NOAA

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

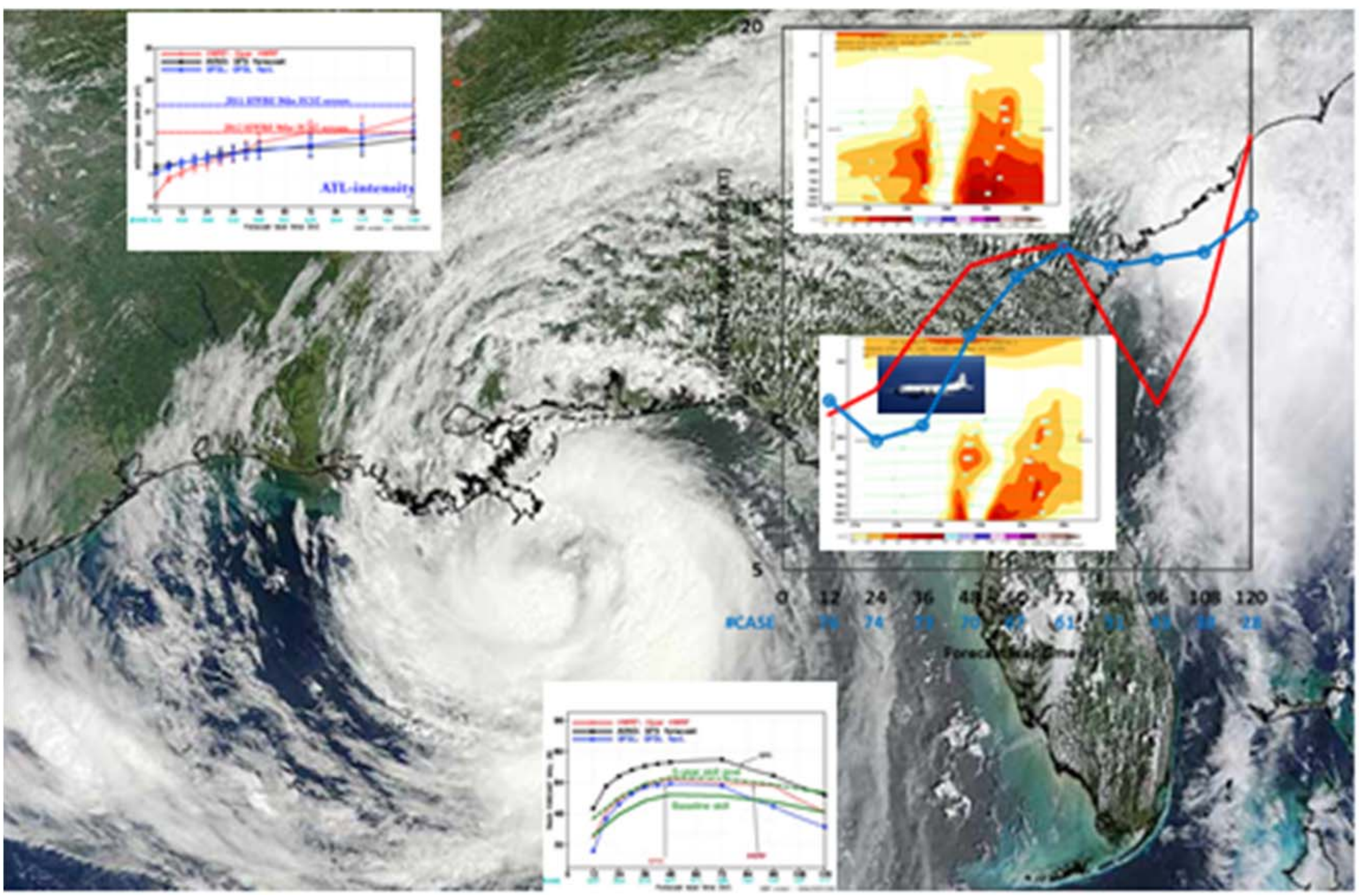
United States Department of Commerce

HURRICANE FORECAST IMPROVEMENT PROJECT

2012 HFIP R&D Activities Summary: Recent Results and Operational Implementation

26 March 2013

HFIP Technical Report: HFIP2013-1



2012 HFIP R & D Activities Summary: Recent Results and Operational Implementation

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

26 March 2013

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

Table of Contents

Executive Summary

This report describes the activities and results of the Hurricane Forecast Improvement Program (HFIP) in 2012. Unlike previous years we organized the report around a discussion of the impact of HFIP development on the operational hurricane forecast system. HFIP is organized around the three "streams": Stream 1 or the operational model development, Stream 1.5 which comprises a group of experimental models that have been evaluated by the National Hurricane Center (NHC) pre-season and then made available to NHC forecasters during their forecast cycle, and Stream 2 representing HFIP experimental models which test and evaluate new techniques and strategies for model forecast guidance before testing is begun for possible operational implementation. Stream 2 also tests techniques that cannot be tested on current operational computers because of size and time requirements, but can be tested on HFIP computer facilities in Boulder, CO. Those studies are looking ahead to possible future operational computational capability. The report outlines the HFIP program, how it is organized, its goals, its models and then results from each of the three Streams.

Stream 1.0 Accomplishments

- The new Hybrid Data Assimilation (DA) Systems, components of which have been tested for several years by HFIP, went operational in the Global Forecast System (GFS) in May 2012. This year, 2012, the operational global model produced hurricane track forecasts which exceeded the 5-year HFIP goal for hurricane track forecasting and approached the 10-year goal.
- A third nest was added to the Hurricane Weather Research Forecasting (HWRF) regional hurricane model allowing an inner core resolution of 3 km. This and other changes led to a 20% improvement in both track and intensity forecasts by HWRF.

Stream 1.5 Results

- The Global Forecast System (GFS) track forecasts were among the best of the dynamical models, even beating the European Centre Medium-range Weather Forecasts (ECMWF) model at most lead times.
- No Stream 1.5 model results for intensity, including those from the HFIP models, reached the skill of the benchmark. These results include runs using radar data from NOAA aircraft but the number of such cases is too small, 19 compared to 183 other cases, to impact the overall statistics. The impact of the radar data is presented in Section 8.
- HFIP plans for improving the model forecasts of intensity include: improved data assimilation by adding a hybrid DA capability to the regional models, further use of high resolution aircraft observations and increased use of satellite data to better define the hurricane core and the environment near the hurricane when aircraft data are not available.

Stream 2.0 Accomplishments

- Perhaps the most promising result from the experimental models this year was the confirmation in real time of the impact that the radar data had on improving the intensity forecasts from the operational version of the HWRF model. Improvement in intensity forecasts of 10%-20% were noted out to 72 hours. That improvement was confirmed with similar results from the Atlantic Oceanographic and Meteorological Laboratory(AOML) running HWRF as an experimental model and with a different data assimilation system.
- A real-time product for potential genesis using the various global ensembles from around the world was implemented. The current operational ensembles tend to over predict genesis at all forecast lead times.
- HFIP began closely working with the hurricane surge models. A forecast for hurricane Sandy made 36 hours before landfall showed surprising skill in forecasting the surge in New York City and along the New Jersey coast.
- The Florida State Multi-Model Ensemble using the stream 1.5 and other models as input provided the best forecasts at almost all lead times.

Future configuration of the Hurricane Forecast System

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

Component	**Specifications**
Global model ensemble with Hybrid Data Assimilation	1) 20 members at 10-20 km 2) Multi Model (at least two – e.g. FIM, GSF)
Regional model ensemble	1) 20 members at 3 km 2) Multi model (at least two – e.g. HWRF, AHW, TC-COAMPS) 3) Using all available aircraft and satellite data in core and near environment of hurricane
Statistical Post Processing	1) LGEM, SHIPS, others

1. Introduction

Tropical cyclone activity in the Atlantic hurricane basin broke records for numbers and impacts during the first decade of the new millennium. A total of 13 hurricanes crossed the contiguous U.S. coastline from 2000-2010, including such now infamous storms as Charley (2004), Katrina (2005), Rita (2005), Wilma (2005) and Ike (2008). In 2005 alone, 27 Atlantic systems reached tropical storm status, far surpassing the previous record of 21. The heightened activity brought an increased awareness of the dangers from tropical cyclones and led to a number of studies concerning NOAA's ability to forecast hurricanes. The additional attention on the nation's hurricane warning program provided opportunities to give visibility to and initiate actions on intensity forecasting, a critical area where no appreciable improvement had been made over the preceding two decades (e.g., Cangialosi and Franklin 2011). To address this issue, NOAA, through its Science Advisory Board (SAB), established a Hurricane Intensity Research Working Group (HIRWG), which documented its recommendations to improve forecasts of hurricane intensity in October 2006 (NOAA SAB 2006). In addition, the National Science Foundation (NSF) National Science Board issued a report in January 2007 on the need for a National Hurricane Research Initiative (NSF 2007) and the Office of the Federal Coordinator of Meteorological Services (OFCM) issued a report in February 2007 calling for a federal investment of $70-85 million annually over the next 10 years for tropical cyclone research and development, transition of research to operations, and operational high performance computing (OFCM 2007).

NOAA's response was the establishment of the Hurricane Forecast Improvement Project (HFIP), as noted in this November 2007 statement: "In response to the HIRWG report, NOAA convened a corporate hurricane summit developing unified strategy to address hurricane forecast improvements. On May 10, 2009 the NOAA Executive Council established the NOAA Hurricane Forecast Improvement Project , a 10 year effort to accelerate improvements in one to five day forecasts for hurricane track, intensity, and storm surge and to reduce forecast uncertainty, with an emphasis on rapid intensity change" (NOAA SAB 2008). In July 2008-July 2009 the President's proposed budget was amended to include $13M for HFIP, and this increment became part of NOAA's base budget.

This report describes the HFIP program, 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.

2. The Hurricane Forecast Improvement Project

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

- Reduce average track errors by 20% in 5 years, 50% in 10 years for days 1 through 5.
- Reduce average intensity errors by 20% in 5 years, 50% in 10 years for days 1 through 5.

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

Forecasts of higher accuracy and greater reliability are expected to lead to higher user confidence and improved public response, resulting in savings of life and property. Reaching these goals, however, requires major investments in enhanced observational strategies, improved data assimilation, numerical model systems, expanded forecast applications based on the high-resolution and ensemble-based numerical prediction systems and improved computational infrastructure. NOAA also recognizes that addressing the challenges associated with improving hurricane forecasts requires interaction with, and the support of, the larger research and academic communities.

It is hypothesized that the ambitious 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.

For FY2012, HFIP program funding was approximately $ 7M, with an additional $4M dedicated to enhancing computer capacity available to the Program. This level was approximately half that of previous years. The funding for computing was used to enhance the HFIP system established in Boulder, Colorado in FY2009, and resulted in machines called t-jet and u-jet with a total of 23,000 processors. The $7M was distributed to: 1) various NOAA laboratories and centers, including the Geophysical Fluid Dynamics Laboratory (GFDL), National Environment Satellite Data and Information Service (NESDIS), the Earth System Research Laboratory (ESRL) and NHC; 2) the National Center for Atmospheric Research (NCAR); 3) the Naval Research Laboratory in Monterey (NRL), and 4) several universities: University of Wisconsin (UW), The Pennsylvania State University (PSU), Colorado State University (CSU), Florida State University FSU, University of Rhode Island (URI), University of Miami (UM), University of Colorado (UC), University of Maryland (UMD), University of California, Los Angeles (UCLA), University of Oklahoma (OU), and the State University of New York (SUNY), Albany (awarded through a NOAA Announcement of Opportunity – see Appendix A) and the National Oceanographic Partnership Program (NOPP). Specifically, $1M was contributed each year for three years to the NOPP, and through an Announcement of Opportunity, for competed proposals

---

[1] POD, Probability of Detection is equal to the total number of correct events forecast (hits) divided by the total number of events observed. FAR false Alarm Ratio is equal to the total number of incorrect events forecast (misses) divided by the total number of events observed.

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

related to improving understanding and prediction of hurricanes. The funding to NOPP from HFIP was matched by funding from the Office of Naval Research.

Distribution of the $7M was based on recommendations from nine teams focused on various components of the hurricane forecast problem. The current teams, made up of over 50 members drawn from the hurricane research, development and operational communities, are listed in Table 1, for 2012, and in Tables 2 and 3, for 2013, along with the team co-leaders and the participating organizations.

Table 1. HFIP Development Teams 2012

FY 2012 Teams	Team Leads and Member's Organization
1. Global Model/Physics	**Stan Benjamin (ESRL), John Brown (ESRL),** AOML, NRL, GFDL, EMC, NRL
2. Regional Model/Physics	**Morris Bender (GFDL), Young Kwon (EMC),** AOML, NRL, ESRL URI, Old Dominion Univ., NCAR
3. Ensembles	**Zoltan Toth (ESRL), Carolyn Reynolds (NRL**), AOML, PSU, EMC, NHC, FSU, NCAR
4. Data Assimilation/Vortex Initialization Team	**Jeff Whitaker (ESRL), Tomi.Vukicevic (AOML),** NRL, CIRA, PSU
5. Verification Team	**Tim Marchok (GFDL), Barb Brown (NCAR),** NRL, NESDIS/STAR, AOML, NHC, EMC, ESRL, NWS/OST
6. Applications Development and Diagnostics	**Mark DeMaria (NESDIS/STAR), Ed Rappaport (NHC),** EMC, NRL, AOML, NCAR, ESRL, OU, FSU, NHC, NWS/OST
7. Hurricane Observations	**Sim Aberson (AOML), John Knaff (NESDIS/STAR),** NHC, EMC, NESDIS/STAR, ESRL, URI, NRL, AOC, RSMAS, NCO, NCAR,NWS
8. Ocean/Wave Models	**Hendrik Tolman (EMC), George Halliwell (AOML),** URI, ESRL, NRL, RSMAS
9. Societal Impacts	**Rick Knabb (NHC), Jennifer Sprague(NWS/OASST),** NWS/SR, NWS/ER, FEMA,CT-EM, TX-EM, NC-EM, FL-EM, Weather Channel

Table2. Strategic Teams 2013

FY 2013 Teams	FY 2013 Team Leads
1. HFIP Model Strategy	**Vijay Tallapragada (EMC), Stan Benjamin (ESRL),** AOML, NRL,GFDL, NCAR, SUNY Albany, URI, EMC, NHC
2. Model Physics	**Young Kwon (EMC), Jian-Wen Bao (ESRL),** GFDL, EMC, AOML, NRL,URI,UM CIMAS, RSMAS, UMD, UCLA
3. Data Assimilation/Initialization	**John Derber (EMC), Xuguang.Wang (OU),** NRL,UM CIMAS ESRL, CIRA, PSU, AOML, U. Utah, NCAR
4. Ensemble Development	**Jeff Whitaker (ESRL), Jiayi Peng (EMC),** NRL, UM CIMAS,PSU,NHC, NCAR, FSU, SUNY Albany
5. Post Processing and Verification Development Team	**Mark DeMaria (NESDIS/STAR), David Zelinsky (NHC), Tim Marchok (GFDL),** EMC,AOML,CIRA/CSU, NRL,NCAR,NHC,ESRL,FSU
6. Societal Impacts	**Jennifer Sprague (NWS), Rick Knabb(NHC),** NWS/SR, NWS/ER, FEMA,CT-EM, TX-EM, NC-EM, FL-EM, Weather Channel, CIRA/CSU

Table 3. Tiger Teams 2013

FY 2013 Teams	Strategic Team	FY2013 Team Leads
1. Web Page Design	5	**Paula McCaslin (ESRL), Thiago Quirino (NHC)** NCAR, NHC, PO, NRL, ESRL, CIRA, GFDL, EMC
2. 3 KM Physics Package	2	**Joe Cione (AOML), Chan Kieu (EMC),** ESRL, AOML, URI, GFDL, NCAR, NRL, UM CIMAS, UCLA
3. Regional Hybrid DA System	3	**John Derber (EMC), Jeff Whitaker (ESRL),** DTC, AOML, GFDL, UM CIMAS, NCAR, OU,
4 Use of Satellite Data in Hurricane Initialization	3	**Xiaolei Zoa (FSU), John Knaff (NESDIS/STAR), Emily Liu (EMC),** AOML, NHC, NRL, ESRL, UM, JPL,UW
5. Stream 1.5 and Demo System Implementation	1	**James Franklin (NHC), Barb Brown (NCAR),** SUNY Albany, NRL, UW
6. Reconnaissance Data Impacts	1	**James Franklin (NHC), Vijay Tallapragada (EMC),** AOML, NCAR, NESDIS/STAR, PSU, SUNY Albany

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

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

A major component of Stream 2 is an Experimental Forecast System (EFS) that HFIP runs each hurricane season. The purpose of the EFS (also known as the Demonstration Project) is to evaluate the strengths and weaknesses of promising new approaches that are testable only with enhanced computing capabilities. The progress of Stream 2 work is evaluated each off-season to identify techniques that appear particularly promising to operational forecasters and/or modelers. These potential advances can be blended into the operational implementation plans through subsequent Stream 1 activities, or developed further outside of operations within Stream 2. Stream 2 models represent cutting-edge approaches that have little or no track record;

consequently NHC forecasters do not use these models to prepare their operational forecasts or warnings.

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

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

3. The HFIP Model Systems

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

a. The Global Models

Global models provide the foundation for all of HFIP's modeling effort. They provide hurricane forecasts of their own, and are top-tier performers for hurricane track. They also provide background data and/or boundary conditions for regional and statistical models, and can be used to construct single-model ensembles, or be members of multi-model ensembles. Throughout 2012 HFIP EFS involved three global models: the Flow-following finite-volume Icosahedral Model (FIM), the Global Forecast System (GFS), and the Navy Operational Global Atmospheric Prediction System (NOGAPS).

Built by NOAA/ESRL, the FIM is an experimental global model that can be run at various resolutions and uses initial conditions from a number of sources (Benjamin et. al. 2004, NOAA ESRL 2011). It is currently using a constant sea surface temperature underneath.

The GFS, the NWS's global model, currently has two versions in use by the HFIP EFS. One of these is the current operational model run at NOAA NCEP. The second is an experimental version developed at ESRL, which differs from the operational GFS by featuring a fixed ocean and an ensemble-Kalman-filter (EnKF)-based initialization system run at higher resolution than the operational GFS hybrid grid-point statistical interpolation (GSI) data assimilation (DA) system (see section 3d below).

HFIP used the operational NOGAPS model (Peng et al. 2004; Hogan and Rosmond 1991) throughout 2012. In February 2013, NOGAPS was replaced by the Navy Global Environmental Model (NAVGEM), developed at NRL. NAVGEM includes semi-Lagrangian advection and more advanced physics than NOGAPS, including 2-species prognostic clouds.

Some specifics of the global models are shown in Table 4.

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	27 km	64	From 2010 GFS - Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	Noah LSM	GFDL/Rapid Radiative TransferModel (RRTM)	ESRL EnKF
GFS/EnKF	27 km	64	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	Noah LSM	GFDL scheme	ESRL EnKF
GFS/GSI	27 km	64	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	Noah LSM	GFDL	GSI
NOGAPS	41 km	42	Emanuel	N/A	NOGAPS	NOGAPS	Harshvardhan/ Fu-Liou	NRL Atmospheric Variational Data Assimilation System-accelerated representer (NAVDAS-AR)

b. The Regional Models

Specifics of the regional models are shown in Table 5. Note that GFDL (OPS) and HWRF (OPS) refer to the current operational regional models. The Weather Research and Forecasting (WRF) modeling system in use by HFIP contains two options for its dynamic core, and several options for physics as well as initialization and post processing systems (for a list of model publications, see DTC 2012a). The two dynamic core configurations are the Advanced Research WRF (ARW), built by NCAR, and the Non-hydrostatic Mesoscale Model (NMM), built by EMC.

The operational NCEP Hurricane WRF (HWRF) derives from the NMM dynamic core and has a movable, triply nested grid capability for one 3-km innermost nest and one 9-km intermediate nest. A coarser outer domain covers an 80° x 80° region at 27-km resolution. The model has 42 vertical layers. Advanced physics include atmosphere/ocean fluxes, coupling with the Princeton Ocean Model, and the NCEP GFS and GFDL physics. HFIP also supports the WRF ARW system, which NCAR runs using a simplified one-dimensional model of the ocean. It features three interactive nests with an inner-nest resolution of 4 km.

The PSU Regional Ensemble constitutes another version of the WRF ARW system, with similarities to the NCAR WRF ARW. It uses a static interactive inner nest of 3 km but no interactive ocean (PSU 2011; Zhang et al 2011; Weng and Zhang 2012; Snyder and Zhang 2003).

The Coupled Ocean/Atmosphere Mesoscale Prediction System – Tropical Cyclone (COAMPS-TC©) (Doyle et al. 2012) and the Wisconsin Model are detailed in the table and have been members of the stream 1.5 suite of models each year. Note that COAMPS-TC© features an interactive ocean (Chen et al. 2010 and Doyle et al. 2012).

Table 5. Specifications for the HFIP Regional Models

Models	**Domains / Horizontal Resolution (km)**	**Vertical Levels core**	**Cumulus Parameterization**	**Microphysics**	**PBL**	**Land Surface**	**Radiation**	**Initial and Boundary Conditions**	**Initialization**
HWRF (OPS)	3 27/9/3	42 NMM	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	GFDL Scheme	GFS	GSI with vortex initialization
GFDL (OPS)	3 55/18/9	42 GFDL	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	Schwarzkopf-Fels (longwave) / Lacis-Hansen (shortwave)	GFS	GFDL Synthetic Bogus Vortex
GFDL (Ens)	3 55/18/9	42 GFDL	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	Schwarzkopf-Fels (longwave) / Lacis-Hansen (shortwave)	GFS for 8 members; GEFS for other 8 members	GFDL Synthetic Bogus Vortex
HWRF-HRD/EMC Basin Scale	3 27/9/3	42 NMM	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	GFDL Scheme	GFS	One-way hybrid GSI-EnKF with vortex initialization
HWRF-HRD	2 9/3	42 NMM	Simplified Arakawa Schubert	Ferrier	GFS Non-Local PBL	GFDL Slab Model	GFDL Scheme	GFS	EnKF with aircraft data
AHW (NCAR)	3 36/12/4	36 ARW	Tiedtke (36/12 km only)	WSM6	YSU	NOAH LSM	RRTMG (LW+SW)	GFS (BC only)	EnKF method in a 6-hour cycling mode
COAMPS-TC©	3 45/15/5 (15/5 km following the storm)	40 COAMPS	Kain Fritsch on 45 and 15 km meshes	Explicit microphysics (5 class bulk scheme)	Navy 1.5 Order Closure	Slab with the NOAH LSM as an option	Fu-Liou	GFS in WATL and EPAC	3D-Var data assimilation with synthetic observations
Wisconsin Model	2 40/8	42 UW-NMS	Modified Emanuel	Tripoli-Flatau Bulk microphysics (1 liquid, 2 ice categories)	1.5 Order Closure	NOAH LSM	RRTM SW and LW	GFS	Bogus vortex with 12-hour dynamic initialization
Penn State ARW	3 27/9/3	42 ARW	Grell-Devenyi ensemble scheme (27 km only)	WSM 6-class graupel scheme	YSU	5-layer thermal diffusion scheme	RRTM (longwave) / Dudhia (shortwave)	GFS	EnKF with NOAA airborne radar

c. Initialization and Data Assimilation Systems

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

1) Global Forecast System: The initial state created for the current operational global model (GFS) is interpolated to the grids used by HFIP global models. The GFS in 2012 used the new Hybrid Ensemble-Variational DA System (HEVDAS—see below) that is a combination of the GSI system formerly used and an ensemble based system to define the background error matrix. The GSI initialization system that has been run operationally since 2006 is a three-dimensional variational approach (3D-VAR) (DTC 2011b; Wu et al. 2002; Parrish and Derber 1992; Cohn and Parrish 1991).

2) HWRF: The 2012 operational HWRF used an advanced vortex initialization and assimilation cycle consisting of four major steps: 1) interpolation of the global analysis fields from the GFS 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 bogus vortex for a cold start); and 4) addition of data observed outside of the hurricane area using GSI. The DA system is capable of ingesting inner core data to optimize the vortex initialization.

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

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

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

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

7) Vortex Initialization: The initial vortex for some of the regional models is produced by a vortex initialization procedure. First, the vortex circulation is filtered from the first guess fields interpolated from global model; then a new vortex modified by the observed intensity is inserted back in the filtered environment. The new vortex is the model balanced vortex cycled from previous six-hour forecast 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. None of the other HFIP global models are currently using vortex relocation.

d. The HWRF Community Code Repository

During 2009-2011 EMC and the Developmental Testbed Center (DTC) unified the operational and research versions of HWRF and created a code management protocol, making the operational model completely compatible with the codes in the central DTC repository. In 2012, researchers had direct access to the repository, encouraging code-sharing and allowing development using the latest code, making improvements in HWRF easily transferable into operations. This was one of the initial goals of the WRF program. DTC maintains a portal (www.dtcenter.org/HurrWRF/users) providing access to HWRF and AHW documentation, datasets, and tutorials. That portal includes a link to the GFDL vortex tracker for users with models other than HWRF.

4. Meeting the HFIP Goals

a. The HFIP Baseline

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

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.

Since a sample of cases from, say, the 2011 season might have a different inherent level of difficulty from the baseline sample of 2006-8 (for example, because it had an unusually high or low number of rapidly intensifying storms), it is necessary to evaluate the progress of the HFIP models in terms of forecast skill, rather than in terms of error. Figure 1 displays the skill of the baseline errors and the 5- and 10-year goals represented in blue and labeled on the right side of the graph is the percentage improvement over the Decay-SHIFOR5 and CLIPER5 forecasts for the same cases. Note the baseline skill 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.

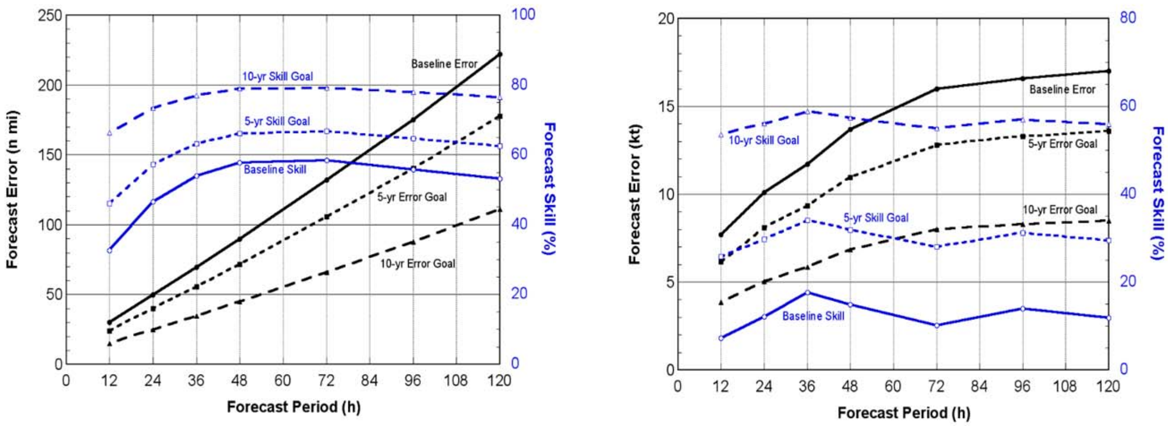


Figure 1. HFIP baseline track (left panel) and intensity errors (righ panel). The baseline errors (black lines) were determined from an average of the top-flight operational models during the period 2006-2008. The HFIP expressed goals (dashed lines) are to reduce this error by 20% in 5 years and 50% 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 also important to note that these HFIP performance goal baselines were determined from a class of operational aids known as "early" models. Early models are those that are available to forecasters early enough to meet forecast deadlines for the synoptic cycle. Nearly all the dynamical models currently in use at tropical cyclone forecast centers, however, (such as the GFS or the GFDL model, referred to as GFDL, for short) are considered "late" models because their results arrive too late to be used in the forecast for the current synoptic cycle. For example, the 1200 Coordinated Universal Time (UTC) GFDL run does not become available to forecasters until around 1600 UTC, whereas the NHC official forecast based on the 1200 UTC initialization must be issued by 1500 UTC, one hour before the GFDL forecast can be viewed. It's actually the older, 0600 UTC run of the GFDL that would be used as input for the 1500 UTC official NHC forecast, through a procedure developed to adjust the 0600 UTC model run to match the actual storm location and intensity at 1200 UTC. This adjustment, or interpolation, procedure creates the 1200 UTC "early" aid GFDI that can be used for the 1500 UTC NHC forecast. Model results so adjusted are denoted with an "I" (e.g., GFDI). The distinction between early and late models is important to assessing model performance, since late models have an advantage of more recent observations/analysis than their early counterparts.

b. Meeting the Track Goals

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

c. Reaching the Intensity Goals

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

5. HFIP Stream 1.5

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

Eight models/modeling systems were provided to NHC in 2012 under Stream 1.5; these are listed in Table 6. Note that most models were admitted into Stream 1.5 based on the models' performance forecasting either track or intensity, but generally not both. For example, forecasters were instructed to consult the COAMPS-TC© model, interpolated ahead 6 hours, (COTI) intensity forecasts but not the COTI track forecasts. Two HFIP Stream 1.5 consensus aids were constructed: the track consensus TV15 comprised the operational models GFSI, EGRI, GHMI, HWFI, GFNI[3], EMXI and the Stream 1.5 models AHWI, APSI, and FM9I, while the intensity consensus IV15 comprised the operational models Decay-SHIP (DSHP), LGEM, GHMI, HWFI and the Stream 1.5 models AHWI, COTI, APSI, and UWNI.

a. Stream 1.5 retrospective testing

The 2012 retrospective testing focused on storms in the Atlantic and eastern North Pacific basins during the period 1 August – 31 October for the 2009, 2010 and 2011 hurricane seasons. Eight modeling groups participated in this year's exercise (Table 6), including advanced HWRF models from NCAR/SUNY-Albany (AHW) and PSU (ARW), the University of Wisconsin Nonhydrostatic Modeling System (UW-NMS), COAMPS-TC, the GFDL ensemble (see Section 6), FIM, and the Statistical Prediction of Intensity from a Consensus Ensemble (SPICE). The evaluation focused on one-to-one comparions between the Stream 1.5 candidate and NHC top-flight models (the European Centre for Medium-range Weather Forecasts model, ECMWF, GFS

[3] GFNI was formally part of the Stream 1.5 TV15 consensus and TVCA, but it was unavailable in 2012.

and GFDL for track; LGEM, Decay SHIPS, DSHP, and GFDL for intensity). For the deterministic models and GFDL ensemble, the evaluation also considered one-to-one comparisons between operational consensus guidance and a consensus based on the operational members plus the Stream 1.5 candidate, whereas the consensus Stream 1.5 candidates were simply compared directly to the operational consensus guidance. And finally, the performance of each candidate was compared to the top-flight models as a group.

Table 6. List of modeling groups that participated in the 2012 HFIP Retrospective Evaluation and the form of the model guidance selected, if any, for Stream 1.5 during the 2012 HFIP Demo.

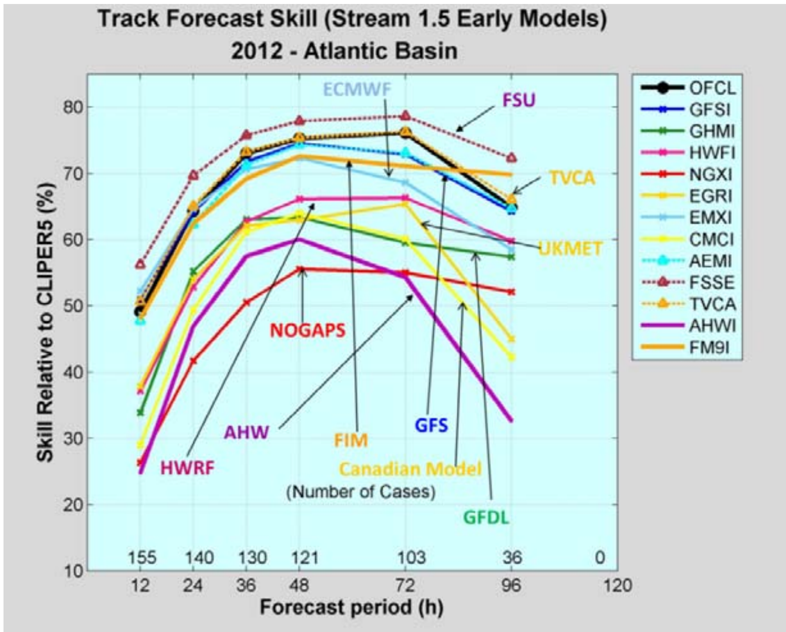
Organization	Model (ATCF ID)	Track	Track Consensus	Intensity	Intensity Consensus
NCAR/SUNY-Albany	Advanced HWRF, AHW (AHWI)		●		●
UW-Madison	UW-NMS (UWNI)				●
NRL	COAMPS-TC (COTI)				●
PSU	ARW (APSI)		●	●	●
GFDL	GFDL ensemble mean (GPMI)	●		●	
	No-bogus member (G01I)	●		●	
ESRL	FIM (FM9I)		●		
NESDIS	SPICE (SPC3)			●	

All evaluations were based on homogeneous samples and applied the appropriate method for assessing statistical significance. A detailed description of the metholodgy used for the evaluation, reports for each participating modeling group and all the verification plots generated during the evaluation are available on the TCMT 2012 Retrospective HFIP Testing website (http://www.ral.ucar.edu/projects/hfip/h2012/), as well as information on the participating models and cases included in the evaluation.

b. Stream 1.5 results

The results presented here reflect the Stream 1.5 runs that were successfully transmitted to NHC in real time during 2012. The Stream 1.5 models arriving at NHC underwent standard operational processing to convert "late" dynamical guidance into "early" interpolated guidance that could be used by the forecasters. Figure 2 (top) presents a homogeneous verification of the Stream 1.5 track models that met availability standards (and regardless of whether they were intended for use explicitly or in a consensus), along with selected operational models. The figure shows that in 2012 the FM9I was competitive with the top-tier operational models, although the

AHWI was not. The GFSI, the operational GFS using the hybrid data assimilation scheme, was among the best performers and outperformed the European Center Model (EMXI) in this sample. Figure 2 (bottom) shows that there was very little impact from adding the Stream 1.5 models to the track consensus through 48 hours, and then a slight negative effect from 72 to 120 hours.



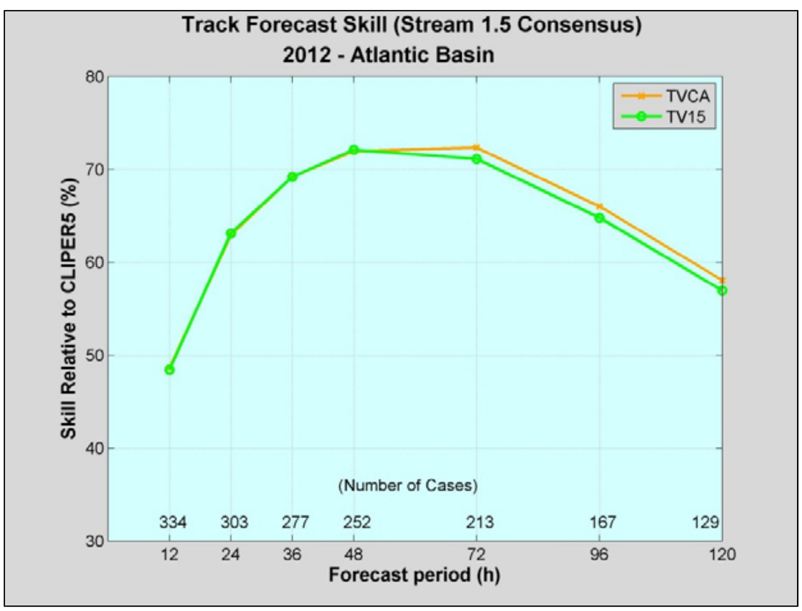
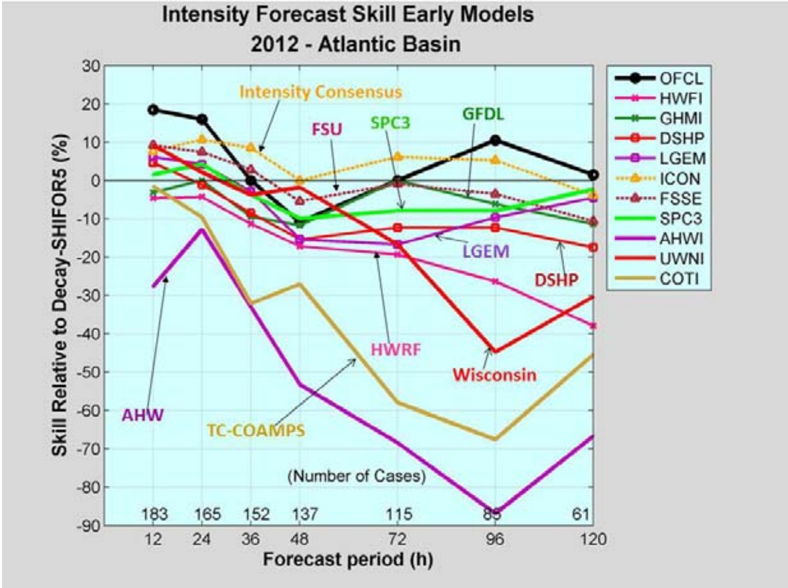


Figure 2. Homogeneous comparison of track skill (error relative to CLIPER5) of Stream 1.5 models and selected operational models for 2012, upper panel and impact of adding Stream 1.5 models to the variable track consensus TVCA, bottom panel. See Appendix for definition of model acronyms.

Stream 1.5 intensity results are shown in Figure 3 (top), for a sample that, due to limited availability, excludes the PSU Doppler runs. The SPC3 was the best performing Stream 1.5 intensity model, an unsurprising result given that SPC3 represents an intelligent consensus of the already top-tier dynamical-statistical models LGEM and DSHP (see Section 11).



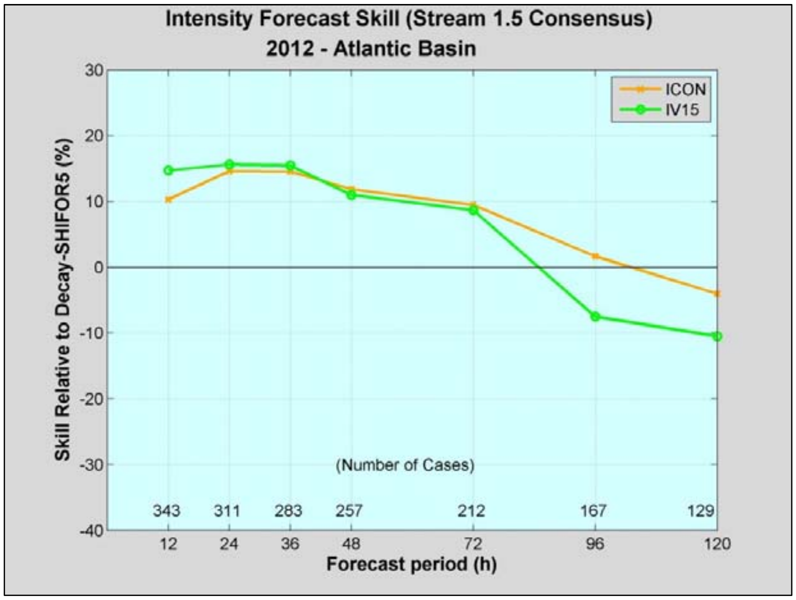


Figure 3. Homogeneous comparison of intensity skill (error relative to Decay-SHIFOR5) of Stream 1.5 models and selected operational models for 2012, upper panel and impact of adding Stream 1.5 models to the fixed intensity consensus, ICON, bottom panel. See Appendix for definition of model acronyms. . In the upper panel, all models available to the forecasters are shown including those with 6 hour interpolations and those with 12-hour interpolations.

The Stream 1.5 models COTI and AHWI performed very poorly; these models had no skill throughout the forecast period and performed worse than all of the operational models. (It should be noted that due to limited availability, the COTI verification contains a relatively high proportion of 12-hr interpolations, which lowers performance. However, even the skill of the 6-hour interpolated COTI was no better than that of the poor-performing HWFI.) UWNI was a better intensity model in 2012 than in 2011, but its skill was still near the HWFI at 96 and 120 hours. The impact of the Stream 1.5 models was slightly positive to the intensity consensus from 12 to 36 hours, but noticeably negative at the longer forecast times (Figure 3, bottom).

The upper panel of Figure 3 contains a mix of 6- and 12-hour interpolated early models, following the standard operational verification methodology. The verification shown in Figure 4 excludes the 12-hour interpolated models, which reduces the sample size but eliminates any disadvantage a model might have due to limited or delayed availability.

.

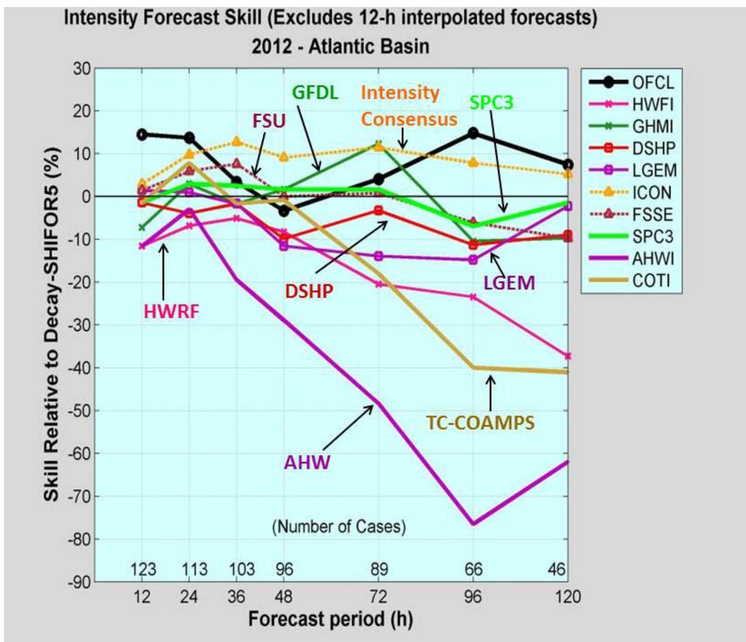
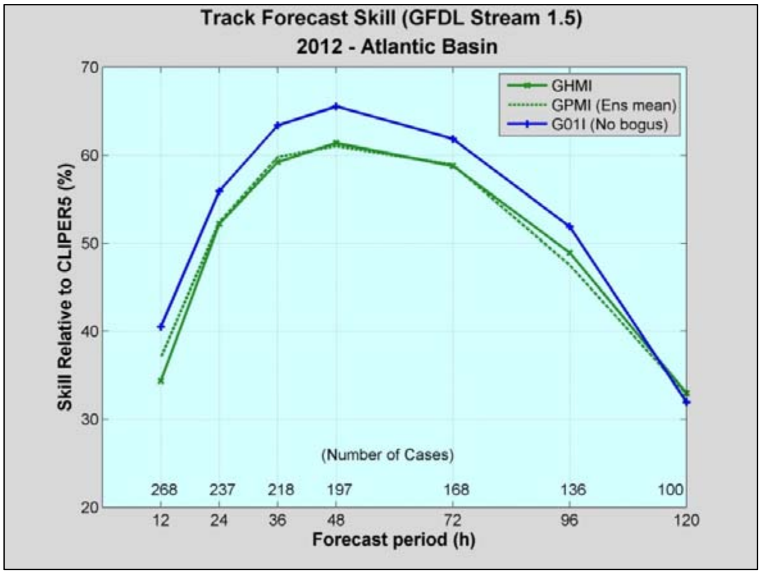


Figure 4. Same as Figure 3, upper panel, but with those cases with 12-hour interpolations removed (hence fewer number of cases). The sample is still homogeneous.

Due to limited availability, the GFDL ensemble (see Section 6 and Table 7 for a description of the ensemble and its members) was evaluated separately from the other Stream 1.5 models. Figure 5 presents the track and intensity forecast skill of GHMI and the Stream 1.5 GFDL ensemble mean (GPMI) and an unbogused GFDL ensemble member (G01I). G01I performed better than GHMI and GPMI for track at all times except 120 hours, with about 5% more skill, than the operation GFDL and its ensemble mean for track in 2012. For intensity, the GFDL ensemble mean was not consistently better than its deterministic run, and G01I performed worse than GHMI and GPMI at most times. It should be noted than none of these models had any skill for intensity prediction throughout the forecast period.



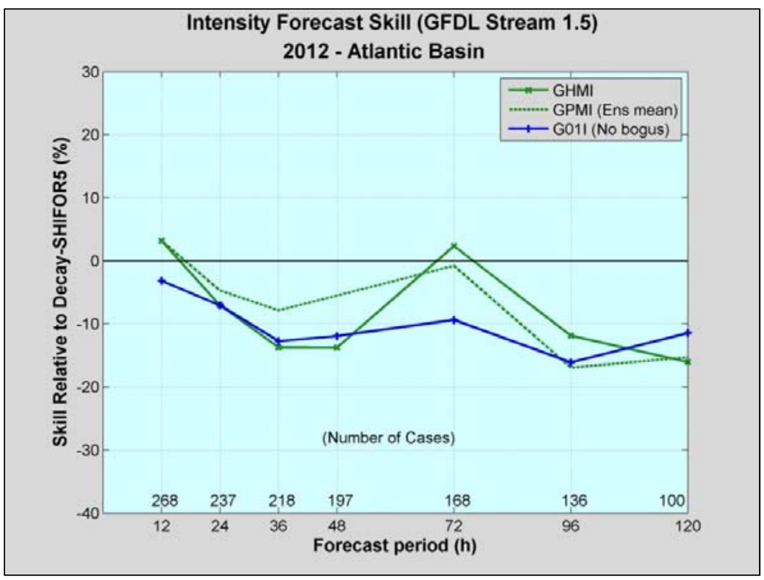


Figure 5. Homogeneous comparison of HFIP Stream 1.5 GFDL ensemble mean and GFDL unbogused ensemble member for track (top) and intensity (bottom)

Overall, the skill of the Stream 1.5 guidance delivered to NHC in 2012 was disappointing. The FM9I did show about equivalent skill to the high performing operational models for track, and the dynamical-statistical consensus SPC3 did have more intensity skill than its individual components.

Figure 3 demonstrates the main issue facing the HFIP goal on intensity. The results shown in Figure 3 are similar, so far, to *any* model forecast displayed in this way. It basically says any model (dynamical or a more complex statistical model) has a hard time beating even the very simplest statistical model. The Decay-SHIFOR basis for its forecasts on only 7 parameters; the current position and intensity, the position and intensity 12 hours earlier and the current date. The weights applied to these numbers are of course determined statistically from historical data.

The HFIP plan to change the situation noted in Figure 3 is:

- Use advanced data assimilation methods in the regional models and
- Incorporate high resolution data taken by aircraft and satellite near the core.

HFIP is in the process of developing a regional hybrid data assimilation system, similar to what was made operational last summer in the GFS global model and which, as noted below, has made a huge improvement in track forecasts from that model. The regional hybrid DA system won't be available until the 2014 season and hence the results shown Figure 3 don't reflect that proposed improvement. This last year we did incorporate aircraft data in those runs for which such data were available. We show results with the aircraft data in the next section. However there were only 19 cases with aircraft data and the impact of those data won't show up in Figure 3 where cases from throughout the season, almost 200, are shown.

6. Operational Hurricane Guidance Improvements

The HFIP goals described in section 4 are only met when the model guidance provided to NHC NCEP reaches those goals. Since 2012 represents the third year of the program we would expect to see progress toward meeting the five year goals in the operational models and not just in the experimental models such as the stream 1.5 models described in the previous section. In this section the emphasis will be on improvements in the hurricane forecasts from the models operational in 2012. This includes the global GFS model, and the HWRF and GFDL operational regional models.

a. Global Model (GFS)

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

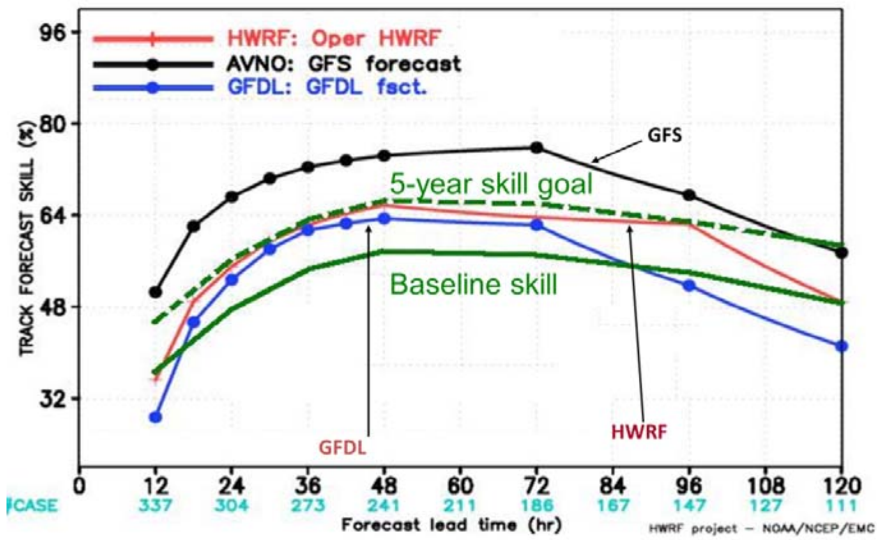


Figure 6. Track Skill (error relative to CLIPER5) for the NCEP Operational models including the global and regional models. Models are indicated on the panel. The green dashed line is the 5-year HFIP goal shown in terms of skill

Figure 6 shows track errors as a percent improvement over CLIPER5 which is often referred to as "Skill". Here we just emphasize the three operational numerical prediction guidance systems that NCEP runs related to hurricanes; the global model GFS, the HWRF and the GFDL. HWRF and GFDL will be discussed further in the next sections.

The baseline and the 5 year goal of a 20% improvement in track guidance is also shown on the plot as skill; i.e., percent improvement over CLIPER5. The data shown in Figures 6 and 7 are a little different than those shown in Figures 2 and 3 because Figures 6 and 7 include more cases.

Note that all three models are showing track skills at or near the 5-year goal and the GFS shows skill considerably above the 5-year goal, approaching the 10-year goal of a 50% decrease in track error (at 72 hours the 10 year goal is an 80% improvement over CLIPER). Since comparing relative to CLIPER errors removes most of the year to year variation from the results, what is displayed in Figure 6 represents a true improvement in skill and is not simply due to 2012 being an easy year for track (if indeed it was).

Figure 7 shows intensity skills relative to the skill of Decay-SHIFOR5. Here, when compared to the improvements in track forecasts, the results are not nearly as impressive. None of the models even come up to the baseline skill let alone the 5 year goal. This has already been noted in the previous section though here we also include the intensity skill of the global model GFS. Surprisingly, the intensity forecasts by the GFS actually exceeded, in skill, the intensity forecasts of the two operational regional models beyond 48 hours. The HWRF did better than either the GFS or the GFDL in the first 36 hours though not at the skill level of the Decay-SHIFOR5

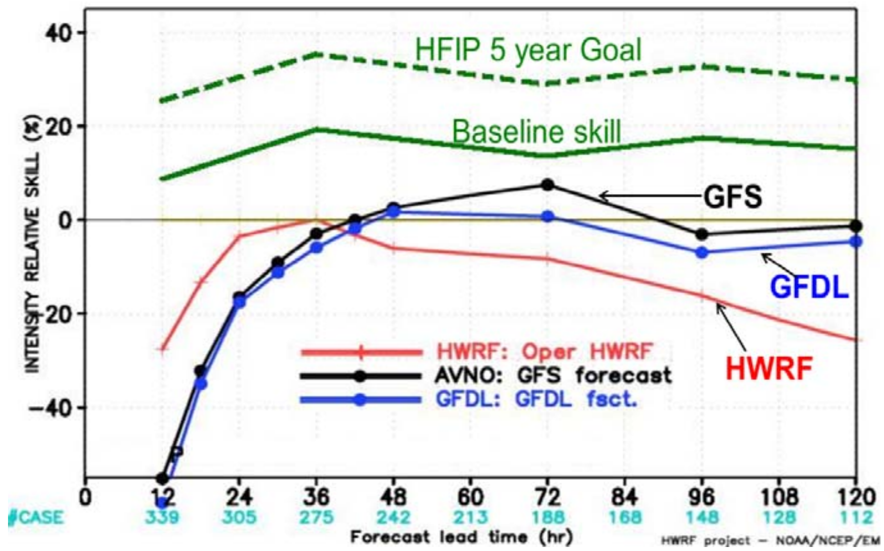


Figure 7. Intensity Skill (error relative to Decay-SHIFOR5) for the NCEP Operational models including the global and regional models. Models are indicated on the panel. The green dashed line is the 5-year HFIP goal shown in terms of skill.

The reason for the apparent lack of progress against Decay-SHIFOR and also the baseline has been described in the previous section on Stream 1.5; the impact of the aircraft data cases were too few in number in 2012 to impact the overall results shown in Figure 7 and we have not yet implemented the hybrid data assimilation system in the regional models.

b. Hurricane WRF (HWRF)

Even though the results for HWRF in 2012 fell short of the 5 year HFIP goal for intensity, Figure 7, the HWRF model has undergone a vast improvement over previous versions of the operational HWRF. There were two major changes in HWRF for the 2012 hurricane season: the introduction of a third nest allowing an inner core resolution of 3 km and a stream 2.0 demonstration of using aircraft data in HWRF. There were other changes including physics package upgrades, bug fixes etc. that also led to improvements in the model but the introduction of the third nest was a game changer.

We discuss the impact of the aircraft data on HFIP forecasts in a later section (Figures 12 and 13) and here will just focus on the upgrading of the HWRF system with the third nest.

Up until this past hurricane season, HWRF had been run with two nests; an outer domain with a 27-km grid spacing and an inner nest of 9 km. The inner nest moves with the hurricane and interacts with the outer domain. HFIP results from previous years and described in annual reports indicated that a higher resolution inner or third nest provided superior results. Hence the HWRF team at EMC undertook a project to include an interactive third nest with a resolution of 3 km that would fit in the operational time slot allotted to the HWRF system. This was a major effort requiring, among other things, making the code more efficient so that it would fit in the time slot which was only modestly expanded for this improvement. EMC noted that many in the HFIP community aided in the effort of making the triple nest HWRF operational:

- **EMC**: Computational efficiency, nest motion algorithm, physics improvements, 3km initialization and pre-implementation Testing and evaluation (T&E)
- **HRD/AOML**: nest motion algorithm, multiple moving nests, Planetary Boundary Layer (PBL) upgrades, interpolation for initialization
- **DTC:** Code management and repository
- **NCAR**: Maximum Potential Intensity (MPI) profiling
- **ESRL**: Physics sensitivity tests and idealized capability
- **URI**: 1D ocean coupling in East Pacific basin
- **GFDL**: Knowledge sharing, joint T&E
- **NHC**: Diagnostics and evaluation of the HWRF pre-implementation tests and real-time guidance

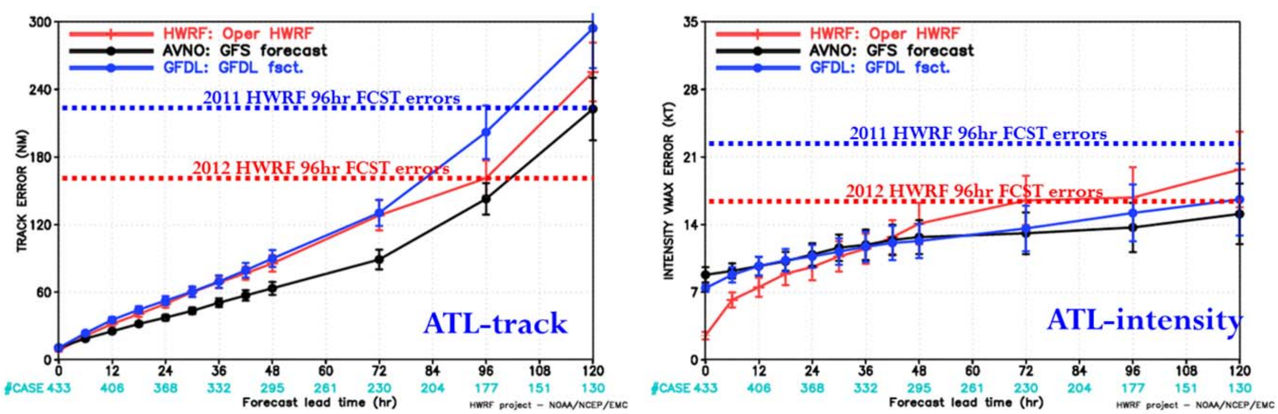


Figure 8. Track (left) and intensity errors for HWRF, GFDL and GFS Operational models for the Atlantic. Dashed lines indicate the improvement of the 2012 HWRF model over the 2011 HWRF. The sample is homogeneous.

Figure 8 shows the 2012 errors for both track and intensity for the GFS, HWRF and GFDL models. This is the same data shown in Figures 6 and 7 but now shown as error, not relative to CLIPER and Decay-SHIFOR, respectively. The main message of Figure 8 is contained in the two dashed lines shown on both panels. The blue dashed lines are the HWRF errors at 96 hours using the 2011 version of HWRF and the red dashed lines, the errors using the 2012 version. The sample is homogeneous. Note that the track errors improved by 27% over the 2011 version and the intensity errors by 25% in the same period.

This level of improvement from one annual version to the next has never been observed before for the operational hurricane models. We attribute this improvement largely to the introduction of the third nest with its higher resolution near the hurricane core.

7. GFDL Ensemble

Results from the 2012 operational GFDL model are shown in Figures 9, 10 and 11. This model is still very competitive with the other dynamical models (see Figures 3, 4 and 5 for a comparison with the stream 1.5 models) even though it is expected to be replaced eventually by the HWRF model. It ranked near the top of the dynamical models for intensity forecasts (see Figure 3).

For the last three 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 7). Working with the forecasters at NHC scientists at GFDL constructed an ensemble by varying various parameters in the initial conditions and sea surface temperatures used by the model. The "unbogussed" forecasts start from the GFS or GEFS without modification to the vortex from what was in the GFS/GEFS initially. The other members use the GFDL initialization scheme but modified as described in Table 7. Members 08-15 use GEFS Mean for constructing the initial conditions and the boundary conditions while 00-07 use the GFS for constructing the initial conditions and the boundary conditions.

Table 7, GFDL Hurricane Model Ensemble Members.

ATCF ID	Description
GP00/GP08	Control forecast (same model as NCEP 2012 operational GFDL)
GP01/GP09	Unbogussed forecast using the 2012 GFDL control model
GP02/GP10	Increase NHC-observed V_{max} +10%, R34 +25%, R50 +40%, ROCI +25%
GP03/GP11	Decrease NHC-observed V_{max} -10%, R34 -25%, R50 -40%, ROCI -25%
GP04/GP12	Increase inner-core moisture by a max of +10%
GP05/GP13	Decrease inner-core moisture by a max of -10%
GP06/GP14	Increase SSTs by a max of +1°C within the initial extent of the TC
GP07/GP15	Decrease SSTs by a max of -2°C within the initial extent of the TC
GPMN	Ensemble mean computed at each lead time where the member availability is at least 6 members (40% threshold)

Figure 9 shows the intensity forecasts for each of the members of the GFDL ensemble, the ensemble mean and the spread of the ensemble. The spread indicates that the model is under-dispersive as has been noted in previous years but that it is much improved over previous years. As usual, the ensemble mean is better than all of its members and in fact produces an intensity forecast that is on the order of 10% better than the GFDL operational deterministic model.

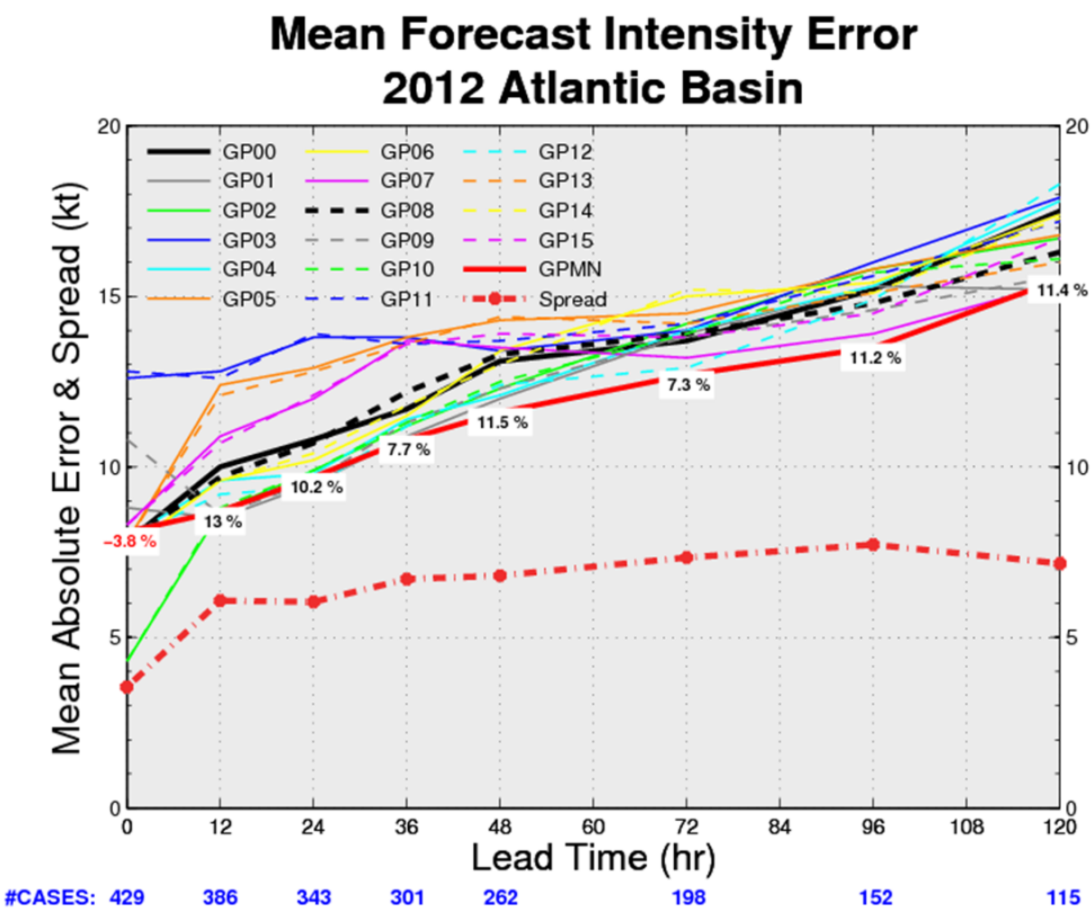


Figure 9. Intensity errors for the 2012 GFDL ensemble. The various members of the ensemble are listed at the top of the chart (see Table 7). The red lines show the ensemble mean and the percentages are the improvement of the ensemble mean over the GFDL operational (control) run. The dashed line is the ensemble spread.

Figure 10 compares the performance of the GFDL ensemble mean (GMNI—red lines) to other forecasts for both track and intensity. The comparison includes the NHC consensus forecast for track (TVCN), the Decay-SHIFOR5 (OCD5), operational HWRF (HWFI), GFS (AVNI) and the two intensity statistical models; DSHP and LGEM. The official forecast is in black. For track the performance of the mean does not compare well with the other models shown. But for intensity it is among the better performers.

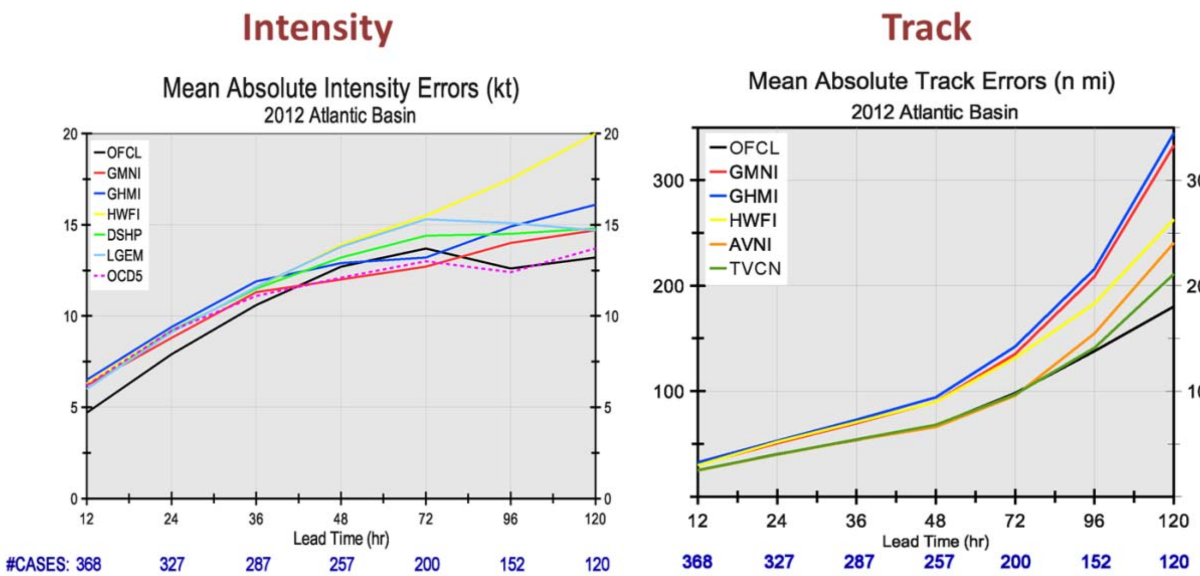


Figure 10. Intensity (left) and track errors (right) for the GFDL ensemble mean (GMNI—red) compared to Operational HWRF (HWFI) and GFDL (GHMI), the GFS (AVNI), two intensity statistical models (LGEM and DSHP) and the NHC Intensity consensus (OCD5) and the NHC track consensus (TVCN). The official forecast (OFCL) is in black.

We can use the GFDL ensemble to test the sensitivity of the various perturbations to the initial conditions on the skill of those members. Figure 11 shows the average error for 3 classes of perturbations (refer to Table 7) plus the ensemble mean. These errors are plotted as a percent improvement over the control run, the operational GFDL. Early in the forecast period the sea surface temperature perturbations had little impact on improving the forecast skill over the operational GFDL but became increasingly important by 5 days. Moisture and size perturbations had roughly the same impact through 48 hours with the moisture perturbations becoming more important later in forecast period.

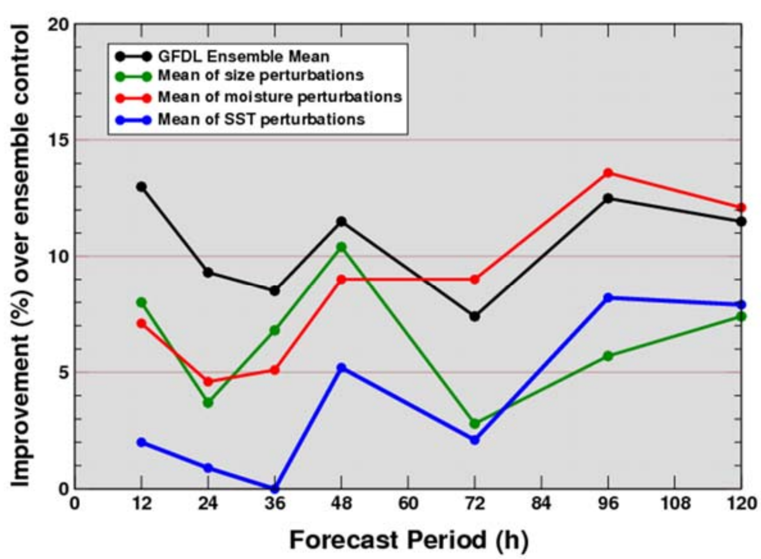


Figure 11. Intensity errors for the 2012 GFDL ensemble. Comparison shows the influence of the various classes of perturbations (see Table 7). This shows the percent improvement over the GFDL operational model (control run) so higher percentages are better

8. Real-time Assimilation of Aircraft data

In previous annual reports we noted several experiments that indicated inclusion of aircraft data in the initialization appeared to make a major improvement in the intensity forecasts even out to 5 days. This year several models both in stream 1.5 (the PSU model) and in Stream 2.0 (HWRF and HEDAS) incorporated aircraft data. As we noted earlier, there were too few cases this year (19) to make a substantial impact on the overall statistics. Results from those 19 cases are available however and shown in Figures 12 and 13.

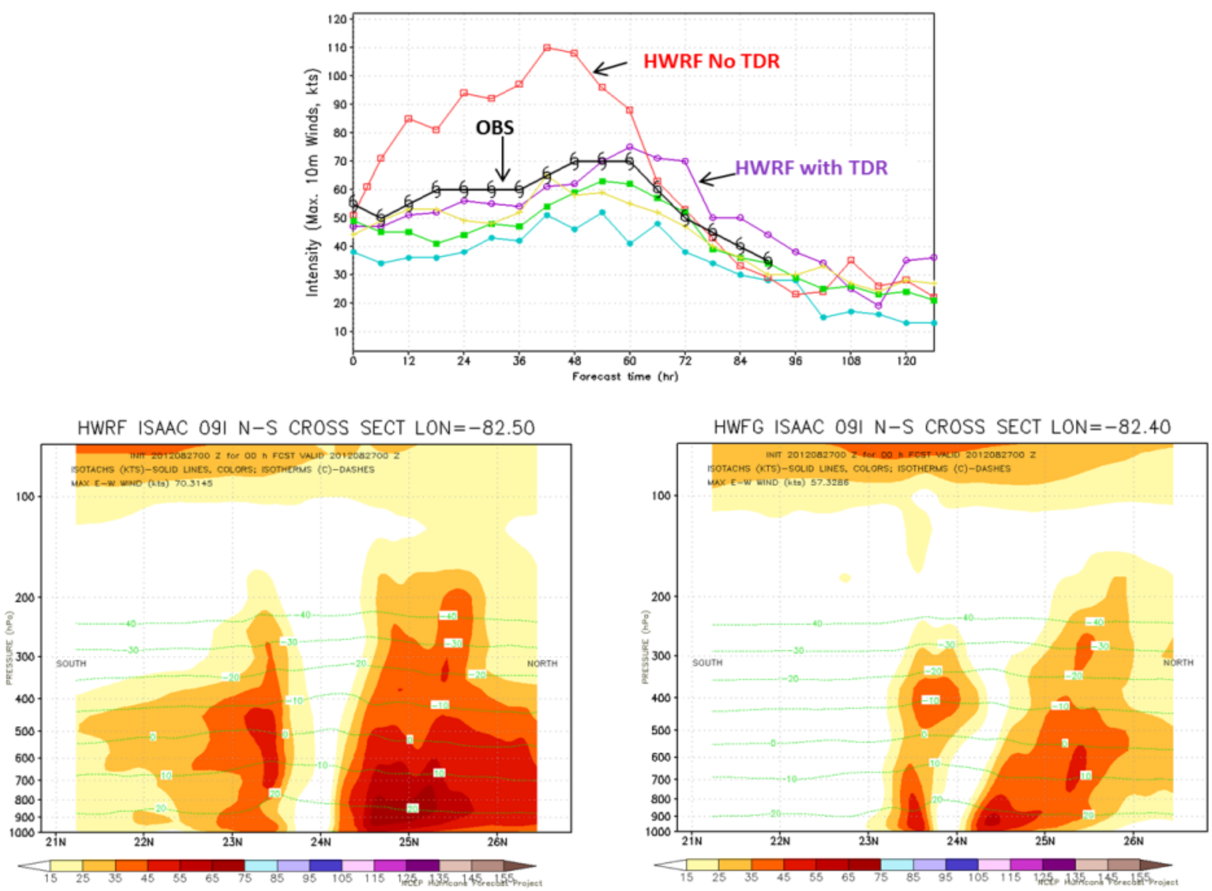


Figure 12. North-South Cross sections through Hurricane Isaac at 82.4W on August 27, 2012 at 00Z. The lower left panel shows the cross section from the initial conditions for HWRF without using radar data, the lower right panel is the same figure but the initial conditions when radar data is included. The green lines are isotherms and colored contours are isotachs, color scale is shown on the bottom. The upper panel shows the forecasts from each of the initial conditions illustrated in the lower panels. The red line shows the forecast from the lower left hand panel and the purple line the forecast from the lower right panel. In the upper panel, the yellow line shows the official forecast, the blue line the ECMWF forecast and the green line, the GFDL forecast using the ECMWF for initial and boundary conditions

Figure 12 is just one example of how inclusion of aircraft data (in this case Tail Doppler Radar, TDR, data) changed the initial conditions for a forecast. The lower left hand panel is a north-south cross section through the center of Hurricane Isaac at a time when it was off the Florida coast. These data come from the initial condition for the operational HWRF. At that time the storm was impacted by shear and had dry air to the south of the center. These factors were preventing the storm from intensifying yet many of the models, the HWRF included, indicated that the storm would strengthen significantly. The lower right hand panel shows the change in the HWRF initial condition for Isaac when the TDR data from the P3 aircraft is included in the initialization. In both panels, the coloring shows the tangential wind.

The cross section in the left hand panel indicates a vertical vortex, and quite a bit stronger than the vortex shown in the right hand panel. In the right panel, in addition, the vortex is strongly tilted and the south side is quite weak. The vortex in the lower right hand panel did not develop as much as the one in the lower left hand panel as one might expect from a tilted vortex with dry air on the up shear side of the center. The development of the vortex with and without the TDR data in HWRF is shown in the upper panel of Figure 12.

The upper panels of Figure 13 were prepared by EMC and show the average error for the same HWRF model used in operations using the operational initialization scheme (red lines) and when that model uses the tail Doppler Radar data (blue lines). Track errors are shown on the left and intensity errors on the right. The total number of cases for each forecast period is shown along the bottom and the sample is homogeneous. On the bottom panels are similar figures prepared by AOML also using the HWRF but using the HEDAS DA system for the runs with aircraft data and the operational initialization scheme when not.

The track errors were about the same for both initializations (with vs. without additional aircraft data) for both the EMC and AOML runs out to about 72 hours. Beyond that the number of cases is low, especially for the upper panels. For intensity, however, there was noticeable reduction in error out to 72 hours, a reduction on the order of 30% for both EMC and AOML. This reduction in intensity error is comparable to the reduction in error over the runs without radar data noted in earlier HFIP annual reports over a much larger set of cases. However we emphasize that the number of cases shown in Figure 13 is low and the reductions shown in Figure 13 may not be statistically significant.

HFIP has begun testing two regional models (HWRF and ARW) using a carefully constructed control to determine the impact of all available aircraft data (TDR, dropsonde, flight level and Stepped Frequency Microwave Radiometer, or SFMR,) as well the TDR data alone. This experiment will more than double the number of cases available, especially for those cases with just dropsonde and flight level data and will hopefully answer the question of the overall impact of the aircraft data and so the impact of high resolution data taken near the hurricane core.

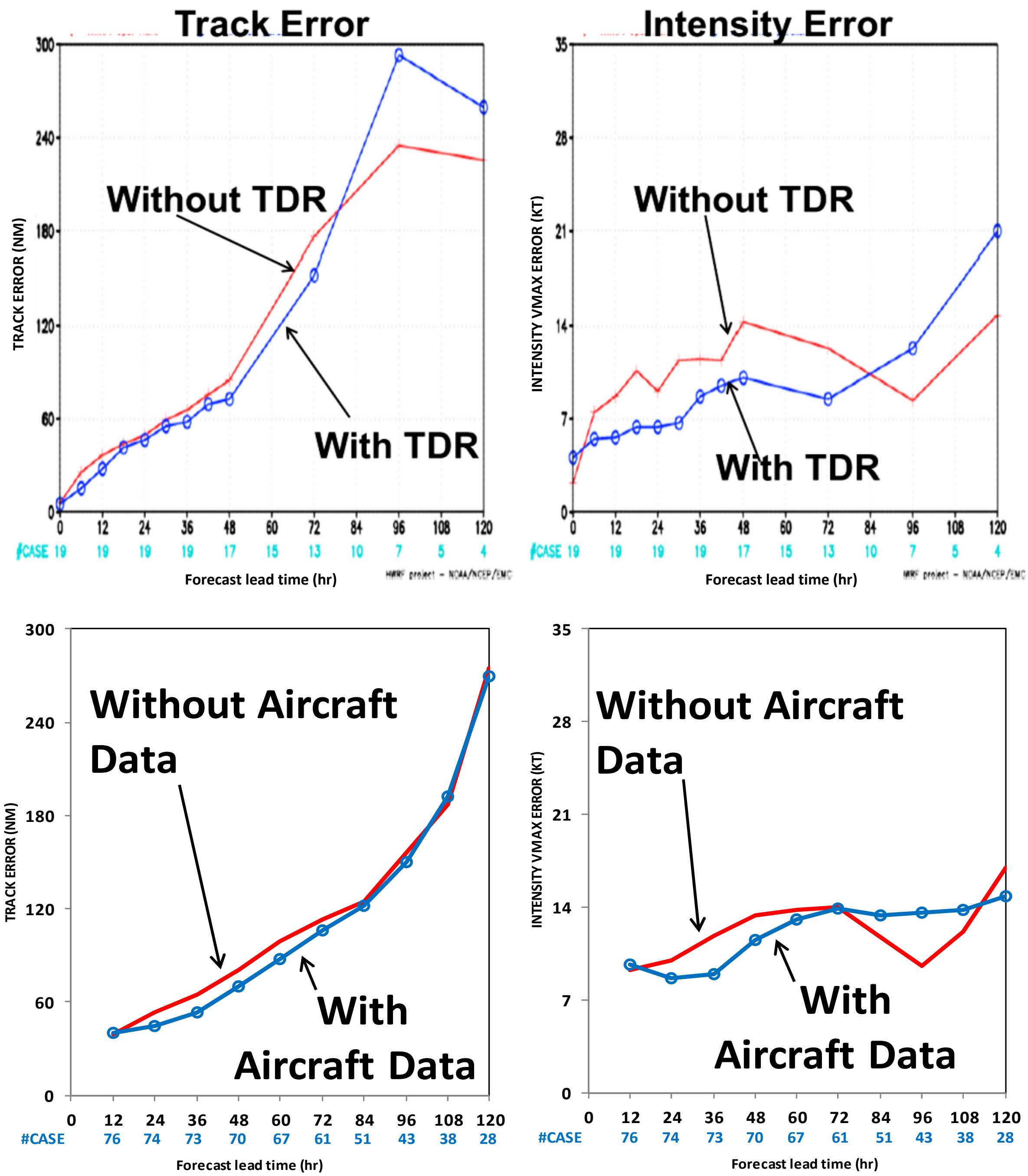


Figure 13. Upper panels; the impact of using Tail Doppler radar data (TDR) in HWRF for all cases for which radar data are available in the Atlantic in 2012. Red line shows the track (left panel) error and Intensity (right panel). The data were from the HWRF model run by EMC using the GSI data assimilation system. Lower panels; same as the upper panels but using all available aircraft data including TDR, flight level, dropsonde and Stepped Frequency Microwave Radiometer (SFMR) data. All cases for which that data were available in 2012. The number of cases is larger for the AOML plots because they include data from the Air Force C130 aircraft which do not have a TRD. For each panel the comparison is homogeneous.

9. Genesis

While not a specific goal of HFIP, hurricane genesis is an implied goal since hurricane forecasts in general extend out to at least 5 days and being able to forecast genesis is important for the longer lead times. Many hurricanes that exist at the end of the 5 -7 day forecast won't exist at the initial time. HFIP has been encouraging the development of methods to diagnose genesis in our experimental and the operational models. This involves the development of a tracker that can identify genesis in an ensemble and then an effort to determine how well the various ensembles perform. Figure 14 shows the genesis prediction for four of the operational models in 2012 (ECMWF, GFS, CMC, NOGAPS) and shows genesis in the first 48 hours of the forecast. Since these are ensembles then if 50% of the members of an individual model's forecasts showed genesis in the first 48 hours then the probability of genesis is 50% for that ensemble forecast. Thus if those cases where 50% probability of genesis was forecast and genesis occurred 50% of the time then the forecast was considered perfect (the thin diagonal line), and so forth.

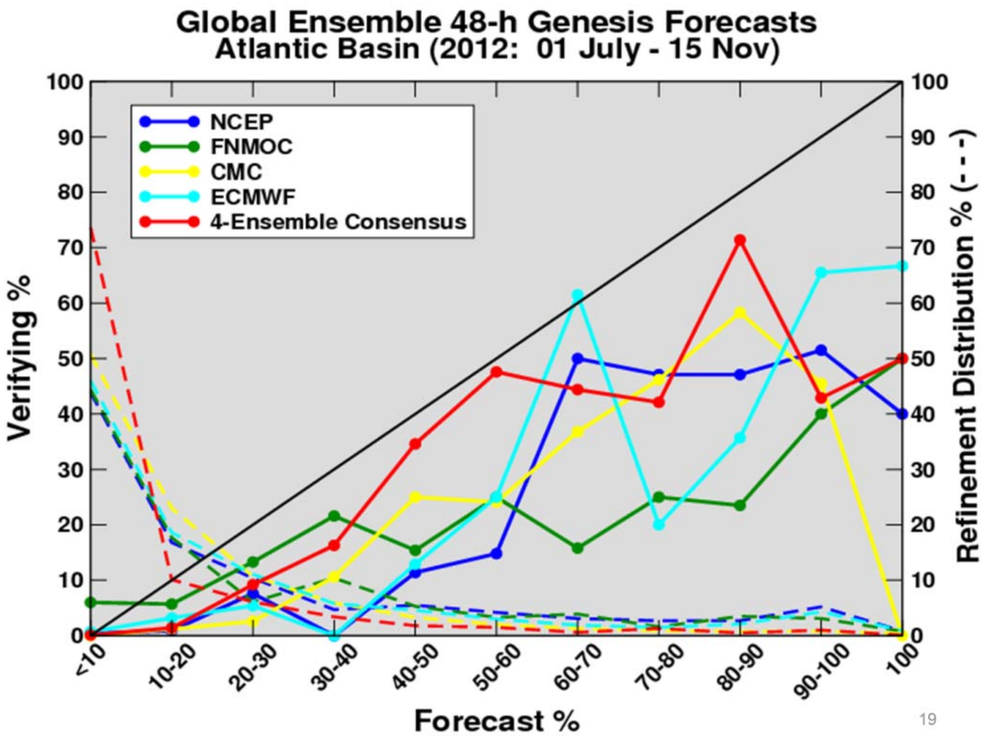


Figure 14. Genesis statistics from 4 operational global models: NCEP GFS model, FNMOC NOGAPS model, Canadian Meteorological Centre (CMC) model, and the ECMWF model ensembles. The 4 member ensemble consensus statistics are shown in red. Solid lines show the percent verifying, the dashed lines are called the refinement distribution and each indicates the percentage of times the forecast probability falls in a particular category. So for example for the ensemble mean (red line) forecasts of 10-20% probability of formation occur 10% of the time.

All four models shown in Figure 14 over predict genesis and there is a slight tendency to over predict more when the number of members in the ensemble showing genesis is larger. In general the Fleet Numerical Meteorology and Oceanography Center (FNMOC) model tends to over predict more than the other three but all vary widely with increasing forecast probability. The four-ensemble consensus shows the best skill in forecasting genesis at least up to a forecast probability of 60%. The dashed lines (the refinement distribution) indicate the percentage of times the predicted genesis probability fell within the various categories. Thus 75% of the time the four-ensemble consensus predicted less than a 10% probability of genesis. This work was done by GFDL.

Figure 15 is a similar figure for the NHC operational genesis forecast. That forecast is rather good for the lower probabilities of genesis, greatly under predicts around a 50% probability of genesis and slightly over predicts above 70%.

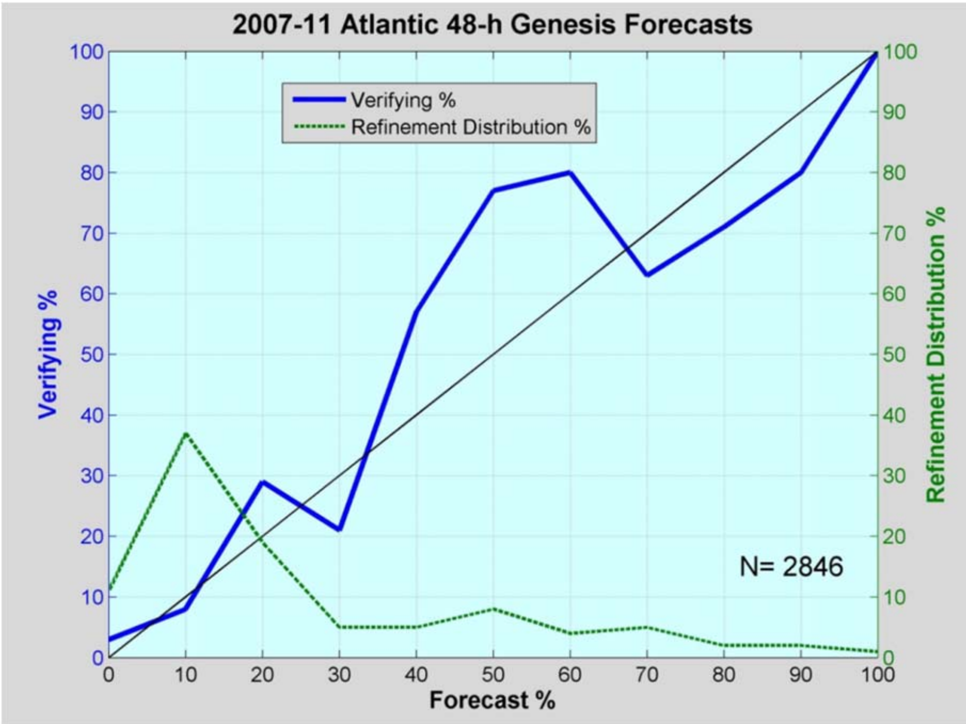


Figure 15. Genesis statistics for the NHC genesis prediction, otherwise same as Figure 13 Data for years 2007-2011 are shown.

10. Storm Surge

One of the goals of HFIP is to increase the skill of the forecast of the storm surge as a result of land falling hurricanes. HFIP provided the Advanced Circulation model (ADCIRC) storm surge development group within NOAA real-time access to the HFIP computers for forecasts during land-falling hurricanes in the 2012 season. The number of cases available in 2012 was quite low since there were few land-falling storms, however real-time forecasts were made for the major land-falling storms—Isaac and Sandy.

Figure 16 shows the HFIP storm surge forecast for Hurricane Sandy made 9:00 am Eastern Daylight time on Saturday October 27, about 36 hours before landfall. The forecast surge heights are shown along the right side of the figure. Figure 16 indicates a forecast storm surge of about 10 feet in New York Harbor and near Atlantic City, NJ. These forecasts were very close to the observed surge in these areas.

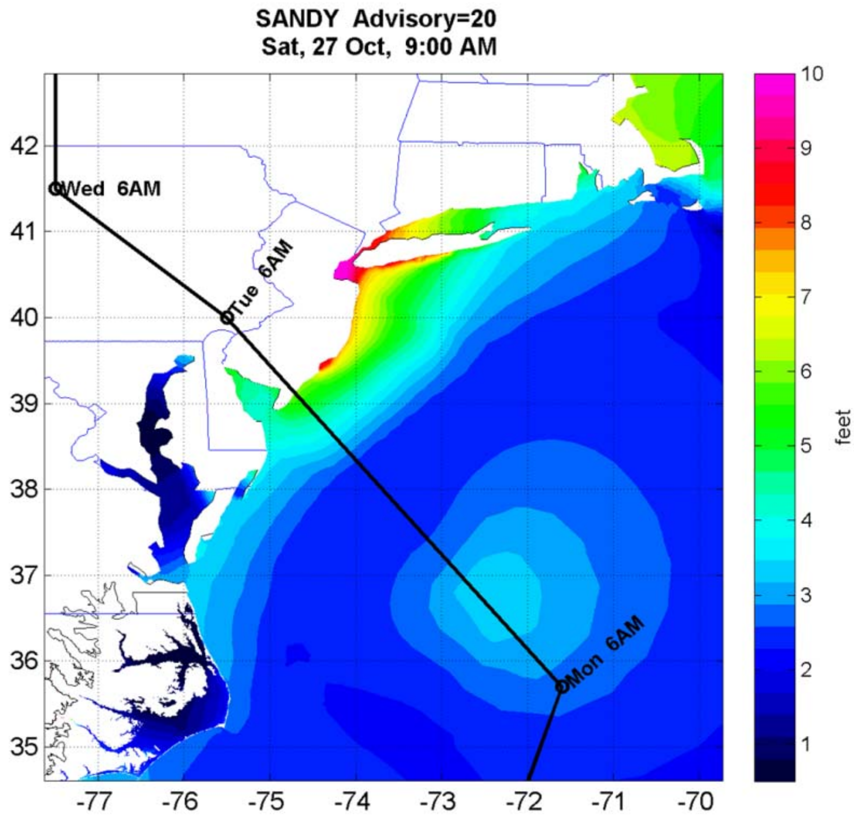


Figure 16. Maximum storm surge forecast for Hurricane Sandy from the ADCIRC storm surge model made 36 hours before landfall. Sandy's track is shown in black

11. Societal Impact Work

Because of severe budget cuts in 2012, HFIP was not able to provide direct funding to its Societal Impacts Team. Regardless, the team continued a modest level of work using funding from other sources and focused on evaluating hurricane storm surge products.

Figure 17 shows a number of possible choices for storm surge products produced from a storm surge forecast. This is for a hypothetical storm striking the Fort Meyers-Cape Coral, Florida, area. The upper left shows the area expected to experience some level of flooding from the hypothetical storm. No depth information is given. The upper panel on the right is a similar figure but now shows the depth of the storm surge. The lower figure is similar to the upper right panel but displays the information as a "surge hazard". In that panel blue indicates that the water will rise to less than 3 feet above ground level (white areas of course experiencing no flooding) while in the red areas, the water level would be expected to exceed 9 feet above ground level.

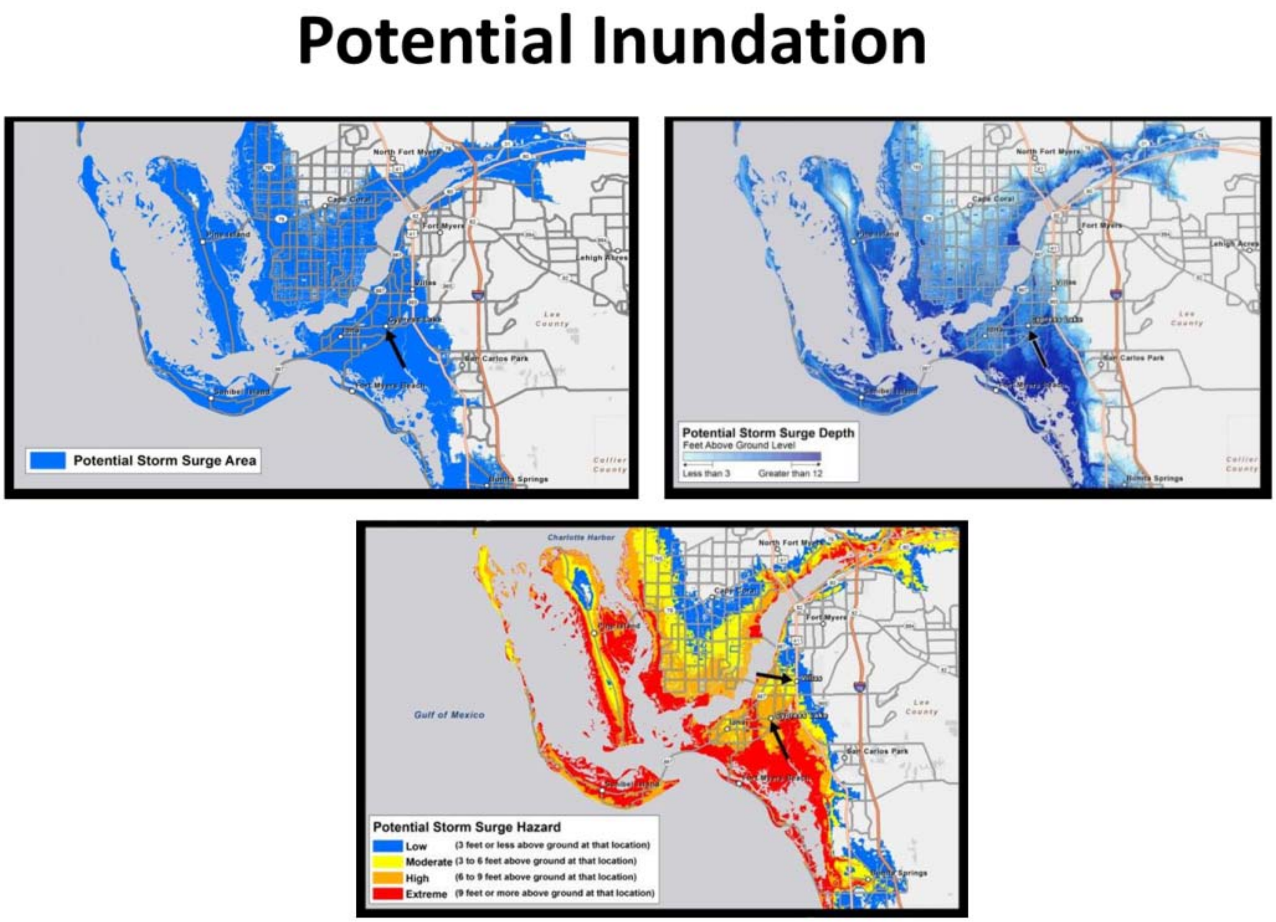


Figure 17. Examples of possible storm surge products for a hypothetical hurricane hitting the Fort Meyers-Cape Coral, FL area. See text for further explanation

The Impacts Team evaluated the lower panel by surveying Emergency Managers, the various Media outlets, the general public and Warning Coordinator Meteorologists (WCMs) for effectiveness of the presentation. The percentage of those responding to the survey and giving an answer of Excellent, Very Good or Good are indicated below;

Easy to Understand:

- 86% by Emergency managers
- 96% by Media
- 77% by Public
- 90% by WCMs

Provides Useful Information:

- 84% by Emergency managers
- 94% by Media
- 98% by public
- 83% by WCMs

HFIP expects to fund the Societal Impacts in the future at former levels. Results such as presented above will be expanded.

12. Statistical Post Processing of Model Output

Much of the discussion above focused on using numerical model improvements to achieve the HFIP goals. Typically statistical models (for example DSHP and LGEM) perform among the best as predictors of hurricane intensity. A statistical model is one where a limited (measured in single to double digits) number of predictors 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. Perhaps the simplest statistical model for intensity is Decay-SHIFOR5 where, as described earlier, the predictors are current position and intensity, position and intensity 12 hours earlier and current date (CLIPER5 is a similar model for track). It is sobering that even a model this simple provides forecasts of intensity almost as good as any of the current dynamical models. More complex statistical models used operationally for intensity are SHIPS and LGEM (NHC 2009).

There is another class of statistical model that takes a particular prediction from a dynamical model (say 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. SPC3, results shown in Figure 5, in 2011 showed improvement compared to the operational statistical and dynamical models, by using multiple operational numerical models as input for the environmental predictors. This year, 2012, however, SPC3 did not perform as well as in 2011. SPC3 uses output from the operational GFS, HWRF and GFDL models for input to both DSHP and LGEM. This gives six variations which are then averaged as a weighted ensemble, where more weight is given to DSHP in the early forecasts and LGEM in the later forecasts.

Figure 18. Intensity errors for all components of the FSU Multi-Model Ensemble for various forecast lead times. The acronyms are shown on the right (see Appendix A for details) and the number of cases is shown across the bottom. The ensemble mean (EM) of the models shown is in orange and the weighted ensemble mean is shown by the red bar. The official forecast is dark purple.

This year the best performer of the statistical models was the FSU Multi-Model Ensemble for track. For intensity only ICON, the NHC intensity consensus (which is an equally-weighted multi-model ensemble) did better (see Figure 5). Figure 18 shows the intensity error for the various models that went into the FSU Multi-Model Ensemble. The orange and red bars on the right side of the groups for each forecast lead time are the equally weighted ensemble mean and the weighted ensemble mean (the FSU model forecast) of the other models shown on the figure. At all lead times the weighted mean was better than the equally weighted mean and at all lead times it was better than all the other models. At some lead times the official forecast was better.

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

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

While it appears that use of aircraft data will likely help HFIP meet its intensity goals for storms for which such data is available, these data 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. A longer-term major focus for HFIP is to improve satellite data assimilation in regional model initialization systems.

We have not addressed the goal of HFIP to improve the forecasting of rapid changes in tropical cyclone intensity because, at this juncture, none of the HFIP dynamical models are capable of providing reliable forecasts of RI. The global models are not able to resolve the inner core processes that are likely to be very important in the RI process and all the regional models have serious spin up (and spin down) problems initially. This was noted in earlier Annual Reports and is not likely to be resolved until cycled DA with a Hybrid DA system is available for the regional models (currently projected to be 2014). 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 end of the initial five years of HFIP:

The longer range predictions, out to one week, of both track and intensity will be accomplished by global models run as an ensembles and initialized with a hybrid data assimilation system and post-processed with various statistical models. Resolution of these global models needs to be no coarser than about 20 km and the results will be improved if more than one global model is used in the ensemble.

The intensity goals for forecast periods out to 48-72 hours will be accomplished with regional models run with resolution at least as fine as 3 km 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 the RI goals may be met with the regional models. More specifically, the end system might include a global model ensemble with hybrid data assimilation, a regional model ensemble with hybrid data assimilation and statistical post processing (Table 8). The ability to run this system, however, will require at least a ten-fold increase in computer resources in operations in order to run the high-resolution ensembles.

Table 8. Numerical Model Hurricane Forecast Guidance System

Component	Specifications
Global model ensemble with Hybrid Data Assimilation	3) 20 members at 10-20 km 4) Multi Model (at least two – e.g. FIM, GSF)
Regional model ensemble	4) 20 members at 3 km 5) Multi model (at least two – e.g. HWRF, AHW, TC-COAMPS) 6) Using all available aircraft and satellite data in core and near environment of hurricane
Statistical Post Processing	2) LGEM, SHIPS, others

HFIP is also currently experimenting with a class of models known as basin scale models. These are simply the regional scale models with a large outer domain (so they cover a couple of basins like Atlantic/ East Pacific) and are capable of running high resolution nests over more than one storm at the same time. These systems have been run successfully during the 2012 hurricane season but still require some work to compete with the models containing a single storm. In both the single storm models and the basin scale models with multiple storms, the various nests interact with the large outer domain. Once we can switch to the basin scale models and all the hurricanes present can be run at once, at least the cost of rerunning the outer model for each storm can be eliminated.

This can then be easily taken one step further; the basin size can be expanded to include the entire globe. Then the regional models shown in Table 8 would be eliminated and only a high resolution global ensemble with multiple nests for each hurricane would appear in the table. Of course the regional models are still there as internal nests in the global model but separate ensembles one for the global model and one for the regional models would coalesce into the global ensemble—that would save considerable computer time.

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

14. Visiting Scientist Program

For the past four hurricane seasons, NHC has hosted a Visiting Scientist Program funded by HFIP. The goals of the annual program are to provide researchers and outside forecasters a better understanding of the NHC hurricane forecasting process, including the tools and techniques utilized by the Hurricane Specialists; and to facilitate additional dialog between NHC and the research/outside forecast community.

Participants in past years have included government researchers, university professors and graduate students, National Weather Service and foreign national forecasters, private sector meteorologists, social scientists, and broadcast meteorologists. Participants during the 2012 hurricane season were the following:

Dave Nolan - University of Miami - genesis/structure -Jul 31-Aug 3
Hayden Frank - WFO Boston - forecaster - Aug 7-10
Clark Evans - UW-Milwaukee - genesis/extratropical transition -Aug 14-17
Vijay Tallapragada - EMC - modeling - Aug 20-23
Orlando Bermudez - WFO Austin - forecaster -Aug 28-31
Scott Prosise - OPC - forecaster - Sep 10-13
Josh Cossuth - FSU - genesis - Sep 19-22
Brad Klotz - HRD/CIMAS - stepped frequency microwave radiometer - Sep 24-27
Andre van der Westhuysen - EMC/MMA Branch - wave/surge modeling - Oct 1-4
Ryan Torn - SUNY-Albany - data assimilation/modeling - Oct 8-11
Yoshihiro Konno - Weathernews - forecaster - Oct 22-25

All participants shadowed the forecasters at NHC during actual tropical storm and hurricane events, learning the analysis and prediction methodologies, technologies employed, observations and models utilized, time constraints, and communications issued. Each visiting scientist spent up to three days with NHC's Hurricane Specialists and one day with the marine forecasters in NHC’s Tropical Analysis and Forecast Branch. The dialog with participants during and subsequent to their visits can lead to improvements in our analysis and prediction methodologies.

15. References

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

Benjamin, S., G. Grell, J. Brown, T. Smirnova, and R. Bleck, 2004: Mesoscale Weather Prediction with the RUC Hybrid Isentropic-Terrain-Following Coordinate Model. *Monthly Weather Review*, 132(2):473–494, (doi:10.1175/1520-0493(2004)132<0473: MWPWTR>2.0.CO;2)

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

Chen, S., T. J. Campbell, H. Jin, S. Gaberšek, R. M. Hodur, and P. Martin, 2010: Effect of Two-way Air–sea Coupling in High and Low Wind Speed Regimes. *Monthly Weather Review*, 138(9):3579–3602, (doi:10.1175/2009MWR3119.1)

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

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

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

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

Doyle, J.D., Y. Jin, R. Hodur, S. Chen. H. Jin, J. Moskaitis, A. Reinecke, P. Black, J. Cummings, E. Hendricks, T. Holt, C. Liou, M. Peng, C. Reynolds, K. Sashegyi, J. Schmidt, and S. Wang, 2012: Real Ttime Tropical Cyclone Prediction Using COAMPS-TC. *Advances in Geosciences*, 28, Eds. Chun-Chieh Wu and Jianping Gan, World Scientific Publishing Company, Singapore, 15-28.

DTC, cited 2012a: WRF for Hurricanes Documents and Publications. [Available online at http://www.dtcenter.org/HurrWRF/users/docs/index.php.]

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

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

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

Hogan, T.F., and T.E. Rosmond, 1991: The description of the Navy Operational Global Atmospheric Prediction System's Spectral Forecast Model. *Monthly Weather Review*, 119(8):1786-1815,(doi: 10.1175/1520-0493(1991)119<1786:TDOTNO>2.0.CO;2).

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

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

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

NOAA ESRL, cited 2011: FIM Documentation. [Available online at http://fim.noaa.gov/fimdocu_rb.pdf.]

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

---------, cited 2008: Proposed Framework for Addressing the National Hurricane Research and Forecast Improvement Initiatives [Available online at http://www.nrc.noaa.gov/plans_docs/HFIP_Plan_073108.pdf]

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

NSF, cited 2007: Hurricane Warning: The Critical Need for a National Hurricane Research Initiative. [Available online at http://www.nsf.gov/nsb/committees/archive/hurricane/initiative.pdf.]

OFCM, cited 2007: Interagency Strategic Research Plan for Tropical Cyclones - The Way Ahead, FCM-P36-2007 [Available online at http://www.ofcm.gov/p36-isrtc/fcm-p36.htm]

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

Peng, M. S., J. A. Ridout, and T. F. Hogan, 2004: Recent Modifications of the Emanuel Convective Scheme in the Naval Operational Global Atmospheric Prediction System. *Monthly Weather Review,* 132(5):1254-1268, (doi: 10.1175/1520-0493(2004)132<1254: RMOTEC>2.0.CO;2).

PSU, cited 2011: PSU WRF/EnKF Real-time Atlantic Hurricane Forecast. [Available online at ocean http://hfip.psu.edu/realtime/AL2011/forecast_track.html.]

Snyder, C., and F. Zhang, 2003: Assimilation of Simulated Doppler Radar Observations with an Ensemble Kalman Filter. *Monthly Weather Review*, 131(8):1663-1677, (doi: 10.1175//2555.1).

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

Weng, Y. and F. Zhang, 2012: Assimilating Airborne Doppler Radar Observations with an Ensemble Kalman Filter for Convection-Permitting Hurricane Initialization and Prediction: Katrina (2005). *Monthly Weather Review*, 140(3):841-859, (doi: 10.1175/2011MWR3602.1)

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

Zhang, F., Y. Weng, J. Gamache and F. Marks, 2011: Performance of Convection-permitting Hurricane Initialization and Prediction During 2008-2010 with Ensemble Data Assimilation of Inner-core Airborne Doppler Radar Observations. *Geophysics Research Letters*, 38(15), (doi:10.1029/2011GL048469)

16. Appendix A: HFIP Funding Opportunity for Academia and Private Industry

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

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

HFIP anticipates releasing a second round of AO in June 2013 with an increased focus on advanced physics for land-air-sea interactions, hurricane intensity predictability, global-to-local scale modeling techniques and continuous improvements to operational hurricane models and data assimilation techniques.

17. Appendix B: Model Acronyms

The following is a list of acronyms used to identify models in this document. Many of the acronyms follow the four-character naming convention in the Automated Tropical Cyclone Forecasting (ATCF) system. For example, 6-hour "earlier" forecasts from "late" models (see Section 4a) are adjusted so that the previous 6-hour forecast matches the conditions at the beginning of the current forecast. This is simply known as an interpolated forecast. Forecasts of those future conditions are denoted with an "I" at the end, for "interpolated" (12-hour interpolations are denoted with a "2").

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

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: AWR with 6-hour interpolation

ARW: Pennsylvania State University Advanced Research WRF

CMC: Canadian Meteorological Centre model.

CMCI: CMC with 6-hour interpolation.

CLIPER5: Climate and Persistence model.

COAMPS-TC: Fleet Numerical Meteorology and Oceanography Center Coupled Ocean/Atmosphere Mesoscale Prediction System-Tropical Cyclone.

COTI: COAMPS-TC with 6-hour interpolation.

Decay-SHIFOR5: Decay Statistical Hurricane Intensity Forecast model.

DSHP: Decay SHIPS.

ECMWF: European Centre for Medium-range Weather Forecasts model.

EGRI: United Kingdom Meteorological Office model, subjective tracker, with 6-hour interpolation.

EMXI:	ECMWF with 6-hour interpolation.
FIM:	Flow-following finite-volume Icosahedral Model.
FM9I:	FIM with 6-hour interpolation
FSSE	Florida State University Super Ensemble
G01I:	GFDL ensemble member 01 with 6-hour interpolation (in general, G##I denotes GFDL ensemble member ## with 6-hour interpolation).
GEFS:	National Centers for Environmental Prediction Global Ensemble Forecast System.
GFDI:	GFDL with 6-hour interpolation.
GFDL:	Geophysical Fluid Dynamics Laboratory model.
GFNI:	Navy version of GFDL with 6-hour interpolation.
GFS:	Global Forecast System.
GFSI:	GFS with 6-hour interpolation.
GHMI:	GFDL adjusted using a variable intensity offset correction that is a function of forecast time, with 6-hour interpolation.
GPMN:	GFDL ensemble mean
GPMI:	GFDL ensemble mean (note all members of the ensemble include 6-hour interpolation).
HWFI:	HWRF with 6-hour interpolation.
HWRF:	Hurricane WRF.
ICON:	National Hurricane Center Intensity Consensus.
IV15:	Intensity forecast ensemble including 2012 stream 1.5 forecasts.
LGEM:	Logistics Growth Equation Model.
NAVGEM:	Fleet Numerical Meteorology and Oceanography Center Navy Global Environmental Model (replaced NOGAPS February, 2013).
NGPI:	NOGAPS with 6-hour interpolation.
NGXI:	Experimental NOGAPS with 6-hour interpolation.

NOGAPS: Fleet Numerical Meteorology and Oceanography Center Navy Operational Global Atmospheric Prediction System (replaced by NAVGEM February, 2013).

NMM: Environmental Modeling Center Nonhydrostatic Mesoscale Model.

SHIPS: Statistical Hurricane Intensity Prediction System.

SPC3: Six member weighted SPICE ensemble using output from GFS, HWRF, and GFDL as input for DSHP and LGEM. The ensemble weights vary with forecast lead time.

SPICE: Statistical Prediction of Intensity from a Consensus Ensemble.

TV15: Track forecast ensemble including 2012 stream 1.5 forecasts.

TVCA: Track Variable Consensus of at least two of AVNI, EGRI, EMXI, NGPI, GHMI, HWFI forecasts

TVCN: National Hurricane Center Track Variable Consensus

UKMI: United Kingdom Meteorological Office model, automated tracker, with 6-hour interpolation.

UWNI: UW-NMS with 6-hour interpolation.

UW-NMS: University of Wisconsin Nonhydrostatic Modeling System

WRF: Weather Research and Forecasting model. It is a regional system with options for the dynamic core, physics, initialization, post processing and verification. Variations include the Hurricane WRF (HWRF), PSU Advanced Research WRF (ARW), and NCAR Advanced Hurricane WRF (AHW)