NOAA

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

United States Department of Commerce

HURRICANE FORECAST IMPROVEMENT PROJECT

2014 HFIP R&D Activities Summary: Recent Results and Operational Implementation

May 2015

HFIP Technical Report: HFIP2015-1

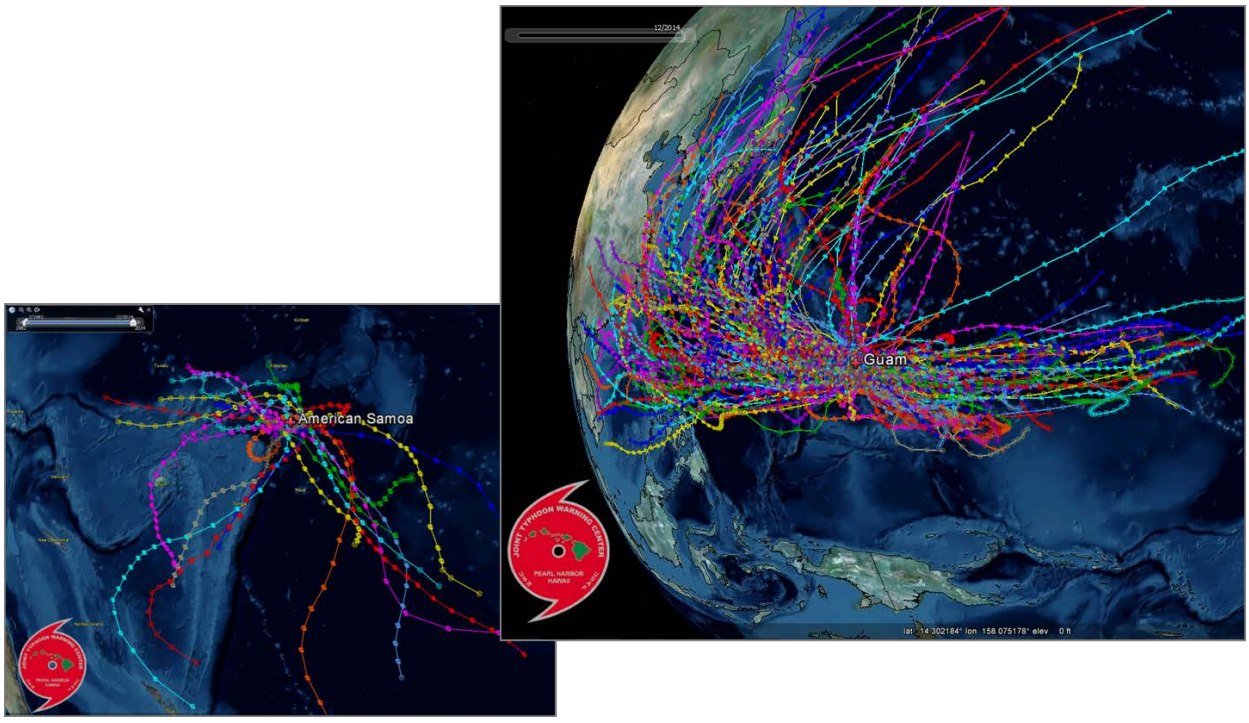


Image on cover page
used with permission from the
Joint Typhoon Warning Center, Hawaii

2014 HFIP R & D Activities Summary: Recent Results and Operational Implementation

R. Gall[14], F. Toepfer[14], F. Marks[1], E. N. Rappaport[12], A. Aksoy[22], S. Aberson[1], J. W. Bao[5], M. Bender[7], S. Benjamin[5], L. Bernardet[3,5], M. Biswas[10], B. Brown[10], J. Cangialosi[12], C. Davis[7], M. DeMaria[12], J. Doyle[13], R. Falvey[9], M. Fiorino[5], S. Forsythe-Newell[18], J. Franklin[12], T. Ghosh[6,] I. Ginis[24], S. Goldenberg[8], S. Gopalakrishnan[1], T. Hamill[5], R. Hodur[17,13], H. S. Kim[4], J. Knaff[11], T. Krishnamurti[6], P. Kucera[10], Y. Kwon[4], W. Lapenta[4], N. Lett[16], S. Lord[4], T. Marchok[7], D. Meléndez[14], M. Morin[7], J. Moskaitis[13], K. Musgrave[2], L. Nance[10,20], A. Reinecke[13], C. Reynolds[13], Brian Strahl[9], V. Tallapragada[4], H. Tolman[4], R. Torn[19], G. Vandenberghe[4], T. Vukicevic[12], X. Wang[23], Y. Weng[15], J. Whittaker[5], R. Yablonsky[24], D. A. Zelinsky[12], D-L Zhang[21], F. Zhang[15] , J. Zhang[22], X. Zhang[22].

May 2015

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

Table of Contents

Executive Summary 1

1. Introduction 3

2. The Hurricane Forecast Improvement Project 3

3. The HFIP Model Systems 6

   a. The Global Models 6
   b. The Regional Models 8
   c. Initialization and Data Assimilation Systems 9
   d. The HWRF Community Code Repository and User Support 10

4. Meeting the HFIP Goals 10

   a. The HFIP Baseline 10
   b. Meeting the Track Goals 12
   c. Reaching the Intensity Goals 12

5. HFIP Stream 1.5 13

   a. Stream 1.5 results 14

6. Performance of Stream 2 models 17

7. Operational Hurricane Guidance Improvements 19

   a. Global Model (GFS)—Operational 19
   b. Global Model Ensembles (GFS based)—HFIP Experimental 20
   c. Hurricane WRF (HWRF) 24
      1) Atlantic 24
      2) West Pacific (WPAC) and Other Basins 25
      3) A Comment on Intensity Forecast Errors 26
      4) The Multi-Model Ensemble 26
   d. GFDL Ensemble 28

8. Impact of Inner Core Reconnaissance Data 31

9. Data Assimilation and Physics Development 32

   a. Data assimilation 32
   b. Physics Development 35

10. Post Processing of Model Output 36

   a. Statistical Post Processing of Model Output: The FSU Multi-Model Ensemble 36
   b. The HFIP Web Page 38

11. Societal Impact Work 39

12. Future Configuration of a Numerical Model Hurricane Forecast Guidance System to meet the HFIP goals 42

References 45

Appendix A: NOAA Opportunity Announcement 47

Appendix B: Model Acronyms 48

List of Tables

Table 1. Future Numerical Model Hurricane Forecast Guidance System 2
Table 2. Strategic Teams 2014 4
Table 3. Tiger Teams 2014 5
Table 4. Specifications of the HFIP Global Models 7
Table 5. Specifics of the regional models used by HFIP in 2014 8
Table 6. Participating modeling groups in the 2014 HFIP Retrospective Evaluation. 13
Table 7. 2014 HFIP Stream 1.5 Real-time runs and model guidance to NHC 14
Table 8. Stream 2 Models 17
Table 9. Automated Tropical Cyclone Forecasting System (ATCF) ID Descriptions 29
Table 10. Numerical Model Hurricane Forecast Guidance System 44

List of Figures

Figure 1. HFIP Track and Intensity Error Baseline and Goals. 11
Figure 2. Tracking and Intensity Forecast Skill in the Atlantic Basin (2014) 15
Figure 3. Track and Intensity Forecast Skills for the East Pacific Basin 16
Figure 4. Track and Intensity Skills for the 2014 Atlantic Basin hurricane season 18
Figure 5. Hurricane track forecast skill errors for GFS, ECMWF, and FIM for the 2011-2013 Atlantic and East Pacific hurricane seasons 20
Figure 6. Operational GEFS (T254) and HFIP GFS Ensemble System (T574) Track and Intensity (Max Wind) Errors 21
Figure 7. EnKF HFIP GFS Ensemble T.S. Force Wind Probabilities 84 Hours for Hurricane Arthur 22
Figure 8. EnKF HFIP GFS Ensemble Systems Forecasts and Ellipses (July 1, 2014 Forecast) 22
Figure 9. Operational GEFS and HFIP Ensemble for Verification of Probability of Storm Force Winds for 2013 and 2014. 23
Figure 10. HWRF Intensity Forecast Improvement for the Atlantic Basin 24
Figure 11. 2014 HWRF Track and Intensity Error Statistics in the Western Pacific 25
Figure 12. OFCL Intensity Error Distribution 26
Figure 13. The HFIP Multi-Model Ensemble for Tropical Storm Gonzalo (2014) 27
Figure 14. Comparing Track & Intensity Errors between Multi-Model Ensemble Components. 28
Figure 15. Mean Track and Intensity Skills in the Atlantic Basin. 30
Figure 16. 2014 Track, Intensity and Bias comparisons with Inner Core Reconnaissance Data 31
Figure 17. Operational HWRF Workflow 32
Figure 18. Cycling HWRF GSI/EnKF Workflow 33
Figure 19. HWRF GSI/EnKF Method for Model Initialization (Hurricane Arthur in 2014) 34
Figure 20. 2014 HWRF Formulations Testing and Evaluations for C_p and C_h 35
Figure 21. 2014 Atlantic Hurricane Season Results using BLUE 96-h^+ Difference Weighting Scheme. 37
Figure 22. Example of the HFIP Webpage 39
Figure 23. Assessment of Potential Storm Surge Product from the NHC. 40
Figure 24. Assessment of how various groups rate the cone of uncertainty for the forecast hurricane track. 40
Figure 25. Onset of Storm Force Winds Product: 41
Figure 26. Example of the experimental global HWRF System in NMMB NEMS framework: 43
Figure 27. Track and Intensity Forecasts for HWRF Basin-Scale and the Operational HWRF. 44

Executive Summary

This report describes the activities and results of the Hurricane Forecast Improvement Project (HFIP) in 2014. It should be generally noted that 2014 was not a representative season for the Atlantic due to very low storm activity, not many rapid intensification (RI) events, and a relatively small sample size of total storms. Like last year, we focused on the improvements in the operational Global Forecast System (GFS) global model and Hurricane Weather Research and Forecasting (HWRF) regional model. HFIP is organized around the three "streams": 1) Stream 1; operational model development, 2) Stream 1.5; a group of experimental models evaluated by the National Hurricane Center (NHC) pre-season and then made available to NHC forecasters during their forecast cycle in real-time, and 3) Stream 2; HFIP experimental models which test and evaluate new techniques and strategies for model forecast guidance prior to testing 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. This report outlines HFIP, how it is organized, its goals, its models and the results from each of the three Streams.

Stream 1.0 Results and Accomplishments

- The 2014 HWRF continued to improve its intensity forecasts (Figure 10) but this year it was in the middle of the pack (Figure 2) for intensity performance with respect to other models, especially the statistical models.
- This year however, HWRF track forecasts were better than other models for the later lead times (Figure 2).
- However, model initialization is still an area needing additional improvements, particularly for initially strong tropical cyclones, where the focus is to obtain more accurate initial vortex structure and environment.
- The Geophysical Fluid Dynamics Laboratory (GFDL) model and the HWRF model are now being run operationally in all hurricane basins, i.e. North Atlantic (NATL), East Pacific (EPAC), the north and south West Pacific (WPAC), and Indian (north and south) oceans.
- The GFS continues to provide excellent track guidance superior to most other models and comparable to the European Centre for Medium-range Weather Forecasts model (ECMWF) guidance and exceeds the 5-year HFIP goal out to 4-day lead time.

Stream 1.5 Results and Accomplishments

- The HWRF was run as a 20 member ensemble and exhibited a 15%-20% improvement of intensity over the control run (operational version of the model).
- The GFDL ensemble was run as a 10 member ensemble. Performance was acceptable but the spread of the ensemble, particularly for track was poor.
- This year a multi-model regional ensemble HWRF, GFDL, and Coupled Ocean/Atmosphere Mesoscale Prediction System-Tropical Cyclone (COAMPS-TC) was inaugurated.
- The best performer of the three ensembles for intensity was the COAMPS-TC ensemble that even outperformed the multi-model ensemble mean.

- The other Stream 1.5 models are shown in Table 7. Performance shown in Figure 3 was mixed.
- A web page http://www.hfip.org/products/ to display HFIP Stream 1.5 and other HFIP sponsored/run models in real time have been developed and maintained for the past two years. New products were added and improvements to existing products were made (e.g., Figure 22).

Stream 2.0 Results and Accomplishments

- The HFIP version of GFS was again run in 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 National Centers for Environmental Prediction (NCEP) Global Ensemble Forecast System (GEFS). As with the previous year, the higher resolution system did not perform as well as the operational system. This disappointing result is likely because the physics and Data Assimilation (DA) packages were not optimized for the higher resolution and the semi-implicit scheme.
- Forecasts of the 35 knot wind probability for ATL, EPAC, and WPAC by the HFIP global GFS ensemble were excellent for days 3-4 (see Figure 9).
- The global version of the HWRF was tested for the first time this year, see Figure 26.
- Performance of the multi-storm basin scale HWRF (from which the global version was developed) showed improvement in intensity over each storm run individually in the current HWRF (not the upgraded operational version).
- DA using the HWRF ensemble to generate background errors is progressing (Figures 17 and 18).
- A new product on presentation of the onset of storm force winds is being developed at the NHC. A prototype is shown in Figure 25.

Future configuration of the Hurricane Forecast System

Based on five years of results from HFIP, the projected future operational hurricane forecast guidance system would be as described in Table 1 below.

Table 1. Future 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	Double nests (9 and 3km), one for each hurricane Using all available aircraft and satellite data in the inner 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, COAMPS-TC)
Statistical Post Processing	Logistics Growth Equation Model (LGEM), Statistical Hurricane Intensity Prediction System (SHIPS), Statistical Prediction of Intensity from a Consensus Ensemble (SPICE), and 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 program 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, and 50% in 10 years for days 1-5.
- Reduce average intensity errors by 20% in 5 years, and 50% in 10 years for days 1-5.
- Increase the probability of detection (POD)[1] for RI[2] to 90% at Day 1 decreasing linearly to 60% at day 5, and decrease the false alarm ratio (FAR) for RI 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 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 obtaining the commensurate resources required for robust real-time use in an operational environment.

In FY2014, HFIP funded approximately $11,000,000 of research and development (R&D) based on recommendations from 4 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

[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] 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 (RW) of a decrease of 25 knots in 24 hours.

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); and 4) 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 International University (FIU), 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), University of Utah (UT), and the University of Wisconsin (UW).

The teams, made up of over 50 members drawn from the hurricane research, development and operational communities, and team co-leaders are listed in Table 2. Tiger teams, focused on specific, shorter-term (1 or 2 years) tasks are listed in Table 3.
Full team membership is available at: http://www.hfip.org/teams/.

HFIP dedicated $2,300,000 to operating and maintaining the high performance computer system in Boulder, Colorado, and an additional $1,600,000 to enhancing computer capacity.

Table 2. Strategic Teams 2014

FY 2014 Teams	**FY 2014 Team Leads**
HFIP Model/Physics Strategy	Vijay Tallapragada (EMC), Jian-Wen Bao (ESRL)
Data Assimilation / Initialization / Ensemble Development	Jeff Whitaker (ESRL), Xuguang Wang (OU)
Post Processing and Verification Development	Mark DeMaria (NHC), David Zelinsky (NHC), Tim Marchok (GFDL)
Societal Impacts	Jennifer Sprague (NWS), Rick Knabb(NHC)

Table 3. Tiger Teams 2014

FY 2014 Teams	**FY2014 Team Leads**
Web Page Design	Paula McCaslin (ESRL), Laurie Carson (ESRL)
Regional Hybrid DA System / Use of Satellite Data	Jeff Whitaker (ESRL), Xiaolei Zou (FSU)
Stream 1.5 and Demonstration System Implementation	James Franklin (NHC), Barb Brown (NCAR)

HFIP's focus and long-term goal is to improve the numerical model guidance that is provided by the NCEP operations to NHC as part of the hurricane forecast process. To accomplish this goal, the project 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, 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 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 improvements of forecast performance. Because the level of computing necessary to perform such a demonstration is larger than can be accommodated by the current operational computing resources, HFIP has developed 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 after each season to identify techniques that appear particularly promising to operational forecasters and/or modelers. These potential advances can be blended into operational implementation plans through subsequent Stream 1 activities, or further developed outside of operations within Stream 2. Stream 2 models represent cutting-edge approaches that have little or no track record; and therefore are not used by NHC forecasters 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 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 – the peak three months of the hurricane season) 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 expedites the availability of real-time advances to forecasters by 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 10-15 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 (NOAA SAB, 2006). Maximizing improvements in hurricane intensity forecasts will therefore require high-resolution regional models or global models with moveable high-resolution nests, 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 forecasts. They also provide the 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 2014 by HFIP are listed in Table 4 along with their characteristics.

Table 4. 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 (ESRL) 5-member ensemble	27 km	64	From 2011 GFS - Simplified Arakawa Schubert	Zhao-Carr	GFS Non-Local PBL	Noah LSM	Rapid Radiative Transfer Model (RRTMG)	ESRL EnKF
FIM – 2014 numerics w/ 1/8 deg. output (ESRL)	15km	64	From 2011 GFS – Simplified Arakawa Schubert	Zhao-Carr	GFS Non-local PBL	Noah LSM	RRTMG	GFS GSI operational hybrid-ensemble variational
GFS/Hybrid EnKF-GSI (NCEP)	27 km	64	Simplified Arakawa Schubert	Zhao-Carr-Moorthy	GFS Non-Local PBL	Noah LSM	RRTMG	Hybrid EnKF-3D-VAR GSI
20-member Global Ensemble GEFS (NCEP)	60 km	42	Simplified Arakawa Schubert	Zhao-Carr-Moorthy	GFS Non-Local PBL	Noah LSM	RRTMG	Bred Vector ETR + STTP
20-member HFIP Global Ensemble GFS/EnKF (ESRL)	27 km	42	Simplified Arakawa Schubert	Zhao-Carr-Moorthy	GFS Non-Local PBL	Noah LSM	RRTMG	ESRL EnKF + New Stochastic Physics
NAVGEM	T359 37 km (T425 ~31 km in transition)	50 (60 in transition)	Simplified Arakawa Schubert	Modified Zhao-Carr	NAVGEM	NAVGEM LSM	RRTMG	NRL Atmospheric Variational DA System-accelerated representer (NAVDAS-AR)

b. The Regional Models

Specifications of regional models used by HFIP in 2014 are shown in Table 5.

Table 5. Specifics of the regional models used by HFIP in 2014

Models	Domains / Horizontal Resolution (km)	Vertical Levels core	Cumulus Parameterization	Microphysics	PBL	Land Surface	Radiation	Initial and Boundary Conditions	Initialization	SST
HWRF (OPS)	3 27/9/3 (9/3 following the storm)	61 NMM	Simplified Arakawa Schubert for 27/9 nests	Ferrier	GFS Non-Local PBL	GFDL Slab Model	GFDL Scheme	GFS	One-way hybrid GSI-EnKF with vortex initialization	3D POM
HWRF in non-NWS basins (WP/SH/SL/IO)	3 27/9/3 (9/3 following the storm)	43 NMM	Simplified Arakawa Schubert for 27/9 nests	Ferrier	GFS Non-Local PBL	GFDL Slab Model	GFDL Scheme	GFS	vortex initialization	GFS (static)
HWRF (Ens) 20 members	3 27/9/3 (9/3 following the storm)	43 NMM	Simplified Arakawa Schubert with Stochastic Perturbations for 27/9 nests	Ferrier	GFS Non-Local PBL with stochastic perturbation	GFDL Slab Model	GFDL Scheme	GEFS	vortex initialization with initial position perturbed	3D POM
GFDL (OPS)	3 55/18/6 (18/6 following the storm)	42 GFDL	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	Schwarzkopf-Fels (longwave) / Lacis-Hansen (shortwave)	GFS	GFDL Synthetic Bogus Vortex	3D POM
GFDL (WP, HFIP version)	3 55/18/6 (18/6 following the storm)	42 GFDL	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	Schwarzkopf-Fels (longwave) / Lacis-Hansen (shortwave)	GFS	GFDL Synthetic Bogus Vortex	3D POM
GFDL (Ens) 10 members	3 55/18/6 (18/6 following the storm)	42 GFDL	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	Schwarzkopf-Fels (longwave) / Lacis-Hansen (shortwave)	GFS.	GFDL Synthetic Bogus Vortex with inner core perturbations	3D POM
HYCOM-Coupled HWRF	3 27/9/3 (9/3 following the storm)	61 NMM	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	GFDL Scheme	GFS	One-way hybrid GSI-EnKF with vortex initialization	3D HYCOM
HWRF-HRD/EMC Basin Scale	3 27/9/3 (9/3 following each storm)	61 NMM	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	GFDL Scheme	GFS	Vortex initialization	GFS (static)
HWRF-HRD (HEDAS)	2 9/3 (3km following the storm)	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	GFS (static)
AHW (NCAR) 15-member ensembles	3 36/12/4	36 ARW	Tiedtke (36/12 km only)	WSM6	YSU	NOAH LSM	RRTMG (LW+SW)	GFS (BC only)	96-member DART EnKF method in a 6-hour cycling mode	Pollard 1-D Column Ocean
COAMPS-TC© (HFIP version)	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	Balanced vortex initialization (4D-VAR, EnKF options)	GFS (static)

Models	Domains / Horizontal Resolution (km)	Vertical Levels core	Cumulus Parameterization	Microphysics	PBL	Land Surface	Radiation	Initial and Boundary Conditions	Initialization	SST
COAMPS-TC© (OPS)	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	NAVGEM	Balanced vortex initialization (4D-VAR, EnKF options)	NAVGGEM (static)
COAMPS-TC© (Ensemble) 10-members	3 27/9/3 (9/3 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 with initial and boundary condition synoptic-scale perturbations	Balanced vortex initialization with perturbations	GFS (static)
Wisconsin NMS	2 45/4,1 (4/1 following the storm)	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	GFS (static)
Penn State ARW	3 27/9/3 (9/3 following the storm)	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	GFS (static)

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. Grid-point Statistical Interpolation (GSI): The GSI system developed by NCEP is a unified 3-dimensional variational (3D-VAR) data assimilation system for both global and regional applications, and has been widely used by many modeling systems across NOAA and other agencies (DTC 2012; Wu, et al, 2002; Parrish and Derber 1992; Cohn and Parrish, 1991). This system has been used by NCEP GFS since 2006 in operations.

2. 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; Houtekamer et al, 1998; Evensen, 1994;). Several HFIP models (e.g., AHW, HFIP GFS ensembles, PSU etc., see Tables 4 & 5 above) are using the EnKF approach for DA. The Hurricane Research Division (HRD)/AOML developed a variant of the EnKF based DA system using the HWRF model, known as Hurricane Ensemble Data Assimilation System (HEDAS) as noted by Aksoy et al, (2012).

3. Vortex initialization: The initial vortex for the regional models is produced by a vortex initialization procedure. In general, 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. An advanced vortex initialization and assimilation cycle for the operational HWRF consists 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 EnKF. This is a version of the Hybrid Ensemble-Variational DA System (HEVDAS) where instead of using an ensemble from the HWRF to create the ensemble part of the background error, it uses ensemble members created by the GFS EnKF. The DA system is capable of ingesting inner core data to optimize the vortex initialization.

4. Global Forecast System (GFS): The initial state created for the current operational global model, the GFS, is interpolated to the grids used by HFIP global models. The GFS in 2012 used the HEVDAS. HEVDAS is a combination of the GSI 3D-VAR and an ensemble-based system to define the background error matrix.

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

d. The HWRF Community Code Repository and User Support

During 2009-2014, 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 3.6a (Bernardet et al, 2014; Tallapragada et al, 2014). 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, and has been supported by HFIP for developing a repository for a community based hurricane modeling system. A community based hurricane modeling system ensures the same code base can be used for research and in operations. Support provided by DTC in 2014 included two HWRF tutorials, taught in College Park, MD, and in Taiwan.

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 and the 5- and 10-year error goals represented in black and labeled on the left side of the graph. A separate set of baseline errors (not shown) was computed for the eastern North Pacific basin.

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 more representative 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's important to remember, however, that normalization by CLIPER or (especially) Decay-SHIFOR5 can fail to adequately account for forecast difficulty in some circumstances. A hurricane season that features extremely hostile environmental conditions will lead to very high Decay-SHIFOR intensity forecast errors (because climatology will be a poor forecast in such years), but relatively low dynamical model and NHC official forecast errors (because few storms will intensify rapidly, making life easy on both models and forecasters). This combination of baseline and model errors yields an unrealistic skill estimate.

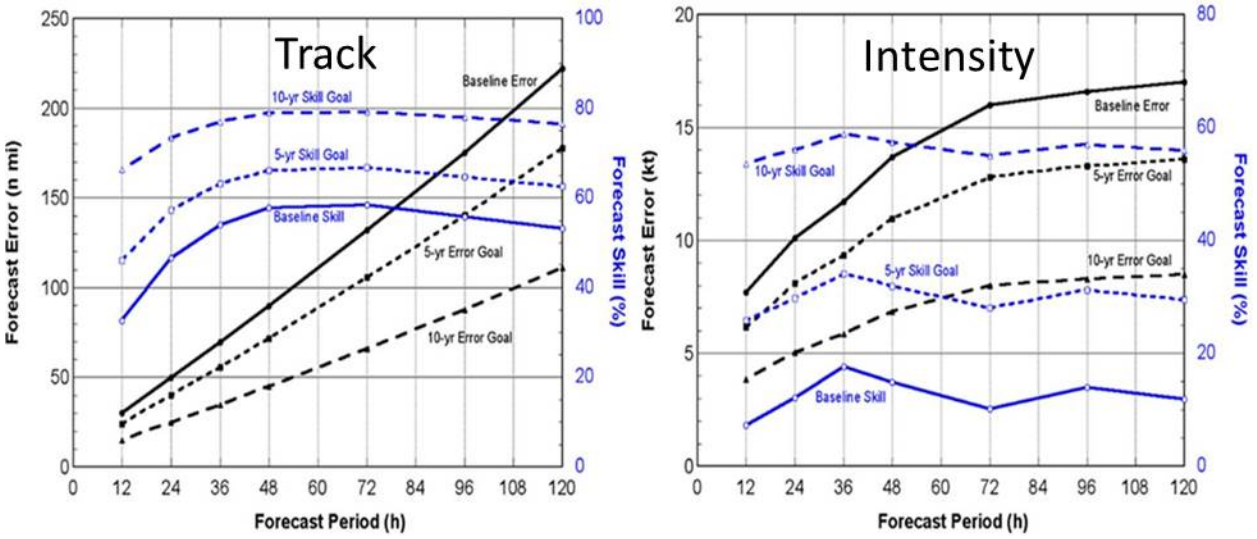


Figure 1. HFIP Track and Intensity Error Baseline and Goals.
The baseline errors (black lines) were determined from an average of the top-flight operational models during the period 2006-2008. The HFIP expressed goals (dashed lines) are to reduce this error by 20% in 5 years and 50% within 10 years. 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).

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 model 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. The procedure also adjusts the forecast position and intensity at some of the forecast times as well and then applies a smoother to the adjusted forecast. This adjustment, called "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. It is interesting to note however that although the early version loses about 3-5% of the skill for track forecasts compared to the late version, the skill for intensity forecasts are virtually the same for late and early versions (for details see Goldenberg et al, 2015).

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 forecasts were improved, particularly in the tropics, by using a more advanced DA scheme than the one employed operationally at that. A version of this advanced DA 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 or global models that have moveable nests with a horizontal resolution near the hurricane's inner 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. The operational HWRF model operating at 3km resolution has started showing its potential for improved intensity forecasts, producing comparable and sometimes superior results versus the statistical models and NHC official forecasts as demonstrated through a large set of retrospective forecasts. Results from HWRF model for intensity forecasts are presented in this report.

Eventually the regional model will be able to interact within the global model. More specifically, there would be one set of nests for each hurricane in the global model, thereby accomplishing the track and intensity forecast goals through a unified global-to-regional scale modeling system.

5. HFIP Stream 1.5

HFIP and NHC agreed in 2009 to establish a pathway to operations known as "Stream 1.5." As explained previously, Stream 1.5 covers improved models and/or techniques that 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 expedite by one or more hurricane seasons the delivery of real-time guidance for a particular model to NHC. 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 2014 season. Those groups and their models are listed in Table 6. Table 7 shows the groups selected for Stream 1.5 and the form of the output from their models provided to NHC.

Note that 4 of the 9 candidate models were admitted into Stream 1.5 based on the models' performance for track and/or intensity. In addition, two HFIP Stream 1.5 consensus aids were constructed: (1) the *TV15 Track Consensus Model* used operational model members: GFSI, EGRI, GHMI, HWFI, GFNI, EMXI and the GFDL Stream 1.5 model, while (2) the *IV15 Intensity Consensus Model* used of operational model members: Decay-SHIP (DSHP), LGEM, GHMI, HWFI and the Stream 1.5 NRL, UW-Madison, and GFDL.

Table 6. Participating modeling groups in the 2014 HFIP Retrospective Evaluation.

Organization	Model	Type
EMC	HWRF coupled with HYCOM	Regional-dynamic-deterministic
EMC	20-member HWRF Ensemble	Regional-dynamic-ensemble
NRL	COAMPS-TC with GFS IC/BC	Regional-dynamic-deterministic
UW-Madison	UW-NMS	Regional-dynamic-deterministic
AOML/HRD	Basinscale HWRF	Regional-dynamic-deterministic
GFDL	10-member GFDL hurricane model Ensemble	Regional-dynamic-ensemble
ESRL	FIM	Global-dynamic-deterministic
FSU	Multi-model superensemble	Weighted-consensus
NESDIS/STAR and CIRA	SPICE	Statistical-dynamical-consensus

Table 7. 2014 HFIP Stream 1.5 Real-time runs and model guidance to NHC

Organization	Model	Track	Track consensus	Intensity	Intensity Consensus
NRL	COAMPS-TC			•	•
UW-Madison	UW-NMS				•
GFDL	10-member GFDL hurricane model ensembles	•	•	•	•
NESDIS/STAR and CIRA	SPICE			•	

All evaluations were based on homogeneous samples and applied the appropriate method for assessing statistical significance. A detailed description of the methodgy used for the evaluation, reports for each participating modeling group (Tables 6 and 7). All the verification plots generated during the evaluation, and information on the participating models and cases included in the evaluation are available on the TCMT 2014 Retrospective HFIP Testing website (http://www.ral.ucar.edu/projects/hfip/d2014).

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 2014. The Stream 1.5 models arriving at NHC underwent standard interpolation processing to convert "late" dynamical guidance into "early" interpolated guidance that could be used by the forecasters, (see Section 4a). Figure 2 (upper left panel) portrays the skill plots for a homogeneous verification of the Stream 1.5 track models that met availability standards (regardless of whether they were intended for use explicitly or in a consensus), along with selected operational models.

Early in the forecast period the track consensus (TVCA) was (marginally) the best model. The GFSI (operational GFS model) was close to the official, and similar to the European model (EMXI) for all times except 120 h when GFSI had much higher skill. So overall, TVCA and HWFI showed the highest skill for track. The GFDL models [green lines: GFDM - GFDL ensemble mean and GHMI (operational GFDL)] had the lowest overall skill for track.

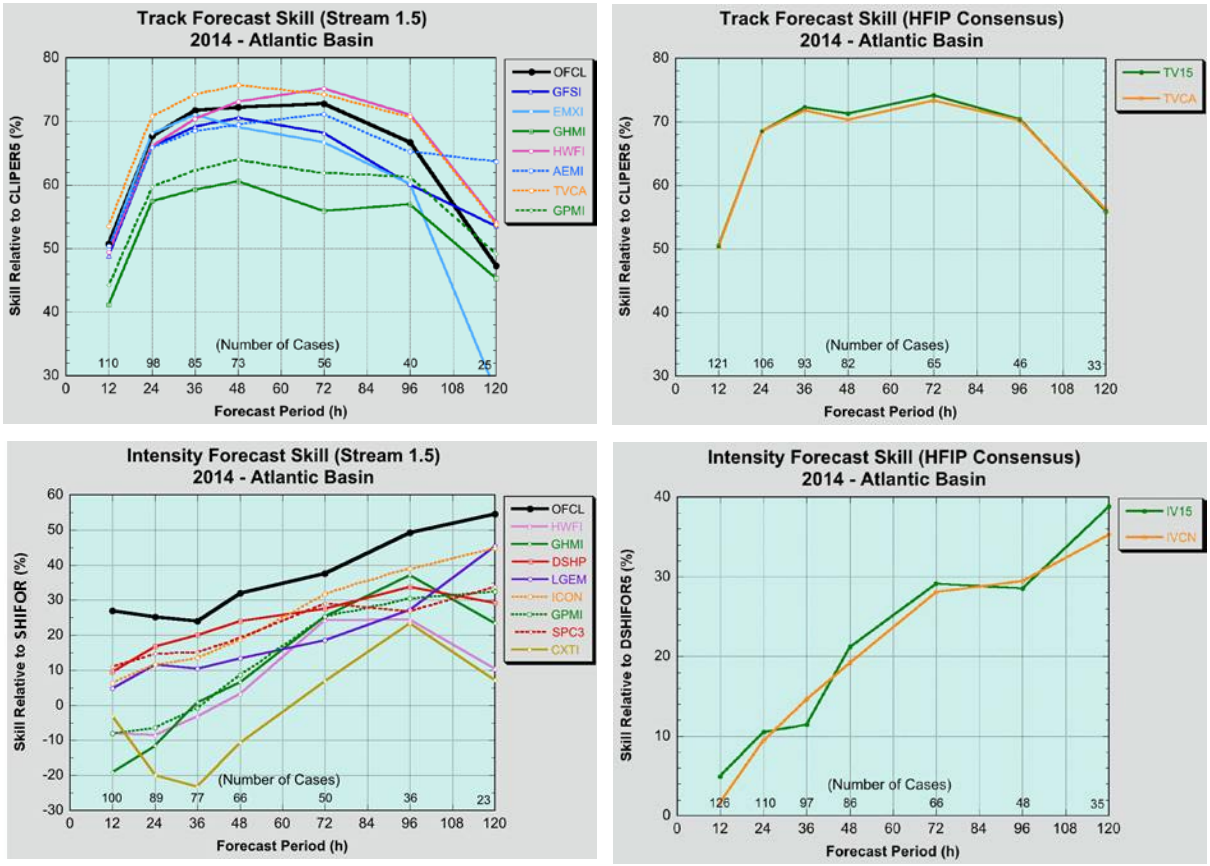


Figure 2. Tracking and Intensity Forecast Skill in the Atlantic Basin (2014)

It should be noted that the results discussed here are for only one (low activity) year. This relatively small sample size (only 110 and 25 cases at 12 and 120 h, respectively) is not necessarily representative of results that would be seen from a multi-year sample.
Figure 3 shows that in 2014 the HWFI was competitive with the top-tier operational models and in fact exceeded the official intensity skill after the first 48 hours and was more skillful in track forecasts than even GFSI from 24 to 36h and after 72 h.

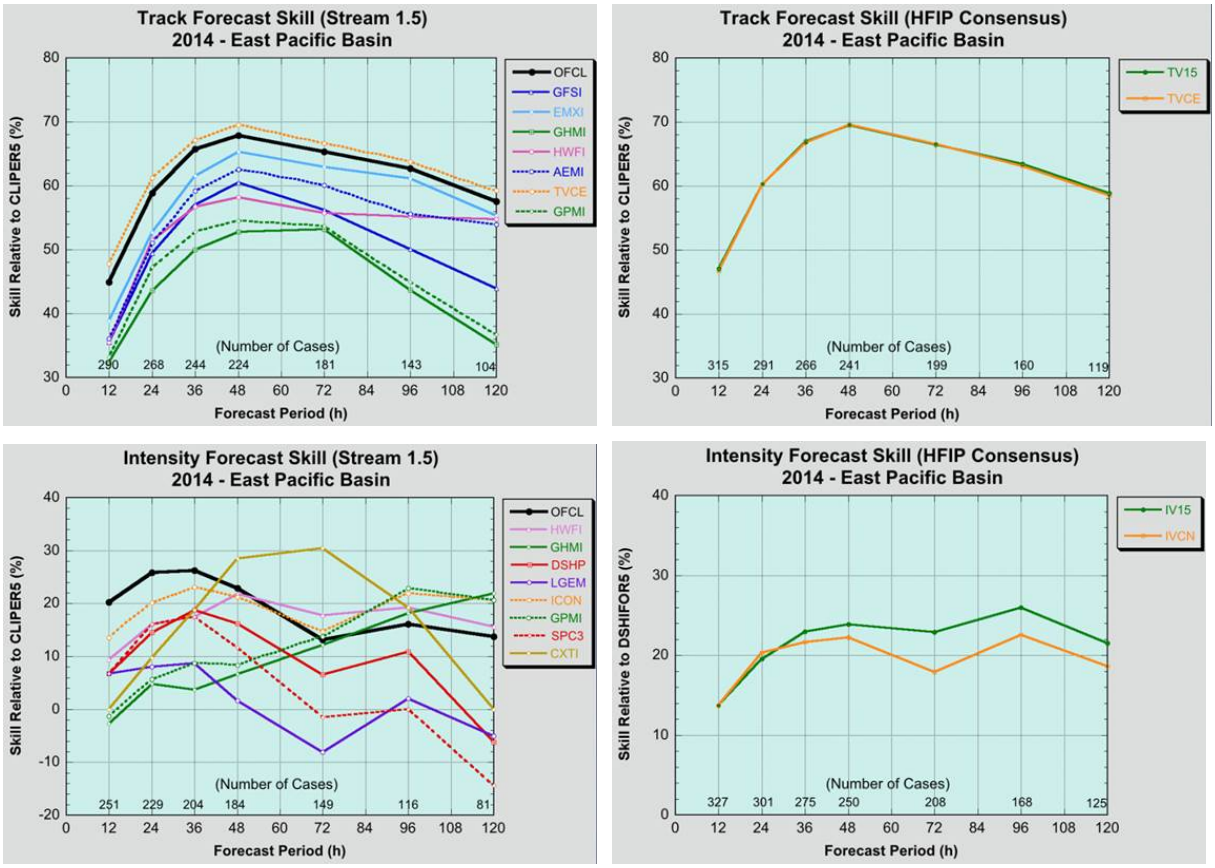


Figure 3. Track and Intensity Forecast Skills for the East Pacific Basin
As in Fig. 2 track and intensity skills are shown in upper and lower panels, respectively. TV15 and TVCE depicted in the upper right panel are Stream 1.5 and operational track consensus respectively, while TV15 and IV15 in the lower right panel are the corresponding consensus with the Stream 1.5 models included.

The intensity results in Figure 3 show that the best performers in the early part of the forecast period were the intensity consensus (ICON) and two statistical models (SPC3 and DSHPS). In general, the LGEM statistical model did not perform as well as the other models (except at 120 hours where there are fewer cases). Although the 2014 HWFI had shown relatively high skill for intensity forecasts in retrospective runs, it did not perform as well for the 2014 sample. HWRF and the Navy model (CXTI) were among the lowest performers. In the upper right panel the consensus with and without the Stream 1.5 models (TV15 and TVCE, respectively) are compared for the East Pacific Basin. The addition of the Stream 1.5 models to the consensus produced almost no change for track forecast skill. However, when comparing the consensus for intensity with and without the Stream 1.5 models (IV15 and IVCN, respectively), the addition of the Stream 1.5 models resulted in increased skill at all times past 24 h (lower right panel).

6. Performance of Stream 2 models

Table 8. Stream 2 Models

Organization	Model Identifier
GSD	FIM9I
University of Utah	A3UI
PSU	APSI
GFDL Ensemble Mean	GTMI
HRD	H3WI, HECI, HEDI
EMC Ensemble Mean	HWMI
FSU	MMSI

In addition to the Stream 1.5 models run as part of the real-time experimental forecast system, HFIP also ran several models as Stream 2 candidates. Table 8 shows the various groups that ran Stream 2 candidates with the identifier for the various models run by each group shown in the right hand column.

The results from track and intensity skill for Stream 2 models for the 2014 Atlantic hurricane season are shown in Figure 4. Note that the sample size is even smaller than for Figure 2 since the Stream 2 models were run for fewer cases. The results are similar in perfomance patterns to those shown for the Stream 1.5 models as well as the operational models. For track, from 72-96 h, HWFI had the highest skill.

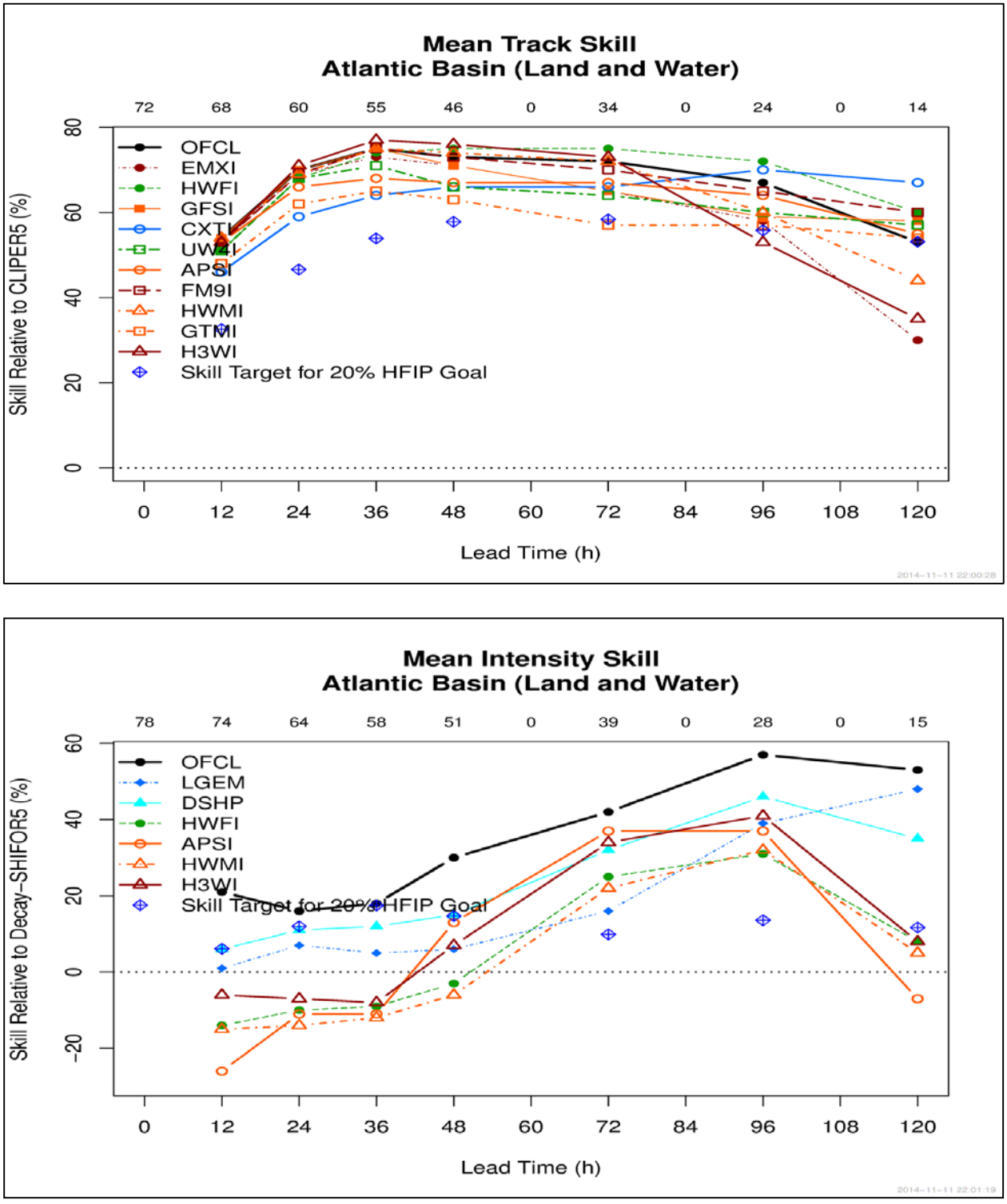


Figure 4. Track and Intensity Skills for the 2014 Atlantic Basin hurricane season
Forecast skill of Stream 2 models compared to operational models. As in Fig. 3 track and intensity skills are shown in the upper and lower panels, respectively. The number of cases is shown across the top of each panel.

For intensity DSHIPS (statistical) is still the best model overall, but at mid-forecast lead times dynamical models from HRD (H3WI) and Pennsylvania State University's APSI model were comparable. However, for intensity forecasts the Official forecast still shows the highest skill.

Several of the dynamical models show problems for intensity for the earliest lead times due to vortex initialization (spin up) issues.

7. Operational Hurricane Guidance Improvements

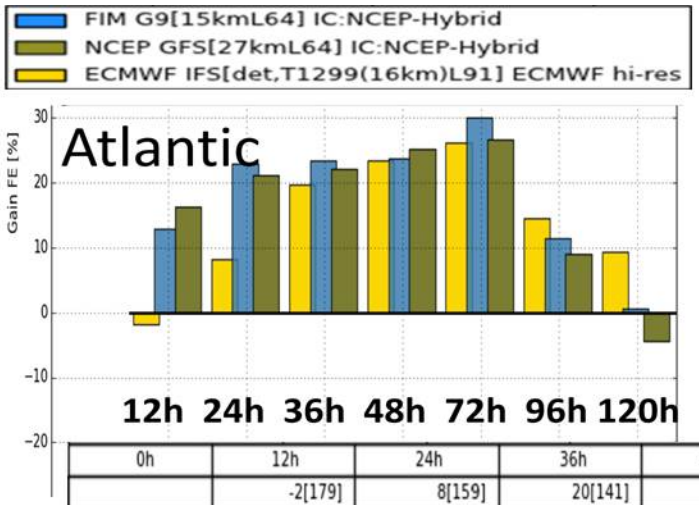
The HFIP goals described in section 4 are considered met when the model guidance provided to NHC by NCEP reaches those goals. Since 2014 represents the fifth year of the project, it is expected to see progress toward meeting the five year goals in both operational models and experimental models (e.g., Stream 1.5 models described in the previous section that provide model guidance to NHC in real time). In this section the emphasis will be on improvements in the hurricane forecasts from the models that were fully operational in 2014. This includes the GFS, and the HWRF operational regional models.

a. Global Model (GFS)—Operational

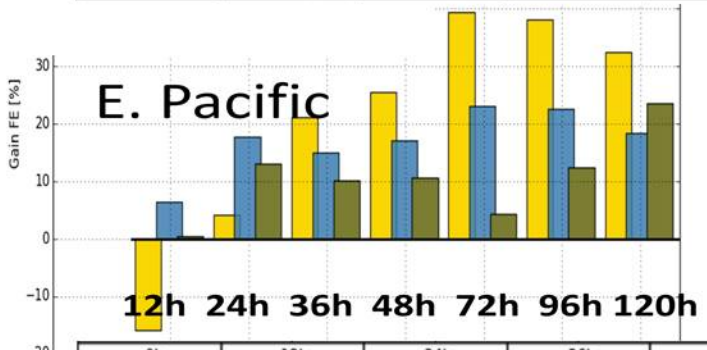
In May of 2012 the GSI data assimilation system in the GFS was replaced by the hybrid data assimilation system. The hybrid system uses an ensemble to generate a flow dependent background error covariance matrix that is then used in the GSI for the analysis. In previous annual reports starting with the first one in 2010 the impact of changing the DA system in the global models was described, particularly the GFS from the 3D-VAR GSI to an ensemble based system, called (EnKF). The hybrid system is basically a combination of the EnKF and the GSI which has shown to provide somewhat better results than EnKF alone. The global hybrid system merge with regional models is considered an important mechanism for the transference of HFIP results into operations. In addition, the GFS has undergone other improvements including improved physics and increased resolution. In 2014 the deterministic operational GFS was run at T574 (~27 km) and the GFS ensemble (GEFS) at T254 (~60 km). Recently (January 2015) the resolution of the GFS has been increased to T1534 (~13 km) but it is too early to show performance results for that configuration.

Using retrospective testing, Figure 5 shows retrospective testing results of the skill for the 2011-2013 Atlantic and East Pacific hurricane seasons of the 2012 GFS operational version as compared to the ECMWF and the FIM models which was run as a Stream 2 model in 2014. The upper panel is similar to a figure shown in the previous annual report but contains one additonal year. The 3-year results are similar to the 2-year results in that the GFS compares favorably to the higher resolution ECMWF model in the Atlantic. The FIM model using the same DA system as the GFS and the same physics performs comparably in the Atlantic to the GFS or ECMWF from 36 to 96 h.

In the East Pacific, FIM has higher skill than GFS for most forecast times. However, starting at 36 h, ECMWF has substantially higher skill than both the GFS and FIM. It will be interesting to see how the higher resolution GFS (T1534) will perform in the upcoming hurricane season for track forecasts.



0h	12h	24h	36h	48h	72h	96h	120h
	-2[179]	8[159]	20[141]	23[122]	26[91]	14[74]	9[56]
	13[179]	23[159]	23[141]	24[122]	30[91]	11[74]	1[56]
	16[179]	21[159]	22[141]	25[122]	27[91]	9[74]	-4[56]



0h	12h	24h	36h	48h	72h	96h	120h
	-16[111]	4[98]	21[83]	25[71]	39[51]	38[34]	32[19]
	6[111]	18[98]	15[83]	17[71]	23[51]	23[34]	18[19]
	0[111]	13[98]	10[83]	11[71]	4[51]	12[34]	23[19]

Figure 5. Hurricane track forecast skill errors for GFS, ECMWF, and FIM for the 2011-2013 Atlantic and East Pacific hurricane seasons
The FIM is an HFIP Stream 2 model. The resolutions used in each model are shown at the top. The number of cases for each time period and each model are shown in parentheses. The percentage gain is relative to the baseline.

b. Global Model Ensembles (GFS based)—HFIP Experimental

The operational NCEP Global Ensemble Forecast System (GEFS) is run at T254, and HFIP (and ESRL) high-resolution version (T574) GFS ensemble system, was run as part of the Stream 2 during the 2014 hurricane season. In this section, which is almost identical to the same section in the 2013 Annual Report except for updated figures, emphasis is placed upon ensemble systems comparison. Both GEFS and fine-grid HFIP ensemble systems use the same physics package and hybrid DA system that the operational GFS employs. But, while GEFS used the bred-vector ensemble transform with rescaling (BV-ETR) and stochastic total tendency perturbation (STTP) method for generating the perturbations, the HFIP system (semi-Lagrangian) used the EnKF based perturbations and a new stochastic physics package in its configuration.

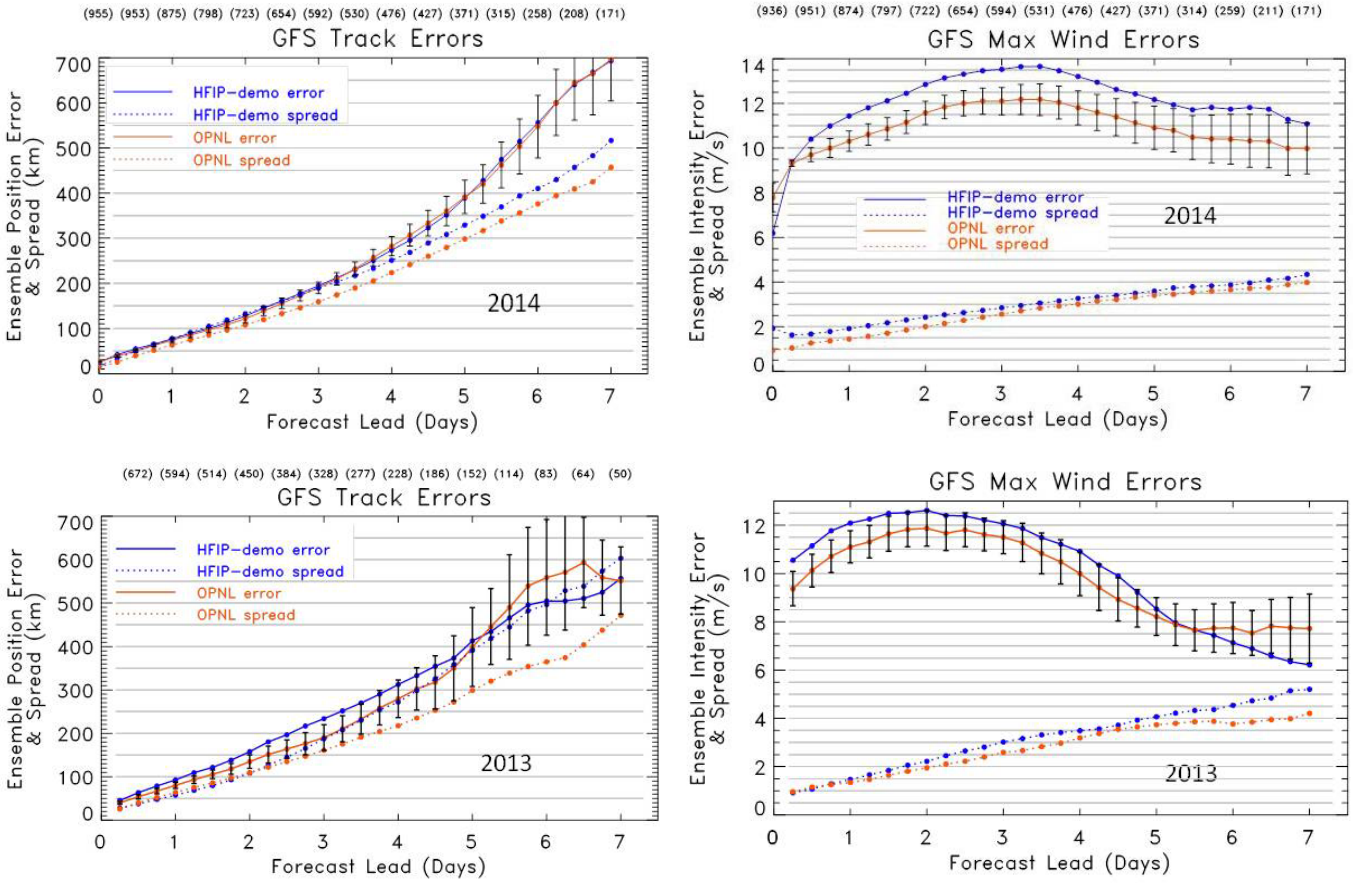


Figure 6. Operational GEFS (T254) and HFIP GFS Ensemble System (T574) Track and Intensity (Max Wind) Errors
For the 2013 and 2014 Atlantic hurricane seasons: Track (left panels) and intensity (right panels) error for the HFIP GFS Ensemble System (T574; HFIP demo; blue lines and the operational GEFS (T254; OPNL; red lines for 2014 (top panels) and 2013 (bottom panels). The HFIP ensemble systems use semi-Lagrangian differencing. The dashed lines show the ensemble spread and the number of cases are shown in parentheses above each panel. Error bars are shown for XX% confidence interval. The samples are homogeneous particular to a given year.

Figure 6 compares the track and intensity errors for the 2013 and 2014 Atlantic hurricane seasons. The results from the two years are similar. Both GEFS and the HFIP system had almost identical average track errors this year (2014) whereas GEFS performed better in 2013 for most lead times than the HFIP system. The spread for the HFIP system is still a little better than for GEFS. The GFS model is not prized for its intensity forecasts, but it is noted here that GEFS has slightly lower average errors than the HFIP ensemble and the ensemble spread of the two systems is rather low but similar.

The poorer performance of the HFIP versus GEFS for intensity is, as last year, related to incomplete tuning of the physics package to the semi-Lagrangian time differencing used in the HFIP system. At the current time, the HFIP developers are not emphasizing the development of global model techniques and 2014 is probably the last time we will run this global system on the HFIP computers.

Figures 7 and 8 show examples of products produced by the HFIP GFS ensemble system. Figure 7 shows the probability of winds greater than 34 knots (at least tropical storm force).

Each of the colored ellipses in Figure 8 contain 80% of the member's positions for each lead time, giving an idea of the probability of the storm position falling within the ellipse. Ellipses with their major axis aligned along the ensemble mean track indicate a speed uncertainty and those aligned perpendicular to the mean track indicate a directional uncertainty. Those aligned at some other angle with respect to the track indicate a speed-bias dependent on tracks to the left of the mean versus those aligned to the right. Both of these products were widely used by the HFIP team during the hurricane season.

Figure 9 shows the verification of Figure 7, probability of winds greater than 34 knots calculated from the HFIP ensemble system. A similar figure was shown in the last year's Annual report for the 2013 season. Here the results are shown from both years. For the 3-4 day forecasts the HFIP ensemble performed with significant improvement in 2014 compared to 2013 but GEFS was worse. The bottom panels show the 5-7 day forecasts. The HFIP ensemble did not perform as well in 2013 while GEFS was unchanged. Both models over forecast the probability of 34 knot winds at 5-7 days.

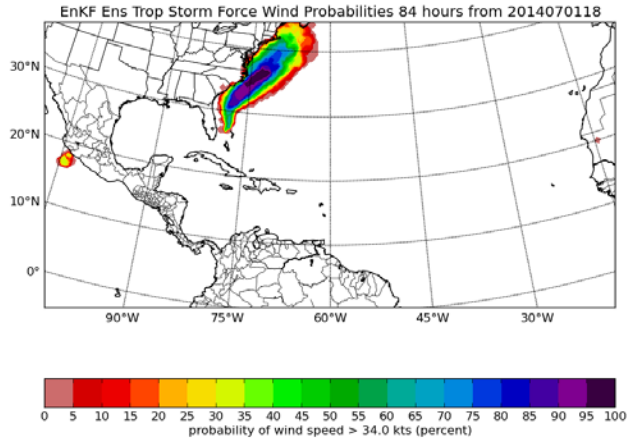


Figure 7. EnKF HFIP GFS Ensemble T.S. Force Wind Probabilities 84 Hours for Hurricane Arthur
This provides an example of a probability of winds greater than 34 knots at the various points during a 3.5 day period starting at 18 Z July 1, 2014 for hurricane Arthur. A 100% probability implies that all 20 members of the ensemble forecast a 34 knot wind at some time during the 84 hours of the forecast. Forecast is from July 1 at 18Z 2014.

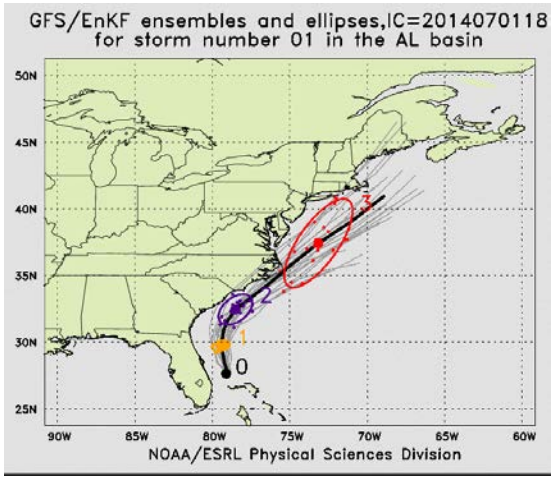


Figure 8. EnKF HFIP GFS Ensemble Systems Forecasts and Ellipses (July 1, 2014 Forecast)
This product created from the HFIP GFS ensemble shows all the tracks from the forecast of July 1, 2014, 18 Z out to 3.5 days for hurricane Arthur. There are twenty ensemble members shown in light gray and the ensemble 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

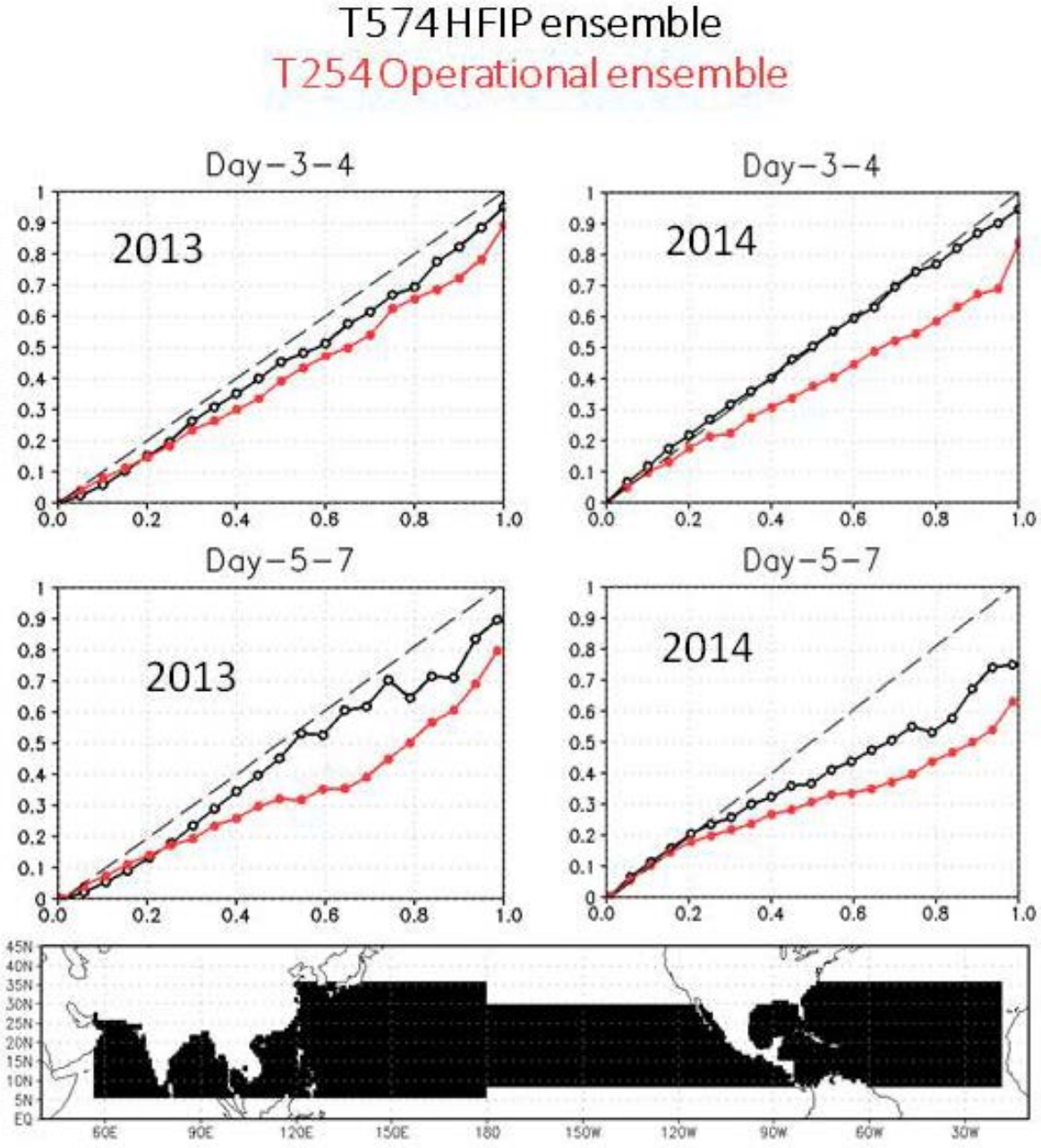


Figure 9. Operational GEFS and HFIP Ensemble for Verification of Probability of Storm Force Winds for 2013 and 2014.

Verification statistics for probability of storm force winds (greater than 34 knots) at points shown in the black area in the lower panel. HFIP GFS (T574) ensemble system and the operational GEFS (T256) are shown in black and red, respectively. 100% (i.e., 1) in the figures would imply that all ensemble members forecast a 34 knot wind at that point during days 3-4 (upper panel) and days 5-7 (lower panel). Probabilities are computed on a 1x1-degree box. Reliability Scores are aggregated over 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.

c. Hurricane WRF (HWRF)

1) Atlantic

Figure 10 reflects the story of where HFIP is with the HWRF hurricane model system. In earlier figures (e.g., Figure 2) it was shown that the most recent version of the HWRF (2014) model was performing very well for track. In fact, in 2014 HWRF provided among the best (if not the best) track guidance beyond 72 hours.

However, intensity is a different story. Again earlier figures (e.g., Figure 2) showed HWRF to be near the middle (or bottom) of the group of models shown in that figure. However, that was for a relatively small (2014 only) sample. The best performers (as is the case most years) were the statistical models, specifically DSHP. In Figure 10 the progress of HWRF in forecasting intensity is shown as error. There has been a steady decrease of intensity error from 2011 to present by 15% to 20% per year; although some of the samples (years over which models were run) shown in the figure are not homogeneous. In fact, in terms of error related to the five-year goal the 2014 version of HWRF has met or exceeded that goal beyond 72 hours. In contrast, there has much work to be done in improving the HWRF intensity forecasts for the first 24 hours; although the 2013 and 2014 operational versions are at least better than the HFIP baseline error in that forecast range.

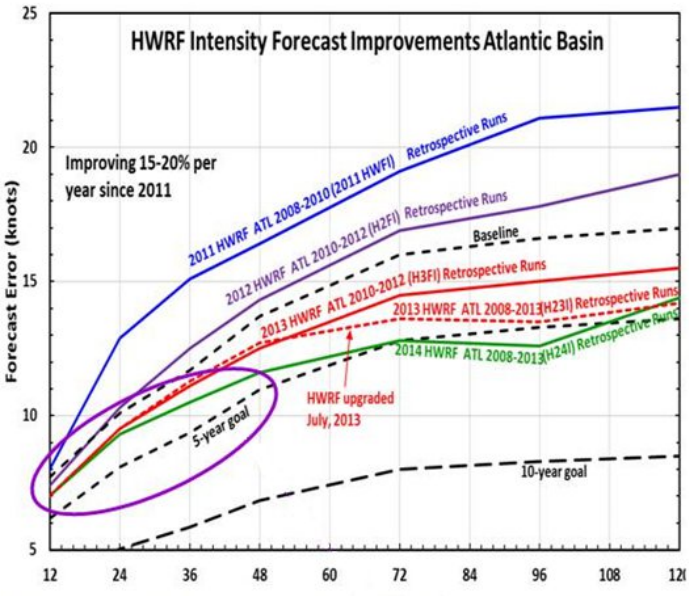


Figure 10. HWRF Intensity Forecast Improvement for the Atlantic Basin
Improvements in the HWRF system from the 2011 to the 2014 are shown as error. Seasons for which models were run are shown on each line. Note that some the samples (years) are not homogeneous between the models.

When the intensity errors for HWRF are shown relative to the Statistical model DSHP the steady improvement is still evident but even in 2014 (the red line in the skill figure) the skill is less than the HFIP 5-year goal. Most striking feature in this figure is the consistently very low skill early in the forecast (purple ellipse). This is related to the ongoing problems with initialization of the models. This problem is not unique to HWRF, see Figure 4. All the dynamical models seem to suffer from a common initialization problem. The initialization of the dynamical models continues to be a high priority for HFIP. Keep in mind however, that statistical models as well as the "early" (aka interpolated) versions of dynamical models', show better results in the first 24 hours since the initial times are adjusted to the operational values. The results shown here are for the "late" (aka non-interpolated) versions and therefore the forecast errors at early lead times are generally larger than with the "early" versions.

2) West Pacific (WPAC) and Other Basins

In 2012 HFIP began running real-time forecasts for the WPAC and in 2013 for the Indian Ocean. In 2014 HWRF runs were also extended into the southern Pacific and Indian Oceans. These runs were done in real time on the HFIP computers in Boulder, CO rather than the operational computers. The forecasts were transmitted to the Joint Typhoon Warning Center (JTWC) where they were used extensively in their forecasts. Figure 11 shows the performance for 2013-2014 for WPAC storms of HWRF compared to various other models. In this basin HWRF track forecasts (red) were as good as the GFS forecasts (black) although the ECMWF (yellow) track forecasts were even better. Traditionally, the ECMWF has provided the best forecasts of track in the WPAC.

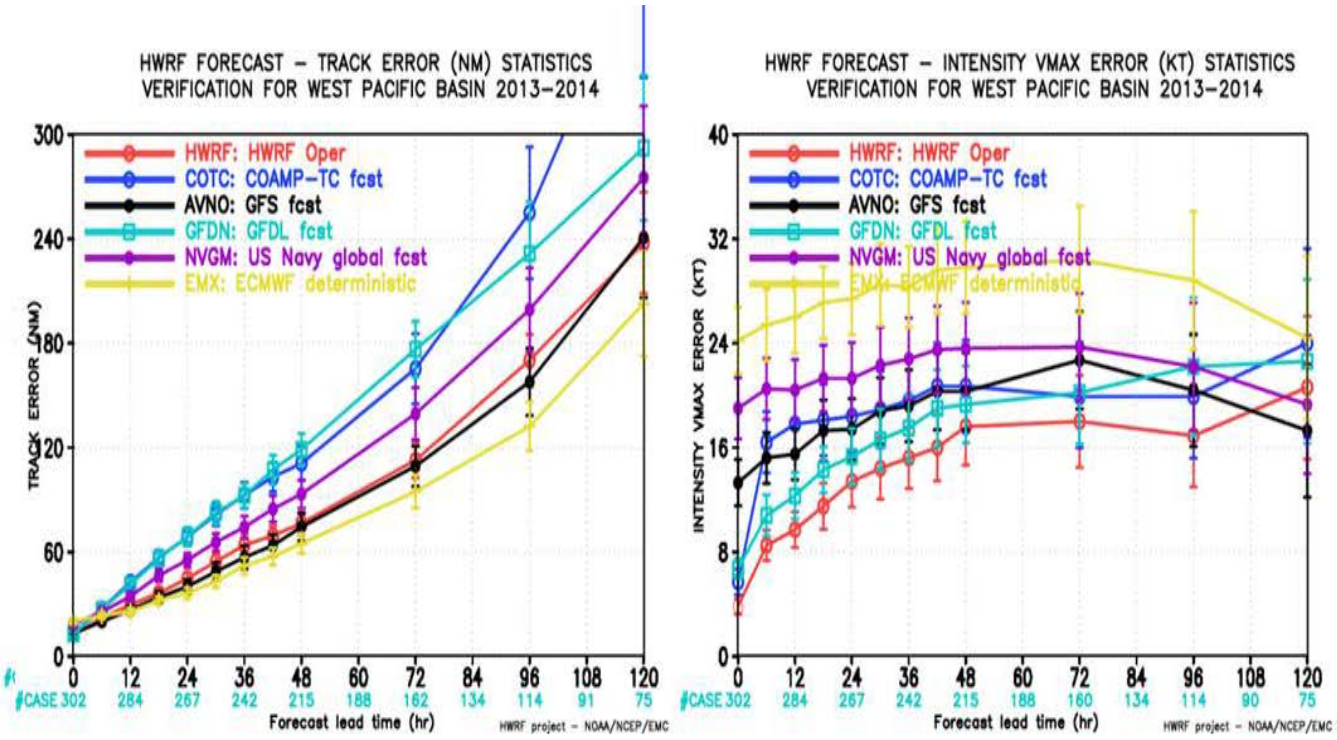


Figure 11. 2014 HWRF Track and Intensity Error Statistics in the Western Pacific
Track and intensity errors in the WPAC for 2013-2014 compared to several other models run in the WPAC.

For the first several years the intensity performance of HWRF in the WPAC has been excellent. Unlike the results for the Atlantic, HWRF provided the best (lowest) intensity forecasts at all forecast times except for 120 h. It is possible that the explanation for the different performance between HWRF in the WPAC and in the Atlantic is frequent occurrence of strong typhoons in the WPAC. Earlier studies show that HWRF tends to do better with initially stronger storms, which perhaps allow the HWRF model to demonstrate the importance and usefulness of high-resolution deterministic models for intensity forecasts in this basin.

3) A Comment on Intensity Forecast Errors

Within HFIP it has been discussed if there is a limit to how much intensity errors can be reduced. Some of the error reduction proposed by the program, especially the 50% 10-year goal may be unrealistic because the errors in the observations (~10 percent or 5-10 knots, see Figure 12) may be larger than the proposed reduction (to errors less than 10 knots) especially in the early forecast times (see Figure 1). Figure 12 shows the cumulative distribution of errors in the NHC official forecasts. The average error over the whole of the distribution is about 14 knots for 48 hours, mainly because of the relatively high number of forecasts with very large errors (up to 70 knots), and 65% of all errors at 48 hours being greater than 10 knots. It is hypothesized that it is possible to make substantial reduction in average error by reducing the magnitude of the extreme errors. Thus a focus on reducing the large errors (termed as outliers) may reduce the average errors to close to the observation error.

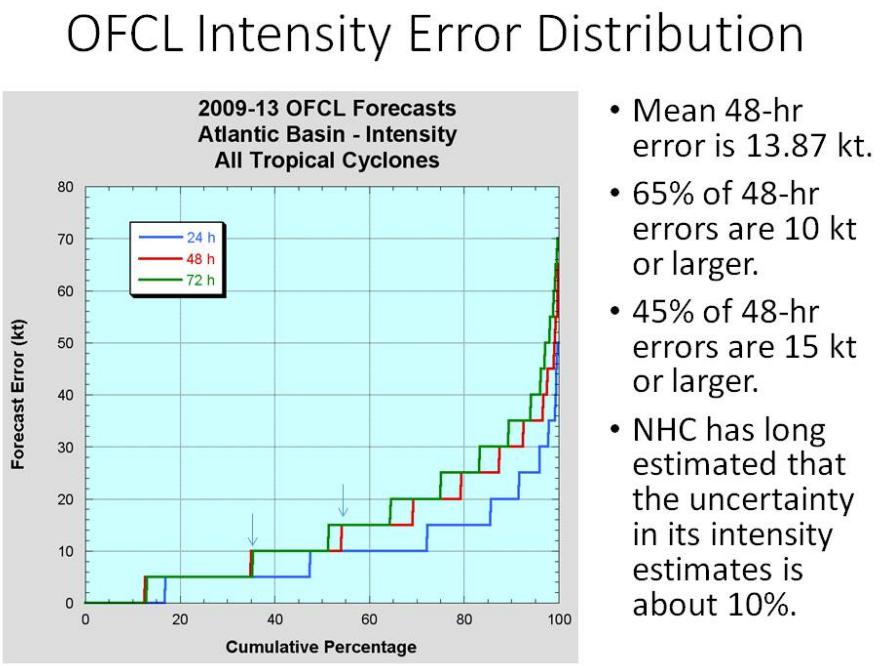


Figure 12. OFCL Intensity Error Distribution
Cumulative percentage of NHC Official Forecast errors are depicted as a function of the magnitude of the error. Forecast lead times of 24, 48 and 72 hours are shown.

4) The Multi-Model Ensemble

This year HFIP began testing a multi-model regional ensemble. Three ensembles were used, the HWRF, COAMPC-TC and the GFDL Ensembles (which will be described in more detail below.

The HWRF ensemble was the same system used for the HWRF ensemble last year with a few additional physics perturbations:

- 27-, 9-, 3-km horizontal grid spacing
- 20 members plus 1 control, the operational HWRF
- IC/BC Perturbations were from the GEFS- and Ensemble Transform with Rescaling (ETR) system was used to create the ensemble members:

- Stochastic boundary layer height perturbations in PBL scheme, -20% to +20%
- Stochastic initial wind speed perturbations with zero mean and -3kts to +3kts

- Model Physics Perturbations (vortex scale):
  - Stochastic Convective Trigger in SAS, -50hPa to + 50hPa white noise

Thus the large-scale perturbations of the ensemble came from the GEFS to initially define each member and then a stochastic convective trigger, initial wind and PBL height perturbations were added within each member.

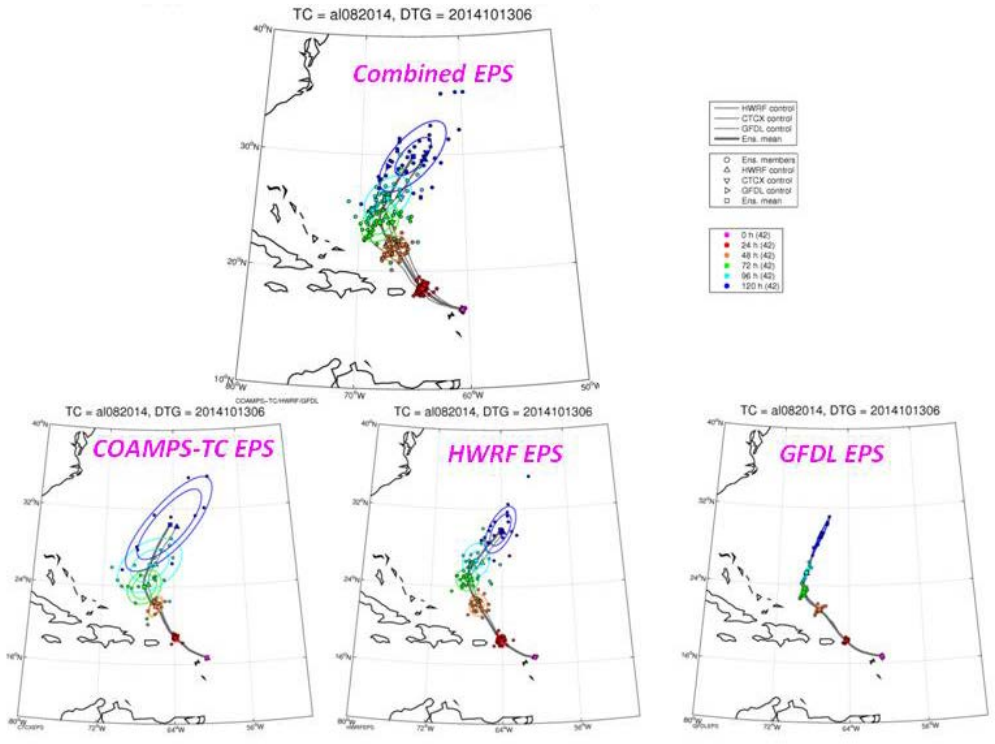


Figure 13. The HFIP Multi-Model Ensemble for Tropical Storm Gonzalo (2014)
This figure depicts an example of a forecast from the HFIP Multi-Model ensemble consisting of HWRF, COAMPS_TC and GFDL. Ellipses enclose 80% of member positions at indicated time. Black lines are the ensemble means for the individual models ensemble (bottom panels) and combined ensemble (top panel). Forecasts shown are for Tropical Storm Gonzalo from initial time on October 13, 2014 at 06 Z.

For COAMPS-TC ensemble:

- 27-, 9-, 3-km horizontal grid spacing
- 1 control + 10 members with initial and boundary condition perturbations
- No physics perturbations
- No data assimilation
- Control forecast:
  - Initialized from the GFS analysis
  - Vortex initialized with a Rankine vortex based on TC vitals
- Ensemble members IC's perturbed about the control:
  - Synoptic perturbations drawn from static covariance (e.g. WRFVAR cv3)
  - Vortex IC's based on perturbed TC vitals

The GFDL ensemble is described in subsection (d).

Figure 13 shows an example of one forecast from the multi-model ensemble for Gonzalo for each of the three component models and the combined model (EPS) with the combined ensemble including the combined mean (black line). It is interesting to note that the dispersion of the tracks for the GFDL model is very narrow while the dispersion is greatest for the COAMPS-TC. Figure 14 shows the performance of control, each of the individual three ensemble means and the

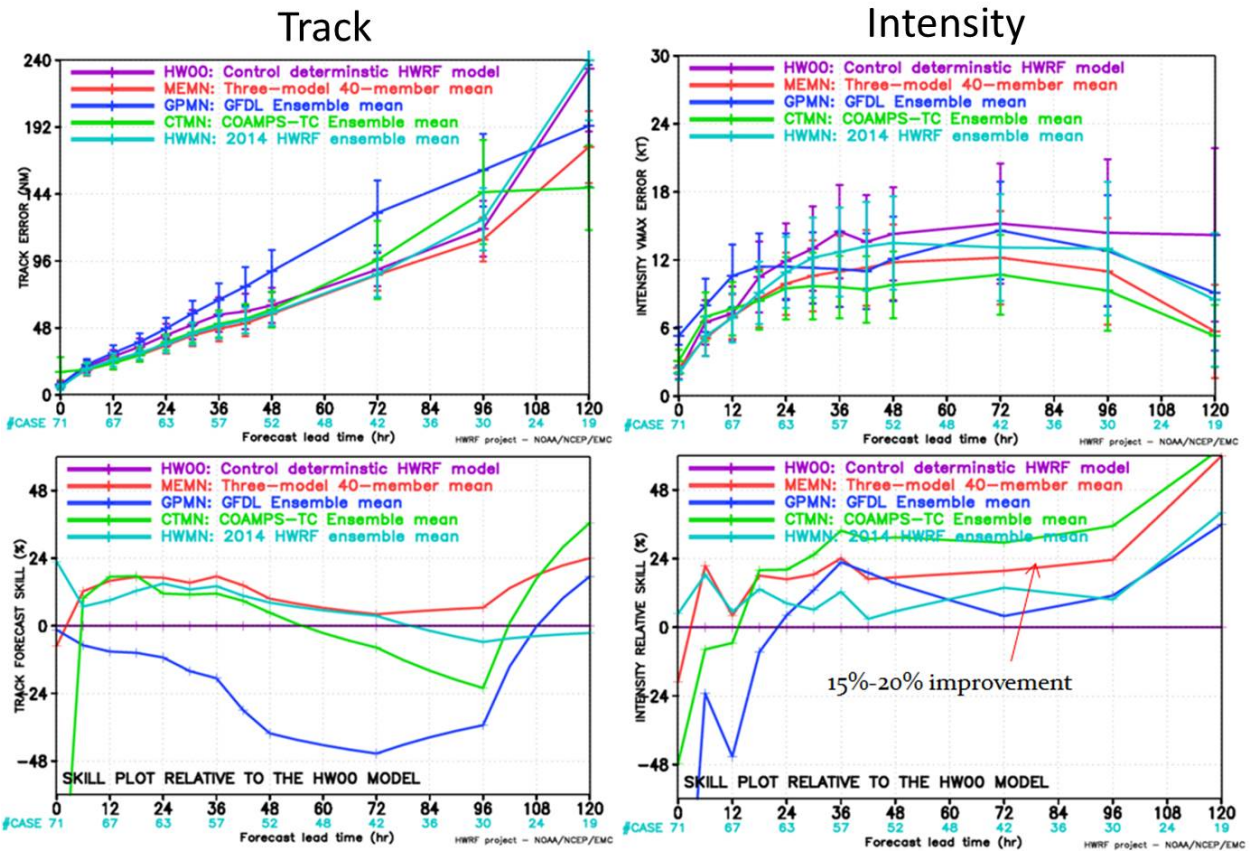


Figure 14. Comparing Track & Intensity Errors between Multi-Model Ensemble Components
Track (left panels) and Intensity (right panels) show errors (top panels) and skill (lower panels) relative to the HWRF deterministic/control model (bottom panels) for 2014 in the Atlantic. The figure compares the various components of the multi-model ensemble: Blue (GFDL ensemble), green (COAMPS-TC), light blue (HWRF ensemble), purple (operational deterministic HWRF) and red (multi-model ensemble mean).

mean of the three-component ensemble. With track the overall ensemble mean was the best performer giving about a 15-20% improvement over the HWRF operational model (control). For intensity the best performer by far was the COAMPS-TC ensemble mean though the three component ensemble mean was still a 15-20% improvement over the operational HWRF. The GFDL ensemble performed relatively poorly for track but comparable to the HWRF ensemble after 24 h.

d. GFDL Ensemble

For the last five 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 9. Working with the forecasters at NHC, scientists at GFDL constructed an ensemble by modifying various parameters in the initial conditions, sea surface

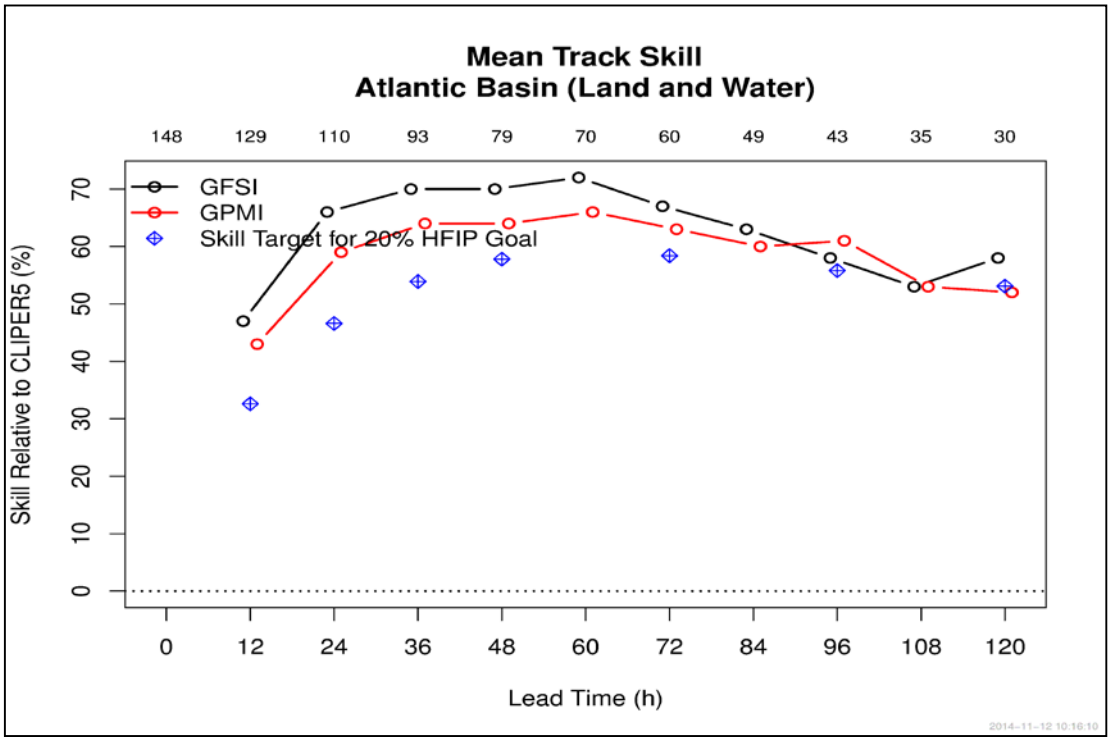
temperatures and surface fluxes used by the model. The *unbogussed* forecasts start from the GFS without modification to the vortex from what was in the GFS initially. The other members use the GFDL initialization scheme but modified as described in Table 9.

Table 9. Automated Tropical Cyclone Forecasting System (ATCF) ID Descriptions.

ATCF ID	Description
GP00	Control forecast (based on NCEP 2014 operational GFDL)*
GP01	**Opposite vortex bogussing** method of the ensemble control model (i.e., runs unbogussed when the control runs bogussed, and vice versa)
GP02	**Increase** NHC-observed V_{max} 10%, R34 25%, R50 40%, ROCI 25% ↑
GP03	**Decrease** NHC-observed V_{max} 10%, R34 25%, R50 40%, ROCI 25% ↓
GP04	**Increase** inner-core moisture by a max of 10% ↑
GP05	**Decrease** inner-core moisture by a max of 10% ↓
GP06	**Increase** SSTs by a max of 2°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 2014 operational formulation* of C_H (surface enthalpy exchange coefficient) ↓
GPMN	Ensemble mean computed at each lead time where the member availability is at least 4 members (40% threshold)

Slide

Figure 15 shows the results for the GFDL model in the Atlantic. The upper panel shows the track skill of the GFDL ensemble mean compared to the control run (the operational GFDL, GFSI) and the bottom panel shows the intensity skill compared to DSHP. Also shown are the five year 20% HFIP goals for track and intensity. Note that for track the ensemble did better than the HFIP goal for all lead times but not as well as the operational GFDL. For intensity the GFDL ensemble mean performed better than the goal starting from 72 hours though it performed very poorly early in the forecast. However, DSHP was the superior intensity model for all lead times. Note in Figure 13 that the GFDL ensemble had very narrow spread in track. This low dispersion is likely why the ensemble did not compare well with the deterministic operational model.



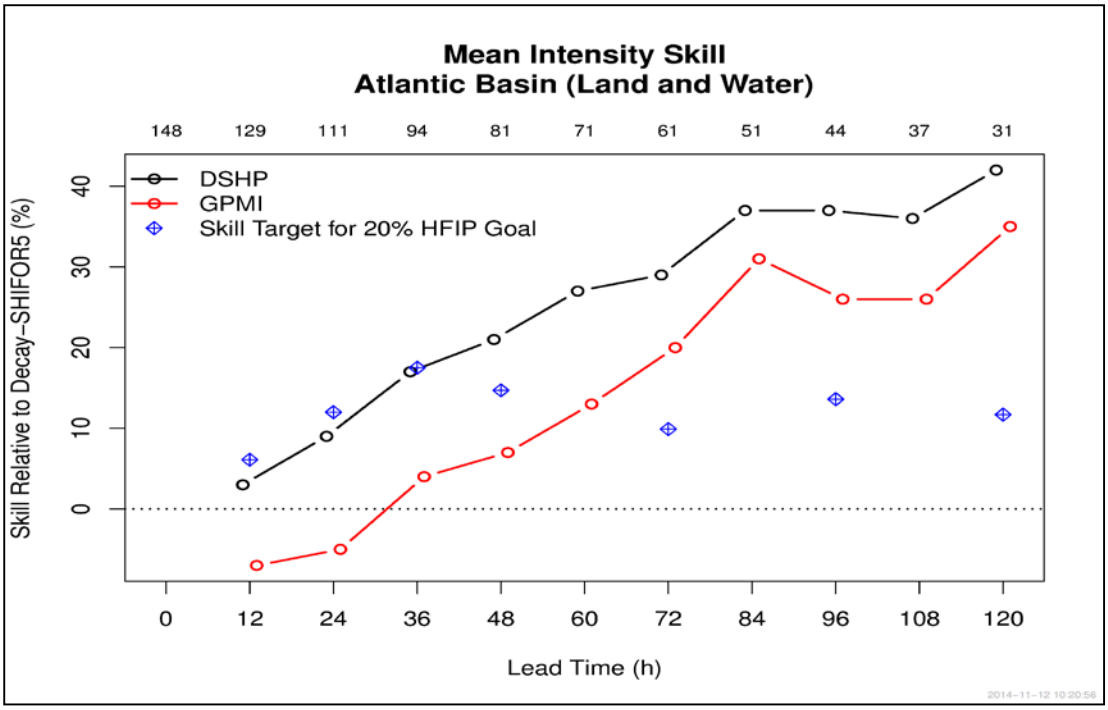


Figure 15. Mean Track and Intensity Skills in the Atlantic Basin.
Results Skill for the GFDL ensemble mean (interpolated version) for 2014 in the Atlantic for track (upper panel) and intensity (lower panel) compared to GFSI (track) and DSHP (intensity). Skill calculated relative to CLIPER5 for track and Decay-SHIFOR5 for intensity.

8. Impact of Inner Core Reconnaissance Data

Figure 16 shows results from that effort using the 2013 operational version of HWRF applied to storms in 2009-2013. Note that these are rather small samples, especially for the two hurricane samples. For tropical storms the aircraft data substantially improved the intensity forecasts at almost all lead times. For category 1-2 hurricanes the results for intensity were mixed and there did not seem to be an overall positive impact. For major hurricanes there is a very large initial error when the aircraft data were included due to an initial rapid spin down from the analysis although for later forecast periods the impact was slight. Including the Tail Doppler Radar (TDR) data had almost no overall impact. For track the impact of including the aircraft, with and without TDR, was minimal. So from this sample, the only overall improvement appears to be for intensity by including aircraft data for the tropical storm strength, perhaps aided by initializing a more realistic vortex. But even for these storms the addition of TDR data did not appear to impact the results.

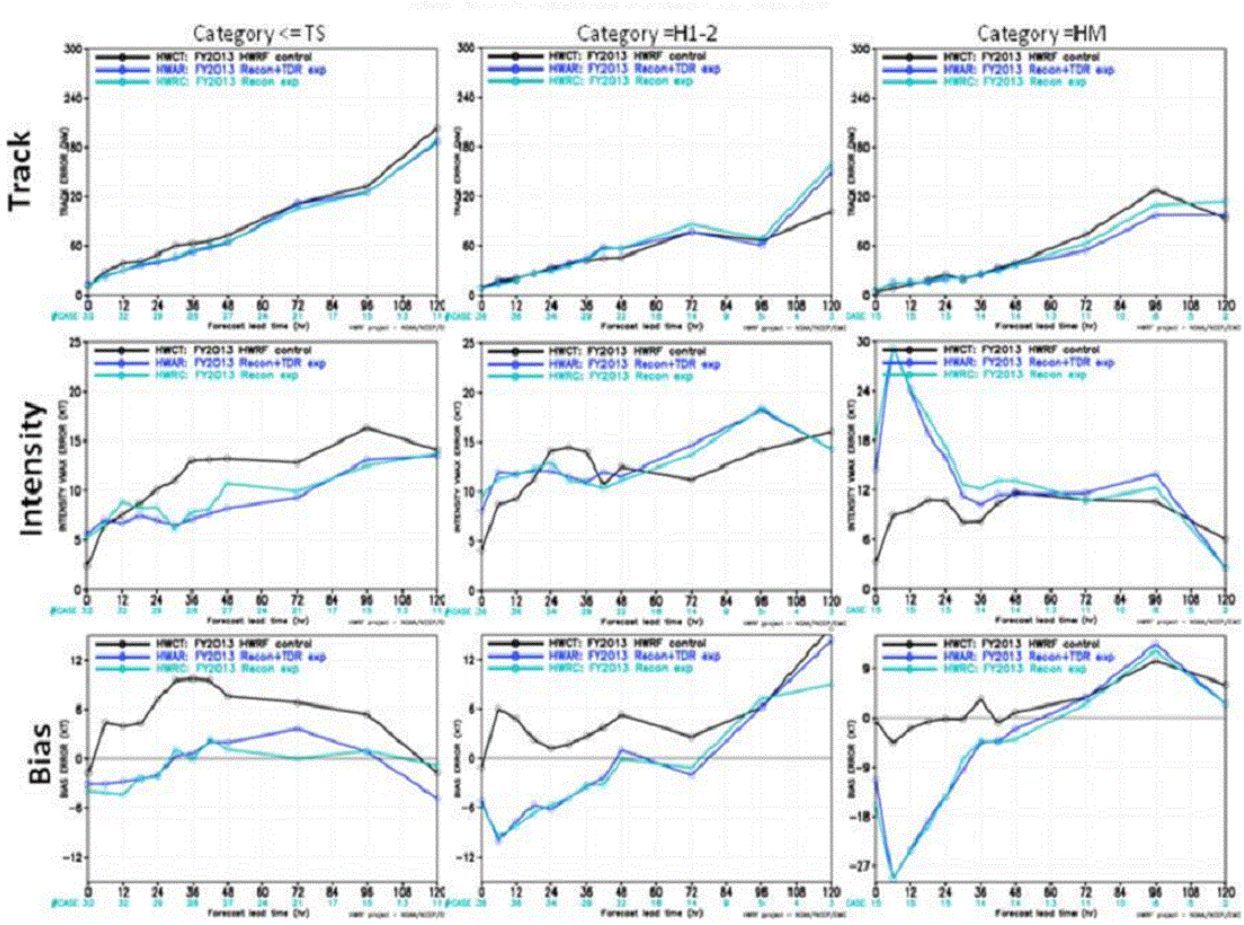


Figure 16. 2014 Track, Intensity and Bias comparisons with Inner Core Reconnaissance Data Comparison between the 2014 operational HWRF model (black lines), the same model but where the reconnaissance data excluding the TDR data are used (dark blue line) and using all the RADAR data including the TDR data (light blue line). Track errors are shown in the top row, intensity errors in the second row, and intensity bias in the third row. The left column includes only tropical storm strength, the middle column-category 1 and 2 hurricanes and the right column category major hurricanes. The number of cases in each panel is shown in blue figures below the panel.

Last year a detailed study of the impact of tail TDR on hurricane forecasts was described. That study was rather inconclusive in that with the three separate models considered (Pennsylvania State University, HWRF, and the AOML model) the impacts were not consistent between the models and overall the impact was small. It was felt that further work with the models, especially improvement in the DA systems, was necessary before we would see a consistent impact. This study was not repeated this year but the HWRF team continued to experiment with including the data in the HWRF initialization.

9. Data Assimilation and Physics Development

a. Data assimilation

Across the HFIP program there are many attempts to develop and test new techniques that will result in future improvements to the operational models. Here we discuss two such attempts.

In Figure 17 the work flow for initialization of the current operational HWRF system is shown. Basically it uses the GFS ensemble to define the background error and adjustments to the position and intensity of the resulting vortex to match observations at the time of initialization. Thus the DA cycle uses perturbations from the large scale coarse resolution ensemble to define the background error.

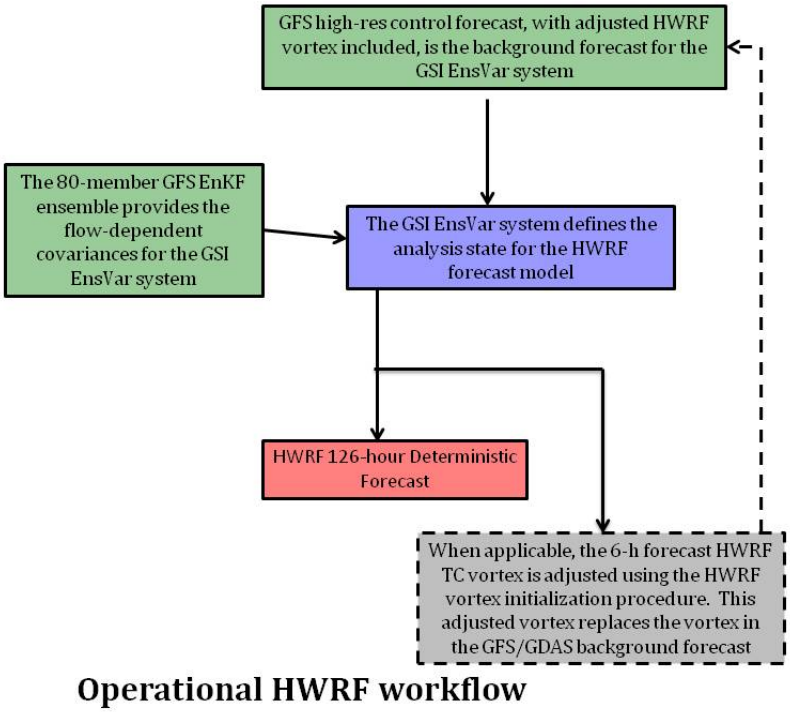


Figure 17. Operational HWRF Workflow
A schematic of the workflow within the current HWRF operational model initialization.

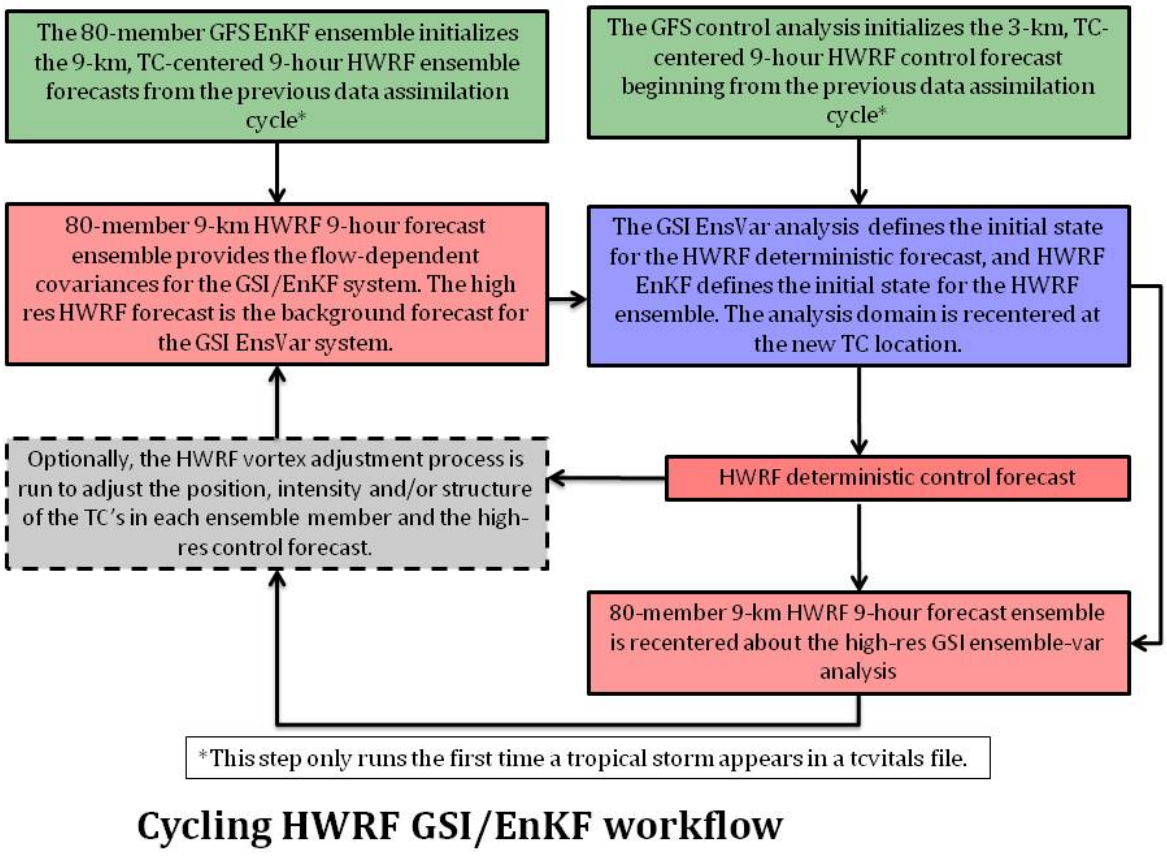


Figure 18. Cycling HWRF GSI/EnKF Workflow
A schematic of the workflow within an experimental version of HWRF Operational model initialization implemented at the University of Oklahoma. The main difference from Figure 18 is the incusion of cycling data using the members of an HWRF ensemble in the initialization.

Figure 18 shows an experimental version of the HWRF initializaton system developed by the University of Oklahoma. The main difference from Figure 17 is the inclusion of a cycling DA system that uses a background error defined by the HWRF ensemble which employs the EnKF DA technique developed within the GSI framework. Figure 1 shows some preliminary results for the experimental system.

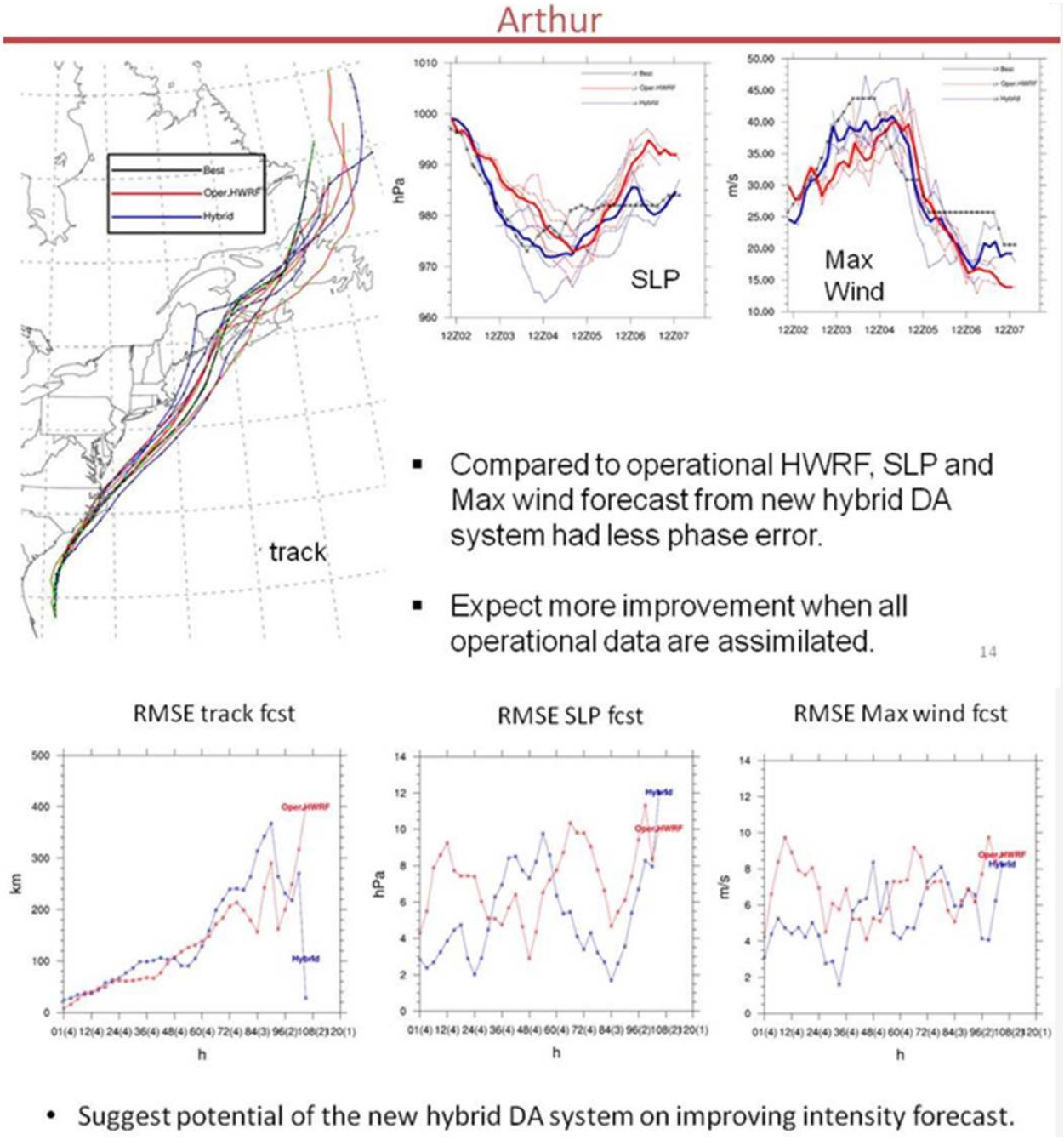


Figure 19. HWRF GSI/EnKF Method for Model Initialization (Hurricane Arthur in 2014)
An example from Hurricane Arthur in 2014 where the method diagrammed in Figure 18 was used to initialize the model.

b. Physics Development

There is an ongoing effort to better parameterize the flux coefficients for heat (C_h) and momentum (C_d). We have discussed earlier attempts to improve these parameterizations in previous annual reports. This year GFDL experimented with a radically different set of values (as a function of wind speed). The old values are shown in the right upper panel of Figure 20 with dashed lines. The new values are shown by solid lines.

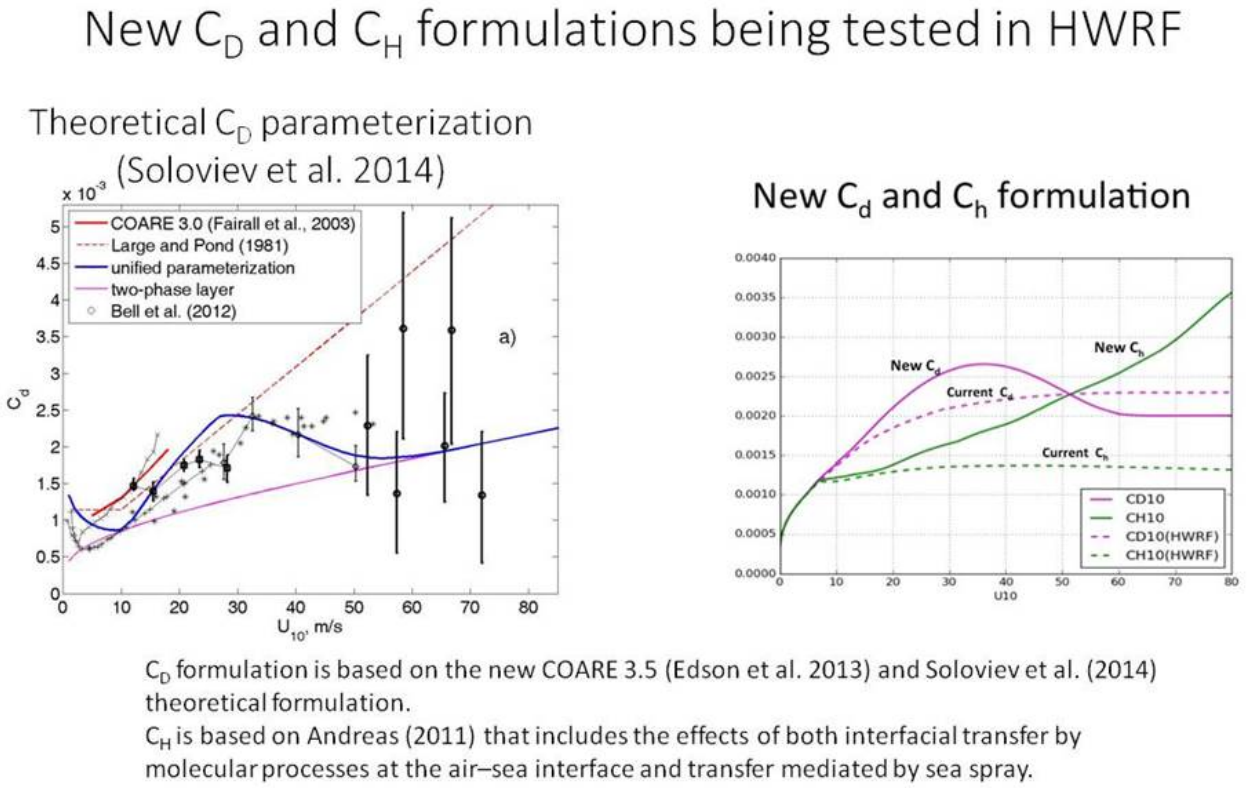


Figure 20. 2014 HWRF Formulations Testing and Evaluations for C_p and C_h
Upper panels show changes to C_D and C_H that were included in the 2014 GFDL model. Left upper panel shows data on C_d and a theoretical parameterization from Soloviev et al (2014). The right panels compare the new C_d and C_h parameterizations where C_d formulation is based on the new COARE 3.5 (Edson et al. 2013; Soloviev et al, 2014) –left upper panel and C_H is based on Andreas (2011) that includes the effects of both interfacial transfer by molecular processes at the air–sea interface and transfer spray. The lower panels compare intensity errors (left lower panel and bias (right lower panel for the 2014 HWRF using the old parameterization for C_d and C_H (red lines) and with the new values (blue line used in the GFDL model).

The impact of these changes in C_h and C_d on intensity are illustrated in the lower panels through forecasts made by the 2014 HWRF model with the old (red lines) and new (blue lines) set of values for D_h and C_d. Note that both the intensity error and the intensity bias are reduced with these new values of C_h and C_d. These results are expected to be somewhat representative since the results for intensity forecasts are from a reasonably large sample size.

10. Post Processing of Model Output

a. Statistical Post Processing of Model Output: The FSU Multi-Model Ensemble

Much of the discussion above focused on using numerical model improvements to achieve the HFIP goals. Typically statistical models (for example DSHP) are among the best predictors of hurricane intensity. A statistical model is one where a limited number of predictors (measured in single to double digits) are combined with weights that are determined by correlation with past data. These predictors are generally selected from parameters describing the current state of the hurricane or various environmental data. Those using environmental data can specify their values from current observations or from model forecasts. There is another class of statistical model that takes a particular prediction from a dynamical model (e.g., track or intensity) and combines it with a weighted average from other models in a multi-model ensemble. The weights are determined by comparing the performance of the various models over a period of years. Perhaps the simplest statistical model for intensity is SHIFOR5 (also called OCD5) where the variables are current position, and intensity including intensity 12 hours prior to the current time (CLIPPER5 is a similar model for track), see Figure 2.

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

Figure 21 on the following page illustrates results for the Atlantic Basin during the 2014 hurricane season.

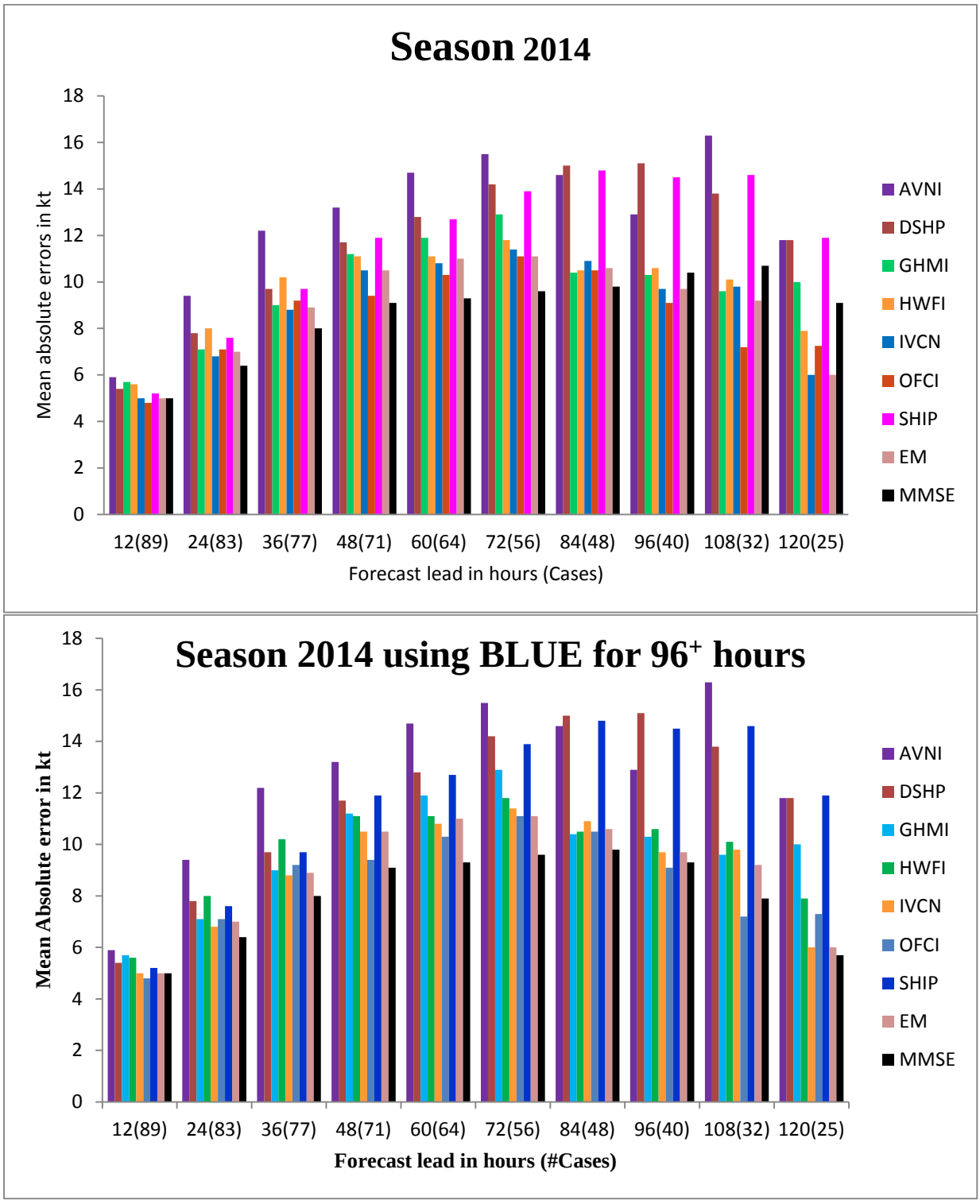


Figure 21. 2014 Atlantic Hurricane Season Results using BLUE 96-h$^+$ Difference Weighting Scheme. The upper panel of depicts intensity errors for all components of the FSU Multi-Model Ensemble for various forecast lead times. The acronyms for the various models are shown on the right, see Appendix B for definitions. The ensemble mean (EM) of the models shown is in orange and the weighted ensemble mean, MMSE, is shown by the black bar. The official forecast is dark red. The bottom panel is the same as the upper panel (but note that the some of the colors have changed so refer to the label on the right of the panel. The difference from the above panel is that a difference weighting scheme is used for MMSE called BLUE. See text for details on BLUE.

In most years the statistical models provide forecasts of intensity almost as good as any of the current dynamical models. As illustrated in Figure 2, last year the dynamical models were comparable.

As in past years the FSU Multi-Model Ensemble was among or the best performer of the statistical models. The FSU system employs a process that finds an optimum linear combination of model outputs to provide a better forecast. This process forms a multiple linear regression equation using least squares. Previous years of data are used to fit the regression equation (known as the training phase). The system was then used to forecast in real-time (forecast phase).

Figure 21, upper panel shows the intensity errors for the various models that went into the FSU Multi-Model Ensemble. The black and purple bars on the right side of the groups for each forecast lead time are the equally-weighted ensemble mean (EM; purple line) and for the variably-weighted ensemble mean (MMSE; black line). At most lead times MMSE was better than EM and from 24-84 h, MMSE was better than all other models. After that the performance of MMSE dropped off and was almost the worst performer.

It was found that the problem with the later times is related to a correlation between errors of different models used in the regression calculations. If two or more regression variables are highly correlated, test of significance of the affected coefficients will give misleading results. In attempt to correct this defect in the FSU system, a new system known as the Best Linear Unbiased Estimator (BLUE): a linear combination of estimators with properties of no bias and minimum variance). The results using BLUE starting at 96 h are shown in the lower panel of Figure 21. Forecasts from 12-84 h are calculated as in the top panel. The use of BLUE for longer lead times greatly reduced the average errors, up to 69% at 120 hours. MMSE (with BLUE) is superior to EM at all lead times past 12 h, and is better than all of the other models at most lead times.

b. The HFIP Web Page

Each hurricane season HFIP runs a real-time system on the HFIP machines in Boulder, CO. 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. For the purpose of presenting results of the runs on the HFIP computer and as presenting data run elsewhere (e.g., NRL and operational models), HFIP has developed a web page (http://www.hfip.org/products/) where the various results are presented. Figure 22 shows an example of a product from the HFIP web site. This product has been available for a couple of years and some enhancements were added in 2014, e.g., intensity color coding for the ensemble tracks and allowing display of various models for comparison as an inset (see Figure 22).

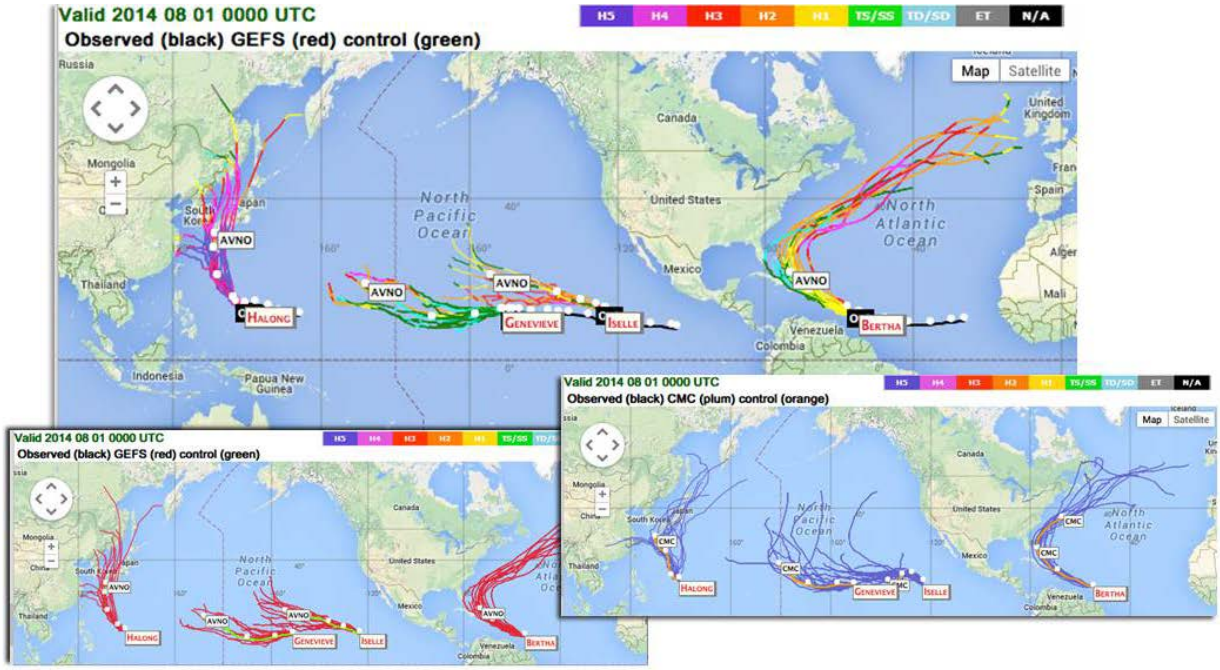


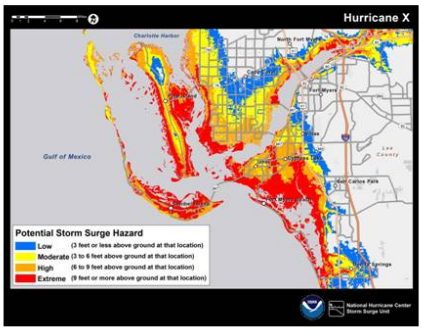
Figure 22. Example of the HFIP Webpage

11. Societal Impact Work

In 2014 HFIP provided modest support for the Societal Impacts Team (SIT) with most of this support going to NHC. Figure 23 shows an example of a storm surge impact product that has been developed at NHC and how it was rated. This type of product was displayed last year in the HFIP annual report with Tampa Bay as the example and was evaluated by SIT 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 rated it very high for ease of understanding and for providing useful information.

Figure 24 is a similar survey for how these groups rate the uncertainty cone for forecast hurricane track that has been used operationally for many years. The results were mixed with the public giving it high marks for ease of understanding and usefulness, while only about half of the WCMs rated it favorably.

Survey Results for the Map



- Most in all surveys preferred this map over solid blue one or graduated blue one
- Problems with using "low" to describe storm surge hazard

Positive Ratings*

- Ease of understanding
  - 86% by EMs
  - 96% by Media
  - 77% by Public
  - 90% by WCMs
- Usefulness
  - 84% by EMs
  - 94% by Media
  - 98% by Public
  - 83% by WCMs

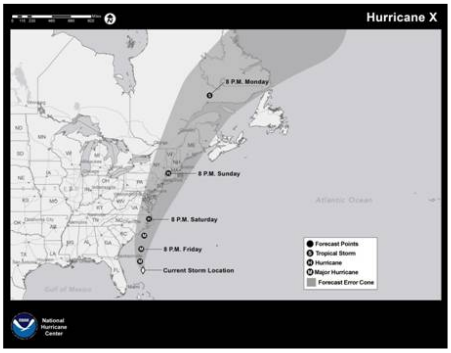
*Excellent, Very Good or Good

Note: Some indication that people pay more attention to legends when at lower left

35

Figure 23. Assessment of Potential Storm Surge Product from the NHC. The storm is hypothetical but was motivated by Sandy.

Track Forecast Cones



Positive Ratings*

- Ease of understanding
  - 70% of EMs
  - 76% of Media
  - 93% of Public
  - 56% of WCMs

Usefulness

- 72% of EMs
- 74% of Media
- 87% of Public
- 46% of WCMs

*Excellent, Very Good or Good

31

Figure 24. Assessment of how various groups rate the cone of uncertainty for the forecast hurricane track.

NWS Service Assessments, such as those for Hurricanes' Irene and Sandy, identified a number of gaps in the way NOAA articulates the risk from tropical cyclones in its various products, information, and services—at both the national and regional levels. Additionally, NWS partners have expressed a strong desire to have NHC create a separate product depicting the potential time of arrival of sustained winds of tropical storm force. Such a product will enable emergency managers and other officials to make more informed decisions on when to conduct and complete preparations as a tropical cyclone approaches. Broadcast meteorologists also have a need for this information for use in their on-air programming, websites, and social media.

The HFIP Socio-Economic Team has begun development of an "Onset of Tropical Storm Force Winds" product (scheduled completion date: 11/30/2015). There are no finalized graphics at this time. However, an early graphic that was tested during the Tropical Storm Surge research is presented here (Figure 25).

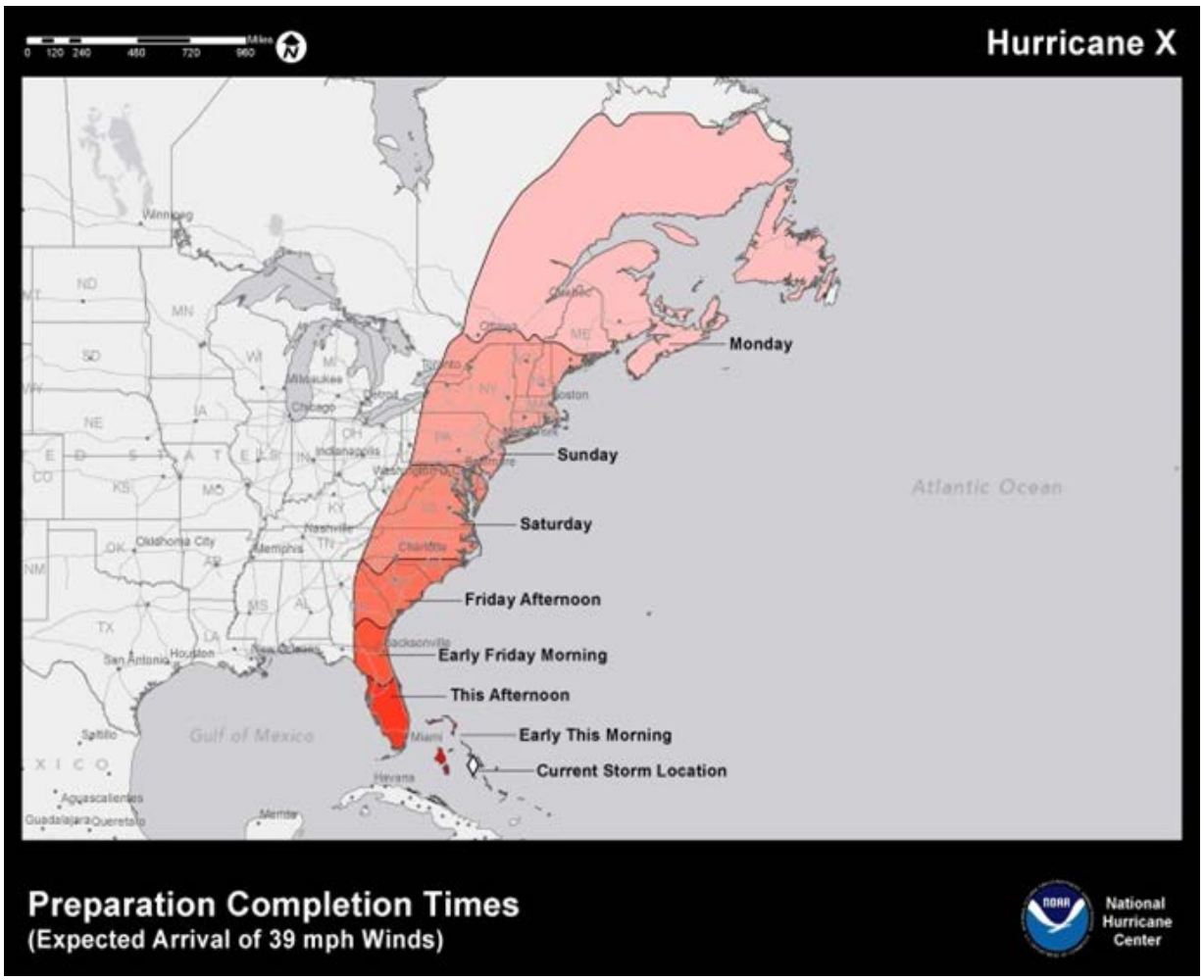


Figure 25. Onset of Storm Force Winds Product:
This figure illustrates a prototype of the arrival of a storm force winds product.

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

It has already noted that the HWRF undergoes considerable testing when new techniques and technology are added to the system. This has led to some significant improvements (Figure 10). HWRF development for the remainder of the HFIP program, originally scheduled to be a 10 year program ending in 2019, is charted out here. The general strategy, for the next 5 years is to gradually evolve the HWRF into the NOAA Environmental Modeling System (NEMS) framework, the infrastructure that has been adopted for other models at EMC. In addition, HWRF will move from EMC Non-hydrostatic Mesoscale Model (NMM) on the E-grid to NMM on the B-grid (NMMB) which is currently 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 (large domain) system where there are multiple moving double nests (one set per hurricane) where each set of nests interact with the large basin-scale domain. This allows for interactions between relatively close hurricanes via the larger scale domain. These interactions can occasionally have an important impact on hurricane track forecasts and perhaps a lesser impact on intensity. It also saves computer time since the basin-scale domain only has to be computed once for all of the hurricanes rather than separately for each hurricane. In addition, since NMMB can easily be made global, the basin-scale domain can be made larger depending on computer resources and even eventually evolve into a global model. Such a global model is currently undergoing development and testing and an example of a forecast with five simultaneous tropical cyclones in the Atlantic and West Pacific is shown in Figure 26. Due to limited computer resources, it has not yet been possible to test the global system over a large number of cases.

Figure 27 shows the comparisons for track and intensity forecasts between the 2013 versions of the basin-scale HWRF (HWHI: red lines) and the operational HWRF (HWFI: black lines) for a large sample size (large number of storms) from 2011-2013. For storms that were simultaneous in time, they were run together in the basin-scale system and individually in the operational HWRF. HWHI out-performs HWFI at almost all lead times for both track and intensity suggesting the potential of the basin-scale version to produced improved forecasts.

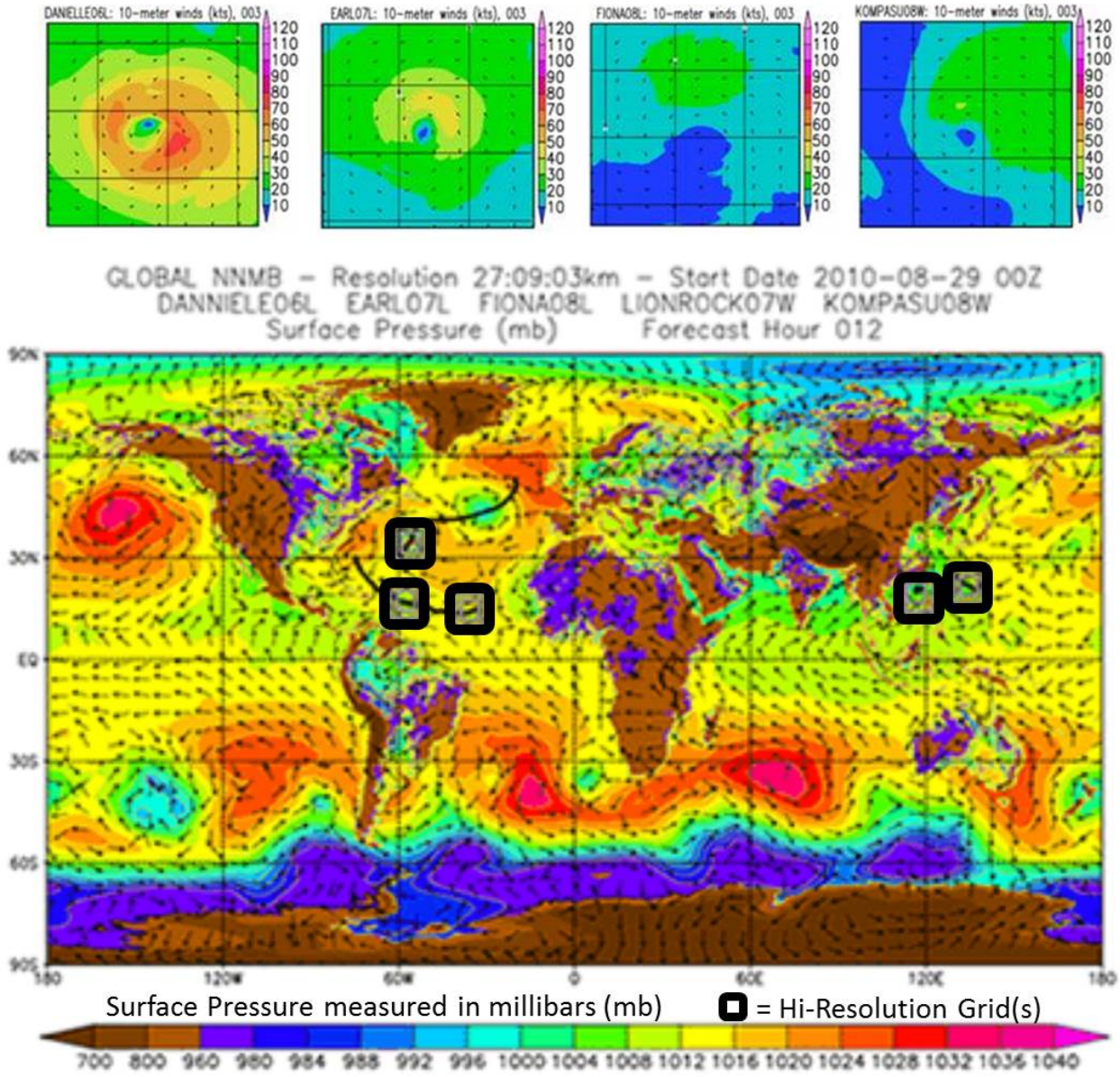


Figure 26. Example of the experimental global HWRF System in NMMB NEMS framework:
This depicts an example of the experimental global HWRF system with multiple moving inner nests about each hurricane. Each hurricane has a pair of moving nests that interact with the global modeling system. This image is from an actual forecast of the five hurricanes shown. Four inner nests are shown in the upper panels.

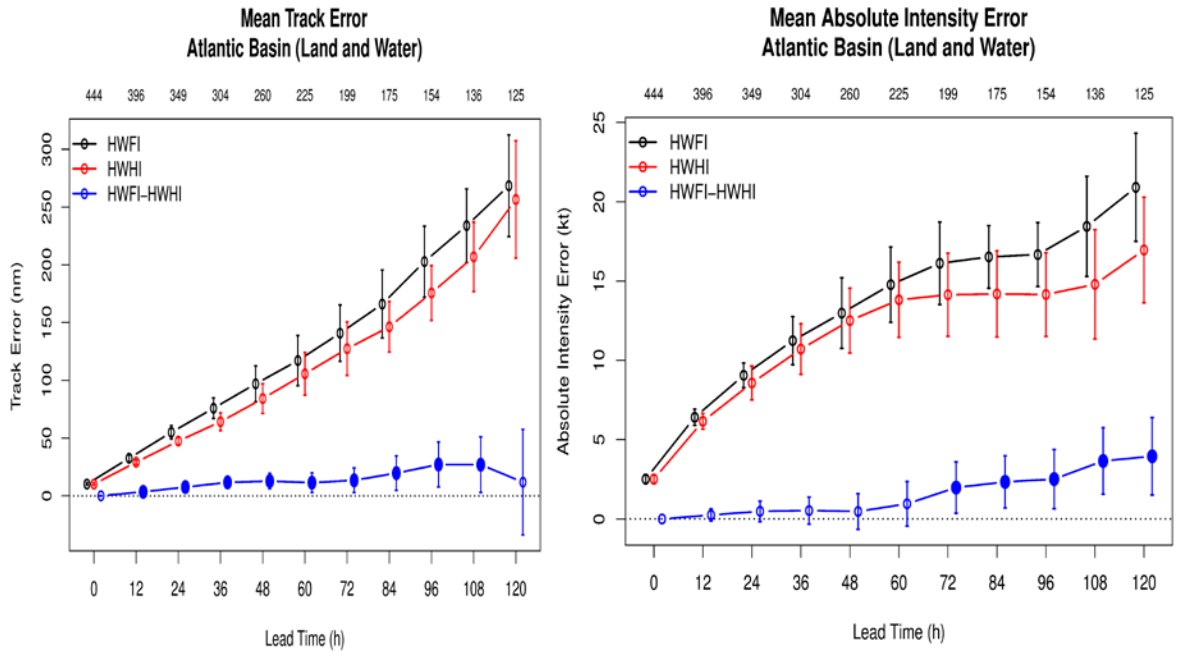


Figure 27. Track and Intensity Forecasts for HWRF Basin-Scale and the Operational HWRF. Comparison of the track (left panel) and intensity (right panel) forecast errors for Atlantic storms from 2011-2013 from the 2013 versions of the basin scale HWRF (HWHI: red lines and operational HWRF (HWFI: black). The blue line is the difference in error HWFI-HWHI). The number of cases is shown across the top of each panel. Error bars for 95% confidence interval are shown for each model and for the difference.

In past years we have speculated on what the final configuration of the hurricane model might look like at the end of the project. Table 10 shows the anticipated 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 10. Numerical Model Hurricane Forecast Guidance System

Component	Specifications
Global model ensemble with Hybrid Data Assimilation	20 members at 10-20 km
Multiple moving nests with the inner nest at 3 km or higher resolution within the global model	Double nests, one set for each hurricane using all available aircraft and satellite data in inner core and near environment of the 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, AHW, TC-COAMPS)
Statistical Post Processing	LGEM, SHIPS, SPICE, and others

References

Aksoy, A., S. Lorsolo, T. Vukicevic, K. J. Sellwood, S. D. Aberson, and 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. *Mon. Wea. Rev.*, **140**(6):1843–1862, doi:10.1175/MWR-D-11-00212.1.

Andreas, E. L., 2011: Fallacies of the Enthalpy Transfer Coefficient over the Ocean in High Winds. *J. Atmos Sci*., **68**(7), 1435-1445, doi:10.1175/2011JAS3714.1.

Bernardet, L., V. Tallapragada, S. Bao, S. Trahan, Y. Kwon, Q. Liu, M. Tong, M. Biswas, T. Brown, D. Stark, L. Carson, R. Yablonsky, E. Uhlhorn, S. Gopalakrishnan, X. Zhang, T. Marchok, B. Kuo, and R. Gall, 2014: Community support and transition of research to operations for the Hurricane Weather Research and Forecast (HWRF) Model. *Bull. Amer. Meteor. Soc*., in press. [Available at http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-13-00093.1]

Cangialosi, J. P., and J. L. Franklin, 2011: 2010 National Hurricane Center Forecast Verification Report. NOAA/NWS/NCEP/National Hurricane Center, 77. [Available online 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. *Mon. Wea. Rev*., 119, 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. *Mon. Wea. Rev*., **129**(4), 869-883, doi:10.1175/1520-0493(2001)129<0869:NFAD>2.0.CO;2. [http://www.dtic.mil/dtic/tr/fulltext/u2/a396883.pdf]

DeMaria, M., 2009: A Simplified Dynamical System for Tropical Cyclone Intensity Prediction. *Mon. Wea. Rev*., **137**, 68–82, doi: http://dx.doi.org/10.1175/2008MWR2513.1

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

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

Edson, J.B., V. Jampana, R. A. Weller, S. P. Bigorre, A. J. Plueddemann, C. W. Fairall, S. D. Miller, L. Mahrt, D. Vickers, and H. Hersbach, 2013: On the Exchange of Momentum over the Open Ocean. *J. Phys. Oceanogr*., **43**, 1589–1610, doi:10.1175/JPO-D-12-0173.1.

Evensen, G., 1994: Sequential Data Assimilation with a Nonlinear Quasi-geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics. *J. Geophys. Res*, **99**(C5):10,143–10,162; doi: 10.1029/94JC0057.

Goldenberg, S., S. Gopalakrishnan, V. Tallapragada, T. Quirino, F. Marks, S. Trahan, X. Zhang, and R. Atlas, 2015: The 2012 Triply-Nested, High-Resolution Operational Version of the Hurricane Weather Research and Forecasting System (HWRF): Track and Intensity Forecast Verifications. *Wea. Forecasting*, doi: 10.1175/WAF-D-14-00098.1, in press.

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. *Mon. Wea. Rev.*, **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. *Mon. Wea. Rev.*, **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. *Mon. Wea. Rev.*, **128**(6):1971-1981, doi: 10.1175/1520-0493(2000)128<1971:DAIAPE>2.0.CO;2.

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

NOAA SAB, 2006: Majority Report: National Oceanic and Atmospheric Administration Science Advisory Board Hurricane Intensity Research Working Group. *NOAA SAB HIRWG*, 05 July, 2006. [Available at http://www.sab.noaa.gov/Reports/HIRWG_final73.pdf]

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

Soloviev, A. V., R. Lukas, M. A. Donelan, B. K. Haus, and I. Ginis, 2014: The air-sea interface and surface stress under tropical cyclones. *Nature Scientific Reports*, **4**, 5306, doi:10.1038/srep05306.

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. [Available athttp://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRFv3.5a_ScientificDoc.pdf]

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

Wu, W., R. J. Purser, and D. F. Parrish, 2002: Three-Dimensional Variational Analysis with Spatially Inhomogeneous Covariances. *Mon. Wea. Rev.*, **130**, 2905–2916, doi: http://dx.doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2.

Appendix A: NOAA Opportunity Announcement

Through an Announcement of Opportunity (AO) in March 2011, 2012, the 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 DA techniques focusing on use of satellite datasets and in-situ measurements obtained through: (a) aircraft reconnaissance; (b) advancements in numerical weather prediction techniques with emphasis on high-resolution physics for more accurate representation of moist processes; (c) air-sea interactions in the hurricane inner-core region; (d) advanced model diagnostics to test model improvements; (e) advanced ensemble techniques for improved track and intensity forecast guidance; (f) enhanced observing strategies; and (g) use of observations for hurricane environment-scale and hurricane inner core-scale circulations.

Of the 32 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. Among the 12 selected proposals: two address DA priorities; four address operational hurricane model improvements; and five address high-resolution physics and ensembles. Progress of each of these proposals is monitored through quarterly reports from the principle investigators (PIs) and through scientific presentations from the PIs 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. Most of these awards are now closing, and final reports being prepared though some investigations have requested and received no-cost extensions. These reports serve to identify results that could be transitioned to operations with varying levels of further development.

In 2014, HFIP launched a second round of awards focused on advanced physics and ensembles for: hurricane intensity predictability, global-to-local scale modeling techniques, air-sea-wave coupling, and continuous improvements to operational hurricane models and DA techniques. The review process for these proposals has been completed and announcements of awards were made in Spring 2014. The list of the winning proposals is made available on http://HFIP.org (http://www.hfip.org/documents/HFIP_Projects_Funded_Round2.pdf). The majority of these proposals are geared towards improving the operational HWRF model by more direct and efficient transition of research to operations. DTC has setup an advanced support system for HWRF model developers (http://www.dtcenter.org/HurrWRF/developers/) to provide training and to facilitate model development in a community repository commonly used by researchers and operational model developers. There are plans for a review of the first year progress of the 14 awards made.

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 "early" (aka "interpolated") forecasts from "late" models (see Section 4a) are adjusted so that the previous 6-hour forecast matches the conditions at the beginning of the current forecast. Forecasts of those future conditions are denoted with an "I" at the end (12-hour interpolations are denoted with a "2").

Other conventions occasionally used in the model naming include 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.

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:	Advanced Weather Research and Forecasting model with 6-hour interpolation.
ARW:	Pennsylvania State University Advanced Research WRF
BLUE:	Best Linear Unbiased Estimator that is a linear combination of estimators with properties of no bias and minimum variance designed to resolve misleading results from a test of significance of affected coefficients associated with two or more highly correlated regression variables.
CMC:	Canadian Meteorological Centre model.
CMCI:	CMC with 6-hour interpolation.
CLIPER:	Climate and Persistence model.
COAMPS-TC:	Coupled Ocean/Atmosphere Mesoscale Prediction System-Tropical Cyclone model at the Fleet Numerical Meteorology and Oceanography Center.
COTI:	COAMPS-TC with 6-hour interpolation.
CXTL:	Experimental COAMPS-TC Forecast Interpolated ahead 6 hours.
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.
EM:	Equally-weighted Ensemble Mean for models used in MMSE.
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).
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).
GTMI:	Geophysical Fluid Dynamics Laboratory model's regional and dynamic ensemble.
H3WI:	Hurricane Research Division's regional, dynamic and deterministic ensemble
HWHI:	Basin-scale HWRF with 6-hour interpolation.
HWMI:	HWRF Ensemble Mean Forecast Interpolated Ahead 6 hr.

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.
IVCN:	Intensity forecast ensemble (operational) without stream 1.5 forecasts.
LGEM:	Logistics Growth Equation Model.
MMSE:	FSU Multi-Model Ensemble (variably weighted mean).
NAVGEM:	Fleet Numerical Meteorology and Oceanography Center Navy Global Environmental Model (replaced NOGAPS February, 2013).
NEMS:	NOAA Environmental Modeling System.
NGPI:	NOGAPS with 6-hour interpolation.
NGXI:	Experimental NOGAPS with 6-hour interpolation.
NMM:	Environmental Modeling Center Non-hydrostatic Mesoscale Model.
NMMB:	NMM on the B-grid.
NOGAPS:	Fleet Numerical Meteorology and Oceanography Center Navy Operational Global Atmospheric Prediction System (replaced by NAVGEM February, 2013).
OFCL:	Official National Hurricane Center Forecast.
RI:	Rapid Intensification (an increase of 30 knots in 24 hours).
RW:	Rapid weakening (a decrease of 25 knots in 24 hours).
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.
UW4I:	University of Wisconsin's Non-hydrostatic Modeling System (4 km).
UWNI:	UW-NMS with 6-hour interpolation.
UW-NMS:	University of Wisconsin Non-hydrostatic 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).