NOAA

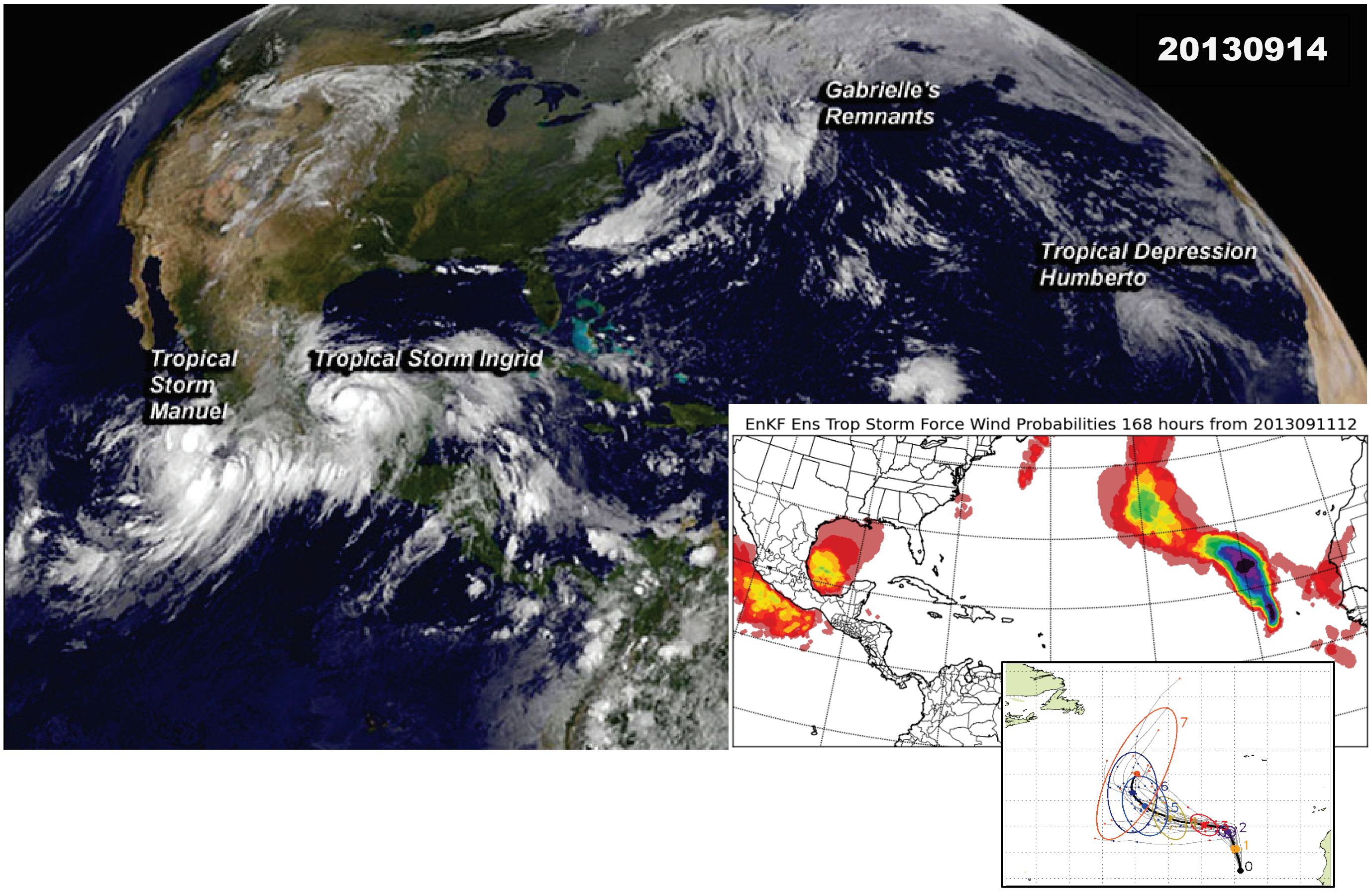
NATIONAL OCEANIC AND
ATMOSPHERIC ADMINISTRATION
United States Department of Commerce

HURRICANE FORECAST IMPROVEMENT PROJECT

2013 HFIP R&D Activities Summary: Recent Results and Operational Implementation

April 15, 2014

HFIP Technical Report: HFIP2014-2



2013 HFIP R & D Activities Summary: Recent Results and Operational Implementation

R. Gall[11], F. Toepfer[11], F. Marks[1], E.N. Rappaport[9], A. Aksoy[20], S. Aberson[1], J.W. Bao[4],
M. Bender[6], S. Benjamin[4], L. Bernardet[2,4], M. Biswas[7], B. Brown[7], J. Cangialosi[9], C. Davis[6],
M. DeMaria[9], J. Doyle[10], M. Fiorino[4], J. Franklin[9], I. Ginis[18], S. Gopalakrishnan [1], T. Hamill[4],
R. Hodur[14], H.S. Kim[3], J. Knaff[8], T. Krishnamurti [5], P. Kucera[7], Y. Kwon[3], W. Lapenta[3],
N. Lett[13], S. Lord[3], T. Marchok[6], E. Mifflin[16], M. Morin[6], K. Musgrave[2], L. Nance[7],
C. Reynolds[10], V.Tallapragada[3], H. Tolman[3], R. Torn[15],G. Vandenberghe[3], T. Vukicevic[9],
X. Wang[21], Y. Weng[12], J. Whittaker[4], R. Yablonsky[18], D-L Zhang[19], F. Zhang[12] , J. Zhang[20],
X. Zhang[20] , D. A. Zelinsky[9]

March 2014

[1] Atlantic Oceanographic and Meteorological Laboratory OAR/NOAA
[2] Cooperative Institute for Research in the Atmosphere, Colorado State University
[3] Environmental Modeling Center NCEP/NOAA
[4] Earth System Research Laboratory OAR/NOAA
[5] Florida State University
[6] Geophysical Fluid Dynamics Laboratory OAR/NOAA
[7] National Center for Atmospheric Research
[8] National Environmental Satellite Data Information Center NOAA
[9] National Hurricane Center NWS/NOAA
[10] Naval Research Laboratory, Monterey
[11] Office of Science and Technology, NWS/NOAA
[12] Pennsylvania State University
[13] Science and Technology Corporation
[14] Science Applications International Corporation
[15] University at Albany, State University of New York
[16] Syneren Technologies Corporation
[17] University Corporation for Atmospheric Research
[18] University of Rhode Island
[19] University of Maryland
[20] University of Miami
[21] University of Oklahoma

Table of Contents

Executive Summary i

1. Introduction 1

2. The Hurricane Forecast Improvement Project 1

3. The HFIP Model Systems 4

    a. The Global Models 4
    b. The Regional Models 5
    c. Initialization and Data Assimilation Systems 5
    d. The HWRF Community Code Repository 8

4. Meeting the HFIP Goals 8

    a. The HFIP Baseline 8
    b. Meeting the Track Goals 9
    c. Reaching the Intensity Goals 10

5. HFIP Stream 1.5 10

    a. Stream 1.5 results 12

6. Operational Hurricane Guidance Improvements 14

    a. Global Model (GFS)—Operational 14
    b. Global Model (GFS)—HFIP Experimental 16
    c. Hurricane WRF (HWRF) 21
        1) Atlantic 21
        2) West Pacific (WPAC) 22
        3) Rapid Intensification 24
        4) HWRF ensemble 25
        5) Evolution of the HWRF System 27

7. GFDL Ensemble 28

8. Impact of Inner Core Reconnaissance Data 32

9. Post Processing of Model Output 36

    a. Statistical Post Processing of Model Output 36
    b. Other Post Processing 38
    c. The HFIP Web Page 40

10. Storm Surge 41

11. Societal Impact Work 42

12. The Configuration of a Numerical Model Hurricane Forecast Guidance System to meet the HFIP goals 43

References 45

Appendix A: HFIP Funding Opportunity for Academia and Private Industry 47

Appendix B: Model Acronyms 48

List of Tables

Table 1. Strategic Teams 2013 2
Table 2. Tiger Teams 2013 3
Table 3. Specifications of the HFIP Global Models 5
Table 4. Specifications for the HFIP Regional Models 6
Table 5. 2013 HFIP Retrospective Evaluation Modeling Groups 11
Table 6. 2013 HFIP Stream 1.5 Real-time Runs Model Guidance Provided to NHC 11
Table 7. HWRF development 2013-2018 28
Table 8. GFDL hurricane model ensemble members 29
Table 9. Numerical Model Hurricane Forecast Guidance System 44

List of Figures

Figure 1. HFIP Track (left panel) and Intensity (right panel) Error Baseline and Goals. 8
Figure 2. Stream 1.5 Track Skills, 2013 Atlantic 12
Figure 3. Stream 1.5 Intensity Errors 2013 Atlantic 13
Figure 4. GFS and ECMWF Track Errors Relative to HFIP Baseline, Atlantic. 15
Figure 5. GFS and EMCWF Track Errors Relative to HFIP Baseline (2012-2013), WPAC 16
Figure 6. Performance Comparison Operational and Experimental GFS Systems 17
Figure 7. Tropical Storm Force Wind Probabilities within 168 Hours 18
Figure 8. Verification of Tropical Storm Force Wind Probabilities 19
Figure 9. Ensemble Member Tracks. 20
Figure 10. HWRF System Improvements 2011-2013 21
Figure 11. HWRF Track and Intensity Errors 2013 WPAC. 22
Figure 12. JTWC Official Intensity Forecast Errors, WPAC, 2000-2013 23
Figure 13. NHC Official Intensity Forecast Errors, Atlantic basin, 1990-2013 23
Figure 14. Forecasted 24-hour Intensity Changes in WPAC, 2013 24
Figure 15. HWRF 2013 Experimental Ensemble Track Performance, Atlantic Basin 25
Figure 16. HWRF 2013 Experimental Ensemble Intensity Performance, Atlantic Basin. 26
Figure 17. Basin Scale System with Three Moving Nests: Kirk, Isaac, and Ileana, 2012 27
Figure 18. 2013 GFDL Ensemble Intensity Errors, EPAC. 30
Figure 19. Intensity Errors for GFDL Ensemble Members: Hurricane Raymond. 31
Figure 20. HWRF Initial Condition Calculation Flow Chart. 33
Figure 21. Example RDITT Test Results: EMC HWRF, All Data. 34
Figure 22. Outlier Example: RDITT Test PSU Model, All Data and Control 35
Figure 23. FSU Multi-model Ensemble: Members and Mean Performance (2012 Atlantic). 37
Figure 24. Post Processing Diagnostics: Hurricane Sandy 250 hPa Geopotential Heights 38
Figure 25. Sensitivity of Intensity to Precipitable Water: Hurricane Bill 39
Figure 26. HFIP.org Products Page. 40
Figure 27. Storm Surge Advisory Product, Tropical Storm Karen 41
Figure 28. Preliminary Storm Surge Advisory Product. 42

Executive Summary

This report describes the activities and results of the Hurricane Forecast Improvement Program (HFIP) in 2013. Since this is the fourth year of the first five years of the project, we, like last year, focus on the improvements in operational Global Forecast System (GFS) global model and the Hurricane Weather Research and Forecasting (HWRF) regional model. HFIP is organized around the three "streams": Stream 1 or the operational model development, Stream 1.5 which comprises a group of experimental models that have been evaluated by the National Hurricane Center (NHC) pre-season and then made available to NHC forecasters during their forecast cycle, and Stream 2 representing HFIP experimental models which test and evaluate new techniques and strategies for model forecast guidance before testing is begun for possible operational implementation. Stream 2 also tests techniques that cannot be tested on current operational computers because of size and time requirements, but can be tested on HFIP computer facilities in Boulder, CO. Those studies are looking ahead to possible future operational computational capability. The report outlines the HFIP program, how it is organized, its goals, its models and then results from each of the three Streams.

Stream 1.0 Accomplishments

- The 2013 HWRF provided intensity forecasts that were the best ever for the model and which beat out the statistical models in the 24- to72-hour forecast lead time, even when run over three years of retrospective data.
- There were few long-lived storms in the Atlantic in 2013 so the error means reported beyond 72-hour forecast lead time are not statistically significant. At the earlier lead times HWRF continued to show the accuracy in intensity forecasts as shown by the retrospective runs.
- The HWRF was run in the western Pacific basin (WPAC) as a real-time system in 2013 and results made available to Joint Typhoon Warning Center (JTWC). The intensity was among the best guidance products for JTWC. Track guidance was also excellent.
- The global GFS continues to provide track guidance comparable to the European Center for Medium-range Weather Forecasts (ECMWF) guidance and exceeds the 5-year HFIP goal out to 4-day lead time.

- At longer lead times the improvement in track accuracy falls off dramatically for unknown reasons so that there has been little, if any, improvement at 7 days. The ECMWF model also did not seem to improve at 7 days indicating that the lack of improvement is likely a problem with global models in general.

Stream 1.5 Results

- The HWRF was run as a 20 member ensemble (using the same model as used operationally—called the control run), with as much as a 50% improvement over the control run.
- It was hard to measure "skill" of intensity forecasts since the model to which all models are usually compared to determine skill, SHIFOR5 (OCD5) performed very poor this year—see Figure 2.
- The six-member weighted Statistical Prediction of Intensity from a Consensus Ensemble (SPC3) statistical model was the best performing Stream 1.5 intensity model in 2012. This year it was among the worst. The best model out to 48 hours was the operational HWRF (HWFI) and the Florida State University super ensemble (FSSE). IV15/IVCI composites performed well as did the Wisconsin model (UW4I) in the 24- to 48-hour forecast lead times.
- A web page http://www.hfip.org/products/ to display HFIP stream 1.5 and other HFIP sponsored/run models has been running for the past two years.

Stream 2.0 Accomplishments

- The Reconnaissance Data Impact Tiger Team (RDITT) was formed to evaluate the impact of inner core aircraft derived data (flight level, dropsondes, SMFR and TDR data) on intensity and track forecasts. The TDR data did not show a clear or consistent benefit beyond what the standard reconnaissance data provided. The RDITT results suggest that inner core observations offer promise for improving operational tropical cyclone guidance, but that much work lies ahead to make optimal use of these data.
- Experimental GFS global deterministic and ensemble systems were again run parallel to the GFS operational model. The HFIP system used a semi-Lagrangian differencing scheme and was run at twice the resolution of both the GFS deterministic model and the GFS ensemble (GEF). Results were disappointing in that the higher resolution system did not perform as well as the operational system, likely because the physics and data assimilation packages had not been optimized for the higher resolution and the semi-implicit scheme because of time constraints in preparing the system for real-time runs in 2013.
- One major improvement noted from the global ensemble runs was that use of a stochastic physics process in the ensemble members gave much better ensemble spread than the current operational system.
- Forecasts of the 35-knot wind probability for Atlantic basin (ATL), eastern Pacific basin (EPAC), and western Pacific basin (WPAC) by the HFIP global GFS ensemble were excellent—see Figure 5.

- The performance of HWRF in forecasting rapid intensification (RI) in the WPAC showed a significant improvement in the probability of detection (POD) for RI in that region over the operational forecasts. The POD was 23%, the false alarm ration near was 0%.
- HFIP began closely working with the hurricane surge models in 2013, running ADCIRC on the Boulder system. There were few land falling systems this year so this year we primarily focused on developing the surge system. A forecast from Karen for southern Louisiana looked very good.
- A new product on presentation of storm surge data has been developed at NHC with partial support from HFIP. This product is likely to be operational this coming hurricane season.

Future configuration of the Hurricane Forecast System

Based on four years of results from HFIP, we project that the future operational hurricane forecast guidance system would be as described in the table below.

Component	Specifications
Global model ensemble with Hybrid Data Assimilation	20 members at 10-20 km
Multiple moving nests to 3 km within the global model	1) Double nests (9 and 3km), one for each hurricane 2) HWRF 3) Using all available aircraft and satellite data in core and near environment of hurricane
Additional models to make a multi model ensemble (possibly run as a global model with internal nests.	Multi model (at least two – e.g. HWRF, TC-COAMPS)
Statistical Post Processing	LGEM, SHIPS, SPICE, others

1. Introduction

This report describes the Hurricane Forecast Improvement Project (HFIP), its goals, proposed methods for achieving those goals, and recent results from the project with an emphasis on recent advances in the skill of the operational hurricane forecast guidance.

The first part of this report is very similar to previous versions of the annual report since it basically sets the background of the program. This year's version is shortened somewhat from previous years but some of the same material is repeated for reference. For more background information the reader is referred to earlier reports available at: http://www.hfip.org/documents/reports2.php.

2. The Hurricane Forecast Improvement Project

HFIP provides the unifying organizational infrastructure and funding for NOAA and other agencies to coordinate the hurricane research needed to significantly improve guidance for hurricane track, intensity, and storm surge forecasts. HFIP's 5 year (for 2014) and 10-year goals (for 2019) are:

- Reduce average track errors by 20% in 5 years, 50% in 10 years for days 1 through 5.
- Reduce average intensity errors by 20% in 5 years, 50% in 10 years for days 1 through 5.
- Increase the probability of detection (POD)[1] for rapid intensification (RI)[2] to 90% at Day 1 decreasing linearly to 60% at day 5, and decrease the false alarm ratio (FAR) for rapid intensity change to 10% for day 1 increasing linearly to 30% at day 5. The focus on rapid intensity change is the highest-priority forecast challenge identified by the National Hurricane Center (NHC).
- Extend the lead-time for hurricane forecasts out to Day 7 (with accuracy equivalent to that of the Day 5 forecasts when they were introduced in 2003).

It is hypothesized that the HFIP goals could be met with high-resolution (~10-15 km) global atmospheric numerical forecast models run as an ensemble in combination with, and as a background for, regional models at even higher resolution (~1-5 km). In order to support the significant computational demands of such an approach, HFIP has developed a high-performance computational system in Boulder, Colorado. Demonstrating the value of advanced science, new observations, higher-resolution models, and post-processing applications is necessary to justify

---

[1] POD, Probability of Detection, is equal to the total number of correct RI forecasts divided by the total number of forecasts that should have indicated RI: number of correctly forecasted ÷ (correctly forecasted RI+ did not, but should, have forecasted RI). FAR, False Alarm Ratio, is equal to the total number of incorrect forecasts of RI divided by the total number of RI forecasts: forecasted RI that did not occur ÷ (forecasted RI that did occur + forecasted RI that did not occur).

[2] Rapid Intensification (RI) for hurricanes is defined as an increase in wind speed of at least 30 knots in 24 hours. This goal for HFIP also applies to rapid weakening of a decrease of 25 knots in 24 hours.

obtaining the commensurate resources required for robust real-time use in an operational environment.

In FY2013, HFIP funded approximately $11,000,000 of research and development based on recommendations from six strategic teams focused on various components of the hurricane forecast problem. That research and development was performed by 1) various NOAA laboratories and centers: Atlantic Oceanographic and Meteorological Laboratory (AOML), Coast Survey Development Laboratory (CSDL), Environmental Modeling Center (EMC), Earth System Research Laboratory (ESRL), Geophysical Fluid Dynamics Laboratory (GFDL), National Environment Satellite Data and Information Service (NESDIS), and National Hurricane Center (NHC); 2) the National Center for Atmospheric Research (NCAR); 3) the Naval Research Laboratory in Monterey (NRL); 4) the Joint Center for Satellite Data Assimilation (JCSDA); and 5) several universities (awarded through a NOAA Announcement of Opportunity – see Appendix A): University of California, Los Angeles (UCLA), Colorado State University (CSU), University of Colorado (UC), Florida State University (FSU), University of Maryland (UMD), University of Miami (UM), the State University of New York (SUNY), Albany, University of Oklahoma (OU), The Pennsylvania State University (PSU), University of Rhode Island (URI), and University of Wisconsin (UW).

The teams, made up of over 50 members drawn from the hurricane research, development and operational communities, and the team co-leaders are listed in Table 1. Tiger teams, focused on specific, shorter term (1 or 2 years) tasks are listed in Table 2. Full team membership is available at: http://www.hfip.org/teams/.

HFIP dedicated an additional $2,000,000 to operating and maintaining the high performance computer system in Boulder, Colorado.

.

Table 1. Strategic Teams 2013

FY 2013 Teams	**FY 2013 Team Leads**
1. HFIP Model Strategy	Vijay Tallapragada, Stan Benjamin
2. Model Physics	Brad Ferrier, Jian-Wen Bao
3. Data Assimilation/Initialization	John Derber, XuguangWang
4. Ensemble Development	Jeff Whitaker, Jiayi Peng
5. Post Processing and Verification Development Team	Mark DeMaria, David Zelinski, Tim Marchok
6. Societal Impacts	Jennifer Sprague, Rick Knabb

Table 2. Tiger Teams 2013

FY 2013 Teams	**Strategic Team**	**FY2013 Team Leads**
1. Web Page Design	5	Paula McCaslin, Laurie Carson
2. 3 KM Physics Package	2	Joe Cione, Chan Kieu
3. Regional Hybrid DA System/Use of Satellite Data	3	Jeff Whitaker, Xiaolei Zou
4. Stream 1.5 and Demo System Implementation	1	James Franklin, Barb Brown
5. Reconnaissance Data Impact	1	James Franklin, Vijay Tallapragada

HFIP's focus and long-term goal is to improve the numerical model guidance that is provided by the National Centers for Environmental Prediction (NCEP) operations to NHC as part of the hurricane forecast process. To accomplish this goal, the program is structured along three somewhat parallel development paths, known as "streams." Stream 1 is directed toward developments that can be accomplished using operational computing resources (either existing or planned). This stream covers development work planned, budgeted and executed over the near term (mostly one to two years) by EMC in collaboration with others, particularly AOML, with HFIP augmenting support to enable participation by the broader modeling community. Since Stream 1 enhancements are implemented into operational forecast systems, these advances are automatically available to the Hurricane Specialists at NHC in the preparation of official forecast and warning products.

While Stream 1 works within presumed operational computing resource limitations, Stream 2 activities assume that resources will be found to greatly increase available computer power in operations above that planned for the next five years. The purpose of Stream 2 is to demonstrate that the application of advanced science, technology, and increased computing will lead to the desired increase in accuracy and other aspects of forecast performance. Because the level of computing necessary to perform such a demonstration is large, HFIP is developing its own computing system at NOAA/ESRL in Boulder, Colorado.

A major component of Stream 2 is an Experimental Forecast System (EFS) that HFIP runs each hurricane season. The purpose of the EFS (also known as the Demonstration Project) is to evaluate the strengths and weaknesses of promising new approaches that are testable only with enhanced computing capabilities. The progress of Stream 2 work is evaluated each off-season to identify techniques that appear particularly promising to operational forecasters and/or modelers. These potential advances can be blended into the operational implementation plans through subsequent Stream 1 activities, or developed further outside of operations within Stream 2.

Stream 2 models represent cutting-edge approaches that have little or no track record; consequently NHC forecasters do not use these models to prepare their operational forecasts or warnings.

HFIP was originally structured around this two-stream approach. However, it quickly became apparent that some Stream 2 research models were producing forecast guidance that was potentially useful to forecasters. Because these models could not be implemented at NCEP due to insufficient operational computing resources, a third activity, known as Stream 1.5, was initiated to expedite the testing and availability of promising new models to forecasters. Stream 1.5 is an approach that accelerates the transfer of successful research from Stream 2 into real-time forecasting, by following a path that temporarily bypasses the budgetary and technical bottlenecks associated with traditional operational implementations.

The Stream 1.5 process for the each hurricane season involves extensive evaluation, by the Tropical Cyclone Modeling Team (TCMT) at NCAR, of the previous season's most promising Stream 2 models or techniques. This testing involves rerunning the models or techniques over storms from the demonstration period (August 1 to October 31) for the three previous seasons involving several hundred cases. If operational computing resources are not available for immediate implementation, those models or techniques that meet certain pre-defined standards for improvement over existing techniques, can be run on HFIP computing resources and the output guidance provided to NHC forecasters in real-time during the upcoming hurricane season as part of the EFS. This process moves forward the real-time availability of advances to forecasters one or more years. It also serves as a proof of concept for both the developmental work (Stream 2) and augmented computational capabilities.

3. The HFIP Model Systems

HFIP believes that the best approach to improving hurricane track forecasts, particularly beyond four days, involves the use of high-resolution global models, with at least some run as an ensemble. However, global model ensembles are likely to be limited by computing capability for at least the next five years to a resolution no finer than about 15-20 km, which is inadequate to resolve the inner core of a hurricane. It is generally assumed that the inner core must be resolved to see consistently accurate hurricane intensity forecasts (e.g., NOAA SAB 2006). Maximizing improvements in hurricane intensity forecasts will therefore likely require high-resolution regional models, perhaps also run as an ensemble. Below we outline the modeling systems currently in use by HFIP.

a. The Global Models

Global models provide the foundation for all of HFIP's modeling effort. They provide hurricane forecasts of their own, and are top-tier performers for hurricane track. They also provide background data and/or boundary conditions for regional and statistical models, and can be used to construct single-model ensembles, or be members of multi-model ensembles. The global models used in 2013 by HFIP are listed in table 3 along with their characteristics.

Table 3. Specifications of the HFIP Global Models

Models	Horizontal resolution	Vertical levels	Cumulus Parameterization	Microphysics	Planetary Boundary Layer (PBL)	Land Surface Model (LSM)	Radiation	Initialization
FIM 5-member ensemble	27 km	64	From 2011 GFS - Simplified Arakawa Schubert	Zhao-Carr	GFS Non-Local PBL	Noah LSM	Rapid Radiative TransferModel (RRTM)	ESRL EnKF
FIM – 2014 numerics w/ 1/8 deg output	15km	64	From 2011 GFS – Simplified Arakawa Schubert	Zhao-Carr	GFS Non-local PBL	Noah LSM	RRTM	GFS GSI operational hybrid-ensemble variational
GFS/EnKF	27 km	64	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	Noah LSM	GFDL scheme	ESRL EnKF
GFS/GSI	27 km	64	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	Noah LSM	GFDL	GSI
NOGAPS	41 km	42	Emanuel	N/A	NOGAPS	NOGAPS	Harshvardhan/ Fu-Liou	NRL Atmospheric Variational Data Assimilation System-accelerated representer (NAVDAS-AR)

b. The Regional Models

Specifics of the regional models used by HFIP in 2013 are shown in Table 4.

c. Initialization and Data Assimilation Systems

A number of approaches are used to create the initial state for the global and regional models in the HFIP EFS:

1. Global Forecast System (GFS): The initial state created for the current operational global model, GFS, is interpolated to the grids used by HFIP global models. The GFS in 2012 used the new Hybrid Ensemble-Variational Data Assimilation System (HEVDAS—see below) that is a combination of the Grid-point Statistical Interpolation (GSI) system formerly used and an ensemble based system to define the background error matrix. The GSI initialization system that has been run operationally since 2006 is a three-dimensional variational approach, 3D-VAR (DTC 2012; Wu et al. 2002; Parrish and Derber 1992; Cohn and Parrish 1991).

Table 4. Specifications for the HFIP Regional Models

Models	**Domains / Horizontal Resolution (km)**	**Vertical Levels core**	**Cumulus Parameterization**	**Microphysics**	**PBL**	**Land Surface**	**Radiation**	**Initial and Boundary Conditions**	**Initialization**
HWRF (OPS)	3 27/9/3	42 NMM	Simplified Arakawa Schubert (27/9 km only)	Ferrier	GFS Non-Local PBL	GFDL Slab Model	GFDL Scheme	GDAS and GFS	One-way hybrid GSI-EnKF with vortex initialization
GFDL (OPS)	3 55/18/9	42 GFDL	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	Schwarzkopf-Fels (longwave) / Lacis-Hansen (shortwave)	GFS	GFDL Synthetic Bogus Vortex
GFDL (Ens)	3 55/18/9	42 GFDL	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	Schwarzkopf-Fels (longwave) / Lacis-Hansen (shortwave)	GFS.	GFDL Synthetic Bogus Vortex
HWRF-HRD/EMC Basin Scale	3 27/9/3	61 NMM	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	GFDL Scheme	GFS	Vortex initialization
HWRF-HRD (HEDAS)	2 9/3	42 NMM	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	GFDL Scheme	GFS	EnKF with aircraft and satellite (AMVs, AIRS and GPS-RO retrievals) data
AHW (NCAR)	3 36/12/4	36 ARW	Tiedtke (36/12 km only)	WSM6	YSU	NOAH LSM	RRTMG (LW+SW)	GFS (BC only)	EnKF method in a 6-hour cycling mode
COAMPS-TC©	3 45/15/5 (15/5 km following the storm)	40 COAMPS	Kain Fritsch on 45 and 15 km meshes	Explicit microphysics (5 class bulk scheme)	Navy 1.5 Order Closure	Slab with the NOAH LSM as an option	Fu-Liou	GFS in WATL and EPAC	3D-Var data assimilation with synthetic observations
Wisconsin NMS	2 45/4,1	42 UW-NMS	Modified Emanuel	Tripoli-Flatau Bulk microphysics (1 liquid, 2 ice categories)	1.5 Order Closure	NOAH LSM	RRTMG SW and LW	GFS	Bogus vortex with 12-hour dynamic initialization
Penn State ARW	3 27/9/3	42 ARW	Grell-Devenyi ensemble scheme (27 km only)	WSM 6-class graupel scheme	YSU	5-layer thermal diffusion scheme	RRTM (longwave) / Dudhia (shortwave)	GFS	Cycling EnKF with all Recon data

2. HWRF: The operational HWRF uses an advanced vortex initialization and assimilation cycle consisting of four major steps: 1) interpolation of the global analysis fields from the Global Data Assimilation System (GDAS) onto the operational HWRF model grid; 2) removal of the GFS vortex from the global analysis; 3) addition of the HWRF vortex modified from the previous cycle's six-hour forecast based on observed location and strength (or use of a corrected GDAS or bogus vortex for a cold start); and 4) addition of observation data outside of the hurricane area using one-way hybrid GSI and Ensemble Kalman Filter. The Data Assimilation (DA) system is capable of ingesting inner core data to optimize the vortex initialization.

3. NRL Atmospheric Variational Data Assimilation System (NAVDAS): This is the system used to provide the initial conditions to NOGAPS. Previously a 3D-VAR system, it was upgraded in September 2009 to NAVDAS-AR, a four-dimensional variational (4D-VAR) approach (NRL 2001, Daley and Barker 2001). The 3D-VAR version of NAVDAS is used to initialize COAMPS-TC.

4. Ensemble Kalman Filter (EnKF): This is an advanced assimilation approach, somewhat like 4D-VAR, that uses an ensemble to create background error statistics for a Kalman filter (Tippett et al 2003, Keppenne 2000, Evensen 1994, Houtekamer et al 1998). This approach has shown considerable promise (Hamill et al 2011).

5. Hurricane Ensemble Data Assimilation System (HEDAS): HEDAS is an EnKF system applied to the HWRF and was developed at the AOML (Aksoy et al 2012).

6. Hybrid Ensemble-Variational Data Assimilation System (HEVDAS): This system combines aspects of the EnKF and 3D- or 4D-Var, such as using the ensemble of forecasts to estimate the covariances at the start of the variational component of the DA system. This technology was developed at EMC, ESRL and AOML/Hurricane Research Division (HRD) and was used in operations for the 2012 season.

7. One Way Hybrid: This is a version of the HEVDAS system that is currently being used by the Operational HWRF regional model. Instead of using an ensemble from the HWRF to create the ensemble part of the background error, it uses ensemble members created by the global GEFS.

8. Vortex Initialization: The initial vortex for some of the regional models is produced by a vortex initialization procedure. First, the vortex circulation is filtered from the first guess fields interpolated from global model; then a new vortex modified by the observed intensity is inserted back in the filtered environment. The new vortex is the model balanced vortex cycled from previous six-hour forecast, from a parent global model, or defined based on a synthetic vortex profile. On the first initialization for a particular storm, the size and intensity of the vortex are modified based on real-time observations. In the HWRF system, the tropical cyclone vortex is cycled from the previous six hour forecast and the vortex is relocated based on the observed position. The one-way hybrid GSI-EnKF DA system assimilates the modified vortex and ambient fields to generate initial conditions for the HWRF system. Vortex relocation is also utilized by the current operational GFS and Global Ensemble Forecast System (GEFS) in NCEP.

d. The HWRF Community Code Repository

During 2009-2013, both EMC and the Developmental Testbed Center (DTC) worked to update the operational version of HWRF from version 2.0 to the current community version of HWRF, version 3.5a (Bernardet et al, 2013; Tallapragada et al, 2013). This makes the operational model completely compatible with the codes in the community repositories, allows researchers access to the operational codes, and makes improvements in HWRF developed by the research community easily transferable into operations. This was one of the initial goals of the WRF program.

4. Meeting the HFIP Goals

a. The HFIP Baseline

To measure progress toward meeting the HFIP goals outlined in the introduction, a baseline level of accuracy was established to represent the state of the science at the beginning of the program. Results from HFIP model guidance could then be compared with the baseline to assess progress. HFIP accepted a set of baseline track and intensity errors developed by NHC, in which the baseline was the consensus (average) from an ensemble of top-performing operational models, evaluated over the period 2006-2008. For track, the ensemble members were the operational aids GFSI, GFDI, UKMI, NGPI, HWFI, GFNI, and EMXI, while for intensity the members were GHMI, HWFI, DSHP, and LGEM (Cangialosi and Franklin 2011). Figure 1 shows the mean errors of the consensus over the period 2006-2008 for the Atlantic basin. A separate set of baseline errors (not shown) was computed for the eastern North Pacific basin.

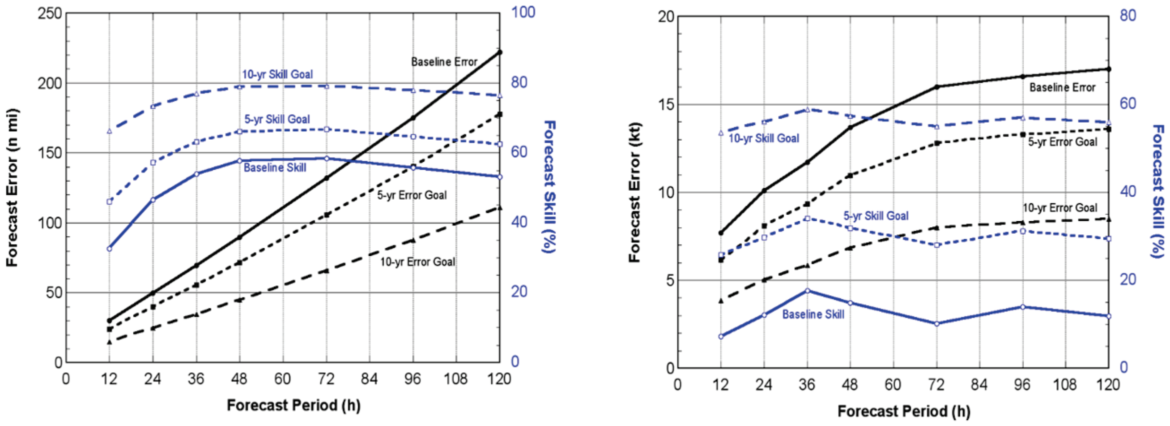


Figure 1. HFIP Track (left panel) and Intensity (right panel) Error Baseline and Goals. The baseline errors (black lines) were determined from an average of the top-flight operational models during the period 2006-2008. The HFIP expressed goals (dashed lines) are to reduce this error by 20% in 5 years and 50% within 10 years. Comparisons of forecasts over non-homogenous samples, however, are best done in terms of skill. To obtain the 5-year and 10-year HFIP goal in terms of skill (blue lines—baseline skill in solid, HFIP goals dashed), the goals are expressed as the percentage improvement over the CLIPER5 errors (track) and Decay-SHIFOR5 (intensity) of the baseline sample (see text).

The baseline errors in Figure 1 are also compared to the errors for the same cases for the climatology and persistence model (CLIPER5) for track and the Decay Statistical Hurricane Intensity Forecast (Decay-SHIFOR5) model for intensity (NHC 2009). Errors from these two models are large when a storm behaves in an unusual or rapidly changing way and, therefore, are useful in assessing the inherent difficulty in a set of forecasts. When a track or intensity model error is normalized by the CLIPER5 or Decay-SHIFOR5 error, the normalization yields a measure of the model's skill.

Because a sample of cases from, say, the 2013 season might have a different inherent level of difficulty from the baseline sample of 2006-2008 (for example, because it had an unusually high or low number of rapidly intensifying storms), evaluating the progress of the HFIP models in terms of forecast skill provides a longer-term perspective. Figure 1 shows the baseline errors and the 5- and 10-year goals as skill, represented in blue and labeled on the right side of the graph. Skill in the figure is the percentage improvement over the Decay-SHIFOR5 and CLIPER5 forecasts for the same cases. Note the skill baseline and goals for intensity at all lead times is roughly constant with the baseline representing a 10% improvement over Decay-SHIFOR5 and the 5- and 10-year goals, representing 30% and 55% improvements, respectively.

It is important to note that the HFIP performance baselines were determined from a class of operational aids known as "early" models. Early models are those that are available to forecasters early enough to meet forecast deadlines for the synoptic cycle. Nearly all the dynamical models currently in use at tropical cyclone forecast centers, however, (such as the GFS or the GFDL model, referred to as GFDL, for short) are considered "late" models because their results arrive too late to be used in the forecast for the current synoptic cycle. For example, the 1200 Coordinated Universal Time (UTC) GFDL run does not become available to forecasters until around 1600 UTC, whereas the NHC official forecast based on the 1200 UTC initialization must be issued by 1500 UTC, one hour before the GFDL forecast can be viewed. It's actually the older, 0600 UTC run of the GFDL that would be used as input for the 1500 UTC official NHC forecast, through a procedure developed to adjust the 0600 UTC model run to match the actual storm location and intensity at 1200 UTC. This adjustment, or interpolation, procedure creates the 1200 UTC "early" aid GFDI that can be used for the 1500 UTC NHC forecast. Model results so adjusted are denoted with an "I" (e.g., GFDI). The distinction between early and late models is important to assessing model performance, since late models have an advantage of more recent observations/analysis than their early counterparts.

b. Meeting the Track Goals

Accurate forecasts beyond a few days require a global domain because influences on a forecast for a particular location come from weather systems at increasing distance from the local region over time. One of the first efforts in HFIP was to improve the existing operational global models. Early in the program, it was shown that using a more advanced data assimilation scheme than the one employed operationally at that time improved forecasts, particularly in the tropics. A version of this advanced data assimilation went operational in the GFS model in May, 2012. Results from that model are presented in this report.

c. Reaching the Intensity Goals

HFIP expects that its intensity goals will be achieved through the use of regional models with a horizontal resolution near the core finer than about 3 km. In addition, early results suggest that output from individual HFIP models can be used in statistical models such as the Statistical Hurricane Intensity Prediction System, SHIPS, (DeMaria and Kaplan, 1994, NHC 2009) or Logistics Growth Equation Model, LGEM, (DeMaria, 2009, NHC 2009) to further increase the skill of the intensity forecasts.

5. HFIP Stream 1.5

The HFIP and the NHC agreed in 2009 to establish a pathway to operations known as "Stream 1.5." Stream 1.5 covers improved models and/or techniques that the NHC, based on prior assessments, wants to access in real-time during a particular hurricane season, but which cannot be made available to NHC by the operational modeling centers in conventional production mode. HFIP's Stream 1.5 supports activities that intend to bypass operational limitations by using non-operational resources to move forward the delivery of guidance to NHC by one or more hurricane seasons. Stream 1.5 projects are run as part of HFIP's annual summertime "Demo Project".

Nine modeling groups provided retrospective tests of their models for consideration by NHC for inclusion as a stream 1.5 model for the 2013 season. Those groups and their models are listed in table 5.

All evaluations were based on homogeneous samples and applied the appropriate method for assessing statistical significance. A detailed description of the methologdy used for the evaluation, reports for each participating modeling group and all the verification plots generated during the evaluation are available on the TCMT 2013 Retrospective HFIP Testing website (http://www.ral.ucar.edu/projects/hfip/d2013), as well as information on the participating models and cases included in the evaluation.

The groups selected for Stream 1.5 in 2013 and the form of the output from their models provided to NHC are listed in Table 6.

Note that most models were admitted into Stream 1.5 based on the models' performance forecasting either track or intensity, but generally not both. For example, forecasters were instructed to consult the ESRL FIM model, interpolated ahead 6 hours, (FM9I—FIM in the tables below) track forecasts but not the FM9I intensity forecasts. Two HFIP Stream 1.5 consensus aids were constructed: the track consensus TV15 comprised the operational models GFSI, EGRI, GHMI, HWFI, GFNI, EMXI and the Stream 1.5 models APSI, GHMI, and FM9I, while the intensity consensus IV15 comprised the operational models Decay-SHIP (DSHP), LGEM, GHMI, HWFI and the Stream 1.5 models COTI, APSI, and UWNI.

Table 5. 2013 HFIP Retrospective Evaluation Modeling Groups

Organization	Model	Type	Configurations
NCAR/MMM and SUNY-Albany	AHW	Regional-dynamic-deterministic	1
NRL	COAMPS-TC	Regional-dynamic-deterministic	1
PSU	ARW	Regional-dynamic-deterministic	3
UW-Madison	UW-NMS 4-km	Regional-dynamic-deterministic/ensemble	3
EMC	HWRF	Regional-dynamic-ensemble	1
GFDL	GFDL hurricane ensemble	Regional-dynamic-ensemble	1
ESRL	FIM	Global-dynamic-deterministic	1
FSU	Multi-model super ensemble	Weighted-consensus	1
NESDIS/STAR and CIRA	SPICE	Statistical-dynamical-consensus	1

Table 6. 2013 HFIP Stream 1.5 Real-time Runs Model Guidance Provided to NHC

Organization	Model	Track	Track Consensus	Intensity	Intensity Consensus
NRL	COAMPS-TC				●
PSU	ARW		●		●
UW-Madison	UW-NMS				●
EMC	HWRF	●		●	
GFDL	GFDL hurricane model	●		●	
ESRL	FIM	●			
NESDIS/STAR and CIRA	SPICE			●	

a. Stream 1.5 results

The results presented here reflect the Stream 1.5 runs that were successfully transmitted to NHC in real time during 2013. The Stream 1.5 models arriving at NHC underwent standard operational processing to convert "late" dynamical guidance into "early" interpolated guidance that could be used by the forecasters. Figure 2 presents a homogeneous verification of the Stream 1.5 track models that met availability standards (and regardless of whether they were intended for use explicitly or in a consensus), along with selected operational models. The figure shows that in 2013 the FM9I was competitive with the top-tier operational models, although the HWFI was not. The TV15, the composite that included the ARW of the stream 1.5 models was the best performer throughout the forecast lead times (though the TVCA -- the consensus without the stream 1.5 models – was almost the same) and outperformed the European Center Model (EMXI) in this sample. The operational GFS (GFSI) was similar to EMXI except for the 36- and 48- hour lead times.

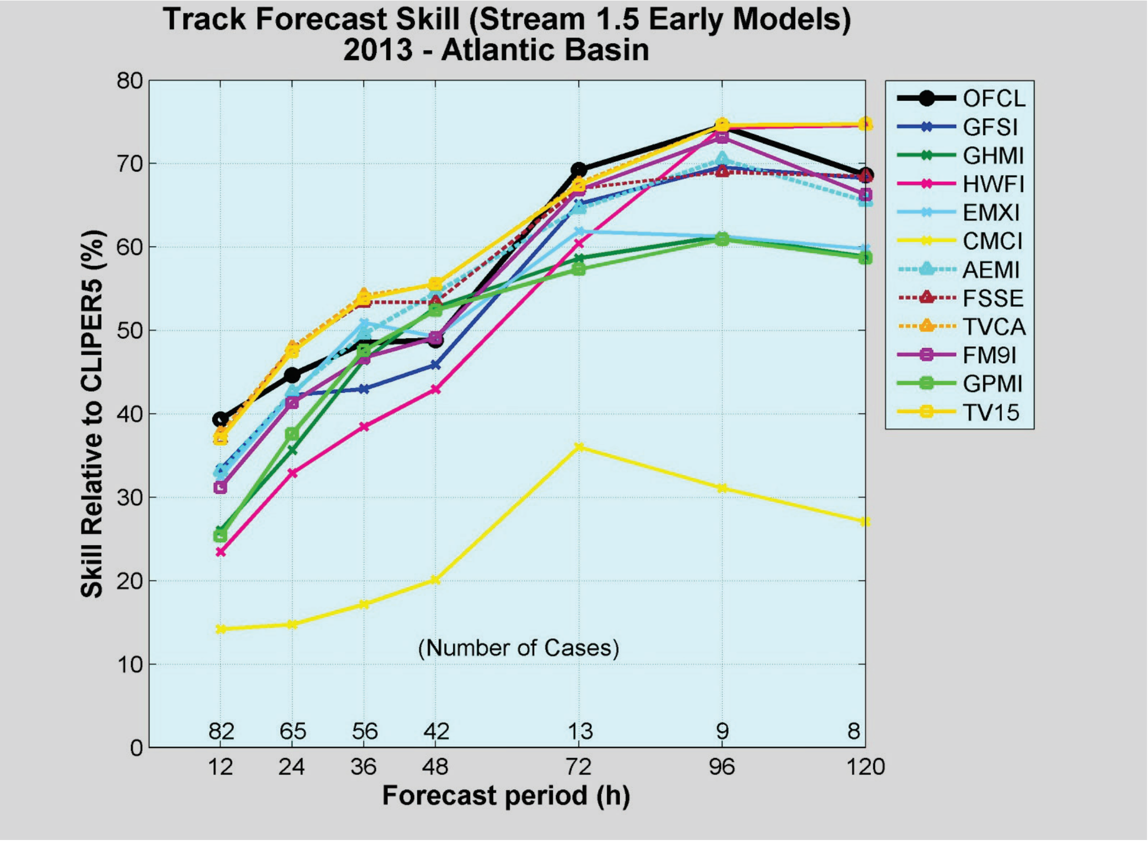


Figure 2. Stream 1.5 Track Skills, 2013 Atlantic

Track skill (relative to CLIPER5) of Stream 1.5 and other models for the Atlantic basin 2013 (homogeneous comparison)

Stream 1.5 intensity results are shown as absolute error in Figure 3, for a sample that, because of limited availability, excludes the PSU Doppler runs. In addition we show this chart as absolute error rather than skill (relative to SHIFOR5, labeled OCD5) because OCD5 was an especially poor performer in 2013 (dotted line in the figure). Note also that there are very few cases at 120 hours—very few of the storms in the Atlantic in 2013 lasted more than three days.

In 2012 the SPC3 statistical model was the best performing Stream 1.5 intensity model. In 2013 it was among the worst. None of the statistical models performed well in 2013. The best model out to 48 hours was the operational HWRF (HWFI) and the FSU super ensemble (FSSE). IV15/IVCI composites performed well as did the Wisconsin model (UW4I) in the 24-48 hour forecast lead times.

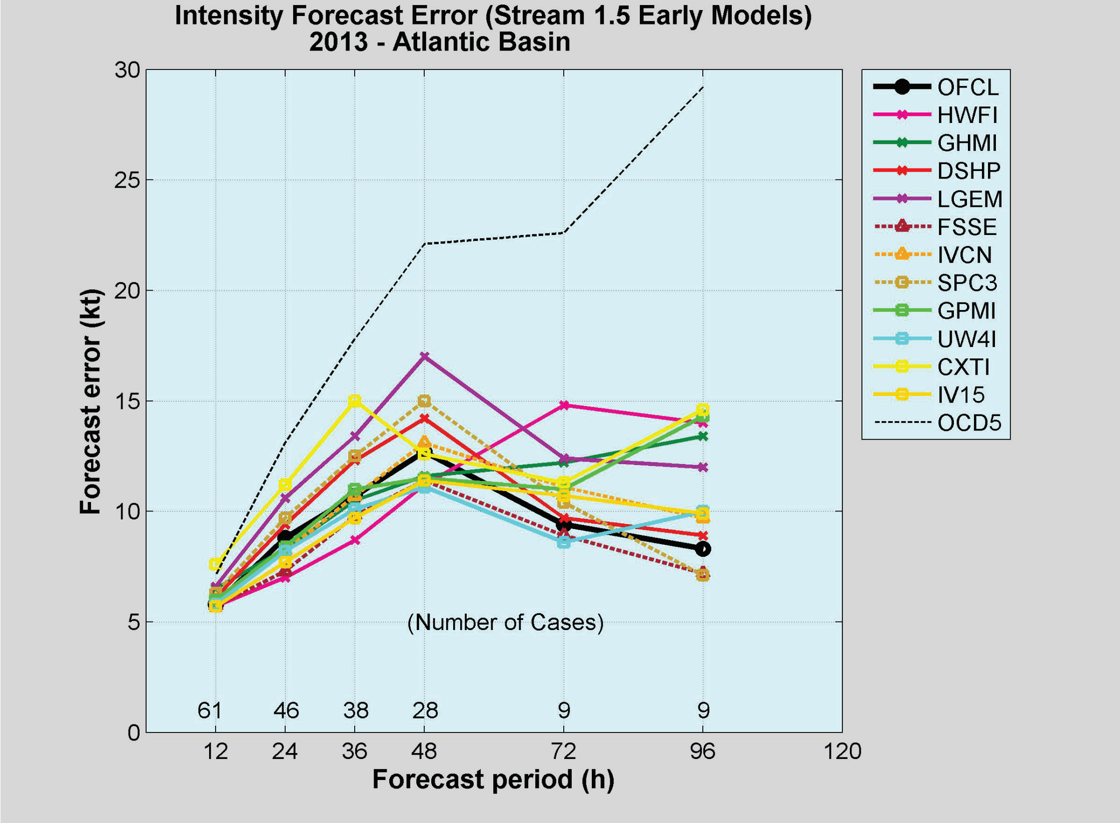


Figure 3. Stream 1.5 Intensity Errors 2013 Atlantic
Intensity errors of Stream 1.5 and other models for Atlantic basin 2013 (homogeneous comparison)

The shining lights in the stream 1.5 models this year were the GFS and FIM for track and the HWRF for intensity. The HWRF and GFS operational systems are discussed in the next section. The GFDL ensemble is discussed in section 7.

6. Operational Hurricane Guidance Improvements

The HFIP goals described in section 4 are only met when the model guidance provided to NHC by NCEP reaches those goals. Since 2013 represents the fourth year of the program we would expect to see considerable progress toward meeting the five year goals in the operational models and not just in the experimental models such as the stream 1.5 models described in the previous section. In this section the emphasis will be on improvements in the hurricane forecasts from the models operational in 2013. This includes the global GFS model, and the HWRF operational regional models.

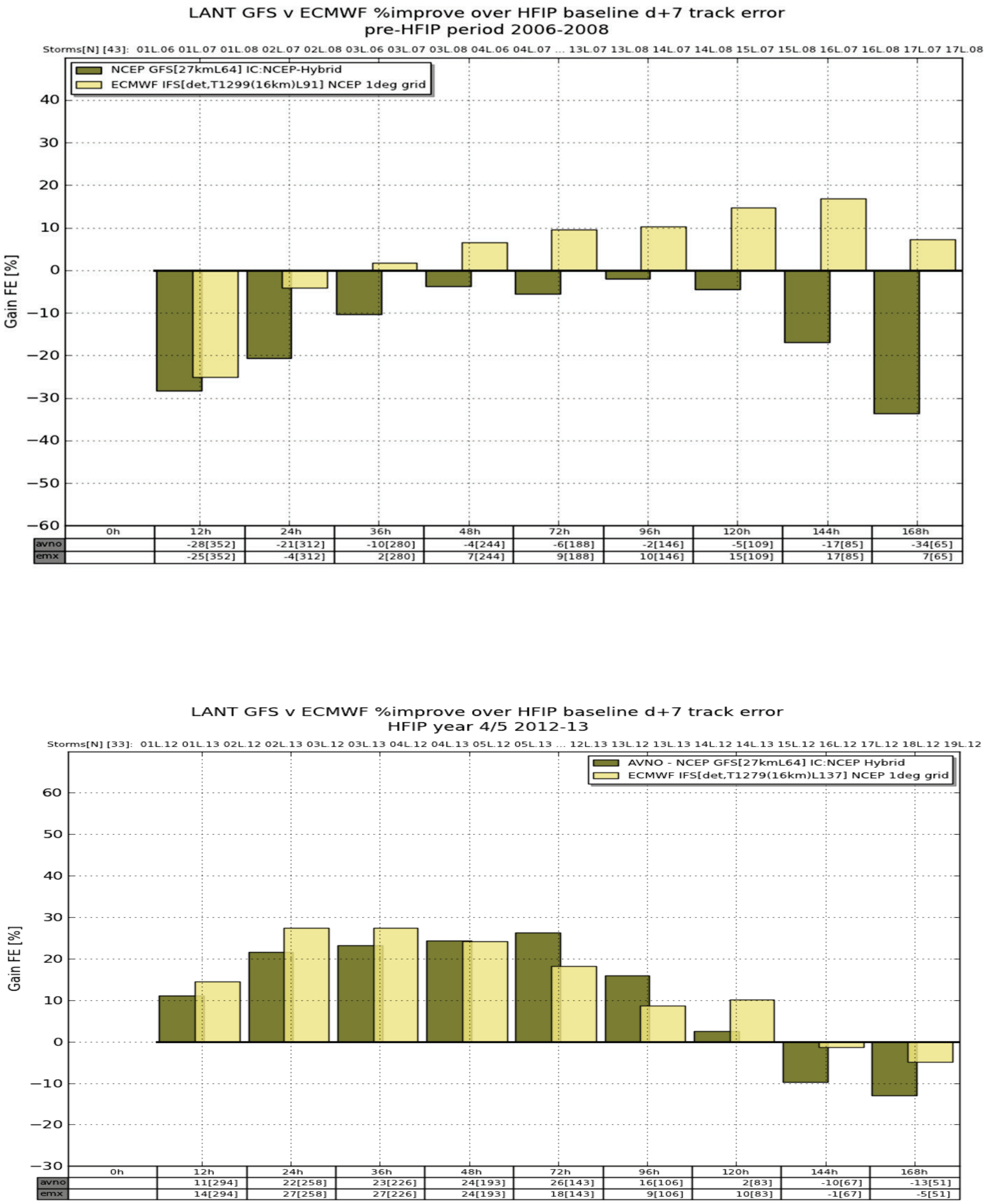
a. Global Model (GFS)—Operational

In May of 2012 the GSI data assimilation system in the GFS was replaced by the hybrid data assimilation system. The hybrid system uses an ensemble to generate a flow dependent background error covariance matrix which is then used in the GSI for the analysis. The reader may note that previous annual reports, starting with the first one in 2010, described the impact of changing the data assimilation system in the global models, particularly the GFS from the 3D-VAR GSI to the ensemble-based EnKF system. The hybrid system is basically a combination of the EnKF and the GSI and has been shown to provide somewhat better results than EnKF alone. The implementation of the hybrid system is an example of transferring HFIP results into operations.

The GFS has undergone other improvements including improved physics and increased resolution. In 2013 the deterministic operational GFS was run at T574 and the GFS ensemble (GEFS) at T265. Figure 4 illustrates the progress that has been made with the GFS since 2006 in forecasting track. The upper panel shows the performance of the GFS and the ECMWF operational models over three years starting in 2006. Data shown on both charts in Figure 4 are plotted relative the HFIP baseline for track. Note that, compared to the HFIP baseline, both the ECMWF and the GFS performed poorly early in the forecast during that three-year period. The ECMWF was better than the HFIP baseline at later forecast lead times. On the other hand, by 120 hours, the GFS forecasts were worse than the forecasts at the earlier lead times.

The bottom panel shows the global model performance in 2012-2013. Again, results from both ECMWF and GFS are shown for the Atlantic. Note that both models have improved dramatically in the first 4 days of forecast lead time. ECMWF beats the GFS in the first 36 hours and the GFS beats ECMWF from 48 to 96 hours. Up to 3 days both models exceed the 5-year, 20% HFIP goal and the GFS does so out to 4 days. However, the charts indicate essentially no improvement in track forecasts beyond 4 days since 2006.

This sudden decrease in track skill improvement is dramatic and apparently is shared by both models shown. It flies in the face of the assumption, often made, that just because the forecasts get better say at day 4 they should also get better at later forecast lead times. This apparently isn’t the case.



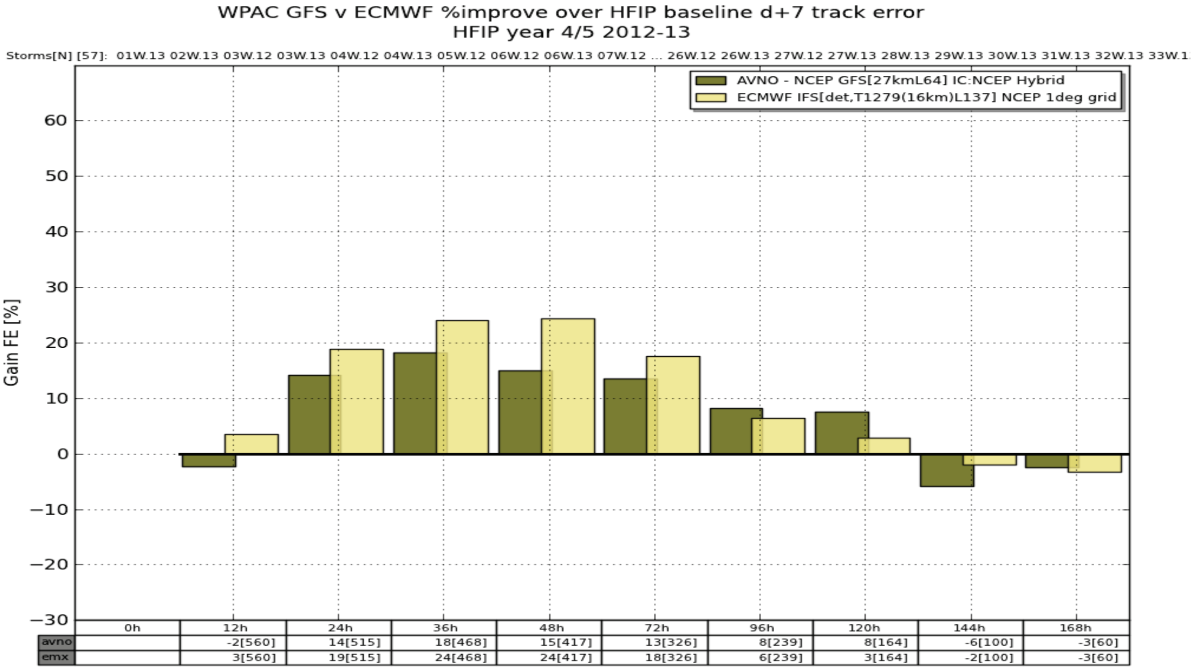
	0h	12h	24h	36h	48h	72h	96h	120h	144h	168h
avno		-28[352]	-21[312]	-10[280]	-4[244]	-6[188]	-2[146]	-5[109]	-17[85]	-34[65]
emx		-25[352]	-4[312]	2[280]	7[244]	9[188]	10[146]	15[109]	17[85]	7[65]

	0h	12h	24h	36h	48h	72h	96h	120h	144h	168h
avno		11[294]	22[258]	23[226]	24[193]	26[143]	16[106]	2[83]	-10[67]	-13[51]
emx		14[294]	27[258]	27[226]	24[193]	18[143]	9[106]	10[83]	-1[67]	-5[51]

Figure 4. GFS and ECMWF Track Errors Relative to HFIP Baseline, Atlantic.

Track Skill (error relative to HFIP baseline) for the NCEP (GFS) and ECMWF Operational models. Upper panel shows error for all storms in the Atlantic for 2006-2008 and number of cases is shown in parentheses at the bottom of the figure. The lower panel is for 2012-2013 and otherwise is the same as the upper panel.

The reasons for the sudden decline in track skill beyond 4 days are an interesting and important problem. There was some speculation that the reason for the sudden drop-off in the Atlantic at day 4 was related to higher observation error propagating into the Atlantic from say the Pacific. But if that were true the performance beyond 4 days should be different in different basins where upstream error properties should be different. Figure 5 shows the same plot as as the bottom panel in Figure 4 but for the West Pacific (WPAC). Note that the drop-off at day 4 is just as apparent in that basin. HFIP will put some resources towards answering this question in 2014.



	0h	12h	24h	36h	48h	72h	96h	120h	144h	168h
avno		-2[560]	14[515]	18[468]	15[417]	13[326]	8[239]	8[164]	-6[100]	-3[60]
emx		3[560]	19[515]	24[468]	24[417]	18[326]	6[239]	3[164]	-2[100]	-3[60]

Figure 5. GFS and EMCWF Track Errors Relative to HFIP Baseline (2012-2013), WPAC

Note also that while Figure 5 is similar to the lower panel on Figure 4, in the WPAC, the performance of the GFS is not generally superior to the ECMWF.

b. Global Model (GFS)—HFIP Experimental

The operational GFS deterministic model was run at a resolution of T547; and the ensemble (GEFS) was run at a resolution of T256. Experimental versions were run on the HFIP high performance computer using a semi-Lagrangian differencing scheme and at higher resolutions: T1149 for the deterministic model; and T547 for the ensemble. Comparisons of the operational and experimental systems performances are shown on Figure 6.

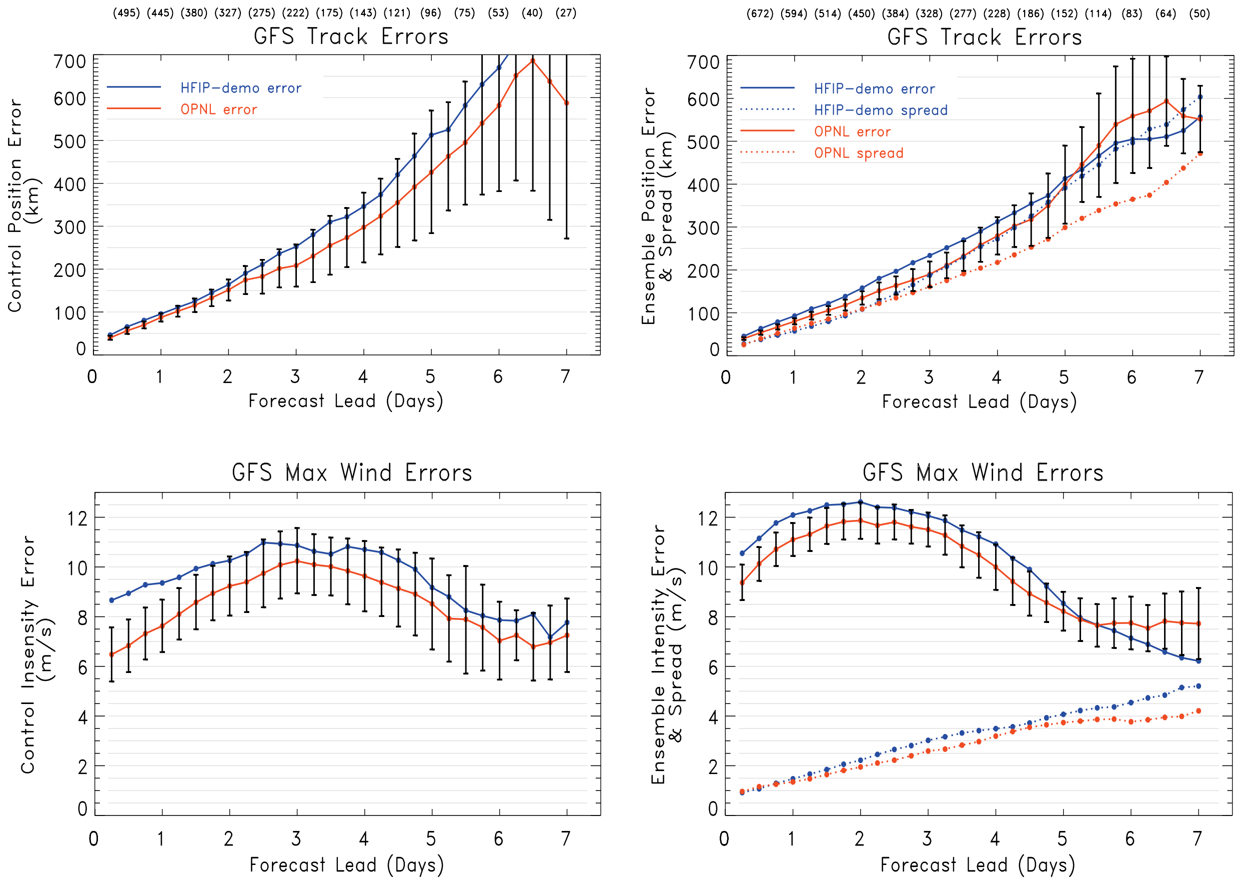


Figure 6. Performance Comparison Operational and Experimental GFS Systems

Track and intensity error for the HFIP experimental GFS T1148 (blue) and operational GFS T574 (red) deterministic model (left panels) and the HFIP experimental GFS T574 (blue) and operational GEFS T256 (red) ensembles (right panels) for the Atlantic in 2013. Track errors above and maximum wind (intensity) errors below. Both the HFIP experimental deterministic and ensemble systems use a semi-Lagrangian differencing. The dashed lines on the right hand panels show the ensemble spread.

Note that the operational versions of the GFS systems outperformed their experimental counterparts in both track and intensity forecasting. The primary reasons for this disappointing performance can most likely be attributed to not adjusting the physics and data assimilation packages to accommodate the revised resolutions and differencing scheme:

1. Increasing resolution does not always improve forecast skill. Rather, careful matching and testing of the physics and DA system to the higher resolution must be performed before the benefits of the higher resolution can be seen. HFIP has noted this in the past for the regional models where a simple test of several regional models at different resolutions did not reveal an advantage of taking the regional model grid to 3 km. However, the recent introduction of a high resolution inner core (3km) in the HWRF along with additional improvements in the overall physics package led to significant improvements over the 9 km version of HWRF.

2. This was the first real-time test of the semi-Lagrangian difference scheme in the GFS and, like the resolution, the physics package may also need to be tuned to the differencing scheme.

The right-hand panels in Figure 6 show the spread of the ensemble errors as well as the mean of the errors. The spread at any forecast lead time is the mean of the distances each of the ensemble members are from the mean location of the forecast. It is encouraging that the spread associated with the ensemble track errors is close to the mean errors. As a rule of thumb, the spread of the ensemble members should approximately equal the error of the ensemble mean otherwise it is said that the ensemble is under dispersed. The track error spread (blue dashed line) is much closer to the mean (blue solid line) for the experimental ensemble than that for the operational ensemble (red lines). That improvement may be the result of using a stochastic physics perturbation process in generating the experimental ensemble. There is, however, no corresponding improvement in the spread of the intensity forecast errors.

One product derived from the experimental HFIP GFS ensemble is the prediction of winds greater than 34 knots. The predictions are in terms of probabilities that tropical storm force winds will be exceeded at a location within a particular time frame. Figure 7 shows the probability forecast for winds 34 knots or greater over a 168-hour period starting 12Z September 11, 2013. The storm in the eastern Atlantic is Humberto which was a hurricane at that time.

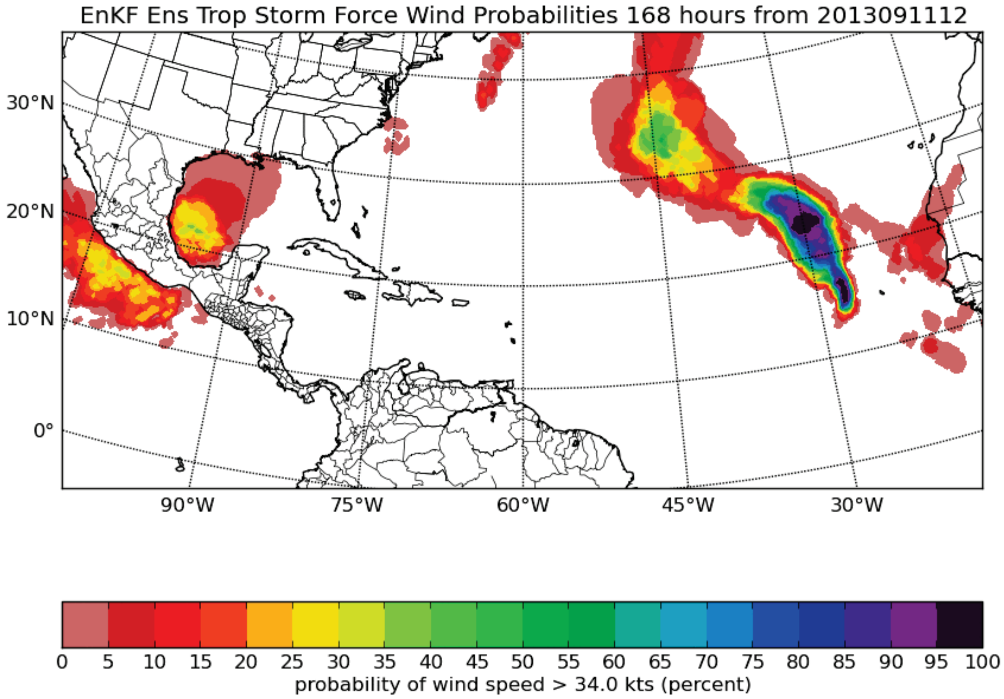


Figure 7. Tropical Storm Force Wind Probabilities within 168 Hours

An example of a probability of winds greater than 34 knots at the various points during a 7 day period starting at 12 Z September 11 2013. A 100% probability implies that all 20 members of the ensemble forecast a 34 knot wind at the point at some time during the 7 days of the forecast. Forecast is from Sept 11 at 12Z 2013. Humberto is the storm in the east Atlantic that formed on Sept 8, 2013, Ingrid in the Gulf of Mexico that formed on September 12, 2013 and Manual in the East Pacific that formed on September 13, 2013.

Humberto later weakened rapidly as indicated in the figure and then regained tropical storm force winds as it turned northeast. The storm in the Gulf formed on the 12th and became hurricane Ingrid on the 14th. The storm in the EPAC is Manuel which formed on the 13th.

Figure 8 shows the verification of probability calculated from the experimental HFIP GFS ensemble of winds greater than 34 knots at a point during the time period indicated on each panel.

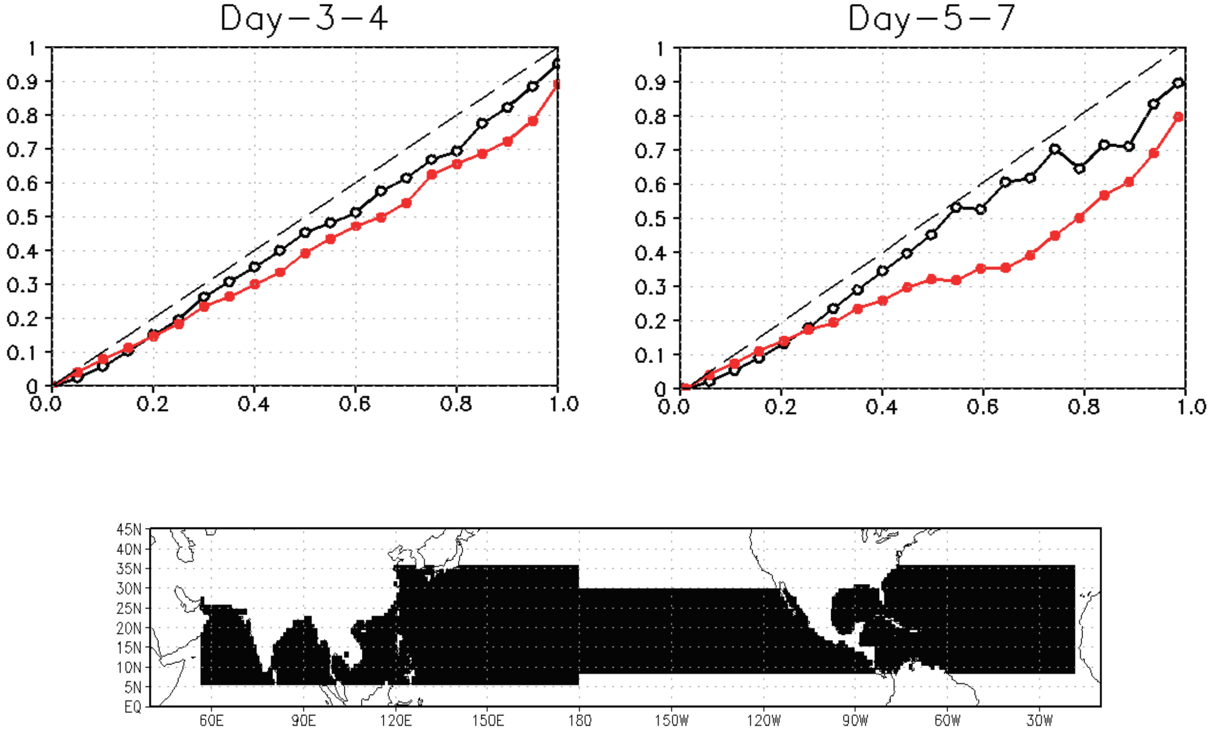


Figure 8. Verification of Tropical Storm Force Wind Probabilities
Verification statistics for probability of tropical storm force winds (greater than 34 knots) at points shown in the black area in the lower panel. This is from the HFIP GFS T574 ensemble and 100% or 1 in the figures would imply that all ensemble members forecast a 34 knot wind at that point during day 3-4 (left panel) and days 5-7 (right panel), black and the operational GEFS (T256), red. Reliability Scores are aggregated over the domain shown below. Observed probabilities are shown on the y-axis and forecast probabilities on the x-axis. A perfect score would be along the dashed line. Forecast probabilities are over forecast below the line and under forecast above.

Another product derived from the experimental HFIP GFS ensemble is a map showing the different tracks forecast by ensemble members. Figure 9 shows the 7-day tracks forecast for Humberto by the 20 ensemble members and the track of the ensemble mean. Ellipses, also shown, are drawn, centered on the mean and enclosing 80-percent of the ensemble members' forecasts associated with each lead time.

Both of these products can be generated from the operational GFS system.

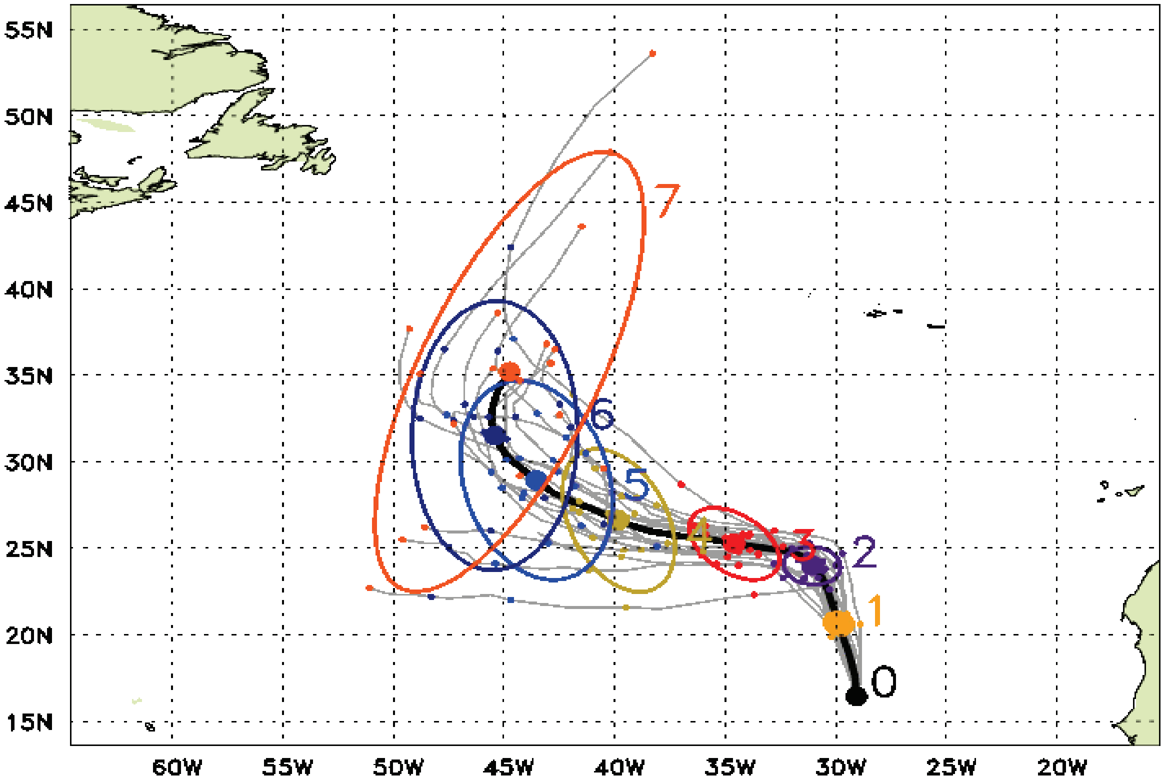


Figure 9. Ensemble Member Tracks.
A product created from the HFIP GFS ensemble that shows all the tracks from the forecast of September 11, 2013 out to 7 days. There are twenty ensemble members shown in light gray and the ensemble mean is shown in black. The various ellipses enclose 80% of the members at various lead times in days from the beginning of the forecast. The storm shown is Humberto.

c. Hurricane WRF (HWRF)

1) Atlantic

In 2012 a third nest was introduced into the HWRF allowing the resolution of the inner core to be increased to 3 km from the 9 km employed in 2011. That change, changes in the DA part of the initialization system in 2013 and various physics changes added in 2012-2013 have led to a 15%-20% per year improvement in intensity forecasts from the 2011 HWRF system. This improvement is illustrated in Figure 10. Although the intensity errors have not yet gotten to the HFIP 5-year goal (green dashed line in top left panel, Figure 10), the rate of improvement suggests that this goal will be met in 2014. Improvements in 2014 will include a higher model top (allowing more satellite data to be included in the initialization) and more frequent calls to the physics routines which preliminary indications suggest may contribute another 10-20% improvement in intensity forecasts.

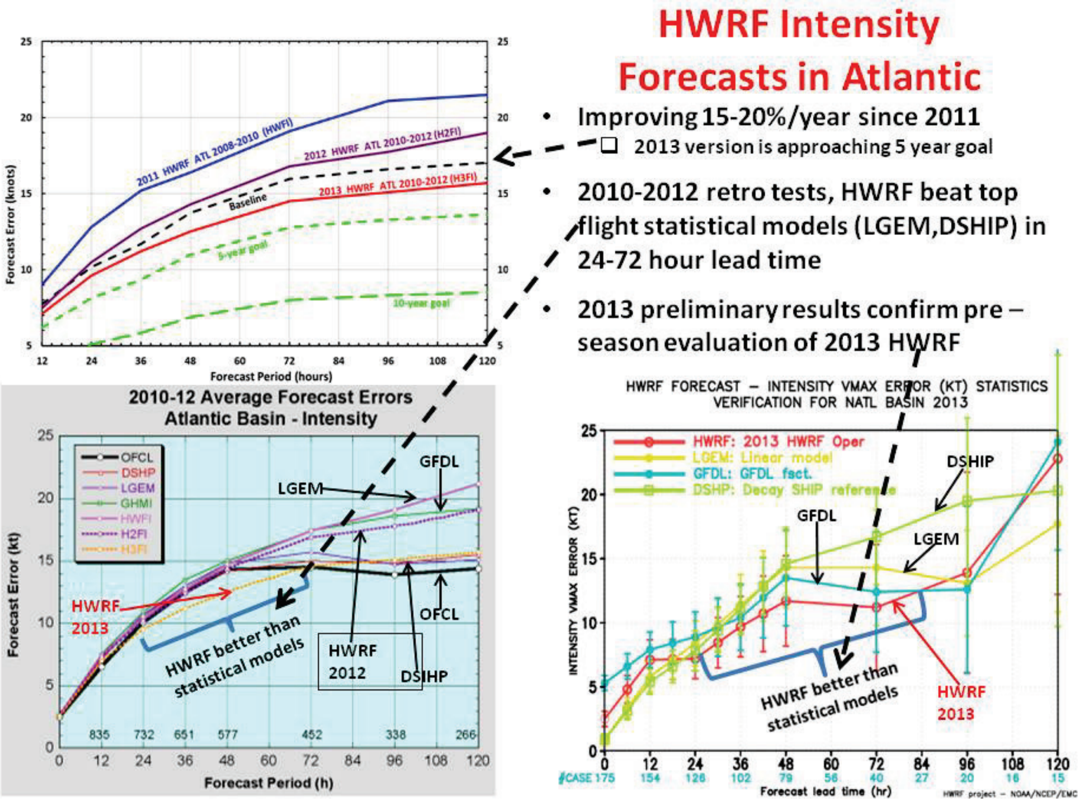


Figure 10. HWRF System Improvements 2011-2013

Improvements in the HWRF system from the 2011 HWRF system to the 2013 HWRF system (upper left panel) and comparison with other models in lower left panel. The right hand lower panel shows results from 2013. The figure is mostly self explanatory.

2) West Pacific (WPAC)

HFIP began running real-time HWRF forecasts for the WPAC in 2012 and for the Indian Ocean in 2013. These runs were done on the HFIP computers in Boulder, CO. Even though they were not run on the operational system, the reliability of those runs was quite high. The forecasts were transmitted to JTWC where they were used extensively in their forecasts. Figure 11 shows the track and intensity forecast performance of HWRF and other modeling systems in WAPC.

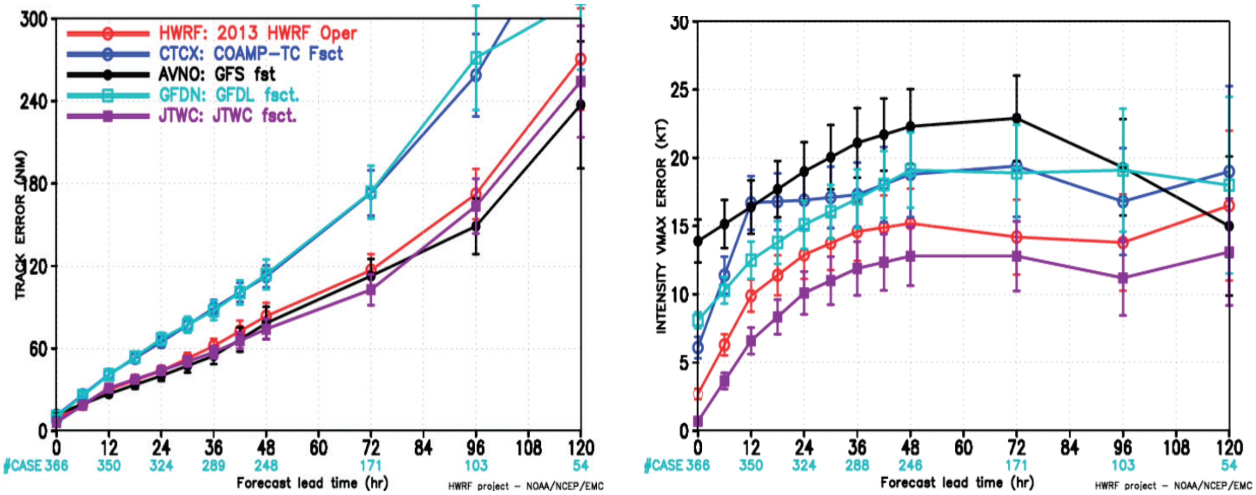


Figure 11. HWRF Track and Intensity Errors 2013 WPAC.
The HWRF in red is compared to COAMPS-TC in dark blue, the GFDLN (GFDL model run by the Navy) in light blue, the GFS in black and the official JTWC forecast in purple.

Figure 12 shows the official forecast errors in intensity for the JTWC forecasts in the WPAC starting in 2000 and going through 2013. The dramatic drop in the error for lead times of 72, 96, and 120 hours in 2013 is noteworthy. We point this out somewhat reluctantly since the drop is only for a couple of years and there have been other years with a similar decrease (2009-2010). However this was an exceptional year in the WPAC with many storms, a relatively large number of which developed rapidly and storms reaching major levels (category 3 or higher). In addition, we have been noticing a similar drop in the NHC intensity forecast error in the Atlantic where the errors curves flatten out at lead times greater than 48 hours (this is the first year we have seen that in the WPAC).

Figure 13 shows the official forecast errors in intensity for the NHC forecasts in the Atlantic basin from 1990 through 2013. Again there has been a tendency for the intensity forecast errors to be greatly reduced at the longer lead times. Whether this trend is real or not remains to be determined. The 2013 season in the Atlantic had few storms, fewer that rapidly intensified, and only two hurricanes. So, 2013 could be regarded as an “easy season” for intensity forecasts. As noted above, that was not true for the WPAC yet we see a similar trend. That there may be a similar trend in the WPAC gives us hope that it is not just a statistical quirk. In addition, at this point it is not possible to tell whether the improvements, if they are real, are related to HFIP activities such as an improved HWRF or something else.

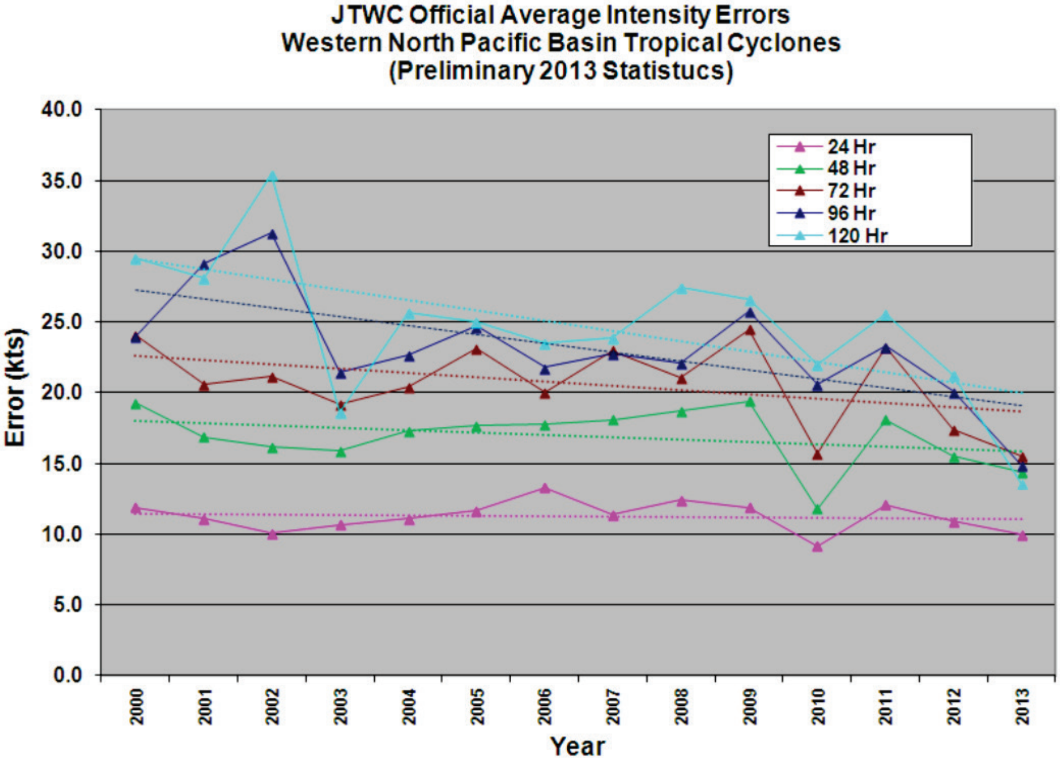


Figure 12. JTWC Official Intensity Forecast Errors, WPAC, 2000-2013

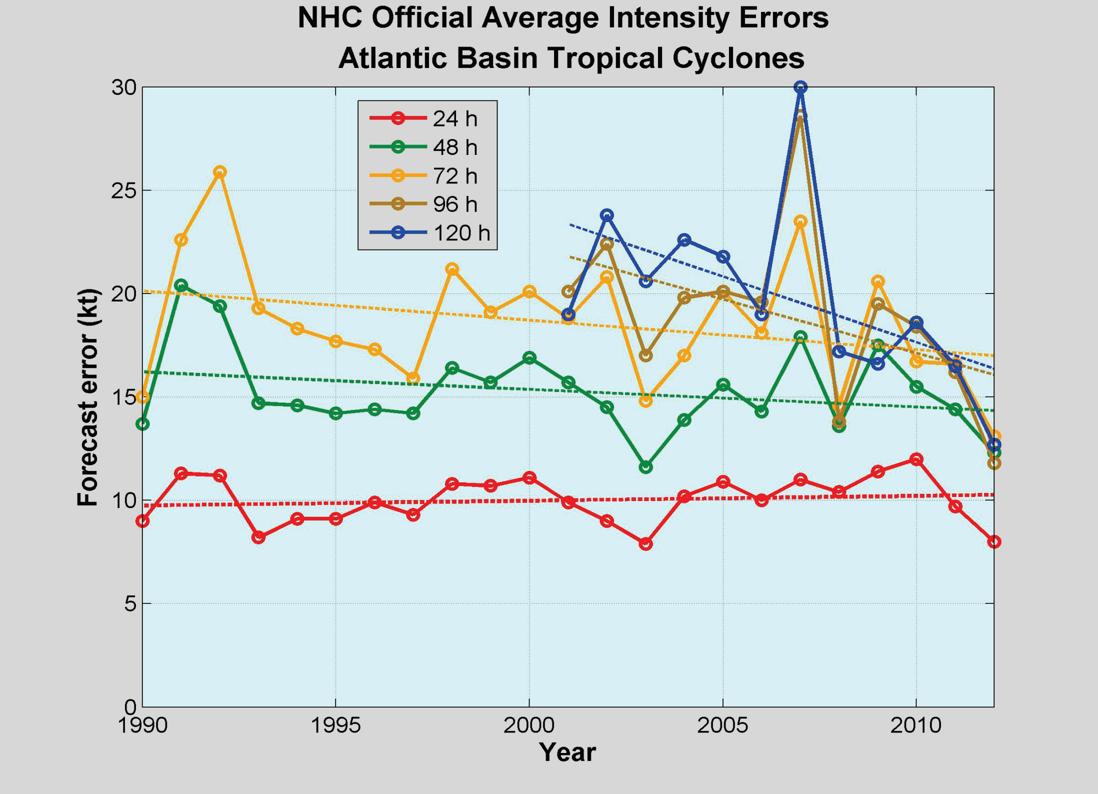


Figure 13. NHC Official Intensity Forecast Errors, Atlantic basin, 1990-2013

3) Rapid Intensification

One of the goals for HFIP is to increase the probability of detection (POD) for rapid intensification (RI) to 90% at Day 1 decreasing linearly to 60% at day 5, and decreasing the false alarm ratio (FAR) for RI to 10% for day 1 increasing linearly to 30% at day 5.

We were not able to test the ability of the 2013 HWRF to forecast rapid intensification in the Atlantic this year because of too few cases. There were enough cases for such a test in the WPAC. Figure 14 shows a comparison of the observed 24-hour intensity change in the West Pacific plotted against the HWRF model forecast, left panel, and observed compared to the JTWC official forecast in the right panel. The numbers in the lower right quadrant of each panel show the POD and FAR for each set of forecasts. The red squares show the points that were correctly forecast as rapid intensifiers. While the POD was still considerably below the 10 year goal for HFIP, the result shown in figure 14 is promising: note that the HWRF RI forecasts are better than the official forecasts.

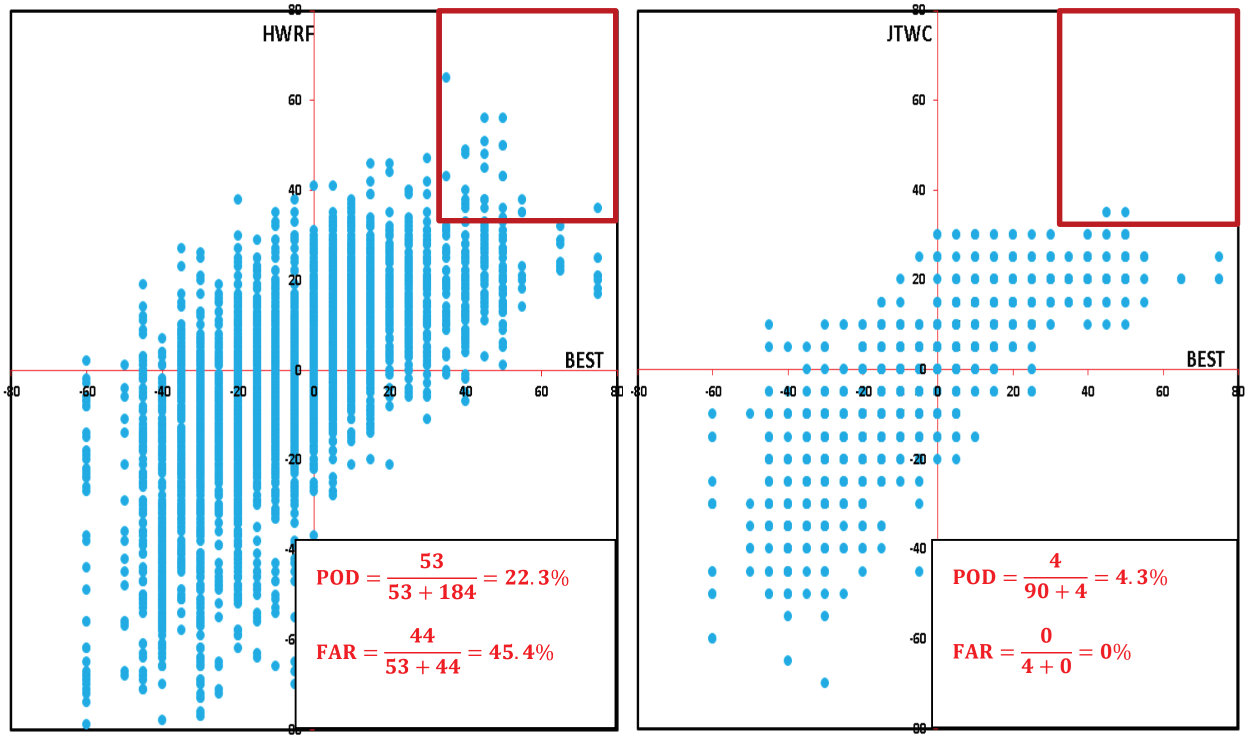


Figure 14. Forecasted 24-hour Intensity Changes in WPAC, 2013
Forecast intensity change along the x-axis and best track intensity change along the y-axes. The red square denotes the region of increased intensity greater than 30 knots in 24 hours—the region regarded as rapid intensification. The Probability Of Detection (POD) and the False Alarm Rates (FAR) are shown. The left panel is for HWRF and the right panel is for the official JTWC forecast. Note that multiple data points with the same forecast-best track value appear as a single blue dot in the plots

4) HWRF ensemble

For the first time HFIP and EMC conducted a real-time experiment on the HFIP machines in Boulder where the HWRF system was run as an ensemble. The basic model used in the ensemble was identical to the operational HWRF for 2013. A 20 member ensemble was run where the initial- and boundary-condition perturbations were from the GEFS. The members were created using the ensemble transform with a rescaling scheme. Additionally, the model physics were perturbed by adding a stochastic component to the convective trigger function in the operational HWRF. In particular, sub-grid scale convection is triggered only when the difference between partial pressure at the convection starting level and that at the level of free convection is less than the (large-scale-vertical-velocity-dependent) trigger function (120 to 180 hPa) plus a random component between -50 and 50 hPa. Random components are generated separately for each member at each cycle thereby avoiding spatial or temporal correlations. Thus the large scale perturbations of the ensemble came from the GEFS to initially define each member and then a stochastic convective trigger was added within each member.

Figures 15 and 16 compare the intensity and track forecasts of the operational model (red line) and the ensemble mean (blue line) from the ensemble experiment in the Atlantic in 2013. Because there were few long-lived storms in the Atlantic this season, there are not enough cases beyond 72 hours to draw reliable conclusions. Figure 15 shows significant improvements in the track forecast out to 72 hours.

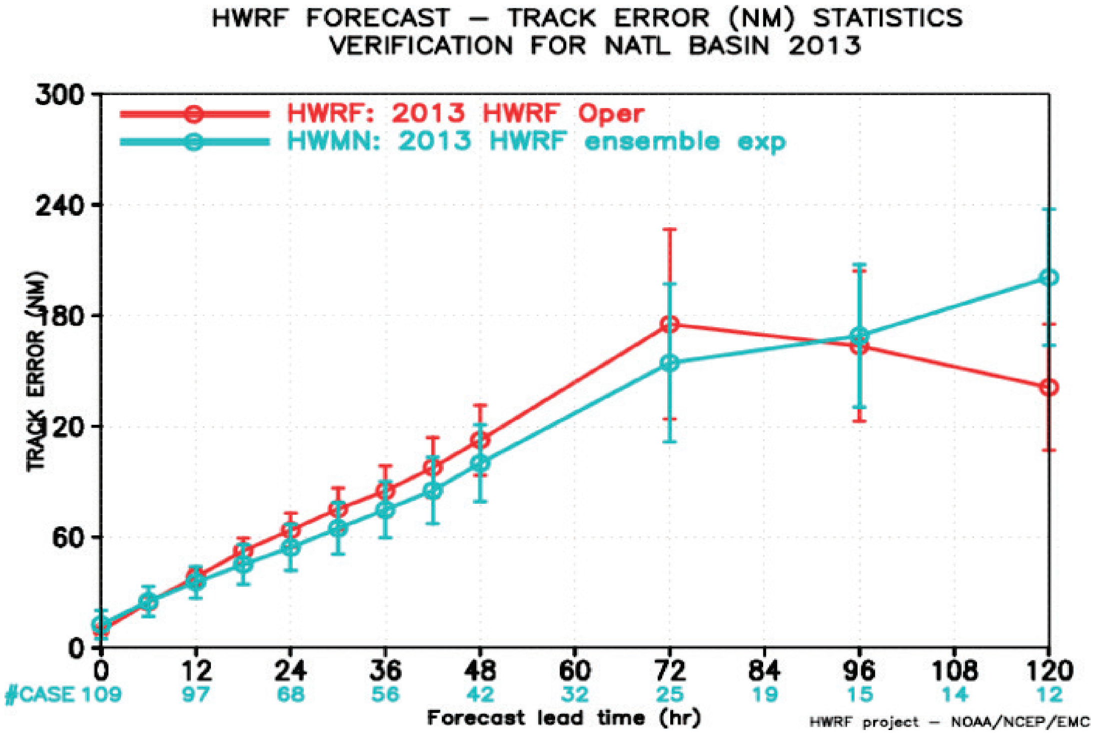


Figure 15. HWRF 2013 Experimental Ensemble Track Performance, Atlantic Basin
Red shows the operational HWRF and blue the ensemble mean. The ensemble model was identical to the operational HWRF, it used an inner grid resolution of 3 km.

Figure 16 shows even more improvement in the intensity forecast, approaching 50% at some lead times. This effect alone (from this type of ensemble) might give the additional improvement overall in intensity guidance from the HWRF system to meet the HFIP goal at 5 years (Figure 10). But there are additional improvements in the model as noted earlier that could contribute to exceeding the 5-year intensity goal.

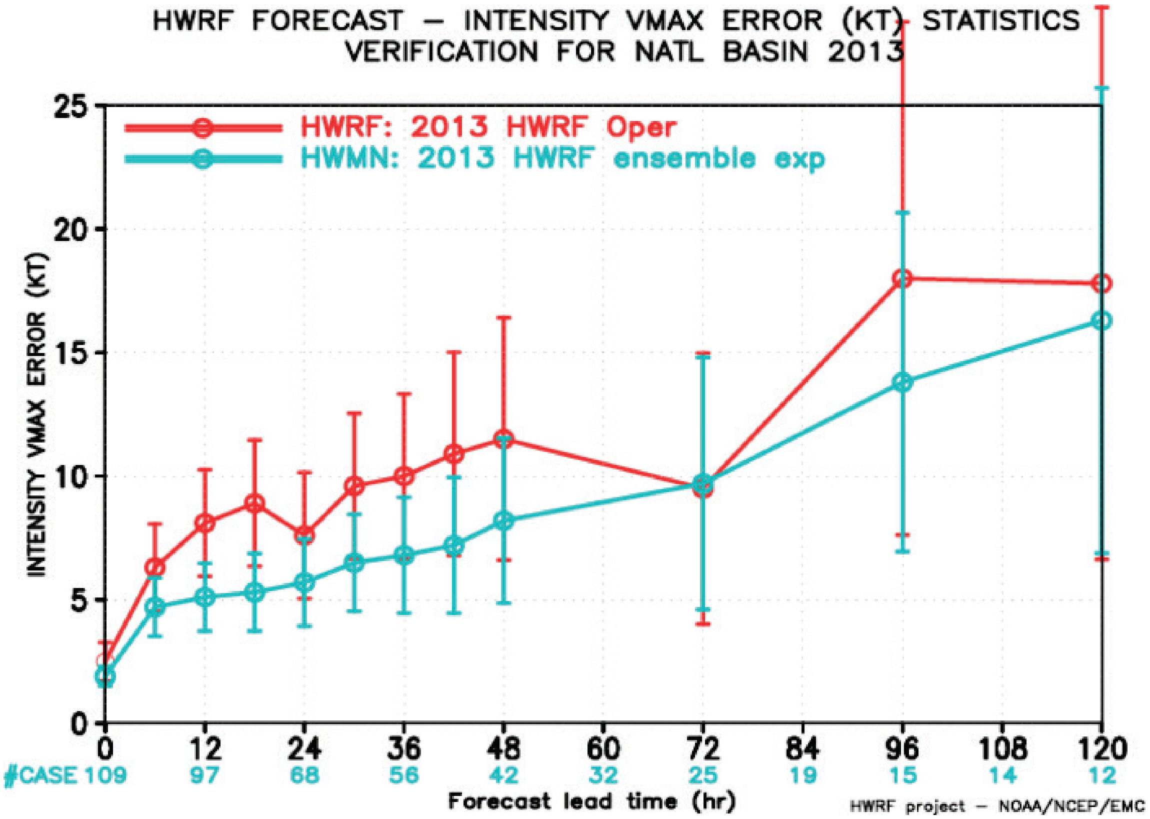


Figure 16. HWRF 2013 Experimental Ensemble Intensity Performance, Atlantic Basin. Again, red shows the operational HWRF and blue the ensemble mean. The ensemble model was identical to the operational HWRF, it used an inner grid resolution of 3 km.

The HWRF ensemble was a stream 1.5 model for the 2013 season. Hence, it was factored into the official forecast.

5) Evolution of the HWRF System

The HWRF undergoes considerable testing each year for adding new techniques and technology. This has led to some extraordinary improvements (Figure 10). The general strategy, for the next 5 years, is to gradually evolve the HWRF into the NCEP Earth Modeling System (NEMS), the infrastructure that has been adopted for other models at EMC. In addition, HWRF will move from the E-grid NMM to the B-grid NMM which is being used in other mesoscale models at EMC. At the same time, in collaboration with HRD, the HWRF system is evolving into a basin scale system where there are multiple moving doubly-nested grids, one for each hurricane, interacting with the larger basin scale domain. Figure 17 shows the extent of the basin scale system in 2012 when Kirk, Isaac, and Ileana were concurrently in the Atlantic, Gulf of Mexico, and Eastern Pacific, respectively.

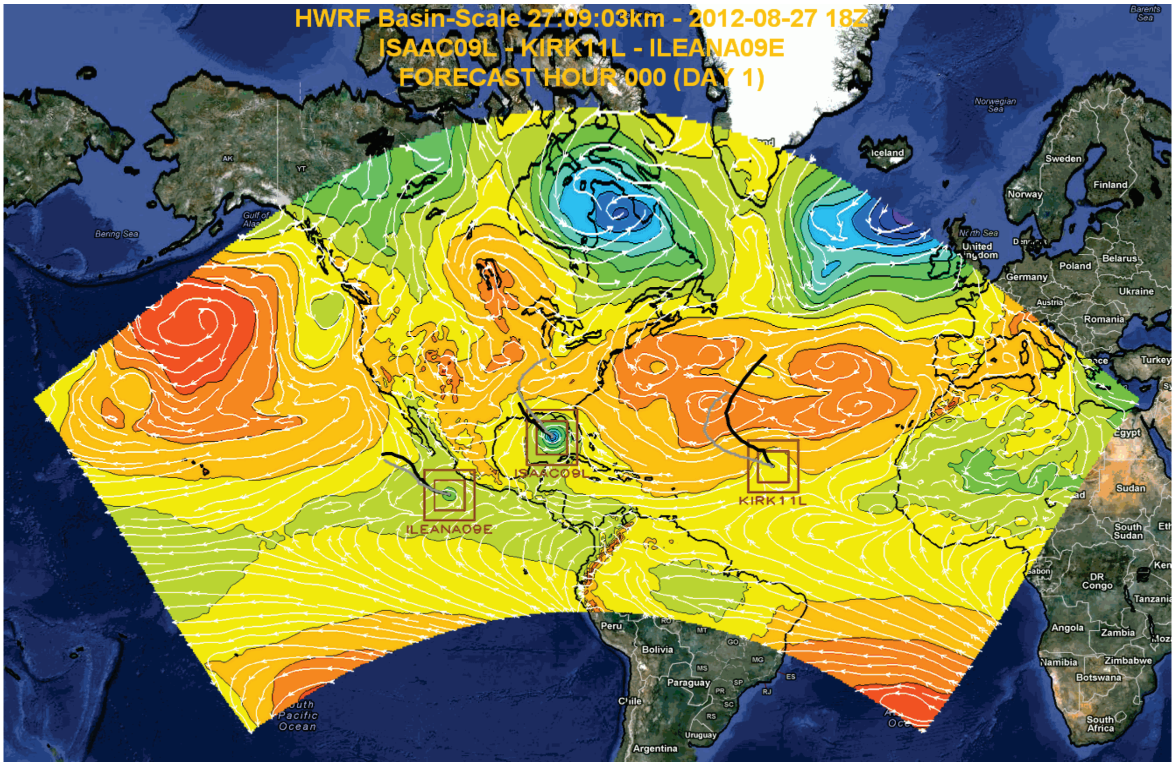


Figure 17. Basin Scale System with Three Moving Nests: Kirk, Isaac, and Ileana, 2012

The basin scale system allows modeling not only the interactions with the larger scale domain and the systems within that domain (which play a large role in hurricane track and perhaps a lesser role on intensity) but, also, interactions between simultaneous, relatively close hurricanes -- an interaction which, occasionally, can be quite important. It saves computation time because the requirement to run a large scale domain for each of the concurrent hurricanes is satisfied by one basin scale run. In addition, since the B-grid NMM can easily be made "global." The basin scale domain can be made arbitrarily bigger depending on computer resources and, eventually, can cover the globe. This will eliminate the need for a separate global model. In summary, the strategy is to smoothly transition hurricane modeling systems from the hurricane scale to the global scale. Table 7 outlines the steps that the HWRF group will undertake to reach a system composed of a global model with multiple moving nests - one for each hurricane. Note also this moving nest approach likely can be made to work well with mid-latitude systems such as mesoscale convective systems.

Table 7. HWRF development 2013-2018

System	Current (Q4FY13)	Q3FY14	Q3FY15	Q3FY18
Atmosphere	27:9:3 km, 42 levels	27:9:3, 61 levels	18:6:2 km, 64 levels	18:6:2 km, 128 levels with 10 member ensembles (in NMMB/NEMS framework)
Ocean	POM (3D ATL and 1D EPAC) 1/6° resolution 23 levels	POM (Trans-Atlantic domain at 1/12° resolution, 23 levels and 3D ocean for East Pacific	HYCOM (1/12° resolution 32 levels Trans-Atlantic and EPAC basin)	Global HYCOM (1/12o resolution, 100 levels)
Waves	None	None	WAVEWATCH III	WAVEWATCH III
Data Assimilation	One-way Hybrid EnKF-3DVAR with vortex initialization, inner core NOAA-P3 TDR DA	One-way hybrid with inner core aircraft recon data (TDR/FL) and clear sky satellite radiance DA	One-way hybrid with inner core recon data (TDR/FL); clear and inner core cloudy radiance DA	Two-way hybrid 3D/4D En-Var with inner core aircraft and all sky satellite radiance DA
Hurricane Physics	Ferrier Microphysics with explicit convection in 3km domain	Advanced Microphysics with high-resolution PBL & convection	Advanced Microphysics, NOAH LSM and land-air-sea-wave interactions	Advanced Microphysics, NOAH LSM and land-air-sea-wave interactions, coupled to wave, hydrology, surge and inundation models
Basins	NATL, EPAC	NATL, EPAC	NATL, EPAC	All Tropical Ocean Basins

7. GFDL Ensemble

For the last four hurricane seasons HFIP has promoted running an ensemble of the operational GFDL model. It uses the same model as the operational GFDL (which forms the control forecast for the ensemble—see Table 8). Working with the forecasters at NHC, scientists at GFDL constructed an ensemble by varying parameters in the initial conditions, sea surface temperatures and surface fluxes used by the model. The "unbogussed" forecasts start from the GFS without initial modification to the vortex from what was in the GFS. The other members use the GFDL initialization scheme but modified as described in Table 8.

Table 8. GFDL hurricane model ensemble members

ATCF ID	Description
GP00	Control forecast (same model as 2013 operational GFDL at NCEP)
GP01	Unbogussed
GP02	**Increase** NHC-observed V_{max} 10%, 34-kt radii 25%, 50-kt radii 40%, ROCI 25%
GP03	**Decrease** NHC-observed V_{max} 10%, 34-kt radii 25%, 50-kt radii 40%, ROCI 25%
GP04	Modification to **increase** inner-core moisture by a max of 10%
GP05	Modification to **decrease** inner-core moisture by a max of 10%
GP06	**Increase** SSTs by a max of 1°C within the initial extent of the TC
GP07	**Decrease** SSTs by a max of 2°C within the initial extent of the TC
GP08	Surface physics modification: *GFDL 2011 operational formulation* of C_D & C_H (surface drag and enthalpy exchange coefficients)
GP09	Surface physics modification: *HWRF 2012 operational formulation* of C_D & C_H (surface drag and enthalpy exchange coefficients)
GPMN	Ensemble mean computed at each lead time where the member availability is at least 4 members (40% member availability threshold)

Figure 18 shows the results for the GFDL model in the Eastern Pacific. We show EPAC instead of the Atlantic since there were many more cases there than in the Atlantic. Figure 18 compares the GFDL ensemble mean to the control run (the operational GFDL) for intensity. The GFDL ensemble mean is compared to other models in the Atlantic basin in Figures 2 and 3. In those figures the GFDL ensemble mean is denoted GPMI since it was a late model and so was interpolated as described in section 5a.

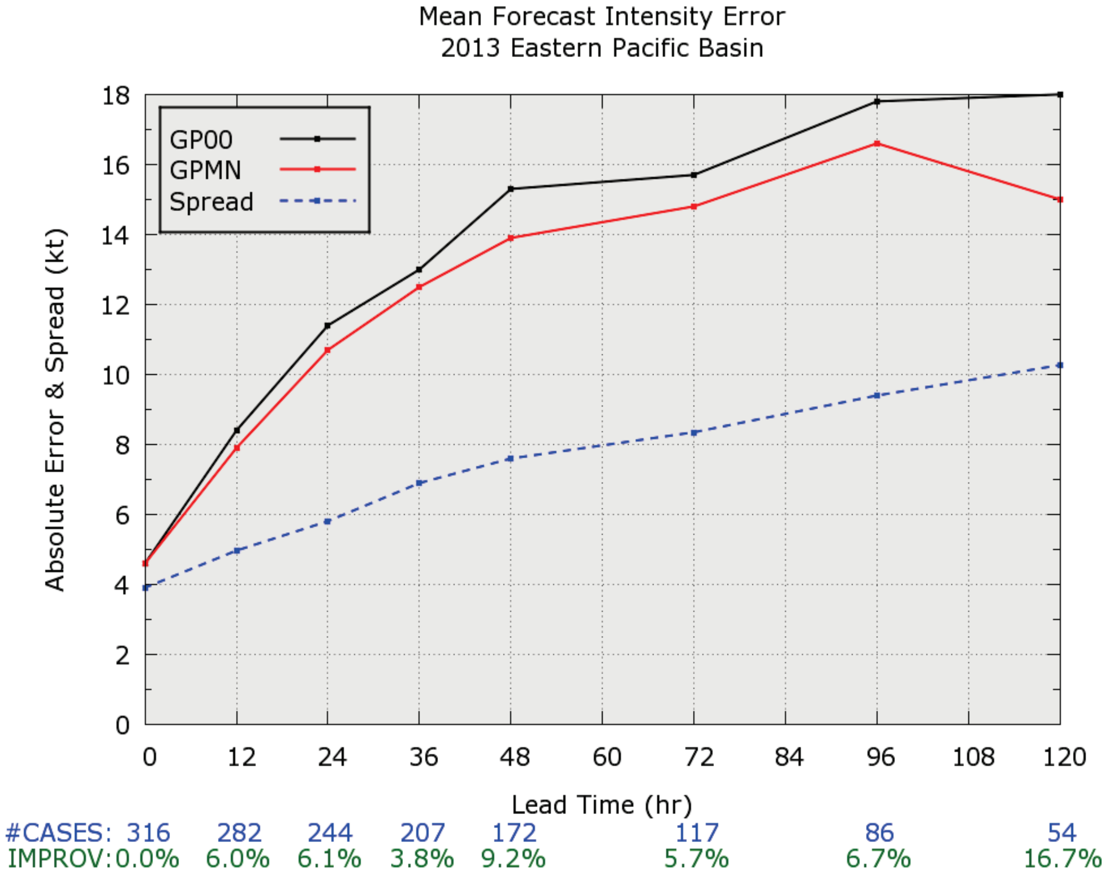


Figure 18. 2013 GFDL Ensemble Intensity Errors, EPAC.
The black line (GP00) is the control run (the operational GFDL) and the red line is the ensemble mean. The blue dashed line is the ensemble spread for intensity. The green numbers along the bottom of the figure show % improvement over the control GP00.

Figure 18 shows that the ensemble mean provided a 6%-10% improvement over the operational deterministic GFDL for intensity. Figure 2 shows that for track in the Atlantic the GFDL ensemble mean was in the middle of the pack for the comparisons shown (remember, ignore the longer lead times as the number of cases is very low).

Of the models shown in figure 3, out to 24 hours, the GFDL ensemble mean (green line) is among the best intensity predictors though the HWRF (pink line) does beat it. The ensemble spread shown by the dotted blue line on Figure 18 is only about half of what is considered to be ideal: where the spread and the mean error are about equal. However also note that the ensemble spread for intensity of the HWRF model (Figure 6 right lower panel) has roughly the same relation to the ensemble mean. We are not sure what that means.

Figure 19 shows the GFDL ensemble performance for hurricane Raymond in the EPAC. Raymond was a notoriously poorly forecast storm. Most models generally forecast no intensification of Raymond though it went through rapid deepening in the middle of its life cycle to a strong hurricane. What is shown is the average intensity error for all cases from that storm of each individual member of the ensemble and the ensemble mean. The best forecasts were by the member using the HWRF surface flux coefficients and the member that decreases the sea surface temperatures within the region of the cyclone. The worst were by the member that used the GFDL exchange coefficients and the member which increased the inner core moisture.

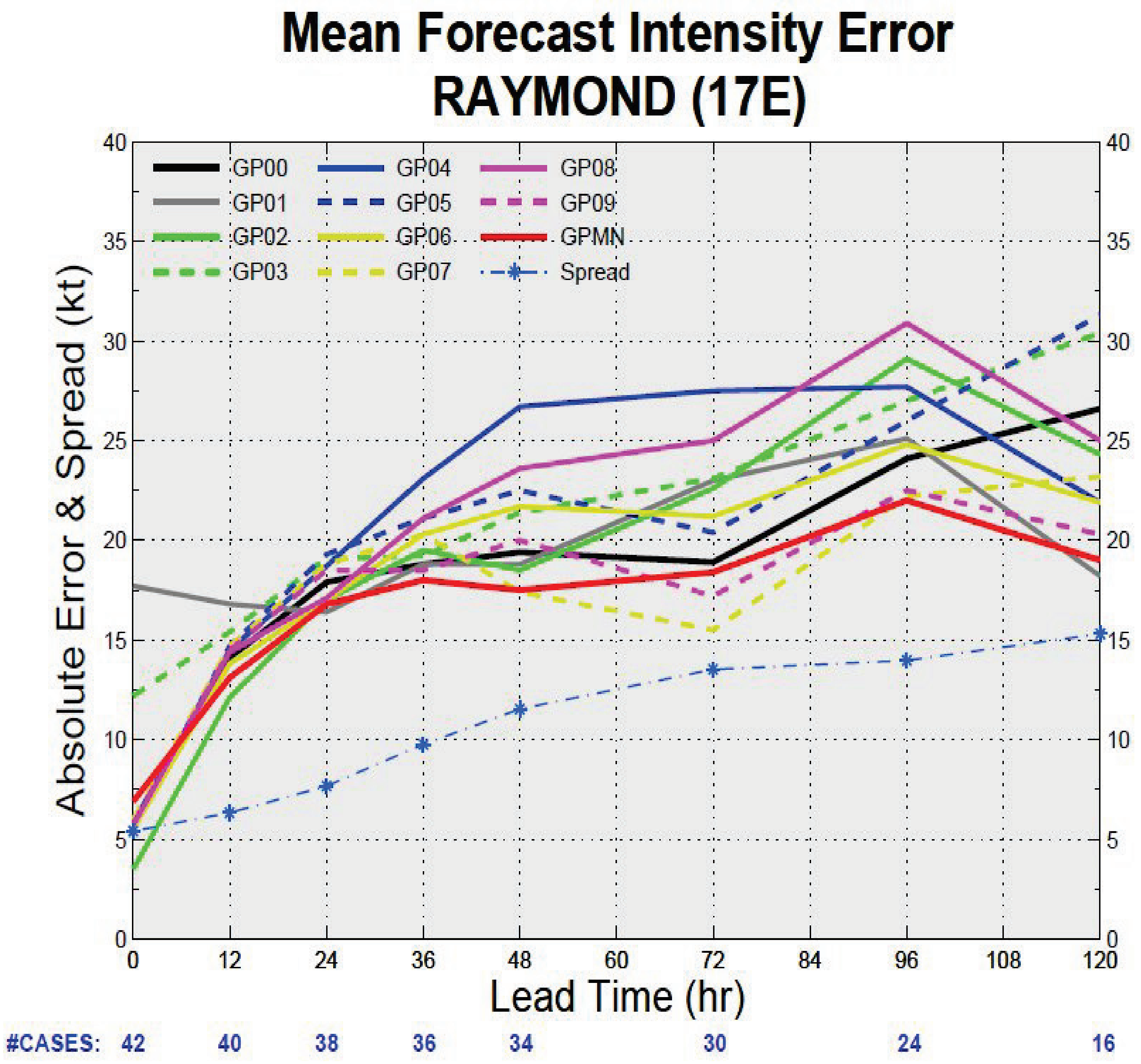


Figure 19. Intensity Errors for GFDL Ensemble Members: Hurricane Raymond.
Light blue dashed line is the ensemble spread; red line (GPMN) is the ensemble mean.

8. Impact of Inner Core Reconnaissance Data

Previous annual reports describe several past experiments that indicated including aircraft data in forecast initialization appeared to substantially improve intensity forecasts -- even out to 5 days. This year, the Reconnaissance Data Impact Tiger Team (RDITT, see table 2) was convened to conduct a systematic investigation of the impact of using aircraft reconnaissance data from the inner core of tropical cyclones on the numerical guidance provided by regional tropical cyclone models. For this investigation:

- 3 regional TC models were selected for comparison of results
  - PSU EnKF ARW hurricane analysis & prediction system
  - EMC operational 3-km HWRF (2013 version)
  - HRD HEDAS-HWRF (v3.2)

- Cases selected were Atlantic storms from 2008-2012 for which conventional aircraft reconnaissance data and/or Tail Doppler Radar (TDR) data are available in the TC inner core

- For all configurations from a particular modeling group, the same end-to-end system was used, each configuration differing only in the data included in the forecast initialization. The configurations were:
  - Control: no inner-core aircraft reconnaissance data assimilated - baseline for evaluating reconnaissance impact
  - Standard reconnaissance: conventional reconnaissance data (flight-level, dropsondes and Stepped Frequency Microwave Radiometer – SFMR, which provides estimates of surface winds) assimilated (PSU did not assimilate data obtained with the SFMR)
  - All reconnaissance: conventional reconnaissance data and tail Doppler radar (TDR) data assimilated
  - TDR only (delivered by EMC and HRD): only TDR data assimilated (no conventional reconnaissance data)

Figure 20 is the flow chart showing how the HWRF was initialized and how the inner core data were incorporated in the retrospective runs for the RDITT test.

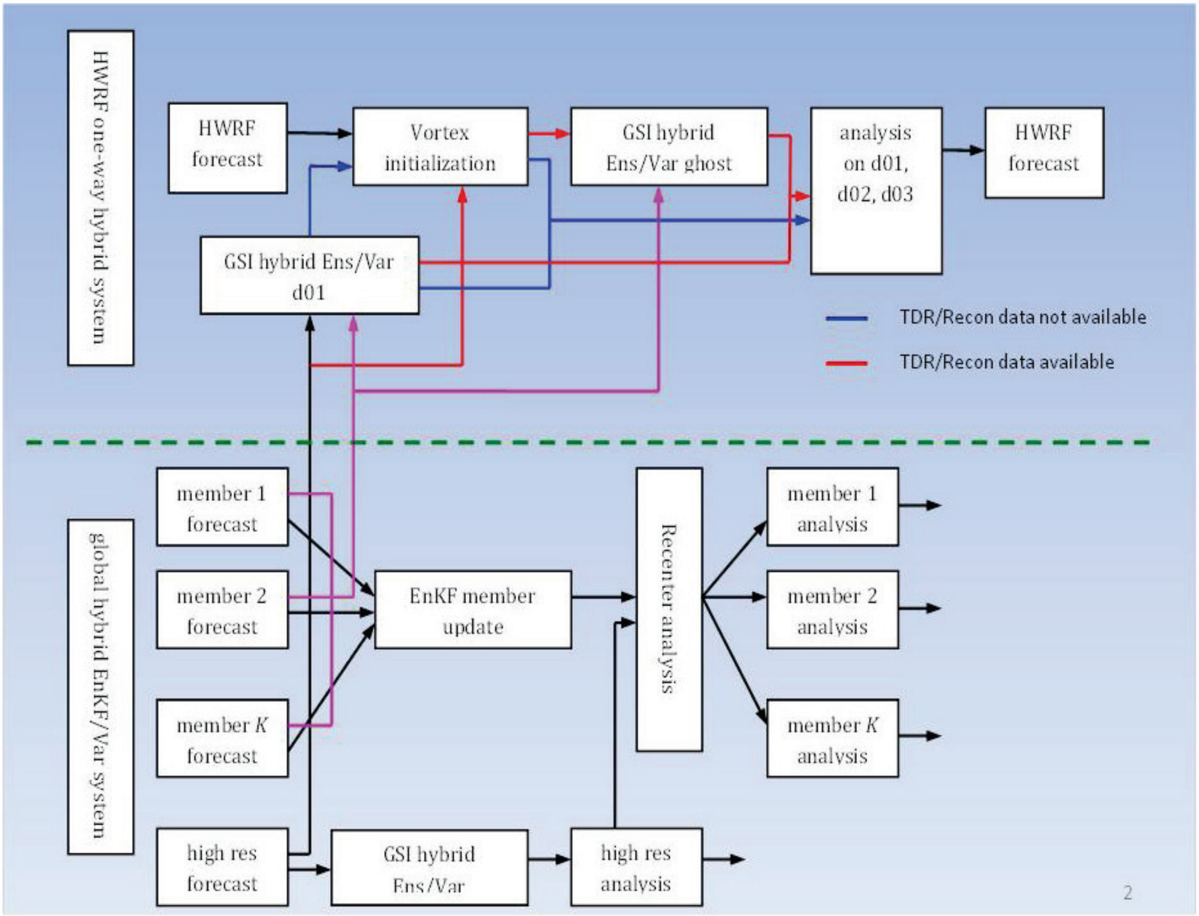


Figure 20. HWRF Initial Condition Calculation Flow Chart.

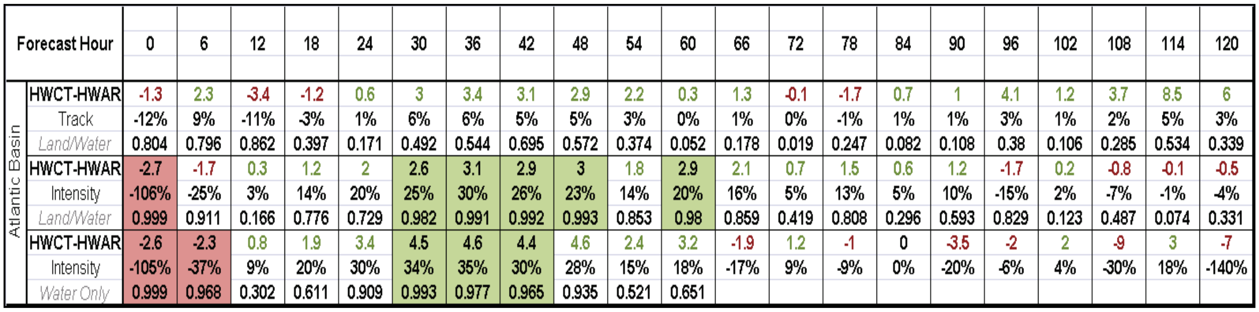
Flow chart for the initial condition calculation for the HWRF one way hybrid DA system when aircraft reconnaissance data is and is not available

For each modeling system, each configuration tested was compared with the control. The comparisons were also partitioned by the initial intensity of storm and whether the storm was over water only or water and land.

Figure 21 shows the comparison of track and intensity errors for the all reconnaissance (flight-level, dropsonde, SMFR and TDR) data in the EMC HWRF test. The formatting convention is shown in the lower left corner of the figure. Green shaded areas are where the addition of the inner core data improved the forecast in a statistically significant way; red shaded areas are where the addition of the data degraded the forecast in a statistically significant way. Here, a p-value exceeding 0.95 is regarded significantly significant

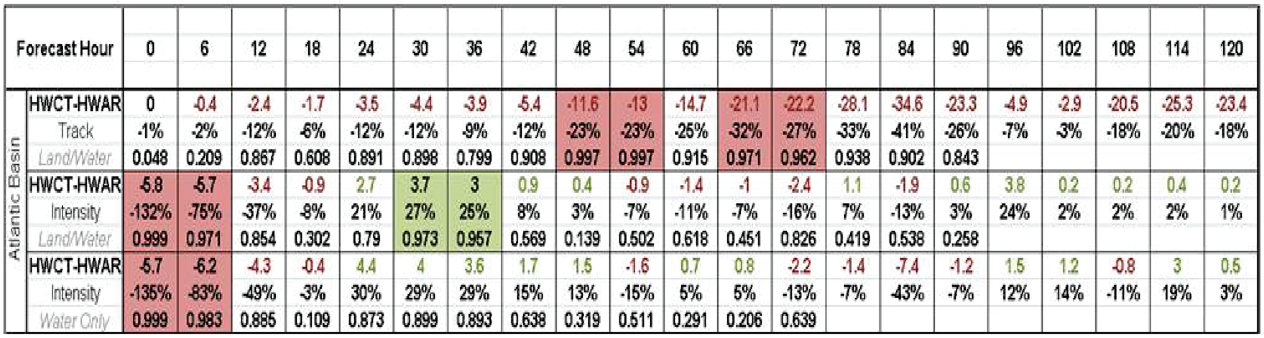
Note that, for intensity, there was a spin-down problem from the initial conditions when the data were added that degraded the forecast in the early lead times. In general, the track forecasts were only slightly improved with the addition of the inner core data -- a result that might be expected because track tends to be determined by the hurricane environment rather than details in the inner core.

Subtropical Depressions, Tropical Depressions,
Subtropical Storms, Tropical Storms Sample Size: 72 -37



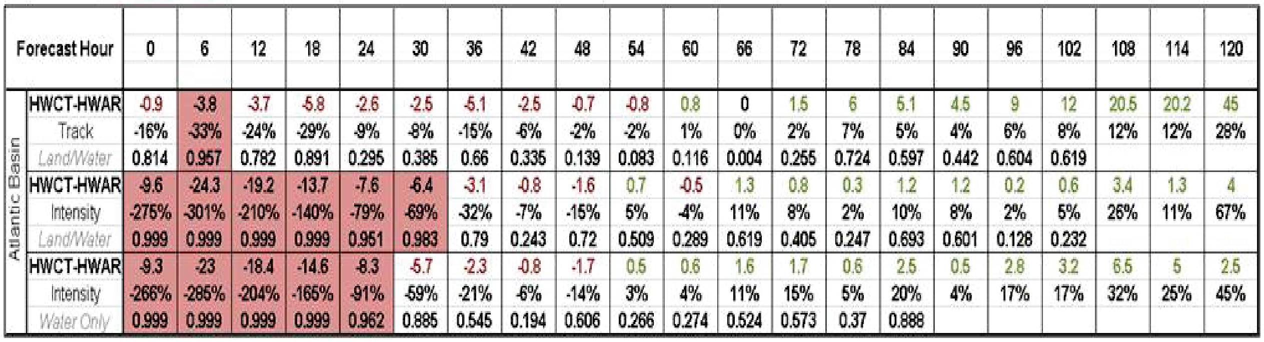
Forecast Hour		0	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102	108	114	120
Atlantic Basin	HWCT-HWAR	-1.3	2.3	-3.4	-1.2	0.6	3	3.4	3.1	2.9	2.2	0.3	1.3	-0.1	-1.7	0.7	1	4.1	1.2	3.7	8.5	6
	Track	-12%	9%	-11%	-3%	1%	6%	6%	5%	5%	3%	0%	1%	0%	-1%	1%	1%	3%	1%	2%	5%	3%
	Land/Water	0.804	0.796	0.862	0.397	0.171	0.492	0.544	0.695	0.572	0.374	0.052	0.178	0.019	0.247	0.082	0.108	0.38	0.106	0.285	0.534	0.339
	HWCT-HWAR	-2.7	-1.7	0.3	1.2	2	2.6	3.1	2.9	3	1.8	2.9	2.1	0.7	1.5	0.6	1.2	-1.7	0.2	-0.8	-0.1	-0.5
	Intensity	-106%	-25%	3%	14%	20%	25%	30%	26%	23%	14%	20%	16%	5%	13%	5%	10%	-15%	2%	-7%	-1%	-4%
	Land/Water	0.999	0.911	0.166	0.776	0.729	0.982	0.991	0.992	0.993	0.853	0.98	0.859	0.419	0.808	0.296	0.593	0.829	0.123	0.487	0.074	0.331
	HWCT-HWAR	-2.6	-2.3	0.8	1.9	3.4	4.5	4.6	4.4	4.6	2.4	3.2	-1.9	1.2	-1	0	-3.5	-2	2	-9	3	-7
	Intensity	-105%	-37%	9%	20%	30%	34%	35%	30%	28%	15%	18%	-17%	9%	-9%	0%	-20%	-6%	4%	-30%	18%	-140%
	Water Only	0.999	0.968	0.302	0.611	0.909	0.993	0.977	0.965	0.935	0.521	0.651										

Category 1 and 2 Hurricanes Sample Size: 82 – 6

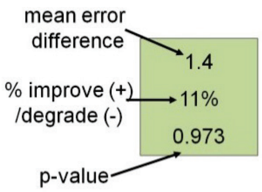


Forecast Hour		0	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102	108	114	120
Atlantic Basin	HWCT-HWAR	0	-0.4	-2.4	-1.7	-3.5	-4.4	-3.9	-5.4	-11.6	-13	-14.7	-21.1	-22.2	-28.1	-34.6	-23.3	-4.9	-2.9	-20.5	-25.3	-23.4
	Track	-1%	-2%	-12%	-6%	-12%	-12%	-9%	-12%	-23%	-23%	-25%	-32%	-27%	-33%	-41%	-26%	-7%	-3%	-18%	-20%	-18%
	Land/Water	0.048	0.209	0.867	0.608	0.891	0.898	0.799	0.908	0.997	0.997	0.915	0.971	0.962	0.938	0.902	0.843					
	HWCT-HWAR	-5.8	-5.7	-3.4	-0.9	2.7	3.7	3	0.9	0.4	-0.9	-1.4	-1	-2.4	1.1	-1.9	0.6	3.8	0.2	0.2	0.4	0.2
	Intensity	-132%	-75%	-37%	-8%	21%	27%	25%	8%	3%	-7%	-11%	-7%	-16%	7%	-13%	3%	24%	2%	2%	2%	1%
	Land/Water	0.999	0.971	0.854	0.302	0.79	0.973	0.957	0.569	0.139	0.502	0.618	0.451	0.826	0.419	0.538	0.258					
	HWCT-HWAR	-5.7	-6.2	-4.3	-0.4	4.4	4	3.6	1.7	1.5	-1.6	0.7	0.8	-2.2	-1.4	-7.4	-1.2	1.5	1.2	-0.8	3	0.5
	Intensity	-135%	-83%	-49%	-3%	30%	29%	29%	15%	13%	-15%	5%	5%	-13%	-7%	-43%	-7%	12%	14%	-11%	19%	3%
	Water Only	0.999	0.983	0.885	0.109	0.873	0.899	0.893	0.638	0.319	0.511	0.291	0.206	0.639								

Major Hurricane Sample Size: 35 – 3



Forecast Hour		0	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102	108	114	120
Atlantic Basin	HWCT-HWAR	-0.9	-3.8	-3.7	-5.8	-2.6	-2.5	-5.1	-2.5	-0.7	-0.8	0.8	0	1.5	6	5.1	4.5	9	12	20.5	20.2	45
	Track	-16%	-33%	-24%	-29%	-9%	-8%	-15%	-6%	-2%	-2%	1%	0%	2%	7%	5%	4%	6%	8%	12%	12%	28%
	Land/Water	0.814	0.957	0.782	0.891	0.295	0.385	0.66	0.335	0.139	0.083	0.116	0.004	0.255	0.724	0.597	0.442	0.604	0.619			
	HWCT-HWAR	-9.6	-24.3	-19.2	-13.7	-7.6	-6.4	-3.1	-0.8	-1.6	0.7	-0.5	1.3	0.8	0.3	1.2	1.2	0.2	0.6	3.4	1.3	4
	Intensity	-275%	-301%	-210%	-140%	-79%	-69%	-32%	-7%	-15%	5%	-4%	11%	8%	2%	10%	8%	2%	5%	26%	11%	67%
	Land/Water	0.999	0.999	0.999	0.999	0.951	0.983	0.79	0.243	0.72	0.509	0.289	0.619	0.405	0.247	0.693	0.601	0.128	0.232			
	HWCT-HWAR	-9.3	-23	-18.4	-14.6	-8.3	-5.7	-2.3	-0.8	-1.7	0.5	0.6	1.6	1.7	0.6	2.5	0.5	2.8	3.2	6.5	5	2.5
	Intensity	-266%	-285%	-204%	-165%	-91%	-59%	-21%	-6%	-14%	3%	4%	11%	15%	5%	20%	4%	17%	17%	32%	25%	45%
	Water Only	0.999	0.999	0.999	0.999	0.962	0.885	0.545	0.194	0.606	0.266	0.274	0.524	0.573	0.37	0.888						



Pairwise difference tables for the EMC model for track and intensity using all aircraft data including the TDR. Here the data have been stratified by intensity with the weakest storms shown on top and the strongest on the bottom.

Figure 21. Example RDITT Test Results: EMC HWRF, All Data.

The addition of the data did provide improvement in the intensity forecasts for the weaker storms (depression though tropical storms) for lead times between 24 and 48 hours, perhaps some improvement for weak hurricanes and little improvement for major hurricanes. For major hurricanes the spin-down problem was the worst of the three categories shown in Figure 21.

The problem with the HWRF behavior in the early part of the forecast for intensity is one that has been a problem for some time and it is hoped that improvements in the initialization system will change that. Once that is fixed, it is possible that inner core data will further improve intensity forecasts.

Figure 22 illustrates one possible advantage of adding the inner core data. The figure shows data provided by PSU as their part of the RDITT test. Shown is a plot with track error data from the control run (APCT--black) of the PSU model and track error data when the model is run with all of the inner core data (flight level, dropsonde and TDR-APAR--red). The difference in the means at each forecast lead time is small though perhaps the track performance was slightly improved when the inner core data were added.

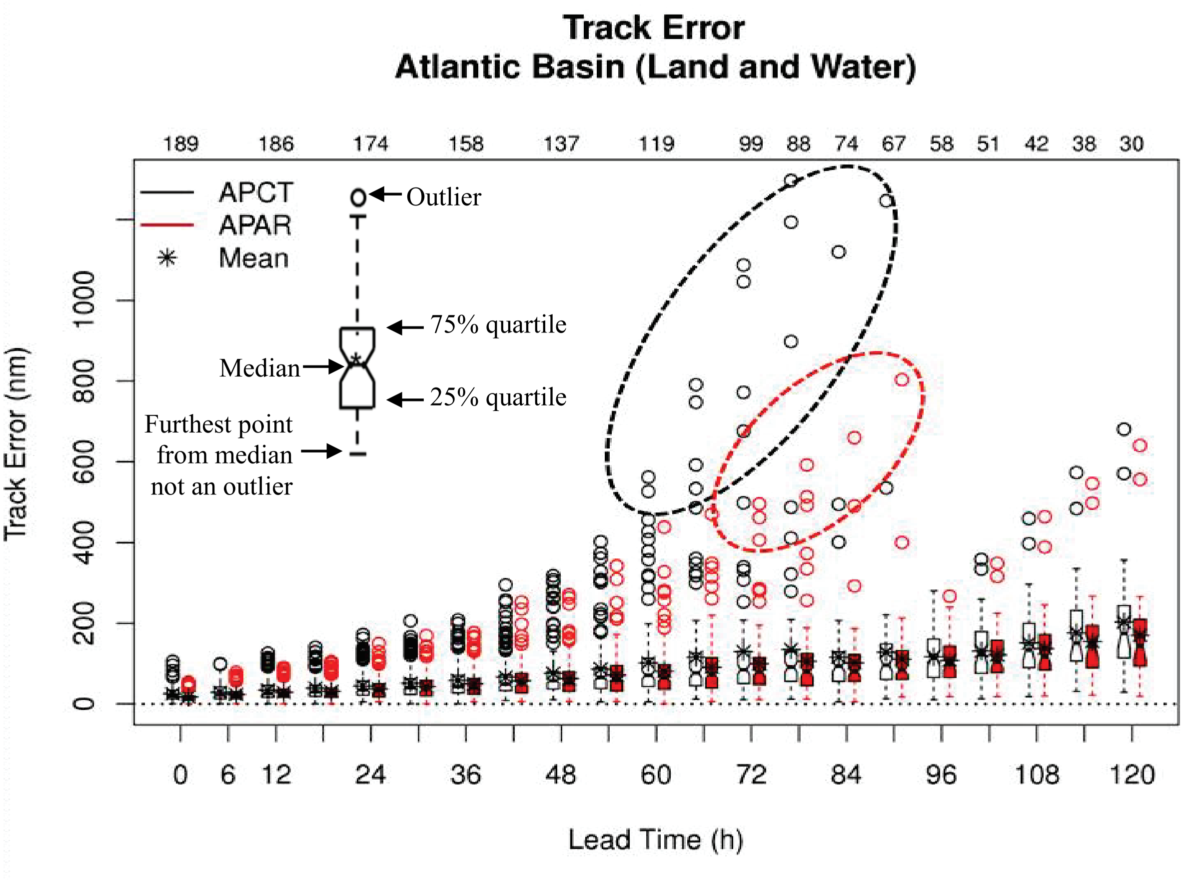


Figure 22. Outlier Example: RDITT Test PSU Model, All Data and Control
Whisker plots from the PSU model comparison using all reconnaissance data (dropsondes, flight level, and TDR data -- red) and control (no inner core data) -- black. Outliers are points greater than 1.5 times the Inter-Quartile Range (difference between the 75% and 25% quartiles) from the median. The outliers within the ellipses are discussed in the text.

Of more interest however is the reduction in the number of large outliers around the 72-hour forecast lead time. Note the outliers within the black dotted ellipse. The red ellipse indicates the reduction in magnitude of those outliers when the inner core data was included. This effect was noted for the other models in the test though not as dramatically. Also, this effect was manifest in just a few cases and may be related to a single storm.

The detailed RDITT committee report is available on the HFIP website (http://HFIP.org) when final. This report provides a summary of the results and a few examples. The main conclusions are:

- Results suggest inner core observations offer promise for improving operational TC guidance, but more work lies ahead to make optimal use of these data. While including standard reconnaissance did result in some statistically significant improvements, particularly for track, in the lead times that were improved, results lacked consistency across the regional models tested.
- The investigation of the impact of TDR data was not definitive, partly due to the small sample size.
- Results varied considerably across the three models.

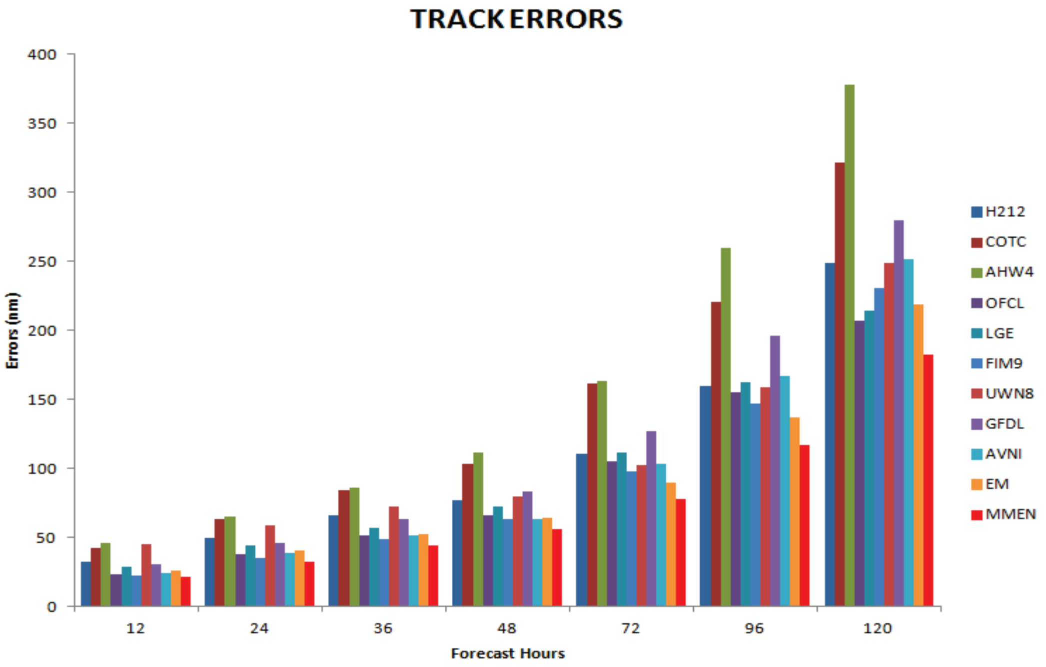
Note that many more inner core radar observations are available from Air Force C130 aircraft reconnaissance. However, because those radars are not comparable to the TDRs on the two NOAA P3s, the C130 observations were not included in the RDITT investigations.

9. Post Processing of Model Output

a. Statistical Post Processing of Model Output

Much of the discussion above focused on using numerical model improvements to achieve the HFIP goals. Typically, statistical models (for example DSHP and LGEM) are among the best predictors of hurricane intensity. A statistical model is one where a limited number of variables (measured in single to double digits) are weighted, through correlation with past data, and combined. The variables are generally selected from parameters describing the current state of the hurricane or various environmental data. Values of environmental data can be specified using current observations or model forecasts. Perhaps the simplest statistical model for intensity is SHIFOR5 (labeled OCD5 in Figure 3) where the variables are current position and intensity, position and intensity 12 hours earlier, and date. CLIPER5 is a similar model for track.

Another class of statistical model takes particular predictions (say track or intensity) from several dynamical models in a multi-model ensemble and combines those predictions as a weighted average. The weights are determined by comparing the performance of the various models over a period of years. The FSU Multi-Model Ensemble is of that class. As in past years, the FSU Multi-Model Ensemble was among or the best performers of the statistical models (see Figures 2 and 3) for 2013. Figure 23 shows the intensity and track errors in 2012 for the various models that went into the FSU Multi-Model Ensemble. The orange and red bars on the right side of the groups for each forecast lead time are the equally weighted ensemble mean (EM) and for the variably weighted ensemble mean MMEN. At all lead times the weighted mean was better than the equally weighted mean and at all lead times it was better than all the other models. At some lead times the official forecast was better.



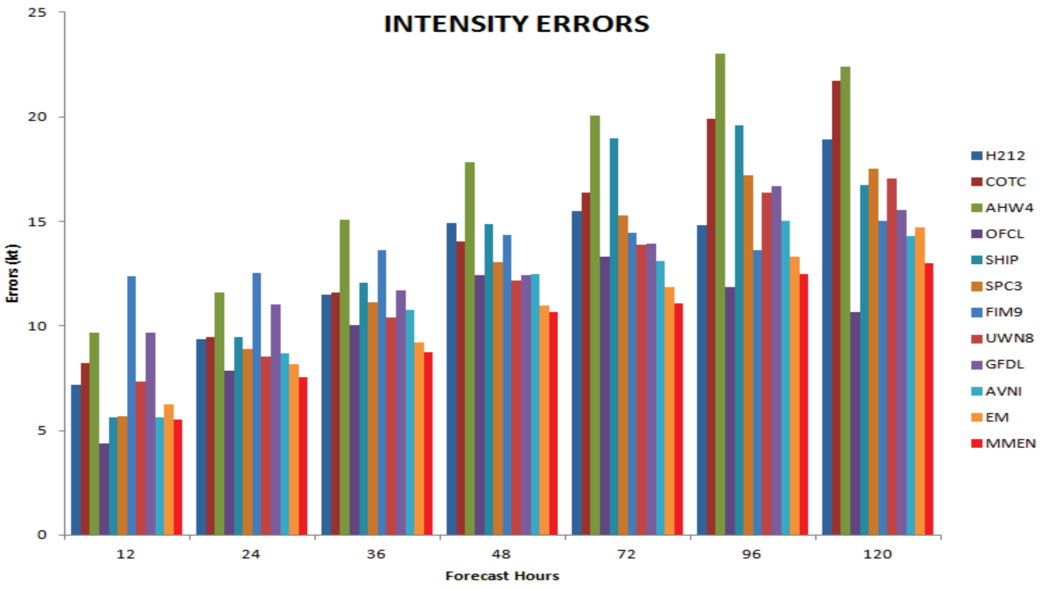


Figure 23. FSU Multi-model Ensemble: Members and Mean Performance (2012 Atlantic). Intensity errors for all components of the FSU Multi-Model Ensemble for various forecast lead times. The acronyms for the various models are shown on the right (see Appendix A for details). The ensemble mean (EM) of the models shown is in orange and the weighted ensemble mean is shown by the red bar. The official forecast is dark purple. Top-to-bottom in legend is left-to-right in charts.

More complex statistical models used operationally for intensity are SHIPS, LGEM and SPICE (SPC3). SPC3, in recent years, showed improvement compared to the operational statistical and dynamical models, by using multiple operational numerical models as input for the environmental predictors of DSHIP and LGEM. SPC3 derives input for both DSHIP and LGEM from the operational GFS, HWRF and GFDL models. This gives six variations which are then averaged as an ensemble.

In most years the statistical models provide forecasts of intensity almost as good as any of the current dynamical models. This year however the traditional statistical models (LGEM, DSHP) did very poorly. See for example Figure 3 where the SHIFOR5 results (again, labeled OCD5) are displayed. This year, the statistical models scored worse than most of the dynamical models (Figure 3). The poorer performing statistical models include LGEM, DSHP and SPC3. SPC3 did perform better at the longer lead times but, again, recall there are few cases at the longer lead times.

Both SPC3 and the FSU Multi-Model ensemble system are examples of statistical post processing being pursued by HFIP.

b. Other Post Processing

HFIP promotes the development of new diagnostic methods that can be applied to understand the source and growth of errors associated with modeling tropical cyclones. Figure 24 shows the difference in 250 hPa geopotential height between high resolution GFS ensemble member forecasts initialized on 0000 UTC 25 October. Here, the color shading shows the difference in geopotential heights for the ensemble members that predicted Hurricane Sandy would move out to sea and the ensemble members that correctly predicted Sandy's westward turn toward the New Jersey coast.

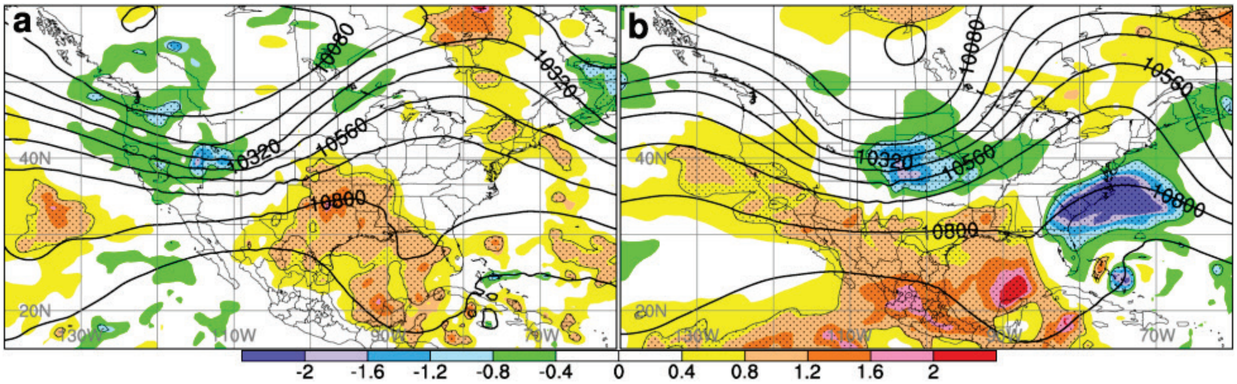


Figure 24. Post Processing Diagnostics: Hurricane Sandy 250 hPa Geopotential Heights

Normalized difference in 250 hPa geopotential height between ensemble members predicting eastward-turning tracks and those predicting westward-turning tracks (shading; units: standardized anomaly) at (a) 0 hours, and (b) 24 hours. Dotted regions indicate where the difference is statistically significant at the 95% confidence level. The contours denote the ensemble-mean 250 hPa height.

At the analysis time, the largest differences were associated with the trough to the west of Sandy over the Gulf of Mexico. By 24 hours, the largest differences were associated with the magnitude of the ridge building to the north of Sandy off the coast of the Carolinas; the ensemble members that predicted Sandy would move out to sea predicted a less amplified ridge. In those forecasts, the less amplified ridge would impart a westerly steering wind on Sandy, forcing the storm out to sea. Additional analysis suggests that the amplification of this ridge is related to the amount of latent heat release predicted along a warm front associated with Sandy, a prediction which in turn is a function of how the cumulus scheme handles upper tropospheric dry-air.

Ensembles can also be used to study how initial condition errors impact subsequent intensity forecasts. Figure 25 shows the sensitivity of the AHW 72-hour maximum wind speed forecast for Hurricane Bill (2009) to the analysis (i.e., initial conditions) of precipitable water. This figure indicates that increasing the precipitable water near the TC core or in the northeast quadrant of the TC leads to a more intense storm 72 hours later in the forecast. During that time, Bill was experiencing northeasterly vertical wind shear, allowing air from the northeast to enter the TC core. Figure 25 suggests that moistening the environmental air before it reaches the core could help mitigate the dry air entrainment which slows down the development of the TC. In addition, this result suggests that taking observations in the northeast quadrant of the storm could have improved the subsequent forecast.

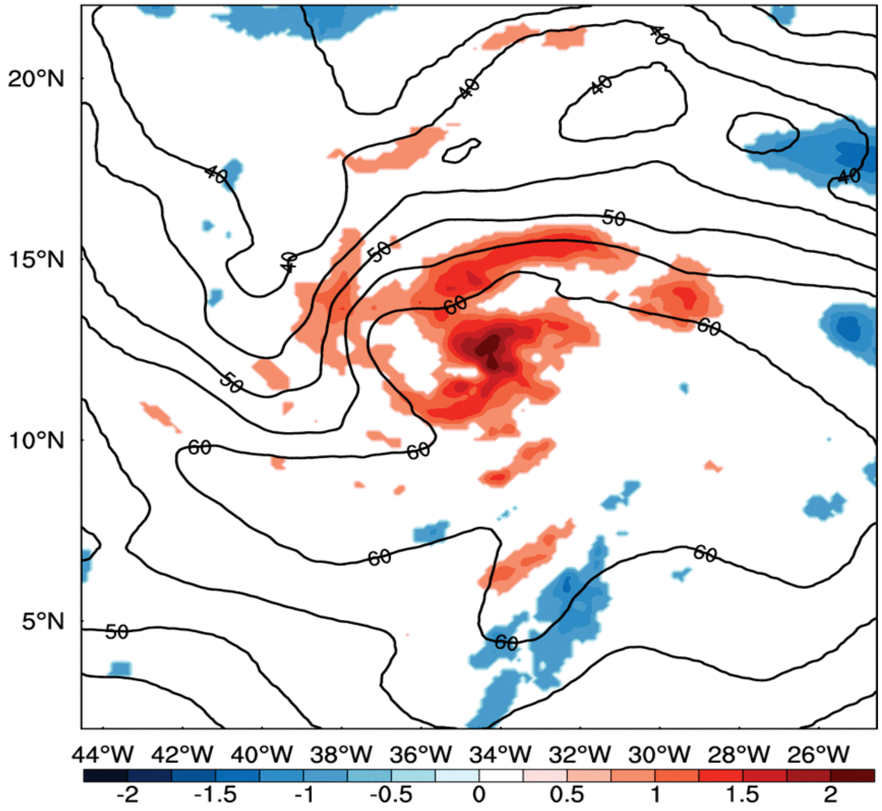
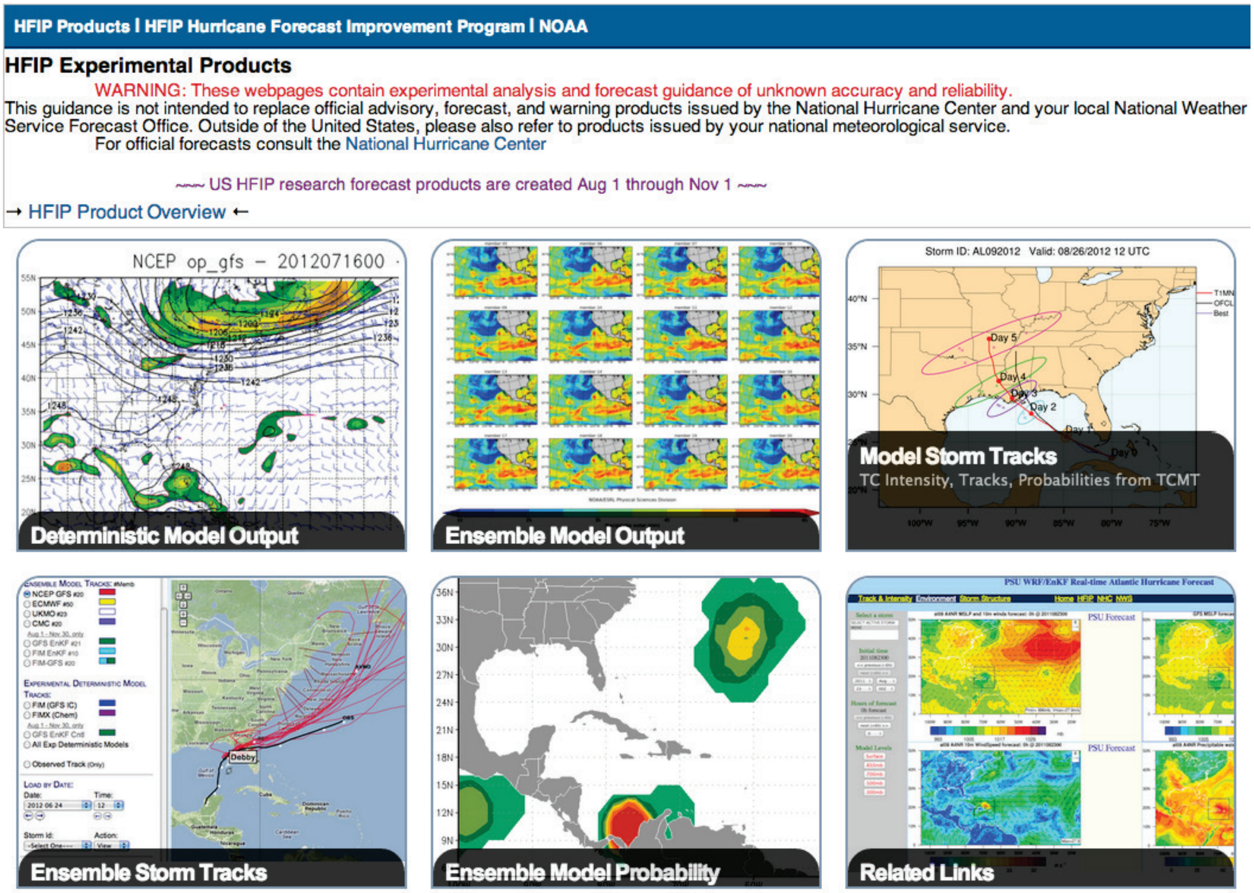


Figure 25. Sensitivity of Intensity to Precipitable Water: Hurricane Bill. Sensitivity of the 72-hour forecast of Hurricane Bill's axisymmetric maximum wind speed to the analysis of precipitable water (shading units: knots per analysis standard deviation of precipitable water) initialized at 0000 UTC 16 August, 2009. The contours are the ensemble-mean analysis fields (mm). The center of Bill is at the center of the plot.

c. The HFIP Web Page

Each hurricane season HFIP runs a real-time system on the HFIP machines in Boulder. This includes global models as well as regional hurricane models. The Stream 1.5 models, described in section 5, are part of this real-time system. HFIP maintains a website where the various results of the runs on the Boulder computer and elsewhere (NRL for example) including the operational models, are presented. This website is located at http://www.hfip.org; the products page is http://www.hfip.org/products. Figure 26 is a "screen-shot" of the products page showing the various options available. The site is open to the public and is active during the hurricane season. Some products remain available through the site during the off season. The site undergoes development each year to add capabilities and include changes to the real-time system output.



HFIP Products | HFIP Hurricane Forecast Improvement Program | NOAA

HFIP Experimental Products

WARNING: These webpages contain experimental analysis and forecast guidance of unknown accuracy and reliability.
This guidance is not intended to replace official advisory, forecast, and warning products issued by the National Hurricane Center and your local National Weather Service Forecast Office. Outside of the United States, please also refer to products issued by your national meteorological service.
For official forecasts consult the National Hurricane Center

~~~ US HFIP research forecast products are created Aug 1 through Nov 1 ~~~

→ HFIP Product Overview ←

Figure 26. HFIP.org Products Page.
~~~

10. Storm Surge

One of the goals of HFIP is to increase the skill of the forecast of the storm surge as a result of land falling hurricanes. HFIP provided the Advanced Circulation model (ADCIRC) storm surge development group within NOAA real-time access to the HFIP computers for forecasts during land-falling hurricanes in the 2013 season. The number of cases available in 2013 was quite low since there were few land-falling storms. Figure 27 shows the HFIP storm surge forecast for Tropical Storm Karen as it passed just south of Louisiana. The forecast surge heights are shown along the right side of the figure.

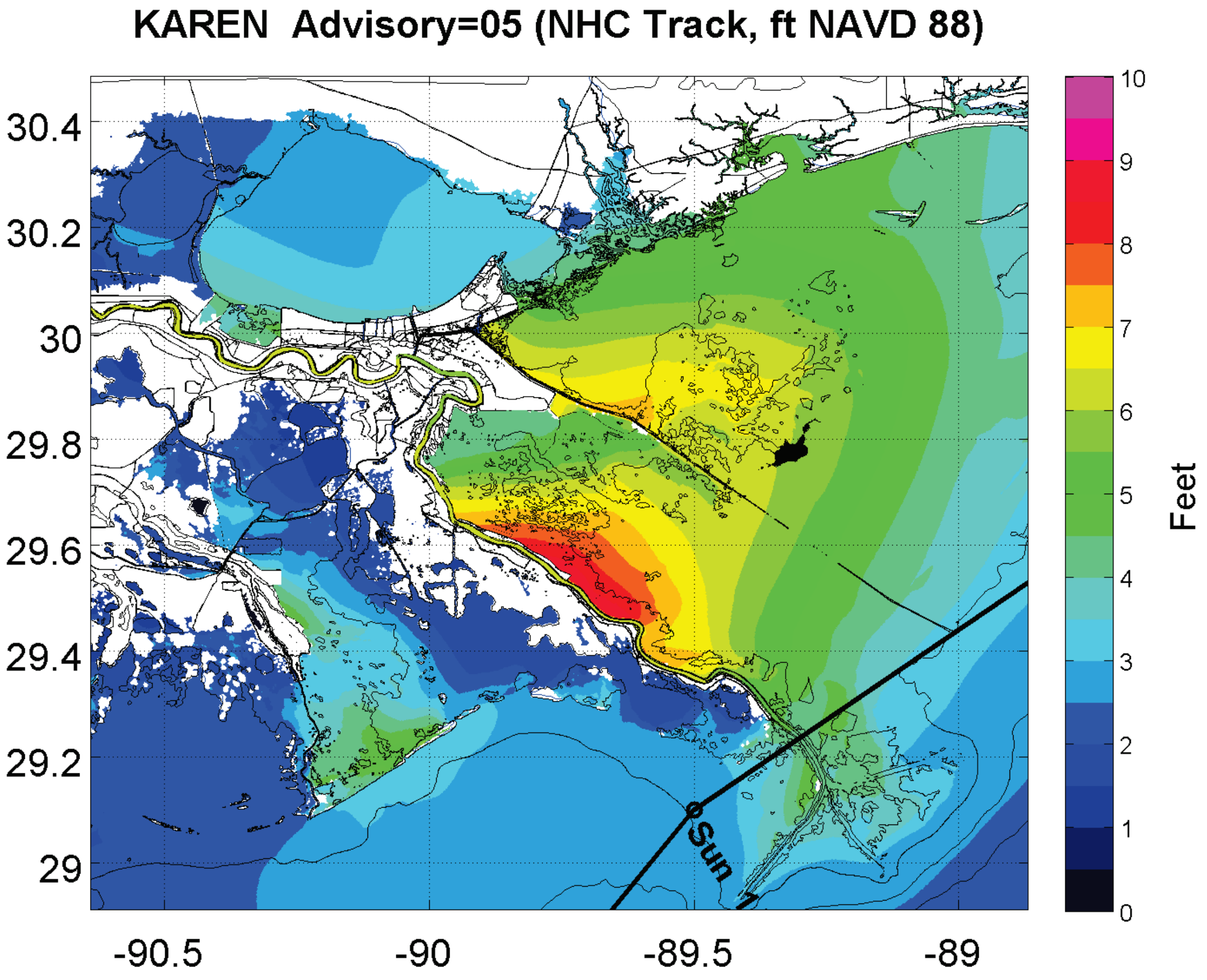


Figure 27. Storm Surge Advisory Product, Tropical Storm Karen

11. Societal Impact Work

In 2013 HFIP provided modest support for the Societal Impacts Team. Most of this support went to NHC. Figure 28 shows an example of a product developed at NHC to depict the impact of storm surge. This type of product was displayed last year in the HFIP annual report with Tampa Bay as the example and it was evaluated by the Impacts Team by surveying Emergency Managers, the various Media outlets, the general public and Warning Coordinator Meteorologists (WCMs) for effectiveness of the presentation. A very large percentage of Emergency managers, the media, public, and WCMs rated it very high in ease of understanding and providing useful information. Figure 28 is a refinement of what was presented last year and this product will likely be operational in 2014.

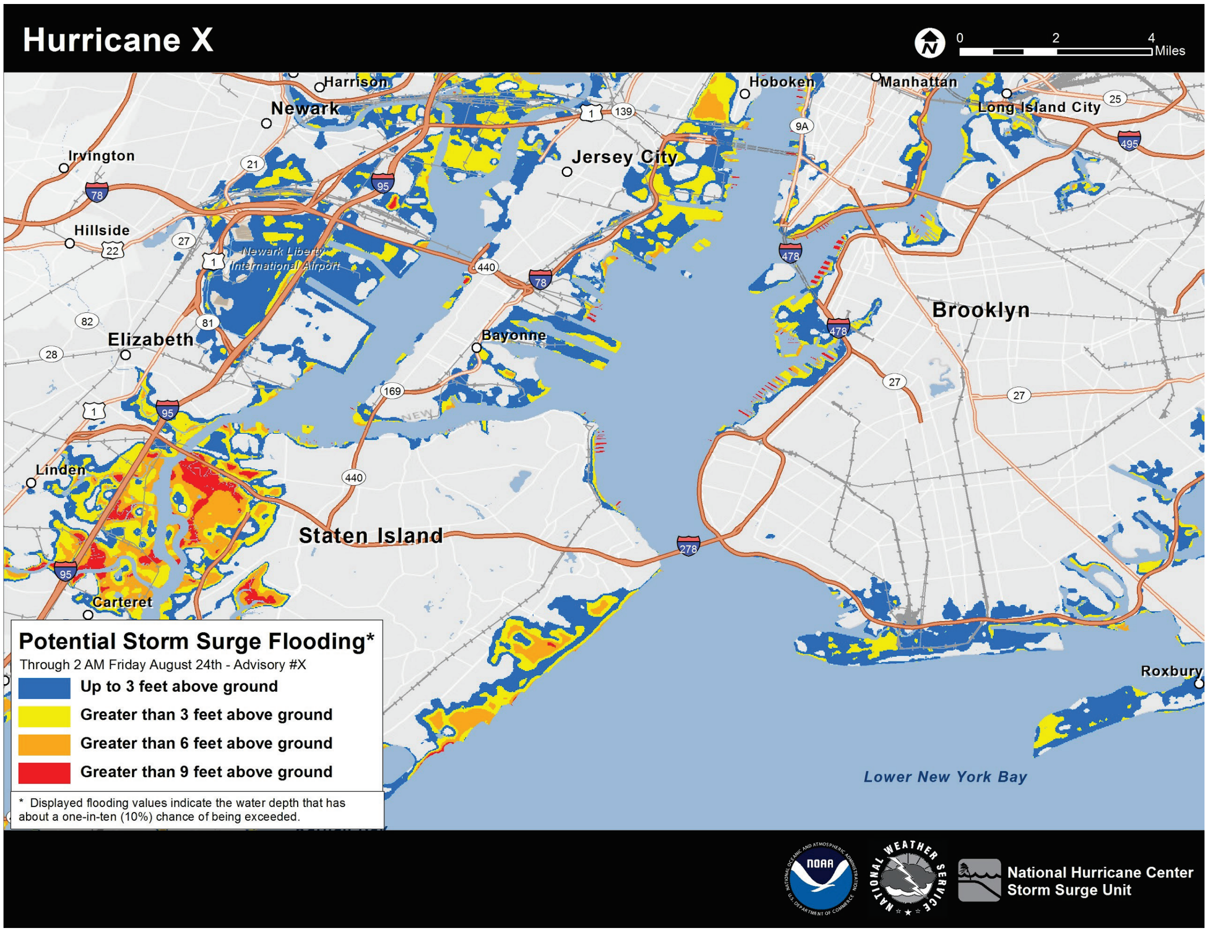


Figure 28. Preliminary Storm Surge Advisory Product.
A potential storm surge product from NHC. The storm is hypothetical but was motivated by Sandy.

12. The Configuration of a Numerical Model Hurricane Forecast Guidance System to meet the HFIP goals

With the new Hybrid Data Assimilation system and increasing resolution in both the global operational deterministic and ensemble models as well as improvements in physics and changes to the data assimilated, we feel that the operational global model will achieve at least the 5-year goals (in fact it appears to already achieved the 5-year track goals at least out to 4 days) by the 5th year of the project. Beyond 4 days there is strong evidence that the global models in general have not improved in the 5-7 day range. The reasons for this are not clear and it will be one of the scientific issues that HFIP will address in the next 5 years. However, given the success in forecasting track out to 4-5 days, we feel that not only the 5-year goals but likely the 10-year goals are within reach in those forecast lead times.

The regional models (particularly the HWRF) are now beginning to reach the track skill of the global models but it is likely that we will have to continue to focus on the global models for track forecasts. The regional models (particularly the HWRF) are just now reaching the 5- year goal for intensity, see Figure 8, and it is likely that planned changes in the HWRF in 2014 including the use of the HWRF as an ensemble, will reach that goal on time.

While HFIP has been emphasizing improving the operational HWRF, HRD and EMC with HFIP support have been experimenting with a class of models known as basin scale models. These are simply the regional scale models with a large outer domain (so they cover a couple of basins like Atlantic/ East Pacific) and are capable of running high resolution nests over more than one storm at the same time (see Section 6.c.5). These systems ran successfully during the 2012-2013 hurricane season and are showing promise. In both the single storm models and the basin scale models with multiple storms, the various nests interact with the larger outer domain so with multiple moving nests within the outer domain, concurrent individual storms can interact with each other.

This can then be easily taken one step further: the basin size can be expanded to include the entire globe. Then the regional models in a sense would be eliminated though they basically remain as the individual nests interacting with the global scale model directly.

We also note that this nested concept with a global model could also be extended to mid-latitude systems like squall lines. In other words, the basin scale concept being tested in HFIP could also likely improve forecasts of other types of weather systems, not just hurricanes.

In past years we have speculated on what the final configuration of the hurricane model might look like at the end of the program. Table 9 shows this configuration and assumes that the hurricane model will be a single global model (likely run as an ensemble) with doubly nested inner nests over each hurricane.

Table 9. Numerical Model Hurricane Forecast Guidance System

Component	**Specifications**
Global model ensemble with Hybrid Data Assimilation	20 members at 10-20 km
Multiple moving nests to 3 km within the global model	1) Double nests (9 and 3km), one for each hurricane 2) HWRF 3) Using all available aircraft and satellite data in core and near environment of hurricane
Additional models to make a multi model ensemble (possibly run as a global model with internal nests.	Multi model (at least two – e.g. HWRF, TC-COAMPS)
Statistical Post Processing	LGEM, SHIPS, SPICE, others

References

Aksoy, A., S. Lorsolo, T. Vukicevic, K. J. Sellwood, S. D. Aberson, F. Zhang, 2012: The HWRF Hurricane Ensemble Data Assimilation System (HEDAS) for High-resolution Data: The Impact of Airborne Doppler Radar Observations in an OSSE. *Monthly Weather Review*, 140(6):1843–1862, (doi:10.1175/MWR-D-11-00212.1).

Bernardet, L., S. Bao, R. Yablonsky, D. Stark, and T. Brown, 2013: Community HWRF Users' Guide V3.5a, NOAA Technical Memorandum OAR GSD-43, August, 2013. Developmental Testbed Center, 139 pp.. (available from http://docs.lib.noaa.gov/noaa_documents/OAR/GSD/NOAA_TM_GSD-43.pdf).

Cangialosi, J. P., and J. L. Franklin, 2011: 2010 National Hurricane Center Forecast Verification Report. NOAA, 77 pp. [Available online at http://www.nhc.noaa.gov/verification/pdfs/Verification_2010.pdf.]

Cohn, S. E., and D. F. Parrish, 1991: The Behavior of Forecast Error Covariances for a Kalman Filter in Two Dimensions. *Monthly Weather Review*, 119(8):1757–1785, (doi: 10.1175/1520-0493(1991)119<1757:TBOFEC>2.0.CO;2).

Daley, R. and E Barker, 2001: NAVDAS: Formulation and Diagnostics, *Monthly Weather Review*, 129(4): 869-883, (doi:10.1175/1520-0493(2001)129<0869:NFAD>2.0.CO;2)

DeMaria, M., J. Kaplan1994: A Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic Basin. *Weather and Forecasting,* 9(2):209-220, (doi: 10.1175/1520-0434(1994)009<0209:ASHIPS>2.0.CO;2).

DeMaria, M., 2009: A Simplified Dynamical System for Tropical Cyclone Intensity Prediction. *Monthly Weather Review*, 137(1):68-82 (doi: 10.1175/2008MWR2513.1)

DTC, cited 2012: Community Grid point Statistical Interpolation GSI Documents and Publications. [Available online at http://www.dtcenter.org/com-GSI/users/docs/index.php.]

Evensen, G., 1994: Sequential Data Assimilation with a Nonlinear Quasi-geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics. *Journal of Geophysical Research*, **99**(C5):10,143–10,162.

Hamill, T.M., J. S. Whitaker, M. Fiorino, and S. G. Benjamin, 2011: Global Ensemble Predictions of 2009's Tropical Cyclones Initialized with an Ensemble Kalman Filter. *Monthly Weather Review*, 139(2):668-688, (doi: 10.1175/2010MWR3456.1).

Houtekamer, P. L., and H. L. Mitchell, 1998: Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Review*, 126(3):796–811 (doi: 10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2)

Keppenne, C.L., 2000: Data Assimilation into a Primitive-Equation Model with a Parallel Ensemble Kalman Filter. *Monthly Weather Review*, 128(6):1971-1981, (doi: 10.1175/1520-0493(2000)128<1971:DAIAPE>2.0.CO;2).

NHC, cited 2009: NHC Track and Intensity Models. [Available online at http://www.nhc.noaa.gov/modelsummary.shtml.]

NOAA SAB, cited 2006: Hurricane Intensity Research Working Group Majority Report. [Available online at http://www.sab.noaa.gov/Reports/HIRWG_final73.pdf.]

NRL, cited 2012: NAVDAS Source Book 2001. [Available online at http://www.nrlmry.navy.mil/docs/_NAVDAS01.pdf.]

Parrish, D.F. and J. C. Derber, 1992: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. *Monthly Weather Review*, 120(8):1747-1763, (doi:10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2)

Tallapragada, V., L. Bernardet, S. Gopalakrishnan, Y. Kwon, Q. Liu, T. Marchok, D. Sheinin, M. Tong, S. Trahan, R. Tuleya, R. Yablonsky, and X. Zhang, 2013: Hurricane Weather Research and Forecasting (HWRF) Model: 2013 Scientific Documentation. Developmental Testbed Center, 99 pp. (available from http://www.dtcenter.org/HurrWRF/users/docs/.)

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble Square Root Filters. *Monthly Weather Review*, 131(7):1485-1490, (doi:10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2).

Wu, W-S., J. Purser, and D. F. Parrish, 2002: Three-Dimensional Variational Analysis with Spatially Inhomogeneous Covariances. *Monthly Weather Review*, 130(12):2905-2916 (doi: 10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2).

Appendix A: HFIP Funding Opportunity for Academia and Private Industry

Through an Announcement of Opportunity (AO) in March 2011, HFIP Program Office invited proposals from academic institutions and private industry offering two-year support of their expertise and experience to contribute towards the advancement of hurricane science and modeling to improve operational hurricane forecasts. The priorities set for this AO include advancements in data assimilation techniques focusing on use of in-situ measurements obtained through aircraft reconnaissance and satellite datasets; advancements in numerical weather prediction techniques with emphasis on high-resolution physics for more accurate representation of moist processes and air-sea interactions in the hurricane core region; advanced model diagnostics to support model improvements; advanced ensemble techniques for improved track and intensity forecast guidance; and enhanced observing strategies and use of observations for hurricane environment- and hurricane core-scale circulations.

Of the 34 proposals received in response to this AO, 12 were selected for funding based on their relevance to the program priorities, feasibility of transition to operations, active participation in HFIP supported activities, and recommendations from the review committee. A complete list of selected proposals is available on the HFIP website at http://www.hfip.org/HFIP%20Grant%20Selected%20Proposals.pdf. Among the 12 selected proposals, four address the data assimilation priorities; three address operational hurricane model improvements; two each address high-resolution physics and ensembles, and one addresses tropical cyclogenesis. Progress of each of these proposals is monitored through quarterly reports from the principle investigators and through scientific presentations from the Principal Investigators in various forums including HFIP bi-weekly teleconferences, monthly EMC-HRD modeling meetings and EMC HWRF weekly meetings. All of the funded efforts showed significant progress during the first year and HFIP extended funding to each of them for the second year.

HFIP released a second round AO in 2013 with an increased focus on advanced physics for land-air-sea interactions, hurricane intensity predictability, global-to-local scale modeling techniques and continuous improvements to operational hurricane models and data assimilation techniques. The review process for these proposals has been completed and announcements of awards will be made in the spring 2014. When they are announced, the list of the winning proposals will be made available on http://HFIP.org.

Appendix B: Model Acronyms

The following is a list of acronyms used to identify models in this document. Many of the acronyms follow the four-character naming convention in the Automated Tropical Cyclone Forecasting (ATCF) system. For example, 6-hour "earlier" forecasts from "late" models (see Section 5.a) are adjusted so that the previous 6-hour forecast matches the conditions at the beginning of the current forecast. This is simply known as an interpolated forecast. Forecasts of those future conditions are denoted with an "I" at the end, for "interpolated" (12-hour interpolations are denoted with a "2").

Other conventions (although not exclusively) in the model naming include using the acronym "A" to denote advanced version, "D" to denote the addition of inland decay, "E" to denote ensemble, "H" to denote hurricane, "R" to denote research, "S" to denote statistical, "T" to denote track, "V" to denote Variable (ensemble of at least 2, for example), and beginning with an "I" to denote intensity.

ADCIRC:	Advanced Circulation Model for oceanic, coastal and estuarine waters
AEMI:	GEFS with 6-hour interpolation.
AVNI:	GFS with 6-hour interpolation.
AHW:	National Center for Atmospheric Research Advanced Hurricane WRF.
AHWI:	AHW with 6-hour interpolation.
APSI:	ARW with 6-hour interpolation
ARW:	Pennsylvania State University Advanced Research WRF
CMC:	Canadian Meteorological Centre model.
CMCI:	CMC with 6-hour interpolation.
CLIPER5:	Climate and Persistence model.
COAMPS-TC:	Fleet Numerical Meteorology and Oceanography Center Coupled Ocean/Atmosphere Mesoscale Prediction System-Tropical Cyclone.
COTI:	COAMPS-TC with 6-hour interpolation.
Decay-SHIFOR5:	Decay Statistical Hurricane Intensity Forecast model.
DSHP:	Decay SHIPS.

ECMWF:	European Centre for Medium-range Weather Forecasts model.
EGRI:	United Kingdom Meteorological Office model, subjective tracker, with 6-hour interpolation.
EMXI:	ECMWF with 6-hour interpolation.
FIM:	Flow-following finite-volume Icosahedral Model.
FM9I:	FIM with 6-hour interpolation
FSSE	Florida State University Super Ensemble
G01I:	GFDL ensemble member 01 with 6-hour interpolation (in general, G##I denotes GFDL ensemble member ## with 6-hour interpolation).
GDAS:	Global Data Assimilation System
GEFS:	National Centers for Environmental Prediction Global Ensemble Forecast System.
GFDI:	GFDL with 6-hour interpolation.
GFDL:	Geophysical Fluid Dynamics Laboratory model.
GFNI:	Navy version of GFDL with 6-hour interpolation.
GFS:	Global Forecast System.
GFSI:	GFS with 6-hour interpolation.
GHMI:	GFDL adjusted using a variable intensity offset correction that is a function of forecast time, with 6-hour interpolation.
GPMN:	GFDL ensemble mean
GPMI:	GFDL ensemble mean (note all members of the ensemble include 6-hour interpolation).
HWFI:	HWRF with 6-hour interpolation.
HWRF:	Hurricane WRF.
ICON:	National Hurricane Center Intensity Consensus.

IV15:	Intensity forecast ensemble including 2012 stream 1.5 forecasts.
LGEM:	Logistics Growth Equation Model.
NAVGEM:	Fleet Numerical Meteorology and Oceanography Center Navy Global Environmental Model (replaced NOGAPS February, 2013).
NGPI:	NOGAPS with 6-hour interpolation.
NGXI:	Experimental NOGAPS with 6-hour interpolation.
NOGAPS:	Fleet Numerical Meteorology and Oceanography Center Navy Operational Global Atmospheric Prediction System (replaced by NAVGEM February, 2013).
NMM:	Environmental Modeling Center Nonhydrostatic Mesoscale Model.
SHIPS:	Statistical Hurricane Intensity Prediction System.
SPC3:	Six member weighted SPICE ensemble using output from GFS, HWRF, and GFDL as input for DSHP and LGEM. The ensemble weights vary with forecast lead time.
SPICE:	Statistical Prediction of Intensity from a Consensus Ensemble.
TV15:	Track forecast ensemble including 2012 stream 1.5 forecasts.
TVCA:	Track Variable Consensus of at least two of AVNI, EGRI, EMXI, NGPI, GHMI, HWFI forecasts
TVCN:	National Hurricane Center Track Variable Consensus
UKMI:	United Kingdom Meteorological Office model, automated tracker, with 6-hour interpolation.
UWNI:	UW-NMS with 6-hour interpolation.
UW-NMS:	University of Wisconsin Nonhydrostatic Modeling System
WRF:	Weather Research and Forecasting model. It is a regional system with options for the dynamic core, physics, initialization, post processing and verification. Variations include the Hurricane WRF (HWRF), PSU Advanced Research WRF (ARW), and NCAR Advanced Hurricane WRF (AHW)