NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

STATISTICAL ANALYSIS OF ENSEMBLE FORECASTS OF TROPICAL CYCLONE TRACKS OVER THE NORTH ATLANTIC

by

Christopher E. Nixon

June 2012

Thesis Advisor:	Patrick A. Harr
Second Reader:	Russell L. Elsberrry

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704–0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202–4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704–0188) Washington DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** June 2012	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE Statistical Analysis of Ensemble Forecasts of Tropical Cyclone Tracks over the North Atlantic		**5. FUNDING NUMBERS**	
6. AUTHOR(S) Christopher E. Nixon			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943–5000		**8. PERFORMING ORGANIZATION REPORT NUMBER**	
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A		**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number _____N/A_____.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited		**12b. DISTRIBUTION CODE**	

13. ABSTRACT (maximum 200 words)

The skill of individual ensemble prediction systems (EPS) is evaluated in terms of the probability of a tropical cyclone (TC) track forecast being within an expected area. Anisotropic probability ellipses are defined from each EPS to contain 68% of the ensemble forecast members. Forecast reliability is based on whether the forecast verifying position is within the ellipse. A sharpness parameter is based on the size of the EPS probability ellipse relative to the main operational forecast probability product, the Goerss Predicted Consensus Error (GPCE). For the 2008–2011 Atlantic TC seasons, the ECMWF ellipses have the highest degree of reliability of the EPSs. Additionally, the ECMWF ellipse has a higher resolution than the GPCE operational product over all forecast intervals. The sizes and shapes of the EPS ellipses varied with TC track types, which suggests that information about the physics of the flow-dependent system is retained compared to isotropic probability circles that may not reflect variability associated with track type. It is concluded that the ECMWF ensemble contributes the most to a combined EPS-based product called the Grand Ensemble (GE), and further modification of the GE to reflect this has a potential for reducing the sizes of warning areas.

14. SUBJECT TERMS Tropical Cyclone Track Errors, Ensemble Prediction Systems, National Hurricane Center forecasts			**15. NUMBER OF PAGES** 87
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540–01–280–5500 Standard Form 298 (Rev. 2–89)
Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

STATISTICAL ANALYSIS OF ENSEMBLE FORECASTS OF TROPICAL CYCLONE TRACKS OVER THE NORTH ATLANTIC

Christopher E. Nixon
Captain, United States Air Force
B.S., Pennsylvania State University, 2003

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY

from the

NAVAL POSTGRADUATE SCHOOL
June 2012

Author: Christopher E. Nixon

Approved by: Patrick A. Harr
Thesis Advisor

Russell L. Elsberry
Second Reader

Wendell A. Nuss
Chair, Department of Meteorology

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The skill of individual ensemble prediction systems (EPS) is evaluated in terms of the probability of a tropical cyclone (TC) track forecast being within an expected area. Anisotropic probability ellipses are defined from each EPS to contain 68% of the ensemble forecast members. Forecast reliability is based on whether the forecast verifying position is within the ellipse. A sharpness parameter is based on the size of the EPS probability ellipse relative to the main operational forecast probability product, the Goerss Predicted Consensus Error (GPCE). For the 2008–2011 Atlantic TC seasons, the ECMWF ellipses have the highest degree of reliability of the EPSs. Additionally, the ECMWF ellipse has a higher resolution than the GPCE operational product over all forecast intervals. The sizes and shapes of the EPS ellipses varied with TC track types, which suggests that information about the physics of the flow-dependent system is retained compared to isotropic probability circles that may not reflect variability associated with track type. It is concluded that the ECMWF ensemble contributes the most to a combined EPS-based product called the Grand Ensemble (GE), and further modification of the GE to reflect this has a potential for reducing the sizes of warning areas.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I. INTRODUCTION....1
A. MOTIVATION....1
B. OBJECTIVE....2

II. BACKGROUND....5
A. NATIONAL HURRICANE CENTER OPERATIONAL METHODS TO DEFINE TROPICAL CYCLONE TRACK UNCERTAINTIES....5
1. Forecast Track Uncertainty Cone....5
2. Wind Speed Probability Swath....6
3. Goerss Predicted Consensus Error (GPCE)....8
B. ENSEMBLE PREDICTION SYSTEMS....9
1. European Center for Medium-range Weather Forecasts (ECMWF)....9
2. United Kingdom Meteorological Office (UKMO)....10
3. National Centers for Environmental Prediction Global Forecast System (NCEP/GFS)....10

III. METHODOLOGY....11
A. DATA....11
1. Data Source....11
2. Data Format....15
3. Data Homogeneity....16
4. Region and Sub-regions....17
5. Developing the EPS Ellipse....18
B. STATISTICAL ANALYSIS TYPES....20
1. Probability within Spread....20
2. Ellipse Reliability....20
3. Mean Area Difference (MAD)....21

IV. ANALYSIS AND RESULTS....23
A. OVERVIEW....23
B. ENSEMBLE MEAN TRACK ERRORS....23
1. FTE....24
2. ATE....27
3. XTE....31
C. STATISTICAL ANALYSIS....35
1. Probability within Spread....35
2. Ellipse Reliability....40
3. Mean Area Difference (MAD)....45
D. SUMMARY....48
E. CASE STUDY IRENE....49
1. Hurricane Irene 0000 UTC 21 August 2011....49
2. Hurricane Irene 0000 UTC 22 August 2011....50
3. Hurricane Irene 0000 UTC 24 August 2011....50

4. Hurricane Irene 0000 UTC 25 August 2011 50
5. Hurricane Irene 0000 UTC 26 August 2011 51
6. Hurricane Irene 0000 UTC 27 August 2011 51

V. CONCLUSIONS AND RECOMMENDATIONS 59
A. CONCLUSIONS 59
B. RECOMMENDATIONS 60

LIST OF REFERENCES 61

INITIAL DISTRIBUTION LIST 65

LIST OF FIGURES

Figure 1. National Hurricane Center official track forecast with cone of uncertainty issued (a) 2300Z 20 August 2011, (b) 0300Z 22 August 2011 (From NHC 2012a). 2

Figure 2. Hurricane Irene 5-day watch/warning and forecast plot for 23 August 2011 (From NHC 2012d). Legend in bottom box explains symbols and gives a distance scale. 6

Figure 3. Wind speed probability graphic that depicts the likelihood of 50-kt winds will occur during the next 120 h issued 23 August 2011 (From NHC 2012g). Legend in bottom box explains color scale and the Hurricane Irene center of circulation is represented as the white dot. 7

Figure 4. Section II of the wind speed probability forecast for Hurricane Irene issued 27 August 2011 (From NHC 2012h). 8

Figure 5. Predicted 70% confidence radius (solid circle) of the 120-hour CONU forecast for Hurricane Isabel on 0000 UTC 13 September 2003. The individual model tracks used to create the CONU consensus model are shown. Notice the GPCE circle is much smaller than the 120-hour radius (dotted circle) used by the NHC potential 5-day track area graphic (Goerss 2007). 9

Figure 6. Official tracks of the 2008 Atlantic hurricane season. Storms are listed in the top-right box with the symbols and track color explained in the legend in the bottom-right box (From NHC 2012j). 12

Figure 7. Official tracks of the 2009 Atlantic hurricane season. Storms are listed in the top-right box with the symbols and track color explained in the legend in the bottom-right box (From NHC 2012l). 13

Figure 8. Official tracks of the 2010 Atlantic hurricane season. Storms are listed in the top-right box with the symbols and track color explained in the legend in the bottom-right box (From NHC 2012n). 14

Figure 9. Official cyclone tracks of the 2011 Atlantic hurricane season. Storms are listed in the top-right box with the symbols and track color explained in the legend in the bottom-right box (From NHC 2012p). 15

Figure 10. Illustration of forecast-track error (FTE), cross-track error (XTE), along-track error (ATE). In this example, the forecast position is ahead and to the right of the best-track position. The XTE in this case will be a positive value to the right of the best track and the ATE will be a positive value ahead of the best track (Neese 2010). 16

Figure 11. Geographic sub-regions of the Atlantic basin used to group Ensemble Prediction System (EPS) forecast track data (Neese 2010). Sub-regions highlighted in red were selected for this study. 17

Figure 12. The Grand Ensemble (GE) probability ellipse (blue) that contains 68% of the ensemble members (black) for Hurricane Igor on 1200 UTC 12 September 2010 (Pearman 2011). 19

Figure 13. Average Forecast-Track Error (FTE) for each Ensemble Prediction System across the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the FTE is represented by a bar graph at each 12 h forecast tau from 0–120 h. 24
Figure 14. Average Forecast-Track Error (FTE) for each of the Ensemble Prediction Systems over the Main Development Region of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the FTE is represented by a bar graph at each 12 h forecast tau from 0–120 h. 25
Figure 15. Average Forecast-Track Error (FTE) for each of the Ensemble Prediction Systems with East Coast Storms (ECS) of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the FTE is represented by a bar graph at each 12 h forecast tau from 0–120 h. 26
Figure 16. Average Forecast-Track Error (FTE) for each of the Ensemble Prediction Systems over the Gulf of Mexico (GOM) of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the FTE is represented by a bar graph at each 12 h forecast tau from 0–120 h. 27
Figure 17. Average Along-Track Error (ATE) for each of the Ensemble Prediction Systems over the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the ATE is represented by a bar graph at each 12 h forecast tau from 0–120 h. 28
Figure 18. Average Along-Track Error (ATE) for each of the Ensemble Prediction Systems over the Main Development Region (MDR) of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the ATE is represented by a bar graph at each 12 h forecast tau from 0–120 h. 29
Figure 19. Average Along-Track Error (ATE) for each of the Ensemble Prediction Systems over the East Coast (ECS) region of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the ATE is represented by a bar graph at each 12 h forecast tau from 0–120 h. 30
Figure 20. Average Along-Track Error (ATE) for each of the Ensemble Prediction Systems over the Gulf of Mexico (GOM) region of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the ATE is represented by a bar graph at each 12 h forecast tau from 0–120 h. 31
Figure 21. Average Cross-Track Error (XTE) for each of the Ensemble Prediction Systems of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the XTE is represented by a bar graph at each 12 h forecast tau from 0–120 h. 32
Figure 22. Average Cross-Track Error (XTE) for each of the Ensemble Prediction Systems over the Main Development Region (MDR) of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the XTE is represented by a bar graph at each 12 h forecast tau from 0–120 h. 33
Figure 23. Average Cross-Track Error (XTE) for each of the Ensemble Prediciton System over the East Coast (ECS) region of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the XTE is represented by a bar graph for each 12 h forecast tau from 0–120 h. 34

Figure 24. Average Cross-Track Error (XTE) for each of the Ensemble Prediction Systems over the Gulf of Mexico (GOM) region of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the XTE is represented by a bar graph at each 12 h forecast tau from 0–120 h.....35
Figure 25. Probability within Spread (PWS) for each EPS across the Atlantic basin for the 2008–2011 Atlantic hurricane seasons. Probabilities on the left correspond to the percentage of the forecasts that the +/- standard deviation of each EPS forecast members includes the TC best-track position at each forecast interval.....37
Figure 26. Probability within Spread (PWS) for each EPS across the Main Development Region (MDR) Atlantic sub-region for the 2008–2011 Atlantic hurricane seasons. Probabilities on the left correspond to the percentage of the forecasts that the +/- standard deviation of each EPS forecast members includes the TC best-track position at each forecast interval. 38
Figure 27. Probability within Spread (PWS) for each EPS across the East Coast Storms (ECS) Atlantic sub-region for the 2008–2011 Atlantic hurricane seasons. Probabilities on the left correspond to the percentage of the forecasts that the +/- standard deviation of each EPS forecast members includes the TC best-track position at each forecast interval. 39
Figure 28. Probability within Spread (PWS) for each EPS across the Gulf of Mexico (GOM) Atlantic sub-region for the 2008–2011 Atlantic hurricane seasons. Probabilities on the left correspond to the percentage of the forecasts that the +/- standard deviation of each EPS forecast members includes the TC best-track position at each forecast interval.....40
Figure 29. Ellipse reliability for each EPS and GE for the entire Atlantic basin. The line graph represents the verification (hit rate) that the TC best-track forecast position falls within the EPS ellipse that contains 68% of its forecast members. The blue line represents the 68% reliability level. The number above the forecast interval is the number of ensemble forecasts from 2008–2011 included. 42
Figure 30. Ellipse reliability for each EPS and GE for the Main Development Region (MDR) of the Atlantic basin. The line graph represents the verification (hit rate) that the TC best-track forecast position falls within the EPS ellipse that contains 68% of its forecast members. The blue line represents the 68% reliability level. The number above the forecast interval is the number of ensemble forecasts from 2008–2011 included. 43
Figure 31. Ellipse reliability for each EPS and GE for the East Coast Storms (ECS) sub-region of the Atlantic basin. The line graph represents the verification (hit rate) that the TC best-track forecast position falls within the EPS ellipse that contains 68% of its forecast members. The blue line represents the 68% reliability level. The number above the forecast interval is the number of ensemble forecasts from 2008–2011 included. 44
Figure 32. Ellipse reliability for each EPS and GE for the Gulf of Mexico (GOM) sub-region of the Atlantic basin. The line graph represents the verification

(hit rate) that the TC best-track forecast position falls within the EPS ellipse that contains 68% of its forecast members. The blue line represents the 68% reliability level. The number above the forecast interval is the number of ensemble forecasts from 2008–2011 included. 45
Figure 33. The Mean Area Difference (MAD) of each EPS and GE compared to GPCE across the Atlantic basin. Positive values indicate the EPS or GE ellipses are smaller than the GPCE circle for each forecast interval. The values above the forecast interval are the number of EPS forecasts included. 46
Figure 34. The Mean Area Difference (MAD) of each EPS and GE compared to GPCE across the MDR sub-region. Positive values indicate when the EPS or GE ellipses are smaller than the GPCE circle for each forecast interval. The values above the forecast interval are the number of EPS forecasts included. 47
Figure 35. The Mean Area Difference (MAD) of each EPS and GE compared to GPCE across the ECS sub-region. Positive values indicate when the EPS or GE ellipses are smaller than the GPCE circle for each forecast interval. The values above the forecast interval are the number of EPS forecasts included. 47
Figure 36. The Mean Area Difference (MAD) of each EPS and GE compared to GPCE across the GOM sub-region. Positive values indicate when the EPS or GE ellipses are smaller than the GPCE circle for each forecast interval. The values above the forecast interval are the number of EPS forecasts included. 48
Figure 37. The TC forecast track ellipses of each EPS for Hurricane Irene at 0000 UTC 21 August 2011. Each ellipse signifies a 12 h forecast interval and is colored to match the individual EPS as defined in the legend at the top right. The large dot inside each ellipse is the corresponding ensemble mean forecast position. The best-track positions are in black. For geographical reference, Puerto Rico is slightly left of center near 18°N, 67°W. 52
Figure 38. The TC forecast track ellipses of each EPS for Hurricane Irene at 0000Z UTC 22 August 2011. Each ellipse signifies a 12 h forecast interval and is colored to match the individual EPS as defined in the legend at the top right. The large dot inside each ellipse is the corresponding ensemble mean forecast position. The best-track positions are in black. For geographical reference, Puerto Rico is slightly bottom-right near 18°N, 67°W. 53
Figure 39. The TC forecast track ellipses of each EPS for Hurricane Irene at 0000Z UTC 24 August 2011. Each ellipse signifies a 12 h forecast interval and is colored to match the individual EPS as defined in the legend at the top right. The large dot inside each ellipse is the corresponding ensemble mean forecast position. The best-track positions are in black. For geographical reference, east coast of the United States is located at the left side of the image. 54

Figure 40. The TC forecast track ellipses of each EPS for Hurricane Irene at 0000Z UTC 25 August 2011. Each ellipse signifies a 12 h forecast interval and is colored to match the individual EPS as defined in the legend at the top right. The large dot inside each ellipse is the corresponding ensemble mean forecast position. The best-track positions are in black. For geographical reference, east coast of the United States is located at the left side of the image. 55
Figure 41. The TC forecast track ellipses of each EPS for Hurricane Irene at 0000Z UTC 26 August 2011. Each ellipse signifies a 12 h forecast interval and is colored to match the individual EPS as defined in the legend at the top right. The large dot inside each ellipse is the corresponding ensemble mean forecast position. The best-track positions are in black. For geographical reference, east coast of the United States is located at the left side of the image. 56
Figure 42. The TC forecast track ellipses of each EPS for Hurricane Irene at 0000Z UTC 27 August 2011. Each ellipse signifies a 12 h forecast interval and is colored to match the individual EPS as defined in the legend at the top right. The large dot inside each ellipse is the corresponding ensemble mean forecast position. The best-track positions are in black. For geographical reference, east coast of the United States is located at the left side of the image. 57

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

Table 1. Radii of NHC forecast cone circles for 2011 based on error statistics from 2006–2010 (From NHC 2012c). 5
Table 2. Summary and comparison of the three Ensemble Prediction Systems (EPS) used in this study. 10
Table 3. Total number of Tropical Cyclones that were included in this thesis for the Atlantic basin and each sub-region. 17

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AFB	Air Force Base
AFMAN	Air Force Manual
ATE	Along-Track Error
CONUS	Continental United States
CXML	Cyclone XML
DoD	Department of Defense
ECMWF	European Center for Medium-range Weather Forecasts
ECS	East Coast Storms
EDA	Ensemble of Data assimilations
EPS	Ensemble Prediction System
ETBV	Ensemble Transform Bred Vector
FTE	Forecast-Track Error
GE	Grand Ensemble
GFS	Global Forecast System
GOM	Gulf of Mexico
GPCE	Goerss Predicted Consensus Error
h	hour
HPP	Hurricane Probability Program
MAD	Mean Area Difference
MDR	Main Development Region
NCEP	National Centers for Environmental Prediction
NHC	National Hurricane Center
n mi	nautical miles
NWP	Numerical Weather Prediction
PWS	Probability within Spread
SV	Singular Vector
TC	Tropical Cyclone
TIGGE	THORPEX Interactive Grand Global Ensemble
UKMO	United Kingdom Meteorological Office
XTE	Cross-Track Error

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

Thank you to my friends and family who have supported and encouraged me, and helped with care of my daughter through this academic process.

A special thank you to my advisor, Professor Pat Harr, for his vast guidance and time in this subject matter, Professor Russell Elsberry for this thorough help in editing, Bob Creasey and Mary Jordan for all the help in coding and troubleshooting program scripts, and all the other NPS professors who have provided valuable knowledge, instruction and background material.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. MOTIVATION

Tropical Cyclones (TCs) routinely affect Department of Defense (DoD) operations with significant adverse weather conditions by either a direct impact on a DoD installation or by restricting air and sea maneuverability. From 2008–2011, there were 63 named storms across the Atlantic basin, which includes the Gulf of Mexico and Caribbean Sea. Of these, 24 came within 300 n mi of 15 DoD installations that are within 100 n mi of the coast across the southeast CONUS and Caribbean regions. These storms caused disruptions to operations, and in some cases relocations of personnel and equipment.

The 26th Operational Weather Squadron at Barksdale AFB, and 612th Support Squadron at Davis-Monthan AFB, provide remote weather support to these 15 DoD installations. This support includes relaying information to the installation Commanders of any potential impacts of approaching TCs. Air Force Manual (AFMAN) 15–129 specifies in paragraph 5.1.1.1. that supporting weather units will not deviate from the official TC forecast position, track, movement, and forecast maximum wind speed (with exceptions to feeder band convective activity, and terrain effects) from the tropical cyclone forecast center (i.e., the National Hurricane Center (NHC) in Miami, FL) guidance. Because of this, the installation Commanders base much of their decisions on the official NHC forecasts and visual aid forecast products.

Improvement in NHC forecasts and forecast products are vital to increase the ability of installation Commanders that are at risk to provide resource and personnel protection. Improvement in these products such that risk may be accurately conveyed largely hinges on the ability to reduce the uncertainty cone in forecast tracks and wind speed probability swaths while maintaining forecast skill. The case of Hurricane Irene (2011) demonstrates how difficult it is to narrow forecast track uncertainty while verifying the actual forecast position within 24 hours of landfall.

Hurricane Irene developed into a tropical storm on 20 August 2011 and the first official forecast not only had the storm center traveling south of Puerto Rico (Figure 1a), but the forecast cone of uncertainty was largely south of the island. During the next 24 h, the track of Irene had shifted northward and Irene made a direct landfall on Puerto Rico on 21 August 2011 (Figure 1b). Furthermore, Irene had strengthened to a minimal hurricane as it crossed over the island. Although Irene was at that point only a minimal hurricane, significant wind and flood damage occurred across the island ($500 million) including the damage to a DoD installation. Irene later moved northwestward, and made several landfalls along the eastern seaboard of the United States, and at one point, poised a significant threat to Langley AFB in Virginia. In all, more than $7 billion in damage was related to Hurricane Irene (*New York Times* 2011).

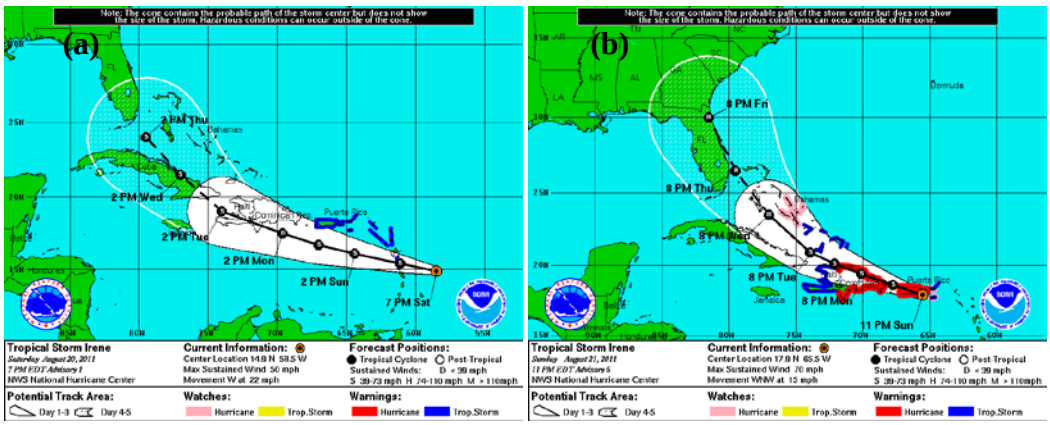

Figure 1. National Hurricane Center official track forecast with cone of uncertainty issued (a) 2300Z 20 August 2011, (b) 0300Z 22 August 2011 (From NHC 2012a).

B. OBJECTIVE

Operational numerical forecast aids are used as a basis for official tropical cyclone track forecasts as in Figure 1. Efforts continue at the NHC to improve the accuracy of these products. The NHC forecasters routinely use consensus forecast aids formed from a suite of operational global atmospheric prediction models (Goerss 2007).

Forecast uncertainty can be based on either a consensus of independent operational models or an Ensemble Prediction System (EPS) that is based on

perturbations to one single model. The NHC frames forecast uncertainty using several methods that are defined in Chapter II. Primarily, their methods are based on the most recent operational forecast errors and the consensus of operational models.

The primary objective of this thesis is to explore the use of forecasts produced by an EPS to convey forecast variability. Statistical characteristics of TC forecast track error distributions for each of the three main operational global EPS are examined. Additionally, the combination of all three EPS, which is called the Grand Ensemble (GE), are also examined. By creating graphical products based on each EPS, uncertainty within the individual model will be represented. To compare the different types of model uncertainty, the statistical characteristics of the EPSs and GE are compared to the Goerss Predicted Consensus Error (GPCE).

Background material is provided in Chapter II. The methodology used in this study is described in Chapter III. The analyses and results are presented in Chapter IV and the conclusions and future recommendations are given in Chapter V.

THIS PAGE INTENTIONALLY LEFT BLANK

II. BACKGROUND

A. NATIONAL HURRICANE CENTER OPERATIONAL METHODS TO DEFINE TROPICAL CYCLONE TRACK UNCERTAINTIES

1. Forecast Track Uncertainty Cone

The NHC TC track forecast cone (Figure 2) was developed in 1983 under the Hurricane Probability Program (HPP) (DeMaria et al. 2009). It depicts the official forecast track as a solid black line for up to 3 days, and dashed line for days 4 and 5. A white cone depicts the geographic uncertainty around the official forecast track for days 1–3. The white cone is replaced by a hashed cone for days 4 and 5. The forecast track uncertainty cone represents the probable track of the center of a tropical cyclone, and is formed by enclosing the area defined by a set of circles (not shown) along the forecast track (at 12, 24, 36 h, etc). The size of each circle is set so that two-thirds of historical official forecast errors over a 5-year sample fall within the circle (NHC 2012b). The circle radii defining the cone sizes in 2011 for the Atlantic and eastern North Pacific basins are given in Table 1. This product is available online to the general public and to installation Commanders as a general depiction of the uncertainty in the TC track. This product has changed very little from 1983 to 2005, except that the forecast periods of 96–120 h were added in 2003 (DeMaria et al. 2009).

Table 1. Radii of NHC forecast cone circles for 2011 based on error statistics from 2006–2010 (From NHC 2012c).

Forecast Period (hours)	2/3 Probability Circle, Atlantic Basin (nautical miles)
12	36
24	59
36	79
48	98
72	144
96	190
120	239

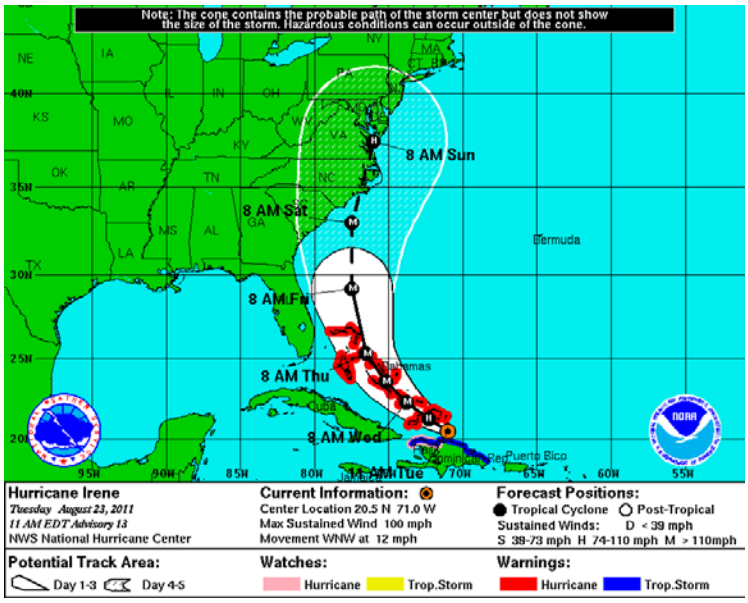

Figure 2. Hurricane Irene 5-day watch/warning and forecast plot for 23 August 2011 (From NHC 2012d). Legend in bottom box explains symbols and gives a distance scale.

2. Wind Speed Probability Swath

The wind speed probability swath (Figure 3) is a graphical display of the overall probability (cumulative probability) that a particular wind speed will occur within the designated timeframe (0–12 h, 0–24 h, 0–48 h, etc.). The white dot defines the current center of circulation of the TC (NHC 2012e).

The current wind speed probability swath was implemented in 2006. This product is created using a Monte Carlo technique that creates 1,000 realizations of TC tracks. Each realization is determined by random sampling from the distribution of official track and intensity errors based on the previous five years. The error samples are then added to the official deterministic forecasts. All 1,000 realizations of TC tracks are then assigned an intensity and wind structure based on a wind profile model. Finally, a linear model is then applied to account for serial correlation and track-intensity dependency (DeMaria et al. 2009).

In addition to the cumulative probability (as in Figure 3), an individual probability text product (Figure 4) that defines the wind-speed threshold probability for a particular location during a specified time frame, which is usually a 6-h increment (NHC 2012f).

The cumulative and individual wind speed probability products give the installation Commander an ability to make the best cost-benefit decisions for resource and personnel protection. However, there are two major limitations with these products. First, the sampling distributions do not account for any background flow dependences, and second, sampling distributions are static for an entire hurricane season since they are based on the previous five hurricane seasons.

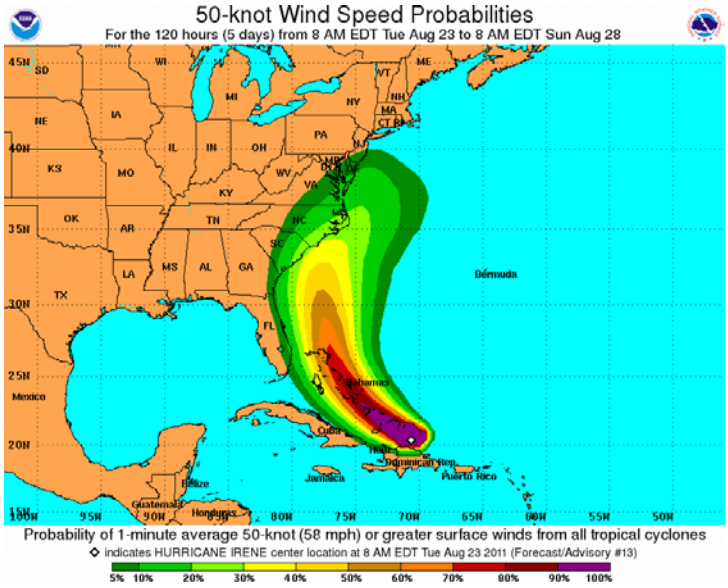

Figure 3. Wind speed probability graphic that depicts the likelihood of 50-kt winds will occur during the next 120 h issued 23 August 2011 (From NHC 2012g). Legend in bottom box explains color scale and the Hurricane Irene center of circulation is represented as the white dot.

II. WIND SPEED PROBABILITY TABLE FOR SPECIFIC LOCATIONS

CHANCES OF SUSTAINED (1-MINUTE AVERAGE) WIND SPEEDS OF AT LEAST
...34 KT (39 MPH... 63 KPH)...
...50 KT (58 MPH... 93 KPH)...
...64 KT (74 MPH...119 KPH)...
FOR LOCATIONS AND TIME PERIODS DURING THE NEXT 5 DAYS

PROBABILITIES FOR LOCATIONS ARE GIVEN AS IP(CP) WHERE
IP IS THE PROBABILITY OF THE EVENT BEGINNING DURING
AN INDIVIDUAL TIME PERIOD (INDIVIDUAL PROBABILITY)
(CP) IS THE PROBABILITY OF THE EVENT OCCURRING BETWEEN
18Z SAT AND THE FORECAST HOUR (CUMULATIVE PROBABILITY)

PROBABILITIES ARE GIVEN IN PERCENT
X INDICATES PROBABILITIES LESS THAN 1 PERCENT
PROBABILITIES FOR 34 KT AND 50 KT ARE SHOWN AT A GIVEN LOCATION WHEN
THE 5-DAY CUMULATIVE PROBABILITY IS AT LEAST 3 PERCENT.
PROBABILITIES FOR 64 KT ARE SHOWN WHEN THE 5-DAY CUMULATIVE
PROBABILITY IS AT LEAST 1 PERCENT.

- - - - WIND SPEED PROBABILITIES FOR SELECTED LOCATIONS - - - -

TIME PERIODS	FROM 18Z SAT TO 06Z SUN	FROM 06Z SUN TO 18Z SUN	FROM 18Z SUN TO 06Z MON	FROM 06Z MON TO 18Z MON	FROM 18Z MON TO 06Z TUE	FROM 06Z TUE TO 18Z TUE	FROM 18Z TUE TO 06Z WED	FROM 06Z WED TO 18Z WED
FORECAST HOUR	(12)	(24)	(36)	(48)	(72)	(96)	(120)	
LOCATION	KT							
BURGEO NFLD	34 X	X(X)	X(X)	1(1)	2(3)	X(3)	X(3)	
PTX BASQUES	34 X	X(X)	X(X)	4(4)	2(6)	X(6)	X(6)	
EDDY POINT NS	34 X	X(X)	X(X)	3(3)	X(3)	X(3)	X(3)	
SYDNEY NS	34 X	X(X)	X(X)	3(3)	X(3)	X(3)	X(3)	
HALIFAX NS	34 X	X(X)	2(2)	4(6)	X(6)	X(6)	X(6)	
YARMOUTH NS	34 X	1(1)	9(10)	1(11)	X(11)	X(11)	X(11)	

Figure 4. Section II of the wind speed probability forecast for Hurricane Irene issued 27 August 2011 (From NHC 2012h).

3. Goerss Predicted Consensus Error (GPCE)

Goerss et al. (2007) determined that the most important predictor in TC track forecast error in the Atlantic was the consensus model spread. That is, the consensus model spread was found to be positively correlated with consensus model TC track forecast error (Goerss 2007). The Goerss Predicted Consensus Error (GPCE) is a circle that represents a 70% probability that a predicted storm position will be within the circle for each forecast interval (Figure 5). This circle is based on the spread of a consensus model called CONU. CONU is a consensus model that is computed when track forecasts from at least two of the following five models are available: GFDI, AVNI, NGPI, UKMI, and GFNI.

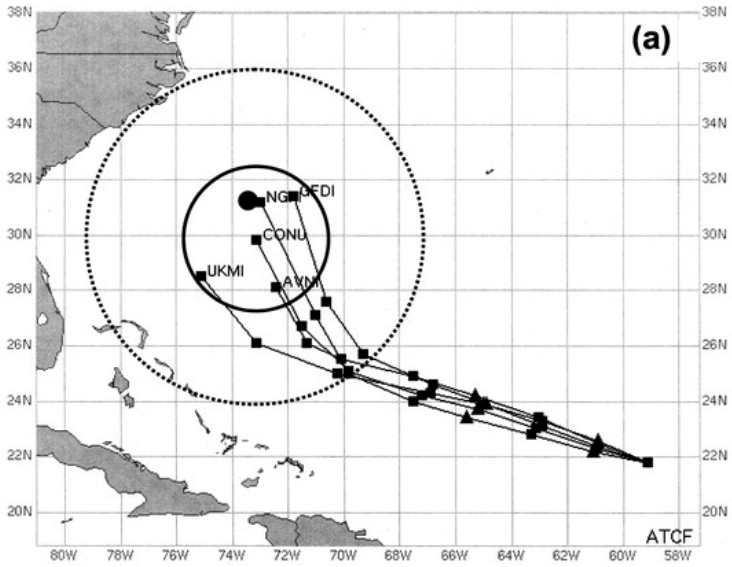

Figure 5. Predicted 70% confidence radius (solid circle) of the 120-hour CONU forecast for Hurricane Isabel on 0000 UTC 13 September 2003. The individual model tracks used to create the CONU consensus model are shown. Notice the GPCE circle is much smaller than the 120-hour radius (dotted circle) used by the NHC potential 5-day track area graphic (Goerss 2007).

B. ENSEMBLE PREDICTION SYSTEMS

Numerous EPSs are in use at forecast centers today. In this study, three primary EPSs are examined. These three include the European Center for Medium-range Weather Forecasts (ECMWF), United Kingdom Meteorological Office (UKMO), and the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS). Each of these EPSs has numerous members created by perturbing the initial conditions of the control forecast. The numbers of members and the techniques for creating these perturbations differ between each EPS and are defined below in Table 2.

1. European Center for Medium-range Weather Forecasts (ECMWF)

The ECMWF EPS has 51 members defined by 50 perturbation members and one control. The EPS forecasts are initialized every 12 h at 0000 UTC and 1200 UTC. The output forecasts extend out to 384 h at an interval of 12 h.

The 50 perturbations are created by three methods: (1) singular vector (SV) technique; (2) using differences between the members of an ensemble of data

assimilations (EDA); and (3) using two different stochastic perturbation techniques (ECMWF 2012). The perturbation technique also varies by latitude.

2. United Kingdom Meteorological Office (UKMO)

The UKMO EPS has 23 members constructed from 22 perturbation members and one control. The EPS is initialized every 6 h at 0000 UTC, 0600 UTC, 1200 UTC, and 1800 UTC. The forecasts are available to 144 h at an interval of 12 h.

The 22 perturbations are created by use of a Kalman ensemble filter. The filter provides estimates of the true state, which is updated by a forecast of the state from the previous time and by observations (Bowler et al. 2008).

3. National Centers for Environmental Prediction Global Forecast System (NCEP/GFS)

The GFS has 21 members constructed from 20 perturbation members and one control. The GFS EPS is initialized every 6 h at 0000 UTC, 0600 UTC, 1200 UTC, and 1800 UTC. The forecasts are available to 384 h at an interval of 6 h.

The 20 perturbations are created by using an Ensemble Transform Bred Vector (ETBV) method that determines the fastest-growing error modes in the model. These perturbations are also subjected to stochastic physics techniques. Note that this perturbation method is generally best suited for the extratropics such that no special perturbations are applied specific to individual tropical cyclones (UCAR 2012).

Table 2. Summary and comparison of the three Ensemble Prediction Systems (EPS) used in this study.

EPS	**Members**	**Forecast Run Times**	**Forecast Duration**
ECMWF	51	00 UTC, 12 UTC	384 hours
UKMO	23	00 UTC, 06 UTC, 12 UTC, 18 UTC	144 hours
GFS	21	00 UTC, 06 UTC, 12 UTC, 18 UTC	384 hours

III. METHODOLOGY

A. DATA

1. Data Source

This study focuses on tropical activity in the Atlantic basin. The 2008, 2009, 2010, and 2011 Atlantic hurricane seasons were all included in this study. A total of 63 named TCs occurred during those four seasons, but only 51 were used due to data availability and track type. Using these 51 storms, a total of 3,422 EPS track forecasts were available. A typical Atlantic hurricane season has 11 named TCs, six of which develop into hurricanes and of those six, two become major hurricanes (category 3 or greater). Three of the four seasons (2008, 2010, 2011) included in this study were above average in terms of activity.

The 2008 Atlantic hurricane season (Figure 6) contained 16 named tropical storms, eight of which became hurricanes and five of those hurricanes strengthened into major hurricanes of category 3 or greater. This season posed a challenge for forecasters as there were six TCs that made landfall across the southeastern United States. Hurricane Paloma was the second strongest November hurricane on record for the Atlantic basin. A high number of casualties were directly caused by these TCs. Approximately 624 people died, with 500 of those occurring from Hurricane Hanna alone across the island of Haiti due to floods caused by heavy rainfall (NHC 2012i) .

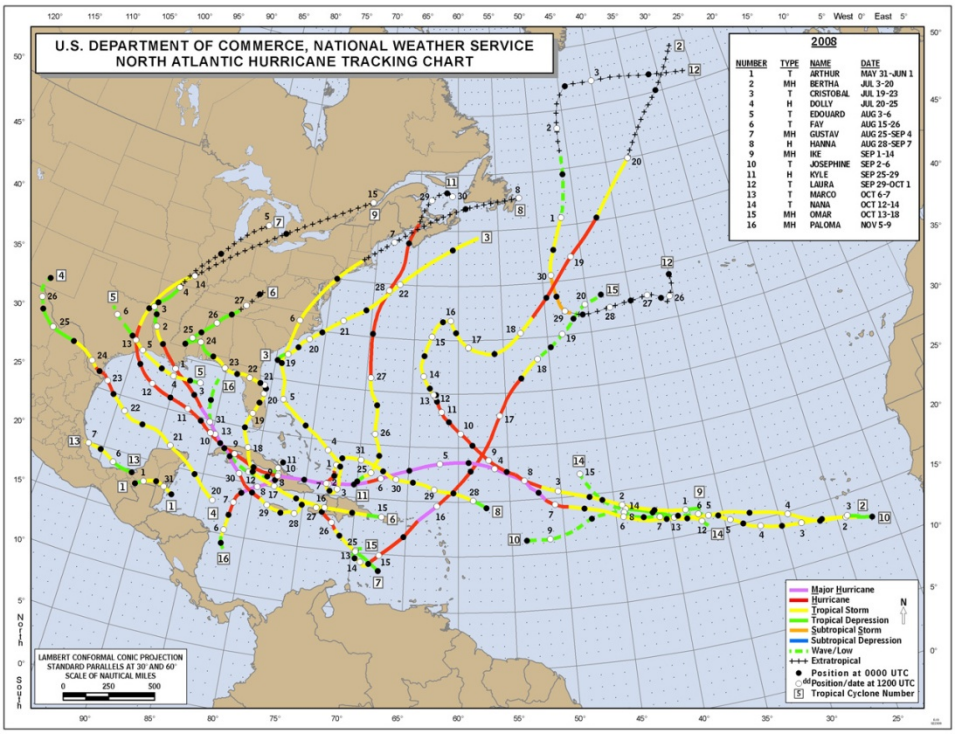

Figure 6. Official tracks of the 2008 Atlantic hurricane season. Storms are listed in the top-right box with the symbols and track color explained in the legend in the bottom-right box (From NHC 2012j).

The 2009 Atlantic hurricane season (Figure 7) was below normal in terms of activity. A total of nine TCs developed, with three becoming hurricanes and two of those strengthening into major hurricanes (NHC 2012k). Only two storms made landfall across the southeastern United States, both of which were tropical storms at the time of landfall. No casualties were experienced and damages from these storms were very minimal.

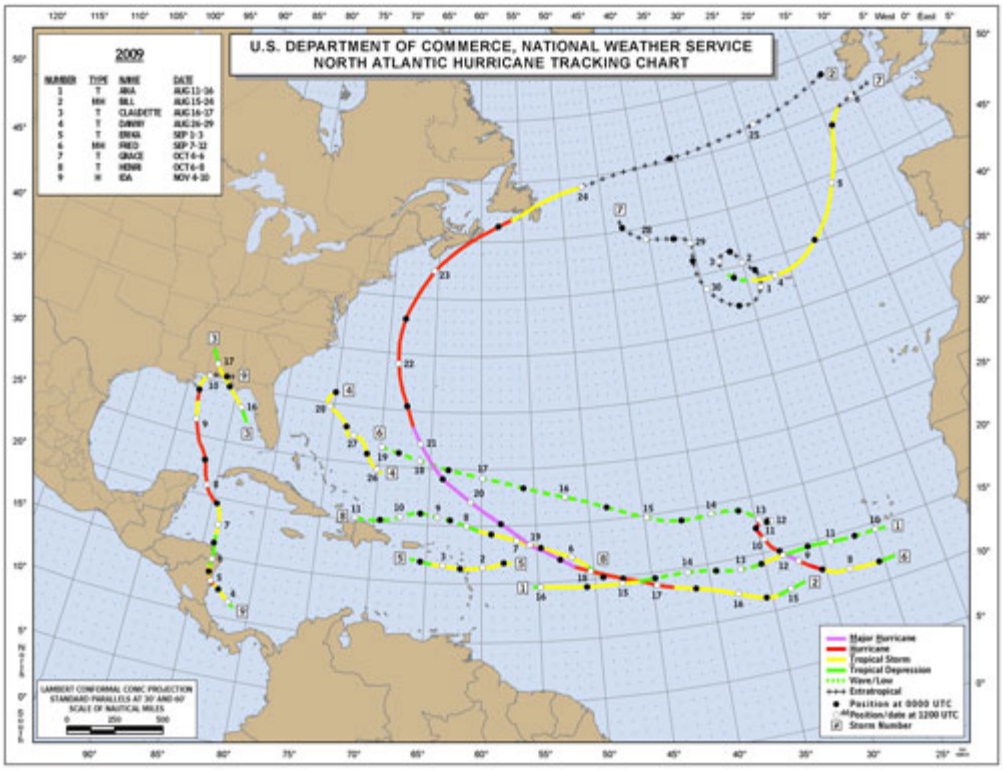

Figure 7. Official tracks of the 2009 Atlantic hurricane season. Storms are listed in the top-right box with the symbols and track color explained in the legend in the bottom-right box (From NHC 2012l).

The 2010 Atlantic hurricane season (Figure 8) was once again an active season in terms of named TCs. In all, 19 named storms developed, 12 of which became hurricanes and five strengthened to major hurricanes. The number of named storms and hurricanes was the highest since the record-setting season of 2005 (NHC 2012m). The bulk of these TCs remained over the central Atlantic, but five did make landfall across central America, and one across the United States.

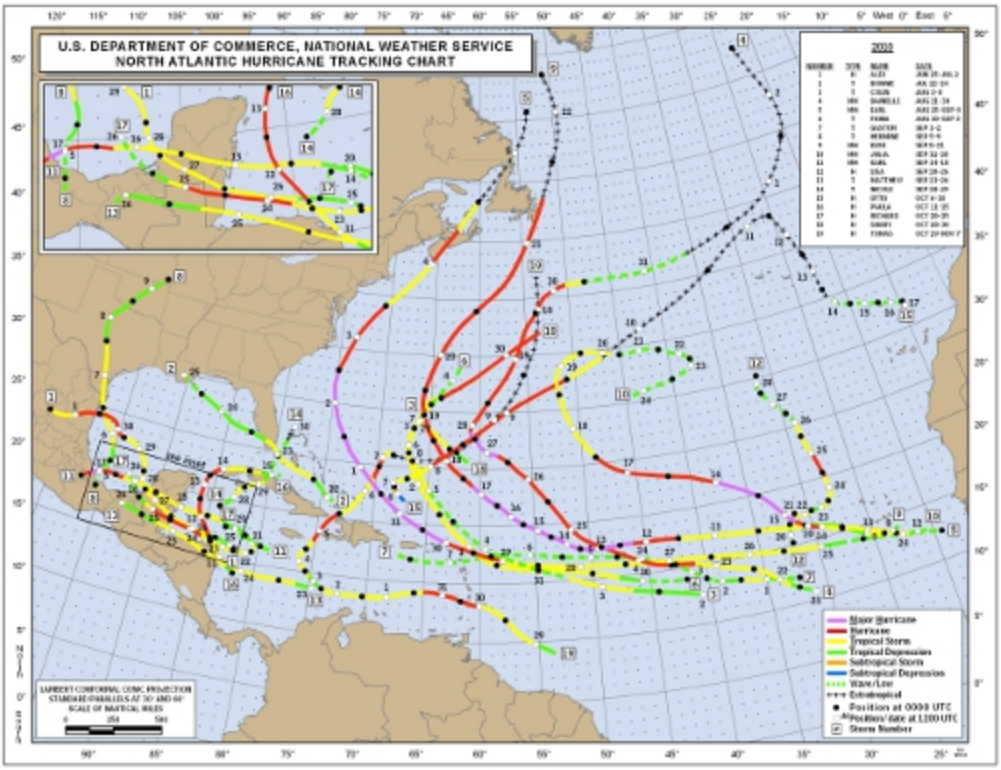

Figure 8. Official tracks of the 2010 Atlantic hurricane season. Storms are listed in the top-right box with the symbols and track color explained in the legend in the bottom-right box (From NHC 2012n).

The 2011 Atlantic hurricane season (Figure 9) was another very active season. There were a total of 19 named TCs, of which seven became hurricanes and four were major hurricanes. Although the vast majority tracked across the central Atlantic, Hurricane Irene was a devastating hurricane across Puerto Rico and parts of the east coast of the United States.

Hurricane Irene made landfall across Puerto Rico as a strong tropical storm and actually strengthened into a hurricane while crossing the island. Irene later made landfall in the Bahamas as a major hurricane but began to gradually weaken. It made landfall in North Carolina as a category 1 hurricane and caused widespread damage across a large portion of the eastern United States as it moved along the coastline. The most severe impact of Irene in the northeastern United States was catastrophic inland flooding in New Jersey, Massachusetts, and Vermont (NHC 2012o).

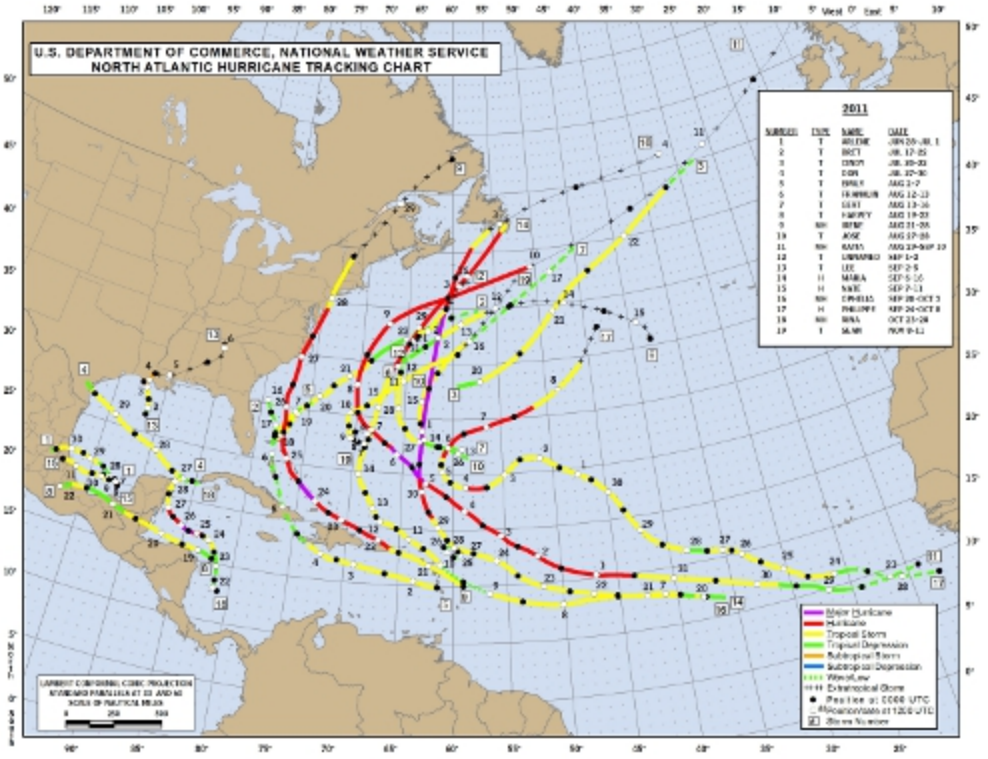

Figure 9. Official cyclone tracks of the 2011 Atlantic hurricane season. Storms are listed in the top-right box with the symbols and track color explained in the legend in the bottom-right box (From NHC 2012p).

2. Data Format

The outputs of the three EPS used in this study were available in the TIGGE (THORPEX Interactive Grand Global Ensemble) database. This database is located online at http://tigge.ucar.edu/home/home.htm. The standard format of these data is in Cyclone XML (CXML). The CXML format was created to be descriptive and human-legible, which makes it easy for users and most automated applications to read. The CXML format is defined such that it contains data from observations and analyses, manual and numerical model forecasts, multiple cyclones and multiple forecasts (ensembles).

The best-track data are a post-storm reanalysis of the cyclone locations and intensities for every six hours (0000 UTC, 0600 UTC, 1200 UTC, 1800 UTC) during the lifespan of the storm. The data collected by NHC to define the best-track analysis include

surface observations, satellite images, aircraft reports, and radar images. The best-track data from the NHC are available through an online directory at ftp://ftp.nhc.noaa.gov/.

Three types of TC track forecast errors are defined in Figure 10 as the along-, cross-, and forecast-track errors. All three of these errors are based on the best-track position. The forecast track error (FTE) is the total great-circle distance between the forecast position and the best-track position. The along-track and cross-track errors are the components of the FTE that results in a 90° angle tangent to the best track as shown in Figure 10.

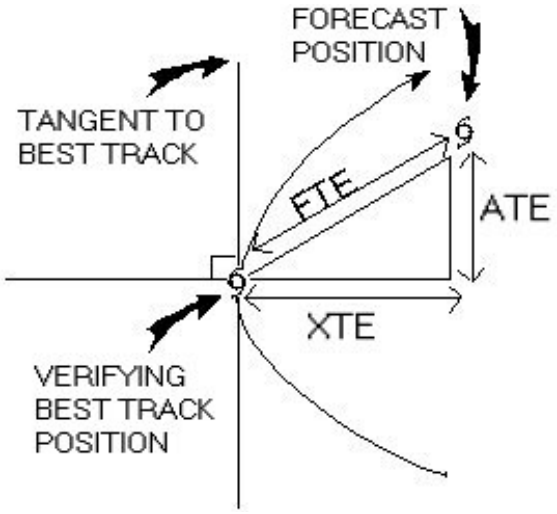

Figure 10. Illustration of forecast-track error (FTE), cross-track error (XTE), along-track error (ATE). In this example, the forecast position is ahead and to the right of the best-track position. The XTE in this case will be a positive value to the right of the best track and the ATE will be a positive value ahead of the best track (Neese 2010).

3. Data Homogeneity

The forecast tracks were organized through grouping of individual models, all regions, and subset regions. Due to the differences in ensemble perturbation techniques and horizontal resolution, the three EPSs may detect and forecast a TC at different times. For example, the ECMWF may begin to develop and forecast a TC 6 h prior to the UKMO and 12 hours prior to the GFS. Only tracks with forecast times that were available for all three EPSs were used. If forecast times were not available for all EPSs,

the forecast data were discarded. The result was a homogeneous dataset of ECMWF, UKMO, and GFS EPS forecasts for each of the TC forecast times.

4. Region and Sub-regions

The database was created for the entire Atlantic basin. However, specific regions of the Atlantic were defined to examine whether there were useful differences in forecast accuracy and uncertainty. Six regions were based on latitude and longitude (Figure 11). Based on the frequency of TCs in each sub-region, three sub-regions were chosen for detailed analysis. The Main Development Region (MDR) had 23 total TCs, the East Coast Storms (ECS) had 29 TCs, and the Gulf of Mexico (GOM) had 14 TCs. Table 3 lists the number of TCs that were included in this study for each sub-region per year.

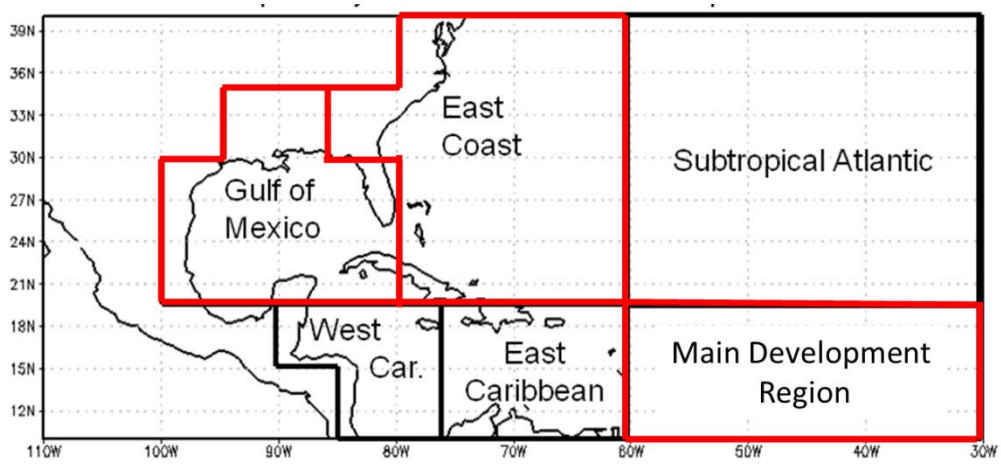

Figure 11. Geographic sub-regions of the Atlantic basin used to group Ensemble Prediction System (EPS) forecast track data (Neese 2010). Sub-regions highlighted in red were selected for this study.

Table 3. Total number of Tropical Cyclones that were included in this thesis for the Atlantic basin and each sub-region.

Year	Atlantic Basin TCs	Main Development Region TCs	East Coast Storms TCs	Gulf of Mexico TCs
2008	13	6	7	5
2009	9	5	2	2
2010	12	6	9	2
2011	17	6	11	5
TOTAL:	**51**	**23**	**29**	**14**

5. Developing the EPS Ellipse

Tropical cyclone forecast track uncertainty can be represented in several ways. The most common method was defined in Figure 1 with the NHC forecast track cone, in which the cone represents a geographic area in which there is a 70% likelihood that the TC will be somewhere within that cone at the projected forecast time. The limitation of an uncertainty cone is that no information can be drawn as to where the TC is most likely to lie in the cone or whether the models have a tendency to over-account or under-account for the background flow. Pearman (2011) addressed this limitation by the creation of the Grand Ensemble (GE) forecast track ellipse.

Pearman (2011) related the uncertainty in the GE forecast position to the principal axis of the spatial distribution of EPS members and centered relative to the GE mean position. The ellipse is defined to contain 68% of the GE member forecast track positions and is centered on the GE mean. These ellipses are calculated for every forecast period that has a homogeneous dataset. In this study, the same approach was used to create the GE ellipse and the ellipses for each EPS.

Since this ellipse calculation is an important aspect in this study, a detailed description is provided. The first step is to create a 2 x n matrix of the latitudes and longitudes of each ensemble member making up the EPS at a particular forecast time. The value of n is defined as the total number of forecasts. A covariance matrix is created from this latitude and longitude matrix that is defined as:

$$\sum_x = \sigma_0^2 (A^T A)^{-1} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \quad (1)$$

where A is the latitude and longitude matrix.

Next, the eigenvalues and eigenvectors of the covariance matrix are calculated by using:

$$\left.\begin{matrix} \lambda_1 \\ \lambda_2 \end{matrix}\right\} = \frac{1}{2}(\sigma_1^2 + \sigma_2^2 \pm \sqrt{(\sigma_1^2 + \sigma_2^2)^2 - 4(\sigma_1^2 \sigma_2^2 - \sigma_{12}^2)}) \quad (2)$$

where λ_1 , λ_2 are the eigenvalues of the covariance matrix. The resulting eigenvectors define the orientation of the ellipse axes and the eigenvalues provide the scaling factors for the semi-major and minor axes.

Lastly, the mean of the latitude and longitude matrix is used to determine the center position of the ellipse. Assuming a Chi-squared distribution, a Chi-squared scaling parameter is applied to ensure that the ellipse captures 68% of the EPS member forecast positions.

An example of an ellipse that captures 68% of the EPS ensemble members at a 95% confidence interval is shown in Figure 12. The semi-major axis, which is not oriented parallel to the track as in Figure 10, is determined by the spatial distribution of the EPS members. This orientation gives detail not only to the spatial uncertainty as does the NHC forecast uncertainty cone, but also to the projected speed uncertainty that is primarily associated with the background steering flow.

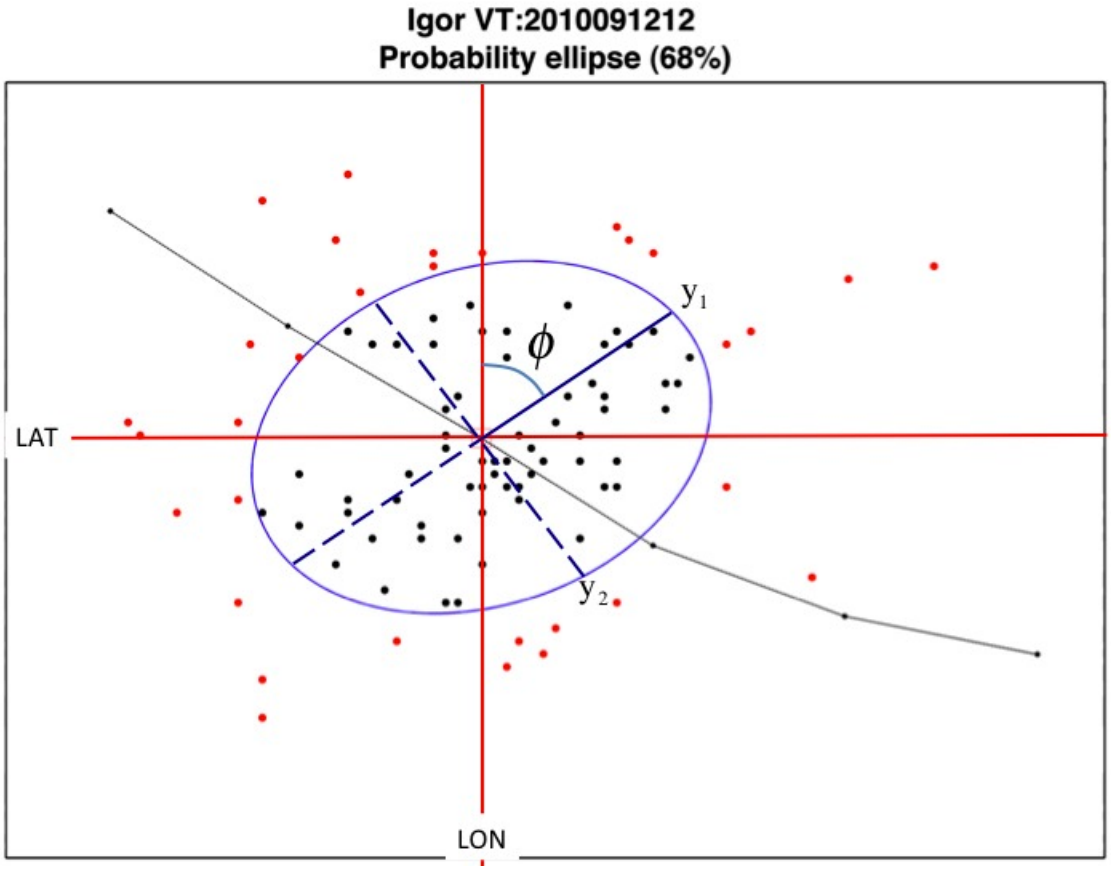

Figure 12. The Grand Ensemble (GE) probability ellipse (blue) that contains 68% of the ensemble members (black) for Hurricane Igor on 1200 UTC 12 September 2010 (Pearman 2011).

B. STATISTICAL ANALYSIS TYPES

Two main statistical characteristics of the ensemble forecasts are the reliability and the resolution. Reliability will be defined as a 68% forecast probability of the occurrence of a particular weather feature of interest (rain, gale-force winds) and verifies 68% of the time. Resolution is defined as a sharpness that is defined with respect to the area of uncertainty in the forecast. Reliability will be calculated for the probability within spread and ellipse reliability, while resolution will be calculated for the mean area difference. The goal is to increase the reliability of a forecast, and increase their resolution (or decrease the uncertainty).

1. Probability within Spread

Probability within Spread (PWS) estimates the likelihood of an observed TC being within the spread of an EPS as:

$$\frac{1}{M}\sum_{m=1}^{M}\left\{\frac{0: s_{obs} > k(\sigma)_m}{1: s_{obs} < k(\sigma)_m}\right\} \tag{3}$$

where k, m are integers, M is the total number of forecasts at a given lead time, s_{obs} is the distance of the observed TC from the EPS mean and σ is the spread of the EPS. If members are sampled from a normal distribution, a standard deviation σ should result in a PWS value of 0.68 (Buckingham et al. 2010).

2. Ellipse Reliability

Ellipse reliability is the percentage of time that the best-track analysis position is within the EPS ellipse at a particular forecast time and is defined as:

$$\frac{1}{M}\sum_{m-1}^{M}\left\{\frac{0: s_{obs} > (68\%\ ellipse)_m}{1: s_{obs} < (68\%\ ellipse)_m}\right\} \tag{4}$$

Due to the definition of the EPS ellipse enclosing 68% of the ensemble forecast track members, the expected ellipse reliability will also be 68%. The EPS reliability percentages above (below) 68% will be defined as over (under) reliable.

3. Mean Area Difference (MAD)

The Mean Area Difference (MAD) is a measure that compares the area of the EPS ellipse with the control ellipse area and calculates a percentage difference in area. In this study, the control ellipse will be the GPCE circle. The formula for MAD is defined as:

$$MAD = \frac{Control\ Ellipse_{area} - Forecast\ Ellipse_{area}}{Control\ Ellipse_{area}} \quad (5)$$

A MAD percentage is positive (negative) if the EPS ellipse area is less (more) than the GPCE circle area (solid circle in Figure 5).

THIS PAGE INTENTIONALLY LEFT BLANK

IV. ANALYSIS AND RESULTS

A. OVERVIEW

The objective of this thesis is to explore the use of forecasts created by individual ensemble forecasts that have different variability. Graphical products based on EPS and the GE will be created to represent uncertainty within the individual models. This section will show the results of analyzing individual EPSs and then comparing the statistical characteristics of combinations of the three EPSs. This sequence of analysis will lead to an understanding of which EPS contributes more to the GE performance at certain forecast intervals. This understanding could lead to further modifications of the GE to increase its reliability and resolution.

All analyses of either the track errors or statistical characteristics will be examined across the entire Atlantic basin, and then within the three sub-regions of the Atlantic as defined in Chapter III. This will allow a greater insight of the potential impact of different TC steering flows that are typical within certain regions of the Atlantic basin.

B. ENSEMBLE MEAN TRACK ERRORS

The ensemble mean track errors are created by calculating the great circle distance between the ensemble mean and the TC best-track position. The mean track error is either represented by an absolute distance between the ensemble mean and best-track positions (FTE), or a component breakdown of the FTE into the Along-Track Error (ATE) and Cross-Track Error (XTE). All track errors are averaged over the four Atlantic hurricane seasons included in this study.

The variability at each forecast interval is represented by plus/minus one standard deviation of the FTE for each individual EPSs. A standard deviation is defined as the average squared difference between the ensemble mean track position and the TC best-track analysis position. Because the standand deviation involves a squaring of the errors, even one very large track error will be reflected strongly in the magnitude of the standard deviation in the bar graph.

1. FTE

Since the FTE is defined as the absolute distance between the ensemble mean and the TC best-track positions, no directionality information can be drawn as to the track error characteristics for each EPS. The FTE does show which EPS forecast tracks are closest on average to the actual best-track analysis positions. The spread about the FTE will be indicated by the standard deviation bar graphs for each forecast interval, and thus will indicate whether significant differences exist between the FTEs of the individual models being compared.

The ECMWF ensemble mean has a consistently lower FTE than the UKMO and GFS throughout all forecast intervals across the entire Atlantic basin (Figure 13). At 120 h, the ECMWF standard deviation is lower than the UKMO and GFS mean FTE. While the FTE does not give any insight as to the size of the ellipse or spread of the ensemble members, it does give a general idea of which EPS more accurately forecasts the overall TC tracks.

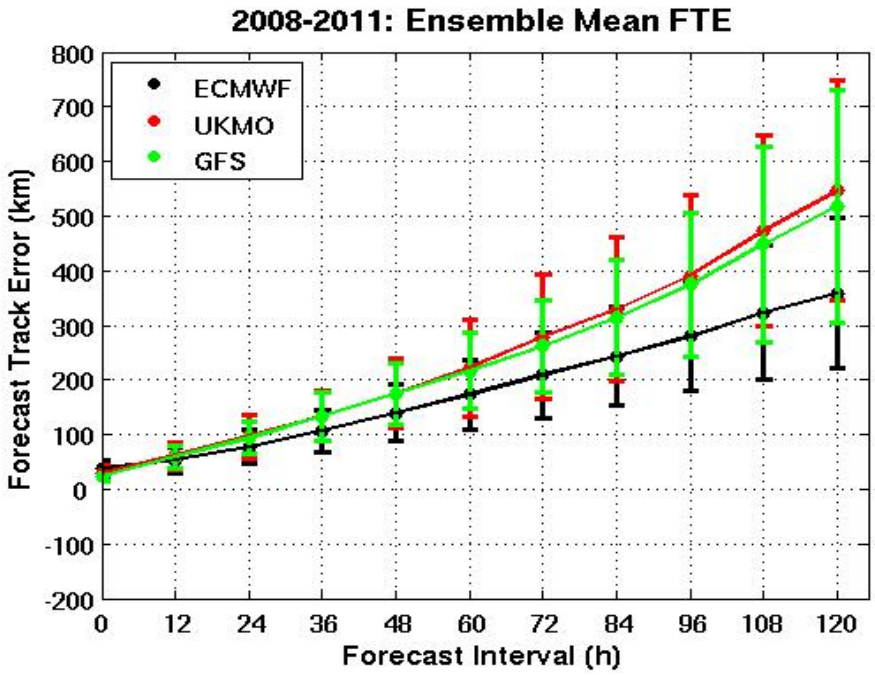

Figure 13. Average Forecast-Track Error (FTE) for each Ensemble Prediction System across the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the FTE is represented by a bar graph at each 12 h forecast tau from 0–120 h.

For the Main Development Region (MDR) (Figure 14), it is noticed that each EPS has a lower ensemble mean FTE than for the entire Atlantic basin in Figure 13. The standard deviations for each EPS are also lower than in Figure 13. This is expected due to the more uniform and steady steering flow in the MDR. During the hurricane season, the steering flow across the MDR is largely due to steady tropical easterlies. This steering flow in the MDR is very zonal with very little north/south component. Because of this lack of variability, these ensembles should have more accurate TC track forecasts. The bulk of the track error is expected to be due to the variations in strength of the zonal tropical easterlies.

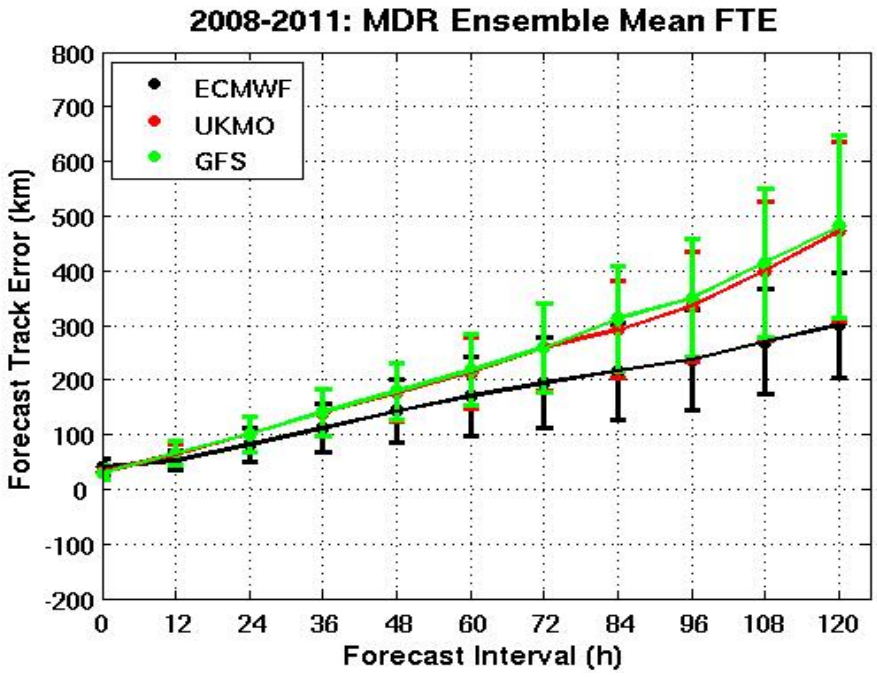

Figure 14. Average Forecast-Track Error (FTE) for each of the Ensemble Prediction Systems over the Main Development Region of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the FTE is represented by a bar graph at each 12 h forecast tau from 0–120 h.

For the East Coast Storms (ECS) sub-region (Figure 15), the FTEs for the ECMWF and UKMO have larger magnitudes, which is expected due to greater variability in the steering flow due to the increasing effects of the mid-latitude westerlies. A surprise was that the FTE for the GFS ensemble mean was smaller at all forecast

intervals compared to its FTEs for the Atlantic basin. Although the GFS was the most accurate EPS past 72 h, the improvement relative to the ECMWF was not statistically significant. The GFS uses the Ensemble Transform Bred Vector perturbation method, which is best suited for use in the extra-tropics. Therefore, the GFS may have a more accurate representation of the uncertainty in the mid-latitude westerlies effects than the ECMWF and UKMO.

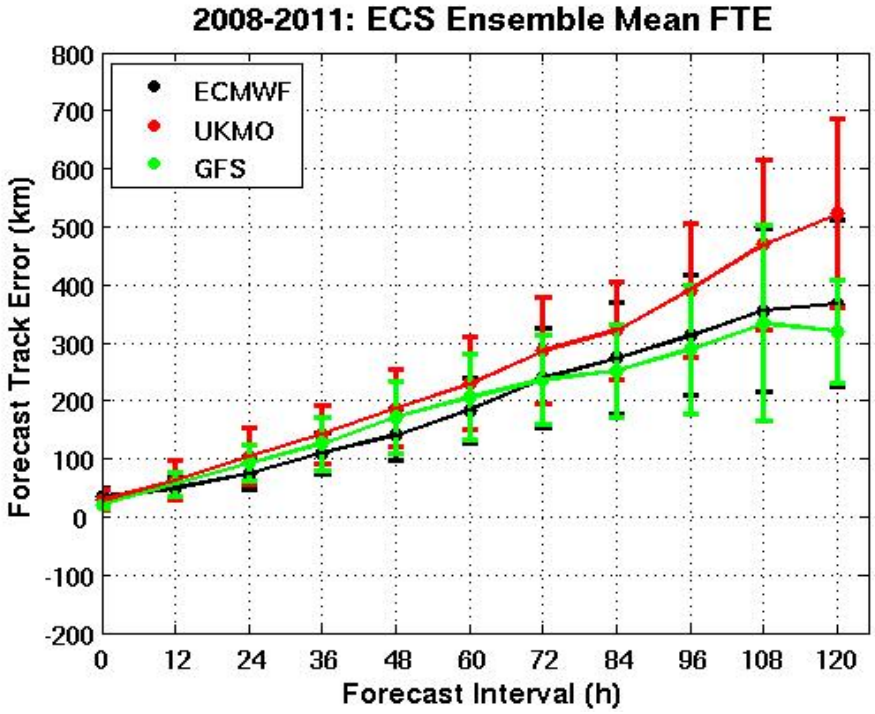

Figure 15. Average Forecast-Track Error (FTE) for each of the Ensemble Prediction Systems with East Coast Storms (ECS) of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the FTE is represented by a bar graph at each 12 h forecast tau from 0–120 h.

Over the Gulf of Mexico sub-region (Figure 16), the GFS was the least accurate in forecasting TC forecast tracks. By 120 h, the mean FTE for GFS was over 700 km, which is approximately the total length of the Texas coastline. The ECMWF and UKMO ensemble means also have higher FTEs and larger standard deviations in this region compared to the entire Atlantic basin. The background steering flow for the Gulf of Mexico TCs is the most variable of the three sub-regions used in this thesis. The

sub-tropical high dominates the steering flow from June – August, with at times very small flow. From September – November, the mid-latitude westerlies become a more dominant steering flow pattern.

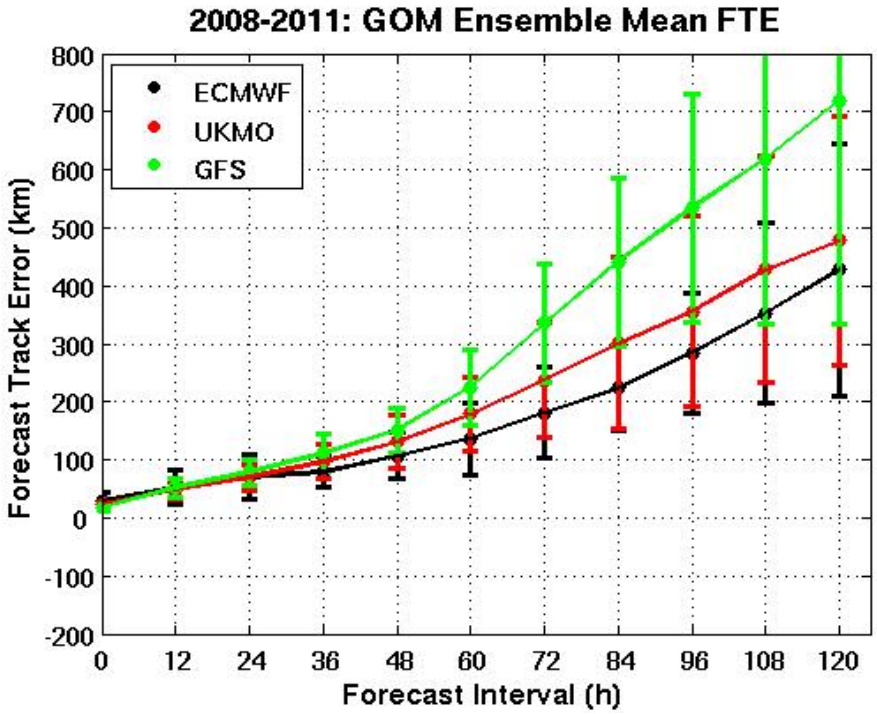

Figure 16. Average Forecast-Track Error (FTE) for each of the Ensemble Prediction Systems over the Gulf of Mexico (GOM) of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the FTE is represented by a bar graph at each 12 h forecast tau from 0–120 h.

2. ATE

The Along-Track Error (ATE) is the ahead or behind component of the Forecast-Track Error (FTE). Positive (negative) values indicate that the forecast position is ahead (behind) of the verifying best track position. The ATE is largely due to the ensemble predictions of the speed of the background steering flow, and not necessarily the orientation of the flow.

In the entire Atlantic basin, each EPS has an ATE within approximately 50 km for each forecast interval in Figure 17. The standard deviations progressively increase with

each forecast interval, which is expected due to greater forecast uncertainty. The ECMWF has the lowest ATE standard deviation values for long-range forecasts (> 84 h), which means it has a higher consistency in predicting the along-track motion than both the UKMO and GFS. All three EPSs have an average ATE that is positive throughout all forecast intervals, which means that on average each ensemble forecasts the TC to be ahead of the verifying best-track position.

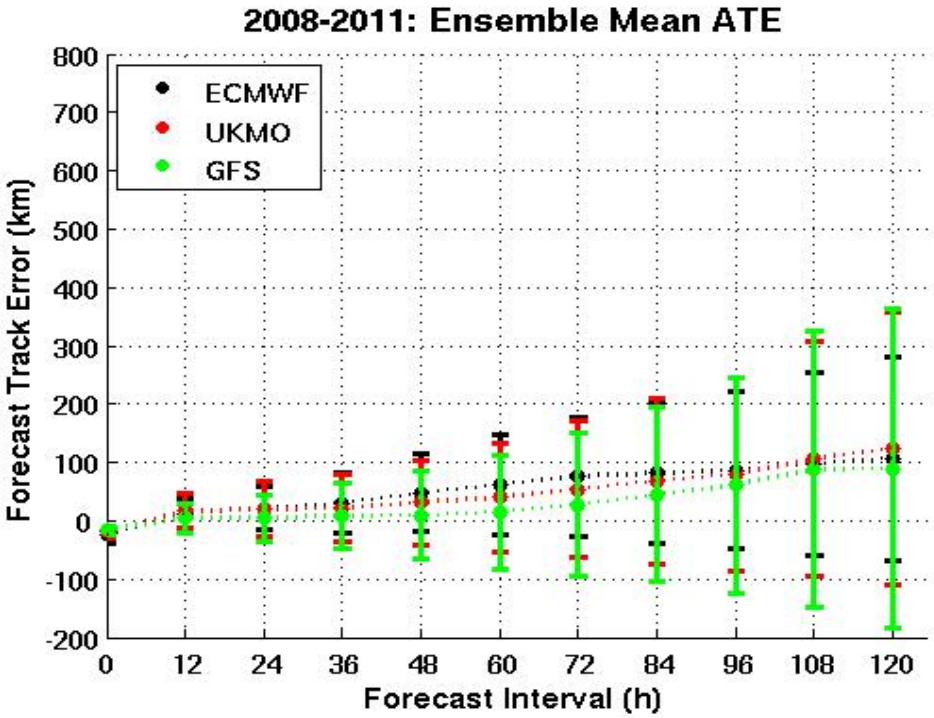

Figure 17. Average Along-Track Error (ATE) for each of the Ensemble Prediction Systems over the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the ATE is represented by a bar graph at each 12 h forecast tau from 0–120 h.

Across the MDR (Figure 18), the UKMO ensemble mean has an average ATE of near zero through the 60 h forecast, which means the UKMO ensemble is more consistent in accurately forecasting the translation speed of TCs across the MDR for short- and mid-range forecasts compared to the ECMWF and GFS. The GFS is the least accurate ensemble in this region, which may be because that system has no special perturbations that are specifically designed for individual tropical cyclones in the deep tropics. The

average ATE for GFS is close to zero for long-range forecasts, but also has the largest standard deviation, which indicates that these ensemble forecasts may be way ahead or way behind the actual track position, but the average error happens to be near zero.

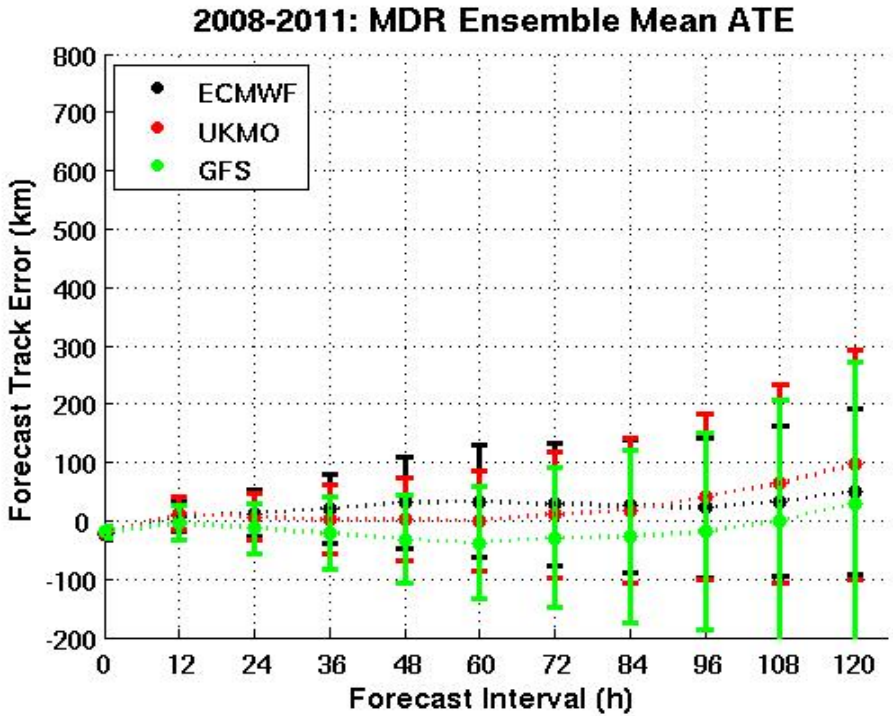

Figure 18. Average Along-Track Error (ATE) for each of the Ensemble Prediction Systems over the Main Development Region (MDR) of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the ATE is represented by a bar graph at each 12 h forecast tau from 0–120 h.

The ATEs for the ECS sub-region (Figure 19) indicate the GFS has the lowest average ATE for all forecast intervals, which is consistent with the result in Figure 15 that the GFS is the most accurate EPS in the ECS sub-region. However, the GFS does have the largest standard deviation past the 48 h forecast interval. These standard deviation values extending below -200 km in the long-range forecasts (>96 h) indicates that the GFS sometimes forecasts the movement of the TC much slower than the actual speed. Such slow forecasts may be the result of the GFS not properly predicting the influence of the strong mid-latitude steering flow as the TC begins to accelerate following recurvature.

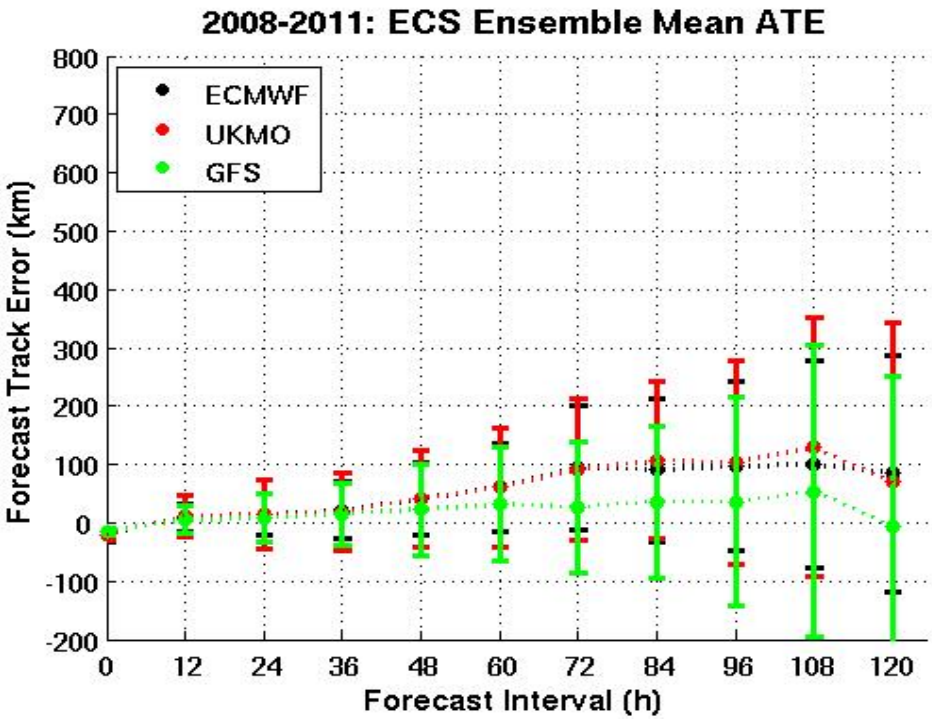

Figure 19. Average Along-Track Error (ATE) for each of the Ensemble Prediction Systems over the East Coast (ECS) region of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the ATE is represented by a bar graph at each 12 h forecast tau from 0–120 h.

Across the GOM sub-region (Figure 20), the ATE is significantly higher for all EPSs beyond 72 h. The increased variability in TC steering flow over the GOM is evident for long-range forecasts beyond 72 h. The GFS again has a very high standard deviation, which indicates the presence of extreme ATE outliers. The ECMWF and UKMO perform relatively the same across this sub-region.

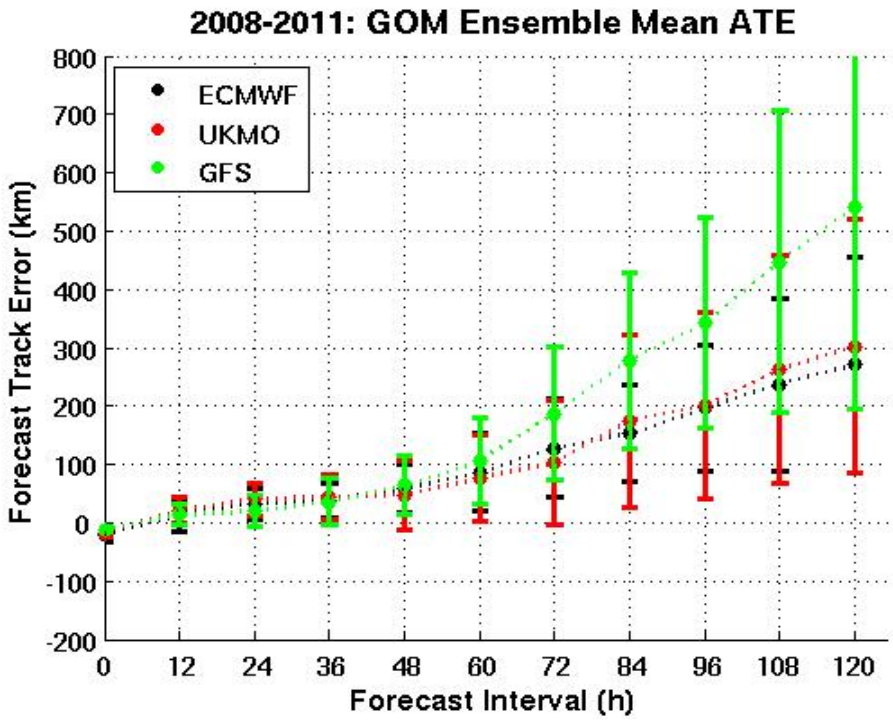

Figure 20. Average Along-Track Error (ATE) for each of the Ensemble Prediction Systems over the Gulf of Mexico (GOM) region of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the ATE is represented by a bar graph at each 12 h forecast tau from 0–120 h.

3. XTE

The Cross-Track Error (XTE) is the left or right component of the Forecast-Track Error (FTE). Positive (negative) values indicate that the forecast position is to the right (left) of the verifying best track position. The XTE can be due to the ensemble error in predicting both the orientation and speed in the background steering flow.

The GFS and ECMWF ensemble means both have very low average XTEs across the Atlantic basin for all forecast intervals (Figure 21). The standard deviation for the ECMWF is smaller than the GFS, which leads to the conclusion that the ECMWF is more consistent in accurately predicting the orientation of the background flow.

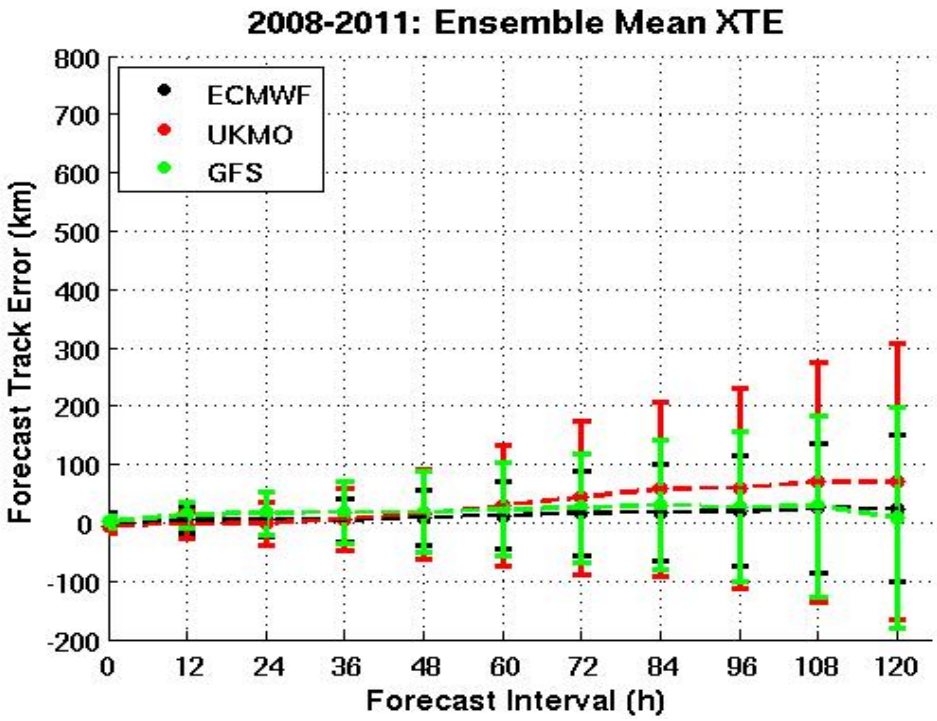

Figure 21. Average Cross-Track Error (XTE) for each of the Ensemble Prediction Systems of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the XTE is represented by a bar graph at each 12 h forecast tau from 0–120 h.

Across the MDR (Figure 22), the ECMWF has the lowest average XTEs and the smallest standard deviations compared to the GFS and UKMO. The GFS is the worst ensemble for long-range forecast XTEs, which may be the result of an inability to forecast when TCs begin to turn toward a north-west direction on the west side of the Bermuda sub-tropical high. By comparison, the UKMO forecasts the TCs to turn toward a more northward component of the actual track.

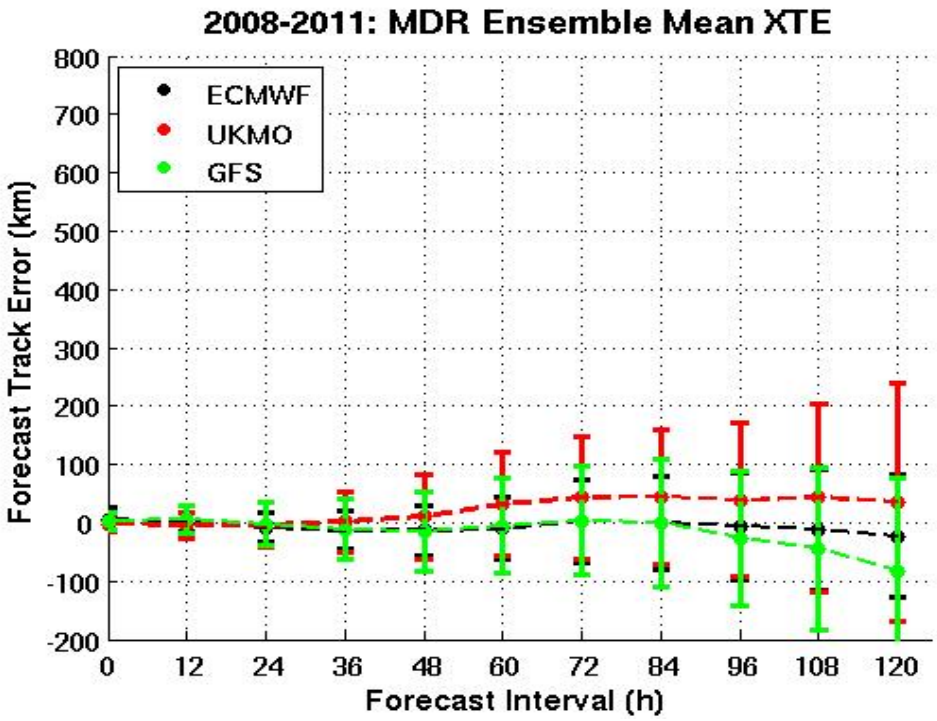

Figure 22. Average Cross-Track Error (XTE) for each of the Ensemble Prediction Systems over the Main Development Region (MDR) of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the XTE is represented by a bar graph at each 12 h forecast tau from 0–120 h.

All three EPSs have a near-zero average XTE throughout the ECS (Figure 23) during short- and mid-range forecasts (< 60 h). Beyond the 60 h forecast, each EPS has a negative average XTE that indicates that the ensembles have a leftward bias. Thus, as the TC moves into the mid-latitude westerlies, the forecast tracks would be more northeastward rather than eastward into the central Atlantic. The ECMWF does have the lowest XTEs and smallest standard deviations for this region.

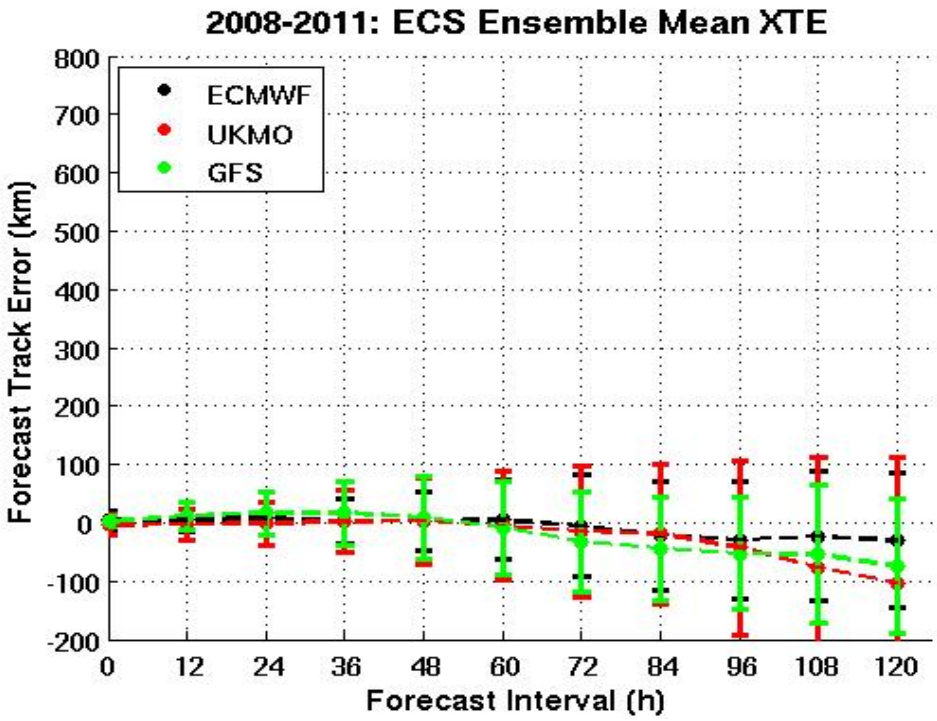

Figure 23. Average Cross-Track Error (XTE) for each of the Ensemble Prediciton System over the East Coast (ECS) region of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the XTE is represented by a bar graph for each 12 h forecast tau from 0–120 h.

Across the GOM sub-region (Figure 24), the average XTEs for the GFS and ECMWF greatly increase beyond the 48-h forecast interval. All three ensembles have large XTE outliers as indicated by the large standard deviations. The positive XTEs indicate each ensemble consistently forecasts the TC track too far to the right of the best-track.

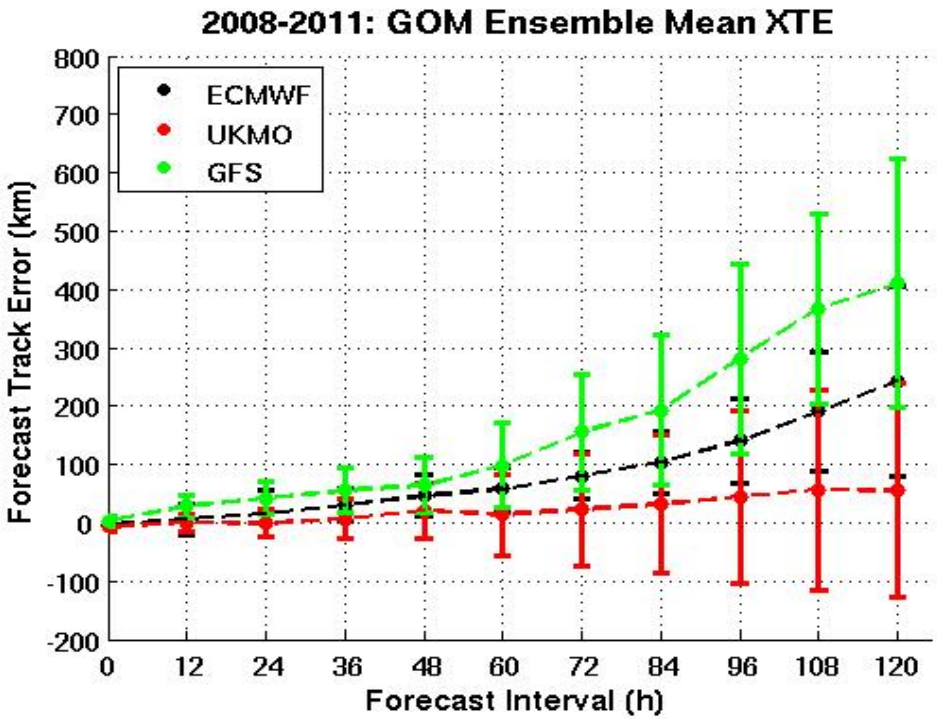

Figure 24. Average Cross-Track Error (XTE) for each of the Ensemble Prediction Systems over the Gulf of Mexico (GOM) region of the Atlantic basin from 2008–2011. A plus/minus one standard deviation of the XTE is represented by a bar graph at each 12 h forecast tau from 0–120 h.

C. STATISTICAL ANALYSIS

Statistical analysis is needed to calculate two main characteristics of each EPS when compared to the GE and GPCE. Resolution and reliability as defined in Chapter III will be measured by calculating the PWS, ellipse reliability, and the MAD. Each statistical analysese will include all TCs in the Atlantic basin or its sub-regions as defined in Table 3. The total of 3,422 EPS track forecasts are included.

1. Probability within Spread

The PWS is a great measure to determine whether an ensemble contains enough spread in its individual perturbed track forecasts to effectively forecast the actual TC track. An EPS that has a high (low) PWS for a particular forecast interval indicates that the individual ensemble forecast track members do (not) have enough spread to reflect the track uncertainty.

In the entire Atlantic region (Figure 25), each EPS has a low PWS starting at the intial 00 h time interval, which indicates the EPSs did not have accurate positions of where the TC was located based on the best-track position. The PWS does increase steadily throughout the forecast intervals when each EPS then has more spread in its forecast members. The ECMWF overall has the highest PWS past the 24 h forecast interval. The GFS consistently has the lowest PWS for the entire Atlantic basin and sub-regions (Figures 25–28), which indicates the GFS has too little spread in its forecast members and high resolution.

Across the MDR (Figure 26), the ECMWF has a much higher PWS compared to the GFS and UKMO. The ECMWF PWS remains relatively constant beyond 24 h, which indicates that the spread in forecast members also remains nearly constant. With the spread of forecast members not increasing much with longer forecast intervals, the ECMWF ensemble has a relatively low level of variability in TC forecast tracks.

For the ECS sub-region (Figure 27), the PWSs are largest among all regions with all EPSs above 80% verification by 120 h. Recall the FTEs were large in this sub-region for each EPS (Figure 15). However, the spread among the forecast members is very large, which results in a high PWS. The ECMWF has the largest PWS throughout the ECS, but high spread in the forecast members indicates increased uncertainty, which will need to be checked.

The GOM sub-region (Figure 28) also has a high PWS at long-range forecast intervals (>96 h), which indicates large spread and variability for all EPS. The ECMWF has a decrease in PWS from 72 h to 84 h, which indicates that the spread among the forecast members did not increase enough to represent the level of forecast uncertainty.

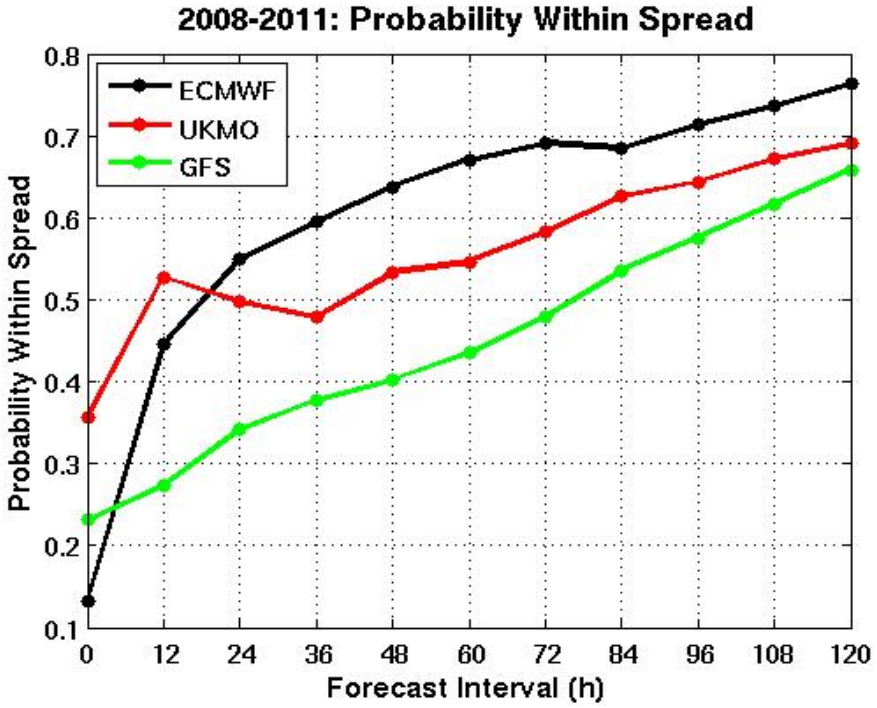

Figure 25. Probability within Spread (PWS) for each EPS across the Atlantic basin for the 2008–2011 Atlantic hurricane seasons. Probabilities on the left correspond to the percentage of the forecasts that the +/- standard deviation of each EPS forecast members includes the TC best-track position at each forecast interval.

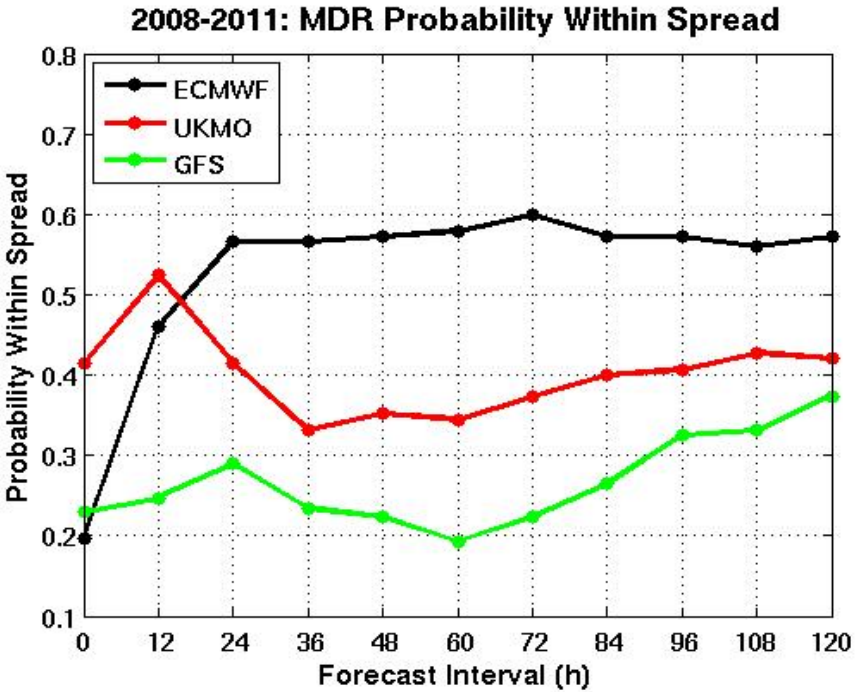

Figure 26. Probability within Spread (PWS) for each EPS across the Main Development Region (MDR) Atlantic sub-region for the 2008–2011 Atlantic hurricane seasons. Probabilities on the left correspond to the percentage of the forecasts that the +/- standard deviation of each EPS forecast members includes the TC best-track position at each forecast interval.

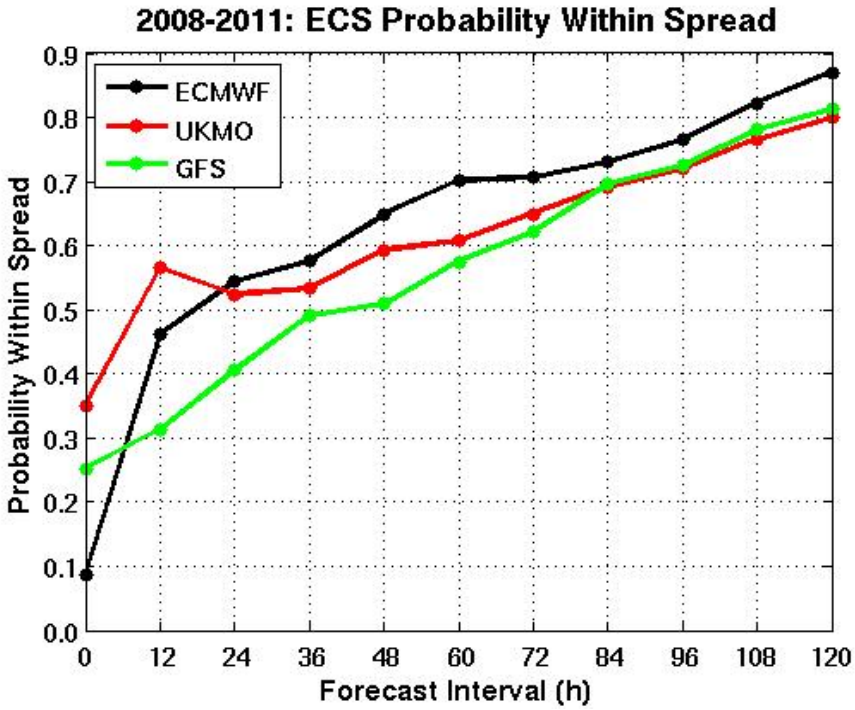

Figure 27. Probability within Spread (PWS) for each EPS across the East Coast Storms (ECS) Atlantic sub-region for the 2008–2011 Atlantic hurricane seasons. Probabilities on the left correspond to the percentage of the forecasts that the +/- standard deviation of each EPS forecast members includes the TC best-track position at each forecast interval.

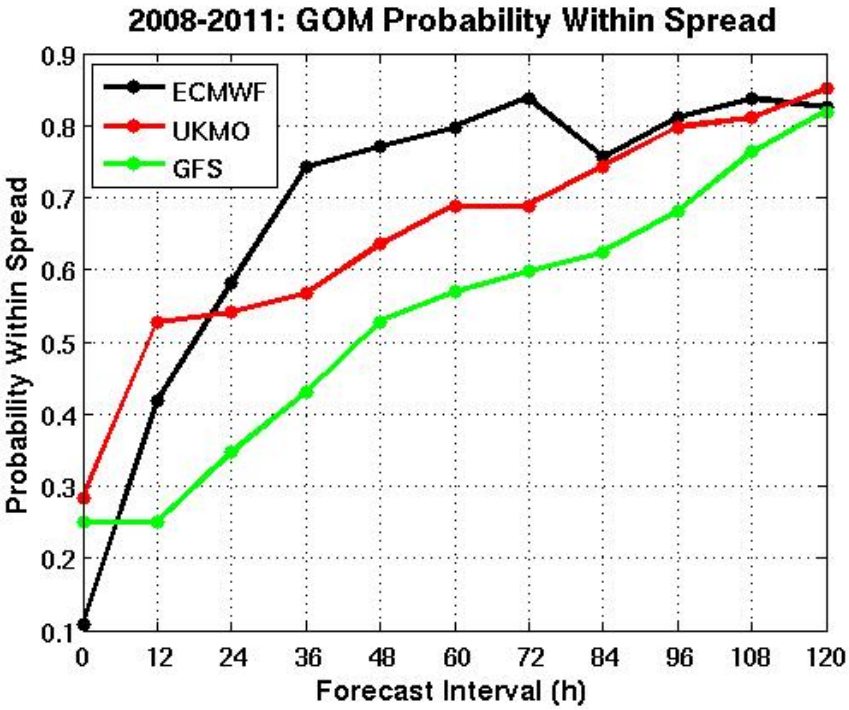

Figure 28. Probability within Spread (PWS) for each EPS across the Gulf of Mexico (GOM) Atlantic sub-region for the 2008–2011 Atlantic hurricane seasons. Probabilities on the left correspond to the percentage of the forecasts that the +/- standard deviation of each EPS forecast members includes the TC best-track position at each forecast interval.

2. Ellipse Reliability

Ellipse reliability is a measure of how dependable the EPS ellipses are with respect to containing the TC best-track position. As discussed in Chapter III, each EPS ellipse is designed to include 68% of the individual forecast track members. Thus, it is expected the TC best-track position will be enclosed within the ellipse 68% of the time.

In Figures 29–32, the reliabilities of individual EPS and GE ellipses are shown by line graphs for each forecast interval. The blue line symbolizes the expected reliability of the ellipses at 68%. The values above the forecast interval are the number of EPS forecasts that were included.

In Figure 29, the reliability for all of the EPSs are below 68%, which indicates the ellipses are under-reliable. Thus, there is not enough spread among the EPS individual forecast tracks to effectively enclose the TC best-track position. The ECMWF has the

highest reliability past 24 h relative to either the GFS and UKMO. Considering the GE is composed of all three EPS, the reliability is higher than the ECMWF alone because of the increased spread among the individual forecast track members. Throughout the Atlantic, it is clear the GE benefits the most from the ECMWF, so that the GE has reliabilities near 60% past 60 h. By contrast, the UKMO and GFS reliabilities are well below 40% past 60 h.

Across the MDR (Figure 30), the GE and ECMWF have ellipse reliabilities near 68% past 48 h. Although the UKMO has a higher reliability than the ECMWF at 12 h, the GE gains the most benefit from the ECMWF during all other forecast intervals. Notice the GFS is far under-reliable.

In the ECS sub-region (Figure 31), the GE has nearly 68% reliability through 84 h. Between 12 h and 48 h, all three EPSs contribute to the GE reliability. The ECMWF EPS then contributes the most to the GE reliability at longer forecast intervals. However, the GE becomes under-reliable during long-range forecast intervals because the ECMWF also drops significantly in reliability past 84 h. In these longer ranges, the GFS and UKMO reliabilities remain between 10–30%.

The GOM sub-region (Figure 32) has similar reliabilities in that the ECMWF has the highest reliability among the individual EPSs and contributes the most to the GE past 24 h. All of the EPSs and the GE become progressively more under-reliable during long-range forecast intervals past 84 h.

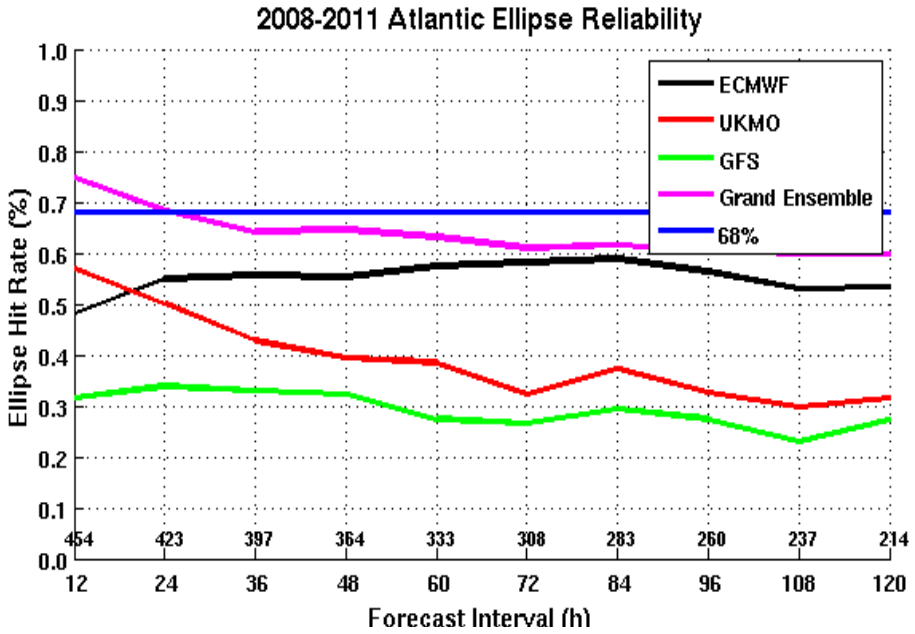

Figure 29. Ellipse reliability for each EPS and GE for the entire Atlantic basin. The line graph represents the verification (hit rate) that the TC best-track forecast position falls within the EPS ellipse that contains 68% of its forecast members. The blue line represents the 68% reliability level. The number above the forecast interval is the number of ensemble forecasts from 2008–2011 included.

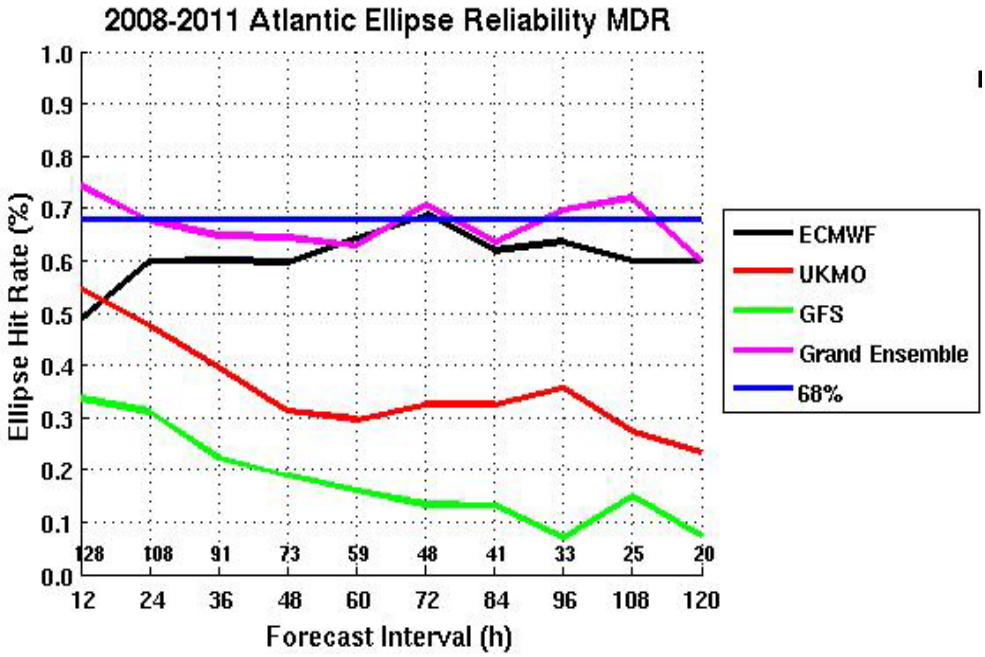

Figure 30. Ellipse reliability for each EPS and GE for the Main Development Region (MDR) of the Atlantic basin. The line graph represents the verification (hit rate) that the TC best-track forecast position falls within the EPS ellipse that contains 68% of its forecast members. The blue line represents the 68% reliability level. The number above the forecast interval is the number of ensemble forecasts from 2008–2011 included.

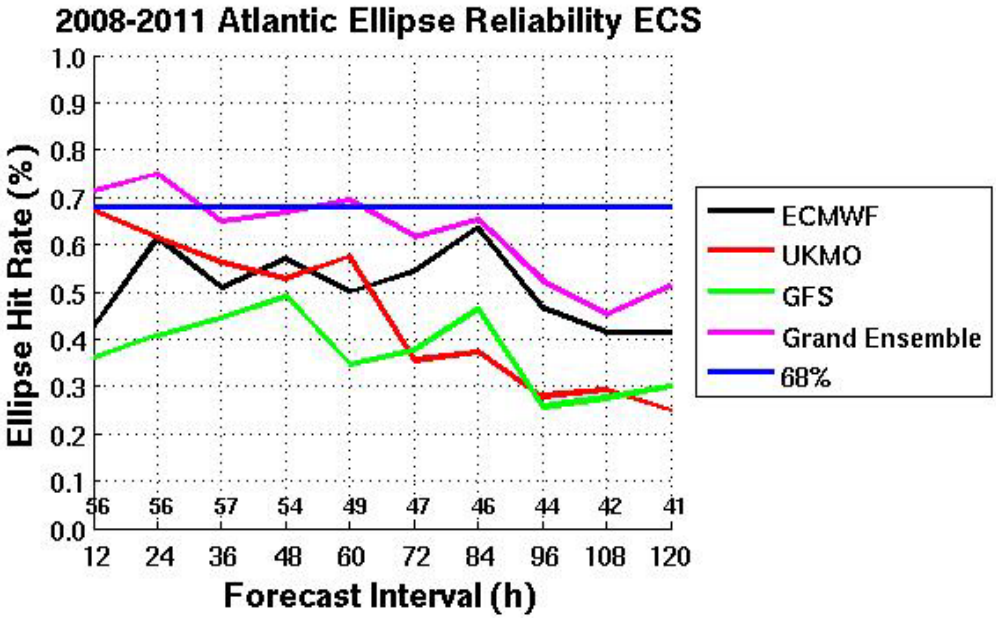

Figure 31. Ellipse reliability for each EPS and GE for the East Coast Storms (ECS) sub-region of the Atlantic basin. The line graph represents the verification (hit rate) that the TC best-track forecast position falls within the EPS ellipse that contains 68% of its forecast members. The blue line represents the 68% reliability level. The number above the forecast interval is the number of ensemble forecasts from 2008–2011 included.

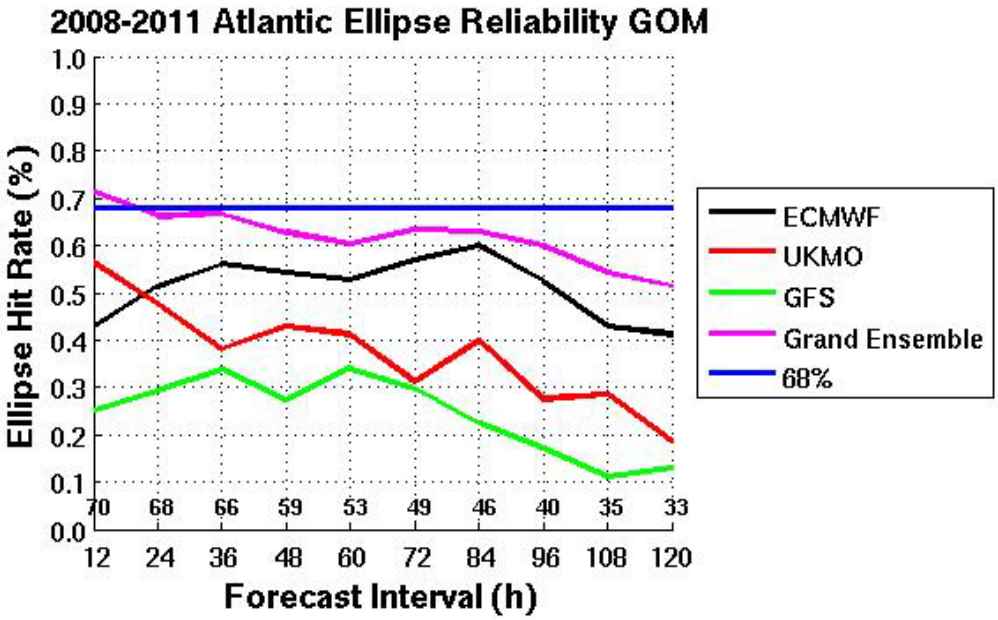

Figure 32. Ellipse reliability for each EPS and GE for the Gulf of Mexico (GOM) sub-region of the Atlantic basin. The line graph represents the verification (hit rate) that the TC best-track forecast position falls within the EPS ellipse that contains 68% of its forecast members. The blue line represents the 68% reliability level. The number above the forecast interval is the number of ensemble forecasts from 2008–2011 included.

3. Mean Area Difference (MAD)

The MAD is a calculation that compares the sizes of the EPS and GE ellipses relative to the GPCE circle. A positive value indicates that the EPS or GE ellipse is smaller in size than the GPCE circle, and thus indicates the EPS / GE has a reduced level of uncertainty or higher resolution. In Figures 33–36, only the positive MAD values are plotted on the vertical axis as the negative MAD values are not of interest. The values above the forecast interval show the number of EPS forecasts included.

Across the Atlantic basin (Figure 33), all three EPS have positive MAD values for all forecast intervals, which means that all EPS ellipses have less spread in their ensemble members than the consensus spread indicated by the GPCE circle. The ECMWF and UKMO have very similar ellipse sizes prior to 36 h and past 96 h. The ECMWF has a slightly larger ellipse than the UKMO during the 48–84 h forecast intervals, but the reliability of the ECMWF is significantly higher than the UKMO. The GFS has the

highest MAD value across all forecast intervals, which means it has the highest resolution. The GE ellipse size is smaller than the GPCE circle for forecast times beyond 12 h and before 96 h.

In the MDR (Figure 34), each EPS and the GE have positive MAD values beyond 12 h. During the long-range forecast intervals (>96 h), the EPS and GE ellipses are between 50–60% smaller than the GPCE circle.

For the ECS sub-region (Figure 35), the EPS ellipses area again smaller than the GPCE circles. Notice the GE has negative MAD values beyond 48 h, which indicates a larger ellipse compared to the GPCE circle.

For the GOM (Figure 36), the number of EPS forecasts beyond 36 h are not sufficient to draw any conclusive results.

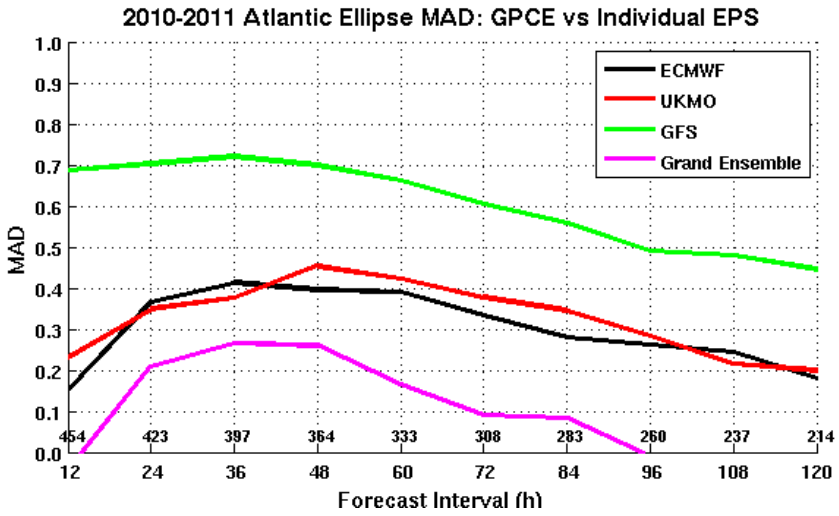

Figure 33. The Mean Area Difference (MAD) of each EPS and GE compared to GPCE across the Atlantic basin. Positive values indicate the EPS or GE ellipses are smaller than the GPCE circle for each forecast interval. The values above the forecast interval are the number of EPS forecasts included.

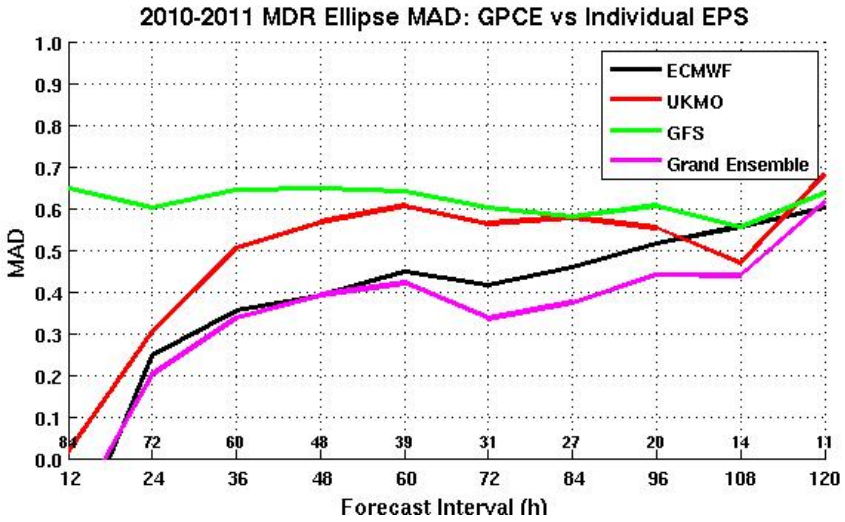

Figure 34. The Mean Area Difference (MAD) of each EPS and GE compared to GPCE across the MDR sub-region. Positive values indicate when the EPS or GE ellipses are smaller than the GPCE circle for each forecast interval. The values above the forecast interval are the number of EPS forecasts included.

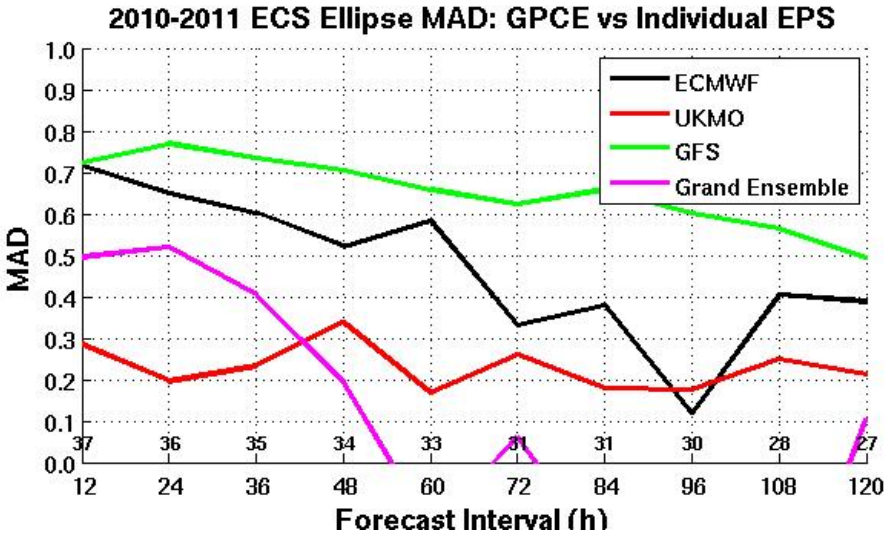

Figure 35. The Mean Area Difference (MAD) of each EPS and GE compared to GPCE across the ECS sub-region. Positive values indicate when the EPS or GE ellipses are smaller than the GPCE circle for each forecast interval. The values above the forecast interval are the number of EPS forecasts included.

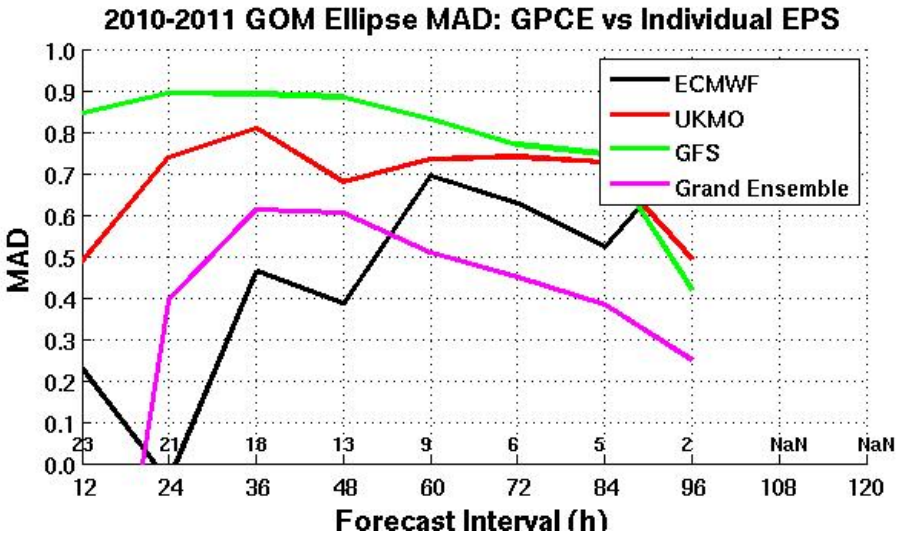

Figure 36. The Mean Area Difference (MAD) of each EPS and GE compared to GPCE across the GOM sub-region. Positive values indicate when the EPS or GE ellipses are smaller than the GPCE circle for each forecast interval. The values above the forecast interval are the number of EPS forecasts included.

D. SUMMARY

The ECMWF ensemble consistently outperforms the UKMO and GFS ensembles when it comes to TC forecast reliability. The ECMWF ensemble mean has the lowest FTE compared to the other EPS across the Atlantic basin for all forecast intervals to 120 h. The ECMWF also has the highest PWS throughout the Atlantic beyond 24 h. However, the ECMWF does not have accurate initial TC positions and does not have enough spread among the forecast tracks to ensure the TC best-track position is consistently within the spread at 12 h. The ECMWF tends to contribute the most to the GE reliability beyond 24 h. In most regions, GE does not benefit in terms of higher reliaibility from inclusion of the UKMO and GFS ensembles.

The GE on average has 5–10% higher reliability than the ECMWF from 36–120 h, but the uncertainty swath for the ECMWF ensemble is lower than the GE by an average of 25%. When comparing the ECMWF to the GPCE consensus error, the ECMWF has a 15% lower reliability on average, but uncertainty is reduced by 30%

through 120 h. The GFS ensemble has the overall highest resolution, or least amount of uncertainty, but has a very low reliability when compared to the ECMWF and UKMO ensembles.

E. CASE STUDY IRENE

Hurricane Irene was chosen as a case study because of the high level of uncertainty surrounding its forecast track during its early development and then the high impact along the east coast of the United States. The NHC consensus-based forecasts had the center of Irene passing south of Puerto Rico, and did not have enough spread among the members to include Puerto Rico within the cone of uncertainty (Figure 1a). Commanders at the two Puerto Rico DoD installations may not have had an accurate sense that a direct landfall was about to occur 24 h later.

On 20 August 2011, Tropical Storm Irene formed 120 n mi south of Martinique. Irene moved northwest and made landfall on Puerto Rico on 22 August 2011. During landfall, Irene was upgraded to hurricane strength and continued to move northwest toward the Bahamas. Hurricane Irene strengthened as it moved across the Bahamas, and reached category 3 strength on 24 August 2011. As Irene moved toward the North Carolina coast, it weakened and made landfall at Cape Lookout, North Carolina on the morning of 27 August 2011 as a category 1 hurricane. Irene moved parallel to the east coast of the United States and made landfall at Atlantic City, New Jersey and later Manhattan island on 28 August 2011. Although the storm weakened to a strong tropical storm during this time, significant rainfall leading to major flooding was experienced from Pennsylvania through the New England states before Irene became absorbed by an extratropical low on 30 August 2011.

1. Hurricane Irene 0000 UTC 21 August 2011

The Hurricane Irene case study begins 0000 UTC 21 August 2011 (Figure 37). The ECMWF was the only EPS that had enough spread in the forecast members such that its probability ellipse enclosed the island of Puerto Rico 24 h prior to making landfall. Both the GFS and UKMO ensembles had very little spread in the individual forecast members.

The orientation of the ECMWF ellipse at the 24 h forecast interval has a much larger cross-track component than the GFS and UKMO ellipses, which indicates that a high level of uncertainty as to how far north Irene's track was turning. The ECMWF had the largest ellipses up to 84 h, which is consistent with the above analysis that the ECMWF has the highest PWS, but lowest resolution of all three EPSs.

2. Hurricane Irene 0000 UTC 22 August 2011

At 0000 UTC 22 August 2011 (Figure 38), Irene was just hours away from making landfall over Puerto Rico. All three EPSs had forecast tracks across Puerto Rico, but the majority of the forecast track members were still south of the best-track positions. All three EPS ensemble means were far ahead of the best-track positions during the 12–60 h forecast intervals. The UKMO had the highest level of uncertainty as indicated by the ellipse major axis being parallel to the best-track movement.

3. Hurricane Irene 0000 UTC 24 August 2011

At 0000 UTC 24 August 2011 (Figure 39), Hurricane Irene was beginning to turn northward and recurve to the central Atlantic by the 84 h forecast. A larger fraction of the ECMWF ellipses contained the best-track positions compared to the GFS and UKMO ellipses, but also had the largest spread in the forecast track members, which resulted in the lowest resolution. The UKMO ensemble was consistenty too slow in predicting Irene's movement, which is consistent with the along-track error statistics in Figure 20. By contrast, the GFS ensemble forecast the storm to recurve too quickly.

4. Hurricane Irene 0000 UTC 25 August 2011

On 0000 UTC 25 August 2011 (Figure 40), Hurricane Irene is passing through the Bahamas and is 60 h from making landfall in North Carolina. The ECMWF continues to be the most accurate (highest PWS) and also has the smallest cross-track error, which is consistent with the XTE summary along the ECS sub-region in Figure 23. Low cross-track error in model forecasts is crucial for reducing the uncertainty cone with TC

forecasts. The GFS ensemble continues to advance Irene into the mid-latitude westerlies with an excessive ATE. Again, the UKMO ensemble underforecasts the Irene translation speed.

5. Hurricane Irene 0000 UTC 26 August 2011

At 0000 UTC 26 August 2011 (Figure 41), Hurricane Irene was 36 h from making landfall in North Carolina. The three EPSs had become into a higher level of agreement as to the left-right uncertainty in Irene's trajectory. The ECMWF ensemble had the lowest XTE through 36 h, but had significantly slowed the forecast speed of Irene, which resulted in high ATEs. For this forecast time, the GFS was the most accurate EPS, although it had had the largest XTEs and FTEs through the previous forecasts. However, it might have been expected from the FTE analysis in Figure 15 that the GFS ensemble would be among the most reliable as Hurricane Irene was entering the ECS sub-region.

6. Hurricane Irene 0000 UTC 27 August 2011

Just 12 h from landfall on 0000 UTC 27 August 2011 (Figure 42), the GFS was again the most reliable EPS with the largest ellipse reliability and resolution through the 48 h forecast period. The UKMO had the highest spread (largest ellipses), but also has the largest FTE and ATE for the 36- and 48-h forecasts. The ECMWF ensemble mean was also slow in the Irene translation speed forecast.

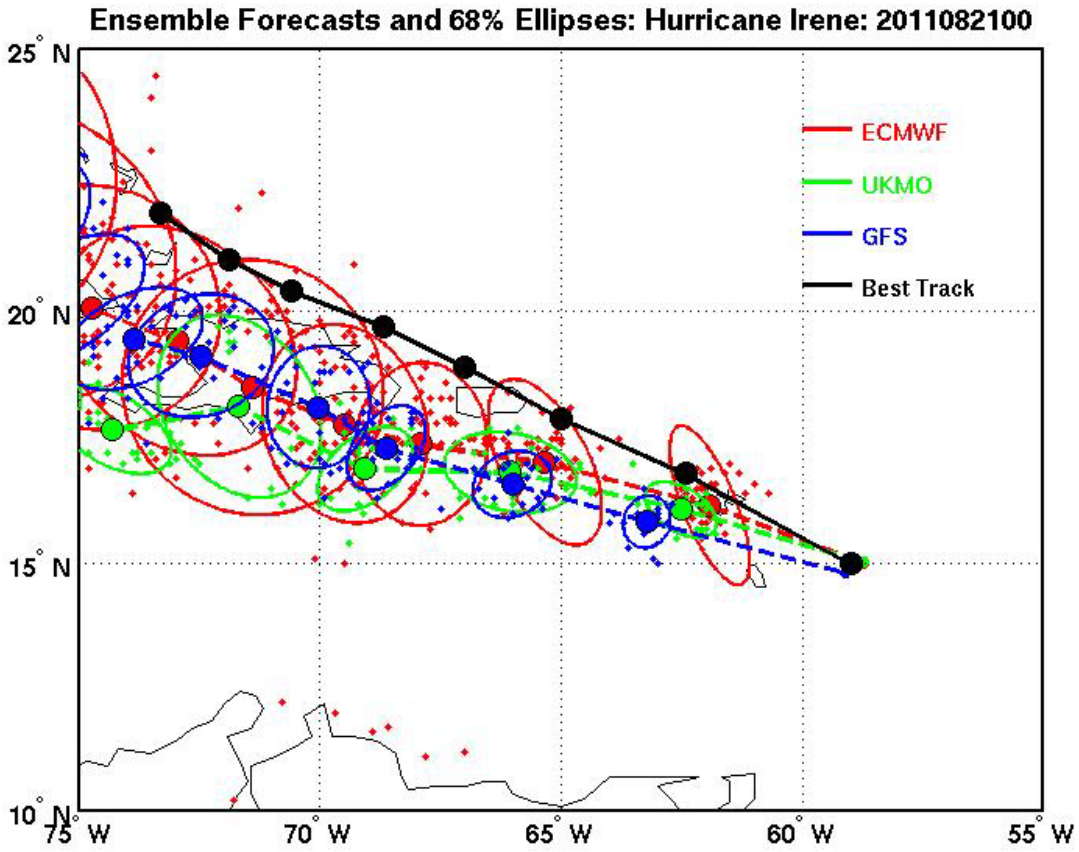

Figure 37. The TC forecast track ellipses of each EPS for Hurricane Irene at 0000 UTC 21 August 2011. Each ellipse signifies a 12 h forecast interval and is colored to match the individual EPS as defined in the legend at the top right. The large dot inside each ellipse is the corresponding ensemble mean forecast position. The best-track positions are in black. For geographical reference, Puerto Rico is slightly left of center near 18°N, 67°W.

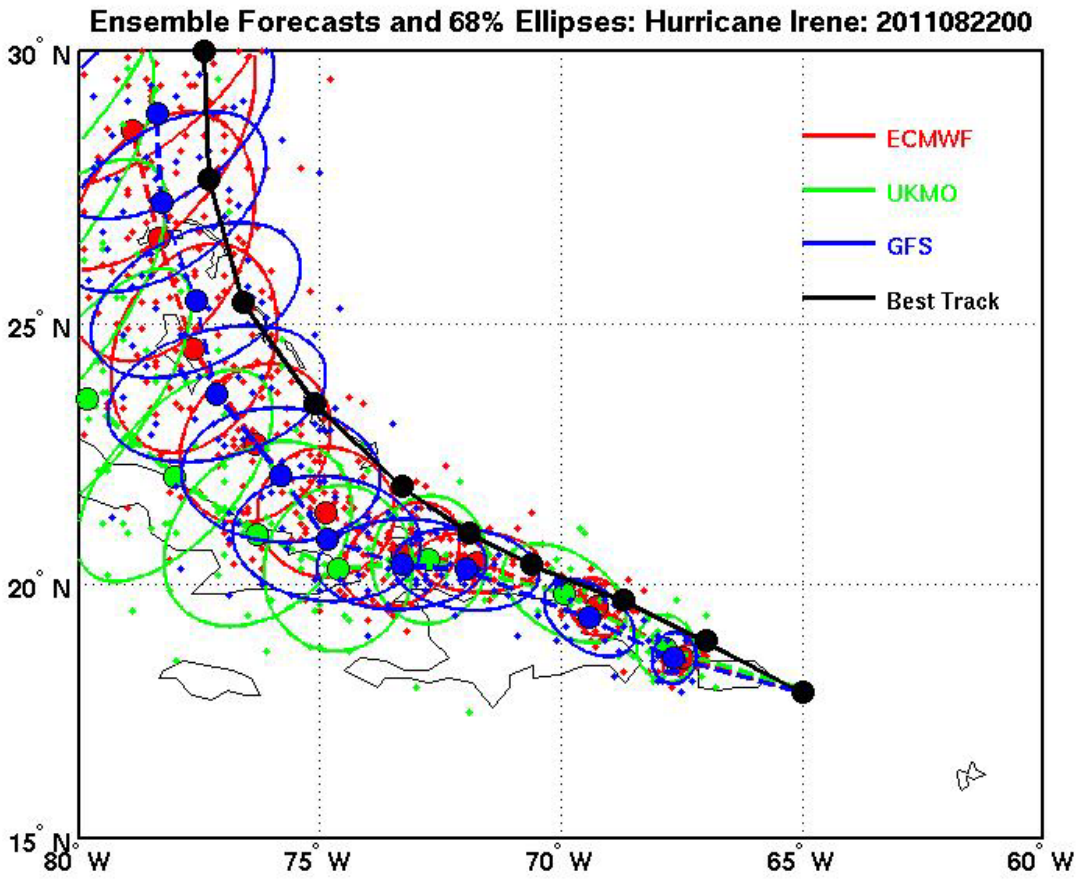

Figure 38. The TC forecast track ellipses of each EPS for Hurricane Irene at 0000Z UTC 22 August 2011. Each ellipse signifies a 12 h forecast interval and is colored to match the individual EPS as defined in the legend at the top right. The large dot inside each ellipse is the corresponding ensemble mean forecast position. The best-track positions are in black. For geographical reference, Puerto Rico is slightly bottom-right near 18°N, 67°W.

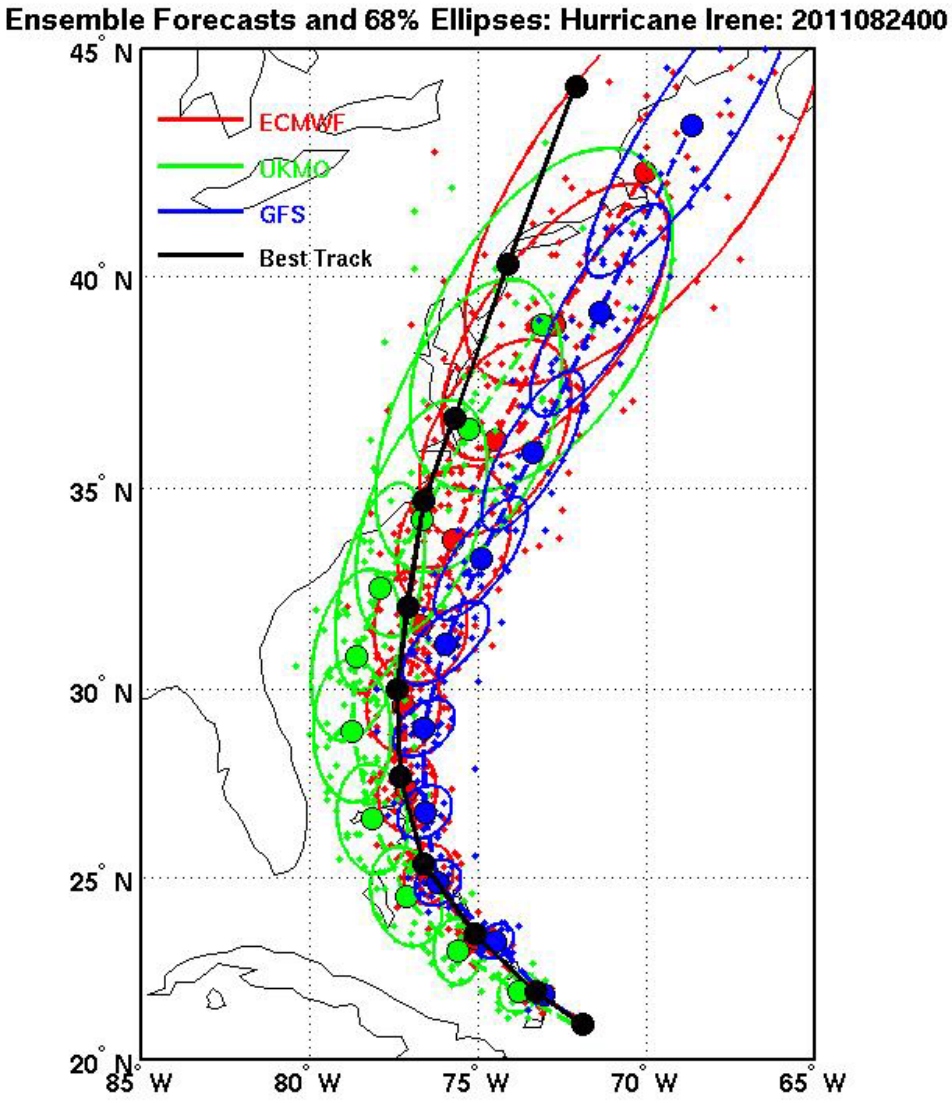

Figure 39. The TC forecast track ellipses of each EPS for Hurricane Irene at 0000Z UTC 24 August 2011. Each ellipse signifies a 12 h forecast interval and is colored to match the individual EPS as defined in the legend at the top right. The large dot inside each ellipse is the corresponding ensemble mean forecast position. The best-track positions are in black. For geographical reference, east coast of the United States is located at the left side of the image.

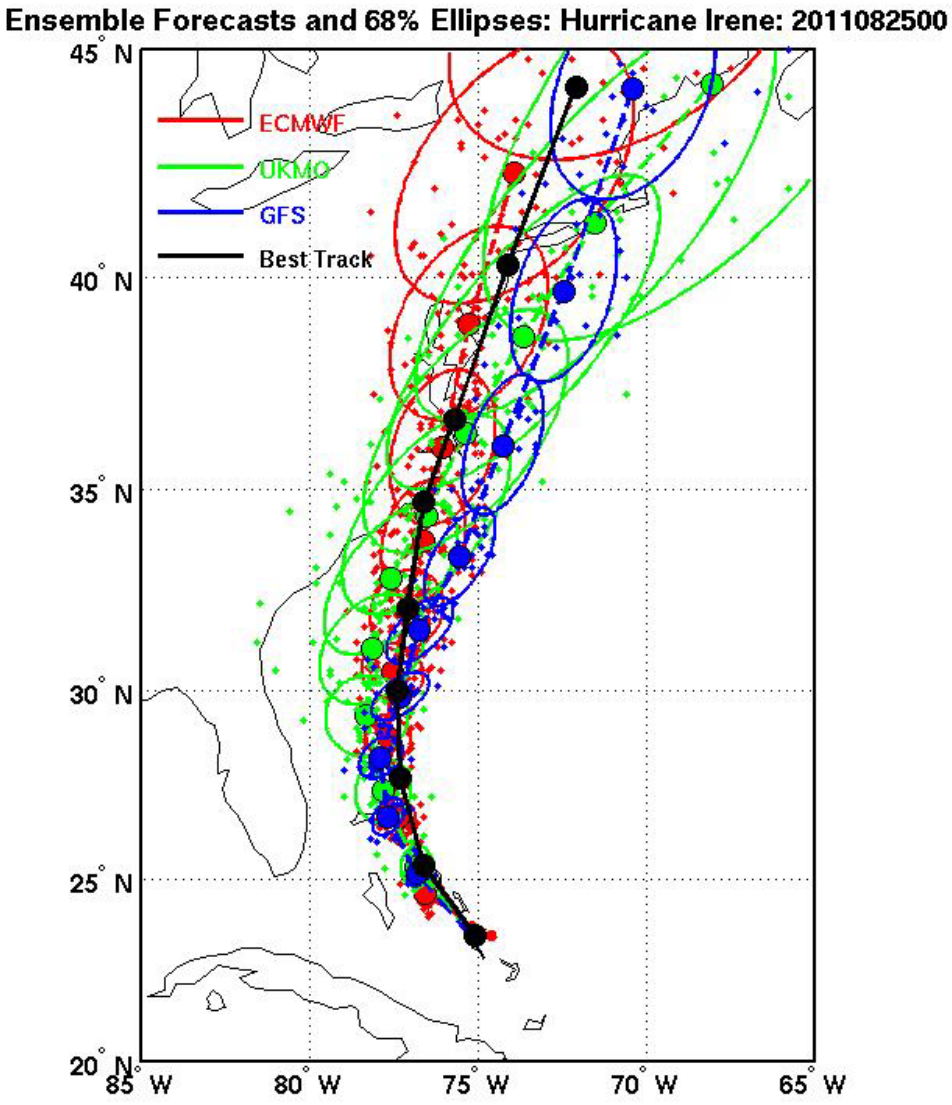

Figure 40. The TC forecast track ellipses of each EPS for Hurricane Irene at 0000Z UTC 25 August 2011. Each ellipse signifies a 12 h forecast interval and is colored to match the individual EPS as defined in the legend at the top right. The large dot inside each ellipse is the corresponding ensemble mean forecast position. The best-track positions are in black. For geographical reference, east coast of the United States is located at the left side of the image.

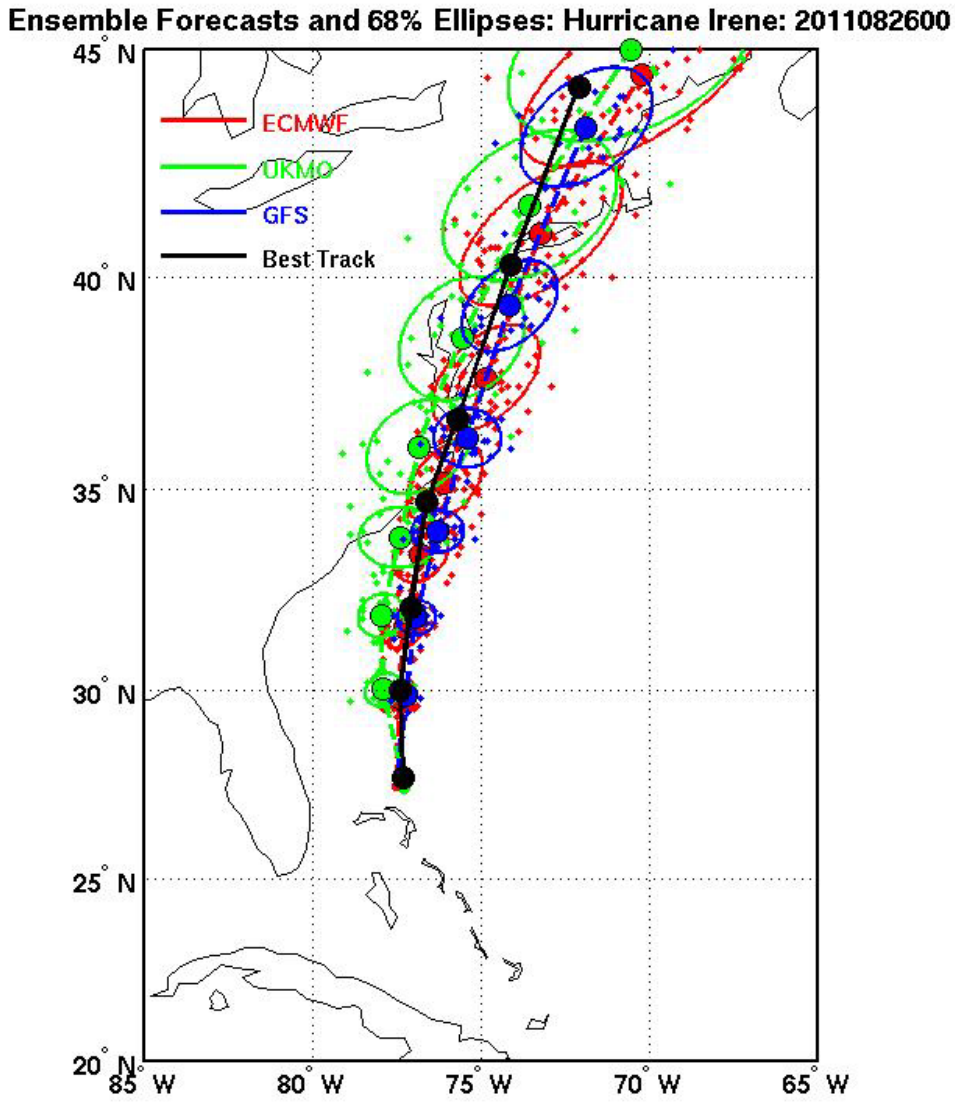

Figure 41. The TC forecast track ellipses of each EPS for Hurricane Irene at 0000Z UTC 26 August 2011. Each ellipse signifies a 12 h forecast interval and is colored to match the individual EPS as defined in the legend at the top right. The large dot inside each ellipse is the corresponding ensemble mean forecast position. The best-track positions are in black. For geographical reference, east coast of the United States is located at the left side of the image.

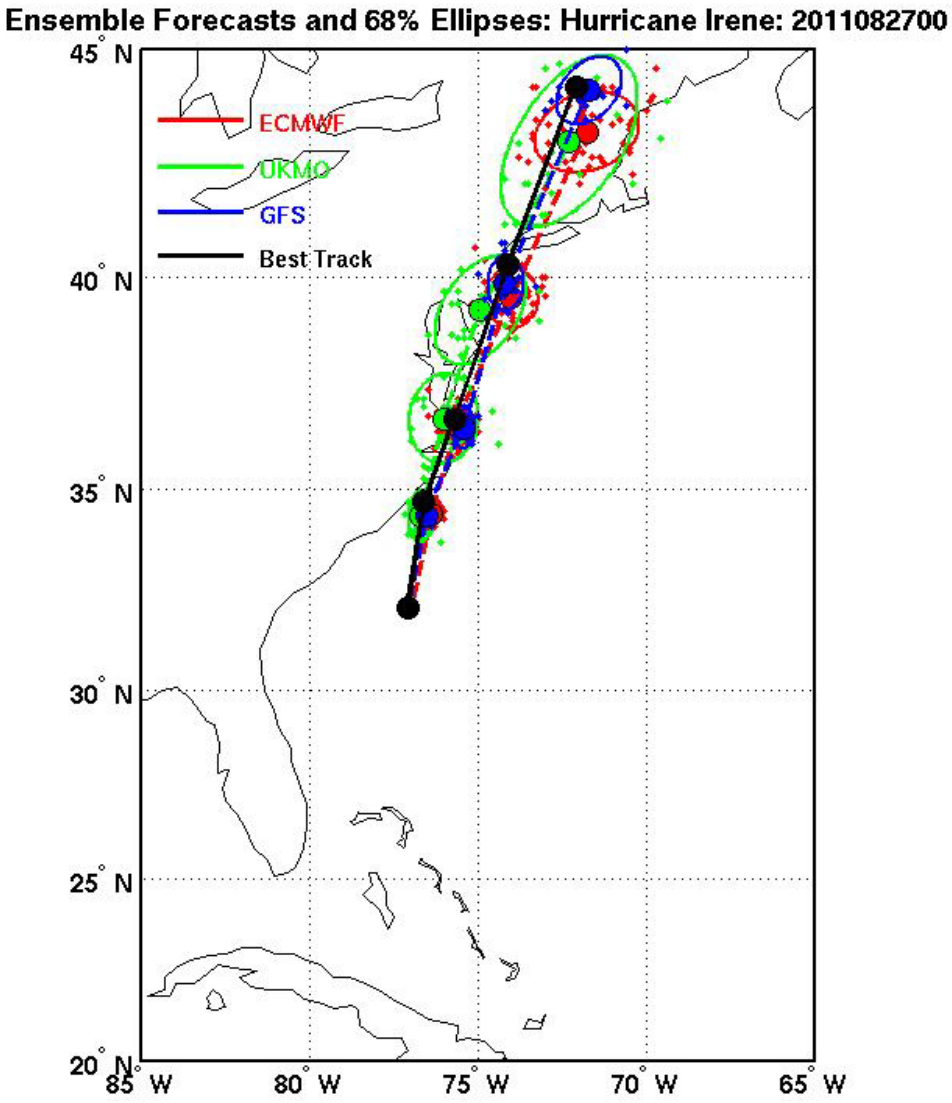

Figure 42. The TC forecast track ellipses of each EPS for Hurricane Irene at 0000Z UTC 27 August 2011. Each ellipse signifies a 12 h forecast interval and is colored to match the individual EPS as defined in the legend at the top right. The large dot inside each ellipse is the corresponding ensemble mean forecast position. The best-track positions are in black. For geographical reference, east coast of the United States is located at the left side of the image.

THIS PAGE INTENTIONALLY LEFT BLANK

V. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

This study has statistically analyzed the error in TC track forecasts with a focus on the error in EPS models. The NHC currently creates TC track forecasts and track uncertainty cones by examining mean and the spread among several deterministic models to form a consensus forecast. The objective of examining these TC track errors from individual ensemble models and a combination of ensembles such as the GE is to increase the reliability and resolution of the ensemble spread.

The first step was to calculate the TC track errors (FTE, ATE, and XTE) for the 2008–2011 hurricane seasons for each EPS mean over the entire Atlantic basin and three sub-regions of the Atlantic basin. The ECMWF ensemble mean has the lowest TC track errors in the overall Atlantic basin. The GFS ensemble mean was the most accurate of the three EPSs when TCs move into the ECS sub-region and begin to recurve to the central Atlantic.

Next, the PWS was calculated to determine which EPS had best predicted the spread or uncertainty among the individual forecast members, with the goal that the spread would contain the best-track position. The ECMWF ensemble consistently had the highest PWS after the 24 h forecast interval. The ECMWF ensemble spread typically did not contain the first 12 h best-track position due to inaccurate placement of the TC during the initialization step. The UKMO ensemble had the most accurate initial TC positions and 12 h forecasts.

Ellipse reliability was calculated to determine whether the ellipse composed of 68% of the individual EPS and GE forecast members was able to consistently enclose the TC best-track position. The ECMWF ensemble had the highest reliability beyond 24 h relative to the GFS and UKMO EPSs. The GE reliability mirrored the trends in the ECMWF reliability during mid- and long-range forecasts, which led to the conclusion that the ECMWF contributes the most to the GE during forecast times beyond 24 h. Within 24 h, the GE benefited most from the UKMO ensemble.

Finally, the MAD was created to compare the resolution of each EPS and the GE relative to the consensus-based forecast tool called GPCE. The ECMWF and UKMO ensembles had smaller areas of TC forecast track uncertainty when compared to GPCE area by an average of 30% over all forecast intervals. The GE reduced the area of uncertainty by an average of only 10% through the 96 h forecast time. Beyond 96 h, the GE had an increase in uncertainty compared to GPCE.

Based on the results of this study, a lot of benefit may be gained by producing TC forecasts based on the spread of individual ensembles. The ECMWF ensemble has the highest reliabilility among the EPSs across the Atlantic basin, and also has a higher resolution than the GE and GPCE. The GE could possibly be improved by applying factors for each EPS based on particular forecast interval. The UKMO ensemble should be the main contributor within 24 h forecasts, but then with the ECMWF ensemble should be the main contributor beyond 24 h. The resulting GE ellipse would be expected to have a higher resolution while also maintaining reliability.

B. RECOMMENDATIONS

Future research should explore whether the track forecasts would be improved by categorizing TC track errors according to the cyclone intensity. That is, stronger hurricanes might have lower EPS track errors due to a better depiction of the cyclone structure. This research could also be expanded to examine the tropical Pacific basin to determine if the results from the Atlantic basin are reproduced.

A modification of the GE composition based on these results could result in a forecast product that has a reduced area of track forecast uncertainty and a near-equivalent reliability to the GPCE product. This would allow forecasters to reduce TC warning areas and reduce government costs in preparation for a possible TC landfall.

LIST OF REFERENCES

Bowler N. E, A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare, 2008: The MOGREPS short-range ensemble prediction system. Q. J. Roy. *Meteorol. Soc.*, **134**, 706.

Buckingham, C., T. Marchok, I. Ginis, L. Rothstein, and D. Rowe, 2010: Short- and medium-range prediction of tropical and transitioning cyclone tracks within the NCEP global ensemble forecasting system. *Weather and Forecasting*, **25**, 1741–1742.

DeMaria, M., J. A. Knaff, R. Knabb, C. Lauer, C. R. Sampson, and R. T. DeMaria, 2009a: A new method for estimating tropical cyclone wind speed probabilities. *Weather and Forecasting*, **24**, 1574–1575

ECMWF, cited 2012: Different perturbation techniques [Available online at http://www.ecmwf.int/products/forecasts/guide/Different_perturbation_techniques.html] accessed 28 March 2012.

Goerss, J S., 2007a: Prediction of consensus tropical cyclone track forecast error. *Mon. Wea. Rev*., **135**, 1985–1990

Neese, J. M., 2010: Evaluating Atlantic tropical cyclone track error distributions for use in probabilistic forecast of wind distribution. M. S. thesis, Naval Postgraduate School, 87 pp. [Available from http://edocs.nps.edu/].

New York Times, cited 2011: "Irene Damage May Hit $7 Billion, Adding to Insurer Woes" [Available online at http://www.nytimes.com/2011/08/29/business/irene-damage-may-hit-7-billion-adding-to-insurer-woes.html?_r=2&scp=20&sq=hurricane%20irene%20new%20jersey&st=cse] accessed 7 March 2012.

NHC, cited 2012a: Irene Graphics Archive [Available online at http://www.nhc.noaa.gov/archive/2011/graphics/al09/loop_5W.shtml] accessed 4 April 2012.

NHC, cited 2012b: NHC Graphical Product Descriptions [Available online at http://www.nhc.noaa.gov/aboutnhcgraphics.shtml?] accessed 9 March 2012.

NHC, cited 2012c: Definition of the NHC Track Forecast Cone [Available online at http://www.nhc.noaa.gov/aboutcone.shtml] accessed 9 March 2012.

NHC, cited 2012d: Irene Graphics Archive [Available online at http://www.nhc.noaa.gov/archive/2011/graphics/al09/loop_5W.shtml] accessed 4 April 2012.

NHC, cited 2012e: Tropical Cyclone Wind Speed Probability Products [Available online at http://www.nhc.noaa.gov/aboutnhcprobs.shtml] accessed 9 March 2012.

NHC, cited 2012f: NHC Graphical Product Descriptions [Available online at http://www.nhc.noaa.gov/aboutnhcgraphics.shtml?] accessed 9 March 2012.

NHC, cited 2012g: Irene Graphics Archive [Available online at http://www.nhc.noaa.gov/archive/2011/graphics/al09/loop_PROB50.shtml] accessed 4 April 2012.

NHC, cited 2012h: Irene wind speed probability forecast 27 August 2011 [Available online at http://www.nhc.noaa.gov/archive/2011/al09/al092011.wndprb.029.shtml] accessed 9 March 2012.

NHC, cited 2012i: Topical Weather Summary [Available online at http://www.nhc.noaa.gov/archive/2008/tws/MIATWSAT_nov.shtml] accessed 11 March 2012.

NHC, cited 2012j: 2008 Atlantic Hurricane Season [Available online at http://www.nhc.noaa.gov/2008atlan.shtml] accessed 11 March 2012.

NHC, cited 2012k: Topical Weather Summary [Available online at http://www.nhc.noaa.gov/archive/2009/tws/MIATWSAT_nov.shtml] accessed 11 March 2012.

NHC, cited 2012l: 2009 Atlantic Hurricane Season [Available online at http://www.nhc.noaa.gov/2009atlan.shtml] accessed 11 March 2012.

NHC, cited 2012m: Topical Weather Summary [Available online at http://www.nhc.noaa.gov/archive/2010/tws/MIATWSAT_nov.shtml] accessed 11 March 2012.

NHC, cited 2012n: 2010 Atlantic Hurricane Season [Available online at http://www.nhc.noaa.gov/2010atlan.shtml] accessed 11 March 2012.

NHC, cited 2012o: Topical Weather Summary [Available online at http://www.nhc.noaa.gov/archive/2011/tws/MIATWSAT_nov.shtml] accessed 11 March 2012.

NHC, cited 2012p: 2011 Atlantic Hurricane Season [Available online at http://www.nhc.noaa.gov/2011atlan.shtml] accessed 11 March 2012.

Pearman, D. W., 2011: Evaluating tropical cyclone forecast track uncertainty using a grand ensemble of ensemble prediction systems. M. S. thesis, Naval Postgraduate School, 63 pp. [Available from http://edocs.nps.edu/].

UCAR, cited 2012: Guide to GFS Ensemble Track Plots. [Available online at http://www.ral.ucar.edu/guidance/guide/gfs/] accessed 28 March 2012.

Wilks, D. S., 1995: Statistical Methods in the Atmospheric Science, 1st edition. Academic Press, 236 pp.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Air Force Weathe Technical Library
   Asheville, North Carolina

4. Professor Patrick Harr
   Naval Postgraduate School
   Monterey, Califoria

5. Professor Russell Elsberry
   Naval Postgraduate School
   Monterey, California

6. Captain Christopher Nixon
   612 Support Squadron
   Davis-Monthan Air Force Base, Arizona

7. William P.Roeder
   45th Weather Squadron
   Patrick Air Force Base, Florida