pool of predictors used to construct the neural networks listed in Table 4.2 is smaller than the more expansive pool of candidate predictors tested with the linear models in Table 3.1. This intentional reduction of predictors is once again done to make it easier to adapt SPIKE2 for real-time forecasting. As such, predictors were removed if they could not easily be calculated from the wind, moisture, and temperature fields that are almost always found in reforecast archives of models such as the GEFS (Hamill et al. 2013).

4.3 Network Setup and Calibration

A) Network Hierarchy:

As stated previously, the neural network used in SPIKE2 to predict changes of IKE in Atlantic TCs will be a two-layer feed-forward network. This network hierarchy will include a hidden layer with multiple neurons and an output layer with a single neuron that will ultimately produce the final IKE tendency projection from the eighteen normalized input variables. Each of

these ANNs will be trained using shared learning, wherein the weights of the network's neurons will adapt from their somewhat random initial value until the error function reaches a minimum. More specifically, the learning algorithm uses a Levenberg-Marquardt backpropagation algorithm (Marquardt 1963) to find this error minimum. This specific algorithm is designed to solve non-linear least squared problems and is typically thought to be an efficient and stable method for converging at an optimal solution in neural network learning (e.g. Hagan and Menhaj 1994).

The final missing aspect from the network hierarchy is to determine exactly how many neurons should be used within the hidden layer of each feed forward ANN. As the size of the hidden layer increases, the ANN becomes able to recognize more signals and patterns between the inputs and the output, resulting in better predictive skill (Berry and Linoff 2004). On the other hand, if there are too many neurons in the hidden layer the network will begin to suffer from overfitting and memorization, and predictive skill will begin to fall off (Berry and Linoff 2004). Ultimately, the optimal number of neurons often depends on the number of inputs and outputs of a neural network and possibly the sample size of the training dataset. Unfortunately, there is no universal guideline to determine the number of neurons that should be within the hidden layer of an ANN (e.g. Kriesel 2007; Stathakis 2009; Berry and Linoff 2004).

Therefore, an exhaustive search approach is taken here to find the optimal number of neurons for the SPIKE network. This approach involves generating hundreds of neural networks, each with a different number of neurons in its hidden layer. Ultimately, the skill of each network over a validation sample is evaluated to determine how the size of the hidden layer affects ANN performance and generalization. Analysis of both median MSE and median correlation coefficients suggest that approximately twenty neurons should be used in the SPIKE network, assuming that there are 18 input parameters (Table 3.2) and a single output parameter (IKE tendency). Therefore, all of the ANNs used to project IKE tendency throughout the remainder of this dissertation will contain 20 neurons in the hidden layer of the network.

B) Training, Validation, and Test Samples:

To avoid overfitting and to promote generalization in the supervised learning algorithm, the historical input and target output data series that are used to construct the ANNs will be randomly split into three independent subsets. The first subset of data, named the training sample, is comprised of 70% of the input and target series. As its name suggests, the training sample is used to train the network by establishing the optimal weights within the neurons. The validation sample is a smaller subset, comprised of 15% of the historical input and target series. This subset is ultimately used to determine when the neural network can stop learning based on the networks ability to generalize effectively. As such, the learning algorithm searches for the point at which the neural network has the least amount of error over the validation subset. Finally, the third subset of input and target data is called the testing sample. This test sample is independent of both of the previous two samples, and is not used in the training of the model in any way. Instead it simply provides an independent measurement of network performance.

C) Neural Network Random Variability:

The methodology used to construct the neural networks introduces random variability into each individual ANN. Specifically, random variability is first introduced when the general population of input and target parameters from 1995 to 2011 is randomly split into the three separate subset samples. Additional random variability is introduced to the neural networks because the weights within the neurons are initialized somewhat randomly before arriving at their optimal weights. Ultimately, the random variability makes it impossible for two neural networks to be exactly identical to one another, even if they are trained on the exact same input and target output datasets. Each neural network weighs connections in the nonlinear system somewhat differently, and as a result some of the networks will seem more accurate in certain

situations but less accurate in other situations. Therefore, it is insufficient to base SPIKE2 off of a forecast from just a single neural network.

Instead, SPIKE2 will utilize a system of one hundred individual neural networks to make its projection of IKE tendency for each of the twelve forecast intervals. As shown by the schematic of SPIKE2 (Figure 4.3), the system of neural networks will produce 100 separate independent projections of IKE tendency from a single set of input parameters. The best deterministic forecast of IKE tendency from SPIKE2 will be taken from the median of these 100 individual projections. Using the median from a large sample of ANNs helps to minimize the random variability present in a single neural network's forecast, thus allowing SPIKE2 to focus on the true skill of the neural networks. The overall skill of this deterministic forecast will be discussed at length in Section 4.4.

In addition to producing a single deterministic forecast, SPIKE2 can be utilized to produce probabilistic projections of IKE tendency. Instead of using perturbed input data from various ensemble datasets, the differing weights included within each of the one hundred neural networks allow for an inexpensive way to compute probabilities for IKE from a single input source. In this manner, probabilistic forecasts will measure the uncertainty present within the ANN methodology, not the amount of forecast uncertainty or initial condition uncertainty present within the input and target output series.

Ultimately, two probabilistic tools are created alongside of SPIKE2. The first of which uses various percentiles as calculated from each of the 100 individual projections to generate error bars for the deterministic IKE tendency projections in an effort to assess forecast uncertainty. The second probabilistic tool uses each individual projection can be used to produce probabilistic projections of exceedance (i.e. Will IKE exceed 25 TJ? 50 TJ? 100 TJ?). The skill of both probabilistic products in SPIKE2 is assessed in Section 4.5.

4.4 Deterministic Network Skill

Overall, the deterministic projections from the SPIKE2 neural network system exhibit a significant amount of skill relative to a persistence benchmark over the 1995 to 2011 calibration interval. Figure 4.4a shows the correlation between the SPIKE2's deterministic median projection of IKE tendency and the observed change in IKE for each forecast interval. As was the case with the linear SPIKE regression model, the ANNs perform better when estimating IKE tendency in the longer forecast intervals. These long-term IKE changes are larger in magnitude than the short-term changes. Thus, the variability of the larger long-term IKE tendency quantities are less susceptible to random noise and biases in the observation techniques that cannot be anticipated by SPIKE2.

Nonetheless, SPIKE2's deterministic projections of IKE tendency are correlated to historical IKE change values at a level greater than r=0.50 in its training sample for even the shortest 6-hour forecast interval. These already significant correlations continue to increase as the forecast interval grows, culminating in a correlation greater than r=0.90 by the 60-hour interval. For comparison, the linear regression SPIKE model projected 6-hour IKE tendency at a correlation of only 0.28. Overall, the improvement over the linear model is quite noticeable at all forecast intervals for the training sample (Figure 4.4). More importantly, the ANNs perform well over their validation and test intervals. In fact, SPIKE2 exhibits more skill at projecting IKE change for its test sample than does the linear SPIKE model over its training interval. Ultimately, the high correlations for SPIKE2 over its test interval suggest that SPIKE2 can retain much of its skill at forecasting IKE tendency when given input parameters outside of its training sample as is required in an operational setting.

As discussed earlier, the deterministic projections of IKE tendency can be added to the known persistence value of IKE at initialization time to produce deterministic forecasts of observed IKE values. Exercises with the linear model in Chapter 3 indicated that forecasting IKE

was more skillful than forecasting IKE tendency because the kinetic energy metric does not change rapidly, making even persistence a good forecast in a 24-hour forecast window or less. Expectedly, the accuracy of a persistence forecast drops off rapidly with time, but the linear SPIKE model is able to use the relationships between IKE and a storm's environment to produce a skillful projection out to at least 72 hours. In fact, the linear model projected total IKE at a 72-hour forecast window with a correlation of r=0.77 and a mean error magnitude of approximately 20 TJ (Figure 4.4b; Figure 4.4c). Ultimately, the linear SPIKE model used its environmental and storm-specific predictors in concert with values of IKE persistence to produce a projection of total IKE that improved upon persistence by more than 50% at 72 hours (Figure 4.5).

Encouragingly, the SPIKE2 neural networks improve upon the linear regression model's abilities to predict deterministic values of total IKE in nearly every measurable aspect. Figure 4.6 shows a sample forecast of total IKE by the SPIKE neural networks for each of the nearly 3000 storm fixes from 1995-2011 on a 36 hour forecast interval. Overall, the 36-hour neural network system shown in Figure 4.6 groups most of its projections very near the dashed center line, which would indicate a perfect forecast. As a result, the mean error magnitude for the SPIKE2 neural networks is only 11 TJ over its training sample at a forecast interval of 36 hours, which is down considerably from nearly 17 TJ in the linear model over the sample forecast window. In total, 41% of the projections are within ±5 TJ of the observed value, 68% of the projections have less than 10 TJ of absolute error, 82% are within ±15 TJ of the observed value, and a whopping 93% of projections have less than 25 TJ of absolute error. Analyses of these errors indicate that the 36-hour SPIKE2 neural network system is more likely to overforecast IKE than it is to underforecasts IKE. For example, 34% of projections have an error greater than +5 TJ and only 25% have an error below -5 TJ. The slope of the best fit line supports the notion that this 36-hour SPIKE2 product has a slight high bias on the whole.

However, SPIKE2's high bias is not consistent. Similar to the linear model, the neural networks tend to underestimate extreme high levels of IKE. Analysis of Figure 4.6 clearly indicates that the largest errors by the 36-hour SPIKE2 neural networks, all of which are negative, occur for large TCs that are observed to have more than 200 TJ of IKE. The failure of SPIKE and SPIKE2 to project these near record levels of IKE can be attributed to the negative relationship between PIKE and IKE tendency, wherein already large TCs are more likely to lose IKE than they are to gain more IKE. The relationship between PIKE and IKE tendency is used by the statistical models to prevent the SPIKE2 projections from running away towards unrealistically high values of IKE. If this relationship were removed, both the linear model and the neural networks would likely have a much higher false alarm rate for TCs exceeding 100 TJ. However, it should be noted that if environmental and storm-specific conditions remain favorable, SPIKE2 still has the potential to accurately project IKE growth in extreme events. For instance, the ANNs in Figure 4.6 correctly identified that IKE would exceed 100 TJ in each of the top 65 (or top 2.1%) storm fixes ranked by total IKE in the observed target dataset. Therefore, in many regards the neural networks' forecast correctly identified a substantial TC that would rank in the top 5% of all TC fixes in terms of IKE even if it did fall several TeraJoules short on its total IKE projection.

On a broader scale, the correlation and mean absolute error metrics for SPIKE2's total IKE projections are plotted for all twelve forecast intervals in Figures 4.4b and 4.4c respectively. These statistics are calculated over each of the neural networks' three data subsamples (training, validation, and test), and they are measured against reproductions of the same error and correlation metrics for a persistence forecast and the SPIKE linear model over its training interval. The neural networks' ability to adapt based on linear and nonlinear relationships allows SPIKE2 to outperform the linear regression models and persistence at nearly every forecast interval for both metrics. For instance, SPIKE2 has a mean absolute error of 12.56 TJ over its training sample at its longest 72-hour forecast interval. The linear regression model, for

comparison, has a mean absolute error of 14 TJ at a much shorter 24-hour forecast interval (Figure 4.4c). Compared to a persistence forecast, SPIKE2 offers a 40% reduction in mean squared error over its training interval at even the shortest six-hour forecast interval (Figure 4.5). At the long 72-hour forecast interval SPIKE2 has nearly 90% less MSE than does a persistence forecast.

Encouragingly, SPIKE2's skill does not sharply degrade over the networks' validation and test samples. Some degradation should be expected because the training sample's statistics measure the networks fit to the input and target data and the test sample's statistics measure the ANNs ability to generalize for input data outside of the training sample. For example, the mean absolute error of SPIKE2 drops off by 5-25% when moving from the training sample to the test sample depending on the length of the forecast interval.

Nonetheless, the neural networks, measured over their test sample, are more skillful than a persistence forecast. At the shorter forecast intervals, SPIKE2 has a similar amount of MSE compared to a similarly short persistence forecast. However, by the 18-hour forecast interval SPIKE2 offers a 35% reduction of MSE on its test sample when compared to a persistence forecast (Figure 4.5). Such an improvement on this short forecast interval is significant at the 90% confidence level based on the results of a two-sample bootstrapping test. By the 36-hour forecast interval, the difference in MSE between a persistence forecast and the SPIKE2 network over its test sample grows large enough such that it is significant at the 99% confidence level. This culminates in SPIKE2 having 78% less MSE than a persistence forecast over the test sample at the longest forecast interval. Ultimately, these results are encouraging because they confirm that SPIKE2 is capable of generalizing at an effective level that is better than a simple persistence forecast.

Compared to a tougher benchmark such as the SPIKE linear regression model, the neural networks still come out on top. For instance, SPIKE2 has a mean absolute error of 16.6 TJ at 72-hours over the test sample (Figure 4.4c). This is about 4 TJ worse than the ANNs over

the training interval, but still better than the linear model by 5 TJ. In fact the 16.6 TJ error of the 72-hour SPIKE2 network is comparable to the linear model's ability to forecast total IKE at a forecast interval that is half as long.

Ultimately, the deterministic projections of total IKE from SPIKE2 are an improvement upon those from its linear regression predecessor in nearly every measurable way. First, the neural networks' deterministic projections have an improved fit over its training sample, when compared to the calibration fit of the linear model. Even more encouragingly, the networks within SPIKE2 are proven to generalize well over their validation and test samples at a level that often exceeds the fit of a persistence forecast or a simple linear regression model. Obviously, this exercise with the ANNs is performed entirely with "perfectly forecasted" input parameters from a historical database. Although SPIKE2 generalizes well to out of sample data, it is more likely than not that the ANNs will have a somewhat noticeable drop in deterministic skill when the perfectly predicted input parameters from this chapter are replaced with imperfectly predicted input parameters from operational models in a real time setting. Therefore, the results in this section are most likely to represent the maximum potential skill of the neural network system, given idealistically perfect calibration and forecast data.

4.5 Probabilistic Network Skill

Considering the imperfect nature of operational meteorology, it is of the utmost importance for meteorologists to quantify and communicate their confidence, or lack thereof, when making a forecast. Therefore, the usefulness of a single deterministic projection from SPIKE2 or any other scheme will be somewhat limited in scope by operational standards. In order to properly assess forecast uncertainty, meteorologists utilize solutions of multiple guidance products (i.e. several deterministic products and/or ensemble guidance). Unfortunately, there are few if any other available tools that are designed to measure IKE in Atlantic TCs. Therefore, one of the only

ways to quantify forecast uncertainty in SPIKE2's deterministic projections will be through the probabilistic products that are created from its system of neural networks.

As previously discussed, the probabilistic projections are derived from a series of 100 individual IKE projections from the system of neural networks in SPIKE2 (Figure 4.3). Each of these independent projections represents an estimate of IKE tendency and total IKE, given the same set of environmental input parameters. In short, these individual projections differ only because random variability affects the weights of the neurons within the neural network during their calibration. As such, these probabilistic forecasts are used to quantify the influence of random variability on projections of IKE from SPIKE2. The two main probabilistic products that are discussed in this section include a forecast uncertainty range product meant to complement SPIKE2's deterministic forecast and a probability of exceedance product that can be used to determine the likelihood of a TC reaching a certain IKE threshold.

A) Uncertainty Range Product

The primary goal for the uncertainty range product, as its name suggests, is to project the amount of confidence in SPIKE2's deterministic forecast. As previously discussed, the deterministic forecast is derived from the median or 50th percentile forecast from the system of ANNs. Therefore, the uncertainty range can be bounded by two other percentiles within the ANN system on either side of the median. One possibility would be to use the 75th and 25th percentiles for the upper and lower bounds respectively. In this example, exactly half of the ANNs' projections would always fall inside of uncertainty range. However, these bounds do not reflect the likelihood of the observed IKE value verifying within the uncertainty range. As a result, less than half of the 1995-2011 observed IKE values fall inside of the 25th to 75th percentile range when using the 36-hour probabilistic SPIKE perfect-prognostic projection.

Instead, the uncertainty range is objectively selected to bound a window inside of which two-thirds of the historical total IKE values would correctly verify. Ultimately, just over two-thirds of historical IKE values between 1995 and 2011 fall between the 87th and 13th percentiles of the 36-hour ANN system. Thus, these two percentiles will be chosen to represent the bounds of the SPIKE2 uncertainty range. Importantly, the uncertainty range can be plotted along with the deterministic projection on the same graph to give forecasters a sense of the best-guess and a reasonable spread for total IKE and IKE tendency. This uncertainty range would be smallest at shorter forecast intervals and would expand at large forecast intervals to capture the growing uncertainty for longer range forecasts when the mean IKE tendency values are larger. An example of the uncertainty range and the deterministic forecast working together is shown in Figure 4.7 for Hurricane Irene in 2011 using a series of consecutive 36-hour SPIKE2 projections.

The average uncertainty range for the 36-hour version of SPIKE is approximately 21 TJ. Assuming that positive and negative forecast errors are equally likely, the uncertainty range allows for on average 10.5 TJ of absolute error if the observed IKE value were to verify within the forecasted range. This is on par with the average error found in 36-hour SPIKE2 projections in the training sample (11.4 TJ; Figure 4.4). Of course, the uncertainty range is not fixed to this average 21 TJ range. Instead, it is designed to expand and contract based on spread of the individual ANN projections. In fact, the smallest uncertainty range is only 9.8 TJ from top to bottom, when SPIKE2 projected a range of 20.0-29.8 TJ for Hurricane Danielle in 2010. Importantly, the observation in this high confidence case fell within the uncertainty range at 22.5 TJ. On the other hand, the largest uncertainty range was for Tropical Storm Olga on November 25, 2001. At that time, Olga was a very large tropical storm with tropical storm force winds extending outward of 600nm from the center on the northern half of the storm. As a result, it had a record level of 432.0 TJ of IKE at the time. The uncertainty range for SPIKE2 using the 36-hour neural networks and perfect input parameters is 139.5-398.5 TJ. Therefore, although the

observed quantity fell on the high side of the uncertainty range, at least the product correctly identifies the high uncertainty and the high likelihood for Olga to have extremely high levels of IKE.

The uncertainty range for Olga is not the only time SPIKE2 has exhibited a low bias for extreme events. The previous section discussed at length that SPIKE2's deterministic projections have a low bias for extreme values of IKE. This systematic bias, which is typical of most statistical schemes trained on historical data, also exists within the uncertainty range. In fact, storms that had a total IKE value greater than 100 TJ from 1995-2011 were slightly more likely to be above the upper bound of the uncertainty range (49%) than they were to be within the uncertainty range itself (45%). Therefore, forecasters should exercise caution for very large TCs, understanding that even with the uncertainty range, SPIKE2 is still likely to underestimate the observed level of IKE, assuming that the environmental input parameters are predicted somewhat accurately.

B) Probability of Exceedance Product

Rather than attempting to project an exact value of total IKE, the Probability of Exceedance (PoE) product instead estimates the likelihood of IKE exceeding a certain threshold. Specifically, PoE uses the 100 individual SPIKE2 ANN projections to project the likelihood of a storm exceeding 25, 50, 100, and 200 TJ of IKE. These four levels represent the top 43%, 22%, 8%, and 1% of all storm fixes from 1990-2011 respectively. By focusing on these four thresholds, PoE ostensibly could be more resilient against SPIKE2's systematic low bias on extreme events. In a deterministic forecast, even if SPIKE projects a very high 300 TJ of IKE for a record breaking storm of more than 400 TJ, there would be a substantial error of 100 TJ. However, in the PoE product's case, it would instead simply indicate that a storm would exceed every threshold, which would adequately alert forecasters and emergency managers to the possibility

of a large and powerful TC. Obviously, the low bias would still negatively impact large storms near the 100 and 200 TJ values.

The output of the PoE product will come in the format of a percentage for each of the four exceedance thresholds. For example, the perfect prognostic 36-hour SPIKE2 PoE product projects that Hurricane Ike would have a 100% chance of exceeding 25 TJ, 99% probability of exceeding 50 TJ, 1% chance of exceeding 100 TJ, and a 0% chance of exceeding 200 TJ at 06Z on September 10, 2008, a time when it actually had 67 TJ.

Since the output of PoE is in terms of a percentage and not a quantity of IKE, skill of this probabilistic product cannot be measured using the same methods that were used to evaluate the deterministic products in the previous section. To create one of the evaluation tools, it is first necessary to bin the numerical output of the PoE product into five categories based on their probability. These categories are “remote” (0-25%), “slight” (25-50%), “moderate” (50-70%), “high” (70-90%), and “very high” (90-100%), in order of increasing probability (Table 4.3). Ultimately, these probability bins allow for the results of the PoE product to be binned with other events that have similar likelihoods for their respective thresholds.

These probabilistic bins are used to make a reliability table to visually identify how the PoE product performs on a 36-hour forecast interval (Table 4.4). In short, the reliability table displays the percentage of events in each predicted probability bin that actually exceed the target threshold. For example, the reliability table shows that 97% of TCs exceed 50 TJ when the 36-hour PoE tool gives that a very high probability of that occurring. In a perfect scenario, the percentage in each column would fall inside each probability bin. For example, one would expect something given a low probability of occurrence to actually occur less than 25% of the time as the bin is defined. On the other hand, the PoE tool would have poor performance if TCs that were given a very high probability of exceeding 100 TJ only did so in rare occasions.

For the most part, the reliability table indicates that the PoE product performs quite well. When storms are given a very high probability of exceeding a certain threshold, those TCs

actually exceed that same threshold the vast majority of the time. Likewise, events rarely occur when they are given a remote possibility of occurring. The two major biases of the reliability table occur in the smallest and largest IKE thresholds. For the smallest threshold, TCs actually exceed 25 TJ less than half the time when they are projected to do so 50-70% of the time by PoE. This result suggests that the individual SPIKE2 ANNs have a slight high bias on smaller storms and tend to overpredict the probability of storms exceeding the lower thresholds. On the other hand, 50% of TCs given a slight chance of exceeding 200 TJ end up exceeding that threshold. This is reflecting the low bias in the SPIKE2 model for extreme events. Therefore, even a 30% probability of exceeding 200 TJ should not be taken lightly, as the neural networks will be hesitant to exceed such a high level of IKE that is only seen in the top 1% of observed storm fixes.

In addition to the reliability table, the 36-hour SPIKE2 PoE product is verified using a series of Brier scores (Brier 1950). Brier scores are a quadratic metric that are widely used to evaluate probabilistic skill. The Brier score (B) is defined as,

$$B = \frac{1}{n}\sum_{i=1}^{n}(f_i - o_i)^2 \,, \tag{4.3}$$

wherein f_i stands for each of the forecasted probabilities and o_i stands for each of the observed outcomes (e.g. Wilks 2005). For binary events such as the exceedance events calculated by the SPIKE2 PoE product, the outcomes are set to one when the event occurs (observed IKE exceeds thresholds) and the outcomes are set to zero when they do not occur (observed IKE fails to exceed threshold). A perfect Brier score is equal to zero (i.e. forecasted probability = 100% and event occurs), and the worst Brier score possible is equal to one (i.e. forecasted probability = 0% and event occurs). In addition, Brier scores are often calculated with reference to another Brier score from a baseline forecast such as a repeating climatology forecast (e.g. Mason 2004). This metric is called the Brier skill score (BSS), and it is defined as

$$BSS = 1 - \frac{B}{B_{ref}}, \quad (4.4)$$

wherein B is the Brier score from the forecast product and B_{ref} is the reference Brier score from a fixed climatology forecast (e.g. Wilks 2005). A BSS greater than zero indicates that the forecast product has skill relative to climatology, and a BSS less than zero indicates that the forecast product is worse than climatology.

The Brier scores, climatology Brier scores, and the Brier skill scores for the 36-hour SPIKE2 PoE product are shown in Table 4.5 for each of the four IKE thresholds. Overall, the brier scores for the SPIKE2 PoE product are quite impressive, as B < 0.1 for each of the four thresholds. The brier scores appear to improve with increasing IKE threshold. However, this apparent improvement is an artificial improvement that arises from the rarity of TCs gaining more than 100 or 200 TJ. In reality, it is fairly easy for the the PoE to simply project a 0% probability of exceeding 200 TJ, considering that the mean value of IKE in all storm fixes from 1990 to 2011 is about 25 TJ. Therefore it is unsurprising that climatology forecasts also do quite well for the higher thresholds, as they repetitively offer a low probability forecasts for the 100 and 200 TJ exceedance levels. As such, it is more appropriate to evaluate PoEs skill with the Brier skill scores relative to climatology than with the raw Brier scores themselves.

In each case, the PoE tool indicates superior skill to climatology. In fact, the Brier scores for the PoE product are 65-80% better than the same scores calculated for climatology. Overall, these impressive Brier skill scores support the results found in the reliability table that the PoE product is an effective probabilistic forecasting tool that can be used to alert forecasters and emergency mangers when storms are likely or unlikely to exceed various arbitrary IKE thresholds. This PoE product adds another useful tool to SPIKE2's growing arsenal, which already includes the deterministic forecasts and the uncertainty range discussed above. This added versatility and the improved determinsitic skill of SPIKE2 over the linear model make it quite apparent that the neural network system should replace the linear regression scheme

discussed in the previous chapter as the primary statistical method when adapting this kinetic energy prediction scheme for operational use. As such, the remaining chapters will build upon the neural network proof of concept presented here by exploring the physical relationships that drive SPIKE2 and by finally verifying the skill of SPIKE2 in an operational setting with input parameters from a global reforecast model.

Table 4.1: Three sets of example input and corresponding output for the sample neuron setup shown in Figure 4.1. The bias value is by definition equal to one, but the other input values are free to vary. The weights within the neuron, which are identical in each example, are detailed in Equations 4.1. The propagation function is a weighted sum, and the activation function is a piecewise function, as explained in Equations 4.2.

	Example 1	Example 2	Example 3
x0	1	1	1
x1	0	-1	1
x2	1	-1	1
x3	2	0	-2
Output	**-1**	**-1**	**1**

Table 4.2: Variables used in the perfect prognostic version of the SPIKE2 neural networks. Many of the variables (e.g. RHLO, SHRD) originated from SHIPS and are averaged over specific areas as noted. Others were specifically created from observations for the SPIKE model.

Variable	Definition	Units
PIKE	persistence of IKE	TJ
dIKE12	previous 12hr change of IKE	TJ
VMAX	maximum sustained wind speed	kts
VMPI	Difference between maximum potential intensity and VMAX	kts
LAT	latitude of storm's center	°N
LON	longitude of storm's center	-°W
MSLP	minimum sea level pressure	hPa
PENV	average surface pressure *(averaged from r=200-800km)*	hPa
VORT	850 hPa vorticity *(r=0-1000km)*	10^{-7} s^{-1}
D200	200 hPa divergence *(r=0-1000km)*	10^{-7} s^{-1}
SHRD	850-200 hPa shear magnitude *(r=200-800km)*	kts
SHTD	850-200 hPa shear direction *(r=200-800km)*	°
RHLO	850-700 hPa relative humidity *(r=200-800km)*	%
RHMD	700-500 hPa relative humidity *(r=200-800km)*	%
T150	150 hPa temperatures *(r=200-800km)*	°C
SST	sea surface temperatures	°C
SDAY	time after tropical storm genesis	days
PDAY	time from peak of season (Sept. 10)	days

Table 4.3: Terms used to describe the probability of a certain event occurring given its likelihood from a probabilistic forecast.

Term	Likelihood
Remote	0-25%
Slight	26-50%
Moderate	51-70%
High	71-90%
Very High	91-100%

Table 4.4: Reliability table for probabilistic projections of IKE exceedance levels from the 36-hour SPIKE2 neural network system. The value in each cell represents the number of observed events that verified in each exceedance level, given a certain probabilistic forecast for that same exceedance level. For example, when the 36-hour SPIKE2 system projects that a storm has a very high likelihood of exceeding 50 TJ of IKE, that storm will end up exceeding 50 TJ 97% of the time, as denoted by the rightmost cell in the second row.

	Remote	Slight	Moderate	High	Very High
> 25 TJ	6%	23%	41%	66%	97%
> 50 TJ	2%	35%	50%	67%	97%
> 100 TJ	2%	40%	57%	74%	95%
> 200 TJ	0%	50%	73%	100%	100%

Table 4.5: Brier scores and brier skill scores relative to climatology for the SPIKE2 Probability of Exceedance (PoE) product. A brier score of zero indicates a perfect forecast. A positive (negative) brier skill score indicates an improvement (degradation) to climatology.

	> 25 TJ	> 50 TJ	> 100 TJ	> 200 TJ
SPIKE2 Brier Score	0.085	0.052	0.033	0.004
Climatology Brier Score	0.245	0.218	0.010	0.010
Brier Skill Score	0.651	0.763	0.670	0.650

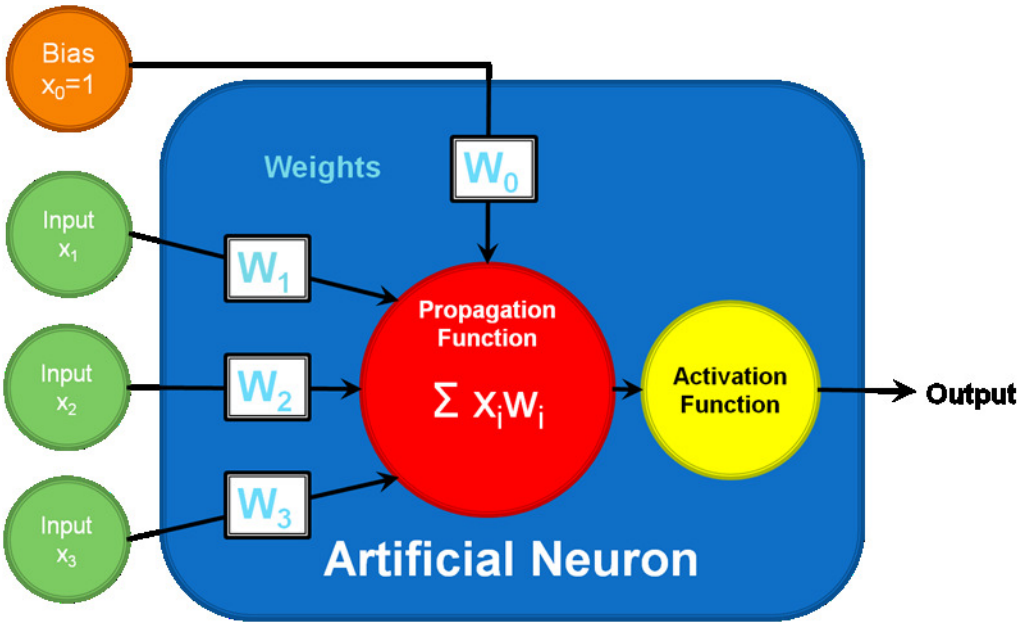

Figure 4.1: Example diagram of a generic artificial neuron and the data that passes through it. The elements within each artificial neuron (weights, propagation function, and activation function) are located within the dark blue polygon in the diagram. An arbitrary number of input neurons (green circles) and a single bias neuron equal to one (orange circle) are passed into the neuron on the left of the diagram. The neuron will process this data and produce a single output on the right.

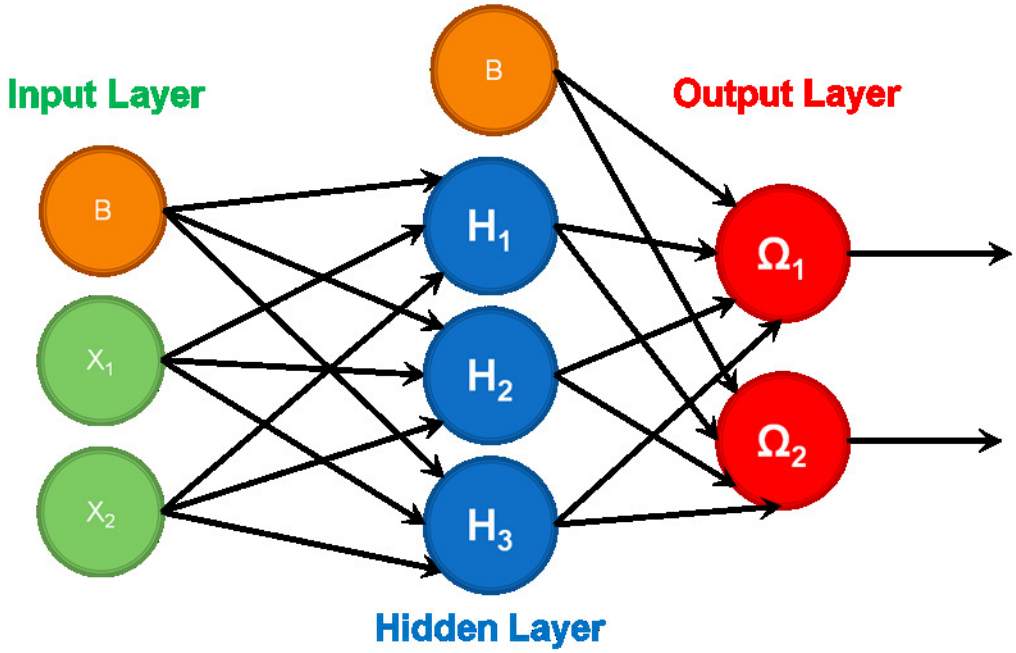

Figure 4.2: Example diagram of a generic two-layer feed-forward artificial neural network hierarchy, with one hidden layer and one output layer. In this example hierarchy, data enters the network from an input layer comprised of two input neurons (green circles) and a single bias neuron (orange circle) on the left. Each hidden neuron processes all of the inputted data much like the sample neuron in Figure 4.1, and then passes on its output along to the output layer. This output layer processes the data from the hidden neurons and a bias neuron to compute the final two outputs of the network on the right.

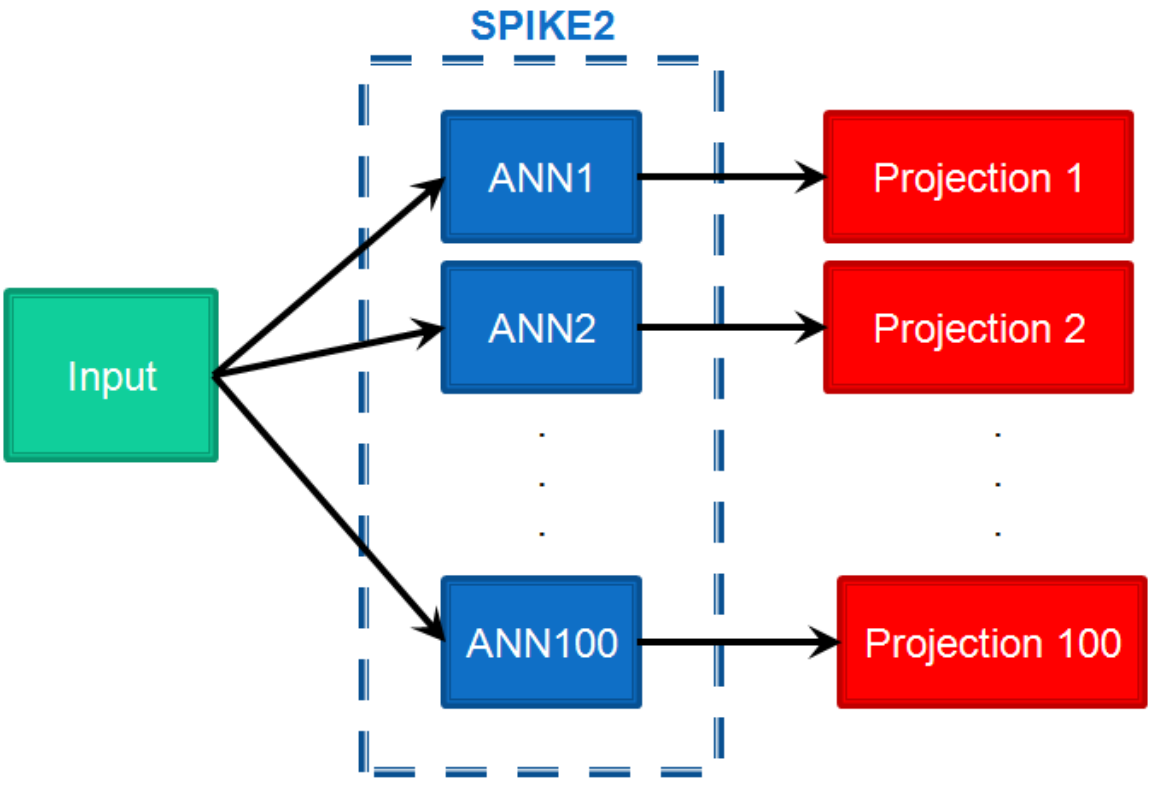

Figure 4.3: Schematic of the SPIKE2 neural network system. A single set of input parameters is passed into each of the one-hundred independent artificial neural networks that make up SPIKE2. Each network produce its own separate projection of IKE tendency based on the same input parameters. The median of these projections is used as a SPIKE2's best deterministic projection, but each individual projection can be used for probabilistic forecasting.

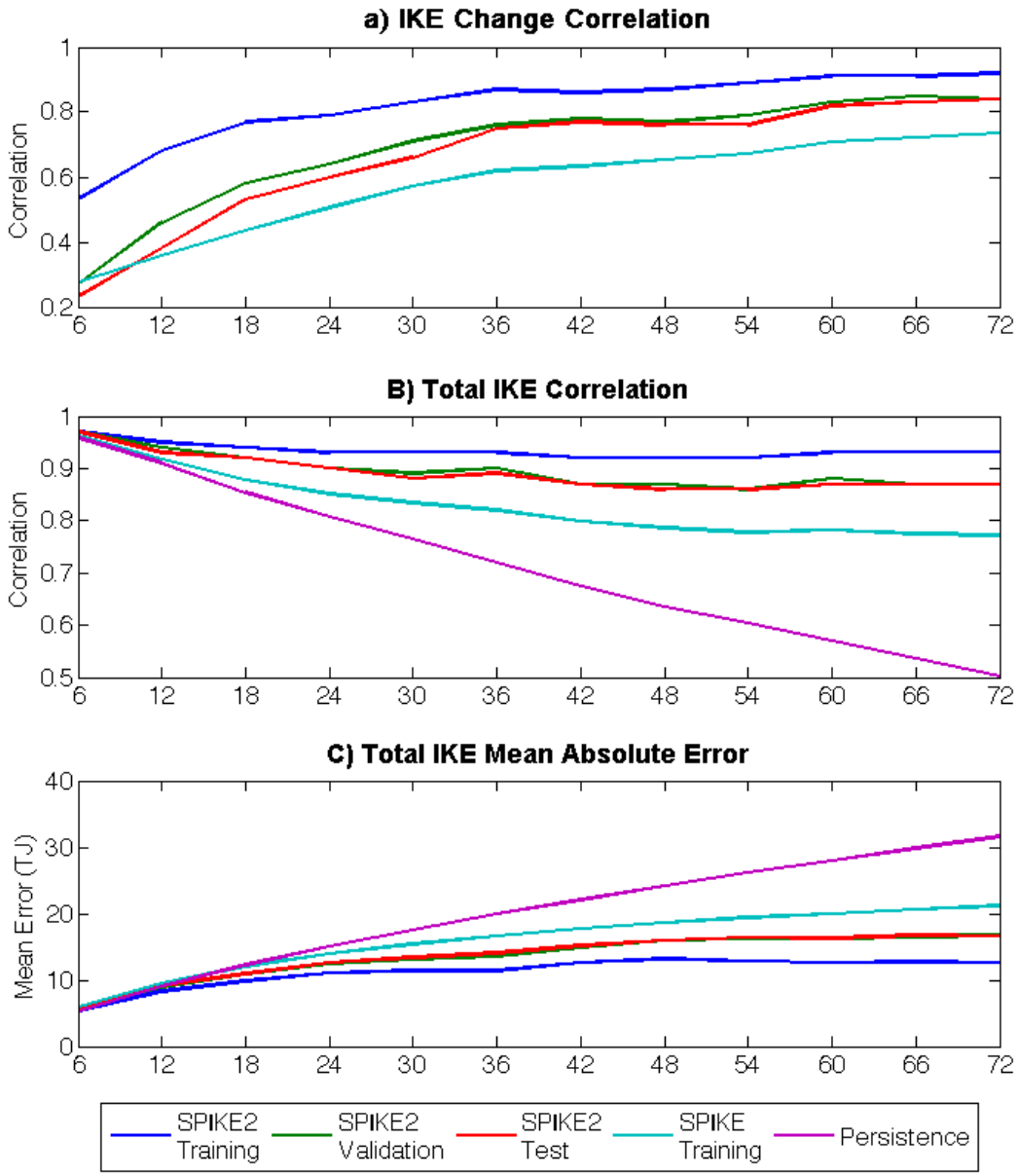

Figure 4.4: Plot of various correlation and error measurements versus forecast hour for deterministic projections from the SPIKE2 neural network scheme. Panel A contains correlations between projections of IKE change and observed IKE change, panel B shows correlations between projections of total IKE and observed total IKE, and panel C includes the mean absolute error between the projections of total IKE and the observations. Each metric is calculated over the training sample (blue lines), validation sample (green line), and test sample (red line) that are used to calibrate the neural networks. The same three metrics from the linear SPIKE regression model over its 1990-2011 training interval (cyan line) are also reproduced from Chapter 3 for comparison to the improved neural networks. Finally, the correlations and error of a persistence forecast (purple line) is also shown to confirm that the IKE prediction models offer an improvement over a simple persistence forecast.

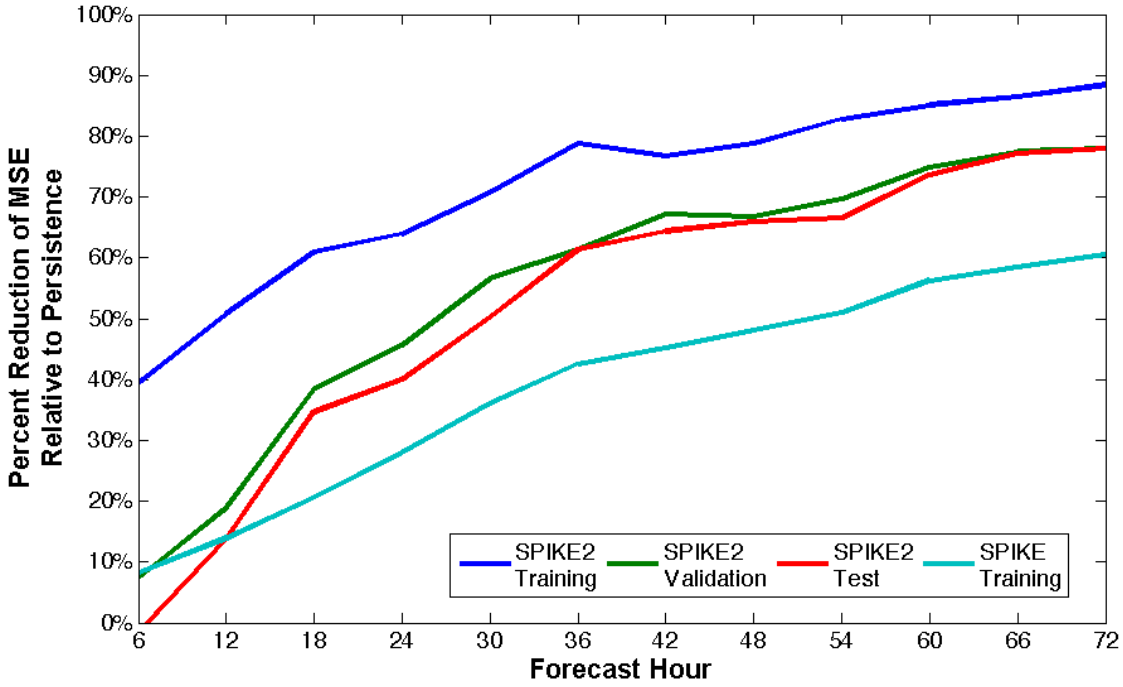

Figure 4.5: Percent reduction of mean squared error (MSE) for SPIKE2's deterministic projections from 1995 to 2011, with respect to a persistence forecast. A reduction of MSE is plotted as a positive percentage, indicating improved model skill. SPIKE2 has significantly lower MSE than persistence at the p=0.05 level for all forecast hours greater than or equal to 12 hours in the training sample, 30 hours in the validation sample, and 30 hours in the test sample. Also plotted is the MSE reduction skill metric of the original SPIKE linear regression model, which is significantly better than persistence at the p=0.05 level for all forecast hours greater than or equal to 24 hours in its training dataset.

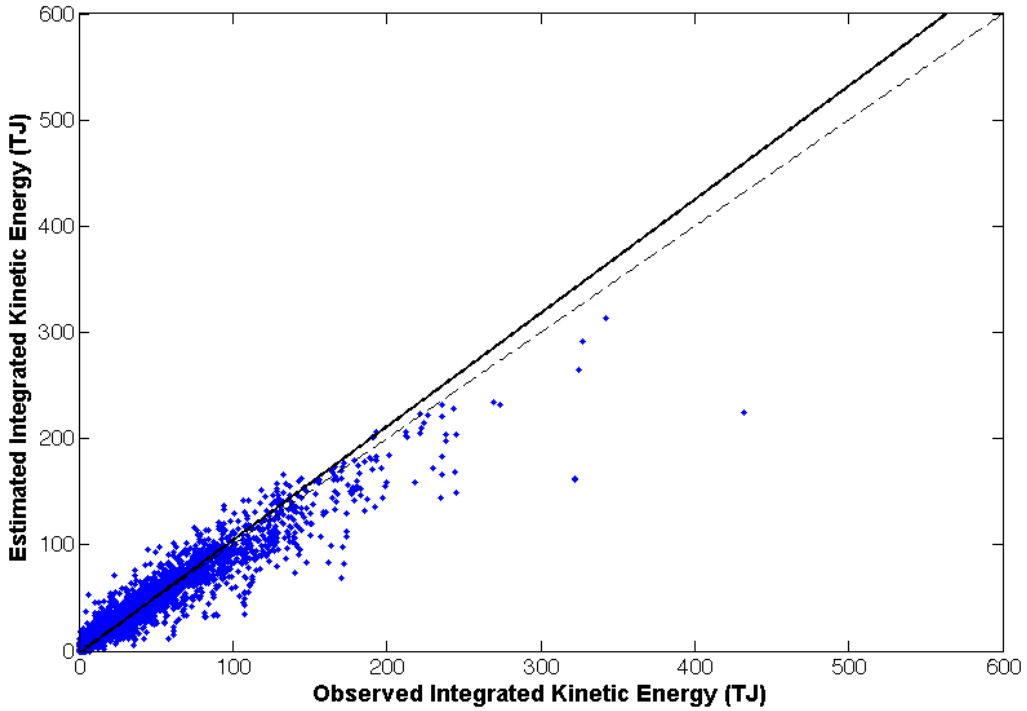

Figure 4.6: Scatter plot of observed total IKE values versus estimated total IKE values at a forecast interval of 36 hours produced by a system of artificial neural networks. The blue dots represent the median forecast for each TC fix from a sample of 100 individual networks. The dark black line represents the best fit line between the 2996 observed and estimated data points. The dashed line represents a perfect forecast (y=x). The correlation between the observed and estimated IKE values is r=0.92, which is significant at the p=0.01 considering the large sample size.

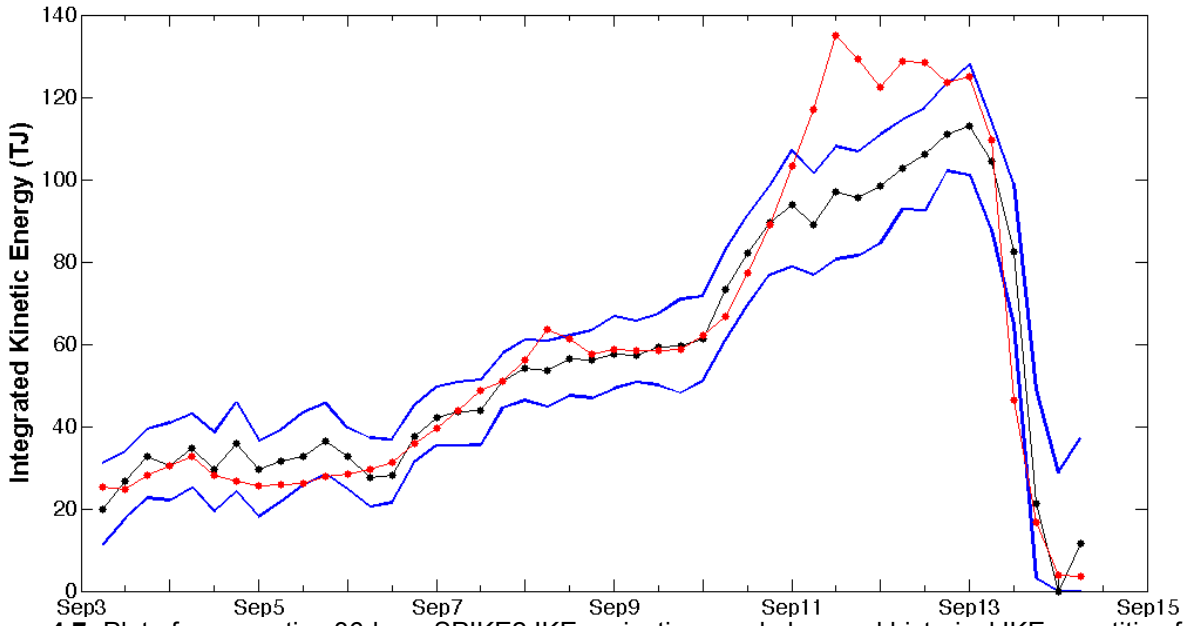

Figure 4.7: Plot of consecutive 36-hour SPIKE2 IKE projections and observed historical IKE quantities for Hurricane Irene in September of 2011. The blue lines represent the uncertainty range generated by the 36-hour probabilistic projections from SPIKE2, the black line represents the deterministic total IKE projection from SPIKE2, and the red dots each represent an observed value of total IKE. The date on the X-axis represents the validation time for each of these consecutive 36-hour projections.

CHAPTER 5
NEURAL NETWORKS SENSITIVITY TESTS

As previously alluded to, this chapter will explore the physical relationships that drive the projections of IKE tendency within the neural networks. In theory, one would expect the neural networks to be governed by many of the same relationships between IKE and the environment that drove the linear regression models (Section 3.4). However, linear regression by definition does not allow a model to properly capture the more complex nonlinear relationships in the storm-environment system. As a result, the ANNs will certainly utilize many different physical relationships that were not resolved by the linear models. Overall, it is important to understand the linear and nonlinear relationships that drive the IKE tendency projections in the neural networks, so the ANNs are not simplified to be black boxes that magically produce a skillful projection of IKE tendency and total IKE.

In the linear regression exercises of Chapter 3, it was possible to isolate the connections between the environment and IKE tendency by simply analyzing the signs and magnitudes of the fixed regression coefficients. However, artificial neural networks contain a complex web of weights and functions that cannot be easily isolated. For instance, each of SPIKE2's neural networks contains hundreds of weights that are distributed across its two layers of neurons. This complex structure makes it nontrivial to analyze the physical relationships between the environment and the storm that are used to predict IKE tendency in the SPIKE2 system of neural networks. In many regards, this can be viewed as a significant weakness of neural networks because a statistical tool is meaningless without an underlying understanding of the physical processes that govern the system.

To alleviate these concerns, and to highlight the nonlinear interactions between the storm and its environment with regards to IKE tendency, the neural networks are put through an exhaustive series of sensitivity tests. More specifically, these tests are performed in an effort to

examine how variability of each input parameter influences the deterministic projections of IKE change produced by SPIKE2. Ultimately, the relationships found within the neural networks themselves will then be compared against our understanding of the physical processes that govern storm structure to ensure that each network relationship can be explained by a known physical process. The first section will describe the methodology of these sensitivity tests by explaining the difference between the control runs and each of the perturbation runs. Afterwards, the remaining sections of the chapter take an in-depth look at the results of each perturbation test, one input parameter at a time.

5.1 Sensitivity Experiment Setup

The sensitivity exercise focuses exclusively on 36-hour deterministic projections of IKE tendency produced by the SPIKE2 neural network system. The exercise is comprised of a single control run and seventy-two perturbation runs, four for each of the eighteen input parameters. In each run, deterministic forecasts are produced by passing variations of the complete observed input dataset through SPIKE2 to produce a projection of IKE tendency for all of the nearly 3000 storm fixes contained within the historical IKE record between 1995 and 2010. Ultimately, this exercise is designed to examine how the SPIKE2's 36-hour deterministic forecasts of IKE tendency change as each input parameter is perturbed.

The control run utilizes input parameters from the unaltered observational record that was used to calibrate the neural networks in the previous chapter. Using the raw observed inputs from each storm fix, SPIKE2 produces projections of IKE tendency for each individual storm fix. Since neither the input variables nor the neural networks have been altered, the control run will include a set of 36-hour projections that are identical in skill to those analyzed in Chapter 4. The control distribution of SPIKE2's projections is shown in Figure 5.1 with a grey shaded curve. Compared to the probability distribution of the historical record of IKE tendency, SPIKE2 tends

to overestimate the likelihood of small positive IKE changes (0 TJ < ΔIKE < + 20 TJ). In addition, SPIKE2 underestimates the probability of negative IKE changes and the probability of large positive IKE changes (ΔIKE > + 40 TJ).

The perturbation runs begin with the same set of historical input parameters that are used in the control run. The sole difference in each of the perturbation runs is that exactly one input parameter per run is perturbed for every storm fix. Specifically, the chosen input parameter is uniformly adjusted up or down by either one or two standard deviations for every storm fix, leaving the other 17 input parameters the same. This yields four different perturbation runs for a single predictor (-2σ, -1σ, +1σ, and +2σ) in an effort to determine how small and large fluctuations of a single input parameter affect the IKE growth in the SPIKE2 neural network system. Ultimately, this process of generating four perturbation runs per input parameter is repeated for each of the eighteen parameters in order to understand the physical relationships in the environment-storm system that govern IKE growth in the SPIKE2 neural networks.

The probability distributions for SPIKE2's deterministic projections of IKE tendency in each perturbation run are shown in Figure 5.2. These plots are useful for determining the nature of the relationships between each predictor and projected IKE tendency. More specifically, an input parameter will have a negative relationship with projected IKE tendency if the positive perturbation runs (red lines) are located to the left of the peaks of distributions for the control run and the negative perturbation runs (blue lines). For instance, as the PIKE parameter is increased, the distribution shifts to the left indicating that SPIKE2 is more likely to project decreasing IKE. On the other hand, an input parameter will have a positive relationship with projected IKE tendency if the peak of the distribution for the positive (negative) perturbation runs are to the right (left) of the peak of the control runs. For example, as VORT increases, the distribution of projections from SPIKE2 shifts towards more positive values of IKE tendency, suggesting that increasing VORT is associated with increasing IKE. The remainder of the discussion in this chapter goes through how the perturbation runs affect each input parameter

individually, focusing on possible physical processes that could explain the fluctuations of SPIKE2's deterministic projections for IKE tendency.

5.2 Persistence Perturbation Tests

The neural networks receive input data from two different persistence predictors, PIKE and dIKE12. The persistence of total IKE variable is one of the more influential input parameters used by the neural networks within SPIKE2. The influence of PIKE on the ANNs is unsurprising considering that this same predictor had the most significant regression coefficients in the linear regression model. As shown in Figure 5.2, when PIKE is perturbed upward (downward), the distribution of SPIKE2 projections are shifted towards more negative (positive) values. The magnitude of this shift can be quantified by calculating the median IKE tendency value within the distributions of each perturbation run and comparing them to the median IKE tendency projection in the control run. The results of such a comparison are included in Table 5.1, where it can be determined that the positive and negative perturbations of PIKE cause the median deterministic projections of IKE tendency to change by roughly -400% and 250% respectively, relative to the control run.

Before continuing with this discussion, it is important to note that the PIKE parameter has a low mean and a high standard deviation during the 1995-2011 training sample. Therefore, a negative perturbation of -2σ is unrealistic for PIKE, as it will require the persistence value of IKE to be impossibly less than zero. Thus, the 250% increase that arises from perturbing PIKE by -2σ compared to the control run is unrealistically high. In reality, a decrease in PIKE will result in a more moderate, yet still positive, shift of projected IKE tendency.

Nonetheless, the results of the sensitivity test are still able to support the findings from Chapters 2.3 and 3.4, wherein larger storms with more IKE are less likely to continue to grow than are smaller developing TCs with less IKE. Once again, this negative relationship between

PIKE and IKE reflects that storms ideally gain IKE for most of their lifetime over the open ocean, even as intensity fluctuates, such that they obtain their maximum value of IKE prior to landfall or the completion of ET (Musgrave et al. 2012). As a result, the SDAY parameter, which measures the time since tropical cyclogenesis has a more important influence over SPIKE2's projection when PIKE is low. If a storm is several days beyond its time of genesis (i.e. SDAY = +2σ), and PIKE is low (i.e. -1σ), then SPIKE2 is less likely to project significant IKE growth when compared to a case in which PIKE is similarly low near the time of its genesis (i.e. SDAY = -1σ).

In addition, this negative relationship between PIKE and IKE tendency is influenced by the fact that a storm with higher values of IKE obviously has more IKE to lose than does a smaller storm. Therefore, a storm with only 20 TJ of IKE at the initialization time can gain a large amount of IKE during the forecast interval, but it cannot lose more than 20 TJ. On the other hand a TC with 100 TJ of IKE can theoretically lose or gain more than 50 TJ of IKE during the forecast interval. As a result, it is of no surprise that changing values of PIKE will affect the variance of SPIKE2's projections. For example, as PIKE increases, the distribution spreads outward from its peak, and the standard deviation for the projections increases. However, as PIKE is perturbed downward, the distribution becomes more narrow and the standard deviation decreases.

On the other hand, the sensitivity exercises with dIKE12 indicate that past changes of IKE have a limited influence on 36-hour projections of IKE tendency from SPIKE2. For example, the differences between the probability distribution of the positive perturbation runs, the negative perturbation runs, and the control run are very subtle (Figure 5.2). The median deterministic projection of IKE tendency within the distribution only significantly responds to large variations of dIKE12 (i.e. ±2σ). This result suggests that if a TC was previously gaining (losing) a large amount of IKE, it will continue to gain (lose) IKE during the subsequent 36-hour interval, providing that the environmental conditions remain somewhat similar. It is likely that this persistence based predictor is more significant for the shorter forecast windows and less

significant for the longer forecast intervals, as was found in analysis of the linear regression SPIKE model.

5.3 Intensity Perturbation Tests

Traditionally, the intensity of a TC is often measured by its maximum wind speed and its minimum central pressure. These two quantities are both used as input parameters (VMAX and MSLP) for the SPIKE2 neural networks. Ultimately, perturbations of either intensity parameter significantly affect projections of IKE tendency by the ANNs in SPIKE2. This relationship does not come as a surprise given the results of the RV experiments in Chapter 2.2. As discussed in that section, a storm with a higher intensity and stronger maximum sustained winds will be expected to have a higher amount of kinetic energy, all else being equal. Therefore, if VMAX is higher at the valid forecast time, it should be expected for the TC to either retain IKE or gain IKE. As a result, the positive perturbation runs for VMAX have a probability distribution that is shifted to the right. Likewise, if MSLP is perturbed downward to indicate a more intense storm, wind speeds and IKE generally increase. Therefore, the distribution of the negative MSLP perturbation runs also shift to the right. Analysis of the median values in both the VMAX and MSLP distributions for the small and large perturbation runs confirm that the shift towards positive IKE tendency values is significant at the 99% level when the intensity of a TC increases (Table 5.1).

On the other hand, if a TC is forecasted to have a weaker intensity (lower VMAX, higher MSLP), the distribution is shifted towards the left, indicating that weakening storms are less likely to gain IKE. Interestingly, the negative VMAX perturbation runs have a very differently shaped distribution when compared to the distributions of the positive MSLP perturbation runs. For instance, the negative VMAX perturbation runs have a more broad distribution than the control run and any other VMAX or MSLP perturbation run. Therefore, a low intensity in terms of

VMAX at the valid forecast time does not prevent SPIKE2 from projecting IKE to increase if other environmental parameters are favorable. In contrast, the distribution for the positive MSLP perturbation runs is shaped very similarly to the control run, albeit shifted to the left. This makes it less likely for TCs to gain an appreciable amount of IKE when its MSLP increases.

The differences between the negative VMAX perturbation runs' distribution and the positive MSLP perturbation runs' distribution arise because VMAX does not have a truly linear relationship with MSLP, despite the fact both metrics are used to quantify TC intensity. For example, in the historical record, there are plenty of storms that have anomalously low minimum pressure levels, despite modest maximum wind speeds (e.g. Hurricane Sandy). Likewise, there are plenty of TCs that have modest values of VMAX (~ 50-70 kts) and anomalously high values of IKE (> 100 TJ). As a result, it is not a surprise that the negative VMAX perturbation runs have a broad distribution.

Many of the instances in which storms have high IKE and low VMAX occur in the upper latitudes of the TNA basin (e.g. Figure 2.4) towards the end of a TCs lifetime as they begin to interact with troughs and/or begin ET. In many of these cases, excluding warm seclusion, a TC will have a lesser intensity in terms of VMAX, but its MSLP may remain relatively low, which explains why the positive MSLP perturbation runs do not have a similarly broad probability distribution. For example, Table 5.2 includes a list of TCs that have high IKE values (> 100 TJ), maximum sustained winds that do not exceed 65 kts, and still relatively low minimum central pressure (< 980 hPa). In each case, these storm fixes occur at the end of a storm's lifecycle prior to the completion of ET in the upper latitudes of the TNA basin.

To test whether or not the neural networks can project increasing IKE for storms under similar conditions, an additional small sensitivity experiment is performed, wherein SPIKE2 is given an experimental set of input parameters containing high LAT and SDAY (+1σ), low VMAX and MSLP (-1σ), and mean states (=0) for the other 14 parameters. Encouragingly, SPIKE2 projects that a TC will have more storm growth under these conditions than it would for a control

test with mean states for all 18 input parameters. Furthermore, the significant increased IKE growth compared to the control mean state would not occur if any one of the following four changes were made to the experimental input setup:

1. The TC was moved out of the high latitudes (LAT changed from 1 to 0)
2. The conditions occur early in a TC's lifecycle (SDAY changed from 1 to -1)
3. It's maximum sustained winds are prohibitively low (VMAX changed from -1 to -2)
4. It's minimum pressure is not below average (MSLP changed from -1 to 0)

Overall, the sensitivity tests with the intensity parameters align well with the understood physical relationship between intensity and IKE. Unlike the linear regression models, the neural networks respond to nonlinear signals in a combination of different predictors allowing for SPIKE2 to properly adapt when making its projection for future IKE change.

5.4 Position Perturbation Tests

The two position variables considered as input parameters for SPIKE2 are unsurprisingly the latitude and longitude coordinates for the center of the TC. The linear regression coefficients in the first SPIKE model, suggest that IKE is more likely to increase as storms move into the higher latitudes, as they are more likely to expand from extratropical transition and other baroclinic influences. Indeed, SPIKE2 also suggests that increasing latitude will result in IKE growth more often than not. The positive LAT perturbation runs have distributions that are skewed to the right when compared to the control run (Figure 5.2). As a result, the median of the distribution for the +1σ and +2σ LAT perturbation tests is significantly greater than that of the control run at the 99% confidence level based on simple two-sample bootstrapping tests (Table 5.1). Interestingly, the distribution is not shifted quite as far away from the control run

when compared with the MSLP, VMAX, and PIKE perturbation runs. This is not terribly surprising, considering that environmental conditions will have a greater influence on the evolution of a TC than will geographical positions in nearly all cases. As a result, the distribution of the positive LAT perturbation tests is somewhat broader than the control run. This ultimately allows for a storm to have decreasing IKE when entering the higher latitudes if the storm is not likely to undergo expansion from extratropical transition or trough interactions. In that regard, if a TC is over the high latitudes (LAT = +1σ), with cold oceans (SSTs = -2σ), with moderately low intensity (MSLP = +1σ; VMAX = -1σ) and its IKE has already begun to fall (dIKE = -2σ), SPIKE2 will project IKE to continue falling.

The distribution of the negative latitude perturbation tests are shifted slightly towards negative values of projected IKE tendency. In addition, the negative LAT perturbation runs have a sharper distribution, such that extreme IKE growth and decay is less likely. Ultimately, this lines up well with the historical record, wherein the largest increases of IKE typically occur during ET over the higher latitudes. Likewise, significant decreases of IKE caused by United States and Canadian landfalls only occur north of the Tropic of Cancer. Overall, IKE can and does vary substantially over the lower latitudes, but it typically takes longer periods of time for a TCs wind field to grow as a result of traditional TC development methods than it does during the often rapid expansion that occurs during either trough interactions and/or ET.

Longitude's influence over SPIKE2's projections is less clear cut than the obvious connection between LAT and IKE. For example, the distributions for the positive and negative perturbation runs do not shift as much away from the control run as do the LAT perturbation runs (Figure 5.2). Nonetheless, the negative LON perturbations are shifted and skewed slightly to the left, and the positive LON perturbations are shifted and skewed slightly to the right. This causes the median of the distribution to increase with increasing longitude (Table 5.1). In this case, increasing longitude is equivalent to eastward movement. As a result, increasing IKE

becomes more likely in the eastern part of the TNA basin when compared to the western half of the basin.

The physical meaning behind this seemingly positive relationship between LON and IKE tendency likely stems from the fact that TCs generally move east to west across the TNA basin as they follow the persistent steering flow surrounding the subtropical high. Therefore, TCs are more likely to be mature in the western half of the TNA basin than they are in the eastern half of the basin. More mature TCs generally have accumulated more IKE over their lifetime than younger TCs closer to their time of genesis (e.g. Maclay et al. 2008). As a result of these increased levels of IKE, mature TCs have more kinetic energy that can be lost. One primary cause for a significant loss of IKE is landfall, which notably occurs primarily on the western edge of the basin. Ultimately, the increased presence of land and the higher values of IKE in western Atlantic storms are likely the cause for the increased probability of IKE decay in the negative LON perturbations. That being said, it is possible for storms to have increasing IKE in the western half of the basin. SSTs are generally more favorable in the western half of the basin, particularly in the Gulf of Mexico. As a result, storms such as Ike in 2008 and Katrina in 2005 grew substantially in the western half of the basin. Likewise, many storms undergo ET and expand near the Eastern Seaboard of the United States as they recurve to the north and eventually east. This could explain why the distribution of the negative LON perturbation runs is not completely shifted to the left like the negative VORT perturbation runs.

In contrast to western Atlantic TCs, storms in the eastern half of the basin are typically less mature, as a result they more likely to have slowly increasing IKE early in their lifetime. Likewise, there are fewer mechanisms to cause storm decay in the eastern half of the basin, provided that a storm is able to reach genesis in the eastern Atlantic during the peak of the season. On most occasions, storms are able to move freely without the threat of landfall. Obviously, inhibiting mechanisms like Saharan dry air layers (e.g. Dunion and Velden 2004) originate in the eastern half of the basin, and SSTs are often cooler in the deeper waters of the

east and central Atlantic when compared to the Caribbean Sea and the continental shelf near the North American coastline. Therefore, storms can and do decay in the eastern Atlantic as well when the environment is unfavorable for strengthening. Again, this may be the underlying reason that explains why the positive shift in the positive LON perturbation runs are somewhat subtle compared to the shifts caused by perturbations of the environmental parameters.

5.5 Upper-level Divergence and Low-level Vorticity Perturbation Tests

Dynamically, for a tropical cyclone to develop and grow, it must obtain cyclonic low-level vorticity and upper-level divergence. Low-level vorticity is tied to the primary closed circulation of the storm, wherein a large area of enhanced positive vorticity is conducive for TC development. Likewise, divergence aloft is a characteristic of a TC's outflow that is required for the storm to maintain its low central pressure. Furthermore, both upper-level divergence and low-level convergence are connected to the AM transport around the storm. Therefore, it is not surprising that the linear SPIKE model previously discussed in Chapter 3 had positive coefficients for both D200 and VORT. For the most part, this positive relationship is also present in the ANNs for SPIKE2.

The sensitivity tests suggest that D200 does have a slightly positive relationship with projected IKE tendency in SPIKE2. However, this relationship between IKE and D200 in the neural networks does not appear to be as significant as the relationships between IKE and the previously discussed persistence, intensity, or location based parameters. The probability distribution of SPIKE2 projections in the positive D200 perturbations runs is shifted to the right of the control distribution (Figure 5.2), suggesting storm growth is more favorable when high amounts of upper level divergence are present and the resulting upper-level export of lower AM parcels is enhanced. As a result, the median deterministic SPIKE2 projection of IKE tendency is higher in these positive perturbation runs than it is in the control run by about 10% in the +1σ

perturbation run and by more than 25% in the +2σ perturbation run (Table 5.1). However, two-sample bootstrapping exercises indicate that only the larger +2σ D200 perturbation run causes a shift in the median SPIKE2 projection that can be considered significant on a two-sided 95% confidence interval.

Unlike the positive D200 perturbation runs, the negative D200 perturbation runs do not appear to shift the probability distribution of the SPIKE2 projections left or right of the control run. Instead, the negative D200 perturbation runs have a sharper distribution of projected IKE tendency, with a higher peak probability at the center of its distribution. In general, it also appears that storm growth is somewhat less likely in the negative D200 perturbation runs than it is in the control run, but that difference is less noticeable than the change in the sharpness of the negative perturbation runs' distributions. As a result, the median SPIKE2 projection only decreases with respect to the control run by around 5% for both the -1σ and the -2σ D200 perturbation runs (neither of those changes are significant at even the two-sided 90% confidence level). Overall, upper level divergence does play some role in determining the likelihood for IKE growth versus the likelihood for IKE decay, but as was the case with the linear SPIKE model, D200's influence over the neural networks is somewhat limited.

On the other hand, the sensitivity tests infer that the low-level vorticity parameter is one of the more significant input parameters for the SPIKE2 neural network system. Overall, VORT has a positive relationship with projected IKE tendency. Positive perturbations of VORT shift the distribution towards more positive SPIKE2 projections, and negative VORT perturbations shift the distribution of IKE changes towards more negative SPIKE2 projections (Figure 5.2). As a result, storm growth is projected to occur more frequently when VORT increases, and storm decay becomes more likely when a storm has either a small or weak vorticity signature.

The shift in the distribution for each VORT perturbation run causes the median of SPIKE2's projections to shift away from the median in the control run by a significant margin at even the 99% confidence level (Table 5.1). For instance, the +2σ VORT perturbation run

causes the median SPIKE2 projection to increase by nearly 100% relative to the control run. Ultimately, these results suggest that VORT is the most influential large-scale environmental input parameter for the neural networks. In fact, only the PIKE perturbation runs and the intensity perturbation runs cause the median SPIKE2 forecast to change more relative to the median of the control run. Overall, it should not be surprising that VORT is a driving force in SPIKE2. First, vorticity is directly tied to the strength of a TC's axisymmetric cyclonic wind field, and as a result it will be connected to the amount of low-level cyclonic RAM contained within the TC. Furthermore, low-level vorticity likely has a strong positive connection with the low-level convergence responsible for drawing higher AM into the storm.

5.6 Relative Humidity Perturbation Tests

There are two relative humidity (RH) input parameters used by the SPIKE2 neural network system. The first parameter (RHLO) measures low-level RH between 850 and 700 hPa, and the second parameter (RHMD) measures mid-tropospheric RH between 700 and 500 hPa. As one might expect, these two RH fields are similar to one another, such that the correlation between RHLO and RHMD is r=0.82. Based on the high correspondence between the two RH fields, one might expect RHLO and RHMD to have a very similar relationship with IKE tendency, wherein the sensitivity tests for these two parameters would produce very similar results (i.e. both RHMD and RHMD have the same sign relationship with IKE tendency in SPIKE2).

However, the results of sensitivity tests for RHLO and RHMD could not be more different from one another. As shown in Figure 5.2, a positive perturbation of RHLO tends to make IKE decay more likely, but a positive perturbation of RHMD tends to make IKE growth more likely. Conversely, a negative perturbation of RHLO tends to make IKE decay more likely, yet a negative perturbation of RHMD tends to make IKE growth more likely. Overall, these plots indicate that RHLO has a negative relationship with IKE tendency and RHMD has a positive

relationship with IKE tendency. Both of these relationships appear to be significant at the 95% level based on the result in Table 5.1, suggesting that the different signs of the relationships between the two RH parameters and IKE tendency are not accidental. Ultimately, the stark differences between the RHLO perturbation runs and the RHMD perturbation runs make the two RH sensitivity tests challenging to interpret.

Past work has indicated that TC size is positively correlated with tropospheric RH. For instance, Hill and Lackmann (2008) argued that higher environmental RH promotes further extension of the outer spiral rainbands, which would in turn help to expand the TC wind field. Similarly, Wang (2009) connected the diabatic heating within the outer spiral rainbands to a decrease of surface pressure below. This decrease in surface pressure outside of the RMW would decrease the horizontal pressure gradient across the eye, and in doing so would force the inner core to expand. Conversely, Wang found that large TCs are unlikely to exist in low RH environments because there would be enhanced evaporational cooling and limited condensational heating. Ultimately, the positive relationship between TC size and tropospheric RH suggests that IKE tendency should be positively correlated with RHLO and RHMD.

On the other hand, studies have noted that a larger system of rain bands could be detrimental to the inner core strength and intensity of TCs. For example, the process that causes the inner core to expand, as detailed by Wang (2009), also causes deintensification by weakening the horizontal pressure gradient within the inner core. Furthermore, Kimball (2006) noted that the formation of outer rain bands in high RH environments would lead to the presence of convective downdrafts in the outer core. The downdrafts would in turn force cooler air with lower equivalent potential temperatures to the surface. Eventually, that sinking cooler air would be transported into the storm's inner core by the secondary circulation and entrained into the eyewall updrafts. These arguments relating higher RH to weaker intensity and strength would then suggest that RHLO and RHMD should be negatively correlated with IKE.

Further complicating the IKE-RH relationship are the baroclinic and ET processes known to affect storm growth. TCs undergoing ET often develop a dry air slot within the storm circulation (Jones et al. 2003), thus causing a RH in the hurricane environment to drop. These dry air intrusions would ultimately imply a negative relationship between IKE tendency and RH, given the propensity of IKE to rapidly increase during the ET process. On the other hand, RH is climatologically higher in the midlatitudes than it is in the tropics, simply because temperatures are lower closer to the poles. Since ET and trough interactions are favored in the higher latitudes of the TNA, this would suggest that wind field expansion from these processes could occur in cooler, higher RH environments. Based on these mixed signals, it is difficult to determine how ET and trough interactions affect the RH-IKE relationship.

Overall, the research reviewed in this section suggests that there is a complex relationship between environmental RH, spiral rain bands, storm size, storm strength, and storm intensity. Storm size and outer core strength may benefit from higher RH, but inner core strength and intensity may be negatively affected by the same moist environments. Since storm size, strength, and intensity all contribute to IKE, the relationship between RH and IKE becomes very complicated. Certainly in some cases, moist environments will be beneficial for development and IKE growth, and in other instances, high RH might be detrimental to intensification and IKE growth. Regardless, it is likely that if the environment is too prohibitively dry (i.e. landfall or Saharan dust plume) TCs of all sizes will be unable to sustain themselves leading to a drop in IKE in all cases (e.g. Kimball 2006; Dunion and Velden 2004).

Ultimately, it is not surprising that RHMD has a positive relationship with IKE tendency given the positive relationship between RH and storm size. Likewise, it is not surprising that RHLO has a negative relationship with IKE tendency, given the somewhat negative relationship between RH and intensity and the increased likelihood of dry air intrusions existing within storms undergoing ET. On the other hand, it is very surprising that the relationships for RHLO and RHMD are not of the same sign, given the correspondence between the two fields. Despite

the mixed RH-IKE signals reviewed in this section, it is puzzling that RH between 850 and 700mb could interact one way with IKE tendency and RH immediately above that layer could interact in the exact opposite manner. Therefore, future work focused on the relationship between tropospheric RH and IKE tendency is clearly needed to determine whether or not this dichotomy is reflective of real world processes or is just an artificial consequence of our ANNs.

5.7 Deep-layer Vertical Wind Shear Perturbation Tests

The SPIKE2 neural networks utilize two input parameters that measure characteristics of environmental deep layer vertical wind shear. The first of those parameters, SHRD, measures the magnitude of vertical wind shear over the layer between 850 hPa and 200 hPa, without removing the TC vortex. It is well understood, that increasing vertical wind shear is generally harmful to the genesis and development of tropical cyclones in the Atlantic basin and around the world (e.g. Gray 1968; McBride 1995). Therefore, one might expect wind shear to have a negative relationship with IKE tendency, wherein IKE increases as shear decreases. However, as discussed in Chapters 2 and 3, increased vertical shear is found in the upper-latitude regions where TCs gain IKE from ET and trough interactions. Therefore, IKE tendency's relationship with SHRD is clearly nonlinear because IKE can increase in both moderately high shear and low shear conditions depending on the other environmental characteristics.

In general, the positive SHRD perturbation tests indicate that as shear increases, the probability of IKE growth as projected by SPIKE2 increases slightly (Figure 5.2). The median SPIKE2 projection increases by 11% and 31% in the +1σ and +2σ SHRD perturbation tests respectively when compared to the control run (Table 5.1). This increase is found to be significant at the two-sided 95% confidence level by a two-sample bootstrapping test. Despite the fact that increased shear is generally harmful to TC development, this result does not come as a surprise considering the results of the positive SHRD regression coefficients in the linear

SPIKE model. However, unlike the linear model, the ANNs used in SPIKE2 are capable of associating increasing IKE tendency with decreasing wind shear. To ensure that the ANNs are capable of resolving these nonlinear signals, SPIKE2 is given data for an additional sample TC that is located in the low latitudes of the TNA basin (LAT=-1), with warm ocean temperatures (SST=1), high tropopause heights (T150=-1), and mean-state conditions for all other input parameters except for SHRD which is allowed to vary. In this environment, TC growth is governed by traditional TC development mechanisms, and IKE tendency has a negative relationship with SHRD, wherein SPIKE2 projects more IKE growth when SHRD is negative than it does if SHRD is positive.

Interestingly, the negative SHRD perturbation runs do not seem to have a significant impact on the distribution of SPIKE2 IKE tendencies. One of the more noticeable changes is that larger IKE increases are less likely than they are when shear is perturbed upward (Figure 5.2). This is not surprising since many of the largest increases in IKE are associated with ET or trough interactions which occur in more shear environments (e.g. Maclay et al. 2008). Nonetheless, the shift in the median IKE tendency value for these negative SHRD perturbation run distributions is not significant at the 95% interval, probably because all of the nonlinear signals between IKE and SHRD

The final wind shear related input parameter is SHTD, which measures the direction of the 850-200 hPa shear vector. A negative normalized SHTD value will indicate that the environment contains mostly easterly shear, and a positive normalized SHTD value will indicate that the environment has a more westerly to northwesterly shear vector. Past work has determined that easterly vertical wind shear is preferential to westerly shear for TC genesis and development (e.g. Tuleya and Kurihara 1981; Lee 1989; Nolan and McGauley 2012). Therefore, it is of no surprise that as SHTD is perturbed lower, and the shear vector becomes more easterly that there is a slight shift in the distribution of SPIKE2 projections (Figure 5.2). Analysis of the median SPIKE2 projection indicates that the negative SHTD perturbation runs make

positive IKE tendency events significantly more likely. In contrast, the positive SHTD perturbation runs do not seem to have a large effect on SPIKE2's projections.

5.8 Sea Surface Temperature Perturbation Tests

The relationship between sea surface temperature and IKE tendency is rather complex. Traditional theories suggest that warm SSTs, typically greater than a threshold of about 26°C, are required for TC genesis and development (i.e. Gray 1968). Warmer SSTs lead to an increased maximum potential intensity all else being equal (Emanuel 1988), and if the ocean is much warmer than the 26°C threshold for genesis (i.e. > 28°C), then rapid development and intensification becomes significantly more likely (Kaplan and DeMaria 2003). Taking all of this into account suggests that IKE tendency should have a positive relationship with SST, similar to the relationship between intensity tendency and SST. However, as previously discussed, IKE also tends to increase over cooler SSTs in the northern latitudes of the TNA basin, where ET and trough interactions are more likely. As a result, IKE can increase over both cooler and warmer SSTs depending on the other dynamical processes that are driving storm growth. Thus, the relationship between IKE tendency and SST is clearly nonlinear, and it is not surprising that SST was not used as a predictor for the linear SPIKE regression model in Chapter 3.

Unlike the linear models, the neural networks used in SPIKE2 should be capable of teasing apart the nonlinear signals between SST and IKE tendency. To test this hypothesis, SPIKE2 is run for a hypothetical TC in two very different test environments. The first test environment is set to be in the low latitudes of the TNA basin (LAT = -1) with low vertical shear (SHRD = -1), high tropopause heights (T150 = -1), and mean conditions for all other input parameters except for SST. The second test environment is located in the higher latitudes (LAT=1) with opposite high vertical shear (SHRD = 1), low tropopause heights (T150 = 1), and mean conditions for all other input parameters except for SST. Overall, it is expected that higher

SSTs would be more favorable for IKE growth in the low-latitude case based on traditional development theories, but lower SSTs could be more favorable in the higher-latitude cases as baroclinic influences take over.

Encouragingly, in the specified low-latitude environment, 36-hour IKE tendency forecasts from SPIKE2 have a positive relationship with the varying SST input parameter. The SPIKE2 neural network system projects that the test TC will lose IKE if the normalized SST parameter is set to negative one in this environment. However, if the normalized SST parameter is set to one or even two, the neural network projects gradual IKE growth that increases in magnitude with increasing SST. On the other hand, the SST input parameter has a general negative relationship with IKE tendency in the high-latitude test environment. As such, IKE grows more rapidly in cooler SSTs than it does in warmer SSTs. However, it should be noted that IKE increases regardless of the sign of SST because of the ET and baroclinic processes that typically cause storm growth in the higher latitudes.

In a more general sense, the same positive and negative perturbation tests were performed for the SST parameter in each of the nearly 3000 storm fixes that were used to train the SPIKE2 neural networks (Figure 5.2). First, the negative SST perturbation tests have a sharper distribution of SPIKE2 IKE tendency projections, and a slight shift towards negative values. However, it is important to note that the cooler SSTs do not make storm growth impossible when conditions suggest ET and trough interactions are likely. Overall the negative SST perturbation runs results in a significant negative shift in the mean IKE tendency forecast in SPIKE2 (Table 5.1).

In contrast, the positive SST perturbation runs have a broader distribution of IKE tendency. For instance, there is an increased probability of SPIKE2 projecting a positive IKE tendency event in the distribution, which is likely associated with lower latitude storms under favorable conditions for TC intensification. The probability of SPIKE2 projecting negative IKE tendency events also becomes more likely in the positive SST perturbation runs. Ultimately,

such a result indicates that that SPIKE2 must complement thermodynamic predictors like SST with dynamic predictors like SHRD, VORT, D200, etc to determine whether or not storm growth is likely or not.

5.9 Time Perturbation Tests

Two of the input parameters for SPIKE2 are simply related to time. Although many environmental conditions follow a somewhat regular seasonal cycle, these two parameters are generally some of the least significant parameters when compared to the parameters that directly measure the storm or its environment. The first of these time parameters, PDAY, measures the time of year when a TC occurs relative to the peak of the season, wherein a negative normalized PDAY is close to mid-September and a large positive value occurs near the beginning or end of the Atlantic Hurricane Season. As discussed in Chapter 3, a small PDAY value is typically associated with the year's most favorable environmental conditions for TC development. As such, the linear SPIKE model had a negative regression coefficient for the PDAY predictor. Unsurprisingly, SPIKE2 also has a slightly negative relationship with PDAY, wherein the negative PDAY perturbation tests have SPIKE2 distributions that are shifted slightly towards more positive IKE tendency values (Figure 5.2). This negative shift in the distribution results in a significant shift of the median SPIKE2 projection for both the -1σ and the -2σ PDAY perturbation runs. In comparison, neither of the positive perturbation runs affects SPIKE2's projections all that much. As a result, it can be concluded that it is more advantageous to analyze the actual environmental conditions during the beginning and ends of the Atlantic Hurricane Season when conditions are less uniform across the basin year to year.

The second time parameter, SDAY, measures the length of a TC's life from its initial genesis. A large SDAY value indicates that a storm has existed for a long period of time, and a negative normalized SDAY parameter indicates that a storm has just recently reached tropical

storm intensity. As discussed in Section 5.2, TCs are more likely to gain IKE early in their lifecycle as they mature, and they are more likely to lose IKE late in their lifecycle as TCs begin to dissipate or make landfall. Overall, the positive SDAY perturbation runs indicate that a storm is slightly more likely to lose IKE late in its lifecycle. However, the negative SDAY perturbation runs do not seem to affect the distribution of SPIKE2 IKE tendency projections all that much. As a result, it can be concluded that SDAY is arguably the least significant individual input parameter used by SPIKE2.

5.10 Potential Intensity Perturbation Tests

The VMPI input parameter used by the SPIKE2 neural network system is related to a TC's maximum potential intensity (MPI). As its name suggests, MPI measures the maximum sustainable wind speed in a TC based on the thermodynamic properties of the storm environment (e.g. Emanuel 1988; Bister and Emanuel 1998). In nearly all cases, VMAX fails to reach this theoretical intensity upper limit because of other processes that limit storm intensification, such as vertical wind shear. Therefore, VMPI quantifies the difference between the storms actual intensity and its maximum potential intensity, wherein,

$$(5.1) \quad VMPI = MPI - VMAX.$$

Based upon this definition, a larger VMPI value suggests that a storm has greater room to intensify based upon the environments thermal properties. On the other hand, a smaller VMPI value suggests that a storm is approaching its MPI and soon will not be able to sustain any further intensification. Knowing whether or not a storm has the potential to intensify is important information when determining whether or not a storm will gain or lose kinetic energy.

Past work by Maclay et al. (2008) found that the sign of a TC's IKE tendency is influenced by the storm's concurrent VMAX tendency. As discussed previously, one would expect intensification to cause a modest increase in IKE. However, since intensity changes occur over

the smaller inner core region of a TC, intensification is not the sole driving factor for IKE tendency all else being equal. Instead, IKE tendency is more influenced by changes in the size or overall strength of the entire wind field, such that IKE will increase more drastically during intensification if the TC is also expanding at the same time. Therefore, Maclay et al. (2008) concluded that most TCs do not intensify and gain significant amounts of IKE simultaneously when conditions are thought to be conducive for traditional TC development (low SHRD and warm SSTs). Instead, TCs traditionally either gain IKE or increase VMAX, but not both under these conditions. The rare instances of simultaneous intensification and IKE growth most commonly occur when TCs are interacting with troughs or are undergoing ET in very different environments.

Therefore, the difference between VMAX and MPI could become very useful for diagnosing whether or not a storm will gain IKE when it is located over a favorable lower-latitude environment with high SSTs and low shear. For example, it is likely that a storm will grow in size and gain IKE in these conditions if VMPI is small and the TC has reached its theoretical intensity limit. However, if VMPI is high and the TC has favorable dynamical conditions, VMAX might be more likely to increase towards its MPI, making it less likely that the storm will utilize the favorable conditions for storm growth.

Encouragingly, the SPIKE2 neural network system captures this relationship between VMPI and changes of IKE. A negative perturbation results in a positive shift of the SPIKE2 probability distribution (Figure 5.2). As a result, when a storm approaches its MPI, it becomes more likely for that storm to gain IKE, all else being equal. The median of the distribution increases by more than 35% relative to the control run for both the -1σ and -2σ VMPI perturbation runs, both of which are significant at the 99% significance level. Interestingly, the positive perturbation runs do not affect the distribution of SPIKE2 projections nearly as much as the positive perturbation runs did. Visually, there is not a huge difference between the control distribution and the distribution of either positive perturbation run in Figure 5.2. Unsurprisingly,

the changes in the median SPIKE2 projection from the positive perturbation runs are less significant such that the +2σ test is not even significant at the two sided 80% confidence interval.

Ultimately, it can be inferred that when a TC approaches its MPI, it is more likely to gain IKE and less likely to intensify, all else being equal. However, since so few storms even approach their MPI, VMPI likely will not influence whether or not a storm will gain IKE if a storm is exceedingly far from its MPI. This is especially true of TCs near their time of genesis. Obviously, if a newly formed TC with weaker intensities is over warm SSTs, it will be far from its MPI. However, as discussed earlier, TCs traditionally gain IKE early in their lifecycle if conditions are favorable despite the large VMPI value.

5.11 Other Perturbation Tests

This section covers how the perturbation runs of the remaining two predictors, T150 and PENV, affect projections of IKE change by the SPIKE2 neural network system. The second of those input parameter was a fairly significant predictor for the linear SPIKE model, but T150 and was not used at all by the final version of SPIKE. In this case, the sensitivity tests reveal that both of these environmental variables have a significant influence of storm growth and decay.

The 150 hPa temperature parameter is related to the height of the tropopause in the models, wherein a warmer (colder) temperature at that pressure level signifies a lower (higher) tropopause. This variable has been used previously to identify instances of storm growth (Maclay et al. 2008), and in most cases it is understood that IKE is more likely to increase over areas of lower tropopause heights. Physically, IKE growth is more likely in areas of higher T150 because trough interactions, baroclinic influences, and ET are more likely to occur in the upper latitudes of the TNA basin, where the tropopause is lower and T150 is higher. As discussed at length in the previous sections, TCs are likely to expand as a result of these processes, making

IKE growth more likely when conditions such as T150 are more favorable for them to occur. Of course, the height of the tropopause is defined by the temperature of the air in the troposphere, and as such the topopause is typically higher in the deep tropics and lower in the midlatitudes and polar regions. Therefore, it is of no surprise that the perturbation runs for T150 nearly mirror those of the LAT perturbation runs. This means that IKE growth is more likely when T150 is perturbed upward, and that the distribution of projected IKE tendency becomes sharper as T150 decreases.

Next, the PENV input parameter is a measure of mean sea-level pressure in the storm's environment. A lower value of PENV would suggest that the storm is more intense and/or that it is a larger storm with a weaker pressure gradient. Therefore, it is not surprising that PENV has a negative relationship with fluctuations of IKE. As PENV decreases, it becomes more likely for IKE to increase and the distribution becomes broader (Figure 5.2). Overall, both of the negative PENV perturbations cause a significant increase in the median SPIKE2 projection (Table 5.1). On the other hand, a positive perturbation of PENV makes these same increases of IKE slightly less likely and the distribution of SPIKE2 projections becomes sharper. As a result, both of the positive PENV perturbation runs cause a decrease in the median SPIKE2 projection, but only the effect of the +2σ PENV perturbation run is actually significant that the two-sided 95% interval. Ultimately, this suggests that a lower area of pressure around the TC makes it more likely for a storm to grow, but an area of higher pressure around the TC will not prohibit a storm from growing unless it is too anomalously high.

Table 5.1: Percent changes for median SPIKE2 deterministic projections of IKE tendency in each perturbation run relative to a control simulation. As discussed in Section 5.1, each perturbation run adjusts exactly one variable (leftmost column) up or down by either one or two standard deviations (topmost row). These changes were tested for significance with a two-sample bootstrapping exercise. Those that are deemed significant at the two-sided 95% level are displayed in a blue font

Peturbed Input Parameter	Perturbations to Variables: -2σ	-1σ	+1σ	+2σ
PIKE	268%	150%	-177%	-406%
dIKE12	-13%	-8%	11%	25%
VMAX	-215%	-61%	34%	59%
RHLO	75%	45%	-39%	-71%
RHMD	-37%	-20%	21%	32%
VMPI	43%	36%	-17%	-11%
SHRD	-2%	-3%	11%	31%
SST	-21%	-16%	18%	1%
D200	-6%	-4%	10%	26%
T150	-23%	-9%	19%	46%
MSLP	157%	79%	-62%	-133%
VORT	-86%	-42%	49%	97%
LAT	-62%	-33%	23%	12%
LON	-69%	-25%	20%	28%
PENV	23%	14%	-7%	-17%
SHTD	24%	12%	-5%	0%
PDAY	43%	17%	-4%	-9%
SDAY	-16%	0%	-1%	-19%

Table 5.2: List of North Atlantic TCs that have 100 TJ or more of IKE, a VMAX that does not exceed 65 knots, and a MSLP less than or equal to 980 hPa. All of these storm fixes occur near the end of each TC's lifecycle.

TC Name	Date	IKE (TJ)	VMAX (kts)	MSLP (hPa)	LAT (°N)
Lili	10/26/96 18Z	192	60	980	42.3
	10/27/96 0Z	189	55	978	44.3
Gert	9/22/99 18Z	116	65	963	40.3
	9/23/99 0Z	113	60	964	42.2
Gabrielle	9/19/01 0Z	108	60	975	41.5
Olga	11/26/01 0Z	592	60	979	30.3
	11/26/01 6Z	592	60	979	30.1
	11/26/01 12Z	193	65	979	30.6
	11/26/01 18Z	118	65	977	31.1
Fabian	9/8/03 12Z	166	65	980	48.7
Kate	10/7/03 18Z	101	60	980	47.5
Karl	9/24/04 18Z	270	65	962	45.5
Florene	9/12/06 6Z	218	65	976	36.3
	9/12/06 12Z	244	65	976	37.4
	9/12/06 18Z	245	65	977	38.5
Helene	9/24/06 6Z	245	65	964	39.5
	9/24/06 12Z	273	65	964	40.6
Bill	8/23/09 18Z	107	65	970	44.4
	8/24/09 0Z	119	65	973	46.3
	8/24/09 6Z	127	60	980	48.0
Danielle	8/30/10 6Z	140	65	970	40.0
	8/30/10 12Z	133	65	970	40.6
	8/30/10 18Z	138	60	973	41.0
Igor	9/20/10 0Z	192	65	953	31.9
	9/20/10 6Z	238	65	957	33.2
	9/20/10 12Z	238	65	960	35.1
	9/20/10 18Z	243	65	960	37.2
	9/21/10 0Z	325	65	960	39.0
	9/21/10 6Z	327	65	960	41.5
	9/21/10 12Z	342	65	955	44.8
Irene	8/28/11 6Z	101	65	958	38.1

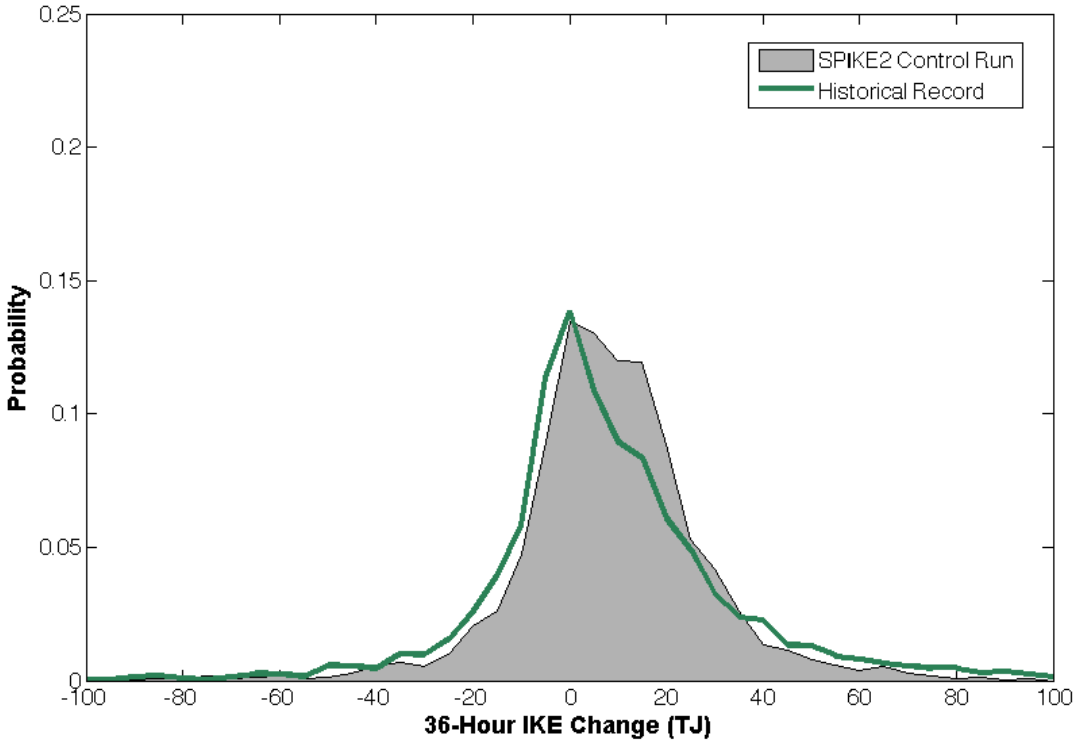

Figure 5.1: Probability distribution of 36-hour IKE change from the control run of the SPIKE2 neural network system and from the historical record. Approximately 3,000 storm fixes from 1995-2011 are included in this distribution.

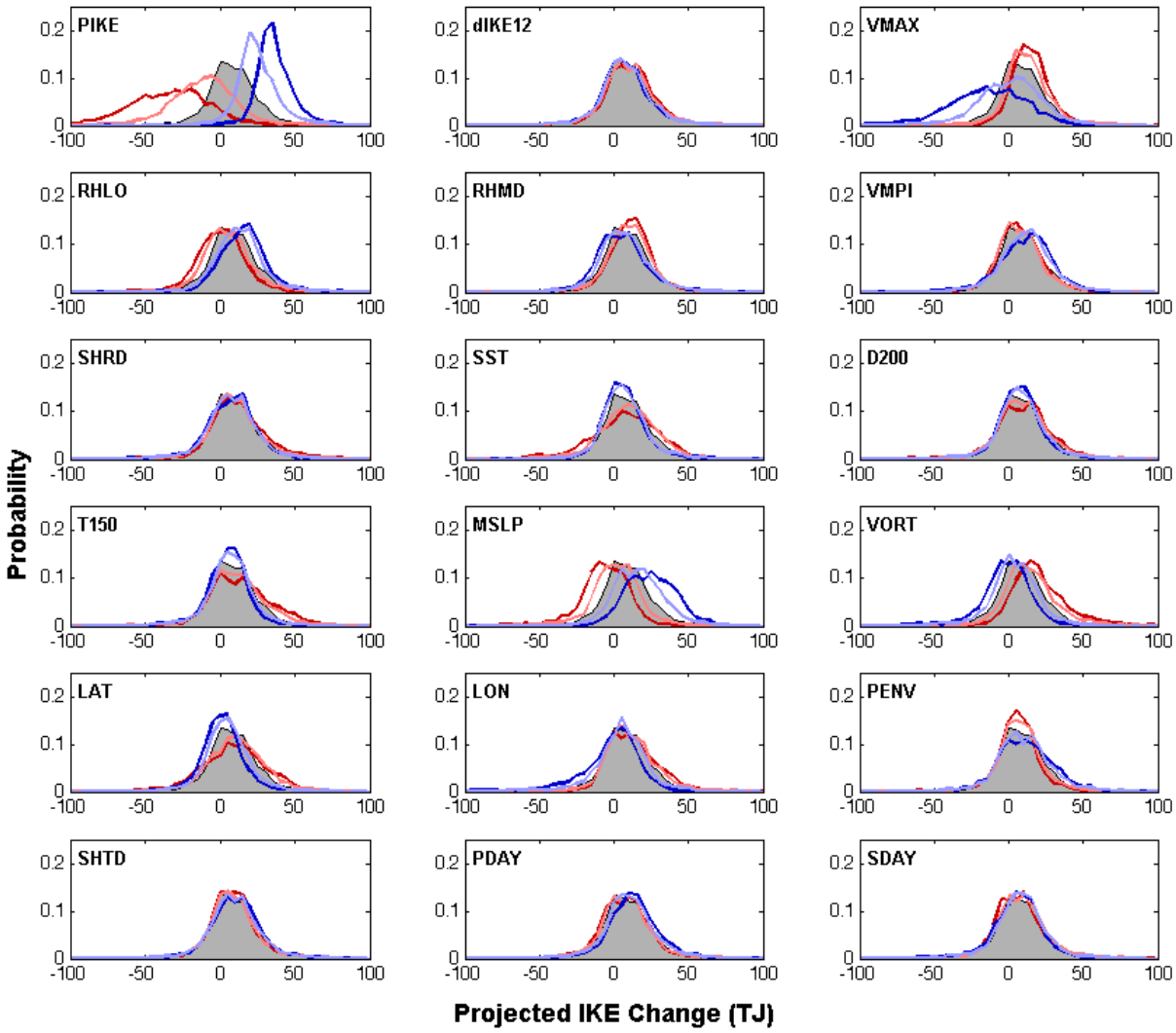

Figure 5.2: The results of the sensitivity exercise are presented by showing how variations of each input parameter affect the probability distribution of a 36-hour deterministic projection of IKE change from SPIKE2 for nearly 3,000 storm fixes. The probability distribution for the control run is reproduced from Figure 5.1 in each panel with a grey shaded polygon. The red curves show the distribution of projected IKE change from SPIKE2 when a single predictor is increased by one standard deviation (light red curve) or two standard deviations (dark red curve) for all historical storm fixes. The blue curves show the distribution of projected IKE change from SPIKE2 when each observed predictor is decreased by one standard deviation (light blue curve) or two standard deviations (dark blue curve) for all historical storm fixes.

CHAPTER 6

EVALUATION OF NEURAL NETWORKS FOR REAL-TIME USE

The proof of concept work detailed in the previous chapters established the foundation for projecting quantities of IKE from environmental predictors by using observed historical data. Ultimately, the perfect prognostic exercises in those chapters were useful for estimating the potential skill of IKE prediction. However, neither the linear regression SPIKE model developed in Chapter 3, nor the neural network SPIKE2 models developed in Chapter 4 are capable of being run in an operational setting. Given that observed historical data is only available in a post-storm capacity, IKE projection models must receive input variables from a numerical weather prediction model if they are to be run operationally in real time. As a result, this chapter will focus on adapting the SPIKE2 neural networks into a "mock-operational" mode by allowing the networks to accept reforecasted predictors in a step towards making SPIKE2 projections available in real time.

The first section of this chapter provides details on the archived analysis and reforecast dataset that is used in the calibration and evaluation exercises for the mock-operational version of SPIKE2. The following section will discuss how the mock-operational neural networks are calibrated and evaluated by using archived historical model analyses and reforecasts. Finally, Section 6.3 will use the ANNs to produce hindcasts of IKE from historical storms using actual reforecasted data from 1990 to 2011 in the GEFS archive to evaluate the predictive skill of SPIKE2 in an operational setting. This exercise will ultimately provide insight on how SPIKE2 will perform when run in real-time for the first time, hopefully during the upcoming 2015 Atlantic Hurricane Season.

6.1 Data Used to Create Mock-Operational Network

To evaluate forecast skill of the SPIKE2 model in a simulated operational setting, this chapter will utilize reforecasted data instead of the observed data that was used to calibrate the proof of concept models in the perfect prognostic approaches in Chapters 3 and 4. The mock-operational version of SPIKE2 still is calibrated and evaluated against the historical record of IKE as a target predictand, but all of the non-persistence input variables will be directly calculated from NOAA's second-generation GEFS reforecast dataset (Hamill et al. 2013). This expansive GEFS dataset contains global analyses and reforecasts from December 1984 to present day, which importantly overlaps with the entire time period during which historical IKE estimates are available. This entire dataset is produced by the Earth System Research Laboratory (ESRL) using a fixed version of the NCEP GEFS from February 2012 (version 9.0.1).

The version of the GEFS used in the reforecast dataset produces forecasts out to 16 days beyond the initialization time. The first eight days of the forecast are run at T254 horizontal resolution (~50km) with 42 vertical levels. The latter half of the forecast is run at a lower T190 resolution (~70km), with the same 42 vertical layers. The reforecast dataset is comprised of eleven ensemble members (1 control run, and 10 perturbation runs) compared to 21 ensemble members in the operational GEFS. Each of the reforecast runs is initialized once daily at 00Z, as opposed to the six-hourly approach for generating operational GEFS forecasts. The initial conditions for the reforecast dataset are produced from the Climate Forecast System Reanalysis (CFSR) prior to February 2011 and operational Grid-Point Statistical Interpolation analyses after that time.

NOAA archives ninety-eight output fields from initial time out to F+384 hrs for each of the daily GEFS reforecasts[1]. The archived data is stored at 3-hourly intervals for the first 72 hours of

[1] http://www.esrl.noaa.gov/psd/forecasts/reforecast2/

the forecast and then at 6-hourly intervals after that time. Each meteorological field is bilinearly interpolated down to a somewhat coarse one-degree resolution global grid. In addition to the one-degree data sets, a smaller selection of twenty-eight fields is also stored in the GEFS's higher resolution native Gaussian grid (~0.5°). However, the higher resolution fields are all single-level variables, primarily near the surface. As a result of this limitation, mid- and upper-atmospheric dynamic and thermodynamic fields (winds, temperatures, humidity, etc) are only available in the one-degree grids. Therefore, to maximize consistency, only the one-degree gridded datasets are used to calculate the input variables for the SPIKE2 forecast model. One might suggest replacing the GEFS with other models that are specifically tailored towards forecasting hurricanes at higher resolutions, since these models may produce better estimations of certain parameters (i.e. VMAX, MSLP, etc). However, by using the lower resolution GEFS reforecast data, we can estimate the lower bounds of skill for a real-time version of SPIKE.

In total, seven dynamical and thermodynamical reforecast fields have been downloaded from the GEFS's control run to adapt SPIKE2 for real-time use (Table 6.1). Analyses and reforecasts of these seven GEFS fields are used to produce hindcasts of the eighteen input parameters that can be used to run SPIKE2 in a mock forecast mode. Previously, the input parameters for the ANNs within the perfect prognostic version of SPIKE2 were mostly derived from the developmental SHIPS database (DeMaria and Kaplan 1999). These SHIPS predictors have very specific definitions because they were developed with operational intensity forecasting in mind. As such, the area over which these SHIPS predictors are averaged can seem somewhat arbitrary (Table 4.2). For example, many of the atmospheric predictors (PENV, SHRD, SHTD, RHLO, RHMD, T150) are averaged over an area between the radii of 200 and 800 km away from the storm center. Without context, it may seem odd that these predictors have an inner circle between 0 and 200km away from the center that is not included in the average. The reason for omitting this inner circle area from the predictor average stems from the fact that the operational version of SHIPS uses a combination of GFS model fields and the

NHC best track forecast to calculate its predictors (DeMaria et al 2005). To account for differences in the storm track between NHC and the GFS, an area around the center is omitted on the scale of the typical GFS forecast errors out to 3 days.

However, in the mock operational version of SPIKE2, the control run of the GEFS is used for both track and environmental forecast when calculating the various input parameters. As a result, the track forecast will be consistent with the placement of the TC vortex in the model solution, such that the omitted inner circle included in the SHIPS predictors is unnecessary. Therefore, the inner circles are no longer omitted from the averaging areas when calculating the eighteen SPIKE2 predictors from the GEFS. A full list of the input parameters that are needed to run the mock-operation version of SPIKE2, along with their updated averaging areas, is found in (Table 6.2). Since SPIKE2 is run at twelve forecast intervals (ranging from 6 to 72 hours), each input parameters will need to be calculated for every individual forecast interval from the GEFS reforecasts.

It should be noted that the GEFS dataset by itself is insufficient to calculate all eighteen of the input parameters. For instance, some of the input parameters require information about the ocean surface (VMPI, SST), time and date of year (SDAY,PDAY) and past values of IKE (PIKE, dIKE12). Therefore, to obtain hindcasts for each of the input parameters the GEFS reforecast dataset will be supplemented with a number of other datasets. Daily one-degree NOAA Optimum Interpolation SST ("OI SST"; Reynolds et al. 2007) is used to estimate observed ocean surface conditions. The historical IKE record is used to produce the persistence parameters, and finally the NHC best track dataset is used to get the time information for each storm fix.

In order to calculate most of the input parameters, it is first necessary to locate the center of the storm in the GEFS reforecast data. The center is identified in the model, by finding the GEFS's minimum in sea level pressure within five degrees latitude and longitude of the best track center position. This minimum pressure location within the GEFS will be used as

coordinates for the model simulated storm, and the LAT and LON input parameters. Once the center of the storm has been identified with in the model, calculating most of the other input parameters is trivial. In most cases, the input parameters can be calculated directly from the GEFS fields. For example, D200 is calculated by averaging the divergence of the wind within 1000 km of the center location in the GEFS. If ocean temperature data is required to calculate an input parameter, persistence SST values are taken from the daily OI SSTs. Finally, the persistence and timing parameters (PIKE, dIKE12, SDAY, PDAY) are calculated directly from the observed historical record of IKE and the NHC best track dataset. Table 6.3 details exactly which GEFS and OI SST fields are needed to obtain hindcasts for each field.

Once the input parameters are calculated from the GEFS for all forecast hours between initial time and T+72 hr, each parameter is normalized by its sample within the GEFS. This normalization will make the input parameters compatible with the SPIKE2 model as described in Chapter 4. Furthermore, normalizing the input parameters offers the benefit of filtering out some of the systematic biases in the GEFS, which in turn should enhance the performance of the operational IKE prediction schemes.

On the subject of biases, it should be noted that the methodology for calculating the SPIKE2 input parameters assumes that the GEFS will have a TC centered within five degrees latitude and longitude of the observed center. In most cases, this is a very reasonable assumption since only analyses and short term GEFS reforecasts (up to F+72 hr) are utilized to create the input parameters. However, in some cases, particularly when the historical storm is weak and disorganized, the GEFS reforecasts do not correctly identify that a TC exists within five degrees latitude and longitude of the expected storm center. In this case, many of the storm-specific parameters (VMAX, MSLP, VORT, etc) could be greatly underestimated, but many of the environmental predictors (e.g. SHRD) could still be somewhat accurate. Unfortunately, this is a common problem in operations, as models are incapable of correctly predicting tropical cyclogenesis in every instance. Therefore, we elected not to remove the

storm fixes in which the GEFS missed a storm. In theory, this will yield a more realistic estimation of operational skill for the real-time version of SPIKE.

There are many other imperfections within the GEFS that could also impact the quality of the input parameters and the mock operational skill evaluation of SPIKE2. For example, as stated earlier all of the model fields are interpolated to a one degree resolution grid. By definition, a one-degree resolution grid will have roughly one grid point per 100 km. As a result, the GEFS will not be able to resolve the small scale features (i.e. eyewall) within a hurricane. Therefore, it is expected that the GEFS will have some substantial biases, particularly with regards to the storm-specific intensity predictors, which in most cases will be greatly underestimated on this grid scale. Normalizing the predictors should help to correct for these mean biases to a certain extent. Nonetheless, it is important to understand how imperfect predictor data from numerical weather prediction models will affect the skill of SPIKE2 in these hindcast exercises.

6.2 Mock-Operational Neural Network Setup and Calibration

In contrast to the proof of concept ANNs used in previous chapters, the mock-operational version of the SPIKE2 neural network system is calibrated using the input parameters from the analyses in the GEFS reforecast dataset (i.e. the reforecast data at initialization time). As their name suggests, these GEFS zero-hour (F00) analyses are not forecasts themselves. Instead, they act as the best estimate for the observed conditions in the GEFS reforecast database's one-degree resolution grid. By using input parameters from the archived GEFS analyses to calibrate SPIKE2, the IKE prediction scheme will be property trained to accept input parameters derived from reforecasts of the same coarse resolution.

Since the GEFS reforecast runs are only initialized at 00Z, the maximum sample size for the F00 calibration dataset is the 1377 storm fixes that occur at 00Z between 1990 and 2011.

However, SPIKE2 requires persistence parameters of varying forecast lead times. Therefore, the sample size of the calibration dataset will decrease with increasing forecast hour because short-lived storms will not have longer-term persistence values. For comparison purposes, there are 1097 fixes at a forecast hour of 12 hours and 614 fixes for the 72-hour forecast interval.

As was the case in the previous chapter, the calibration dataset is randomly broken down into the training, validation, and evaluation subsets. These three calibration subsamples, comprised of both GEFS F00 input parameters and corresponding known target values of historical IKE tendency, are used to calibrate the weights of the neurons within the 100 ANNs that make up the SPIKE2 neural network system.

Upon calibrating the new ANNs, it would be prudent to quantify the maximum potential performance of the mock-operational version of SPIKE2. The proof-of-concept ANNs constructed in Chapter 4 provide a rough estimate for the potential level of skill of SPIKE2. However, those perfect prognostic exercises used historical input parameters that were not taken from a coarse resolution model analysis. It is distinctly possible that deriving input parameters from a coarse resolution model, such as the GEFS, could limit the potential performance of the neural networks without even considering the effects of forecast errors, all else being equal. Therefore, the maximum potential performance of the mock-operational ANNs is calculated by running the calibration input parameters from the GEFS F00 analyses through the mock-operational version of SPIKE2. As shown by Figure 6.1, the potential performance of the mock-operational version of SPIKE2, as measured by mean squared error, is very similar to the performance of the proof-of-concept, perfect-prog version of SPIKE2. Considering that the SHIPS and the GEFS F00 analyses are attempting to resolve the same observed conditions without any forecast errors, this result is not terribly surprising. Nonetheless, it is important to know that the SPIKE2 neural network system will not have significantly degraded performance when using data from a model with coarse resolution like the GEFS.

The expected performance of the mock-operational version of SPIKE2 should be lower than this potential performance level. Obviously, forecasted input parameters required in a real-time setting will contain inherent errors that are not found in the GEFS zero-hour analyses. Therefore, to properly assess the expected performance of SPIKE2 in an operational environment, the ANNs are given input parameters from 12- to 72-hour hindcasts in the control run of the GEFS reforecast database. These hindcasted input parameters include inherent forecast errors and can be used to simulate the imperfect input parameters that would be received in an operational environment because the variables are not dependent on observations or post storm analyses. By feeding these input parameters into the mock-operational version of SPIKE2, the first true hindcasts of IKE in Atlantic TCs will be created. Ultimately, the results from this experiment can finally be used to gauge the plausibility of using SPIKE2 to predict quantities of IKE in real time.

6.3 Mock-Operational Network Skill Evaluation

This section evaluates the expected skill of the SPIKE2 neural network system in a mock-operational mode, and determines whether or not SPIKE2's forecast products have the potential to be used in an operational setting. Both the deterministic predictions of total IKE and the two probabilistic products (uncertainty range and probability of exceedance) are evaluated throughout this section using various different runs of the SPIKE2 neural network system.

A) Deterministic Skill for IKE Hindcasts

Similar to previous chapters, hindcasts of IKE are constructed by adding the hindcasted value of IKE tendency from SPIKE2 to a persistence IKE value at initialization time. The results of the deterministic total IKE forecasts from each of the experiments are shown in Figures 6.2 and 6.3. These two figures include plots of two previously used metrics: correlations and mean

absolute errors. In all cases, the expected mock-operational errors and correlations (SPIKE2 with GEFS hindcasted input parameters) are plotted against the potential mock-operational errors and correlations (SPIKE2 with GEFS F00 analysis input parameters) and the same two metrics for a persistence forecast.

Unsurprisingly, the expected deterministic performance of the mock-operational version of SPIKE2 falls between its high potential deterministic performance and the lower performance of a persistence forecast in terms of both correlation and mean absolute error at all forecast intervals. Within the shortest forecast intervals (i.e. 12-hours or less), the GEFS SPIKE2 deterministic hindcasts correlate to the observed total IKE values at approximately r=0.90 with a mean error of 8 TJ. These levels of performance are somewhat comparable to the maximum potential deterministic performance that is possible by this operational model in the event that the hindcasted input parameters exactly match the analyses. In fact, the estimated error of the mock-operational SPIKE2 system is only about a single TJ higher than its lowest potential error.

On the other hand, the mock-operational version of SPIKE2 does not greatly improve over persistence within the shortest forecast periods. At the 12-hour forecast interval, the GEFS SPIKE2 deterministic hindcasts are nearly indistinguishable from a persistence forecast in terms of mean absolute error. Both persistence and the mock-operational version of SPIKE2 have mean errors of just under 9 TJ at this short interval. As such, a deterministic forecast from the operational version of SPIKE2 is only expected to have about a 5% reduction of MSE relative to persistence within a time frame of 12-hours (Figure 6.4). As discussed previously, it might be unreasonable to expect SPIKE2's deterministic forecasts to be significantly better than persistence at such short forecast intervals, especially when given imperfectly hindcasted input parameters. IKE is an inertial quantity, and as a result, IKE often does not change much in a the span of twelve-hours or less, Therefore, persistence forecasts are typically fare very well in this limited short-range window.

However, as the initialization time of these SPIKE2 hindcasts becomes further removed from its validation time, the performance of a persistence forecast deteriorates quite rapidly. In a 48-hour forecast window, the expected mean absolute error of an operational version of SPIKE2 is approximately half that of a 48-hour persistence forecast (Figure 6.3). More impressively, a 72-hour hindcast from the mock-operational version of SPIKE2 has a higher expected correlation to the historical IKE record than does a much shorter 24-hour persistence forecast (Figure 6.2). Furthermore, the error expected by a 72-hour operational version of SPIKE2 is comparable to the amount of error in a persistence forecast with a lead time half as long. Overall, the expected reduction of MSE relative to persistence by the mock-operational version of SPIKE2 (shown in Figure 6.4) is significant at the two sided 90% confidence interval starting at a lead time of 24 hours. For a lead time of two-days or more, the mock-operational version of SPIKE2 reduced MSE by more than 60% compared to persistence, which is significant at the two sided 99% confidence interval.

Obviously, deterministic forecasts in an operational version of SPIKE2 are expected to fall short of their maximum potential deterministic performance baselines. As forecast hour increases, the input parameters will contain larger biases and errors, such that the errors of an operational statistical-dynamical model will also begin to increase. Therefore, it is not surprising that the gap between the expected and potential error levels from an operational version of SPIKE2 increases with increasing forecast hour. By 72-hours, an operational version of SPIKE2 is expected to exhibit an average error of about 20 TJ, compared to a minimum potential error of just 14 TJ. Overall, this modest drop in skill relative to the minimum potential error is not that large when considering that SPIKE2 is moving from a perfect scenario in which it is given nearly perfect GEFS F00 predictors to a realistic operational scenario with imperfect 72-hour hindcasted input parameters. Importantly, at these larger forecast intervals, the expected performance of the operational version of SPIKE2 more closely resembles its maximum

potential performance metrics than it does the performance metrics for a lowly persistence forecast.

A simple persistence forecast makes for a good starting benchmark when evaluating the expected operational skill of SPIKE2, but it is possible to measure the deterministic skill of SPIKE2 relative to a more challenging climatology and persistence baseline. For instance, it is possible to create a simple linear regression model that uses nothing more than a series of persistence and climatological variables at initialization time in lieu of the hindcasted environmental parameters from GEFS used in previous exercises. A simple statistical model such as this one would follow in the footsteps of the Statistical Hurricane Intensity Forecast model (SHIFOR), which uses seven known parameters at initialization time to set the baseline skill for operational intensity forecasts (Jarvinen and Neumann 1979; Knaff et al. 2003). The exact parameters of SHIFOR include: Julian day, initial storm intensity, previous 12-hour intensity change, initial latitude, initial longitude, initial zonal component of storm motion, and initial meridional component of storm motion. These SHIFOR climatology and persistence predictors are somewhat relevant to IKE tendency as well. Therefore, an IKE statistical persistence model named the “Benchmark of Integrated Kinetic Energy (BIKE)” is created to project IKE tendency in a simple linear regression model using the same seven input parameters, with two exceptions. First, the 12-hr intensity change parameter will be switched out for a 12-hr IKE change parameter. Second, the initial or persistence value of IKE will be added as an eighth predictor. This BIKE regression model is trained using all 00Z storm fixes from 1990-2011, such that its calibration fit will be compared to the GEFS-SPIKE2 hindcasts at various lead times for the same 1990-2011 interval.

The calibration fit of BIKE is shown against other metrics in Figure 6.5. Overall, BIKE is indeed a tougher benchmark than a simple IKE persistence forecast, especially at the longest forecast interval. Nonetheless, the calibration fit of BIKE, in terms of both correlations and mean absolute error, falls well short of the expected and potential skill of SPIKE2 as calculated in the

hindcast exercises. At 48-hours, BIKE has an average error of 23 TJ which is about 40% higher than the expected mean error in the GEFS-SPIKE2 hindcasts. At the longest 72-hour forecast interval, the statistical persistence model's mean errors are twice that of the potential GEFS-SPIKE2 run.

Overall, it can be concluded that the operational version of SPIKE2 is expected to produce deterministic total IKE forecasts that are significantly better than any forecasts derived from persistence and climatology guidance. The GEFS-SPIKE2 hindcasts clearly improve upon forecasts from both a simple persistence forecast and the more complicated persistence linear regression model, BIKE. This improvement over persistence is particularly noteworthy at longer forecast intervals, where the added environmental and internal storm input parameters become necessary to properly forecast quantities of IKE. Perhaps more impressively, the skill levels exhibited by the deterministic total IKE hindcasts exceed those produced by the proof-of-concept linear regression model that used perfect input parameters in Chapter 3. Overall, this speaks well for the power of ANNs because linear models coupled with imperfectly hindcasted predictors undoubtedly would have exhibited significantly less skill than was exhibited by the mock-operational version of SPIKE2 in this section.

Finally, it is also possible to measure the skill of SPIKE2's deterministic hindcasts relative to existing operational benchmarks that could potentially be tougher to beat than a simple persistence forecast or linear model. Currently, the National Hurricane Center does not specifically forecast values of IKE. However, it is possible to create an "NHC-based IKE forecast" because the official NHC forecast advisories include forecasts of operational radii and intensity. Together, the data from these NHC forecast advisories can be used to produce an approximate forecast of IKE using the equations from Powell and Reinhold (2007). Operational radii are forecasted by NHC at least four times daily (03Z, 09Z, 15Z, 21Z) for active TCs, out to 33 hours for r_{64} and out to 69 hours for r_{50} and r_{34}. Therefore, the NHC-based IKE forecasts would be limited to at most a 33 hour forecast for any storm at or exceeding minimal hurricane

intensity. NHC's forecast advisories are archived online, such that it is possible to generate NHC-based IKE forecasts for most past storms in the last two decades as comparison to the mock-operational GEFS-SPIKE2 hindcasts.

Four three-day case studies involving United States landfalling hurricanes (Katrina 2005; Floyd 1999, Ike 2008, and Irene 2011) are selected for this comparison, in an effort to evaluate the skill of the hindcasts and forecasts for societally impactful storms. The deterministic GEFS-SPIKE2 hindcasts are run in a mock-operational mode for a full 72-hour window in the days leading up to each hurricane's landfall. Again, the NHC-based IKE forecasts can only be calculated for a shorter 33-hour forecast interval, so those forecasts will be taken from multiple advisories issued throughout the hindcast window to cover the entire three day span.

The results of these four case studies are shown in Figure 6.6. Overall, the deterministic GEFS-SPIKE2 hindcasts perform quite well for each of these four storms. In each case, the hindcasts correctly identified that IKE values would increase as these hurricanes approached the United States. The hindcast for Hurricane Irene was particularly impressive, as it nearly matched the observations at each forecast hour. However, the deterministic hindcasts unsurprisingly still exhibit somewhat of a low bias for the high IKE values in the other three hurricanes.

In comparison, the shorter NHC-based forecasts do not seem to perform as well in this small sample. In fact, the NHC-based forecasts seem to closely resemble a persistence forecast, wherein the forecasted value of IKE does not deviate much from the initial value of IKE at the time the advisory is issued. For instance, in Hurricane Ike, the NHC-based forecast issued at 03Z on 9/10/08 was too low because the hurricane gained IKE over the subsequent 33-hours. However, the next forecast issued on 9/11/08 at 15Z was too high because the hurricane lost IKE in the subsequent hours.

Ultimately, the examples provided by these four landfalling hurricane case studies suggest that a 72-hour hindcast from the GEFS-SPIKE2 model can outperform a much shorter

33-hour NHC-based forecast. These results are not meant to be a definitive comparison of the model's skill relative to current operational guidance, especially given the very small sample of storms. Nonetheless, considering the impacts of these four storms, at the very least, it can be determined that the deterministic SPIKE2 model is worth investigating further because it appears to have the potential to be an operationally relevant piece of guidance for determining the IKE values and resulting damage potential of significant landfalling storms.

B) Probabilistic Skill for Total IKE Hindcasts

In this subsection, the expected operational skill of the probabilistic SPIKE2 products is evaluated within the framework of the GEFS-SPIKE2 hindcasts for all storms between 1990 and 2011. The first of those probabilistic products is the uncertainty range tool, which is used to measure the inherent forecast uncertainty present within the ANNs that make up SPIKE2. This uncertainty range tool can be coupled with deterministic IKE projections to present a best-estimate for IKE alongside a series of properly scaled error bars in a single product.

Example hindcast graphics for the uncertainty range are produced for selected notable hurricanes over the past two decades in Figure 6.7. These six hindcasts were produced using the GEFS input parameters, and overall, they seem to perform well. For instance, the 72-hour GEFS-SPIKE2 hindcasts correctly track the evolution of IKE in Hurricanes Irene (2011) and Fabian (2003), such that the observed IKE amounts consistently fall within the uncertainty range across all forecast hours. On the other hand, the hindcasts for Katrina (2005) and Floyd (1999) help to reiterate a limitation of this statistical forecast tool, wherein the ANNs and as a result the uncertainty range tool are likely to underestimate extreme events. That being said, the Katrina and Floyd hindcasts are still useful, as they correctly project steadily increasing IKE throughout the forecast window, culminating in an uncertainty range upper bound that nears or exceeds 100 TJ by the end of the forecast.

The six example products suggest that the uncertainty ranges will grow in size as forecast interval increases. In general, this relationship holds true, as a wider analysis of all storms between 1990 and 2011 indicates that the mean uncertainty range grows with increasing forecast hour (Figure 6.8). At a short 12-hour forecast window, the mock-operational version of SPIKE2 has an uncertainty range of nearly 17 TJ on average. At the much longer, 72-hour forecast interval that average uncertainty range grows to 46 TJ. Ultimately, this positive relationship between forecast interval and uncertainty range size is connected to the larger errors by the SPIKE2 forecast system at larger lead times. Therefore, it is not surprising to discover that the average uncertainty ranges are about twice as large as the mean absolute error of the deterministic forecast for each forecast interval. By design, this ensures that most of the historical IKE values fall within the uncertainty range whether they are above or below the deterministic forecast by a somewhat typical magnitude of error.

The pie charts beneath each of the twelve-hourly points on Figure 6.8 indicate the percentage of historical IKE values that fall within or outside of their hindcasted uncertainty range. As previously discussed, the 87th and 13th percentiles were chosen as the bounds of this uncertainty range because two-thirds of all historical IKE values fell within the proof-of-concept version of this product, given perfect predictors. Despite the inclusion of imperfect GEFS input parameters, the mock-operational version of the uncertainty range tool still captures more than 60% of its targeted historical IKE values at each forecast interval. In fact, the length of the forecast interval does not seem to affect the likelihood of the uncertainty range capturing the expected historical IKE quantity. This occurs because the uncertainty range uses percentiles and is allowed to grow or contract depending on forecast length and environmental conditions.

Interestingly, at shorter forecast intervals, observed IKE quantities are just as likely to be above or below the bounds of the uncertainty range. However, at the longer forecast intervals, it becomes evident that the uncertainty range tends to exhibit a low bias, wherein a higher percentage of storms exceed the upper bound of the uncertainty range. This low bias is

expected because the previously discussed inability of SPIKE2 to properly predict extreme levels of IKE without the help of an already high persistence value.

The second probabilistic tool that is discussed in this subsection is the Probability of Exceedance product. As discussed in Section 4.5, the PoE tool measures the probability of IKE eclipsing various thresholds at each forecast hour. Evaluation of the PoE tool in a mock-operational setting centers on the Brier Skill Score metric. As discussed previously, the BSS measures a probabilistic forecast's skill relative to a reference forecast, which in this case will be climatology. The BSS is calculated for each of the four thresholds (25, 50, 100, and 200 TJ) used in the SPIKE2-PoE product (Figure 6.9). In general, the expected BSS for the mock-operational PoE tool decreases with increasing forecast hour. For example, the mock-operational PoE tool has an expected BSS of 0.81 when projecting the likelihood of a TC exceeding 50 TJ at a short 12-hour forecast window. That BSS continues to drop to a BSS of 0.66 at 72-hours for the same 50 TJ threshold.

Compared to climatology, the PoE tool remains more skillful at projecting whether or not a TC will exceed its lowest two thresholds, 25 and 50 TJ for every forecast interval. These two thresholds are exceeded much more frequently than the larger thresholds. As a result, a simple fixed climatological forecast will do quite poorly. Therefore, it is encouraging to see that SPIKE2's expected mock-operational BSSs always exceed 0.60 for these two thresholds regardless of the forecast hour. On the other hand, a fixed climatology forecast for the highest 200 TJ threshold has a more skill, simply because such large IKE values are rare. As such, a simple fixed low-percentage forecast will often be very close to the actual outcome (which equals zero when a TC does not exceed the threshold). Nonetheless, the neural networks acknowledge that TCs rarely exceed the higher thresholds, and as a result, the SPIKE2-PoE product often predicts that the probability for TC to exceed 200 TJ will be equal to 0%. Therefore, the 200 TJ PoE product is still more skillful than climatology, but the improvement is somewhat limited. The lack of a significant improvement over climatology likely arises because

of the SPIKE2 system's tendency to under predict extreme events. Ultimately, this would lead the SPIKE2-PoE tool to underestimate the probability of a TC exceeding 200 TJ of IKE in most cases.

Nonetheless, both of the probabilistic forecast tools fare reasonably well in a mock-operational setting with independently hindcasted predictors. The uncertainty range tool appropriately scales to fit the historical values of IKE based on the input parameters and the forecast lead time. As such, it is a useful tool in an operational setting to assess the confidence of the neural network methodology. The PoE tool offers additional value to forecasters by assigning the likelihood of exceeding certain IKE thresholds at a skill level that exceeds a simple fixed climatological forecast. Overall, the skill levels exhibited by the deterministic and probabilistic SPIKE2 products indicate that this ANN system can serve as a powerful forecast tool for meteorologists in a realistic real-time setting.

Table 6.1: One-degree resolution gridded fields from the GEFS reforecast archive that are used to produce hindcasted input variables for the mock-operational SPIKE2 model. The abbreviation of each GEFS field is in the leftmost column. A description and the units of each variable are in the middle column. The vertical pressure levels of each field that are used to calculate SPIKE2 input parameters are shown in the right most column.

Abbreviation	Variable Description	Vertical Levels (hPa)
spfh_pres	specific humidity (kg/kg)	1000,925,850,700,500,300
tmp_pres	air temperature (K)	1000,925,850,700,500,300,250,200,100,50,10
ugrd_pres	zonal wind (m/s)	850,700,500,300,250,200
vgrd_pres	meridional wind (m/s)	850,700,500,300,250,200
ugrd_10m	10-meter zonal wind (m/s)	10m
vgrd_10m	10-meter meridional wind (m/s)	10m
pres_msl	Mean sea-level pressure (Pa)	mean sea level

Table 6.2: Input parameters for the neural networks within the mock-operational version of the SPIKE2. These input parameters are obtained from GEFS reforecasts and analyses, NOAA OI SSTs, and the historical record. The definitions and abbreviations for many of the predictors are identical to those listed in Table 4.2 for the perfect prognostic version of SPIKE2. However, since these predictors are not restricted to the definitions established by SHIPS, many of the averaging areas have been slightly altered.

Variable	Definition	Units
PIKE	persistence of IKE	TJ
dIKE12	previous 12hr change of IKE	TJ
VMAX	maximum sustained wind speed	kts
VMPI	Difference between maximum potential intensity and VMAX	kts
LAT	latitude of storm's center	°N
LON	longitude of storm's center	-°W
MSLP	minimum sea level pressure	hPa
PENV	average surface pressure *(averaged from r=0-800km)*	hPa
VORT	850 hPa vorticity *(r=0-1000km)*	10^{-7} s^{-1}
D200	200 hPa divergence *(r=0-1000km)*	10^{-7} s^{-1}
SHRD	850-200 hPa shear magnitude *(r=0-800km)*	kts
SHTD	850-200 hPa shear direction *(r=0-800km)*	°
RHLO	850-700 hPa relative humidity *(r=0-800km)*	%
RHMD	700-500 hPa relative humidity *(r=0-800km)*	%
T150	150 hPa temperatures *(r=0-800km)*	°C
SST	sea surface temperatures	°C
SDAY	time after tropical storm genesis	days
PDAY	time from peak of season (Sept. 10)	days

Table 6.3: Description of which reforecast datasets (top row) were used to calculate each input parameter for the mock operational version of SPIKE2 (leftmost row). The GEFS reforecast datasets include analyses at initial time and several forecasts out to 72 hours. The NOAA OI SST dataset is a persistence value taken from the initial time. The NHC Best track dataset is used primarily to calculate the timing and persistence predictors. In addition, the best track storm coordinates are used as a starting point to search for the center of each storm within each run of the GEFS reforecast database.

Reforecast Fields

SPIKE2 Input Parameters	spfh pres	tmp pres	ugrd pres	vgrd pres	ugrd 10m	vgrd 10m	pres msl	NOAA OI SST	Best Track
PIKE									X
dIKE12									X
VMAX					X	X	X		X
VMPI	X	X					X	X	X
LAT							X		X
LON							X		X
MSLP							X		X
PENV							X		X
VORT			X	X			X		X
D200			X	X			X		X
SHRD			X	X			X		X
SHTD			X	X			X		X
RHLO	X	X					X		X
RHMD	X	X					X		X
T150		X					X		X
SST							X	X	X
SDAY									X
PDAY									X

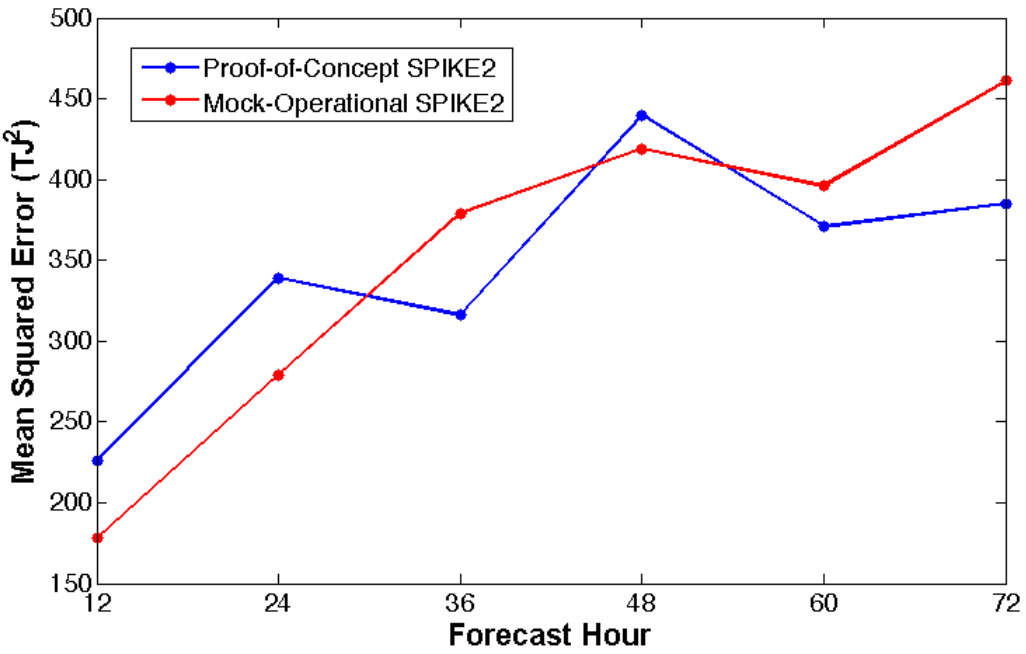

Figure 6.1: Plot of mean squared error from two different versions of the SPIKE2 neural network system. The blue curve shows the errors of the proof-of-concept version of SPIKE2 created in Chapter 4 when given perfect input parameters from SHIPS and HURDAT. The red curve shows the minimum potential error of the mock-operational version of SPIKE2 when given input parameters from the GEFS zero-hour analyses.

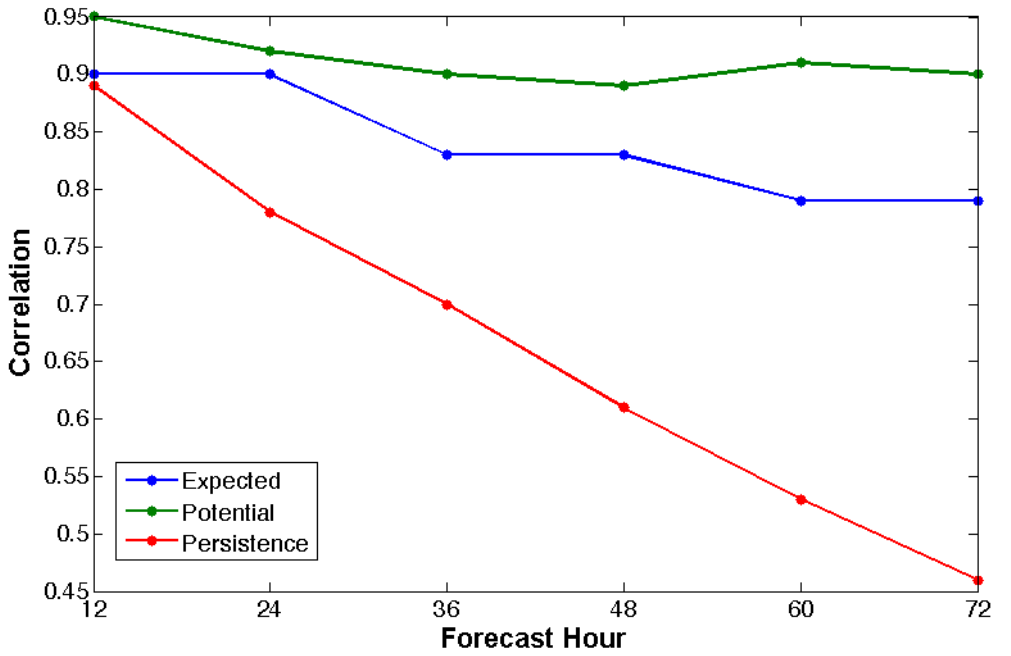

Figure 6.2: Correlations for total IKE between the mock-operational version of SPIKE2 and the observed historical IKE quantities in Atlantic TCs from 1990 through 2011. The blue curve represents the expected correlation of a real-time deterministic total IKE forecast, which is measured by the mock-operational version of SPIKE2 forced with hindcasted GEFS input parameters. The green curve represents the maximum potential correlation of a deterministic total IKE forecast, which is measured by the mock-operational version of SPIKE2 forced with GEFS F00 analysis input parameters. The red curve represents the correlation of a persistence forecast.

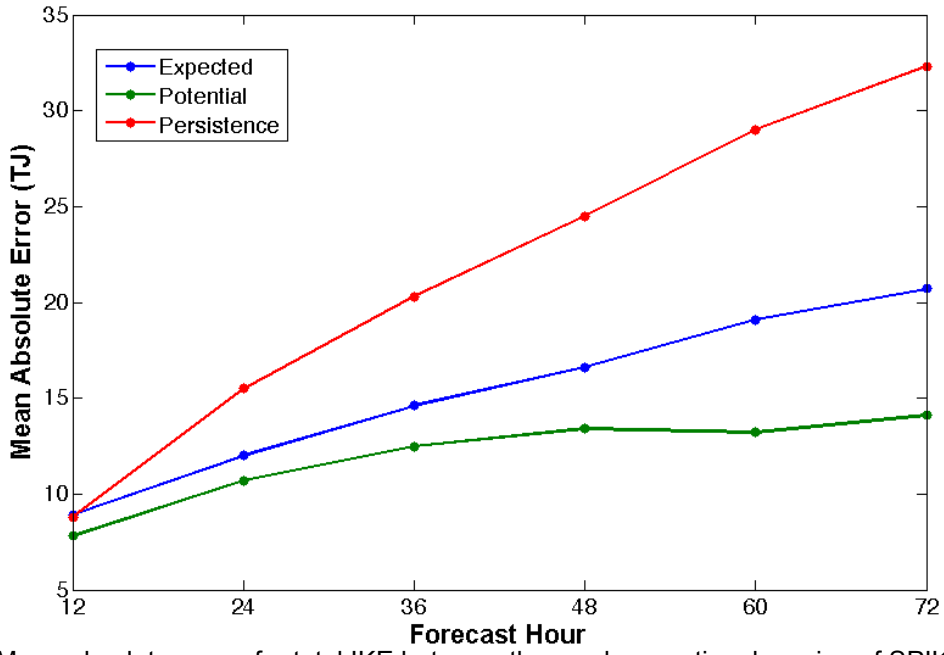

Figure 6.3: Mean absolute errors for total IKE between the mock-operational version of SPIKE2 and the observed historical IKE quantities in Atlantic TCs from 1990 through 2011. The blue curve represents the expected error of a real-time deterministic total IKE forecast, which is measured by the mock-operational version of SPIKE2 forced with hindcasted GEFS input parameters. The green curve represents the minimum potential error of a deterministic total IKE forecast, which is measured by the mock-operational version of SPIKE2 forced with GEFS F00 analysis input parameters. The red curve represents the error of a persistence forecast.

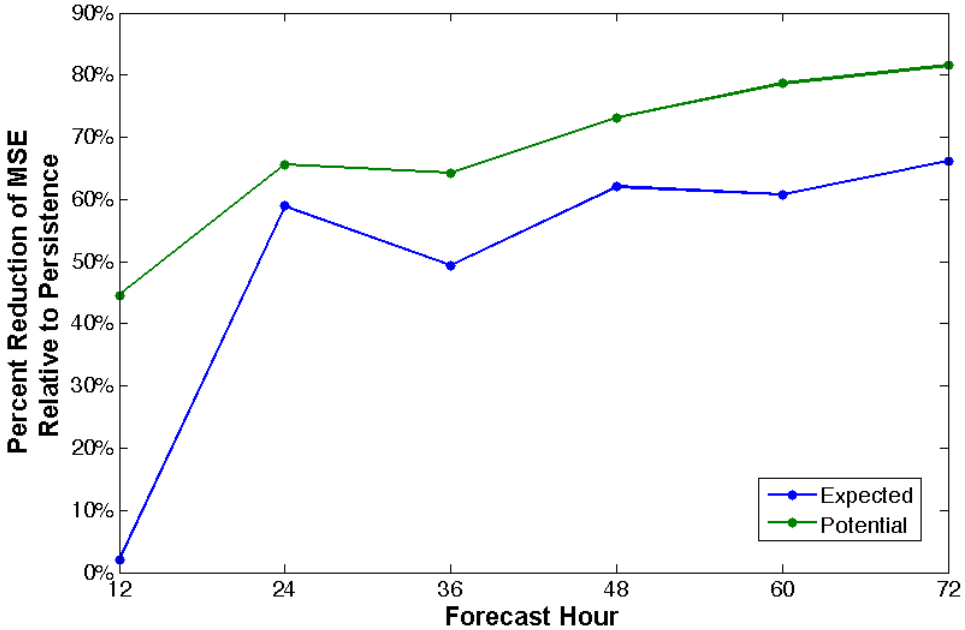

Figure 6.4: MSE reduction relative to persistence for the mock-operational version of SPIKE2 from 1990 through 2011. The blue curve represents the expected skill of a real-time deterministic total IKE forecast, which is measured by the mock-operational version of SPIKE2 forced with hindcasted GEFS input parameters. The green curve represents the maximum potential skill of a deterministic total IKE forecast, which is measured by the mock-operational version of SPIKE2 forced with GEFS F00 analysis input parameters.

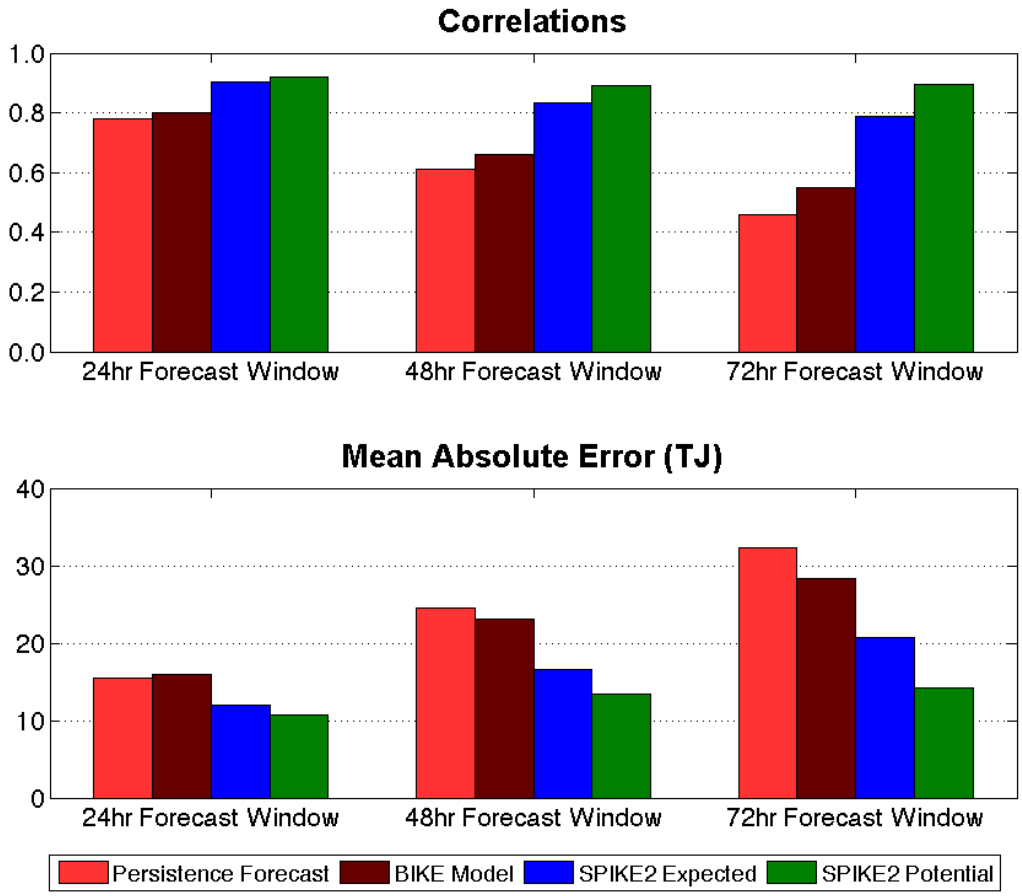

Figure 6.5: Comparison of correlation and error metrics between the SPIKE2 hindcasts (blue and green bars), persistence (light red bars), and a statistical model trained solely on persistence and climatological predictors (dark red bars) at various forecast hours.

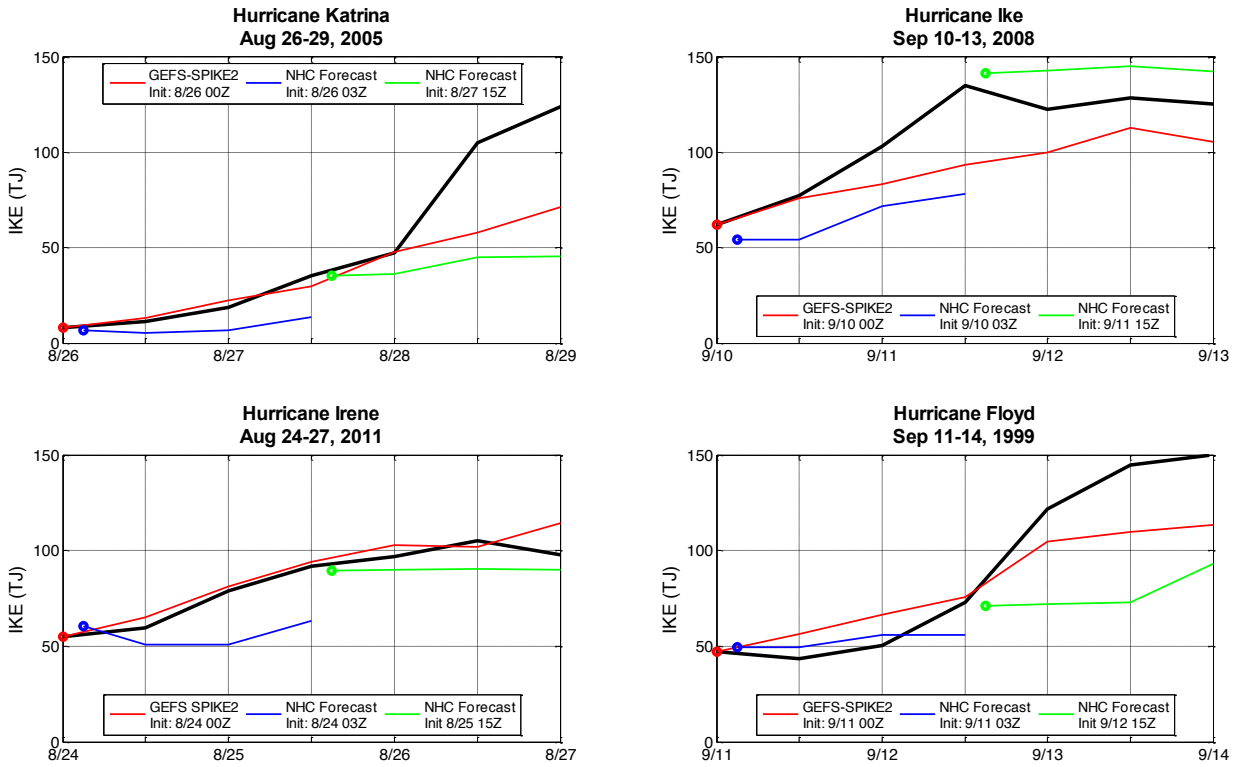

Figure 6.6: Comparison between 72-hour GEFS-SPIKE2 deterministic hindcasts of IKE and 33-hour NHC-based forecasts in four landfalling United States hurricanes. The black line represents the observations, the red line represents the SPIKE2 deterministic GEFS-SPIKE2 hindcast, and the other color lines represent the NHC-based forecasts. The initialization time for each hindcast and forecast is shown by an open circle on the plot.

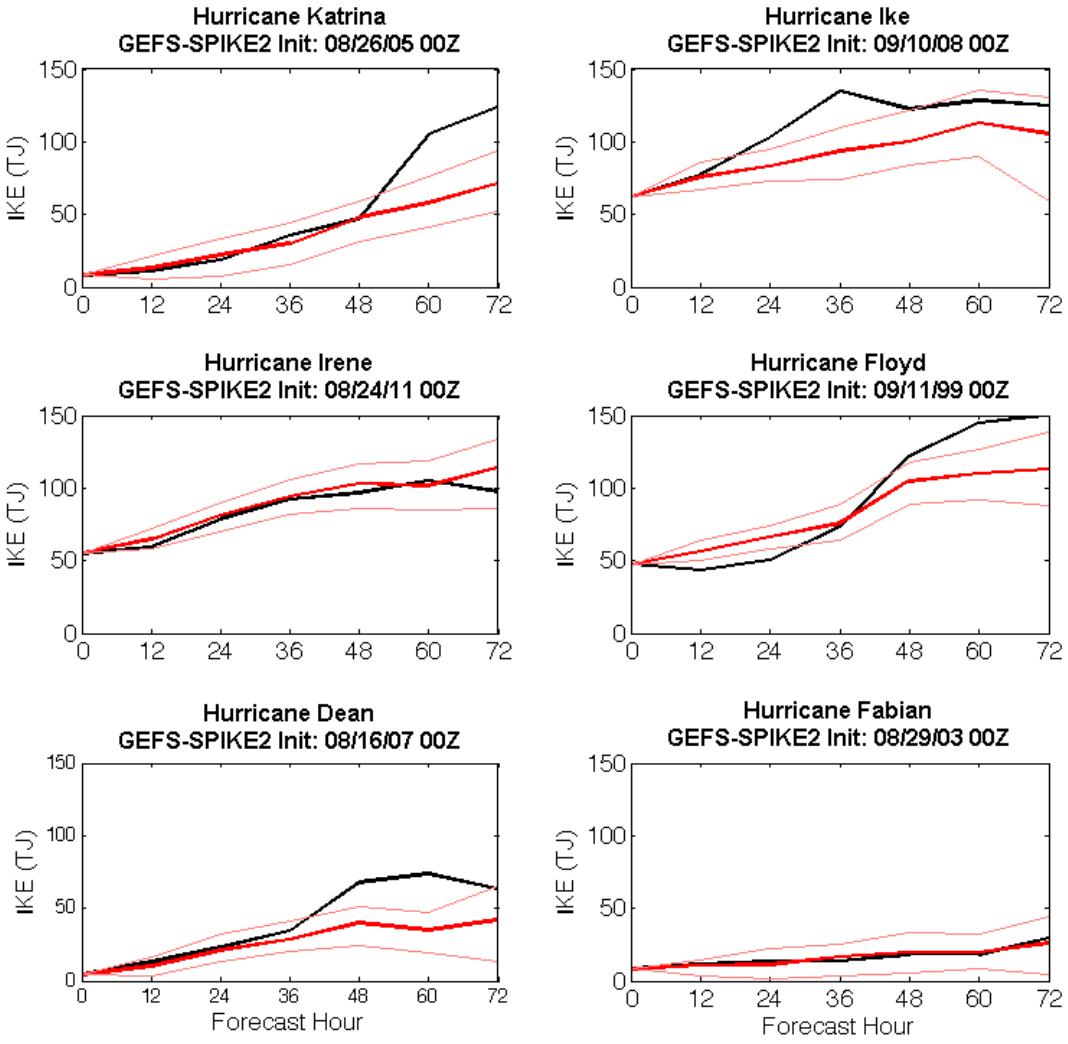

Figure 6.7: Example 72-hour hindcasts of IKE for TCs between 1990 and 2011, using the mock-operational GEFS-SPIKE2 runs. In each panel, the uncertainty range (light red lines) is presented alongside of the deterministic SPIKE2 forecasts (bold red line) and the actual observed value of IKE (bold black line).

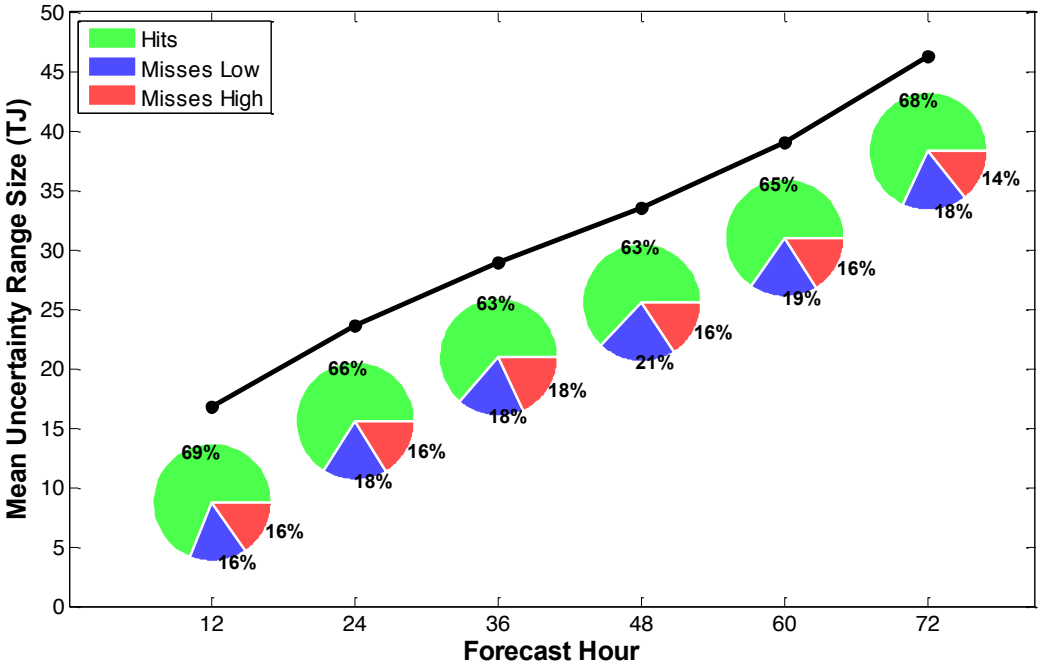

Figure 6.8: Various evaluation metrics for the GEFS SPIKE2 uncertainty range tool. The black curve gives the average size of the uncertainty range for every IKE hindcast as denoted by the left axis. This average size is defined by subtracting the bottom error bar from the top error bar for every GEFS SPIKE2 probabilistic hindcast from 1990-2011. Note that dividing these uncertainty range sizes by two corresponds well with the mean absolute error metric shown by the blue line in Figure 6.3. The pie graphs beneath each data point show the distribution of the observed historical IKE events from 1990-2011 with relationship to the uncertainty range. A hit occurs when the observed total IKE values are found to be correctly located within the uncertainty range of the GEFS SPIKE2 hindcasts. A miss low (high) occurs when the uncertainty range falls below (above) the observed value of IKE.

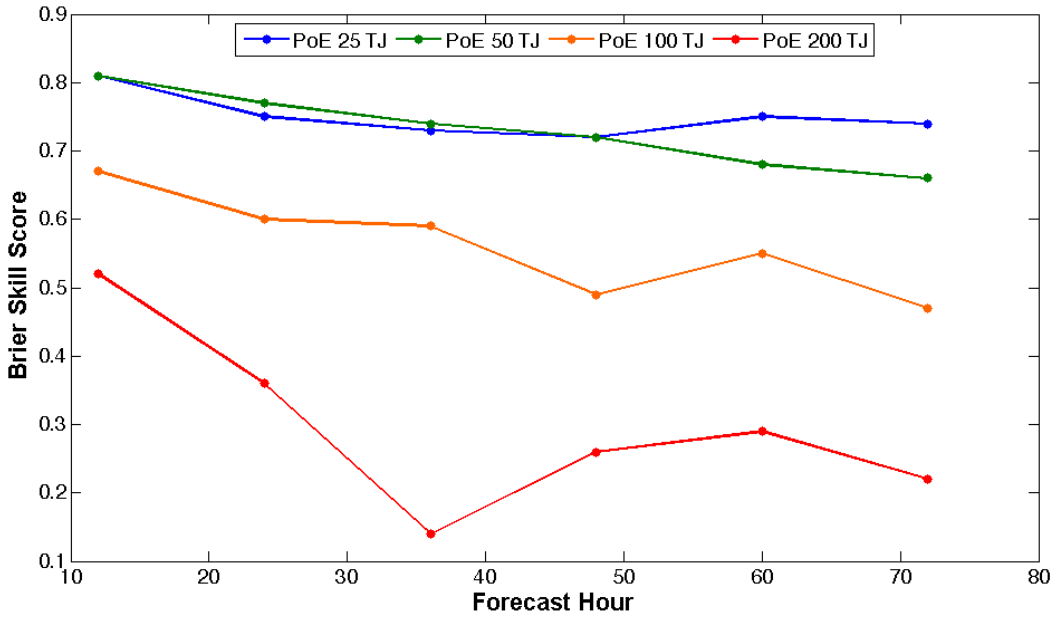

Figure 6.9: Brier skill scores for the GEFS SPIKE2 hindcasts' probability of exceedance (PoE) tool. The Brier skill scores are calculated with respect to climatology for each IKE threshold, color coded from cool colors for the smaller thresholds to warm colors for the larger thresholds. A positive BSS indicates that the PoE tool exhibits more skill than climatology

CHAPTER 7

APPLYING IKE FORECASTS TO PREDICT STORM SIZE

As discussed throughout this dissertation, IKE is closely related to the size of a TC. By definition, IKE is an integration of a TC's wind field, such that a larger storm will have more IKE than a smaller storm, all else being equal. Figure 7.1 clearly indicates that the radii of the three measured wind speeds (34, 50, and 64 kts) increases nonlinearly with increasing IKE. Based on IKE's relationship with storm size, it is ostensibly possible that IKE could be used to approximate operational wind radii used by the National Hurricane Center.

Any such relationship between the operational radii and IKE would increase the value of specifically tailored IKE forecasts to weather forecasters. Currently, IKE is used in very few if any operational decisions, limiting the value of the IKE forecasting tools presented in earlier chapters. A common criticism against the operational use of IKE is that TeraJoules of kinetic energy might be difficult to disseminate to the general public, despite the metric's well known correspondence to damage potential. Obviously, there is still value in categorizing the damage potential from a certain TC through IKE (Powell and Reinhold 2007; Kozar and Misra 2014), especially considering the well known shortcomings of categorizing storm surge damage with existing maximum intensity metrics such as VMAX (Irish et al. 2008). However, if these IKE forecasts can be applied to forecast existing operational metrics, SPIKE2 would be more versatile and desirable to operational forecasters.

Therefore, a method is presented for approximating the sizes of observed TC wind fields from estimates of IKE with a simple axisymmetric vortex model. Through this new IKE-based wind radii model, the SPIKE2 forecasts will be useful for not only assessing damage potential through values of IKE, but also for estimating the size of a storm in terms of its operational wind radii. Such wind radii forecasts can be used directly by forecasters and emergency responders

to assess certain risks of a TC that cannot be assessed with simple intensity and position forecasts.

The design of the new IKE-based axisymmetric wind model is presented in the first section. The second section discusses the potential performance of the symmetric wind model using historical cases in a perfect prognostic mode. The third section expands upon those results by exploring the plausibility of predicting wind field asymmetries using information about the translational motion of a TC. Finally, the fourth section will discuss the expected forecast skill of this wind radii tool using input data from the GEFS and the SPIKE2 neural networks. This operational skill is ultimately compared to radii projections made by the National Hurricane Center to determine if this tool has any future operational value.

7.1 Symmetric Wind Model Design

The algorithm for estimating TC wind fields from IKE assumes a simple Rankine vortex radial wind distribution (Depperman 1947). As discussed in Chapter 2, the RV equations define the wind speed at any radius inside an axisymmetric circulation to be a function of VMAX, RMW, and the scaling parameter, A. More specifically, the wind field increases linearly inside the eyewall from zero to VMAX at the RMW and it subsequently decreases exponentially outside of the RMW, such that

$$v(r) = \left(\frac{VMAX}{RMW}\right) r \qquad r < RMW$$
$$v(r) = VMAX\left(\frac{RMW}{r}\right)^{A} \qquad r \geq RMW. \tag{7.1}$$

In a pure Rankine vortex, the scaling parameter in the equation for winds outside of the RMW would be set to exactly one. However, in most cases other numbers are used to match observations and to account for, among other processes, angular momentum loss by friction (e.g. Holland et al. 2010). It should be noted that the Rankine Vortex equations presented above

are quite primitive in comparison to some of the more recent approaches to analyzing a wind profile of a TC (e.g. Holland 1980; Shapiro 1983; Vickery et al. 2000; Holland et al. 2010). Although these modern approaches improve upon the basic Rankine Vortex model, they also utilize complex equations that require a larger number of known variables. In an effort to limit the number of variables that are required as input parameters, the simplicity of the Rankine vortex works to our advantage.

Ultimately, the Rankine vortex based approach presented here is designed to estimate storm size using only known values of IKE, VMAX and RMW as inputs. Based upon equation 7.1, if VMAX and RMW are known, the scaling parameter becomes the only remaining unknown quantity needed to solve for the entire wind field. The missing scaling parameter can be approximated by fitting a Rankine vortex wind profile to match the known inputted value of IKE. This is accomplished by first generating hundreds of candidate RV wind projections. Each of these candidate RV projections uses the same VMAX and RMW values, but they each have a unique scaling parameter that is evenly spaced between zero and two. Integrated kinetic energy is computed for all of these individual candidate wind projections to find the most appropriate scaling parameter. In the end, the candidate wind projection that most closely matches the inputted known value of IKE is selected as the optimal Rankine Vortex wind projection in our IKE-based approach.

However, by allowing the scaling parameter to vary, it becomes possible to generate a RV wind field that is inertially unstable. Since relative angular momentum at any one point is equal to the product of the radius and the tangential wind speed, the equation for RAM in a Rankine Vortex will be equal to:

$$RAM(r) = \left(\frac{VMAX}{RMW}\right) r^2 \qquad r < RMW$$
$$RAM(r) = VMAX \cdot RMW^{A} \cdot r^{1-A} \qquad r \geq RMW. \tag{7.2}$$

Within the RMW, the flow will always be intertially stable, as both RAM and EAM are a function of r^2.On the other hand, it is possible for RAM to decrease with increasing radius outside of the RMW, if and only if the scaling parameter is greater than one. If the earth angular momentum does not offset the decreasing RAM in these cases, then absolute angular momentum will decrease with radius and the wind profile will be inertially unstable. Future work aims to restrict the scaling parameter such that an estimated storm profile cannot become inertially stable in any scenario. However, in this proof of concept exercise, the scaling parameter is allowed to make the RV profile inertially unstable over a small portion of the storm, in an effort to match the amount of observed IKE in a storm. Fortunately, inertially unstable profiles from the RV model are rare (< 10% of examined cases), so the results discussed in the following section should still be a somewhat accurate estimation of the wind radii model's future skill.

7.2 Evaluation of Symmetric Wind Model in a Perfect-Prog Mode

The potential skill of this wind field model is evaluated in this section using the large sample of historical storms from 1990-2011. As such the wind model is tested using historical VMAX and RMW values from the National Hurricane Center's best track dataset (Jarvinen et al. 1984) and estimates of historical IKE values derived from the operational wind radii within Colorado State's extended best track dataset (Demuth et al. 2006). This observation-driven approach is useful for determining how closely observed TC wind fields follow the idealized RV equations.

Ultimately, the optimal wind field produced by the symmetric RV model using historical IKE, VMAX and RMW data will be evaluated against historical tropical cyclone radii data from the extended best track dataset (Demuth et al. 2006). This dataset includes the radii of three specific wind speeds (34, 50, and 64 kts) over each of the four quadrants of a TC at six-hourly intervals throughout the life of all North Atlantic TCs since 1988. For the purpose of this study,

the historical wind radii data will be averaged radially across all of the four quadrants to focus on the axisymmetric size of a tropical cyclone for a direct comparison to the output of the axisymmetric RV model. Overall, the skill of the axisymmetric IKE-based wind model is evaluated against the historical record in 5498 North Atlantic tropical storm or hurricane fixes between 1990 and 2011 for each of the three operational wind radii.

It should be noted once again, that the baseline historical archive of wind radii is not perfect, as it contains significant uncertainties and biases, particularly for open ocean storms lacking aircraft reconnaissance. As discussed in previous sections, Landsea and Franklin (2013) estimated that each radii measurement from the 2010 Atlantic Hurricane season contains uncertainty that could exceed 30% of the mean size for each operational radius. Therefore, the evaluation of the wind radii model's skill with respect to the historical record will also be imperfect. However, given the large sample size, it is likely that this historical dataset will be at the very least sufficient to determine the adequacy and appropriateness of using a simple RV wind radii model to approximate storm size when given an approximate value of IKE, which once again is the goal of our observation-driven approach

An example of a TC wind field produced by the axisymmetric RV model for Hurricane Ivan in 2004 is shown by Figure 7.2. In this specific case, the observed intensity was 115kts, with a RMW of 25 nm. As explained in Section 7.1, the scaling parameter is allowed to vary until the IKE of the RV wind estimate approaches the observed IKE value of 137 TJ. The optimal scaling parameter for Ivan at this time was A=0.59, which resulted in an IKE value within the RV model of 137.1 TJ, nearly identical to the historical value. The resulting wind field estimate can then be used to obtain estimates for the operational 34-knot, 50-knot, and 64-knot radii (r_{34}, r_{50}, and r_{64} respectfully) for a direct comparison to observations. In this case, the estimated radii are quite accurate, as the wind model has a bias of just +12nm, -3nm, and -13.5nm compared to observed axisymmetric means of r_{34}, r_{50}, and r_{64} respectfully. Given the observed position coordinates of Hurricane Ivan at this time, the estimated operational radii can be plotted on a

map projection to assess the size of the wind field with respect to landmasses and other operational interests (Figure 7.3).

In a broader sense, the IKE-based wind model corresponds quite well to historical operational wind radii for all storms between 1990 and 2011. Figure 7.4 shows a scatter plot of estimates of 34-knot wind radii from the wind model plotted against the corresponding mean 34-knot wind radii from the extended best track. In this case, the RV model's estimates of 34-knot radii correspond to the historical record with a remarkable correlation of r=0.96. There are some significant outliers in Figure 7.4, (i.e. the points located furthest from the central dashed diagonal), wherein the model's estimate was off by more than 50 nautical miles (1 nm = 1.852 km) when compared to the historical record. Further analysis indicates the majority of these outliers occurred in less intense tropical storms (VMAX < 60 kts), particularly when the RMW was abnormally large (greater than 100 nm). This result indicates that less organized and weaker TCs are less likely to follow the theoretical Rankine Vortex wind profile. Therefore, the simplistic model is expected to have increased error in its estimates of wind radii for these less intense TCs, particularly for those that are observed to have their maximum sustained winds situated in a rain band well outside of the center of circulation.

Linear correlations and mean absolute error statistics for the wind radii model when compared to the historical wind radii for each of the three wind radii thresholds are shown in Table 7.1. The high correlations between the IKE-driven wind model's estimated radii and the historical radii indicates that the axisymmetric wind model excels at projecting TC wind fields when given accurate RMW, VMAX, and IKE data over the large sample size between 1990 and 2011. Furthermore, the mean absolute error of the wind model when compared to the historical record is relatively small for all three of the operational radii. For the 34-knot radii, the mean error in the large sample is just 11 nautical miles for r_{34}, compared to the mean historical r_{34} of 97.6 nm over the same large evaluation sample size. The mean errors of the higher wind speed radii are similarly low (< 10nm), but these radii are typically smaller to begin with, suggesting

that the idealistic RV model fares slightly worse for r_{50} and r_{64}. Nonetheless, the correlations between the estimates and the historical record are all near or above 0.90., suggesting that the model performs well at reproducing the variability of these radii given accurate input values.

Overall, the success of the idealistic wind model in a perfect-prog approach with historical data suggests that in most cases a Rankine vortex is appropriate for modeling symmetric wind fields, despite its simplistic nature. In fact, the mean biases and errors of this model are well within the magnitude of the uncertainty of the historical database to which it is compared. An advantage of this methodology is that the wind model does not use any sort of regression techniques to estimate a TC's wind field. Thus, when given values for IKE, VMAX, and RMW, the only assumption needed to approximate a TC's wind field with this approach is that the storm has an idealistic RV wind profile, which is generally a reasonable assumption for most well-organized TCs.

7.3 Considering Wind Field Asymmetries

Despite the IKE-based wind model's ability to accurately locate symmetric operational wind thresholds, the RV model still has some notable limitations. One of the most obvious shortcomings lies in the fact that a RV model can only produce estimates of the axisymmetrical wind circulation. However, analyses of observed TC wind plots over the past several years (e.g. H*wind; Powell 1998) easily confirm that real TCs are not remotely symmetric in nature. Instead, most TCs have complex wind fields that contain a wide range of asymmetric features. For instance, even a somewhat symmetrical intense hurricane, such as Hurricane Wilma (Figure 7.5), contains wave-like asymmetries within its inner core and some larger asymmetries in the outer portion of its wind field (i.e. gale force winds are skewed to the east).

Some of the asymmetries within a TC's structure can be resolved by considering other, more complex, vortex models (e.g. Vickery et al. 2000; Knaff et al. 2007). However, by

attempting to use these newer methodologies to approximate storm asymmetries, new unknowns and/or assumptions will be introduced into the equations. The main justification for using the RV wind model in the first place was to limit the number of assumptions that go into the model in an effort to produce a simple proof-of-concept wind projection from IKE. Although, a symmetric forecast is less than ideal, it is still operationally relevant, considering that NHC's verification of operational storm size forecasts is focused on radially symmetric radii as opposed to the radii within each individual storm quadrant (Cangialosi and Landsea 2014).

In an effort to include at least some asymmetric features into the SPIKE wind projections, it is possible to consider the relationship between translational storm movement and TC asymmetries (e.g. Shapiro 1983). Translational speed can easily be calculated by tracking the position of a storm's center through time and applying a center finite differencing scheme. Therefore, as a first attempt to capture TC asymmetries, a fraction (70%) of the TC's translational movement will be added to the axisymmetric RV wind field across the entire storm domain. In doing so, the resulting wind field projection will exhibit stronger winds to the right of the translational motion vector, where the translational motion has an additive effect to the axisymmetric wind circulation. As the translational motion of a TC increases, the skewness of the storm's wind field will also increase. On the other hand, if the storm is stationary, the storm's wind field will be identical to the axisymmetric projection produced by a RV.

A sample asymmetric wind field is produced using this methodology in Figure 7.6 for Hurricane Wilma on October 24, 2005 at 1800Z. At this time, Hurricane Wilma had a maximum intensity of 105 kts and an IKE value of approximately 121 TJ. The storm was traveling 15 kts in a northeast direction, causing the wind field in the sample projection to skew towards the southeast. Observational analyses (not shown) encouragingly confirm that Wilma contained its strongest winds in its southeast quadrant as predicted by the asymmetric model. However, the asymmetric simulation still fails to capture some of the other more complex asymmetric features observed throughout the storm.

Unfortunately, some other problems arise by simply adding a fraction of the translational wind speed to the axisymmetric RV circulation. For instance, the Rankine Vortex equations utilize observed values of VMAX and IKE to produce an estimate of the axisymmetric wind circulation. However, in reality, these intensity and energy quantities contain a contribution from the storm circulation and another contribution from the translational motion of the TC. Therefore, they cannot be used to accurate measure the strength of just the axisymmetric circulation in the RV model.

For example, if the observed intensity is given to the RV equations, then the maximum wind speed of the axisymmetric circulation will be equal to VMAX. This maximum sustained wind value will occur in a ring located at the RMW in all directions from the storm's center. Once the translational speed of the storm is added to the axisymmetric vortex, wind speeds will be adjusted upwards on the right side of the storm relative to storm motion. Therefore, the asymmetric wind field will contain wind speeds to the right of the eye that erroneously exceed the observed VMAX value. Similarly, the IKE of the asymmetric wind field will be too high since translational motion has a net increase on windspeeds around a storm, particularly to the right of the storm motion vector.

Ultimately, if the asymmetric wind model overestimates VMAX and IKE, the resulting projections of the various radii will also be artificially too large. Overestimating VMAX will cause the asymmetric model to overestimate the inner core radii (i.e. r_{64}), and overestimating IKE will result in the outer core being too large as well. This systematic high bias in each of the operational wind radii is plainly evident when producing asymmetrical wind estimates for all of the storms between 1990 and 2011. For instance, the 34-knot radii are larger than their observed counterparts in each quadrant by an average of 20-30 nm.

Correcting the systematic biases present in the asymmetric model is not necessarily a simple process. On one hand, it is trivial to account for the translational motion's impact on VMAX by simply subtracting the translational speed from the VMAX input that goes into the RV

model, such that the RV model's maximum wind speed will be adjusted downward to compensate for the contributions of the translational storm motion. On the other hand, there is no easy way to account for the amount of IKE added to the TC by storm motion. The contribution to IKE by the translational speed appears to be a quadratic function based on the magnitude of the translational speed. This function is also affected by the magnitudes of RMW, VMAX, and IKE, making it nearly impossible to consistently compensate for the translational motion's contribution to IKE. Despite the desire to produce asymmetric wind fields, these biases are too significant to overcome for the purposes of this proof-of-concept study. Therefore, the remainder of this chapter will focus exclusively on the symmetrical Rankine Vortex model (without considering the translational motion of the storm) in an effort to see if IKE can be used to project the sizes of symmetric operational wind radii in a mock-operational environment. However, by all means, future work should focus on attempting to solve the issues discussed in the past few paragraphs so that IKE can be used to produce more useful asymmetrical wind projections in the future.

7.4 Assessment of Operational Skill in Symmetric Wind Model

The analysis in this section uses the IKE-based axisymmetric RV wind model in a hindcast mode to evaluate the expected skill of the simple wind model in an operational setting. As discussed in Section 7.2, the axisymmetric RV wind model performs very well when given historical data in a perfect prognostic approach. However, the operational version of this same RV wind model will have less skill when predicting storm size in real time, simply because it will be dependent upon imperfectly predicted values of RMW, VMAX, and IKE.

Of the three required input parameters, VMAX forecasts are widely available, and IKE forecasts can be obtained from SPIKE2. Conversely, there are few reliable existing methods to forecast the RMW, considering that many operational models lack the high resolution necessary

to accurately resolve the convective features of a hurricane eyewall. Therefore, an operational version of the wind radii tool will be limited to simple persistence RMW forecasts. Thankfully, short-term changes of RMW are historically small, and as a result, RMW persistence values at the initialization time can be used to force the RV wind model without a significant loss of skill. In fact, the average error of the IKE-based RV model's radii estimates only increases by one nautical mile for r_{34} and 4 nautical miles for r_{64} over the 1990 to 2011 sample, when using 72-hour persistence forecasts of RMW instead of the actual historical value. Ultimately, imperfect persistence RMW values should not entirely derail a wind radii forecast using the RV model, except in rare circumstances when a TCs eyewall rapidly expands or contracts.

Building upon these results, a series of sensitivity experiments are run to evaluate the impacts of using forecasted VMAX and IKE inputs on the RV wind model, when already given persistence RMW values. In total, there will be seven sensitivity experiments using various combinations of observations, persistence, and hindcast datasets for the VMAX and IKE input values (Table 6.2). These sensitivity tests are once again run using storm fixes from 1990 to 2011. Since some of the experiments require GEFS reforecasted data, the sample will be restricted to storm fixes with valid GEFS reforecast data. Therefore, the sample size for the following sensitivity studies will be about 25% of the sample size used in the perfect prognostic exercises in Section 7.2 because the GEFS control run reforecasts are only initialized once daily at 00Z.

A) Persistence Sensitivity Tests

In the persistence sensitivity runs, the observational input values used in the control run are replaced with 36-hour persistence data for either VMAX or IKE (Table 7.2). The mean absolute error and correlation statistics for each of the sensitivity runs are shown in Table 6.3 for a 36-hour forecast interval. As anticipated, using a simple 36-hour persistence forecast for IKE

or for VMAX has an apparent detrimental effect on the accuracy of the RV wind model, all else being equal.

The results of the "Persist IKE" sensitivity run indicates that IKE forecast errors negatively impact the ability of the RV model to predict a TCs outer wind field. For instance, the mean absolute error for r_{34} in the Persist IKE run is 31.68 TJ, which is nearly three times as large as the mean error for the control run (Table 7.3). This error is equal to about 30% of the mean 34kt radii in these storms, which is still well within the uncertainty range of the observations themselves (Landsea and Franklin 2013). The correlation between the predicted r_{34} and the observed r_{34} similarly decreases from 0.97 in the control run to 0.70 in the Persist IKE run. On the other hand, the introduction of imperfect IKE input values does not drastically increase the mean r_{50} or r_{64} errors. In fact, the mean absolute error and the correlations for r_{50} and r_{64} in the Persist IKE run are both comparable to those of the control run. Overall, these results suggest that a good forecast of IKE is necessary to accurately predict the radius of tropical storm force winds. On the other hand, forecasts of the inner radii do not seem to be similarly degraded by the inaccuracy of a persistence IKE forecast.

Instead, r_{50} and r_{64} seem to be more sensitive to the accuracy of the VMAX input variable. For example, the mean absolute error for those two radii increases by more than 30% in the "Persist VMAX" run compared to the control run (Table 7.3). The 18.61 nm average error for r_{50} is equal to about 50% of the average observed 50-knot radii, and the 13.40 nm average error for r_{64} is roughly equal to the historical average r_{64} value for all TCs between 1990 and 2011. The observed standard deviations for these more intense radii are actually larger than their mean values, indicating that the inner radii fluctuate quite substantially from storm to storm. This large observed variability of r_{64} is not surprising considering that tropical storms will not have hurricane force winds at all, and other larger storms can have hurricane force winds extending more than 100 nm out from its center. Nonetheless, the RV model still captures more than 42% of the variance for r_{64} (r=0.65) and more than 57% of the variance for r_{50} (r=0.76) when given

imperfect VMAX data from a persistence forecast. Overall, however, it remains very clear that the inner core of a TC's wind field cannot be resolved correctly by the RV model if the inputted intensity values are inaccurately forecasted. It should be noted that a 36-hour persistence forecast is not a good intensity forecast, as TCs often undergo drastic intensity changes in short periods of time. Therefore, better intensity forecasts should lead to improved RV model skill.

As all of these persistence sensitivity tests indicate, the RV model's sensitivity to IKE forecast error is quite different from its sensitivity to VMAX forecast error. To explain the different characteristics of these sensitivities it is necessary to recall the Rankine Vortex equations that govern the wind model (Equations 7.1). Ultimately, the VMAX input variable controls the magnitude of the peak of the distribution, and the IKE input variable controls the scaling factor that governs how quickly winds fall off outside of the radius of maximum sustained winds. Based on these equations, the winds fall off most rapidly closest to the strongest winds near the RMW. Therefore, r_{64} is typically restricted to the area immediately surrounding the RMW. If VMAX is underestimated, the peak of the wind field is lower and the entire inner core is underestimated accordingly. To accommodate the decreased area of intense winds in this scenario, the RV model slightly overestimates the area of tropical storm force winds to maintain the same amount of IKE. On the other hand, if IKE is underestimated, the winds field profile has a sharper radial distribution. As a result, winds will decrease too rapidly in the outer portion of the wind field, and r_{34} is greatly underestimated. r_{50} and r_{64} are also underestimated in this scenario, but those radii are more closely tied to the accuracy VMAX and the peak of the wind field.

This discussion can be visualized in Figure 7.7 using a sample hurricane with 60 TJ of IKE and an intensity of 100 kts. Given these inputs, and a RMW of 35 nm, the RV model produces the wind field shown in the top plot. As IKE is perturbed up or down by 20 TJ, the 34-knot wind field changes substantially, as seen by the size of the blue shaded area in the second row of plots. On the other hand, if VMAX is perturbed upwards by 20kts, r_{64} increases as seen by the

larger area of red in the bottom left plot. However, if VMAX is perturbed down by 20kts, the area of hurricane force winds decreases dramatically and the area of tropical storm force winds increases to maintain the 60 TJ of IKE. Ultimately, this figure supports the notion that accurate VMAX inputs are needed to correctly assess the inner core of a TC's surface wind field, and that accurate IKE inputs are needed to correctly assess the outer extent of the tropical storm force winds.

B) Hindcasted Sensitivity Tests

The sensitivity runs analyzed in this section, utilize VMAX and IKE input values from a series of 36-hour hindcasts. The IKE input data utilized by the RV model in the "Hindcast IKE" sensitivity run is taken directly from the 36-hour mock-operational version of SPIKE2 (Table 7.2). As discussed in the previous section, the SPIKE2 projections of IKE are more accurate than a persistence IKE forecast. Therefore, one would expect that the improved input would enable the RV model to perform more skillfully in the Hindcast IKE run than in the Persist IKE run. For the most part, this expectation is met.

Some results of the Hindcast IKE run are shown in Table 7.4 along side of the results of the control run. Compared to the Persist IKE run, mean absolute error statistics are lower and correlations improve for both r_{34} and r_{50} in the Hindcast IKE run. Nonetheless, the RV model still has mean absolute r_{34} errors of about 25 nm when given IKE values from SPIKE2, which is still double those produced by the RV model in the control run (Table 7.4). Overall, this 25 nm mean error for the 34-knot radius in the Hindcast IKE run is roughly comparable to the biases and errors that are found in the historical wind radii dataset. Thus, it may be unrealistic to expect a simple forecast model to have less error than the operational and post-storm analyses themselves.

Interestingly, upgrading the IKE inputs from persistence to hindcasted values from SPIKE2 does not seem to significantly change the wind model's ability to forecast r_{64}. The correlations for r_{64} in both the Hindcast IKE run and the Persist IKE run are nearly identical to those for the control run. Similarly, all three runs have mean absolute errors for r_{64} between 9 nm and 10 nm. As discussed previously, the accuracy of the 64-knot radius is more sensitive to the accuracy of VMAX than it is to the accuracy of IKE. However, the results of the Hindcast IKE and Persist IKE runs suggest that the RV model's ability to project r_{64} in a hurricane is barely impacted by most normal IKE forecast errors at all.

The next set of hindcast sensitivity runs focuses on the relationship between the VMAX inputs and the radii outputs within the RV model. The VMAX data used in the "Hindcast VMAX" sensitivity run is taken from the one-degree resolution GEFS reforecast dataset (Table 7.2). Unlike the IKE hindcast data, the VMAX values from the GEFS reforecast dataset are quite inaccurate. The coarse resolution of the GEFS dataset precludes the model from resolving the high wind speeds within the inner core of a TC. As a result, the GEFS systematically underestimates the maximum sustained winds contained in Atlantic TCs by an average of 30 kts. Figure 7.8 serves as visual evidence for this low bias, by plotting raw 36-hour VMAX reforecasts from GEFS against the corresponding best track VMAX values for all available storm fixes in the 1990-2011 sample. Inputting these raw GEFS intensities into the RV model (as is done in the Hindcast VMAX run) will enable the extreme biases in the VMAX dataset to artificially ruin any wind radii projections. As a result, the RV model unsurprisingly performs quite terribly in the Hindcast VMAX sensitivity run (Table 7.4).

The main goal of these sensitivity tests is to analyze the operational biases of the RV wind radii model, not the intensity biases of the GEFS reforecast dataset. Therefore, an additional sensitivity test is performed for VMAX, wherein the GEFS intensity input values are bias corrected to have the same mean and standard deviation intensity as the best track dataset from 1990 to 2011. The improved VMAX input values enable the RV model to perform

substantially better in this "Corrected VMAX" run when compared to the Hindcast VMAX run. For example, the r_{34} mean errors are about 15 nm, which is only slightly higher than the average r_{34} errors found in the "Control" run. Similarly, the r_{34} correlations (r=0.92) are also very close to those found in the "Control" run.

The r_{50} and r_{64} statistics are also improved in the Corrected VMAX run compared to the Hindcast VMAX run. However, in both VMAX sensitivity runs, the mean errors for r_{50} and r_{64} are nearly double those found in the control run. Based on these results, it is obvious that the GEFS simply does not produce sufficiently accurate VMAX reforecast for the RV model. Even after the bias correction, the GEFS reforecasts still have an average intensity error of 20 kts. At this magnitude, the GEFS will be off by at least one Saffir Simpson category on average at this short 36-hour forecast interval. Despite these large errors, the RV model still performs well for r_{34}. With the ingestion of better intensity guidance, it is likely that the performance of the RV model's r_{50} projections, and to a lesser extent its r_{64} projections, will improve.

The final sensitivity run combines the IKE inputs from SPIKE2 and the VMAX inputs from the bias corrected GEFS output. This "Mock-Operational" run similarly suffers from the poor accuracy of the VMAX input values, but it is nonetheless important to see how imperfect VMAX data and imperfect IKE data combine to affect the accuracy of the RV model. Overall, the Mock-Operational run helps to confirm previous hypotheses. The introduction of poorly forecasted VMAX inputs does not seem to affect the RV model's ability to forecast r_{34}, when compared to the Hindcast IKE run. Likewise, the introduction of imperfectly forecasted IKE data does not seem to negatively affect the RV model's ability to forecast r_{50} and r_{64} data, when compared to the Hindcast VMAX run. Therefore, these results once again support the notion that the IKE input is critically important to the accuracy of the r_{34} projection and the VMAX input is critically important to the accuracy of the r_{50} and r_{64} projections in the RV model.

Ultimately, in its current state the low resolution GEFS hindcast dataset negatively affects both the VMAX input parameters and to a much lesser extent the IKE projections from SPIKE2.

Therefore, when these imperfect input values are given to the RV model, the hindcasts of the operational radii contain undesirably high levels of mean error. Ideally, the mean absolute error statistics in the Mock-Operational run would be closer to the control run," but instead they are near or above 20nm for r_{34}, r_{50}, and r_{64}.

On the other hand, these results indicate that the RV model should have room to improve towards the higher performance of the control run. To achieve that higher level of performance, it will be necessary to pair an operational model with reasonably accurate intensity projections with SPIKE2. Therefore, it is abundantly clear that additional testing using higher resolution models is needed to determine whether or not this product has any potential for real time use.

C) Comparison of Hindcasts to National Hurricane Center Radii Forecasts

Although the performance within the "Mock-Operational" hindcast runs of the RV model seems low compared to the control run, it is worthwhile to compare these wind radii hindcasts to real operational forecasts produced by the National Hurricane Center. Considering the lack of emphasis on forecasting storm size, the GEFS-SPIKE wind radii projections may still be competitive with existing forecast guidance. If this is the case, the IKE-based RV tool would be a very desirable operational tool, especially considering its potential for improvement by using higher-resolution model data to drive the wind model and by possibly adding asymmetric features to the simple RV wind field in the future.

As discussed in Section 6.3, NHC projects storm size in its forecast advisories, which are traditionally issued four times daily: at 03Z, 09Z, 15Z, and 21Z. The 34-knot and 50-knot radii are projected for each quadrant of the storm out to 69 hours beyond the issuance time. Since the radii in these NHC forecasts are not symmetric around the center of the storm, they cannot be directly compared to the skill of the axisymmetric wind model. Therefore, the NHC wind

forecasts must first be radially averaged by taking the mean of the four quadrants. Doing so allows for a better comparison with the axisymmetric results from IKE-based RV model.

The NHC forecasts are compared against observations and the “mock-operational” GEFS-SPIKE2 wind radii projections in Table 7.5 for selected Atlantic TCs. To focus on societally impactful storms, only billion dollar landfalling United States hurricanes between 2003 and 2011 were analyzed in this comparison. In total, ten storms were selected: Hurricanes Isabel (2003), Frances (2004), Ivan (2004), Jeanne (2004), Katrina (2005), Rita (2005), Wilma (2005), Gustav (2008), Ike (2008), and Irene (2011). For each of these storms, the validation time is set as 00Z on the day of landfall. The NHC forecasts are taken from the advisory issued at 03Z three days prior to the validation time, and the 72-hour GEFS-SPIKE2 RV hindcasts are initialized three hours prior to that at 00Z. Therefore, the comparison will place a 69-hour NHC forecast against the 72-hour GEFS-SPIKE hindcasts at a time when an accurate three-day storm size forecasts could be significantly helpful to emergency managers as they plan evacuations ahead of an impending hurricane watch and/or warning. Before making this direct comparison, it is important to note that an operational version of SPIKE2 using the operational GEFS model will not actually be issued at 00Z when the GEFS is initialized. Unfortunately, modern dynamical weather prediction models take hours to run, and therefore it is distinctly possible that a 00Z SPIKE2 forecast will not be available in advance of the NHC’s forecast advisory at 03Z. Nonetheless, measuring the 00Z GEFS-SPIKE2 wind radii hindcasts against the 03Z NHC advisories is a good first order comparison of the operational relevancy of the SPIKE2 wind radii product.

Somewhat surprisingly, the GEFS-SPIKE2 wind radii hindcasts outperform the NHC forecasts for nearly all of the ten landfall examples. Considering the poor quality of the GEFS data used in the "mock-operational" RV runs, the NHC forecasts were expected to be the runaway winner in this comparison. However, in nine out of ten cases, the RV model made a better three-day prediction of r_{34} than did the NHC forecast advisories (Table 7.5). In total, for

these ten landfalling cases, which together caused nearly a quarter of a trillion dollar in damage, the "mock-operational" RV model runs had a mean absolute error of 15nm, compared to 42nm of mean error by the NHC forecasts. For all ten storms, the NHC forecasts of tropical-storm force wind radii were far too conservative, as they systematically fell below the observed r_{34} value by more than 10 nm in each of the ten 69-hour forecasts. In contrast, the 72-hour "mock-operational" RV hindcasts tended to overpredict r_{34} by an average of only 6nm.

The comparison between NHC forecasts and the SPIKE2 RV Mock-Operational hindcasts were much closer in terms of the 50-knot operational radii. Nonetheless, the GEFS-SPIKE2 hindcasts were more accurate than NHC in six out of ten cases (Table 7.5). The RV model's mean absolute hindcast error at 72-hours was 23nm for these ten storms, compared to a mean absolute error of 30nm by NHC. Overall, both NHC and the GEFS-SPIKE2 model tended to underestimate the radius of 50-knot winds. In the case of the RV hindcasts, the poor intensity reforecasts from the low-resolution GEFS model undoubtedly played an effect in the poorer performance for the inner radii.

Looking more closely at individual storms, the GEFS-SPIKE2 model performed the best for Hurricane Isabel, prior to its landfall in the Mid-Atlantic United States. On September 19, 2003 at 00Z, Isabel had tropical storm force winds out to 231nm and 50-knot winds out to 114nm on average. Amazingly, the 72-hour hindcast from the Mock-Operational RV run nearly matched the observations with a prediction of 226nm for r_{34} and 110nm for r_{50}. The 03Z NHC advisory on 9/16/03 was also quite good for this storm, as it pinned the 50-knot radii within 15nm at the verification time, but it fell just short of the high benchmark set by the GEFS-SPIKE2 model.

In contrast, NHC's worst storm size forecast was for Hurricane Katrina, which coincidentally became the costliest and deadliest United States hurricane in recent memory. For instance, the 03Z NHC forecast advisory on August 26, 2005, predicted that Katrina would have tropical-storm force winds extending out only 75nm from the center at 00Z on the 29th. In reality Katrina's 34-knot wind field extended beyond 183 nm out from the center at that time. In

comparison, a 72 hour hindcast from the "mock-operational" GEFS-SPIKE2 wind model also underestimated the size of Hurricane Katrina prior to its historic landfall, but its predicted r_{34} and r_{50} values were substantially closer to the observed value (Table 7.5).

Overall, this exercise suggests that the GEFS-SPIKE2 RV model may be a useful addition to the NHC's guidance package, even in its current state with the low-resolution GEFS input parameters. Ten storms is a very small sample size, and as such, it is impossible to determine whether or not the IKE-based radii model can consistently compete with the NHC forecast advisories in a 72-hour forecast window, especially considering that the hindcasts do not attempt to account for the time constraints present in operational real-time forecasting. Nonetheless, these ten storms represent some of the most impactful storms in the past two decades, and any tool that could potentially better predict the size of these storms should be investigated further. Encouragingly, the performance of this RV model can likely be improved upon in future (especially for r_{50} and r_{64}) by pairing it with a higher resolution model that is more capable of accurately predicting observed intensity values than the low-resolution GEFS data used here. Ultimately, continued developmental support of this wind radii model is strongly recommended in an effort to assist forecasters and emergency managers as they attempt to determine the size of landfalling tropical cyclones.

Table 7.1: Correlation and mean absolute error statistics for the axisymmetric wind model using observed VMAX, RMW, and IKE data to find the optimal scaling parameter in the RV formulas. The statistics are calculated over 5498 storm-fixes from 1990 through 2011, and the wind model is evaluated against symmetrical means of the observed radii data from the extended best track data.

	Correlation	Mean Absolute Error
r_{34}	0.96	11.4 nm
r_{50}	0.89	9.9 nm
r_{64}	0.90	5.1 nm

Table 7.2: List of the sensitivity experiments performed with the RV wind radii model. All seven of the experiments utilize persistence RMW data in addition to the listed VMAX and IKE data to produce projections of operational wind radii. For consistency, these experiments are run for storm fixes that have available GEFS data between 1990 and 2011. The persistence and GEFS data is calculated for a lead time of 36 hours.

Control Experiments	**VMAX**	**IKE**
1. Control	Observations	Observations
Persistence Experiments	**VMAX**	**IKE**
2. Persist IKE	Observations	Persistence
3. Persist VMAX	Persistence	Observations
Hindcast Experiments	**VMAX**	**IKE**
4. Hindcast IKE	Observed	SPIKE2
5. Hindcast VMAX	GEFS Hindcasts	Observed
6. Corrected VMAX	Bias Corrected GEFS	Observed
Mock-Operational Experiments	**VMAX**	**IKE**
7. Mock-Operational	Bias Corrected GEFS	SPIKE2

Table 7.3: Correlations and mean absolute error statistics for the control and persistence wind radii model experiments run on a 36-hour forecast interval. These statistics are calculated for three operational radii threshold for storm fixes between 1990 and 2011.

a) Correlations

	R34	R50	R64
Control	0.97	0.82	0.77
Persist IKE	0.70	0.77	0.75
Persist VMAX	0.96	0.76	0.65

b) Mean Absolute Error

	R34	R50	R64
Control	11.53	14.23	9.58
Persist IKE	31.68	16.95	9.38
Persist VMAX	13.81	18.61	13.40

Table 7.4: Correlations and mean absolute error statistics for the control and hindcast experiments run on a 36-hour forecast interval. These statistics are calculated for three operational radii threshold for storm fixes between 1990 and 2011.

a) Correlations

	R34	R50	R64
Control	0.97	0.82	0.77
Hindcast IKE	0.81	0.79	0.76
Hindcast VMAX	0.80	0.64	0.41
Corrected VMAX	0.92	0.66	0.54
Mock-Operational	0.79	0.61	0.52

b) Mean Absolute Error

	R34	R50	R64
Control	11.53	14.23	9.58
Hindcast IKE	25.98	15.44	9.82
Hindcast VMAX	43.36	39.02	20.88
Corrected VMAX	15.06	22.74	19.08
Mock-Operational	24.86	23.57	17.80

Table 7.5: Comparison between the mock-operational axisymmetric GEFS-SPIKE2 wind model hindcasts and archived NHC forecast advisories. The hindcasts and forecasts are verified at 00Z on the same day when each hurricane makes a United States landfall. The damage produced by each storm is also listed, as estimated in 2015 US Dollars by the ICAT Damage Estimator (http://www.icatdamage estimator.com/). The SPIKE2 hindcasts are initialized 72 hours prior to the validation time, and the NHC Forecast advisories were issued 69 hours prior to the validation time. The observations and the NHC forecasts were radially averaged across the four quadrants to produce an estimate of axisymmetric radii for a direct comparison to the output axisymmetric SPIKE2 wind model.

Radii Skill Comparison in Landfalling Hurricanes

Storm Details				Mean Observations		GEFS-SPIKE2 72hr Hindcast		Mean NHC 69hr Forecast	
Name	Valid Time	Landfall State	Damage ($ 2015)	R_{34}	R_{50}	R_{34}	R_{50}	R_{34}	R_{50}
Isabel	9/18/03 00Z	NC	5.6 Billion	231	114	226	110	188	125
Frances	9/5/04 00Z	FL	14.3 Billion	155	109	176	78	125	88
Ivan	9/16/04 00Z	AL	21.3 Billion	188	113	211	92	150	100
Jeanne	9/26/04 00Z	FL	12.2 Billion	151	88	157	66	125	65
Katrina	8/29/05 00Z	LA	88.4 Billion	183	104	162	67	75	33
Rita	9/24/05 00Z	LA	11.3 Billion	145	98	163	83	135	73
Wilma	10/24/05 00Z	FL	29.3 Billion	175	108	195	92	106	50
Gustav	9/1/08 00Z	LA	4.8 Billion	175	100	157	59	128	75
Ike	9/13/08 00Z	TX	23.1 Billion	190	120	199	84	163	94
Irene	8/27/11 00Z	NC	7.6 Billion	183	104	196	94	165	73

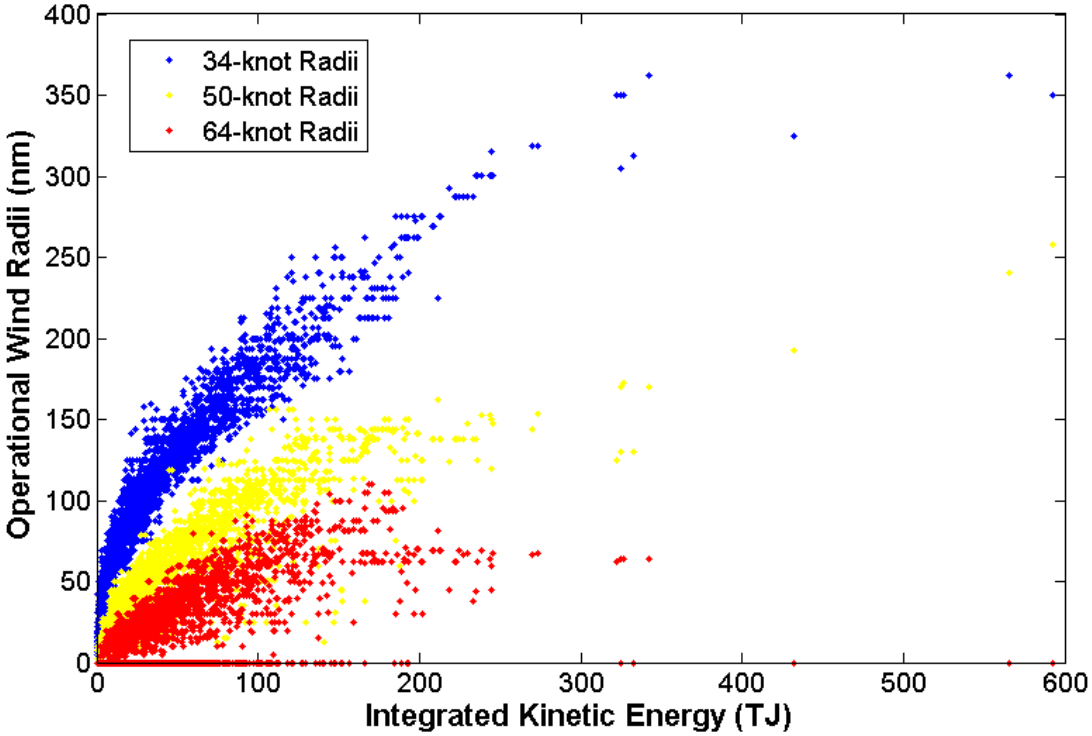

Figure 7.1: Scatter plot of historical operational wind radii (nm) versus historical integrated kinetic energy (TJ) in 5498 six-hourly storm fixes between 1990 and 2011. As expected, IKE increases with increasing storm size. It should be noted that storms with intensities below 50 and 64 knots will have 50-knot and 64-knot radii set to exactly zero.

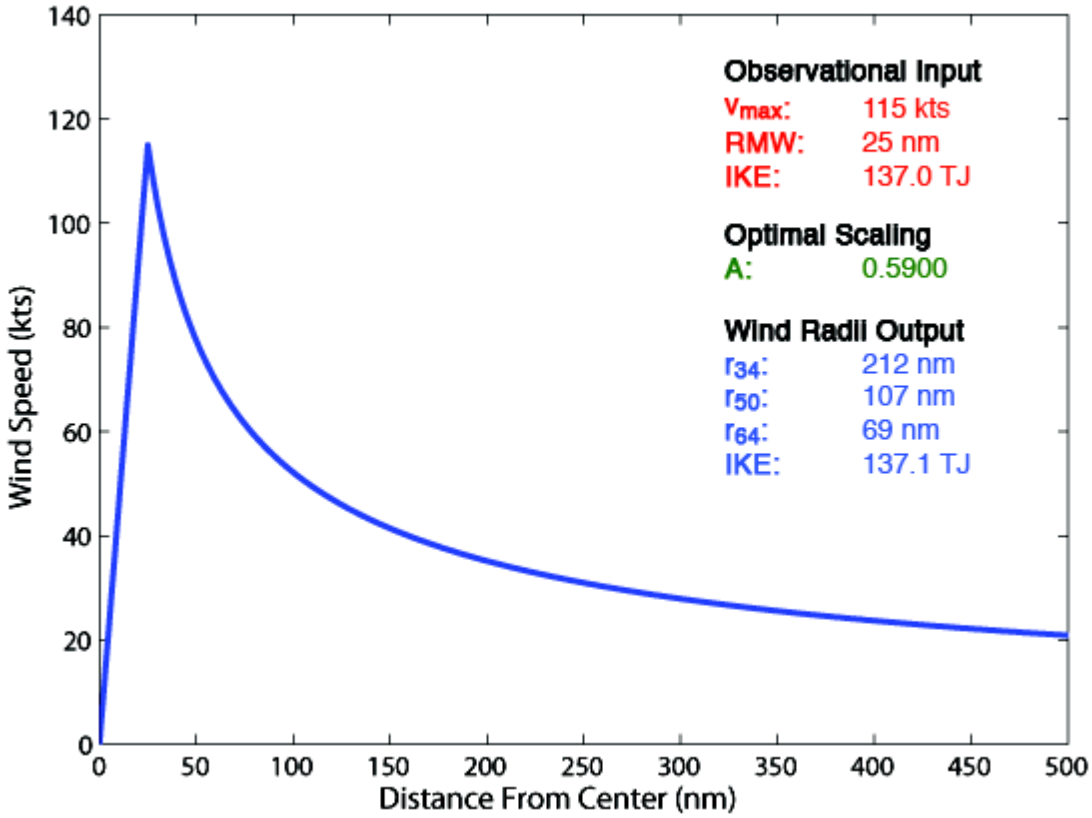

Figure 7.2: Sample wind field produced by the symmetric wind model from Hurricane Ivan in September of 2004. In this case, the observed VMAX and RMW data (red text) are fed into the simple Rankine Vortex equations. The scaling parameter, A, is allowed to vary until it reaches an optimal value (green text) wherein the integrated kinetic energy of the estimated vortex most closely resembles the observed IKE levels. In this case the optimal scaling parameter is A=0.59. Given this scaling parameter, the wind field of the storm is estimated with respect to the distance from the storm center (blue curve). Estimates from the operational wind radii can then be obtained from the plot (blue text).

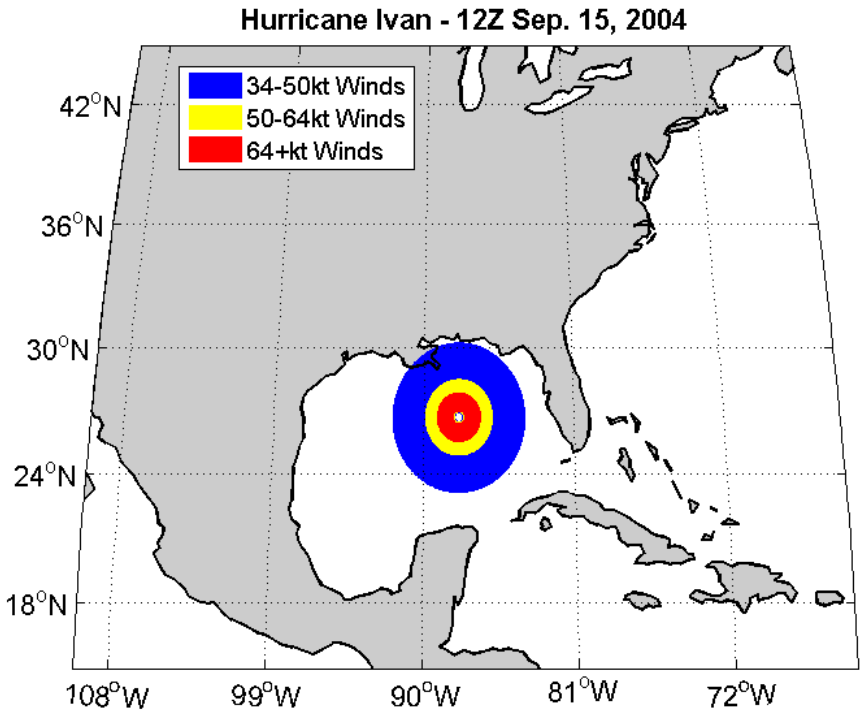

Figure 7.3: Estimated symmetric wind field of Hurricane Ivan on September 15, 2004 obtained from the RV wind model. To produce this plot, the same estimated wind field in Table 7.2 is projected onto a map of the basin, given observed latitude and longitude coordinates

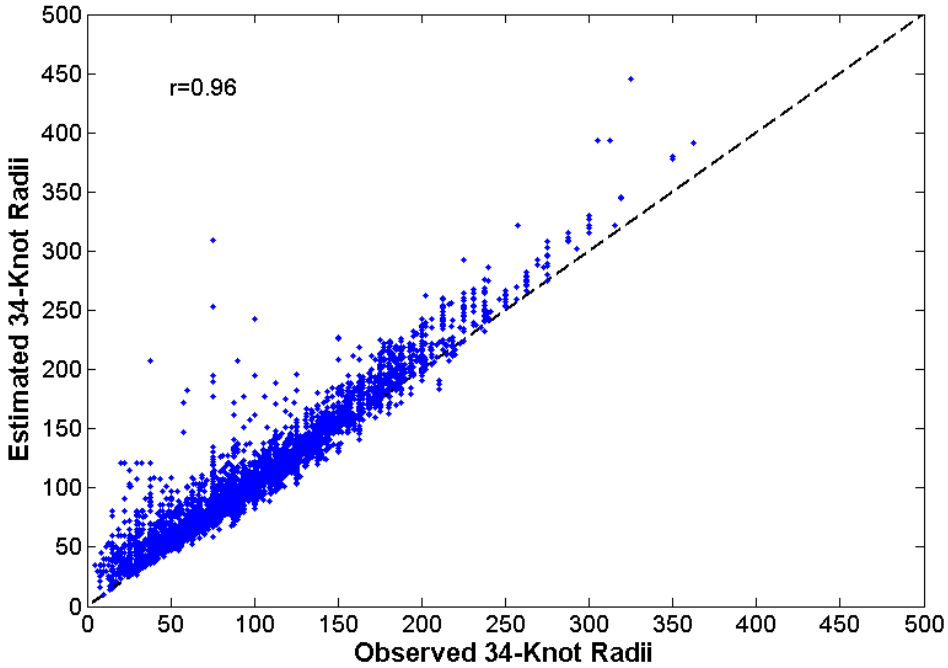

Figure 7.4: Scatter plot comparing observed 34-knot wind radii (x-axis) to estimated 34-knot wind radii from the symmetric RV wind model (y-axis) for 5498 storm fixes between 1990 and 2011. The observed wind radii plotted here represents a radial symmetrical mean such that the observed quantity is directly comparable to the axisymmetric estimates from the RV model. The estimates are correlated to the observed radii at r=0.96. A perfect fit would fall on the dashed black line (y=x).

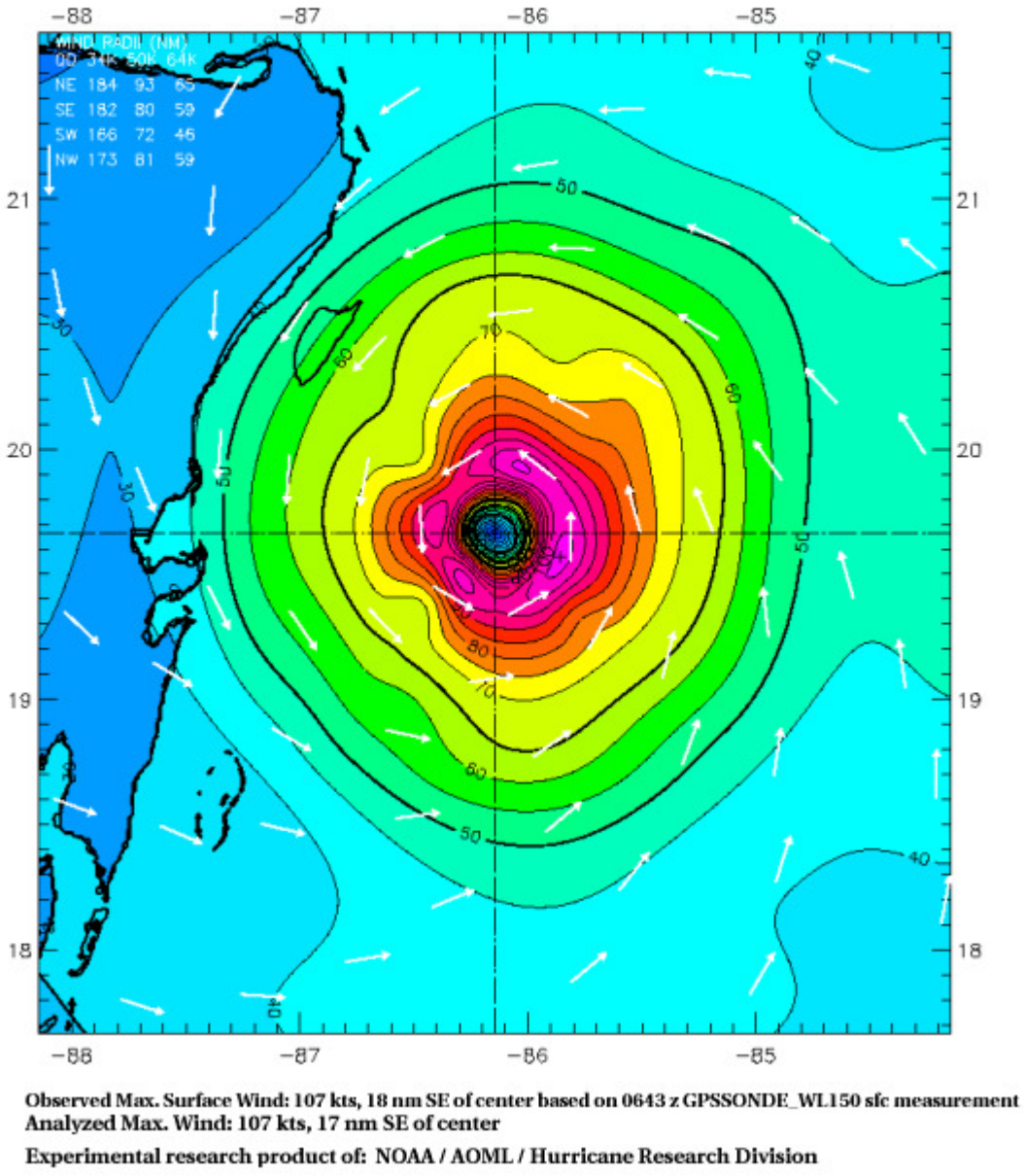

Figure 7.5: Example of TC asymmetries as shown by the H*wind analyses of Hurricane Wilma on 10/21/2005 at 0730Z. Legacy H*wind products are made available online by NOAA and H*wind Scientific.

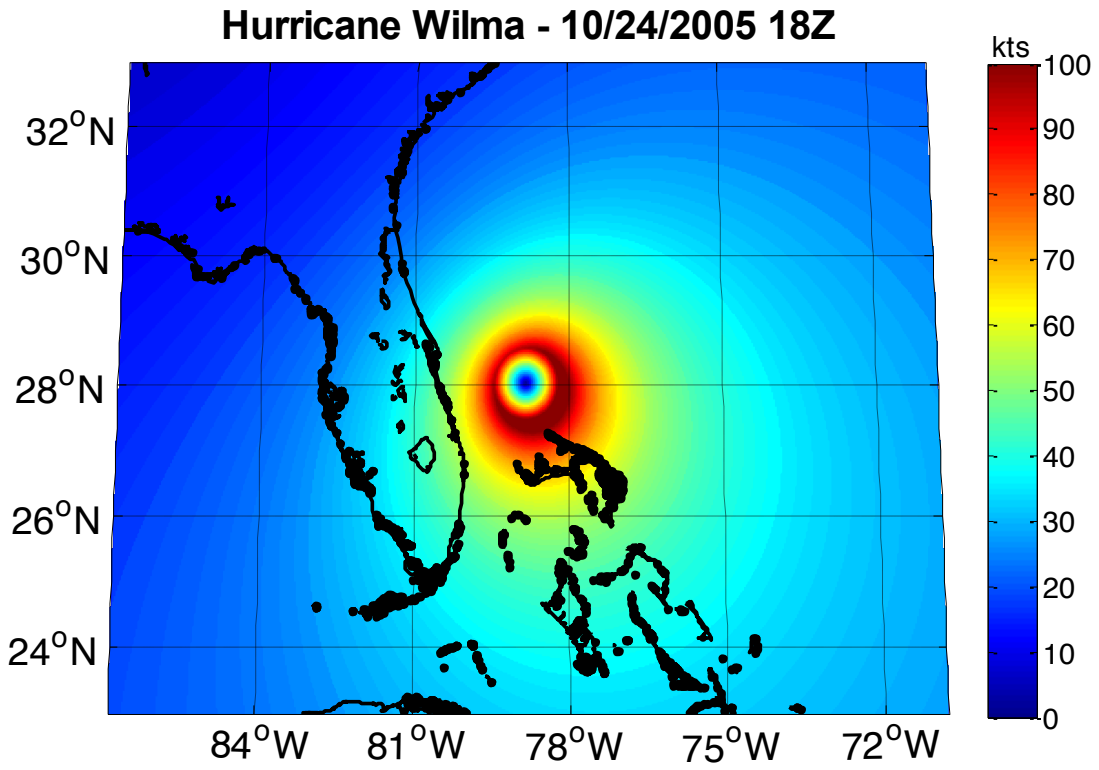

Figure 7.6: An example of an asymmetric SPIKE2 wind model product that combines a portion of the translational storm motion with the axisymmetric Rankine Vortex circulation.

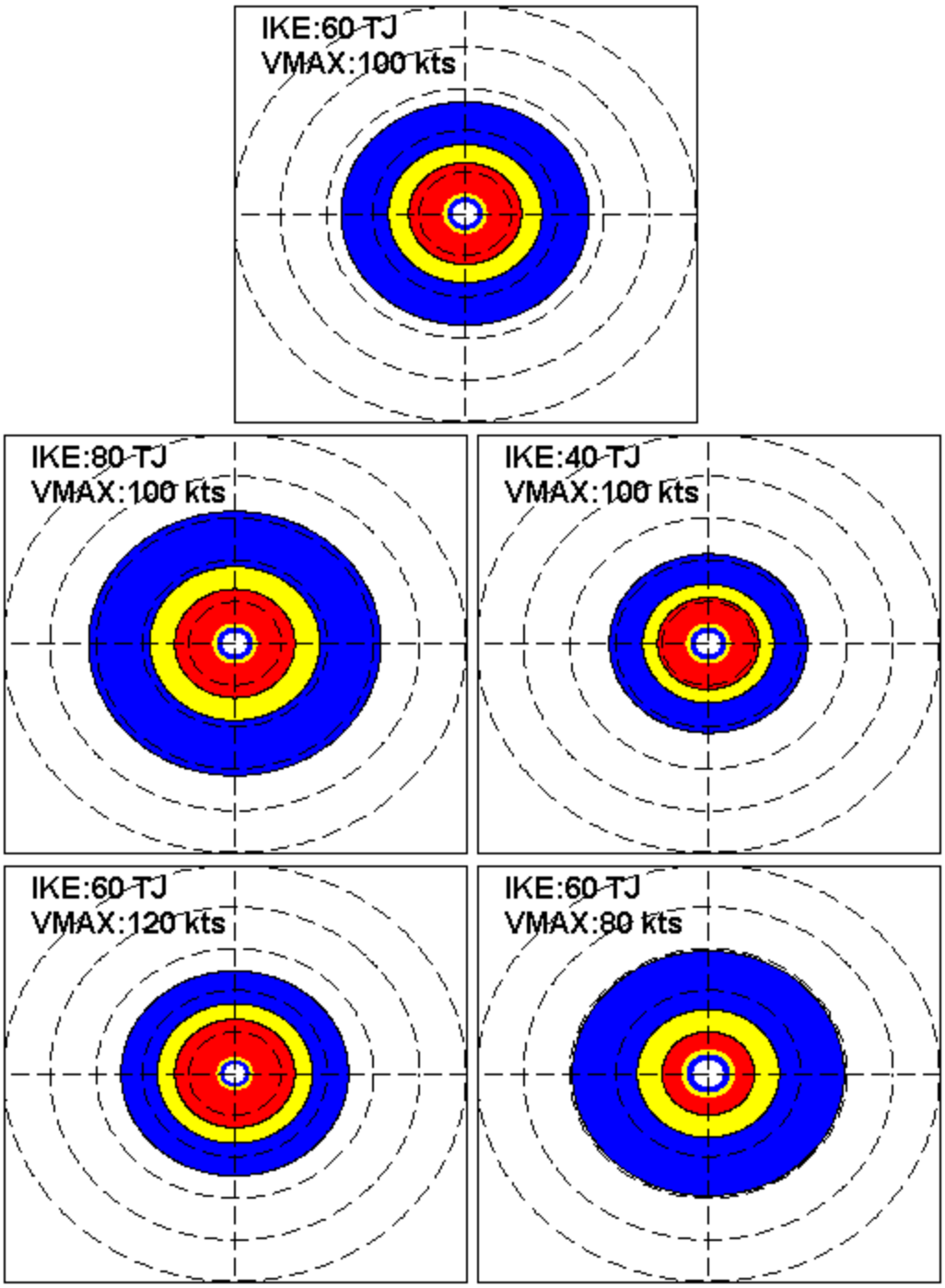

Figure 7.7: Sample wind radii plots designed to show how perturbations of VMAX and IKE affect the results of the RV model. In each panel the blue circle indicates winds exceeding 34 kts, the yellow circle indicates winds exceeding 50 kts, and the red indicates winds exceeding 64 kts. The dashed circles are range rings that measure the distance from the center of the storm at 50nm intervals. The topmost panel contains a hurricane with maximum sustained winds of 100 mph, a radius of maximum winds of 35 nm, and 60 TJ of IKE. The panels below perturb either IKE or VMAX by ±20 TJ or kts as denoted, without changing RMAX

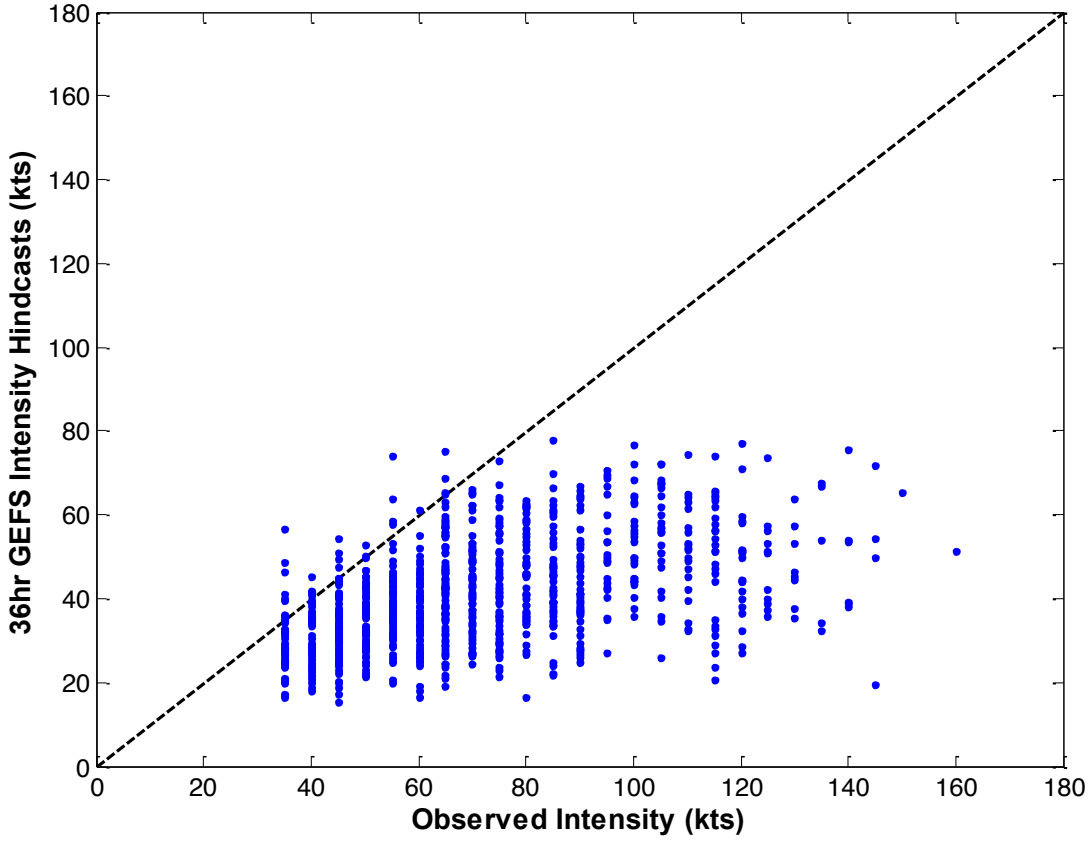

Figure 7.8: Scatter plot comparing observed maximum sustained winds to 36-hour GEFS reforecasted maximum sustained winds in 869 storm fixes between 1990 and 2011. The dashed line indicates a perfect one-to-one relationship (y=x). In nearly every case, the GEFS underestimates the observed intensity, largely because of its course one-degree resolution.

CHAPTER 8

SUMMARY AND CONCLUSIONS

Throughout this dissertation, a series of SPIKE products have been presented in an effort to produce a scheme that is capable of forecasting integrated kinetic energy in Atlantic TCs. Forecasts of IKE are sought after in this work because of the metric's known relationship to damage potential, particularly with regards to storm surge (Powell and Reinhold 2007). IKE is found to be an inertial quantity since it involves integrating a storm's entire wind field. As such, IKE does not change as quickly as do intensity metrics like VMAX, which are notoriously difficult to forecast in real time. Ultimately, the slow changing nature of the IKE metric can be exploited by statistical models in an effort to provide useful guidance on the damage potential of a tropical cyclone.

Early versions of SPIKE were admittedly primitive, as they used simple linear regression to project IKE change from a series of observed environmental and storm-specific predictors in historical TCs between 1990 and 2011. Despite the obvious nonlinearities within the atmospheric system, the normal regression models showed promise, indicating that IKE could be projected skillfully relative to persistence by a simple statistical scheme when given accurate predictors. However, it was immediately apparent that these linear models could be improved upon since they were incapable of handling many of the complex relationships that govern IKE growth.

As a result, the linear models were eventually replaced by a system of one hundred complex artificial neural networks. The ANNs have the advantage of not being restricted by a closed set of linear equations, and as a result they are better equipped to project IKE from a nonlinear set of input parameters. In a perfect prognostic setup, the neural network system produced more accurate deterministic projections of IKE change than did the older linear

models. Probabilistic products were added to the skillful deterministic products in an effort to quantify the uncertainty present in the system of neural networks. Working in concert, the probabilistic and deterministic ANN products found in the second version of SPIKE made the prediction scheme quite versatile.

All of the initial work with SPIKE and SPIKE2 indicated that IKE could be predicted on the restrictive condition that the environmental predictors were nearly perfect. However, in an operational setting, the input parameters will be forecasted, and they surely will contain inherent errors and biases. Subsequently, SPIKE2's deterministic and probabilistic products were tested using imperfect input parameters from the GEFS reforecast database. As expected, the imperfect hindcasted predictors precluded SPIKE2 from achieving the same levels of skill that it exhibited in the perfect prognostic exercises. Nonetheless, the mock-operational version of SPIKE2 proved to be skillful relative to persistence forecasts and other benchmarks such as the BIKE model out to at least 72 hours. More impressively, the hindcasts outperformed a limited number of simulated operational IKE forecasts that were derived from archived NHC wind radii forecasts, given the distinction that a 00Z forecast of SPIKE using 00Z GEFS data likely would not have been made until an hour or two after the issuance of the 03Z NHC forecasts that same day.

In addition to simply projecting quantities of IKE, a new methodology for projecting storm size from SPIKE2 was also presented. The methodology utilizes a simple axisymmetric Rankine Vortex to project the radial distributions of winds within a TC from inputs of IKE, VMAX, and RMW. The RV wind model worked well in a perfect prognostic experiment, suggesting that SPIKE2 could be used to project symmetrical TC wind fields if the input values given to the RV model were accurate. Mock-operational tests of the wind radii model using SPIKE2 hindcasts and GEFS reforecasts of VMAX were unable to replicate the high levels of skill within the perfect prognostic runs. The wind radii projections proved to be very sensitive to inputs of IKE and VMAX. Nonetheless, the GEFS-SPIKE2 wind radii hindcasts greatly outperformed archived

operational forecasts of storm size produced by NHC, given again that a real-time version of SPIKE likely would not be available to NHC until just after the issuance of their forecast advisories because of the runtime required for dynamical weather prediction models. Overall, the initial results suggest that this wind radii product has the potential to be helpful to forecasters once it is fully developed for real-time use.

Ultimately, the work in this dissertation serves as a proof of concept for a methodology that can be used to effectively forecast IKE and storm structure in real time using artificial neural networks. In fact, the skill of the mock-operational models documented throughout the latter stages of this dissertation was likely underestimated because of the somewhat poor quality of hindcasted input parameters taken from the low resolution control runs found in the GEFS reforecast dataset. Therefore, it would be prudent to build upon this work by testing SPIKE2 and the RV model with a set of input parameters from other models. For example, improved input parameters from a higher resolution model could lead to an increase in SPIKE2's already impressive deterministic and probabilistic skill. Additionally, better estimations of TC intensity from a high resolution dynamical model could potentially allow the RV wind radii model to approach its theoretical maximum skill levels in an effort to make it even more viable for operational use.

Acquiring improved input parameters from a more sophisticated model is not the only way to improve the usefulness of SPIKE2. In fact, it would be wise to adapt SPIKE2 to accept predictors from ensembles and an array of multiple regional and global weather forecasting models. By utilizing predictors from multiple atmospheric scenarios, SPIKE2 will be able to produce probabilistic forecasts that not only account for the uncertainty in the neural networks but also the uncertainty in the inputted atmospheric predictors. Furthermore, obtaining predictors from different model solutions will ensure that the biases and errors from a single model (such as those found in the one-degree GEFS data) do not cripple the ability for a SPIKE2 to accurately assess the kinetic energy in Atlantic TCs.

It should also be noted that statistical models like SPIKE2 are limited by a requirement that they have a sufficiently long historical database for calibration. This limitation is one of the primary reasons for using the control run in the available GEFS reforecast database throughout this dissertation. Despite its coarse one-degree resolution, the GEFS archive contained a long record of data from a static version of the same model. Unfortunately, few operational models have long archives of forecasts or hindcasts that are readily available. Therefore, adapting SPIKE2 to be used with a higher resolution operational model or model ensembles is dependent upon securing an archive for the desired model. As such, adapting SPIKE2 to the rest of the GEFS ensemble members would be easier to accomplish than would be adapting SPIKE2 to work with predictors from the Hurricane Weather and Forecasting Model (HWRF).

Future work must also focus on actually testing SPIKE2 in real time during the Atlantic hurricane season. The hindcast exercises suggest that SPIKE2 should retain its skill with the GEFS, but further testing must be done to ensure that SPIKE2 can produce timely IKE forecasts in an operational setting. As alluded to previously, SPIKE2's products almost certainly cannot be issued instantaneously at initialization time. Although the neural networks themselves can be run fairly quickly, an operational version of SPIKE2 still requires dynamically forecasted input parameters, and unfortunately, the output from most modern dynamical models is not available until a few hours after their initialization time. For instance, NCEP's Global Forecast System (GFS) model is initialized four times daily (00Z, 06Z, 12Z, 18Z), but its data is traditionally not available until two to five hours after those initialization times. Therefore, the statistical-dynamical models using this data including Model Output Statistics (MOS) and, in the future, SPIKE2 cannot come out until after this multiple-hour delay. As a result, a 72-hour SPIKE2 forecast using GFS or GEFS data would be already a few hours into its forecast period by the time it was issued, thus shortening it to a 66-70 hour forecast depending on the actual issuance time.

Ultimately, a large delay between issuance and initialization would be detrimental to the usefulness of SPIKE2 because most operational forecasters are required to issue their forecasts at regular intervals. For instance, NHC issues regular forecast advisories at 03Z, 09Z, 15Z, and 21Z. If the new version of SPIKE2 is not ready for those advisories, forecasters will be forced to use older runs of the product if they are interested in evaluating the IKE of a system. To alleviate this concern some operational statistical-dynamical models, such as SHIPS, are run in a so called “early cycle” mode, wherein each product uses environmental predictors from the previous dynamical model run, which is typically initialized six hours earlier (i.e. the 00Z SHIPS forecast uses dynamical predictors from the 18Z GFS). Adapting this early cycle approach to SPIKE2 will ensure that its IKE forecasts are in advance of each forecast advisory. Consequently, SPIKE2’s dynamical predictors in an early cycle mode would be several hours old before SPIKE2 is even issued. Ultimately, the need to forecast an additional few hours would likely result in a slight degradation to model skill. Therefore, it will be necessary to compare the pros and cons of running SPIKE2 in this early cycle mode against simply releasing SPIKE2 shortly after the release of each new dynamical model run.

Experimental tests of SPIKE2 using the operational GEFS during the 2015 Atlantic Hurricane Season would go a long way towards answering these timing questions, all while bringing SPIKE2 from a mock-operational mode to a fully operational mode. Ideally, SPIKE2 can be issued less than three hours after the issuance of each model run, but whether or not this is feasible remains to be seen. In addition, SPIKE2 will contend with the fact that operational weather forecast models are constantly being changed and upgraded in these real time tests. In theory, these upgrades could affect the model’s ability to predict each individual predictor. As a result, normalizing the predictors could become troubling since the historical hindcast sample mean and sample standard deviation could be out of sync with the operational sample statistics. In all likelihood, the small changes to the predictors in each model upgrade should not affect the skill of the SPIKE2 model, so long as the physical relationships that govern

the statistical equations remain constant. However, over time, it is likely that the SPIKE2 model will need to be recalibrated with newer models to keep up with the operational forecast models.

If the operational tests are successful in the Atlantic, it could be advantageous to adapt SPIKE2 for other tropical basins. Unfortunately, it would be difficult to replicate the approach performed for the Atlantic in this study for other basins, simply because there is no global long-term extended best track dataset. However, the SPIKE2 model may not need to be recalibrated for each basin. In theory, the dynamics and thermodynamics processes that govern tropical cyclone development do not change across the basins. Therefore, adapting SPIKE2 for the Pacific and the world's other tropical basins could be as simple as using a global model like the GEFS to produce input parameters for storms across the world.

In addition to running SPIKE2 operationally, considerable effort must be taken to produce improved products that will make the forecasts produced by SPIKE2 more appealing and more informative to operational meteorologists and the general public whom otherwise might not be familiar with the kinetic energy metric. To this point, most of the SPIKE2 results have been shared via complex tables or simple line plots that would probably be unfit for operational use. One exception would be the wind radii plots, which are similar to wind field products already produced by NHC. Therefore, it would be wise to improve the graphic fidelity of the simple IKE forecasts, both deterministic and probabilistic. Additionally, it would be beneficial to relate IKE forecasts to other TC impacts that are more readily understood by the general public. For example, considering its relationship to storm size, it is distinctly possible that IKE is also related to the size of a TCs precipitation shield. Therefore, it is worth investigating whether or not rainfall or flooding risk can be measured through IKE in any meaningful way. Furthermore, it would be beneficial to continue to explore how IKE can be related to storm surge damage, whether through a simple categorical scale like Powell and Reinhold (2007) or through a more complex process of using the projected SPIKE2 RV wind field in a storm surge model.

The most important future research topic, however, should be focused on trying to better

understand the physical processes that control IKE and storm size in TCs. Any statistical model is meaningless without an understanding of the underlying physical processes that govern the relationships within the statistical equations. An interpretation for the physical processes that govern the SPIKE2 model is given earlier in Chapters 2, 3, and 5. However, there is still much to learn about the physical processes that govern changes in storm structure. Any future discoveries of relationships between storm size and the environment surrounding a storm could greatly benefit the SPIKE2 models, and the scientific community as a whole.

Finally, future work must also include outreach and education of the risks of TCs. Currently, storm size is not prioritized in the meteorological community, and as such, forecasts of IKE may not be well accepted initially, despite the inherent uses of knowing a storm's destructive potential. Even more concerning, is the public's belief that storm surge is primarily controlled by the maximum sustained winds in a TC, despite the evidence provided by recent hurricanes proving otherwise. The National Hurricane Center's storm surge unit and many other organizations in the field have made novel strides in studying storm surge and the impacts of storm size on storm surge. However, more work is needed to study the effects of storm size on the risks of a TC and to educate the public about these risks so that they can properly protect lives and property during a tropical cyclone landfall event.

REFERENCES

Abdul-Wahab, S.A., and S.M Al-Alawi, 2002: Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks. *Environmental Modelling & Software,* **12**, 219-228.

American Meteorological Society, 2012: "Conservation of Angular Momentum". Glossary of Meteorology. [Available online at http://glossary.ametsoc.org/wiki/Conservation_of_ angular_momentum]

Atkinson, P.M., and A.R.L. Tatnall, 1997: Introduction Neural networks in remote sensing. *International J. of Remote Sensing*, **18**, 699-709.

Avila L.A. and J. Cangialosi, 2011: Tropical Cyclone Report Hurricane Irene (AL092011) 21-28 August 2011. *National Hurricane Center*. [available online: http://www.nhc.noaa.gov/data/tcr/ AL092011_Irene.pdf]

Bell, G. D., and Coauthors, 2000: Climate assessment for 1999. *Bull. Amer. Meteor. Soc.,*81, 1328–1378.

Berg, R., 2009: Tropical Cyclone Report Hurricane Ike (AL092008) 1-14 September 2008. *National Hurricane Center*. [available online: http://www.nhc.noaa.gov/pdf/TCR-AL092008_Ike_3May10.pdf].

Berry, M.J.A., and G.S. Linoff, 2004: Artificial Neural Networks. *Data Mining Techniques for Marketing, Sales, and Customer Relation Second Edition.*, Wiley Publishing Inc., 211-256.

Bister, M. and K.A. Emanuel, 1998: Dissipative heating and hurricane intensity. *Meteor. Atm. Phys.,* **52**, 233-240.

Blake, E.S., T.B. Kimberlain, R.J. Berg, J.P. Cangialosi, and J.L. Bevin II, 2013: Tropical Cyclone Report Hurricane Sandy (AL182012) 22-29 October 2012. *National Hurricane Center*. [available online: http://www.nhc.noaa.gov/data/tcr/AL182012_Sandy.pdf].

Brier G.W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*,**78**, 1-3.

Cangialosi J.P. and C.W. Landsea, 2014: National Hurricane Center Wind Radii Forecast Verification. *31st Conference on Hurricanes and Tropical Meteorology,* San Diego, CA, Amer. Met. Soc., 56.

Cao Q., B.T. Ewing, M.A. Thompson, 2012: Forecasting wind speed with recurrent neural networks. *European J. of Operational Res.*, **221**, 148-154.

Cawley G.C. and S.R. Dorling, 1996: Reproducing a subjective classification scheme for atmospheric circulation patterns over the United Kingdom using a neural network. *Proceedings International Conference on Neural Networks,* 281-286.

Chan, K.T.F and J.C.L. Chan, 2013: Angular Momentum Transports and Synoptic Flow Patterns Associated with Tropical Cyclone Size Change. *Mon. Wea. Rev.*, **141**, 3985–4007.

—— and ——, 2014: Impacts of initial vortex size and planetary vorticity on tropical cyclone size. Q.J.R. Meteorol. Soc., **140**, 2235–2248.

Dean. L., K.A. Emanuel, and D.R. Chavas, 2009: On the size distribution of Atlantic tropical cyclones. *Geophys. Res. Lett.,* 36, L14803.

DeMaria, M., and J. Kaplan, 1994: A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin. Wea. Forecasting, 9, 209 220.

——, and ——, 1999: An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic and eastern north Pacific basins. Wea. Forecasting, 14, 326-337.

——, M. Mainelli, L.K. Shay, J.A. Knaff and J. Kaplan, 2005: Further Improvements in the Statistical Hurricane Intensity Prediction Scheme (SHIPS). Wea. Forecasting, 20, 531-543.

——, J-J. Baik, and J. Kaplan, 1993: Upper-level eddy angular momentum fluxes and tropical cyclone intensity change. *J. Atmos. Sci.*, **50,** 1133–1147.

——, 2010: Tropical Cyclone Intensity Change Predictability Estimates Using a Statistical-Dynamical Model, *29th Conf. on Hurr. And Trop. Meteor., May 10-14, Tuscon, AZ, Amer. Meteor. Soc., 9C.5.*

Demuth, J.L., M. DeMaria., and J. A. Knaff, 2006: Improvement of advanced microwave sounder unit tropical cyclone intensity and size estimation algorithms. *J. Appl. Meteorol. Climatol.,* **45**, 1573-1581.

Depperman, C.E. (1947), Notes on the origin and structure of Philippine typhoons, *Bull. Amer. Meteor. Soc.,* **28**, 399-404.

Dunion, J.P., and C.S. Velden, 2004: The impact of the Saharan Air Layer on Atlantic tropical cyclone activity. *Bull. Amer. Meteor. Soc.*, vol. 85, no. 3, 353-365.

Emanuel K., 1988: The Maximum Intensity of Hurricanes. *J. Atmos. Sci.*, **45**, 1143-1155.

——, 2005: Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, **436**, 686–688.

Evans, C., and R.E. Hart, 2008: Analysis of the Wind Field Evolution Associated with the Extratropical Transition of Bonnie (1998). *Mon. Wea. Rev.*, **136**, 2047-2065.

Fogarty, C., 2010: Hurricane Igor 2010 Newfoundland's Most Damaging Hurricane in 75 Years Preliminary Storm Summary, *Canadian Meteor. and Ocean. Soc. Bull.,* **38**, 209-217.

Gardner M.W., and S.R. Dorling, Artificial neural networks (the multilayer perceptron)-a review of applications in the atmospheric sciences. *Atmos. Environment,* **32**, 2627-2636.

Gray, W. M., 1968: Global view of the origins of tropical disturbances and storms. *Mon. Wea. Rev.*, **96**, 669-700.

Hagan, M.T. and Menhaj, M.B., 1994: Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, **5**, 989-993.

Hamill, T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M.l Fiorino, T. J. Galarneau, Y. Zhu, and W. Lapenta, 2013: NOAA's Second-Generation Global Medium-Range Ensemble Reforecast Dataset. *Bull. Amer. Meteor. Soc.*, **94**, 1553–1565.

Hanley D., J. Molinari, and D. Keyser, 2001: A Composite Study of the Interactions between Tropical Cyclones and Upper-Tropospheric Troughs. *Mon. Wea. Rev.,* **129**, 2570-2584.

Hapuarachchi, H.A.P., Q.J. Wang, and T.C. Pagano, 2011: A review of advances in flash flood forecasting. *Hydrol. Process.,* **25**, 2771-2784.

Hart R.E., and J.L. Evans, 2001: A Climatology of the Extratropical Transition of Atlantic Tropical Cyclones. *J. Climate*, **14**, 546-564.

Hill, K.A., and G.M. Lackmann, 2009: Influence of Environmental Humidity on Tropical Cyclone Size. *Mon. Wea. Rev.*, **137**, 3294–3315.

Holland, G.J., 1980: An analytic model of the wind and pressure profiles in hurricanes. *Mon. Wea. Rev.,***108**, 1212-1218.

——, 1983: Angular momentum transports in tropical cyclones. *Quart. J. R. Met. Soc.,* **109**, 187-209.

——, and R.T. Merrill, 1984: On the dynamics of tropical cyclone structural changes. *Quart. J. R. Met. Soc.,* **110**, 723-745.

——, J.I. Belanger, A. Fritz, 2010: A Revised Model for Radial Profiles of Hurricane Winds. *Mon. Wea. Rev.*, **138**, 4393–4401.

Irish, J.L., D T. Resio, and J.J. Ratcliff, 2008: The influence of Storm Size on Hurricane Surge. *J. Phys. Oceanogr.,* **38**, 2003-2013.

Jarvinen, B. R., and C. J. Neumann, 1979: Statistical forecasts of tropical cyclone intensity for the North Atlantic basin. NOAA Tech. Memo. NWS NHC-10, 22 pp.

——, ——, and M. A. S. Davis, 1984: A tropical cyclone data tape for the North Atlantic Basin, 1886-1983: Contents, limitations, and uses. NOAA Technical Memorandum NWS NHC 22, Coral Gables Florida, 21 pp.

Jones, S.C., and Coauthors, 2003: The Extratropical Transition of Tropical Cyclones: Forecast Challenges, Current Understanding, and Future Directions. *Wea. Forecasting*, **18**, 1052–1092.

Jones, T. A., D. J. Cecil, and M. DeMaria, 2006: Passive Microwave-Enhanced Statistical Hurricane Intensity Prediction Scheme. Wea. and Forecasting, **21**, 613-635.

Kantha, L., 2006: Time to replace the Saffir-Simpson Hurricane Scale? Eos, *Trans. Amer. Geophys. Union,***87**,3–6.

Kaplan, J., and M. DeMaria, 2003: Large-Scale Characteristics of Rapidly Intensifying Tropical Cyclones in the North Atlantic Basin. *Wea. Forecasting*, **18**, 1093–1108.

Kimball, S.K., and J.L. Evans, 2002: Idealized Numerical Simulations of Hurricane-Trough Interaction. *Mon. Wea. Rev.*, **130**, 2210-2227.
——, 2006: A Modeling Study of Hurricane Landfall in a Dry Environment. *Mon. Wea. Rev.,* **134**, 1901-1918.

Knabb, R.D., J.R. Rhome, and D.P. Brown, 2005: Tropical Cyclone Report Hurricane Katrina 23-30 August 2005. *National Hurricane Center.* [available online: http://www.nhc.noaa.gov/ data/tcr/ AL122005_Katrina.pdf]

Knaff J.A., M. DeMaria, C.R. Sampson, and J.M. Gross, 2003: Statistical, 5-Day Tropical Cyclone Intensity Forecasts Derived from Climatology and Persistence. *Wea. Forecasting*, **18**, 80–92.

——, C.R. Sampson, M. DeMaria, T.P. Marchok, J.M. Gross, C.J. McAdie, 2007: Statistical Tropical Cyclone Wind Radii Prediction Using Climatology and Persistence. *Wea. Forecasting*, **22**, 781–791.

——, S.P. Longmore, D.A. Molenar, 2014: An Objective Satellite-Based Tropical Cyclone Size Climatology, *J. Climate,* **27**, 455-476.

Kriesel D., 2007: A Brief Introduction to Neural Networks. [available online: http://www.dkriesel.com/en/science/neural_networks]

Kozar, M.E. and V. Misra, 2014: Statistical Prediction of Integrated Kinetic Energy in Atlantic Tropical Cyclones, *Mon. Wea. Rev.,* e-view.

Landsea, C.W., and J.L. Franklin, 2013: Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format. *Mon. Wea. Rev.*, **141**, 3576-3592.

Lee C.S., 1989: Observational Analysis of Tropical Cyclogenesis in the Western North Pacific. Part I: Structural Evolution of Cloud Clusters. *J. Atmos. Sci.*, **46**, 2580–2598.

Maclay K.S., M. DeMaria, and T.H. Vonder Haar, 2008: Tropical Cyclone Inner-Core Kinetic Energy Evolution. *Monthly Weather Review* **136**:12, 4882-4898.

Marquardt D., 1963: An algorithm for least squares estimation of non-linear parameters. *J. Soc. Ind. Appl. Math,***11**, pp.431 -441

Mason S.J., 2004: On Using “Climatology” as a Reference Strategy in the Brier and Ranked Probability Skill Scores. *Mon. Wea. Rev.*, **132**, 1891-1895.

McBride, J. L., 1995: Tropical cyclone formation. *Global Perspectives on Tropical Cyclones,* WMO/TD No. 693, Rep. TCP-38, World Meteorological Organization, 63–105.

Merrill, R.T. 1984: A Comparison of Large and Small Tropical Cyclones. *Mon. Wea. Rev.*, **112**, 1408–1418.

Misra V., S. DiNapoli, and M. Powell, 2013: The Track Integrated Kinetic Energy of Atlantic Tropical Cyclones. *Monthly Weather Review* **141**:7, 2383-2389.

Musgrave K.D., R. K.Taft, J. L.Vigh, B. D.McNoldy, *and* W. H.Schubert, 2012: Time evolution of the intensity and size of tropical cyclones, J. *Adv. Model. Earth Syst.,* **4**, M08001.

Neumann, C.J., 1987: Prediction of tropical cyclone motion: Some practical aspects. Extended Abstracts, *17th conference on Hurr. And Trop. Meteor.*, April 7-10, 1987, Miami, FL., Amer. Meteor. Soc., 266-269.

Nolan, D. S., and M. G. McGauley, 2012: Tropical cyclogenesis in wind shear: Climatological relationships and physical processes. *Cyclones: Formation, Triggers and Control,* K. Oouchi and H. Fudeyasu, Eds., Nova Science Publishers, 1–34.

Pfeffer R.L., 1958: Concerning the Mechanics of Hurricanes, *J. Meteor.,* **15**, 113-120.

Price, J.F., 1981: Upper Ocean Response to a Hurricane. *J. Phys. Oceanogr.*, **11**, 153–175.

Powell, M.D., and T.A. Reinhold, 2007: Tropical Cyclone Destructive Potential by Integrated Kinetic Energy. *Bull. Amer. Meteor. Soc.*, **88**, 513–526.

——, S.H. Houston, L. R. Amat, and N. Morisseau-Leroy, 1998: The HRD real-time hurricane wind analysis system. J. Wind Eng. Ind. Aerodyn.,77–78,53–64.

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily high-resolution blended analyses for sea surface temperature. J. Climate, 20, 5473-5496.

Saffir, H., 1975: Low cost construction resistant to earthquakes and hurricanes. *ST/ESA/23, United Nations*, 216 pp.

Schade L.R. and K.A. Emanuel, 1999: The Ocean's Effect on the Intensity of Tropical Cyclones: Results from a Simple Coupled Atmosphere–Ocean Model. *J. Atmos. Sci.*, **56**, 642–651.

Shapiro, L.J., 1983: The Asymmetric Boundary layer Flow Under a Translating Hurricane. *J. Atmos. Sci.*, **40**, 1984–1998.

Shukla R.P., K.C. Tripathi, A.C. Pandley, and I.M.L. Das, 2011: Prediction of Indian summer monsoon rainfall using Niño indicies: A neural network approach. *Atmos. Res.*, **102**, 99-109.

Simpson, R. H., 1974: The hurricane disaster potential scale. *Weatherwise*, **27**, 169–186.

Stathakis, D., 2009: How many hidden layers and nodes?, *Inter. J. of Remote Sensing,* **30**, 2133-2147.

Tuleya, R.E., and Y. Kurihara, 1981: A Numerical Study on the Effects of Environmental Flow on Tropical Storm Genesis, *Mon. Wea. Rev.*, **109**, 2487-2506.

Vickery, P., P. Skerlj, A. Steckley, and L. Twisdale, 2000: Hurricane Wind Field Model for Use in Hurricane Simulations, *J. Struct. Eng.*, **126**, 1203–1221.

Wilks, D.S., 2005: Statistical Methods in the Atmospheric Sciences: An Introduction, 2nd Edition, Academic Press.

BIOGRAPHICAL SKETCH

Michael Kozar was raised in the borough of Park Ridge in Northern New Jersey, along with his brother Steven, by his parents Frank and Maria Kozar. He developed a passion for science and math at an early age while attending West Ridge Elementary School. During the active 1998 Atlantic Hurricane Season, he became obsessed with studying and forecasting the weather and has wanted to become a meteorologist ever since. Appropriately, Michael graduated from Park Ridge High School in June of 2007 as a line of strong thunderstorms interrupted and ultimately derailed the outdoor ceremony.

Following high school, he moved to University Park, PA to attend Penn State University. At Penn State, Michael had the opportunity to take his first formal meteorology courses and attend his first meteorology conference, all while also gaining invaluable experience as a forecaster in the student run Campus Weather Service. While in Happy Valley, Michael also received his first exposure to the meteorology profession through multiple volunteer internships and eventually a student position at the local National Weather Service offices. After four years, Michael graduated with distinction, earning his B.S. and M.S. in meteorology after completing undergraduate and graduate studies related to hurricanes and climate with his major professor, Dr. Michael Mann.

Following his time at Penn State, Michael decided to continue his graduate studies by attending Florida State University to earn his doctorate degree in Meteorology. While in Tallahassee, he had the opportunity to take graduate level classes in both oceanography and meteorology while working with his major professor, Dr. Vasu Misra, at the Center for Ocean-Atmospheric Prediction Studies. Throughout Michael's doctoral studies, he had an opportunity to do research across a number of disciplines within the field of meteorology, ranging from tropical meteorology to climate, before finally settling on a dissertation topic related to his

primary interest in the field, operational tropical meteorology. Michael is set to graduate after four years at Florida State with his PhD. Afterwards, Michael is looking forward to beginning his professional career in meteorology, where ever that may take him.