2010 HFIP R & D Activities Summary: Accomplishments, Lessons Learned, and Challenges

R. Gall[11], F. Toepfer[11], S. Aberson[1], J.W. Bao[4], M. Bender[6], S. Benjamin[4], L. Bernardet[4,14], M. Biswas[5], B. Brown[7], C. Davis[6], M. DeMaria[8], J. Doyle[10], M. Fiorino[4], J. Franklin[9], I. Ginis[13], J. Goerss[10], S. Gopalakrishnan [1], T. Hamill[4], H.S. Kim[3], T. Krishnamurti [5], Y. Kwon [3], W. Lapenta[3], S. Lord[3], T. Marchok[6], F. Marks[1], L. Nance[7], P. Kucera[7], E. Rappaport[9], C. Reynolds[10], N. Surgi[3], V.Tallapragada[3], H. Tolman[3], G. Vandenberghe[3], Y. Weng[12] J. Whittaker[4], F. Zhang[12], M. Zupanski[2]

30 March 2011

[1]Atlantic Oceanographic and Meteorological Laboratory OAR/NOAA
[2]Colorado State University
[3]Environmental Modeling Center NCEP/NOAA
[4]Earth System Research Laboratory OAR/NOAA
[5]Florida State University
[6]Geophysical Fluid Dynamics Laboratory OAR/NOAA
[7]National Center for Atmospheric Research
[8]National Environmental Satellite Data Information Center NOAA
[9]National Hurricane Center NWS/NOAA
[10]Naval Research Laboratory, Monterey
[11]Office of Science and Technology, NWS/NOAA
[12]Pennsylvania State University
[13]University of Rhode Island
[13]University of Colorado

TABLE OF CONTENTS

2010 Hurricane Forecast Improvement Program (HFIP) R & D Activities Summary: Accomplishments, Lessons Learned, and Challenges

1. Executive Summary

This report describes the activities and results of the Hurricane Forecast Improvement Program in 2010. It is organized around three themes, *Encouraging results* from 2010 HFIP testing, evaluation and development activities, *lessons learned* from negative results and the *challenges* faced by the program to achieve its goals. The main topics from each of these three categories are:

1.1. Encouraging Results

- *The operational HWRF codes were fully synchronized with the community HWRF repositories, allowing operations and research to work from a common code.*

- *The operational GFS model initialized with EnKF gave the best track forecasts in the Atlantic out to 5 days even beating the ECMWF.* *The improvement exceeded the 20% 5 year goal for HFIP on track demonstrating the potential to reach this goal.*

- *High resolution global ensembles (30 km, 20 members) can be run in real time on available computing resources meaning higher resolution is definitely possible.*

- *Using all aircraft data (including tail radar data) collected in 57 cases over 3 years, the Penn State regional model with an EnKF data assimilation scheme improved intensity forecasts beyond 24 hours by more than 20% relative to the HFIP baseline demonstrating the potential to reach the HFIP 5 year intensity goal.*

- *The multi-model regional ensemble using a "Correlation Based Correction" scheme substantially outperformed any of its components and showed a 40-50% gain over the 5-day Statistical Hurricane Intensity Forecast model. (SHF5).*

- *When run as a seasonal prediction system, the Cubed Sphere model produced a remarkable prediction of the number of hurricanes and tropical storms for the period 2000-2010. This is powerful evidence of the ability of the global models to forecast genesis.*

- *For some 2010 storms, the possibility of using ensemble information to narrow the "cone of uncertainty" from the operational cone for track was demonstrated.*

- *The stream 1.5 testing and evaluation concept was successfully demonstrated. Four stream 1.5 candidate models were evaluated through a retrospective analysis study. Two of the models, AHW and GFDL, were accepted by NHC as part of stream 1.5 demonstration during the 2010 hurricane season.*

1.2. Lessons Learned

- *The major problem with initialization of regional models identified last year remains. It is a particularly severe problem for the first 24 hours of the forecast, which results in a severe problem in forecasting rapid intensification.*

- *Global models consistently provide significantly better track forecasts than regional models.*

- *Advanced data assimilation systems, particularly EnKF, are essential for improving both track and intensity forecasts.*

- *Statistical post processing of ensemble output is likely to provide additional skill improvement.*

- *The delivery of complete, consistent Tier 1 (ATCF files) and Tier 2 (gridded model) data is necessary to provide consistent evaluations of the model forecasts for the same forecast challenge using a common (operational) tracker.*

1.3. Challenges

- *Initialization systems for regional models need to be improved for forecasts in the first 24 hours.*

- *HFIP needs to develop systems to make better use of satellite data in regional models.*

- *Implementing the Hybrid data assimilation (DA) system in operations as soon as possible.*

- *Development of appropriate statistical post processing systems for ensemble output.*

- *Operational computing not likely to be sufficient to meet HFIP goals within 3 years.*

In the following sections, we discuss each of the above lessons and challenges and our plans for addressing each.

2010 HFIP R & D Activities Summary:
Accomplishments, Lessons Learned, and Challenges

2. *Background on HFIP*

HFIP provides the basis for NOAA and other agencies to coordinate hurricane research needed to significantly improve guidance for hurricane track, intensity, and storm surge forecasts. It also engages and aligns the inter-agency and larger scientific community efforts towards addressing the challenges posed to improve hurricane forecasts. The goals of the HFIP are to improve the accuracy and reliability of hurricane forecasts; to extend lead time for hurricane forecasts with increased certainty; and to increase confidence in hurricane forecasts. These efforts will require major investments in enhanced observational strategies, improved data assimilation, numerical model systems, and expanded forecast applications based on the high resolution and ensemble-based numerical prediction systems.

HFIP began in 2009 with specific goals to reduce the average errors of hurricane track and intensity forecasts by 20% within five years and 50% in ten years with a forecast period out to 7 days. In addition, HFIP has a goal to increase the probability of detecting rapid intensification at day 1 to 90% and 60% at day 5. The benefits of HFIP will significantly improve NOAA's forecast services through improved hurricane forecast science and technology. Forecasts of higher accuracy and greater reliability (i.e., user confidence) are expected to lead to improved public response, including savings of life and property.

NOAA recognizes that addressing the broad scope of the research and technology challenges associated with improving hurricane forecasts requires interaction with, and support of, the larger research and academic community. It is hypothesized that these very ambitious goals of the HFIP can only be met using high-resolution (~5-15 km) global atmospheric forecasting numerical models run as an ensemble in combination with regional models at even higher resolution (~1-5 km). Demonstrating this is very expensive computationally and hence HFIP has been building up a computational system at the NOAA facilities in Boulder, Colorado, called t-jet, where HFIP can demonstrate the techniques necessary to meet its goals. Only by demonstrating the value of high resolution is there any opportunity to obtain such a computational resource for operational hurricane forecasts.

For FY10, the HFIP program consisted of about $23 M with $3 M dedicated to enhancing computer capacity available to the HFIP program. The funding for computing was used to enhance the t-jet system by the addition of processors. The t-jet system currently has 10,000 processors available for HFIP computing. About $10M of the $23M is part of the base funding for the Atlantic Ocean and Meteorology Laboratory (AOML) in Miami and the Environmental Modeling Center (EMC) at NCEP for hurricane model development. The remaining $11M was distributed to various NOAA laboratories and centers (Earth System Research Lab (ESRL), Geophysical Fluid Dynamics Laboratory (GFDL), National Environment Satellite Data and Information Service (NESDIS), and National Hurricane Center (NHC). Funding was also provided to the National Center for Atmospheric Research (NCAR), Naval Research Laboratory in Monterey (NRL), and several universities: University of Wisconsin, The Pennsylvania State University, Colorado State University, University of Arizona, Florida State University, University of Wisconsin, University of Miami, and State University of New York-Albany, and

University of Rhode Island. Finally, $1M was contributed for second year funding to the National Oceanographic Partnership Program (NOPP), Announcement of Opportunity for competed proposals related to improving understanding and prediction of hurricanes. The funding to NOPP from HFIP was matched by funding from the Office of Naval Research (ONR).

Distribution of the $11M was accomplished through recommendations from 8 teams focused on various components of the hurricane forecast problem.   Current teams are listed in Appendix A with the team leads shown in bold type. For reference, a list of acronyms for the various organizations is shown in Appendix B. These teams are made up of over 50 members drawn from the hurricane research, development, and operational community.

HFIP is primarily focused on techniques to improve the numerical model guidance that is provided by NCEP operations to NHC as part of the hurricane forecast process. It is organized along two paths of development called Streams (see Table 1). Stream1 assumes that the computing power available for operational hurricane forecast guidance will not exceed what is already planned by NOAA. The development for this stream has been in planning for several years by EMC. HFIP activities at the NOAA labs and centers will help accelerate this development.

Table 1. The HFIP Stream Strategy.

Stream	Description
Stream 1	Development to directly improve the current operational global and regional hurricane models. Assumes that the computing that will be available for operations is that currently being planned.
Stream 2	Assumes that operational computing can be substantially increased above current plans. Seeks computing resources from major supercomputing centers for testing and evaluation. Emphasis is on high resolution global and regional models run as ensembles. It will include a demonstration system run in real time each summer to test and evaluate promising new technology.
Stream 1.5	This will be part of the summer demonstration system and will be forecaster defined. Components from Stream 2 that forecasters see as particularly promising in one year will be configured to run in real time the next year, with products made available to NHC.

HFIP Stream 2 does not put any restrictions on the increases in computer power available to NWS operations, and in fact, assumes that resources will be found to greatly increase available computer power in operations above that planned for the next 5 years. The purpose of Stream 2, therefore, is to demonstrate that the application of advanced science and technology developed under the auspices of HFIP along with increased computing will lead to the expected increase in accuracy and other aspects of forecast performance. Because the level of computing necessary to perform such a demonstration is so large, the Program is applying to resources outside NOAA in addition to trying to increase internal computing for development.

A major part of Stream 2 is a demonstration system, or Demo System, that is run in testing mode each hurricane season. The purpose of this system is to evaluate strengths and weaknesses of promising new technology. As a result of the Demo System testing, some components may be

found to be of particular interest to the operational forecasters, and, if resources do not permit its implementation in the operational infrastructure, the Demo system for the following season will emphasize those components and will provide specific output that is made available to NHC forecasters for evaluation. We refer to this component of the Demo System as Stream 1.5.

Roughly half of the HFIP funding is going toward Stream 2 development activities. In Stream 2 we are assuming that the best approach to improving the forecast hurricane track beyond 4 days is through the use of high resolution global models run as an ensemble. We describe below the logic behind this assumption. For improvements in forecast of hurricane intensity, especially in the 1 to 4 day time range, the best approach is likely to be high resolution regional models, also run as an ensemble. The global models are likely to be limited in resolution to about 10 km for at least the next 5 years, because of computer limitations, especially when they are run as an ensemble. Thus the only way to achieve the very high resolution of about 1 km necessary for resolving the inner core of the hurricane is with regional models. It is generally assumed that the inner core must be resolved before we can expect to see consistently accurate hurricane intensity forecasts.

To facilitate the transition of research to operations, HFIP has recognized the importance of having research and operations share the same code base, and HFIP has co-sponsored the Developmental Testbed Center (DTC) to make available and support HWRF to the community. This support started in February 2010 with the DTC/EMC/MMM Joint Hurricane Workshop and WRF for Hurricanes Tutorial. During FY10, EMC and DTC established a common version of the operational codes for HWRF in the community code repositories. This version has been thoroughly tested and shown to give comparable results to the operational codes. Thus with the repository in Boulder at the DTC, both the operations and research community will be drawing from the same central code repository facilitating the transitions of research results to operations.

3. The HFIP Model Systems

3.1. The Global Models:

FIM – Refers to the Flow-following finite-volume Icosahedral Model. The FIM is an experimental global model that can be run at various resolutions and uses initial conditions from a number of sources. It is currently using a fixed ocean underneath. It has been built by the NOAA Earth System Research Laboratory (ESRL).

Cubed Sphere – Refers to a finite volume model developed by GFDL using a grid where the sphere is projected onto a cube. The governing finite difference equations are very similar to those used in the FIM

GFS – Refers to the Global Forecast System. There are two versions of this model currently running in the demonstration system. This includes a version of the current operational model run at the NOAA National Centers for Environmental Prediction (NCEP) and an experimental version of that model at ESRL. The main difference between the two versions is the initialization system; GSI for the operational model and EnKF for the ESRL version.

NOGAPS – Refers to Navy Operational Global Atmospheric Prediction System. Currently a semi-Lagrangian version of NOGAPS is being developed, which will allow for efficient high-resolution forecasts.

A summary of some of the main attributes of the global models is shown in Table 2.

Table 2. Specifications of the HFIP Global Models.

Models	Horizontal resolution	Vertical levels	Cumulus Parameterization	Microphysics	PBL	Land Surface	Radiation	Initialization
FIM	20 km	64	From 2010GFS - Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	Noah LSM	GFDL/RRTM	ESRL EnKF
GFS	27 km	??	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	Noah LSM	GFDL scheme	ESRL EnKF
Cubed Sphere	25 km	32	Shallow only	Modified Lin, 6-class	Lock (AM2)	GFDL LM3	GFDL	nudging to NCEP analysis
NOGAPS	41 km	42	Emanuel	N/A	NOGAPS	NOGAPS	Harshvardhan/ Fu-Liou	NAVDAS-AR

3.2. The Regional Models:

WRF – Refers to Weather Research and Forecasting model. This is actually a modeling system with options for the dynamic core (ARW—Advanced Research WRF built by NCAR and NMM—Non-hydrostatic Mesoscale Model, built by EMC) and several options for physics as well as initialization systems, post processing systems, and verification systems.

The NCEP Hurricane WRF (HWRF) is based on the Non-hydrostatic Mesoscale (NMM) dynamic core and has a movable, two-way nested grid capability for the 9 km inner nest. The coarse domain has 27-km grid spacing and covers a 75° x 75° region with 42 vertical layers. Advanced physics include atmosphere/ocean fluxes, coupling with the Princeton Ocean Model (POM) and the NCEP GFS boundary layer and deep convection.

Two configurations of the WRF ARW system were also run. The ARW system run by NCAR used a simplified one dimensional model of the ocean with two interactive nests within the outer regional model. FSU also ran a version of the ARW without an interactive ocean.

The Penn State Regional Ensemble – This was another version of the WRF ARW system similar to the NCAR WRF ARW. It used a static interactive inner nest but no interactive ocean. It was run as a 30 member ensemble.

COAMPS-TC – It is a Navy model run by NRL Monterey. It is a version of their COAMPS regional prediction system that is being run operationally and has an interactive ocean. The model run for 2010 had horizontal resolution of 45, 15, and 5 km for the three nested grids.

The Multi-Model Ensemble – The multi-model ensemble is organized by Florida State University and was made up of a total of 7 models run by different organizations. The various models and their resolution are indicated in Table 3. Two of the members are the operational models, GFDL at 7.5 km and HWRF at 9 km. GFDL is the old operational model that is still being run in parallel with the current operational model HWRF. HWRF is constructed from the NMM core of the WRF (see above), and both GFDL and HWRF models are coupled to the Princeton Ocean Model (POM) in the Atlantic Basin. HWRF-x is an experimental version of the operational HWRF run by the Hurricane Research Division of OAR. It did not have an interactive ocean model associated with it but did have an interactive nest. GFDL Parallel, also referred to as GFD5 is a somewhat updated version of the GFDL operational model and run at the same resolution.

Table 3. Regional models run in the Demonstration System during 2010.

Models	Nesting / Horizontal Resolution (km)	Vertical Levels	Cumulus Parameterization	Microphysics	PBL	Land Surface	Radiation	Initial and Boundary Conditions	Atmospheric Initialization
HWRF (OPS)	2 27/9	42	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	GFDL Scheme	GFS	GSI 3DVAR with cycled vortex initialization
GFDL (OPS)	3 30/15/7.5	42	Arakawa Schubert	Ferrier	GFS Non-Local PBL	Slab Model	Schwarz-kopf-Fels Scheme	GFS	GFDL Synthetic Bogus Vortex
HWRF IC	2 27/9	42	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	GFDL Scheme	GFS	GSI 3DVAR with inner core data
HWRF -x	2 9/3	42	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	GFDL Scheme	GFS	HWRF (OPS)
HWRF HYCOM	2 27/9	42	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	GFDL Scheme	GFS	HWRF (OPS)
HWRF-HRD EnKF DA	2 9/3	42	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	GFDL Scheme	GFS	EnKF with aircraft data
HWRF NOAH LSM	2 27/9	42	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	NOAH LSM	GFDL Scheme	GFS	HWRF (OPS)
GFDL Parallel	3 30/15/7.5	42	Arakawa Schubert	Ferrier	GFS Non-Local PBL	Slab Model	Schwarz-kopf-Fels Scheme	GFS	GFDL synthetic Bogus Vortex
WRF ARW FSU	2 12/4	40	Simplified Arakawa Schubert	WSM5	YSU	5-Layer Thermal Diffusion soil Model	RRTM (longwave) / Dudhia (shortwave)	GFS (initial and boundary condition)	Initialized from GFS
WRF ARW (NCAR)	3 12/4/1.3	36	New Kain Fritsch (12 km only)	WSM5	YSU	5-Layer Thermal Diffusion soil Model	RRTM (longwave) / Dudhia (shortwave)	GFS	EnKF method in a 6-hour cycling mode
WRF ARW Utah	3 27/9/3	31	Betts-Miller	Lin	MYJ	5-Layer Thermal Diffusion Scheme	RRTM	Dudhia	WRF 3DVAR
COAMPS-TC	3 45/15/5 (15/5 km following the storm)	40	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	Harshvardhan	NOGAPS and GFS	3D-Var data assimilation with synthetic observations
Wisconsin Model	UW NMS (3D enstrophy/ entropy/KE conserving dynamics core)	3-4 90/45/9 km 90/45/9/3 km	52	Modified Emanuel	Explicit bulk microphysics (cloud/rain/pristine/ aggregate/graupel)	1.5 Order Closure	WRF vegetation/land surface/Andreas emulsion layer	RRTM	GFS/GFDL
Penn State ARW	3 40.5/13.5/4.5 for ensemble forecast 1.5-km nest for control	35	Grell-Devenyi ensemble scheme (40.5 km only)	WSM 6-class graupel scheme	YSU	5-layer thermal diffusion scheme	RRTM (longwave) / Dudhia (shortwave)	GFS	EnKF with NOAA airborne radar

8

3.3. Initialization systems:

A number of approaches were used to create the initial state for the global and regional models in the experiments, which are described below. The choices include:

- The initial state created for the current operational model (Global Forecast System or GFS) interpolated to the higher resolution grid. The GFS uses the Grid point Statistical Interpolation (GSI) initialization system that has run operationally for many years. It is a three-dimensional variational approach (3D-VAR).
- NRL Atmospheric Variational Data Assimilation System (NAVDAS). This is the system used to provide the initial conditions to the Navy global model. It has been a 3D-VAR system but starting late September 2009, it was upgraded to NAVDAS-AR (for accelerated representor), a four-dimensional variational approach (4D-VAR). The COAMPS-TC makes use of NAVDAS as well as synthetic observations to initialize the tropical cyclone structure and intensity.
- Ensemble Kalman Filter (EnKF). This is also an advanced data assimilation approach (somewhat like 4D-VAR) that uses an ensemble to create background error statistics for a Kalman Filter. While this approach is still in the experimental stage in the U.S. (though operational in Canada), it has shown considerable promise.
- Hybrid Variational-Ensemble Data Assimilation System (HVEDAS). This system combines aspects of the EnKF and 3D- or 4D-VAR for example, using the ensemble of forecasts to estimate the covariances at the start of a 4D-VAR assimilation window. This technology is under development at NOAA/NCEP/EMC and NOAA/OAR/ESRL. It was not ready for testing in the 2010 season but may be available for subsequent seasons. This hybrid approach is likely to define the operational global data assimilation system for NOAA in the 5-year time frame.
- The initial state for the regional models was generally produced by downscaling the global models' analysis and forecasts. In addition, the Penn State Regional Ensemble model, the WRF/ARW/NCAR model used an EnKF initialization system.

- The operational HWRF utilizes an advanced vortex initialization and assimilation cycle consisting of four major steps: 1) interpolate the global analysis fields from the Global Forecast System (GFS) onto the operational HWRF model grid; 2) remove the GFS vortex from the global analysis; 3) add the HWRF vortex modified from the previous cycle's 6-hour forecast (or use a synthetic bogus vortex for cold start); and 4) add satellite radiance and other observation data in the hurricane area (9 km inner domain) – no observations are used in the inner core of the storm. The major differences from the GFDL model initialization are steps 3) and 4).

4. The HFIP Baseline

HFIP 10-year goals for Atlantic guidance were set by the HFIP Executive Oversight Board (HEOB) as:
- Reduce average track errors by 50% for days 1 through 5
- Reduce average intensity errors by 50% for days 1 through 5
- Increase the probability of detection (POD) for rapid intensity change to 90% at Day 1 decreasing linearly to 60% at day 5, and decrease the false alarm ratio (FAR) for rapid

intensity change to 10% for day 1 increasing linearly to 30% at day 5. The focus on rapid intensity change is the highest forecast challenge indentified by the NHC.
- Extend the lead time for hurricane forecasts out to Day 7 (with accuracy of Day 5 forecasts in 2003).

HFIP 5-year goals are to improve track and intensity guidance errors by 20% over the next 5 years.

To measure progress toward meeting these goals, HFIP established a baseline against which results from experimental and operational HFIP model guidance will be measured. These HFIP Performance Goals Baselines were developed in a white paper authored by James Franklin dated 5 May 2009 and summarized here.   Tables 4 and 5 define the baselines for track and intensity, respectively. CONS in Table 4 is a consensus of operational models and is the primary baseline for measuring improvements. OCD5 is a simple statistical model baseline for measuring improvements in forecast skill, and OFCL is the average NHC official forecast errors from the baseline period.

Table 4. HFIP Track Performance Baseline (nautical miles).

VT (h)	N	OFCL	OCD5	CONS
0	818	7.4	7.7	7.8
12	741	29.4	44.5	30.0
24	663	49.6	93.3	49.8
36	586	69.9	150.9	69.5
48	518	91.2	212.2	89.6
72	411	135.0	317.2	132.0
96	313	173.0	396.5	175.2
120	247	218.6	473.0	221.9

Table 5. HFIP Intensity Performance Baseline (knots).

VT (h)	N	OFCL	OCD5	CONS
0	820	1.9	2.2	2.2
12	745	7.2	8.3	7.7
24	667	10.4	11.5	10.1
36	590	12.6	14.2	11.7
48	522	14.6	16.1	13.7
72	415	17.0	17.8	16.0
96	316	17.5	19.3	16.6
120	250	19.0	19.3	17.0

5. *FY2010 Demonstration System*

During hurricane season of 2010 HFIP ran a number of models in real-time on its t-jet computer in Boulder. Much of the results reported below were from that demonstration system. During the period just after last year's hurricane season and the start of the season in 2010, t-jet was used largely for development of the systems used in 2010 and for retrospective runs in part to establish the stream 1.5 components and for the development of statistics for the Multi Model Ensemble. An example of results for the retrospective runs is shown in figure 4 and figures 5 and 6 for the Multi Model Ensemble.

As discussed before, Table 2 lists details of the global models in use by HFIP. Table 6 lists those that took part in the HFIP 2010 demonstration system.

Table 6. Global Models run in the 2010 Demonstration System.

Global Models Run in 2010
FIM Deterministic models • **30 km (Initialized with EnKF)** **Global Ensembles:** • **60 km GFS (initialized with EnKF) 20 members** • **55 km NOGAPS (4DVAR), 9 members**

As shown above, Table 3 lists the various regional models used by HFIP during this last year. Included in this list are all the models used in the multi-model ensemble (MMEN). Table 7 briefly lists those models that were part of the MMEN. Other regional models were part of stream 1.5 (WRF ARW NCAR, and GFD5). The GFDL model in addition was run as a 20 member ensemble by varying the initial conditions provided to the model. Table 8 lists the various members of the GFDL ensemble.

Table 7. Multi Model Mesoscale Ensemble.

Multi Model Regional Ensemble	
• **HWRF**	**9 km**
• **GFDL**	**7.5 km**
• **HWRF-x (H3HW)**	**3 km**
• **WRF/ARW/NCAR (AHW1)**	**1.3 km**
• **WRF/ARW/FSU (ARFS)**	**4 km**
• **COAMPS-TC (COTC)**	**5 km**
• **GFDL parallel (GFD5)**	**7.5 km**

Table 8. Eleven member ensemble: 10 perturbed members and a control forecast.

GFDL Regional Ensemble
• **GP0: Control forecast (GFD5 run on Jet)**
• **GPA: Unbogussed forecast**
• **GPB: GFD5 with no asymmetries included**
• **GPC: GFD5, but with the use of old environmental filter**
• **GPD: Increase storm size (ROCI-based) by 25%**
• **GPE: Decrease storm size (ROCI-based) by 25%**
• **GPF: Increase wind radii 25%, increase storm size 25%**
• **GPG: Decrease wind radii 25%, decrease storm size 25%**
• **GPH: Old filter (GPC), plus both size increases from GPF**
• **GPJ: Old filter (GPC), plus both size decreases from GPG**
• **GPK: Set Rmax minimum to 45 km (GFD5 control uses 25 km)**

6. Encouraging Results from FY10

Table 9 gives a high level overview of some encouraging results from the HFIP work during FY10. A more detailed description for each encouraging results is further expanded upon in the discussion below.

Table 9. Encouraging Results from FY10 Demonstration System.

- *The operational HWRF codes were fully synchronized with the community HWRF code repositories, allowing operations and research to work from a common code.*

- *The operational GFS model initialized with EnKF gave the best track forecasts in the Atlantic out to 5 days even beating the ECMWF. This was achieved with a relatively low resolution version of the GFS (T256~60km). The improvement exceeded the 20% 5 year goal for HFIP on track demonstrating the potential to reach this goal.*

- *High resolution global ensembles (30 km, 20 members) can be run in real time on available computing resources, which mean higher resolution is definitely possible.*

- *Using aircraft tail radar data collected in 57 cases over 3 years, the Penn State ARW regional model with an EnKF data assimilation scheme improved intensity forecasts beyond 24 hours by more than 20% relative to the HFIP baseline demonstrating the potential to reach the HFIP 5 year intensity goal.*

- *The multi-model regional ensemble using a "Correlation Based Consensus" scheme substantially outperformed any of its components and showed a 40-50% gain over the 5-day Statistical Hurricane Intensity Forecast model. (SHF5). This demonstrates the possibility that statistical models combined with use of additional data (especially aircraft data) can give even greater improvements over the HFIP 5 year goal.*

- *When run as a seasonal prediction system, the Cubed Sphere model produced a remarkable prediction of the number of hurricanes and tropical storms for the period 2000-2010. This is powerful evidence of the ability of the global models to forecast genesis.*

- *For some 2010 storms the possibility of using ensemble information to narrow the "cone of uncertainty" from the operational cone for track was demonstrated.*

- *The stream 1.5 testing and evaluation concept was successfully demonstrated. Four stream 1.5 candidate models were evaluated through a retrospective analysis study. Two of the models, AHW and GFDL, were accepted by NHC as part of stream 1.5 demonstration during the 2010 hurricane season.*

The operational HWRF codes were fully synchronized with the community HWRF code repositories, allowing operations and research to work from a common code.

During this past year both EMC and DTC worked to update the operational version of HWRF from version 2.0 to the current version of WRF (version 3.2.1). This makes the operational model completely compatible with the codes in the community code repositories and vice versa. This 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 now been fully implemented for the first time with the hurricane models.

Figure 1 shows schematically how the process of moving the operational model into the community WRF code repository was achieved. The new version 3.21 codes were carefully compared to the original operations code until the two versions gave comparable answers. Subsequent changes to the operational code will now be entered directly into the community code repositories.
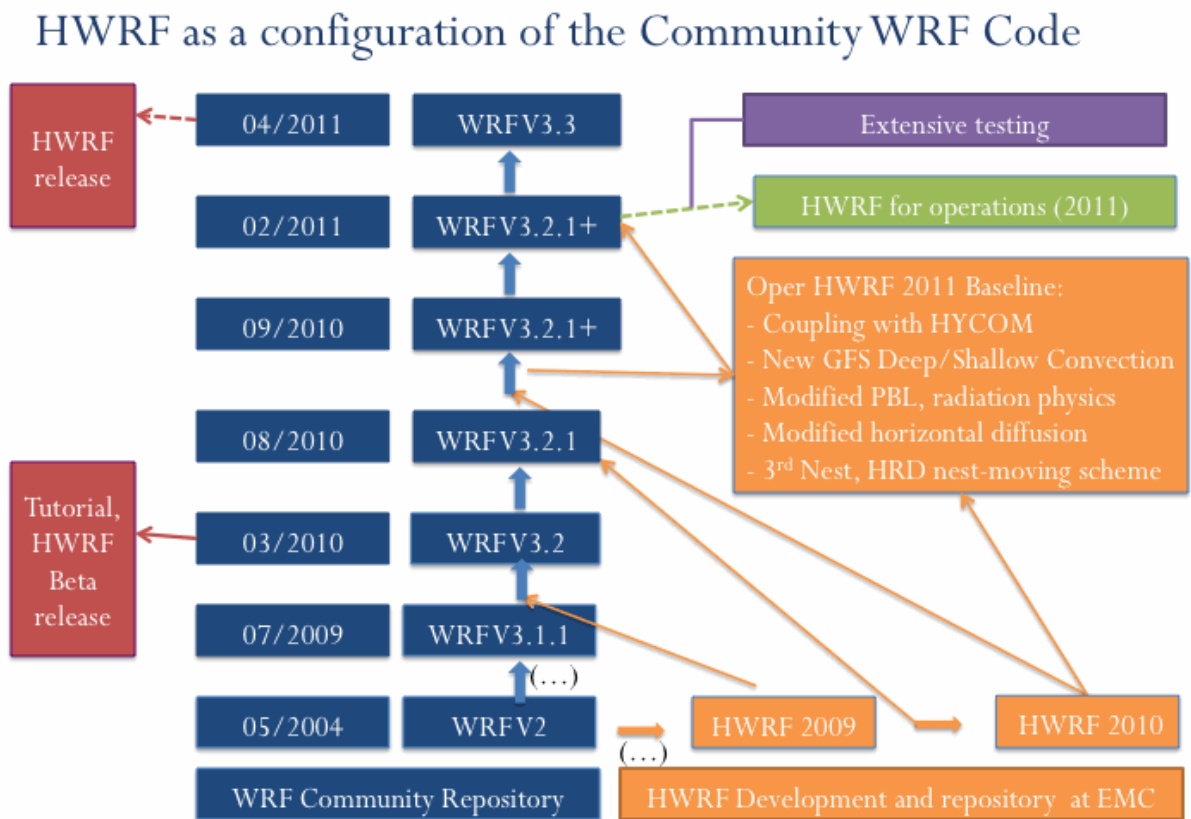

Figure 1. Schematic of the process for integrating the atmospheric forecast model component of the HWRF code onto the WRF community code repository.

The operational GFS model initialized with EnKF gave the best track forecasts in the Atlantic out to 5 days even beating the ECMWF.

During the 2010 Demo project run on the HFIP t-jet system, a 20 member ensemble T256 (~60km) version of the operational GSF global model was run. It is identical to the operational GFS including its physics package. The only exception was a different initialization system was used. For the operational GFS (deterministic T574 ~29km), the operational GSI (essentially 3DVAR) data assimilation system was used for initialization. For the HFIP ensemble, an EnKF approach was used. The track verification for the Atlantic season is shown in Figure 2. Note:

- Between 36 and 72 hours of forecast time the HFIP ensemble in green equals or exceeds the HFIP goal of a 20% improvement over the HFIP baseline track errors that it is trying to achieve within 5 years (we are in the second year of that 5 year period). This demonstrates that given the computer resources to run such a global ensemble operationally, HFIP may achieve its track goals at least out to 5 days.
- In the Atlantic the HFIP ensemble performed significantly better than the much higher resolution ECMWF T1299 ~15km model as well as the higher resolution operational GFS T547 ~20 km.
- The global models performed considerably better on track at all lead times than the two operational regional models, HWRF and GFDL. This tendency for the global models to outperform the regional models on track has been frequently noted.

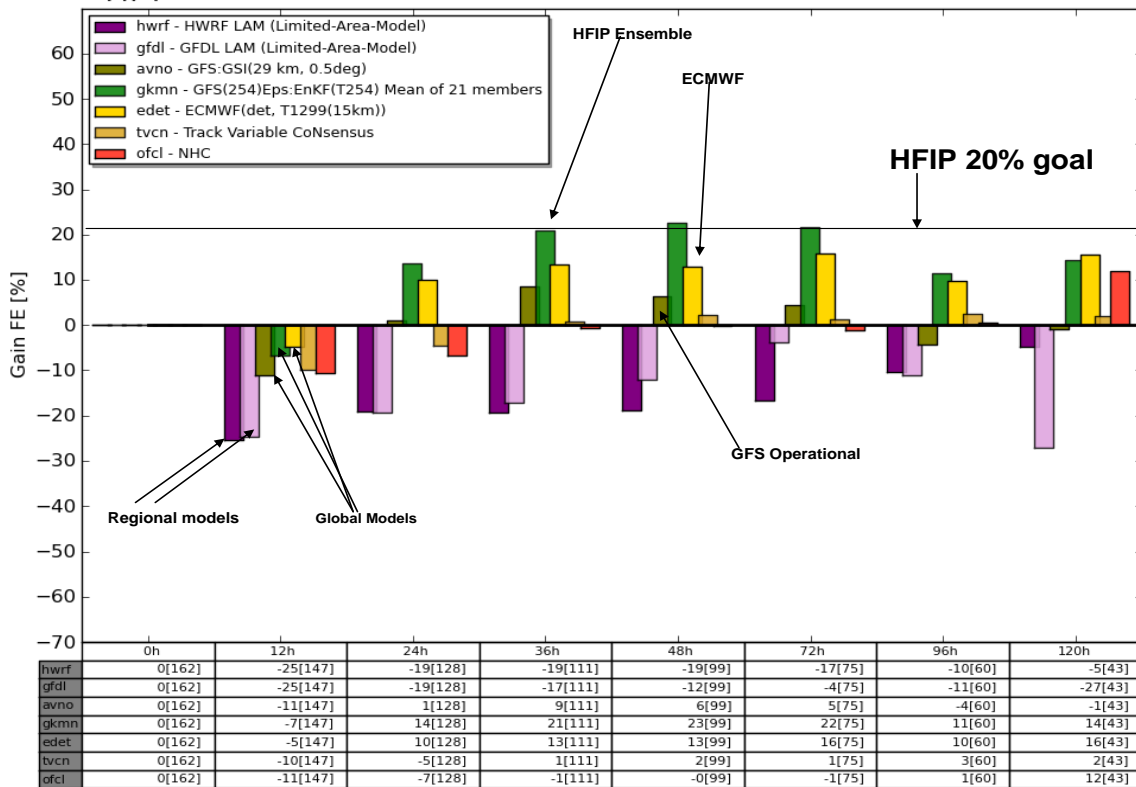|	0h	12h	24h	36h	48h	72h	96h	120h
hwrf	0[162]	-25[147]	-19[128]	-19[111]	-19[99]	-17[75]	-10[60]	-5[43]
gfdl	0[162]	-25[147]	-19[128]	-17[111]	-12[99]	-4[75]	-11[60]	-27[43]
avno	0[162]	-11[147]	1[128]	9[111]	6[99]	5[75]	-4[60]	-1[43]
gkmn	0[162]	-7[147]	14[128]	21[111]	23[99]	22[75]	11[60]	14[43]
edet	0[162]	-5[147]	10[128]	13[111]	13[99]	16[75]	10[60]	16[43]
tvcn	0[162]	-10[147]	-5[128]	1[111]	2[99]	1[75]	3[60]	2[43]
ofcl	0[162]	-11[147]	-7[128]	-1[111]	-0[99]	-1[75]	1[60]	12[43]

Figure 2. Average track errors for various models relative to the HFIP baseline in 2010 in the Atlantic. All storms in 2010 are included. HFIP 20% goal refers to the 5 year goal of reducing track errors by 20% relative to the HFIP baseline defined earlier.

Figure 3 shows the same as Figure 2, but for the West Pacific. We note the following result:

- The HFIP ensemble performance is similar to the Atlantic.

15

- The ECMWF model performs better than the HFIP ensemble at least out to 72 hours. In the Pacific in 2010, many of the typhoons were strongly affected by mid-latitude systems. It is known that the ECMWF is known to perform better in these situations. This is likely due to its much higher resolution and therefore the better resolution of mid-latitude frontal processes.
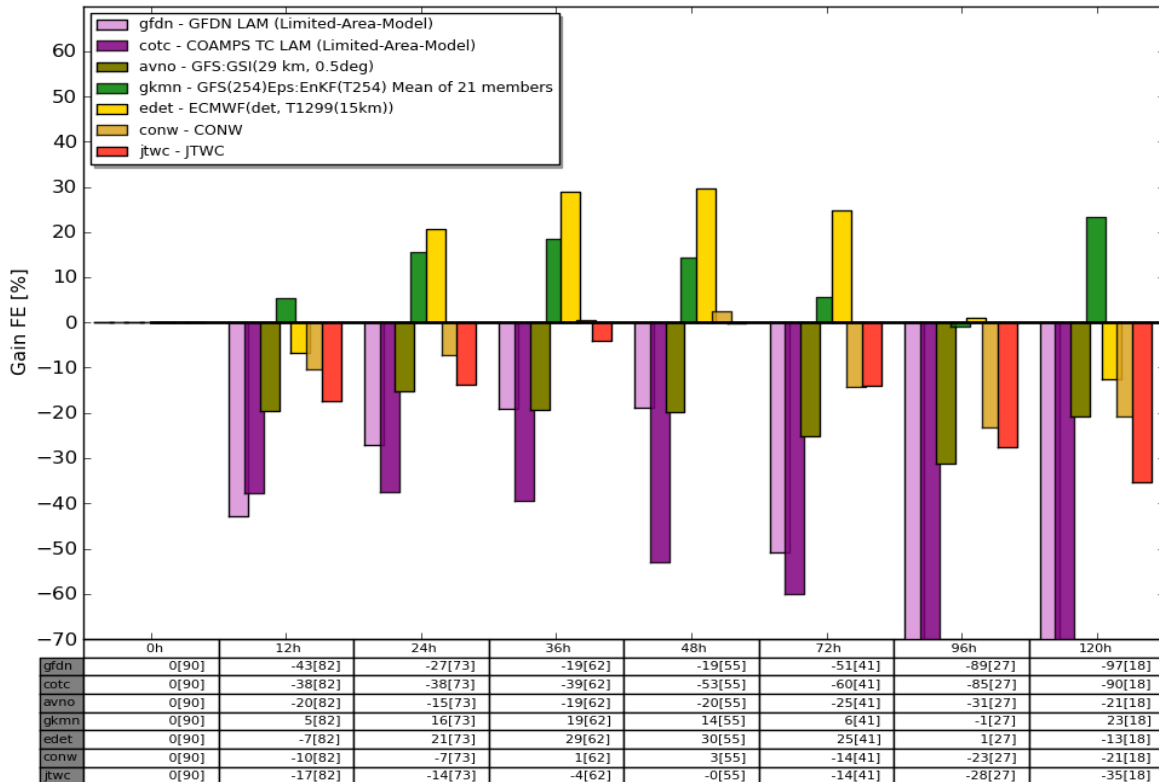- Again, the global models outperform the regional models on track.

Figure 3 legend:
- gfdn - GFDN LAM (Limited-Area-Model)
- cotc - COAMPS TC LAM (Limited-Area-Model)
- avno - GFS:GSI(29 km, 0.5deg)
- gkmn - GFS(254)Eps:EnKF(T254) Mean of 21 members
- edet - ECMWF(det, T1299(15km))
- conw - CONW
- jtwc - JTWC

	0h	12h	24h	36h	48h	72h	96h	120h
gfdn	0[90]	-43[82]	-27[73]	-19[62]	-19[55]	-51[41]	-89[27]	-97[18]
cotc	0[90]	-38[82]	-38[73]	-39[62]	-53[55]	-60[41]	-85[27]	-90[18]
avno	0[90]	-20[82]	-15[73]	-19[62]	-20[55]	-25[41]	-31[27]	-21[18]
gkmn	0[90]	5[82]	16[73]	19[62]	14[55]	6[41]	-1[27]	23[18]
edet	0[90]	-7[82]	21[73]	29[62]	30[55]	25[41]	1[27]	-13[18]
conw	0[90]	-10[82]	-7[73]	1[62]	3[55]	-14[41]	-23[27]	-21[18]
jtwc	0[90]	-17[82]	-14[73]	-4[62]	-0[55]	-14[41]	-28[27]	-35[18]

Figure 3. Same as figure 2 but for the Western Pacific. In this case, the regional models (in shades of purple) are the Navy regional models.

High resolution global ensembles (30 km, 20 members) can be run in real time on available computing resources meaning higher resolution is definitely possible.

A major accomplishment of the HFIP program over the last two years has been the development of a research computer system located at the NOAA facilities in Boulder, Colorado. This system is referred to as t-jet and is devoted to HFIP testing, evaluation and development. Currently, it consists of about 10,000 processors and is about equivalent to all of the operational computing at NCEP. During the demonstration program conducted during the summer hurricane season, it is focused on real-time experimental forecasts. In 2010, this included a 20 member 60 km global ensemble, results from which are displayed in the previous section. HFIP is planning to increase the number of processors on t-jet for the 2011 season to 16,000 processors. It will be possible given our experience with the 60 km global ensemble to increase the resolution to 30 km and run the global model ensemble with an EnKF initialization in real-time.

We note that the total cost of t-jet when it has been increases to 16,000 processors will have been $10M so it seems reasonable that even a relatively modest increase in size of a computer like t-jet would allow global ensembles to reach at least 20 km resolution and perhaps considerably

16

higher. The global models are useful for forecasts far beyond hurricane forecasting so we feel high resolution global ensembles will be the future centerpiece of forecasting in general.

Using aircraft tail radar data collected in 57 cases over 3 years, the Penn State ARW regional model with an EnKF data assimilation scheme improved intensity forecasts beyond 24 hours by more than 20% relative to the HFIP baseline.

This was the first demonstration within HFIP to show that it will be possible to achieve its 5 year intensity goal. Figure 4 shows the results from the PSU ensemble when all aircraft data including the tail radar data are included in the initialization. We note the following from the results shown in Figure 4.

- Beyond 24 hours the errors meet or exceed the 20% goal set by HFIP as its 5 year intensity goal for storms for which there are aircraft radar data.
- The intensity errors at 0 hours are roughly the same as the errors later on. The initial errors are a result of the ensemble spread in creating the initial states and it apparently takes about 24-36 hours for the initial error to settle down.
- The right hand panel illustrates the possible use of statistical post processing to improve the forecasts. Here a very simple bias correction is used that removes the initial bias and uses this correction out to 36 hours with the correction being reduced linearly to zero at 36 hours.
- Note that very similar results to those shown in figure 4 (but for just the 2010 season) were obtained by HRD using their version of the HWRF model and EnKF DA.
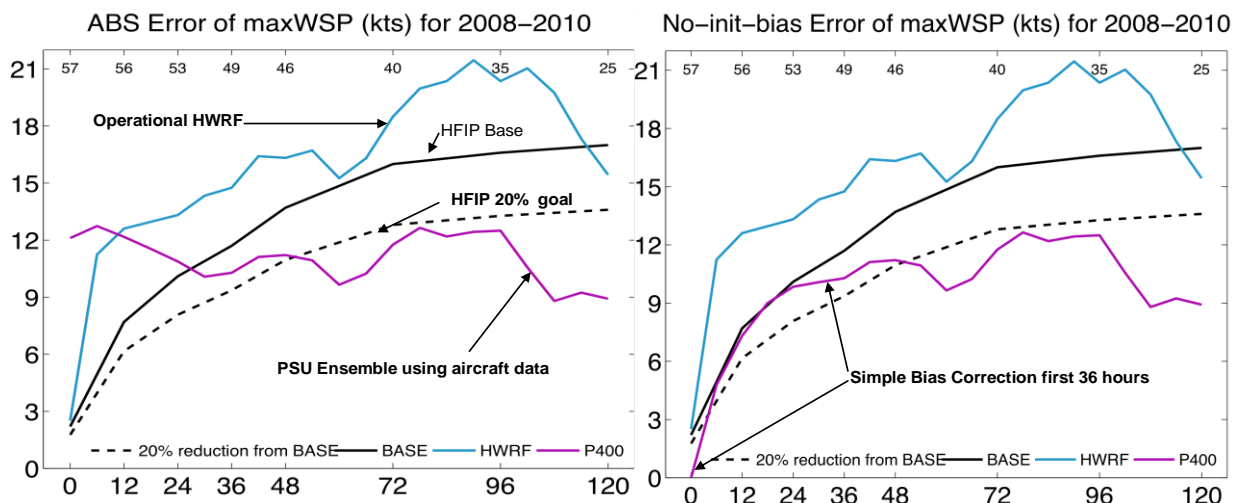
Figure 4. Intensity error from forecasts with the PSU ensemble run at 4 km and initialized with an EnKF system using all available aircraft radar data including radar data. All storms with aircraft tail radar data for three seasons, 2008-2010 are included in the statistics, purple line. Blue line is the forecast from the operational HWRF, black line is the HFIP baseline for track and the black dashed line represents a 20% improvement over the baseline. Numbers at the top of the figures indicate the number of cases. The left figure shows the raw forecasts in the purple line. The right figure is the same except that a simple bias correction is applied that reduces the bias to zero at zero hour and this correction is applied to other times decreasing linearly to zero at 36 hours.

17

The multi-model regional ensemble using a "Correlation Based Consensus" scheme substantially outperformed any of its component models and showed a 40-50% gain over the 5-day Statistical* Hurricane Intensity *Forecast model. (*SHF5*).

Figure 5 shows results from the Multi model Ensemble defined in Table 7 where the red line indicates the Correlation Based Consensus (CBC) where weights are applied to each model result based correlation coefficients derived from a training phase where all storms but the storms being forecast are correlated with the observations. These coefficients are then used to compute a weighted ensemble mean. Note that in a real forecast situation the training would be done on storms from earlier years, however, for these calculations only the storms for 2010 were available. The calculation of the CBC is illustrated in Figure 7

In Figure 5 just the mesoscale models of the Multi Model Ensemble are shown are shown for both track and intensity. For most lead times the skill of the CBC relative to the statistical models (CLIPER for track and SHF5 for intensity) meets or exceeds the skill for even the most skillful of the regional models.

Figure 6 is just focused on intensity and now includes other models including the operational global models (GFS and NOGAPS) and two operational statistical models for intensity (SHF5 and DSHP). Here the comparison is against the HFIP baseline defined in Table 5. The red bars are again the CBC result calculated from all the models shown in Table 7 and again it has the lowest error of any of the component models except at 120 hours. Furthermore the error of the CBC is less than the baseline and in fact 25% less at 72 hours. This exceeds the HFIP goal of 20% improvement within 5 years.

As discussed earlier, we noted the introduction of aircraft data into the initialization of the regional models through EnKF provided a 20% improvement ads well. The results in Figure 6 do *not* include any models that were initialized with radar data. We can speculate here that combining the two; use of a multi model ensemble with CBC and where each model is initialized with aircraft data should provide an even greater improvement. This demonstrates the possibility that statistical post processing combined with use of additional data (especially aircraft data) can give even greater improvements over the HFIP 5 year goal.
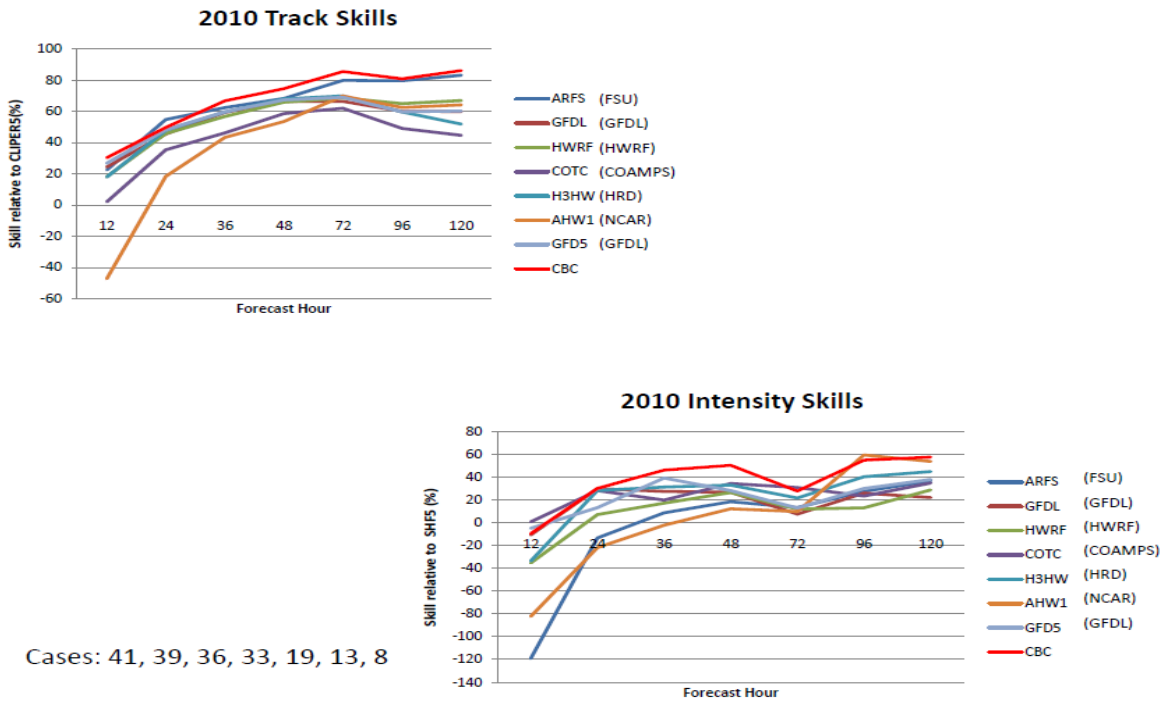
Figure 5. Track and intensity skill for the 2010 Multi-Model Ensemble (see Table 7). The various models are identified in the figure. CBC (red curve) is the correlation based consensus described in the text.
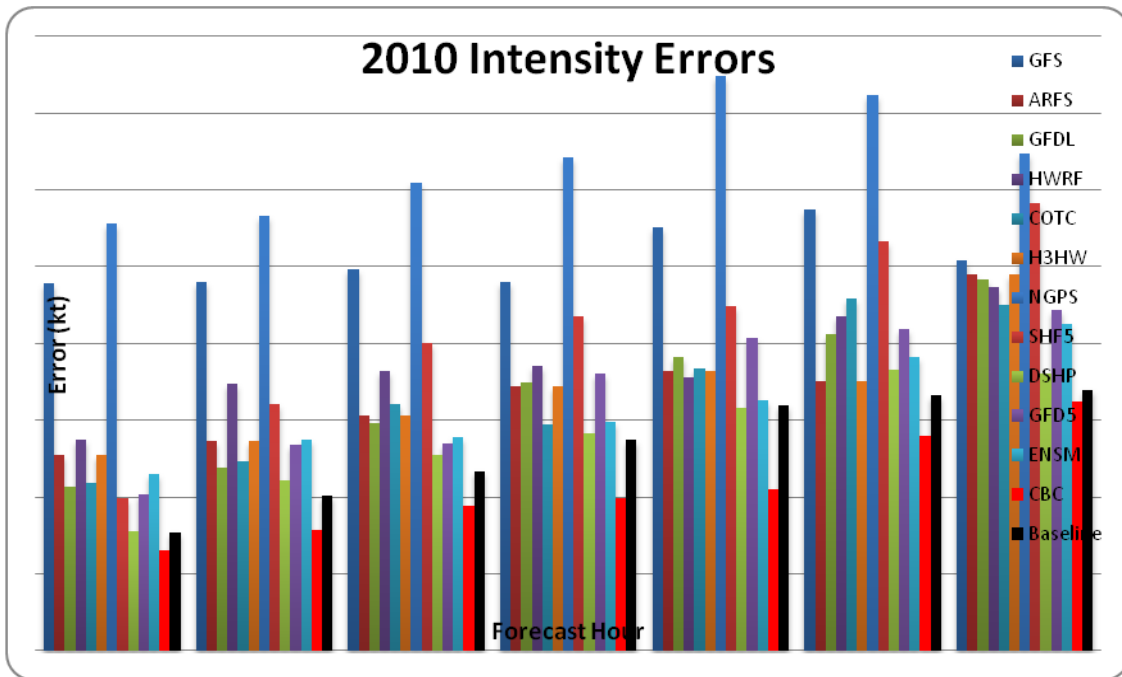

Figure 6. Another example of the CBC results but here including global models and two statistical models used in operations (SHF5 and DSHP). The error using CBC is shown in red, GFS is the operational global GFS, NGPS is the Navy global NOGAPS model and the black bar is the HFIP baseline.

Correlation Based Consensus (CBC)

Observed increment values (Lat, Lon, Int) for each lead time

Model increment forecasts (Lat, Lon, Int) for each lead time

Correlation coefficients for each model for Lat, Lon, Int at each lead time

Training phase

➤ 2010 storm cases (Total 149 cases)

➤ The storm to be forecasted is taken out to calculate the correlation coefficients

Normalize the coefficients using available member models for Lat, Lon, Int at each lead time

Utilize the above coefficients during the forecast phase and construct a new forecast

Figure 7. Illustration of the Correlation Based Consensus (CBC) method.

When run as a seasonal prediction system, the Cubed Sphere model produced a remarkable prediction of the number of hurricanes and tropical storms for the period 2000-2010.

Figure 8 shows results from the Cubed Sphere model (see Table 2) of GFDL was run coupled with the Princeton Ocean Model (POM) starting with ocean and atmospheric conditions in June then integrated through the hurricane season so it is a true prediction of the number of storms and hurricanes for the each season. Years 2000-2010 are included and the system was run several times to produce a small ensemble. Shown in the figure are the counts of hurricanes and tropical storms from each of the ensemble as well as the ensemble mean (red). The black line shows the observed numbers.

The match is remarkable indicating that the model is not significantly over or under predicting genesis so we get nearly correct predictions of the number of tropical storms and hurricanes. Since the breakout between tropical storms and hurricanes is approximate, the model system is able to forecast intensity at least at two intensity levels—tropical storm and hurricane.   Granted this is a seasonal forecast, not a forecast of individual storms. However it does indicate the ability of global models to provide reasonable forecasts of genesis and at least the large scale controls on intensity. CONS in Table 4 is a consensus of operational models and is the primary baseline for measuring improvements. OCD5 is a simple statistical model baseline for measuring improvements in forecast skill, and OFCL is the average NHC official forecast errors from the baseline period.

Figure 8. Number of hurricanes (upper figure) and tropical storms (lower figure) from a ocean/atmosphere coupled run using the Cubed Sphere global model initialized in June.

For some 2010 storms the possibility of using ensemble information to narrow the "cone of uncertainty" from the operational cone for track was demonstrated.

One of the more popular products from the national hurricane center is the "cone of uncertainty" that displays their estimate of the possible error in the forecast track. An example of this type of forecast is shown for Earl in Figure 9. The error bars for each forecast time are the average error for those lead times for all forecasts for all storms for the previous three years and so are constant at the various lead times throughout the hurricane season. This is a reasonable estimate of the uncertainty and is probably the best available at this time.

However, it is well known that the uncertainty of any forecast is widely variable from one forecast to another where there must be times when the error bars are less than the averages and there must be times when the uncertainty is much larger. One application of ensembles is to use the spread of the ensemble tracks at the various forecast times to estimate the uncertainty. An example of this is shown on the left panel of Figure 9 for the HFIP GFS ensemble with 20 members and initialized with EnKF. Note that the error ellipse for the 2-day ensemble forecast off Norfolk, VA is roughly half the size the operational error range at that position. Thus the ensemble could, in principle have been used to narrow the cone of uncertainty in that region. On the other hand there will be times, such as late in the forecast period of Figure 9 that the uncertainty is much larger than shown by the three year average. So again, in principle, this information could be used to indicate that the confidence in the official forecast must be considered to be low.
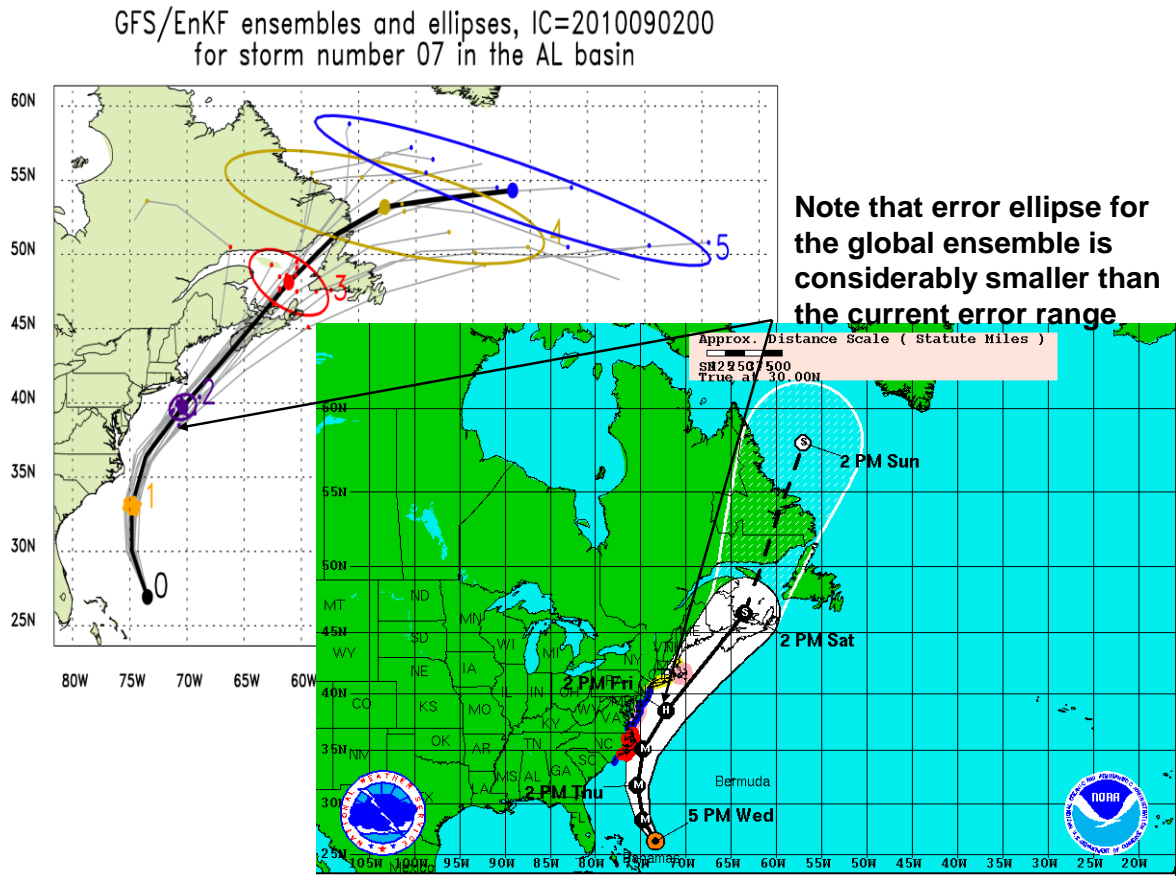
21

Figure 9. Left figure shows the ensemble of tracks from the GFS 20 member ensemble initialized with EnKF for Hurricane Earl in 2010. The elipses at the various forecast lead times in days encompase 80% of the tracks and repersent a track uncertainty given by the ensemble. Right figure shows the operational forecast from NHC with the error cone defined as the average error over the last three years.

The stream 1.5 testing and evaluation concept was successfully demonstrated. Four stream 1.5 candidate models were evaluated through a retrospective analysis study. Two of the models, AHW and GFDL, were accepted by NHC as part of stream 1.5 demonstration during the 2010 hurricane season.

As presented in Table 1, the concept of stream 1.5 is to identify models that look promising by the NHC forecasters as future operational models. These promising models are then configured to run in real time during the summer demonstration study, with products made available to NHC. The specific goals of the retrospective analysis is to (1) provide adequate statistics for assessing the skill of the stream 1.5 model candidates, (2) help identify modeling systems to could be included in future operational forecast guidance, and (3) provide information that may help to calibrate the multi-model ensemble forecasts. To determine if a model is a candidate for stream 1.5, the retrospective study is conducted using cases selected by NHC that were observed during previous seasons.

The TCMT successfully conducted the 2010 retrospective analysis study, which was coordinated with NHC and the HFIP Project Office. The candidate models were evaluated using storms

observed during the 2008 and 2009 hurricane seasons in the Eastern Pacific and Atlantic basins. A total of 27 storms were used for the evaluation. Four modeling groups participated in the retrospective analysis. The models included two configurations of the Weather Research and Forecasting (WRF) model, a new version of the NOAA Geophysical Fluid Dynamic Laboratory's (GFDL) model, and the Navy's tropical cyclone model.

The models were evaluated using a variety of baselines for the comparisons (see Table 10). The study focused on evaluating (1) track error and (2) absolute intensity errors. For verification, we used the NHC Best Track analysis product. Verification was conducted using the NHC Verification System (NHCVx) and with evaluation tools developed by TCMT. The cases were verified and samples homogenized via case matching so the evaluation was conducted over the same sample. For the evaluation, pairwise differences were computed between the model forecast and NHC Best Track, error distributions were created and results were reported in graphics and statistical significant (SS) tables.

Table 10. Stream 1.5 baseline comparisons.

Operational Baseline	Stream 1.5 Configuration	Lead Times Evaluated
GFDL	Stream 1.5	Every 6 h out to 120 h
Consensus (at least 2 available)	Consensus + Stream 1.5	Every 6 h out to 120 h
GFS (Track) DSHP (Intensity)	Stream 1.5	Official Forecast Times
Homogeneous average of previous year's top flight models	Stream 1.5	Official Forecast Times
Consensus Track: GFS, UKMET, NOGAPS, GFDL, HWRF Consensus Intensity: DSHP, LGEM, GFDL, HWRF Homogeneous Track: GFS, UKMET, GFDL Homogeneous Intensity: GFDL, HWRF, DSHP, LGEM Official Forecast Times: 00, 12, 24, 36, 48, 72, 96, 120 h		

An example result from this evaluation is shown in Figure 10 for the Atlantic Basin. Similar results were also created for the Eastern Pacific Basin (not shown). The plots show the number of lead times that were statistically significant for each of the stream 1.5 model candidates. The results are stratified by baselines (GFDL and Consensus) for track and intensity errors. The blue bars indicate the number of lead times that the baseline was statistically significant. The red bars show the number of lead times the stream 1.5 model has statistically significant results. For example, the AHW (MMM in the plot) had no statistically significant results for track compared to the GFDL baseline (upper left plot of Figure 10). However, the operational GFDL baseline

was statistically significant for 13 lead times. When compared to the Consensus Baseline, the AHW model had statistically significant (improved) results for 14 (21 – 7) lead times.

To summarize, two candidate models were accepted for stream 1.5: GFDL and AHW (MMM). GFDL was accepted prior to the TCMT evaluation based on prior assessment by NHC. The model showed some improvement when added to the operational consensus. The AHW model showed statistically significant improvements at numerous time periods when added to the operational consensus with no statistically significant degradations. The FSU (ARW) and NRL (COAMPS-TC) models were not accepted as stream 1.5 candidates. The FSU model had a neutral impact on the consensus. There was only limited sample size for evaluation. The NRL results were not sufficiently strong or consistent enough to warrant inclusion.

The stream 1.5 evaluation study was successful. The TCMT developed the procedures and successfully conducted the evaluation of the candidate models. Based on the evaluation, NHC selected two models (AHW and GFDL) that were included in the stream 1.5 for the 2010 Demo. Products from these models were delivered to NHC in real time or near real time during the hurricane season. TCMT is currently refining the retrospective analysis procedures for the 2011 stream 1.5 analysis.
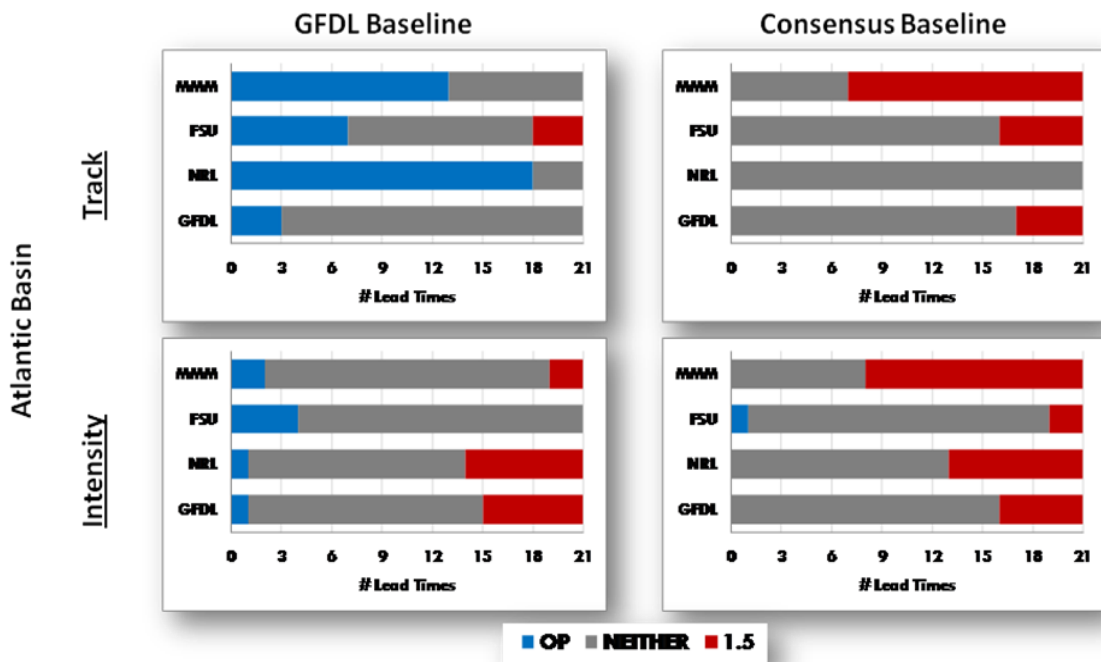

Figure 10. Plots showing the statistically significant differences between the baselines (GFDL: left; consensus: right) and the candidate models for the Atlantic Basin. The blue bar indicates that the operational model was the better model for the displayed number of lead times and the red bar indicates the stream 1.5 candidate model was the better model for the number of lead times. The gray bar indicates that neither model was statistically significant.

7. Lessons Learned from FY09 HFIP Activities

Table 11 gives a high level overview of some lessons learned from the HFIP work during FY09. A more detailed description of the lessons learned is provided in the following section.

Table 11. Lessons Learned from FY09 HFIP Activities.

> - ***The major problem with initialization of regional models identified last year remains. It is a particularly severe problem for the first 24 hours of the forecast which results in a severe problem in forecasting rapid intensification.***
>
> - ***Global models consistently provide significantly better track forecasts than regional models.***
>
> - ***Advanced data assimilation systems, such as y EnKF, is essential for improving both track and intensity forecasts.***
>
> - ***Statistical post processing of ensemble output likely to provide additional skill improvement.***
>
> - ***Significant improvement in the intensity forecast skill of dynamical models can be attained through advancements to the model physics and data assimilation, as illustrated by the 2010 COAMPS-TC performance in the W. Atlantic.***
>
> - ***The delivery of complete, consistent Tier 1 (ATCF files) and Tier 2 (gridded model) data is necessary to provide consistent evaluations of the model forecasts for the same forecast challenge using a common (operational) tracker.***

The major problem with initialization of regional models identified last year remains. It is a particularly severe problem for the first 24 hours of the forecast which results in a severe problem in forecasting rapid intensification.

Most of the regional models used operationally and in the Multi Model Ensemble use some sort of synthetic initial vortex adjusted to come close to the observed position and intensity of the storm at the initial time. This requires removing the vortex from the original global model fields given to the model and introducing the synthetic vortex. There are a number ways of developing the latter including using an axis-symmetric version of the model to spin up a vortex while forcing it to the correct intensity or using a sophisticated system to specify a vortex based on the previous run of the regional model (see last bullet in section 3.3). All this leads to some well known issues including a miss-match between the fields given the model and what the model would have developed itself or a mismatch between the introduced hurricane and the environment of the storm in the model.

This often leads to a rapid initial adjustment in the vortex to something either stronger or weaker than what was initially introduced. Note in Figure 11 that several of the models show an initial rapid increase in the vortex strength in the first 6 hours. This is quite common in regional model predictions, often more pronounced than shown in Figure 11 and in fact often occurs in the first few minutes of the forecast probably because of the model adjusting the vortex to the

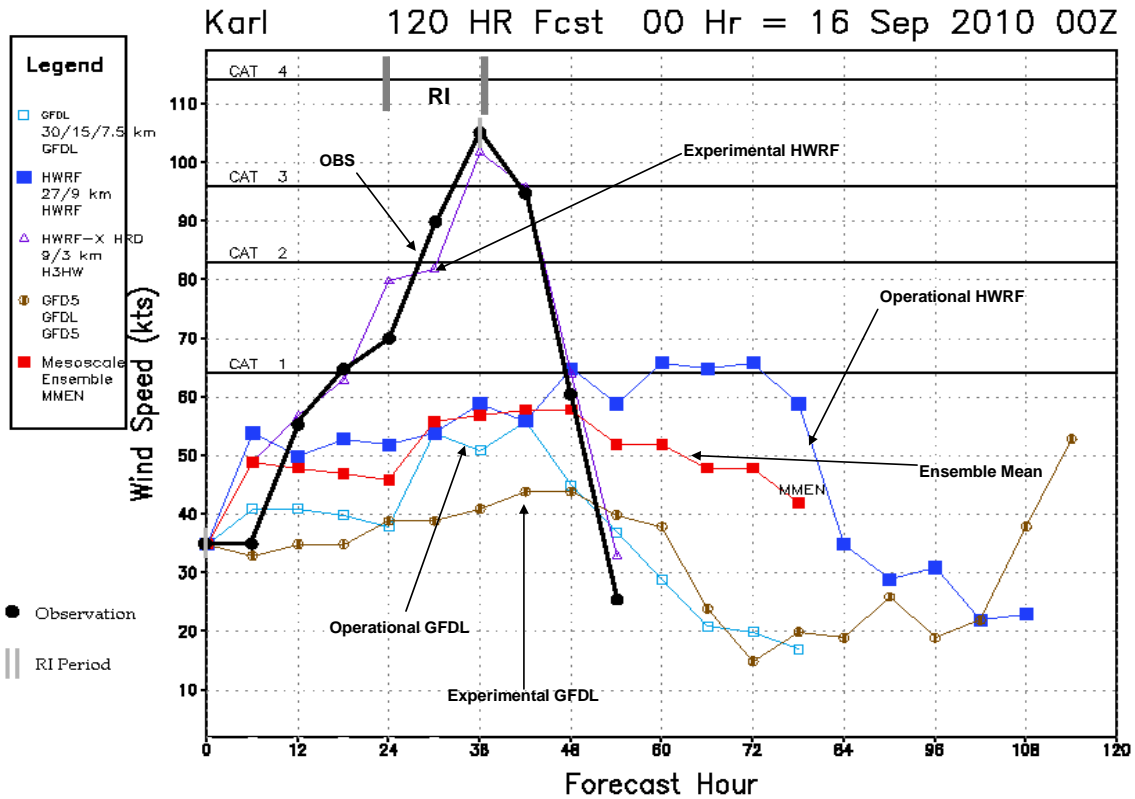environment or to agree with the model physics. We often also see sudden decreases in the initial intensity.

Figure 11. Wind speed forecasts from various members of the Multi Model Ensemble for hurricane Karl. The observed wind is shown as the black line.

In Figure 12, we illustrate one recent result from our examination of the problems related to initialization of regional models. The upper panel shows results from the operational HWRF and the lower panel from the operational GFDL. Shown are the wind speed and pressure bias after 6 hours of forecast time binned by wind speed from weak to strong storms. Blue shows the wind speed bias and red the pressure error. Note that, for weak storms, at 6 hours both models have stronger winds and lower central pressure than observed. Thus the weaker storms have spun up (note that the scales in the two figures differ). For HWRF, at stronger initial wind speeds, the opposite is true. The winds have weakened considerably, spun down by as much as 30 knots for the strongest storms plus decrease in the central pressure. In the GFDL model, the results also shows an initial spin down of the vortex but an increase in the central pressure,
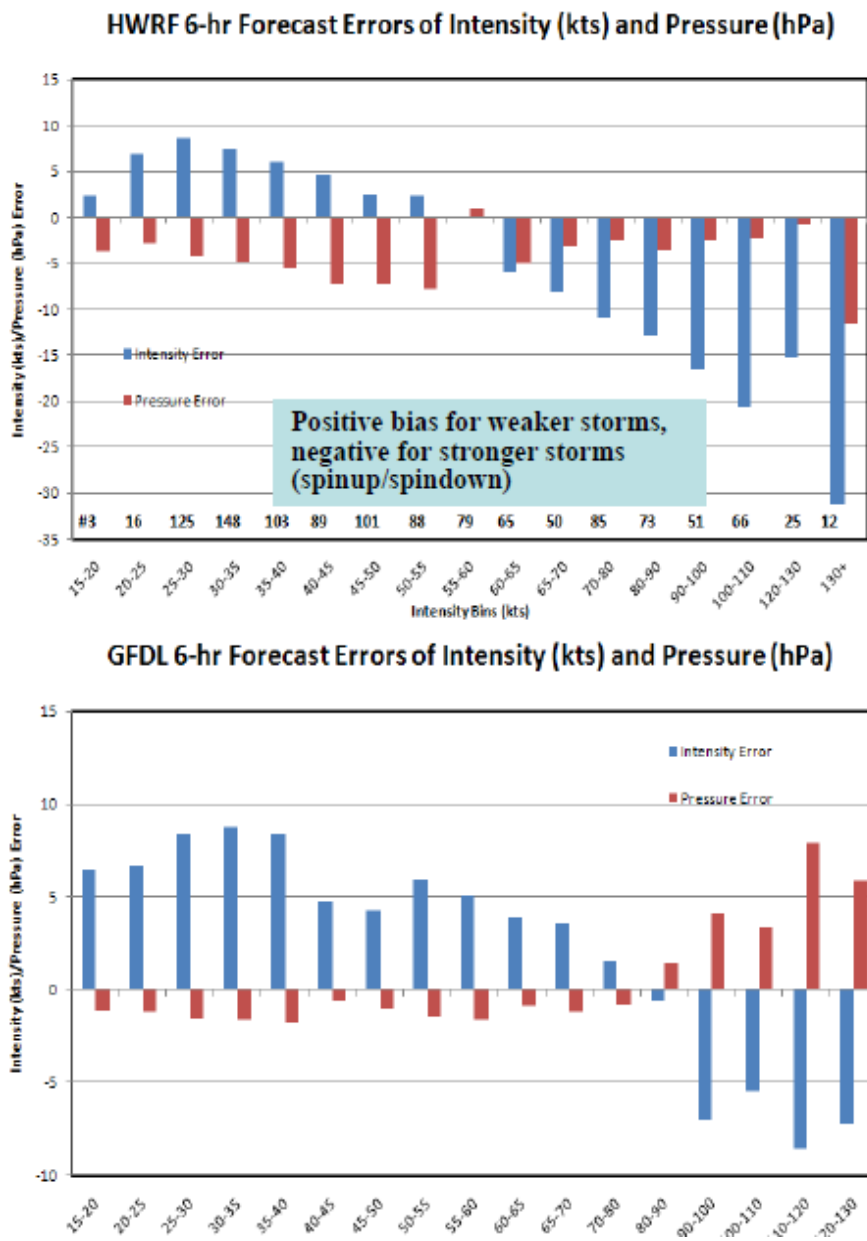
Figure 12. Wind Speed (blue bars) and minimum pressure (red bars) after six hours grouped by wind speed for HWRF (upper panel) and GFDL Model (lower panel).

One problem that can be attributed to the issues with initialization will be prediction of rapid intensification (RI) in the first 24 hours. One of the goals of HFIP is to reduce the false alarm rate and increase the probability of detection of RI especially during the first day of the forecast. Given the problems of initialization of the regional models, we have not yet been able to demonstrate an ability to meet that goal. Occasionally a model will forecast rapid intensification quite accurately. An example is shown in Figure 11 where the HWRF-x (e.g., experimental HWRF) run by HRD did a remarkable job forecasting the intensity changes of hurricane Karl including the period of rapid intensification. This of course wasn't the case for the other models shown in the figure and in fact is not always true for the HWRF-x as well. Still Figure 11 does provide hope that we will eventually be able to meet the RI goal of HFIP. But this is likely only to happen once the general problem with initialization is resolved.

Some of the regional models used in HFIP (NCAR/ARW and PSU/ARW for example) use an EnKF initialization process and there is some hope that a system that uses the model itself for initialization and avoids the use of a synthetic initial vortex may perform better than those that do. This would allow the initial vortex to be consistent with the model and consistent with the hurricane environment in the model. However, if the PSU/ARW model results shown in Figure 4 are reexamined, that model uses the EnKF approach yet the initial errors remain quite large.

Finally it is generally accepted that much of the physics associated with RI occurs at small, perhaps the convective scales. To resolve these scales the models will require at least 3 km resolution and this will only be possible with the regional models for the next decade. Thus we will probably need high resolution to resolve the problem with initialization of the regional models. Note in Figure 11, of all the models shown, only HWRF-x used 3 km resolution.

Global models consistently provide significantly better track forecasts than regional models

Figures 2 and 3 are a clear demonstration that the regional models do not do as well as global models in predicting hurricane track at any lead time but particularly at the longer lead times. This is perhaps not surprising since the regional models use the global models for predicting their lateral boundary conditions so it might be expected that the regional models can be no better than the global models in predicting track. Track is largely determined by the large scale environment in which the hurricanes are imbedded which is better predicted by the global models than the regional models. Still the differences shown in Figures 2 and 3 are surprisingly large. Figure 13 is another example of the same result where the two operational regional models (HWRF and GFDL) are compared to the operational GFS global model. Here the differences are shown relative to CLIPPER. The differences between the global and regional models are similar to those shown in Figures 2 and 3.
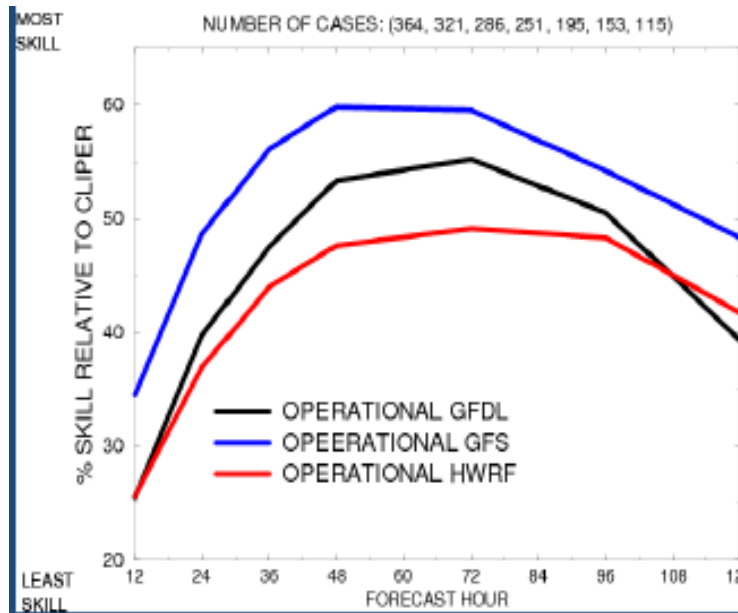

Figure 13. Track forecast skill relative to CLIPPER (Climatology and persistence statistical model) for GFS operational global model (blue), GFDL Model (black) and HWRF (red). Here higher numbers are better.

In Figures 2 and 3, four different regional models are compared to two different global models. One can only conclude that the global models (even at relatively low resolution—the GFS EnKF was 60 km) out performs the regional models in track. Thus emphasis on improving track forecasting must focus on improving the global models. The since the global models will not reach the resolution necessary to account for inner core process for at least the next 5 years, regional models will be required for intensity forecasting, particularly RI as noted above with emphasis on forecast times out to 3-4 days.

Advanced data assimilation systems, such as EnKF, are essential for improving both track and intensity forecasts.

The 2009 HFIP final report contained several examples of the relative value of the more advanced data assimilations systems (over 3DVAR). Though we didn't show any examples of the results again this year, the same conclusions were evident for the 2010 results. Currently NCEP is in the process of replacing their GSI (3DVAR) system with a hybrid system involving the ensemble approach combined with first GSI and later possibly with 4DVAR. The results from HFIP suggest this is a good approach.

Statistical post processing of ensemble output likely to provide additional skill improvement

Much of the discussion above focused on using model improvement to achieve the HFIP goals. We have already noted in a previous section where we discussed results from the Multi Model Ensemble that statistical post processing adds considerably to the skill of the ensemble. In particular we noted the results of the CBC process (Figure 6) which gave improvements in intensity forecasts that are apparently similar to adding aircraft data to the initialization of a regional model (Figure 4). This is a very preliminary result and needs to be further explored using statistical data sets that are completely separate from the forecasts.

There is another class of statistical models that combine various sources of data beyond that derived from models and they have proved superior to the models historically. The one shown in Figure 14, SHIPS (Statistical Hurricane Intensity Prediction System), is one that is used operationally. These models can and do use data from model forecasts. Figure 14 shows the impact of using several of the HFIP models (the operational GFDL, HWRF and COAMPS-TC)) in SHIPS. Note that in the cases of HWRF and GFDL the statistical model provided better forecasts than the respective models that were included in SHIPS and in fact gave slightly better forecasts than the original SHIPS model beyond 24 hours except at the longest lead time.

Statistical post processing taken together with model improvements, use of ensembles, improved data assimilation and inclusion of more data such as aircraft and satellite data promises to be the keys to meeting the HFIP goals.
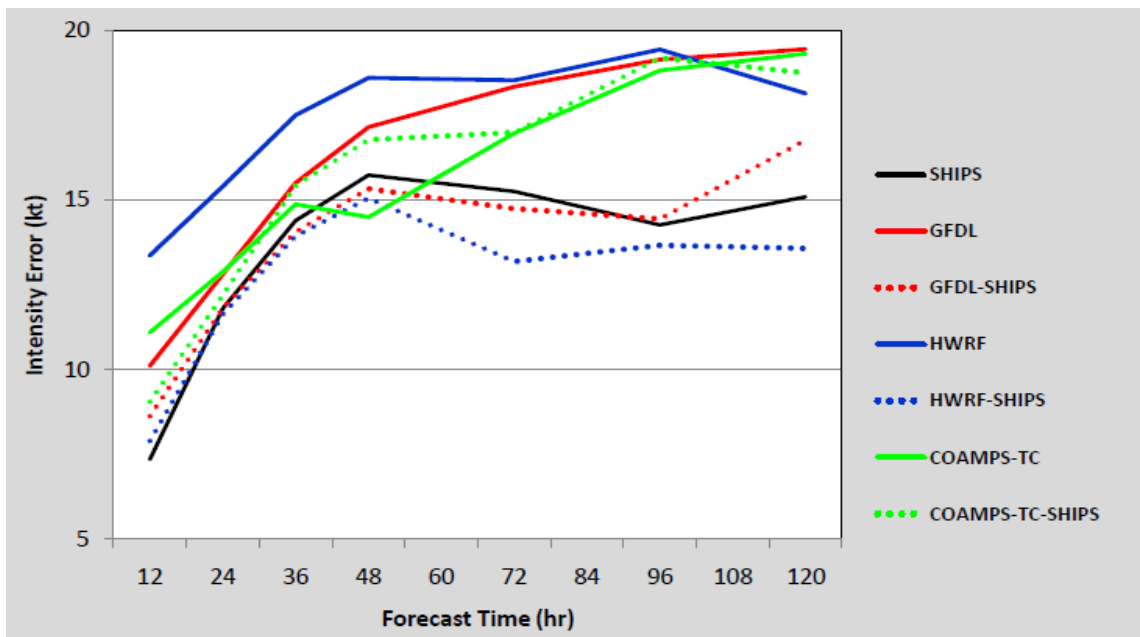
Figure 14. The impact of including three of the operational regional models (HWRF, GFDL and TC-COAMPS) in the SHIPS statistical intensity model.

Significant improvement in the intensity forecast skill of dynamical models can be attained through advancements to the model physics and data assimilation, as illustrated by the 2010 COAMPS-TC performance in the W. Atlantic.

Recent model upgrades following the FY09 demonstration forecasts to COAMPS-TC resulted in improvements to the overall intensity forecast skill. These upgrades included improved boundary layer and microphysics parameterizations, as well as inclusion of additional observations into the Navy 3D-VAR, NAVDAS. Figure 15 shows a homogeneous comparison of the intensity error for HWRF, COAMPS-TC, GFDL, GFDN, HWRF-X and the statistical models LGEM and DSHIPS. The COAMPS-TC performance is on par or superior to the statistical models in 36-60 h forecast periods, a significant improvement over of FY09 statistics resulting from model improvements to the physics and data assimilation. The statistical models and HFIP baseline are lagged by 6 h to reflect the delay in attaining the dynamical model output. The results highlight the sensitivity of intensity forecasts to the physical parameterizations and data assimilation. Another interesting lesson learned is that the COAMPS-TC intensity forecasts were superior to all other dynamical models in the 36-60 h forecast range, however, the track forecasts were not as skillful as the global or other regional models. The results suggest that reasonably skillful intensity forecasts can be attained using simple vortex initialization methods. Improvements to the track skill have been achieved through assimilation of additional observations including total precipitable water (TPW) and new averaging methods appropriate for the mesoscale for scatterometer observations, as illustrated in homogeneous track error statistics shown in Figure 16.
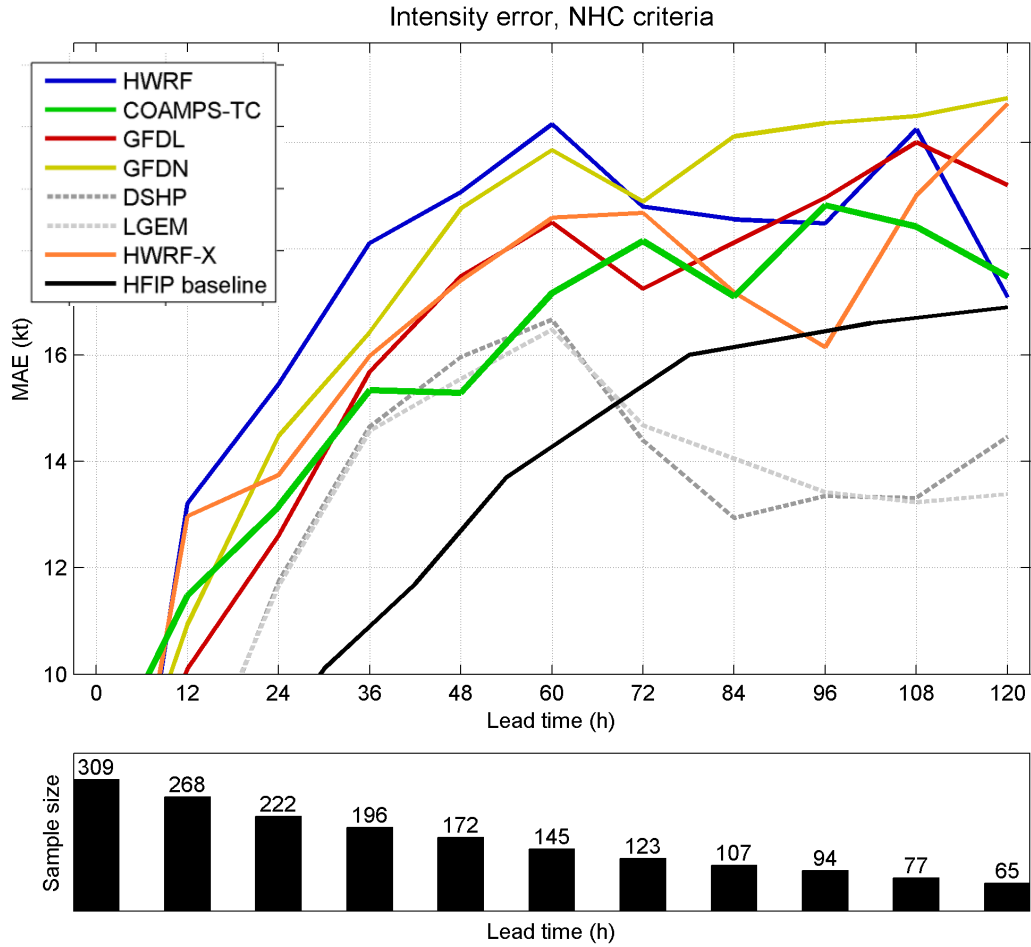
Figure 15. Homogeneous comparison of the intensity error for HWRF, COAMPS-TC, GFDL, GFDN, HWRF-X and the statistical models LGEM and DSHIPS. Note the COAMPS-TC performance is on par with the statistical models in 36-60 h forecast periods, a significant improvement over of FY09 statistics resulting from model improvements to the physics and data assimilation. The statistical models and HFIP baseline are lagged by 6 h to reflect the delay in attaining the dynamical model output.
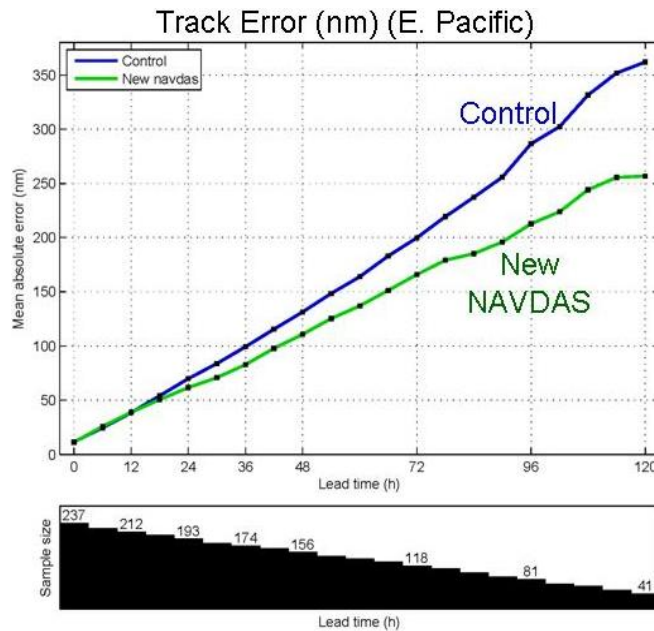
Figure 16. Homogeneous comparisons of the track error for the E. Pacific are shown. Track error results are shown for a series of control forecasts and for a new version of the Navy 3D-Var, NAVDAS, that makes use of additional observations including total precipitable water (TPW) and new averaging methods appropriate for the mesoscale for scatterometer observations. The sample size is shown at the bottom as a function of forecast hour.

The delivery of complete, consistent Tier 1 (ATCF files) and Tier 2 (gridded model) data is necessary to provide consistent evaluations of the model forecasts for the same forecast challenge using a common (operational) tracker.

The TCMT conducted several studies to show the importance of having a complete, consistent dataset for both Tier 1 (ATCF files) and Tier 2 (gridded model products) data. To show the impact on sample size, the NRL (COAMPS-TC), which had an almost complete dataset, was subset to match the number of cases that were delivered by other modeling groups. The stream 1.5 analyses results were re-evaluated with the smaller sample size. The results (not shown) indicated an improvement in the number of lead times for both intensity and track that the NRL model was statistically significant compared to the consensus baseline. Clearly, the models that did not provide a complete, consistent dataset had different forecast challenges, which influenced the results.

Figure 17 shows the impact of using different trackers. In this example, we compare the HWRF track errors from the tracker used at NCEP/EMC (Tier 1 products) with the track errors using the common GFDL tracker (Tier 2 products). The results shown in Figure 17 indicate there are significant differences as a result of the different trackers used. These results highlight the importance in having a complete Tier 2 dataset available to make consistent comparisons of the models with a common tracker.
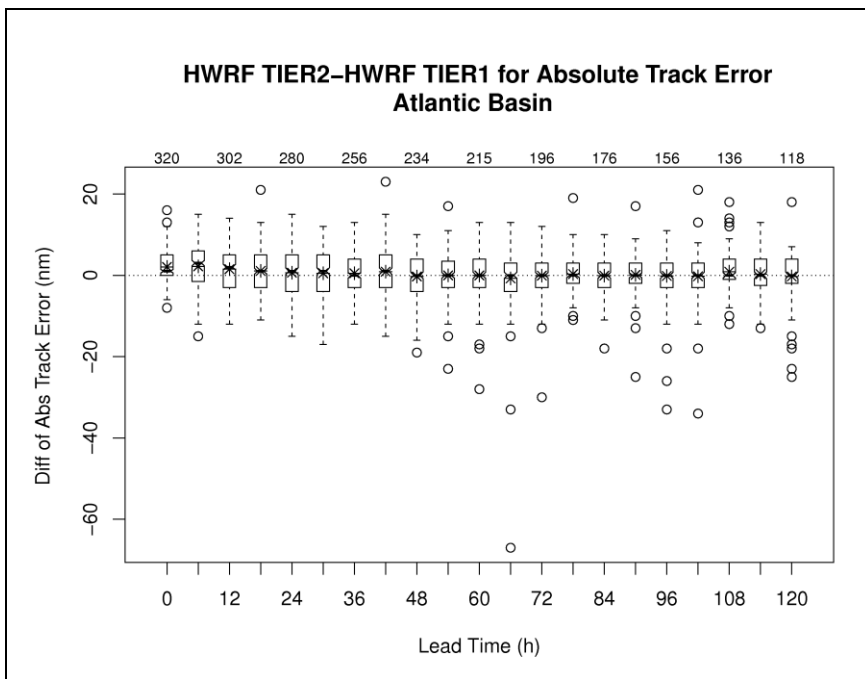
Figure 17. Results showing differences in absolute track errors when different trackers are used.

8. Challenges for HFIP beyond FY10

Table 12 lists what HFIP considers to be its main challenges for 2011 and beyond. All these points have been covered in the earlier discussion so we simply list them here.

Table 12. HFIP Challenges.

- *Initialization systems for regional models need to be improved for forecasts in the first 24 hours.*

- *HFIP needs to develop systems to better make use of satellite observations in regional models.*

- *Implementing the Hybrid DA system in operations as soon as possible.*

- *Development of appropriate statistical post processing systems for ensemble output.*

- *Operational computing not likely to be sufficient to meet HFIP goals within 3 years.*

9.   Meeting the HFIP Performance Goals

In the discussion above, we showed evidence that it is possible to meet the 5 year HFIP intensity and track goals. It is less clear whether we can or cannot make the RI goals. The problem with the latter is that we still need to solve a major problem with regional model initialization. Most techniques provide the model with a vortex that is not consistent with the model or the hurricane environment so there is an initial adjustment that can be larger than a typical example of RI. A major HFIP focus this next year will be trying to solve the initialization problem.

While it appears that use of aircraft data will likely help HFIP meet its intensity goals for storms for which such data is available, it will not be available for storms for a large majority of model initializations. For those, we will need to rely on better use of satellite data taken in the near vicinity of the hurricane. Another major focus this next year will be on improving satellite data assimilation in regional model initialization.

Except for the RI issue, we can now say with considerable confidence what a final end state operational configuration of the hurricane numerical prediction system should look like in 2014.

The longer range predictions, out to one week, of both track and intensity will be accomplished by global models run as an ensemble and initialized with a Hybrid data assimilation system and post processed with various statistical models. Resolution of these global models needs to be at least 20 km and the results will be improved if more than one global model is used in the ensemble.

The earlier forecast periods—out to 48-72 hours will be accomplished with regional models run with at least 3 km resolution as a multi- model ensemble. All models will use all available aircraft and satellite data. These will also be post processed with statistical models. The focus with the regional models will be on intensity and with the high resolution there is confidence in the community that the RI goals can be met with the regional models. More specifically the end system might include:

- Global model ensemble with Hybrid Data Assimilation
    - 20 members at 20 km
    - Multi Model (at least two—e.g.: FIM, GSF, Cubed Sphere)

- Regional model ensemble
    - 20 members at 3 km
    - Multi model (at least two—e.g.: HWRF, AHW, COAMPS-TC)
    - Using all available aircraft and satellite data in core and near environment of hurricane

- Statistical Post processing
    - Bias correction, Correlation Based Correction, LGEM, SHIPS

The ability to run this system will however require about a 10 times increase in computer resources in operations with a major emphasis on the ability to run the high resolution ensembles.

10. List of HFIP Supported Publications and Presentations

This extensive list can be view and downloaded at the follow link:

http://www.hfip.org/documents/HFIP_pubs_2010-012311.pdf

11. *Appendix A: List of HFIP Teams*

Bold face type denotes team leads.

1. Global Model/Physics Development Team
Stan Benjamin (ESRL) Stan.Benjamin@noaa.gov
John Brown (ESRL) John.M.Brown@noaa.gov
Kevin Yeh (AOML) Kevin.Yeh@noaa.gov
Melinda Peng (NRL) melinda.peng@nrlmry.navy.mil
Shian-Jiann Lin (GFDL) Shian-Jiann.Lin@noaa.gov
Steve Lord (EMC) Stephen.Lord@noaa.gov
Jim Ridout (NRL) james.ridout@nrlmry.navy.mil
Jian-Wen Bao (ESRL) Jian-Wen.Bao@noaa.gov
John M Ward (EMC) John.M.Ward@noaa.gov

2. Regional Model/Physics Development Team
Morris Bender (GFDL) Morris.Bender@noaa.gov
Young Kwon (EMC) young.kwon@noaa.gov
Steve Lord (EMC) Stephen.Lord@noaa.gov
Gopal (AOML) Sundararaman.G.Gopalakrishnan@noaa.gov
Rich Hodur (NRL) hodur@nrlmry.navy.mil
Shaowu Bao (ESRL) shaowu.bao@noaa.gov
Isaac Ginis (URI) iginis@gso.uri.edu
Jim Doyle (NRL) james.doyle@nrlmry.navy.mil
Robert Rogers (AOML) Robert.Rogers@noaa.gov
Jian-Wen Bao (ESRL) Jian-Wen.Bao@noaa.gov
Bob Tuleya (Old Dominion Univ) Robert.Tuleya@noaa.gov
Ligia Bernardet (ESRL) ligia.bernardet@noaa.gov
Chris Davis (NESL) cdavis@ucar.edu

3. Ensemble Systems Development Team
Zoltan Toth (ESRL) zoltan.toth@noaa.gov
Carolyn Reynolds (NRL) carolyn.reynolds@nrlmry.navy.mil
Sim Aberson (HRD) Sim.Aberson@noaa.gov
Tom Hamill (ESRL) tom.hamill@noaa.gov
Jeff Whitaker (ESRL) Jeffrey.S.Whitaker@noaa.gov
Fuqing Zhang (PSU) fzhang@psu.edu
Yuejian Zhu (EMC) Yuejian.Zhu@noaa.gov
Jun Du (EMC) Jun.Du@noaa.gov
Mike Brennan (NHC) Michael.J.Brennan@noaa.gov
Mrinal K Biswas (FSU) Mrinal.K.Biswas@noaa.gov
Krish (FSU) tkrishnamurti@fsu.edu
Teddy Holt (NRL) teddy.holt@nrlmry.navy.mil

4. Data Assimilation/Vortex Initialization Team)
Jeff Whitaker (ESRL) Jeffrey.S.Whitaker@noaa.gov
Bill Lapenta (EMC) Bill.Lapenta@noaa.gov
Mingjing Tong (EMC) mingjing.tong@noaa.gov
Jim Doyle (NRL) jim.doyle@nrlmry.navy.mil
Altug Aksoy (AOML) Altug.Aksoy@noaa.gov
Yuanfu Xie (ESRL) yuanfu.xie@noaa.gov
Milija Zupanski (CIRA) zupanskim@cira.colostate.edu
Liyan Liu (EMC) Liyan.liu@noaa.gov
Fuqing Zhang (PSU) fzhang@psu.edu
Tomislava Vukicevic (AOML) Tomislava.Vukicevic@noaa.gov

5. Verification Team
Tim Marchok (GFDL) timothy.marchok@noaa.gov
Barb Brown (RAL) bgb@ucar.edu
Jim Goerss (NRL) jim.goerss@nrlmry.navy.mil
Mark DeMaria (NESDIS/STAR) DeMaria@cira.colostate.edu
Robert Rogers (AOML) Robert.Rogers@noaa.gov
James Franklin (NHC) james.franklin@noaa.gov
Vijay Tallapragada (EMC) vijay.tallapragada@noaa.gov
Michael Fiorino (ESRL) Michael.Fiorino@noaa.gov
Hao Jin (NRL) hao.jin@nrlmry.navy.mil
Ligia Bernardet (ESRL) ligia.bernardet@noaa.gov
Louisa Nance (RAL) nance@ucar.edu
Tony Eckel (NOAA/OST) t.eckel5162@sbcglobal.net

6. Applications Development/Diagnostics Team
Mark DeMaria (NESDIS/STAR) mark.demaria@noaa.gov
Ed Rappaport (NHC) edward.n.rappaport@noaa.gov
Vijay Tallapragada (EMC) vijay.tallapragada@noaa.gov
Yi Jin (NRL) yi.jin@nrlmry.navy.mil
Buck Sampson (NRL) buck.sampson@nrlmry.navy.mil
Robert Rogers (HRD) Robert.Rogers@noaa.gov
Barb Brown (RAL) bgb@ucar.edu
Richard Pasch (NHC) richard.j.pasch@noaa.gov
Michael Fiorino (ESRL) Michael.Fiorino@noaa.gov
Louisa Nance (RAL) nance@ucar.edu
Bob Tuleya (Old Dominion Univ) Robert.Tuleya@noaa.gov
Jim Hansen (NRL) jim.hansen@nrlmry.navy.mil
John Knaff (NESDIS/STAR) john.knaff@noaa.gov
Gopal (AOML) Sundararaman.G.Gopalakrishnan@noaa.gov
Krish (FSU) tkrishnamurti@fsu.edu
Tony Eckel (NOAA/OST) t.eckel5162@sbcglobal.net

7. Hurricane Observations Team
Sim Aberson (AOML) Sim.Aberson@noaa.gov
John Knaff (NESDIS/STAR) john.knaff@noaa.gov
Robert Atlas (AOML) Robert.Atlas@noaa.gov
Jack Beven (NHC) john.l.beven@noaa.gov
John Derber (EMC) john.derber@noaa.gov
Michelle Mainelli (NCO) michelle.m.mainelli@noaa.gov
George Halliwell (AOML) ghalliwell@rsmas.miami.edu
Daniel Melendez (NWS/OST) Daniel.Melendez@noaa.gov
Jeff Hawkins (NRL) Jeffrey.Hawkins@nrlmry.navy.mi
Chris Fairall (ESRL) chris.fairall@noaa.gov
Isaac Ginis (URI) iginis@gso.uri.edu
Nick Shay (RSMAS) nshay@rsmas.miami.edu
Paul Chang (NESDIS)Paul.S.Chang@noaa.gov
Jim McFadden (AOC) jim.d.mcfadden@noaa.gov
Tara Jensen (RAL) jensen@ucar.edu
Thiago Quirino (AOML) Thiago.Quirino@noaa.gov
Yuanfu Xie (ESRL) yuanfu.xie@noaa.gov

8. Coupled Ocean/Wave model Team
Hendrik Tolman (EMC) hendrik.tolman@noaa.gov
George Halliwell (AOML) george.halliwell@noaa.gov
Isaac Ginis (URI) iginis@gso.uri.edu
Shaowu Bao (ESRL) shaowu.bao@noaa.gov
Sue Chen (NRL) sue.chen@nrlmry.navy.mil
Jian-Wen Bao (ESRL) Jian-Wen.Bao@noaa.gov
Chris.Fairall (ESRL) Chris.Fairall@noaa.gov
Nick Shay (RSMAS) nshay@rsmas.miami.edu
Daniel Melendez (OAR) Daniel.Melendez@noaa.gov

12. *Appendix B: Organization Acronyms*

AOC—Aircraft Operations Center, NOAA
AOML—Atlantic Oceanographic and Meteorological Laboratory, OAR/NOAA
EMC—Environmental Modeling Center, NCEP/NOAA
ESRL—Earth System Research Laboratory, OAR/NOAA
FSU—Florida State University
GFDL—Geophysical Fluid Dynamics Laboratory, OAR/NOAA
NCAR—National Center for Atmospheric Research
NCEP—National Centers for Environmental Modeling, NWS/NOAA
NESDIS—National Environmental Satellite Data Information Service, NOAA
NHC—National Hurricane Center, NWS/NOAA
NOAA—National Oceanic and Atmospheric Administration
NRL—Naval Research Laboratory, Monterey
NWS—National Weather Service, NOAA
OAR—Ocean and Atmospheric Research, NOAA
ODU—Old Dominion University
OST—Office of Science and Technology, NWS/NOAA
RAL—Research Applications Laboratory, NCAR
RSMAS—Rosenstiel School of Marine and Atmospheric Science, University of Miami
URI—University of Rhode Island