2009 HFIP R & D Activities Summary: Accomplishments, Lessons Learned, and Challenges

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

27 April 2010

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

TABLE OF CONTENTS

2009 Hurricane Forecast Improvement Program (HFIP) R & D Activities Summary: Accomplishments, Lessons Learned, and Challenges

Executive Summary

This report describes the activities and results of the Hurricane Forecast Improvement Program in 2009. It is organized around three themes, *Encouraging results* from 2009 HFIP testing, evaluation and development activities, *lessons learned* from negative results and the *challenges* faced by the program to achieve its goals. The main topics from each of these three categories are:

Encouraging Results

- *Advanced data assimilation systems (EnKF, 4DVAR) appear to improve global forecasts over the current operational data assimilation system.*
- *High resolution global and regional ensemble systems are showing promise but require further testing and evaluation.*
- *High resolution global ensembles (30 km, 20 members) can be run in real time on available computing resources. Higher resolution is definitely possible.*
- *There is some preliminary evidence that airborne radar data used to initialize the hurricane vortex in regional models can improve forecasts of track and intensity.*
- *The Multi-Model regional ensemble showed promise.*
- *Various physics parameterization schemes were examined using HWRFX and COAMPS. Great sensitivities of storm structure to variations in the model physics were noted, strongly suggesting possible routes for improving the physics package in the operational HWRF model.*
- *Assimilation of pseudo sea-level pressure observations in GDAS improved track forecasts out to day 5.*
- *The FY10 NCEP GFS upgrade including increased horizontal resolution (35 km to 27 km) and upgrades to shallow and deep convection physics and the PBL has a significant positive impact on track and intensity forecasts.*
- *The FY10 NCEP operational HWRF configuration including coupling to HyCOM using more realistic surface flux exchange coefficients has a significant positive impact on track and intensity forecasts.*
- *Most components of the HWRF model are now publicly available and supported, and HWRF can be configured from community code repositories.*

Lessons Learned

- *Initialization of Regional models is a major problem.*

- *Both the regional and global models greatly over-predict tropical cyclo-genesis.*

- *Simply increasing the resolution of the regional models alone does not lead to improvements in model guidance.*

- *The research community must do a better job conveying the value and use of ensemble information to the forecast community and in developing value added products from ensembles.*

- *Model performance metrics for hurricanes must include more than just track and intensity metrics.*

Challenges

- *A vast majority of model forecasts will be initialized for storms for which no aircraft data are available. In those cases we need to better utilize available satellite data particularly for the regional models.*

- *For storms with aircraft data (particularly radar data) more use of that data is crucial for hurricane initialization especially for hurricanes near the coast.*

- *Development and tuning of physics packages for hurricane models at high resolution is critical.*

- *Advanced DA systems in both regional and global models appear to lead to substantial forecast improvement. We need to get these advanced assimilation systems into operation as soon as possible.*

- *For a limited sample, high resolution global and regional ensemble systems are showing promise but require further testing and evaluation.*

- *We need to develop better products to convey ensemble information to forecasters.*

- *We need to fully engage the whole hurricane community in improving the operational HWRF. A major step in this direction is the adoption of the HWRF community code by NCEP, so that research and operations use the same code base.*

- *We need to emphasize coordination between the HFIP modeling, observations and evaluation components to determine observational requirements for the improvement of model physics packages.*

In the body of the text we discuss each of the above lessons and challenges and our plans for addressing each.

2009 HFIP R & D Activities Summary: Accomplishments, Lessons Learned, and Challenges

1. Background on HFIP

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

The specific goals of the HFIP are to reduce the average errors of hurricane track and intensity forecasts by 20% within five years and 50% in ten years with a forecast period out to 7 days. The benefits of HFIP will significantly improve NOAA's forecast services through improved hurricane forecast science and technology. Forecasts of higher accuracy and greater reliability (i.e., user confidence) are expected to lead to improved public response, including savings of life and property.

NOAA recognizes that addressing the broad scope of the research and technology challenges associated with improving hurricane forecasts requires interaction with, and support of, the larger research and academic community. It is hypothesized that these very ambitious goals of the HFIP can only be met using high-resolution (~5-15 km) global atmospheric forecasting numerical models run as an ensemble in combination with regional models at even higher resolution (~1-5 km). Demonstrating this is very expensive computationally so HFIP requires access to resources currently only available at the few supercomputing centers in the country. Only by demonstrating the value of high resolution is there any opportunity to obtain such a computational resource for operational hurricane forecasts.

For FY09 the HFIP program consisted of about $27 M with $6 M dedicated to enhancing computer capacity available to the Program, including $4.5 M for a dedicated machine with 3500 processors located in Boulder, Colorado. The other $1.5 M was for development and testing and an additional 160 processors for the development machine at the National Centers for Environmental Prediction (NCEP) to facilitate transition of HFIP technology into National Weather Service (NWS) operations. About $10M of the $27M is part of the base funding for the Atlantic Ocean and Meteorology Laboratory (AOML) in Miami and the Environmental Modeling Center (EMC) at NCEP for hurricane model development. The remaining $11M was distributed to various NOAA laboratories and centers (Earth System Research Lab (ESRL), Geophysical Fluid Dynamics Laboratory (GFDL), National

Environment Satellite Data and Information Service (NESDIS), and National Hurricane Center (NHC). Funding was also provided to the National Center for Atmospheric Research (NCAR), Naval Research Laboratory in Monterey (NRL), and several universities: University of Wisconsin, The Pennsylvania State University, Colorado State University, University of Arizona and University of Rhode Island. Finally, $1M was contributed to the National Oceanographic Partnership Program (NOPP), Announcement of Opportunity for competed proposals related to improving understanding and prediction of hurricanes. The funding to NOPP from HFIP was matched by funding from the Office of Naval Research (ONR).

Distribution of the $11M was accomplished through recommendations from 11 teams focused on various components of the hurricane forecast problem. In late 2009 we combined some of the teams and assigned co-leads to each team (as opposed to a single lead initially) to increase leadership across the various HFIP organizations. Current teams are listed in Appendix A with the team leads shown in bold type. For reference a list of acronyms for the various organizations is shown in Appendix B. These teams are made up of over 50 members drawn from the hurricane research, development and operational community.

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

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

A major part of Stream 2 is a demonstration system, or Demo System, that is run in testing mode each hurricane season. The purpose of this system is to evaluate strengths and weaknesses of promising new technology. As a result of the Demo System testing, some components may be found to be of particular interest to the operational forecasters, and, if resources do not permit its implementation in the operational infrastructure, the Demo system for the following season will emphasize those components and will provide specific output that is made available to NHC forecasters for evaluation. We refer to this component of the Demo System as Stream 1.5. Table 1 outlines these various streams.

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

To facilitate the transition of research to operations, HFIP has recognized the importance of having research and operations share the same code base, and has co-sponsored the Developmental Testbed Center (DTC) to make available and support HWRF to the community. This support started in February 2010 with the DTC/EMC/MMM Joint Hurricane Workshop and WRF for Hurricanes Tutorial.

Table 1. The Two Stream Strategy

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

2. High-Resolution Ensemble Approach

A single "deterministic" forecast by a particular numerical model has an inherent but unknown level of uncertainty; any two model forecasts starting from infinitesimally different initial states will grow differently with time, the amount of difference depending upon the weather situation. If the forecast is reproduced many times, each time introducing small initial differences, the result is called an ensemble, and the different model forecasts can potentially provide information on the confidence one should place in a particular forecast. Frequently, but not always, the highest probability is that the correct forecast is near the mean, median or mode of the ensemble, though other ensemble realizations have a finite probability of being correct. Because the various forecasts diverge with time, emergency managers should be able to make more effective decisions when provided with ensemble guidance compared to being provided with a single forecast.

High resolution is hypothesized to be necessary in these ensembles in order to adequately resolve the hurricane structure, for the hurricane can alter the flow in which it is embedded and, in turn, this altered flow will impact the hurricane track and so also its intensity. To begin to get structures in the forecast model which resemble actual hurricanes, resolutions of 5-15 km are necessary. Ideally, each ensemble member will have this resolution, and ideally 20-30 members are computed to provide adequate estimates of the uncertainty.

Beyond about three days, forecast guidance must come from global ensembles since the planetary-scale patterns interact with and influence the steering of the storm. After about three days, it has been shown that the evolution of the atmospheric flow at a given location depends on atmospheric features distributed globally. Therefore, forecasts that extend out to 4-7 days require that the forecast models be global.

The potential value of high-resolution global ensembles has been demonstrated in part through forecasts from international competitors such as the European Centre for Medium-Range Weather Forecasts (ECMWF). However, there is still much to be learned about high-resolution global modeling. The best way for the U.S. to make progress is to run the ensembles over enough cases such that statistical significance of the computed skill of the forecasts can be determined. Generally this requires at least that the high-resolution ensemble be run over the most active few months of the hurricane season and every forecast period from genesis to decay (with 2 to 3 years of cases being even better at capturing the full range of tropical cyclone characteristics associated with inter-annual changes in environment, e.g., associated with El Nino events). This is an enormous computing challenge, but it needs to be performed to demonstrate the value of the high-resolution forecast guidance over the guidance that is operationally available today.

Much the same can be said for regional ensembles, but here the emphasis shifts from track forecasts at longer forecast leads, to intensity forecasts at medium forecast leads. Much of the control of the intensity of the storm is thought to reside in the dynamics of the inner core region of the hurricane. If this is true, then the inner core must be resolved to account for these dynamics requiring a resolution of at least 3-5 km. We will show some preliminary results from HFIP activities that demonstrate the value of high resolution ensembles in the intensity forecast problem. Ideally, regional high-resolution ensembles are nested within high-resolution global ensembles, which provide an ensemble of lateral boundary conditions that describe the influence of global flow patterns. Also, it is preferable to use similar advanced data assimilation approaches and model physics.

3. The HFIP Model Systems

3.1. The Global Models:

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

GFS—Refers to the Global Forecast System. There are two versions of this model currently running in the demonstration system. This includes a version of the current operational model run at the NOAA National Centers for Environmental Prediction (NCEP) and an experimental version of that model.

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

3.2. The Regional Models:

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

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

The Multi-Model Ensemble—The multi-model ensemble was organized by Florida State University and was made up of a total of 7 models run by different organizations. The various models and their resolution are indicated in Table 5. Two of the members are the operational models, GFDL at 7.5 km and HWRF at 9 km. GFDL is the old operational model that is still being run in parallel with the current operational model HWRF. HWRF is constructed from the NMM core of the WRF, and both GFDL and HWRF models are coupled to the Princeton Ocean Model (POM) in the Atlantic Basin. A version of the HWRF was also run at 4 km resolution (HWR4) during the 2009 hurricane season. HWRF-x is an experimental version of the operational HWRF run by the Hurricane Research Division of OAR. It did not have an interactive ocean model associated with it but did have an interactive nest within it. HyHWRF is the HWRF model coupled to the Hybrid Coupled Ocean Model (HyCOM), and it was run in parallel for the 2009 hurricane season for the Atlantic Basin.

In addition to HWRF, two versions of the WRF ARW system were also run. The ARW system run by NCAR used a simplified one dimensional model of the ocean. It used two interactive nests within the outer regional model. FSU also ran a version of the ARW without an interactive ocean.

COAMPS-TC is a Navy model run by NRL Monterey. It is a version of their COAMPS regional prediction system that is being run operationally and has an interactive ocean.

The Penn State Regional Ensemble

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

3.3. Initialization systems:

A number of approaches were used to create the initial state for the global and regional models in the experiments we describe here and below. The choices include:

- The initial state created for the current operational model (Global Forecast System or GFS) interpolated to the higher resolution grid. The GFS uses the Grid point Statistical Interpolation (GSI) initialization system that has run operationally for many years; it is a three-dimensional variational approach (3D-VAR).
- NRL Atmospheric Variational Data Assimilation System (NAVDAS). This is the system used to provide the initial conditions to the Navy global model. It has been a 3D-VAR system but starting late September 2009, it was upgraded to NAVDAS-AR (for accelerated representor), a four-dimensional variational approach (4D-VAR).
- Ensemble Kalman Filter (EnKF). This is also an advanced assimilation approach (somewhat like 4D-VAR) that uses an ensemble to create background error statistics for a Kalman Filter. While this approach is still in the experimental stage in the U.S. (though operational in Canada), it has shown considerable promise. The system in FY09 HFIP used the GFS model to provide the ensemble forecasts for the EnKF data assimilation, although the FIM may also be used in FY10.

- Hybrid Variational-Ensemble Data Assimilation System (HVEDAS). This system combines aspects of the EnKF and 3D- or 4D-Var for example, using the ensemble of forecasts to estimate the covariances at the start of a 4D-Var assimilation window. This technology is under development at NOAA/NCEP/EMC and NOAA/OAR/ESRL. It will not be ready for testing in the 2010 season but may be available for subsequent seasons. This hybrid approach is likely to define the operational global data assimilation system for NOAA in the 5-year time frame.

- The initial state for the regional models was generally produced by downscaling the global models' analysis and forecasts. In addition, the Penn State Regional Ensemble model, the WRF/ARW/NCAR model used an EnKF initialization system.

- The operational HWRF utilizes an advanced vortex initialization and assimilation cycle consisting of four major steps: 1) interpolate the global analysis fields from the Global Forecast System (GFS) onto the operational HWRF model grid; 2) remove the GFS vortex from the global analysis; 3) add the HWRF vortex modified from the previous cycle's 6-hour forecast (or use a synthetic bogus vortex for cold start); and 4) add satellite radiance and other observation data in the hurricane area (9 km inner domain). The major differences from the GFDL model initialization are steps 3) and 4).

4. The HFIP Baseline

HFIP 10-year goals for Atlantic guidance were set by the HFIP Executive Oversight Board (HEOB) as:

- Reduce average track errors by 50% for days 1 through 5
- Reduce average intensity errors by 50% for days 1 through 5
- Increase the probability of detection (POD) for rapid intensity change to 90% at Day 1 decreasing linearly to 60% at day 5, and decrease the false alarm ratio (FAR) for rapid intensity change to 10% for day 1 increasing linearly to 30% at day 5. The focus on rapid intensity change is the highest forecast challenge indentified by the NHC.
- Extend the lead time for hurricane forecasts out to Day 7 (with accuracy of Day 5 forecasts in 2003).

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

To measure progress toward meeting these goals, HFIP established a baseline against which results from experimental and operational HFIP model guidance will be measured. These HFIP Performance Goals Baselines were developed in a white paper authored by James Franklin dated 5 May 2009 and summarized here.

For both the track and intensity goals, a consensus (equally-weighted average) of operational guidance models was utilized, evaluated for the Atlantic Basin over the period 2006-2008. This 3-year average was determined to be feasible and adequate based upon the trend data presented above. That is, this shorter 3-year period is adequate to determine the HFIP Performance Goal Baselines because there has been a significant reduction in track error in recent years, and because increased tropical cyclone activity in the Atlantic Basin in the last few years allows for more stable statistics over this shorter time period.

For track error, this consensus was a particularly good choice because the mean skill of the official NHC forecast is very close to that of the consensus. Consequently, a 20% improvement in any HFIP guidance over this baseline could reasonably be expected to translate to a 20% improvement in the official forecast. This would not be the case if an individual operational model were used as a baseline.

The track baseline was a consensus of GFSI, GFDI, UKMI, NGPI, HWFI, GFNI, and EMXI, which were computed whenever at least one of the consensus members was present. This is essentially the membership of NHC's current operational consensus model TVCN. Even though HWFI was only available in 2007-8, the recommendation was to include 2006 as well. The additional year of data will provide a more representative assessment of the current state of the forecast guidance. Evaluation of this consensus over 2006-8 is given below (CONS). For comparison, climatology and persistence skill baseline errors are also shown (OCD5), as are the official forecast errors (OFCL). Forecast errors are in nautical miles.

Table 2. HFIP Track Performance Baseline (nautical miles)

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

For intensity, the consensus members were: GHMI, HWFI, DSHP, and LGEM (note that GHMI is the GFDL model), with the consensus computed whenever at least one of these models was present. This is the same set of models used in the operational intensity consensus ICON (except that ICON is not computed unless all the member models are present). Evaluation of this consensus over 2006-8 is given below, along with climatology/persistence and the official forecast. Forecast errors are in knots. The table shows that the intensity consensus is actually slightly better than the official forecast, at least beyond 24 h or so. In part, this is because this intensity consensus has only been operationally utilized for one year. NHC's operational practice of only making incremental changes to the official forecast, from forecast to forecast, may also contribute.

Table 3. Proposed HFIP Intensity Performance Baseline (knots)

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

When evaluating HFIP forecast products it will be important to consider differences in forecast difficulty between the cases used by HFIP and those of the 3-year baseline sample. In order to compute the "true" percentage improvement represented by the HFIP products, the OCD5 skill baseline will be computed for the HFIP sample, and the HFIP errors will be normalized so that the OCD5 errors for the two samples are equivalent.

HFIP also has a goal to introduce a 7-day track forecast. In this case, a baseline is not required, however a metric is needed to judge whether a 7-day forecast can be introduced. The metric decided upon is that this metric would be for HFIP to produce guidance at least as good as the 5-day official forecast was, on average, when it was introduced in 2003. A linear fit to the official 5-day errors (ignoring varying sample sizes from year to year), evaluated at 2003, yields 314 nm.

There are no official forecasts issued for rapid intensification (RI), unless it is inferred from the official intensity forecast values. However, the official forecast virtually never shows a 24 h increase in intensity of 30 kts or more, so the performance of the official forecast would not make a good baseline. Hence, it was decided that it was better to use the existing model guidance. Both deterministic and probabilistic frameworks were used. The GFDL model provided a baseline for deterministic HFIP models, using metrics such as POD/FAR and related measures. For consistency with the baselines discussed above, these guidance models were evaluated over the same period 2006-8.

In the future all results that refer to track or intensity errors will contain the baseline for reference.

5. Stream 1 FY09 Activities at NCEP/EMC

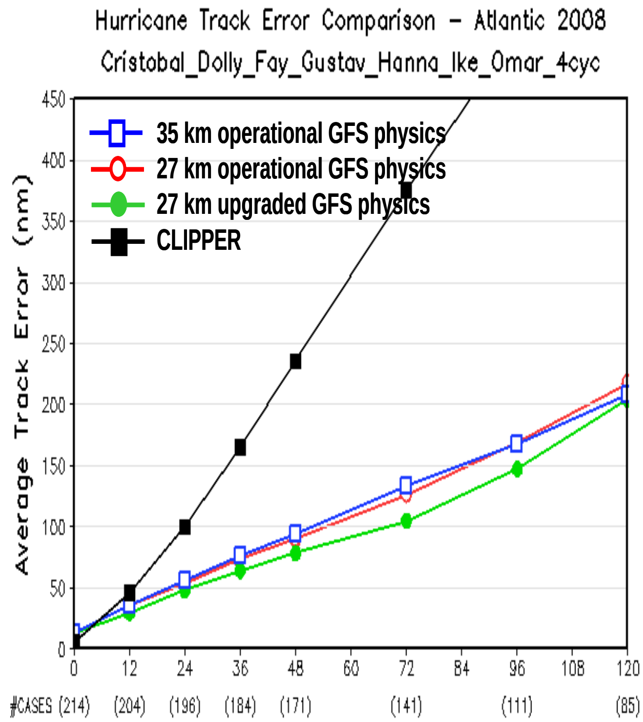
The NCEP operational modeling systems supported by or in part by HFIP include the GFS, GEFS, HWRF and RTOFS modeling systems. FY09 upgrades to these systems along with impacts on tropical system forecasts are provided below.

5.1. The Global Data Assimilation and Forecast System (GDAS/GFS)

These include some major changes in the GFS system from which the HWRF and GFDL models get their initial and boundary conditions.

There are two phases for GFS upgrades, one implemented in December 2009 and one to be implemented in June 2010. These changes are:

- GFS upgrades (Implemented **Dec 2009**)
    - Added tropical storm pseudo sea-level pressure observations
    - Added NOAA19 hirs/4,AMSU-A, & MHS brightness temperature observations
    - Added EUMETSAT-9 atmospheric motion vectors
- GFS upgrades (Planned for **May 2010**)
    - Resolution increase (27 km from 35 km)
    - Upgrade radiation to AER RRTM2
    - Revised Gravity Wave Drag and Mountain Blocking
    - Removal of negative water vapor with a positive-definite tracer transport scheme (enhances impact of satellite radiance data)
    - Higher resolution hurricane relocation
    - Major upgrades to shallow convection, PBL, deep convection with overshooting cloud tops (minimizes grid point storms)



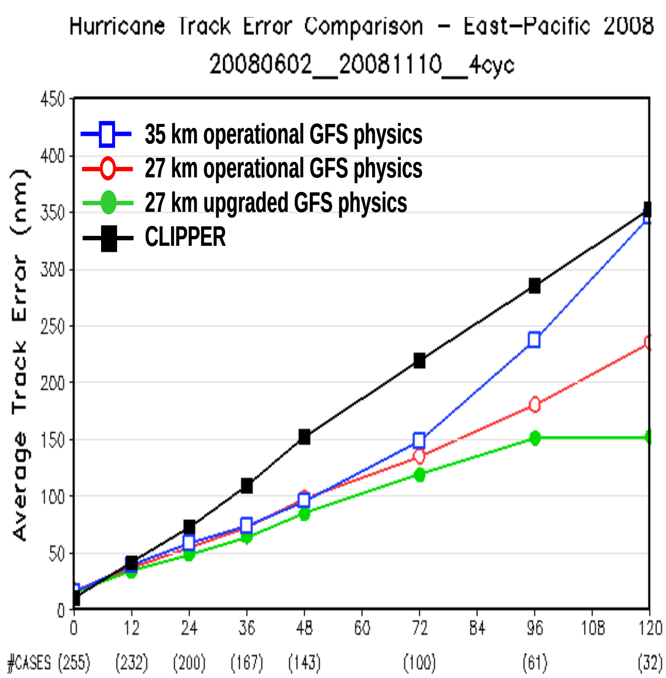


Figure 1. Improvements in track forecasts in the GFS model as a result of resolution and physics upgrades out to 5 days. The blue line represents the operational model before the listed changes. The red line shows the impact of increased resolution. The green line shows the impact of increased resolution and upgrades to the physics. The black line is the benchmark track model CLIPER. Numbers in parentheses at the bottom of the figures indicate sample size.

Although a major change slated for June 2010 is an increase in resolution to 27 km from 35 km, the improvement in GFS hurricane track error is primarily due to the physics changes in PBL and the deep and shallow convection. Figure 1 shows the track errors for the Atlantic and East Pacific for the resolution increase (red) and the inclusion of the improved physics (green). In both basins, the improved physics accounts for the majority of the forecast improvement.

5.2. The Global Ensemble Forecast System (GEFS)

There was a major implementation for GEFS (Global Ensemble Forecast System) on 23 February 2010. This upgrade mainly includes:

- Increasing horizontal resolution to 70 km from 90 km
- Using 8th order horizontal diffusion instead of 4th order, at all resolutions
- Adding a stochastic perturbation scheme to account for random model errors

This upgrade is significantly improving the skill of uncertainty forecasts out to 16 days. There is a one-day extended skill (60% anomaly correlation) from current operation for the Northern Hemisphere 500 hPa geopotential height (see Figure 2). For tropical storm prediction, there is about a 25% improvement for ensemble mean track errors during the 2009 hurricane season (see Figure 3 for selected 4 major storms during the 2009 season).

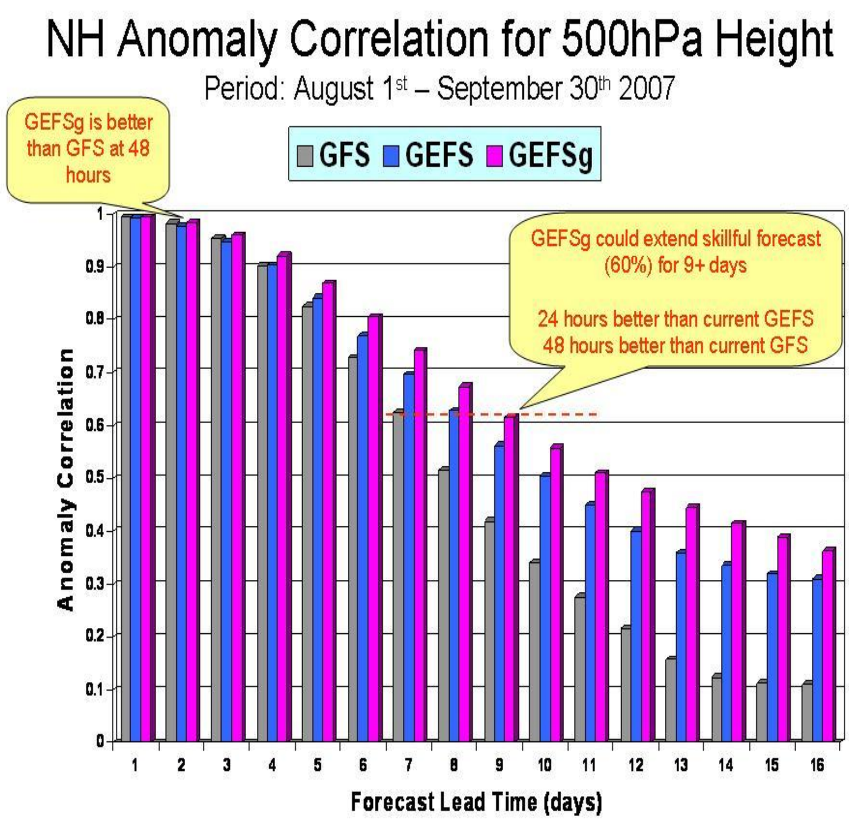


Figure 2. Northern Hemisphere 500hPa geopotential height anomaly correlation for the period of Figure 2. Northern Hemisphere Anomaly Correlation for 500 hPa Height, 08/01 – 09/30/2007. GFS (grey) is NCEP high resolution deterministic forecast at T382L64. GEFS (blue) is NCEP operational ensemble at T126L28. GEFS (pink) is NCEP off-line ensemble parallel (T190L64).

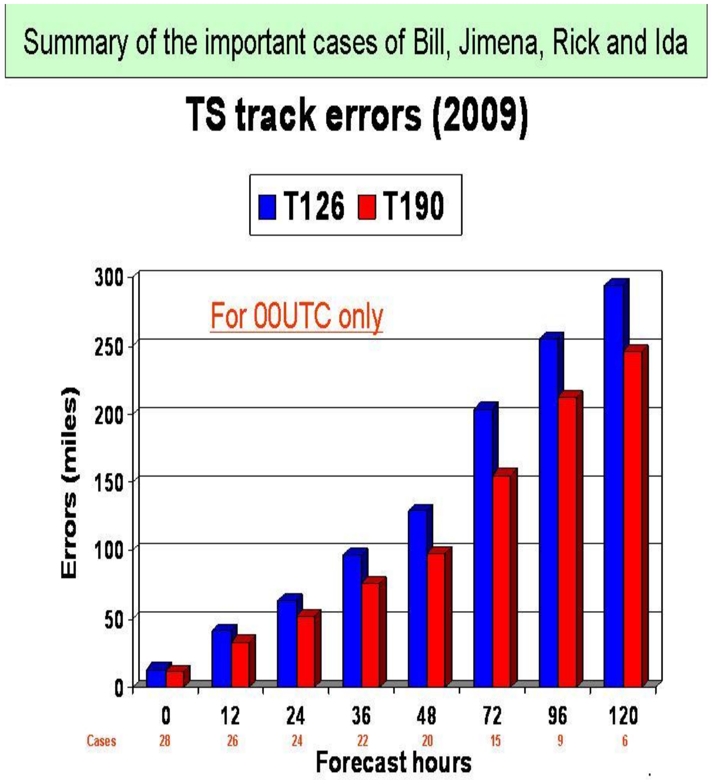


Figure 3. Ensemble mean track errors for important cases of Bill, Jimena, Rick and Ida during the 2009 hurricane season. Blue bars are for operation T126L28 GEFS forecast, and red bars are for retrospective experiments for planned new GEFS (T190L28).

5.3. Hurricane Weather and Research Forecast (HWRF) Model

At the time of this writing, NCEP is in the process of concluding HWRF FY10 pre-implementation testing to determine the final configuration for the 2010 hurricane season. The following HWRF upgrades have been systematically tested and evaluated by EMC and TPC since the fall of 2009 to support the December 2009 GFS implementation and the proposed GFS May 2010 upgrade:

- *Upgrade to surface exchange coefficients (closer to observations from CBLAST field experiment and dropsonde data)*
- *Modifications to the vortex initialization procedure through use of conventional and satellite based observations in the inner nest (9 km domain)*
- *Inclusion of Gravity Wave Drag parameterization*
- *Coupling with HyCOM ocean model in the Atlantic Basin*

The HWRF FY10 pre-implementation test plan consists of a series of systematic experiments requiring more than 600 runs of 2008-2009 Atlantic and East Pacific storms. The first step in defining the FY10 operational configuration was to define a new benchmark for the HWRF. The FY10 benchmark is identical to the operational HWRF used in 2008 and 2009 in every way except that there were several corrections made to rectify code errors in the treatment of downward shortwave radiation. Each of the proposed FY2010 upgrades listed above were tested in (and compared to) the FY10 benchmark and evaluated independently to ensure the upgrades had the expected impact on track and intensity forecasts. Then the individual upgrades were tested in combination using the operational 2008 GFS configuration, the December 2009 GDAS/GFS upgrade, and the planned June 2010 GDAS/GFS resolution increase and physics upgrade.

Results from the pre-implementation testing are presented in Figures 4 and 5 for both the Atlantic and East Pacific Basins, respectively. In the Atlantic Basin the combination of all 3 upgrades listed above results in a reduction in track error compared to the baseline throughout the 5-day forecast period. Analysis of the track error shows that Hurricane Bill (2009) track was not simulated well by the GFS May 2010 upgrade. Removal of Bill from the sample results in additional reduction of track error compared to the FY10 baseline (not shown). The proposed upgrades to both the HWRF and GFS had the highest percentage of superior track forecasts, the lowest average intensity error, and a significant reduction in the intensity bias for 72 hours and beyond. It is noted that the intensity bias is now negative from 12 to 48 hours. However, NCEP considers this to be a reasonable result for the 9 km horizontal resolution. It was found that the new enthalpy exchange coefficient obtained from the observational study of CBLAST, and the new drag coefficient obtained from the observational study of Powell et al (2003) had the largest positive impact on the reduction of intensity errors. In the East Pacific Basin the planned HWRF and GFS upgraded systems resulted in a significant (10% to 20%) reduction of track errors at all forecast times. The proposed upgrades to both the HWRF and GFS had the highest percentage of superior track forecasts, the lowest average intensity error, and a near-zero intensity bias for 72 hours and beyond.

At the time of this writing the final testing of the coupled HWRF-HYCOM system was not complete. Therefore, the results of those tests were not included in this report. However, in 2009 a parallel real-time run of HWRF coupled with HYCOM was conducted at NCEP along with a parallel HWRF-POM configuration that included corrections to the downward shortwave radiation and gravity wave drag parameterization for 7 tropical storms and hurricanes in the Atlantic Basin. Figure 6 shows that coupling with HYCOM resulted in a reduction of track and intensity errors for this limited sample compared to the parallel HWRF-POM system.

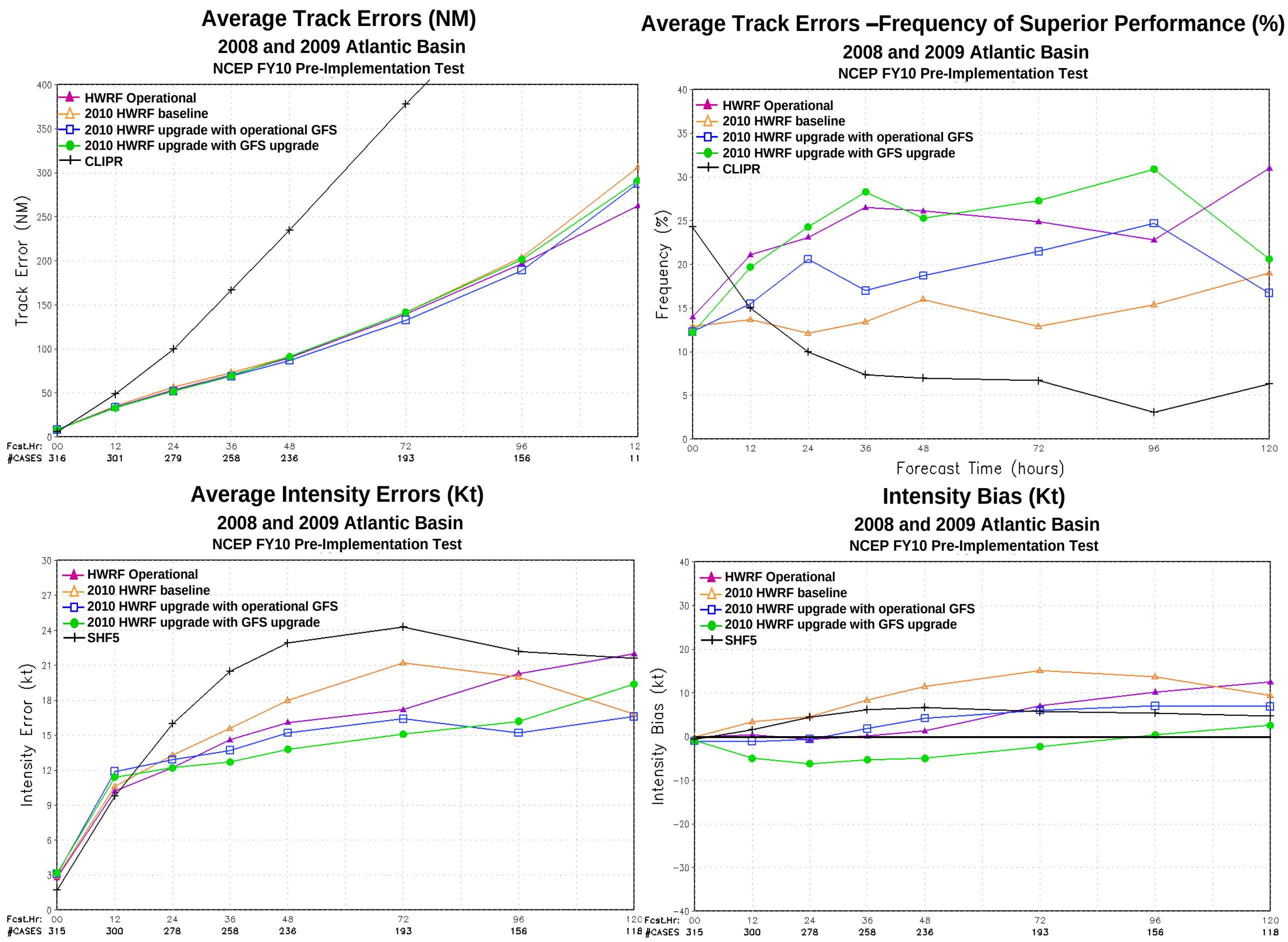


Figure 4. HWRF FY2010 pre-implementation test results for 2008-2009 Atlantic Basin. The 2010 baseline HWRF is identical to the operational HWRF and includes corrections to radiation calculations. The 2010 HWRF upgrade configuration builds off the 2010 HWRF baseline and includes assimilation of satellite data on the 9 km nest, Gravity Wave Drag parameterization and updated surface flux exchange coefficients.

5.4. Real Time Ocean Forecast system (RTOFS)

An initiative is also underway at NCEP to improve and maintain the Real-Time Ocean Forecasting System (RTOFS) in the Atlantic in support of the coupled HWRF-HYCOM system. The RTOFS has been significantly improved by removing spurious features in the Sargasso Sea. The model initialization will soon be upgraded by assimilating Jason-2 altimeter data, which will be implemented operationally in May 2010. Experiments have been conducted to show positive impact of synthetic TS profiles from altimeter data on the ocean model mixed layer and deep water characteristics, with modest impact on hurricane forecasting. Work on a global RTOFS model is also underway as EMC's Ocean Modeling Branch is adopting the 1/12 degree Navy HYCOM model for operational implementation to provide a real-time global backbone ocean modeling capability that can be used to provide boundary data for coupled hurricane models in the Pacific (or globally if so required). This is on track for parallel testing at NCEP Central Operations (NCO) in FY2010Q4.

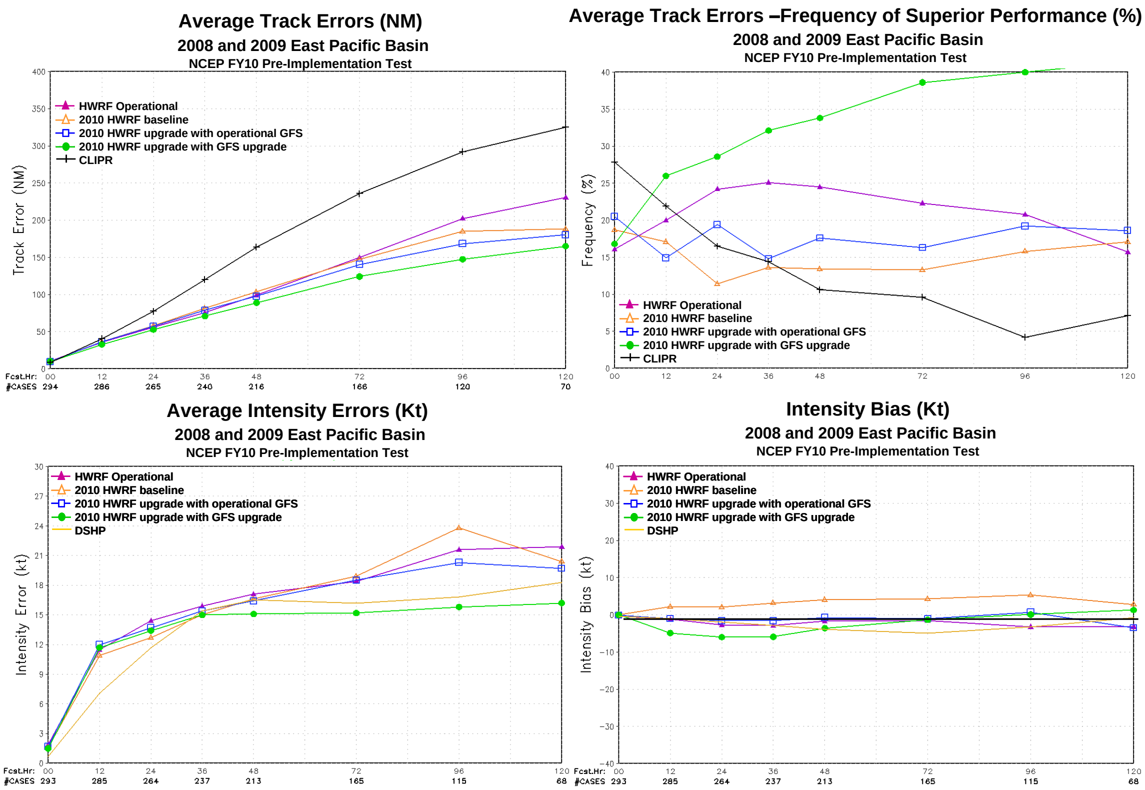


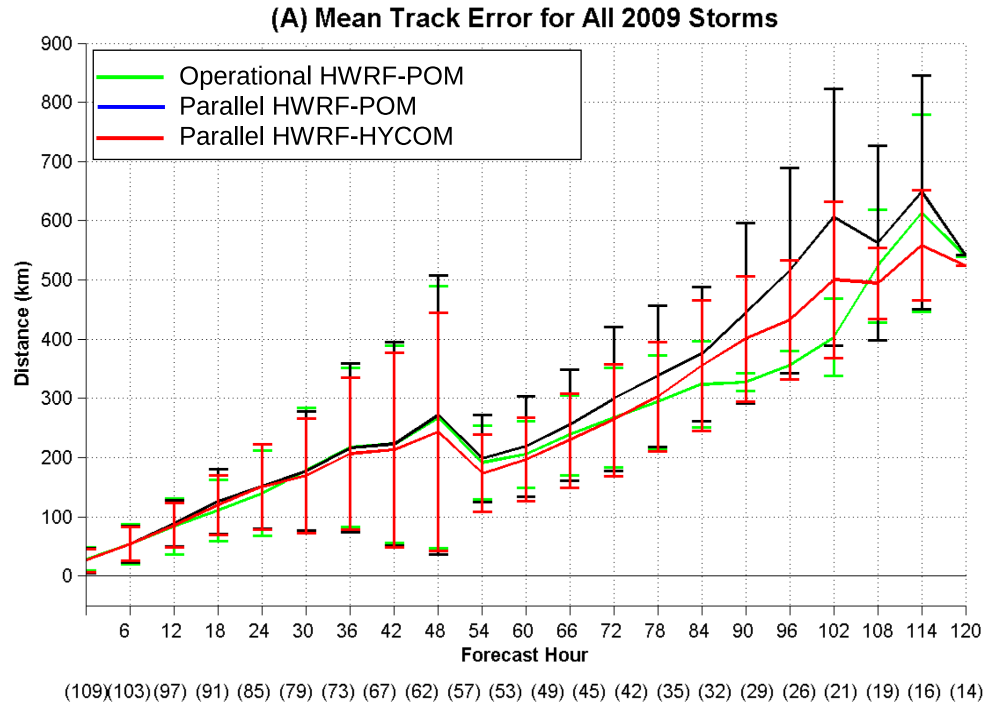
Figure 5. HWRF FY2010 pre-implementation test results for 2008-2009 East Pacific Basin. The 2010 baseline HWRF is identical to the operational HWRF and includes corrections to radiation calculations. The 2010 HWRF upgrade configuration builds off the 2010 HWRF baseline and includes assimilation of satellite data on the 9 km nest, Gravity Wave Drag parameterization and updated surface flux exchange coefficients.

For 2011, EMC will consider use of the NOAH Land Surface Model (LSM) with the HWRF forecast system. Several preliminary cases have been run which indicate a positive impact on track and rainfall prediction.

5.5. Wave Modeling (WAVEWATCH III)

The HWRF-HYCOM-WAVEWATCH III coupled model has been built at NCEP, and URI is fitting in the advanced coupling approach developed previously with the GFDL-POM-WAVEWATCH codes. This coupling includes using sea spray to improve and link various surface fluxes. OSIP 06-096 has directed NCEP to centrally maintain and develop capabilities for running a site-specific version of the WAVEWATCH III wave model with applications at arbitrary U.S. coastal areas. Typically such models would be run at WFOs or NCEP partners. This project is now funded by HFIP through NOS HFIP storm surge funding, and in this context has been expanded to become a coupled relocatable wave-surge capability. As part of this project NOS is implementing ADCIRC inundation capability for the U.S. West Coast using NCO computers. The project is scheduled to begin

1 April 2010 and will be accelerated with additional NCEP manpower in the summer of 2010.



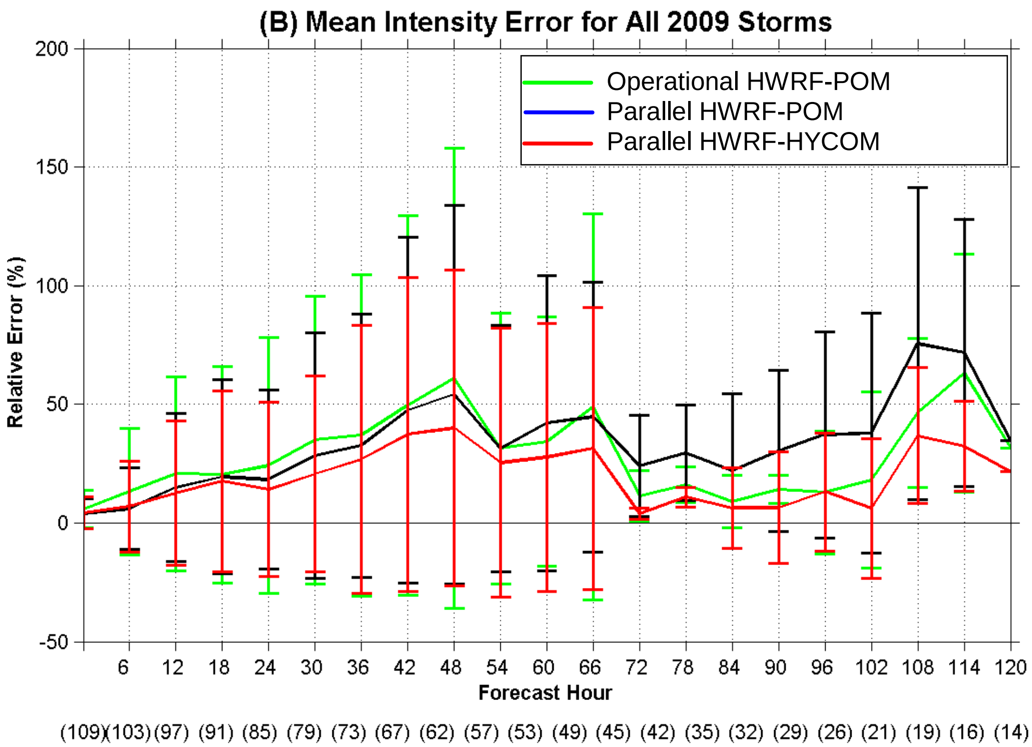


Figure 6. Coupled HWRF-HYCOM results for 2009 tropical cyclones: (A) Mean track error and standard deviation; (B) mean intensity error and standard deviation. Number of samples is listed on x-axis, along with forecast hour.

6. Stream 1 FY09 Activities at HRD

The Hurricane Research Division of NOAA conducts a yearly field program in hurricanes as a component of the HFIP program. Listed below are the components of that program conducted during the 2009 season.

- Participated in the Intensity Forecast Experiment (IFEX) 2009: WP-3D (17 missions, all 1 P-3/double crew) and G-IV (14 missions) with HRD crews flew ~200 h to gather data in Hurricanes Paloma (2008) and Bill, plus TS Ana & Danny
- Deployed 756 dropwindsondes and 238 AXBTs. 55 real-time TDR analyses transmitted to NHC and Doppler radial wind files transmitted to EMC. TDR super-observations transmitted in Paloma, Bill and Danny provided real-time assimilation of TDR data into ARW model. 60 real-time H*Wind surface analyses.

Figure 7a shows the wind field computed from the P3 tail radar data taken in Hurricane Bill. These data were transmitted in real time to Penn State and used to initialize their ARW ensemble model. Figure 7b shows the wind field analyzed by the Penn State EnKF system through assimilating the radar data. Figure 19, to be described later, shows the impact of this radar data on regional model forecasts.

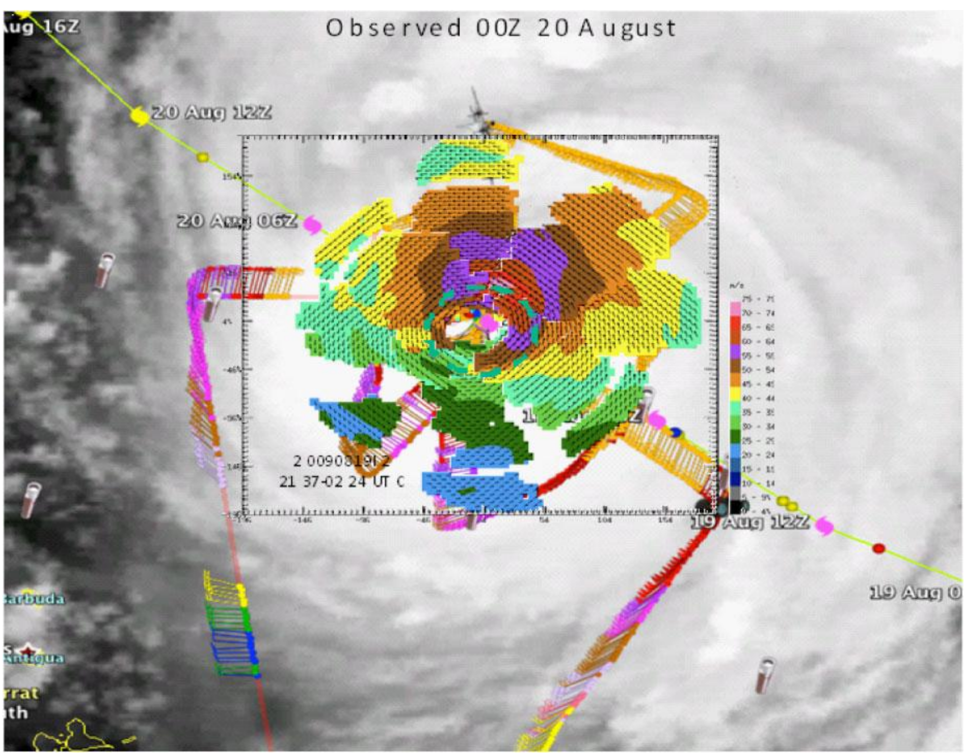


Figure 7a. Wind observations retrieved from the NOAA P3 tail radar for Hurricane Bill 20 August 2009. Color coding for the winds is shown on the right in m/s.

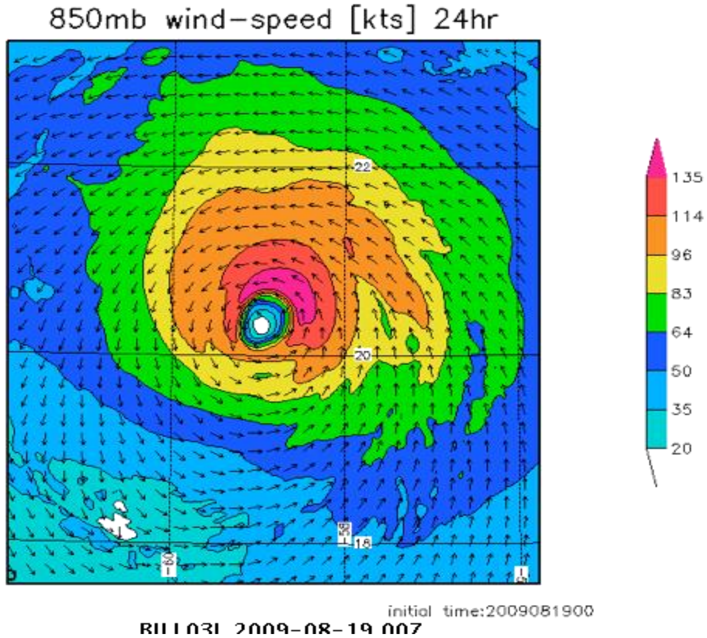


Figure 7b. Initial wind fields at 850 mb in the ARW model run in real time using the radar data shown in the upper figure. Plots are for Bill at 00Z 20 August 2009. Color coding for the winds is in knots.

7. Stream 2 activities: Results from FY08 and FY09 Demonstration System

7.1. The 2008 hurricane season

HFIP performed a series of experiments during the summer of 2008 which provided some preliminary results on a limited number of cases. These experiments were accomplished using the Texas Advanced Computing Center (TACC) computers in CY 2008. Because of the limited number of cases, the conclusions are only tentative. However, there is evidence that:

- Global ensembles can provide improved guidance for forecasters at lead times of at least 5 days.
- Introducing radar data taken by aircraft improves the initial state of the model, and hence, improves subsequent forecasts.

7.2. The 2009 hurricane season

During the 2009 hurricane season HFIP conducted the Stream 2 Demonstration program noted above. In 2009, we were not yet ready to implement Stream 1.5 (see Table 1)) thus the bulk of the work this last year focused on Stream 2 which included the near real-time demonstration system. There will be additional retrospective runs for the 2008 season to increase the statistical sample for the performance evaluation.

Table 4 summarizes the global models run during 2009. Table 5 is the same for the regional models.

Table 4: Global models run in the Demonstration System during 2009

30 km EnKF Data assimilation System was run during Aug. and Sept.
FIM Deterministic models run each day during Aug-Sep: • **30 km (Initialized with GSI-3DVAR and EnKF).** • **15 km (Initialized with EnKF).** • **10 km (Initialized with EnKF).** **Global Ensembles:** • **30 km FIM (initialized with EnKF) 20 members.** • **27 km GFS (Initialized with GSI-3DVAR) 5 members.** • **55 km NOGAPS (3DVAR then 4DVAR), 9 members**

Table 5: Regional models run in the Demonstration System during 2009

Multi Model Ensemble (various initialization schemes):
• **(run for all storms—not all models present for all run times)** – **HWRF** **9km** – **HWRF4** **4km** – **GFDL** **7.5km** – **HWRF-x** **3km** – **WRF/ARW/NCAR** **1.3km** – **WRF/ARW/FSU** **4km** – **COAMPS-TC** **5km** – • **Single model Ensemble (run for most storms)** – **WRF/ARW/PSU** **4.5 km 30 members** a) **Initialized with an EnKF system** b) **Initialized with P3 radar data when available**

In the following section we will describe some results from the Demo System. Later we will summarize lessons learned from the demo system and other HFIP work followed by a summary of challenges for HFIP learned from these lessons.

During September and November 2009, NCEP set up experimental high resolution ensemble through the TACC high performance computer system. The configuration includes:

- GFS 2008 version (T382L64)
- Use T382L64 GSI analysis running ETR (Ensemble Transform with Rescaling) with cycling every 24 h (every 6 h for current operation) to generate 4 ensemble initial perturbations
- Upgrade to T574L64 resolutions for integration
- 5 members (including control) forecast out to 168 h
- No stochastic perturbations
- Period: 09/01 – 09/20/2009, 09/25-09/26/2009, 11/06-11/09/2009

Unfortunately, there were no strong storms during the period when the NCEP high resolution ensemble experiment was running. However, there were several cases where significant differences were found between the high resolution ensembles and the operational Global Ensemble Forecast System (GEFS) as shown in Figure 8. The high resolution ensemble successfully captured storm development reasonably well. Figure 9 shows the statistics (small samples) of ensemble mean track errors from different resolutions. Note that there are only 5 members of T574L64 ensembles, and there is no significant difference from T190L28 ensembles with 20 members and well tuned model physics.

Preliminary results from the high resolution GEFS ensemble experiments, indicate that enhanced ensemble performance may require tuned initial perturbations, optimum cycling (every 6 h), and stochastic perturbations.

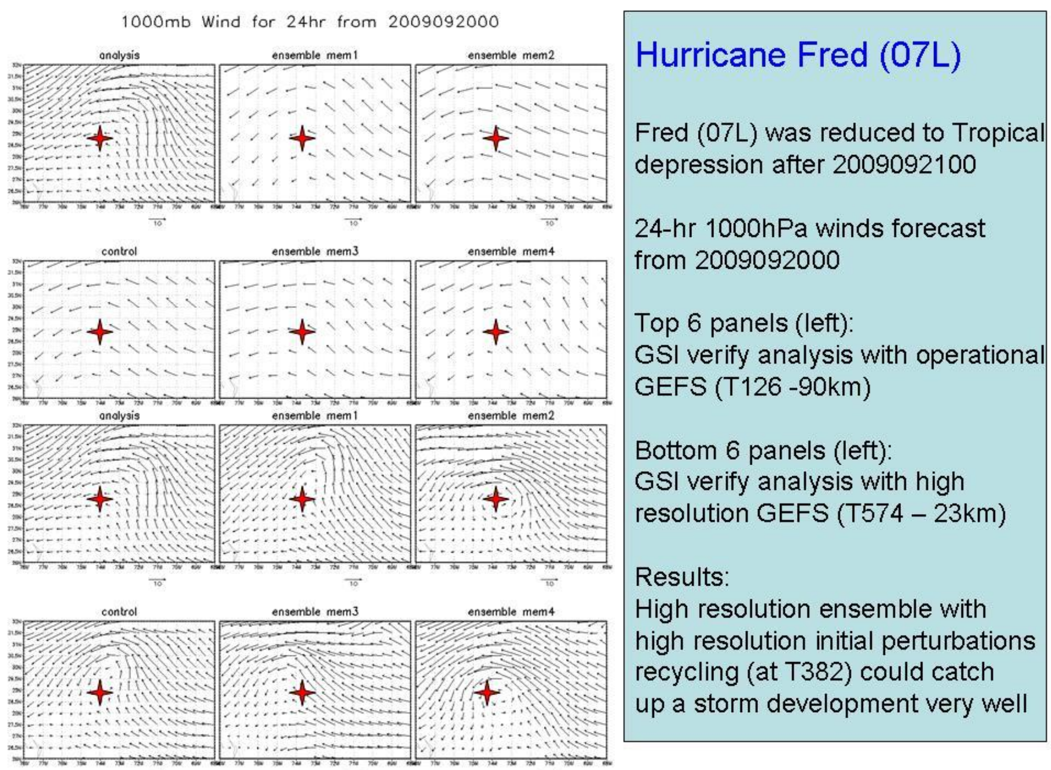
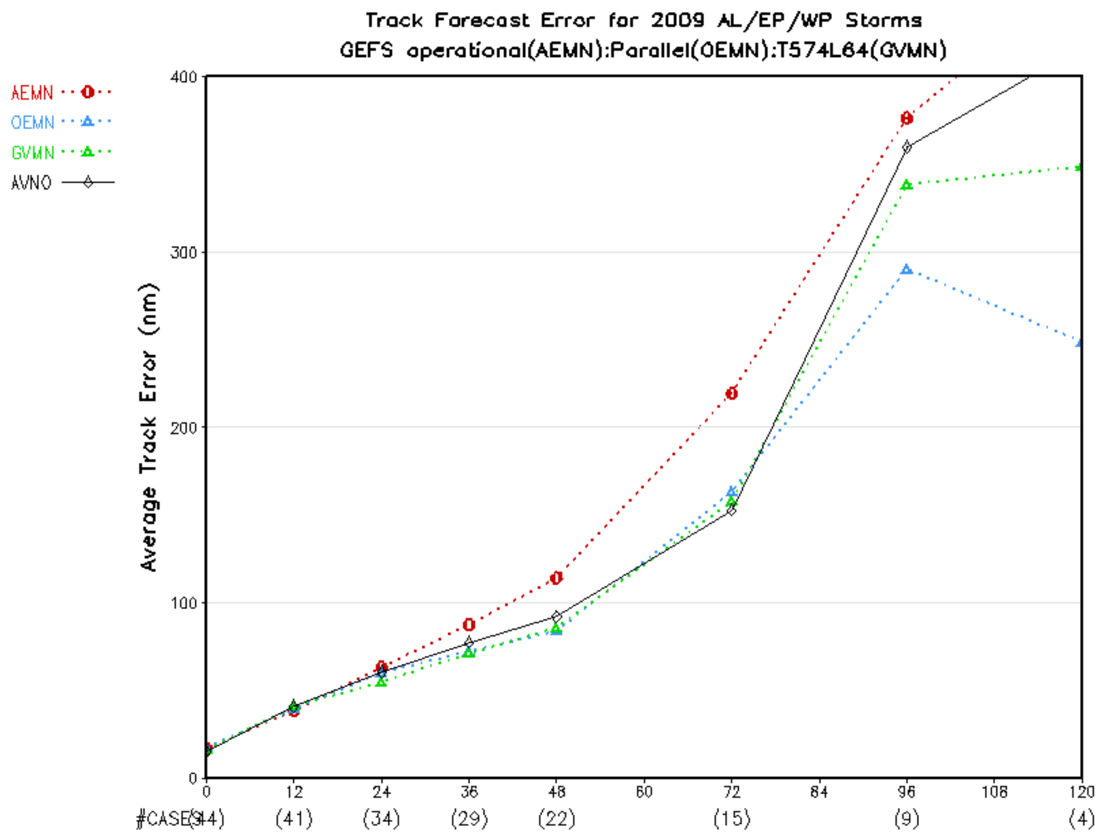


Figure 8. 24-hour 1000hPa wind forecast from 00UTC 20 September 2009 for hurricane Fred (07L). The top 6 panels show the high resolution analysis (T382L64/GSI analysis, top left) and low resolution ensemble forecast (T126L28). The bottom 6 panels show the high resolution analysis (T382L64/GSI analysis, top left) and high resolution (T574L64) ensemble forecast.



In Figure 10 we show ensemble forecast tracks for two different times during Hurricane Bill.

Figure 9. Ensemble mean track errors for the period of 09/01 – 09/20/2009. AEMN (red) is for operational (T126L28 and 20 members). OEMN (blue) is for parallel (T190L28 and 20 members) implemented in February 2010. GVMN (green) is for experiment (T574L64 and 5 members). AVNO (black) is for NCEP deterministic forecast at T382L64.

similar to the ensemble.

For the second time period, the spread of the tracks for all members of the ensemble and the deterministic models was very narrow indicating a high degree of confidence in the forecasts. The verification results were excellent for this case.

In Figure 11 the lower panel shows a period in the storm when the ensemble of forecast tracks were to the right of the observed track although perhaps another way of looking at the forecast is that the verifying track was near the left edge of the ensemble of forecast tracks.

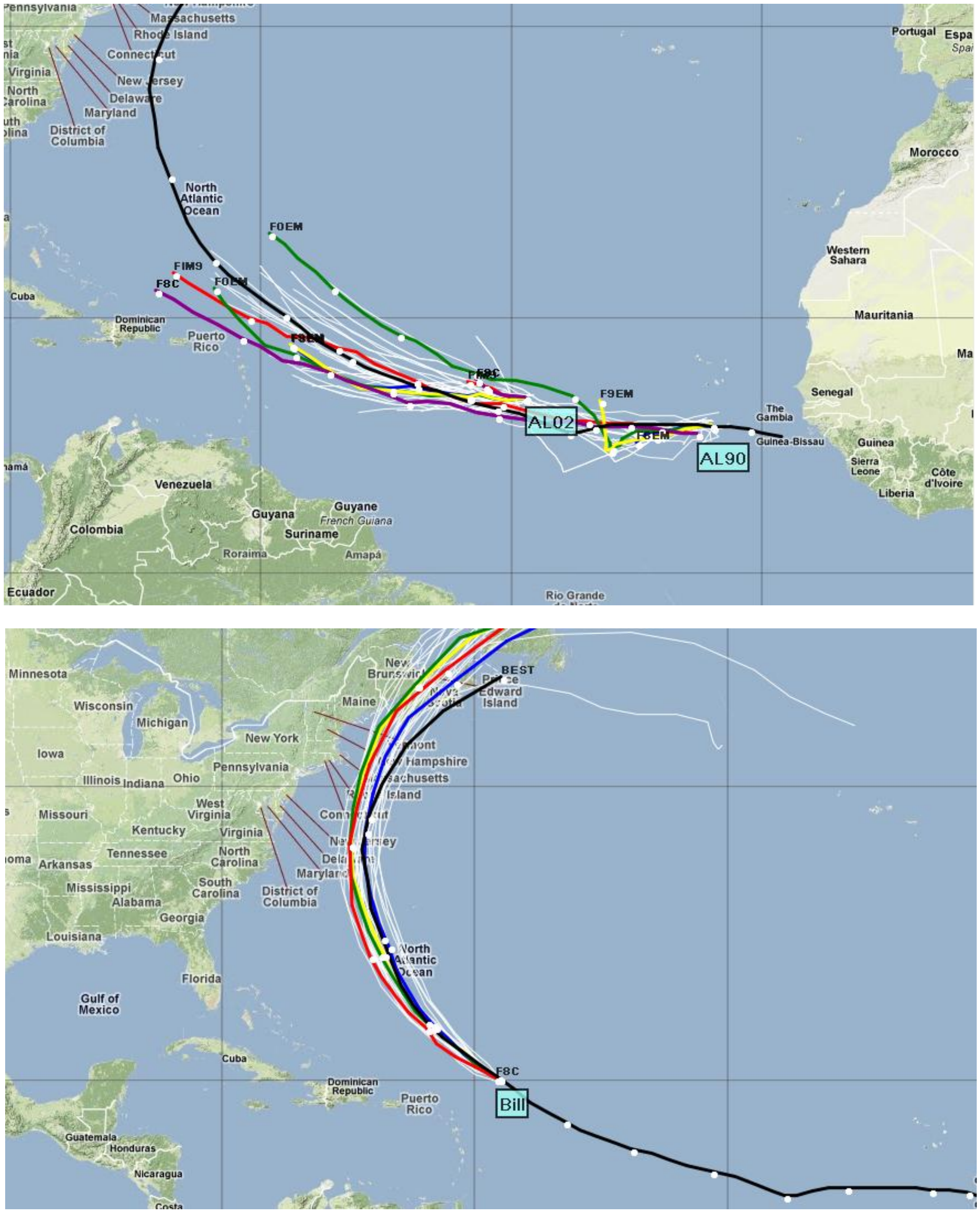


Figure 10. Forecast tracks by the FIM global model runs and FIM global ensemble. The upper figure is for an initial time of 0000 UTC 8/14/09 and the lower figure was initialized at 0000 UTC 8/20/09. Note that in the upper figure tracks for two storms are shown: AL02 was the disturbance that became tropical storm ANA and AL90 the disturbance that became Hurricane Bill. The longer tracks are those associated with Bill. In the lower figure only Bill is shown. The black track in both figures is the observed track for Bill. The white tracks are from the 20 member 30 km ensemble, magenta and green are 30 km deterministic runs with different initial conditions (EnKF vs. GFS), the yellow and red tracks are from the 15 km deterministic model with the different initial conditions and the blue line is from the 10 km deterministic run using the EnKF initial conditions. In all cases the forecast length was 168 h or 7 days.

The upper panel in Figure 11 shows each member of the 30 km ensemble run from the same initial time as the lower part of the figure and verifying 5 days later. Each “postage stamp” is centered on the observed location of the hurricane. This is another way to look at the ensemble information and can be used to give some idea of the rate of intensification forecast by the various members. However, note that at 30 km resolution, the deep central pressure observed in hurricanes will not appear since the central region of the hurricane is not resolved. The track forecast information is also available in this presentation, but note that all forecasts are to the east and north of the observed location. This information is also available in the lower part of the figure.

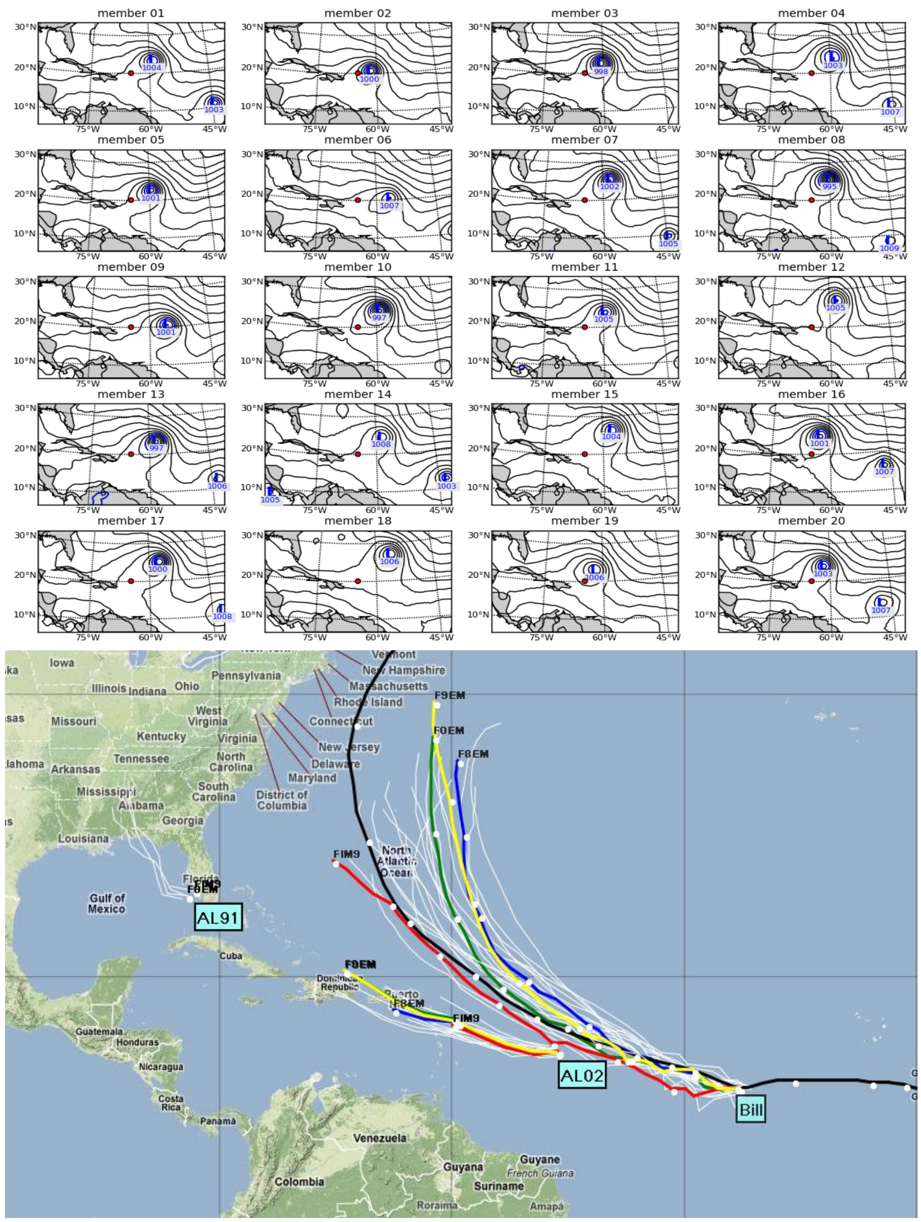


Figure 11. The lower figure is the same as for Figure 10 except it was initialized at 0000 UTC 08/16/09. The upper figure shows the 20 members of the 30 km ensemble initialized at 0000 UTC 08/16/09. The upper figure is a 120 (5 day) forecast verifying at 0000Z UTC 08/21/09. The contour field is sea level pressure every 4 mb and the forecast central pressure is indicated. The red dot on each figure is the verifying location at 0000 UTC 08/21/09.

The presentation in the upper part of Figure 11 also provides an indication of the probability of genesis. The hurricane that appears in the lower right in about half of the members was the one that eventually slowly and briefly intensified to Tropical Storm Danny. This ensemble would indicate about a 50% chance that the system following Bill would intensify into at least a tropical storm. So far we have not yet finished developing the software to verify the genesis statistics (that is underway) but a cursory examination of the results indicate that the global model used in these experiments is over active with respect to hurricane genesis. At this point that is true of almost all global and regional models. The reason is not known at this time but may be related to physics parameterization or the lack of an active ocean coupled to the atmosphere model, particularly with the global models.

Figure 12 shows two snapshots from the 10 km FIM global model started at the same time as the lower panel of Figure 10; the upper panel of Figure 12 is a 72 h (3 day) forecast and the lower panel is a 168 h (7 day forecast). These panels illustrate the global nature of the 10 km forecasts. The top panel shows Hurricane Bill off the East Coast of the U.S. and Typhoon Vamco in the central West Pacific. At the end of the forecast period, bottom panel, both storms had been absorbed into, respectively, the Icelandic Low and the Aleutian Low. The red lines on the figures show their tracks. To the east of Bill, the development of the next system that became Danny is apparent. These forecast indicated a very active genesis period in the Eastern Pacific with four tropical depressions or tropical storms active on 08/27/09 according to the model. This verified to some extent in that during this forecast period three named storms developed and one, Jimena, became a category 4 hurricane. But generally it was noted that the FIM over-predicted genesis.

Figure 13 shows verification statistics for models run from the GSI initial state (the operational data assimilation system) and the EnKF system. These data assimilation systems were used to initialize the operational GFS and the FIM. Shown in the figure are the RMS error and anomaly correlation coefficient for 72 h forecasts with both models. Here we show verification of just one field: the tropical winds (30N-30S) at 250 mb. It turns out that this field showed one of the biggest differences between the models and is important in determining hurricane environment. Other fields such as the 250 mb wind in polar regions showed much less difference.

Recall that smaller RMSE is better, and larger anomaly correlation is better. Note that when the EnKF initialization replaced the GSI initialization in both models there was a substantial improvement in the forecasts for both models. Comparing the models when the same initialization system is used for both, the FIM model was better than the GFS system. Note however, that the FIM was run at 15 km and the GFS at 30 km. It is not known at this time whether this improvement was due to the increased resolution of the FIM over the GFS or to differences in forecast models or initialization schemes. In addition the EnKF system introduced a vortex based on observations of the individual hurricanes which led to initial vortices that were initially considerably stronger than when the models were initialized by the GSI. We return to this issue in the “Encouraging Results” section below.

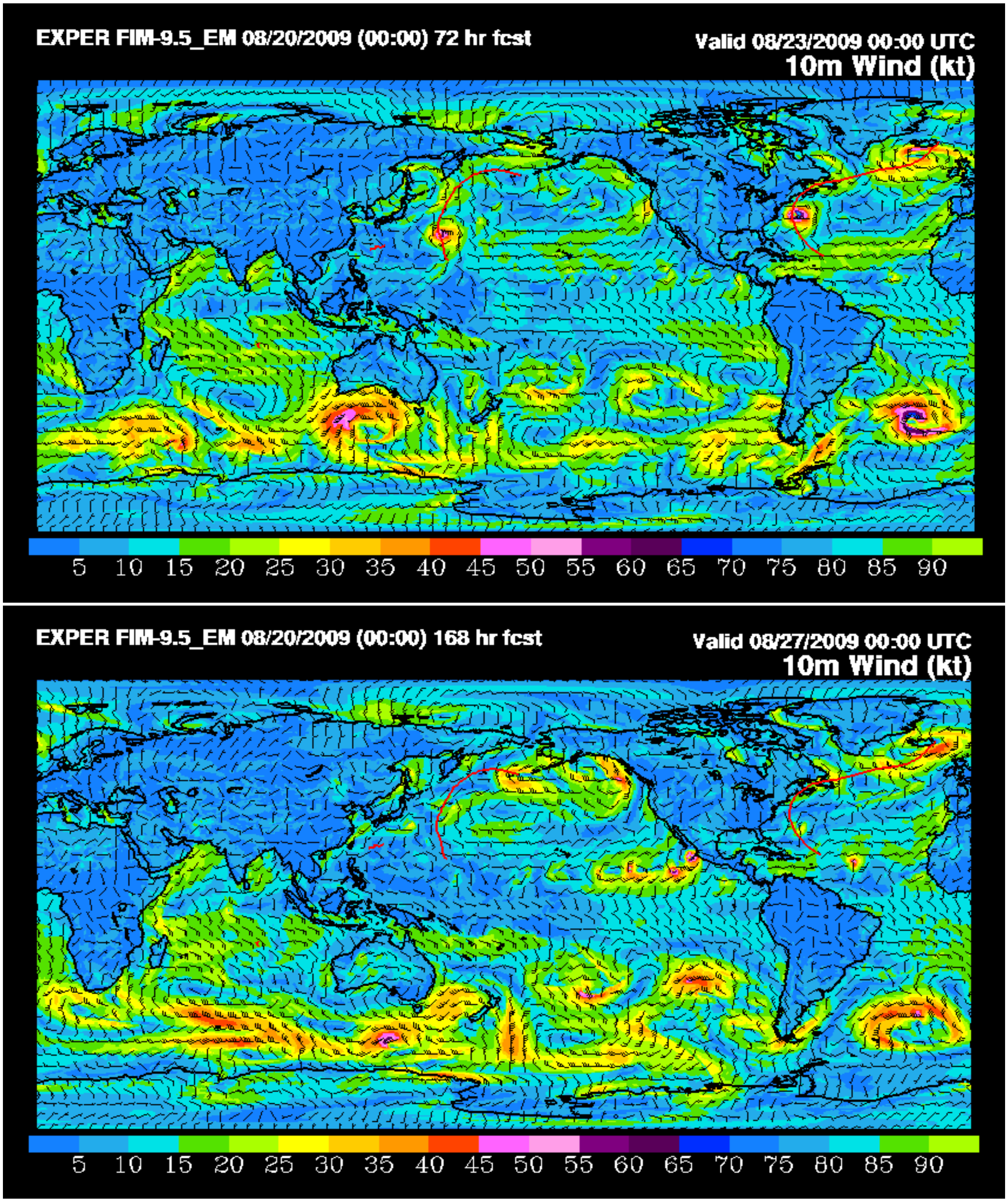


Figure 12. The above two plots are snapshots of the 10 wind speed and wind barbs from the 10 km global model initialized at 0000 UTC 082009, the same as the lower panel in Figure 10. The color shading shows the 10m wind speed in knots. Red lines are the forecast tracks from the initialization time. The upper figure is a 72 h (3 day forecast) for 0000 GMT 082309; the lower panel a 168 h (7 day) forecast for 0000 UTC 08/27/09.

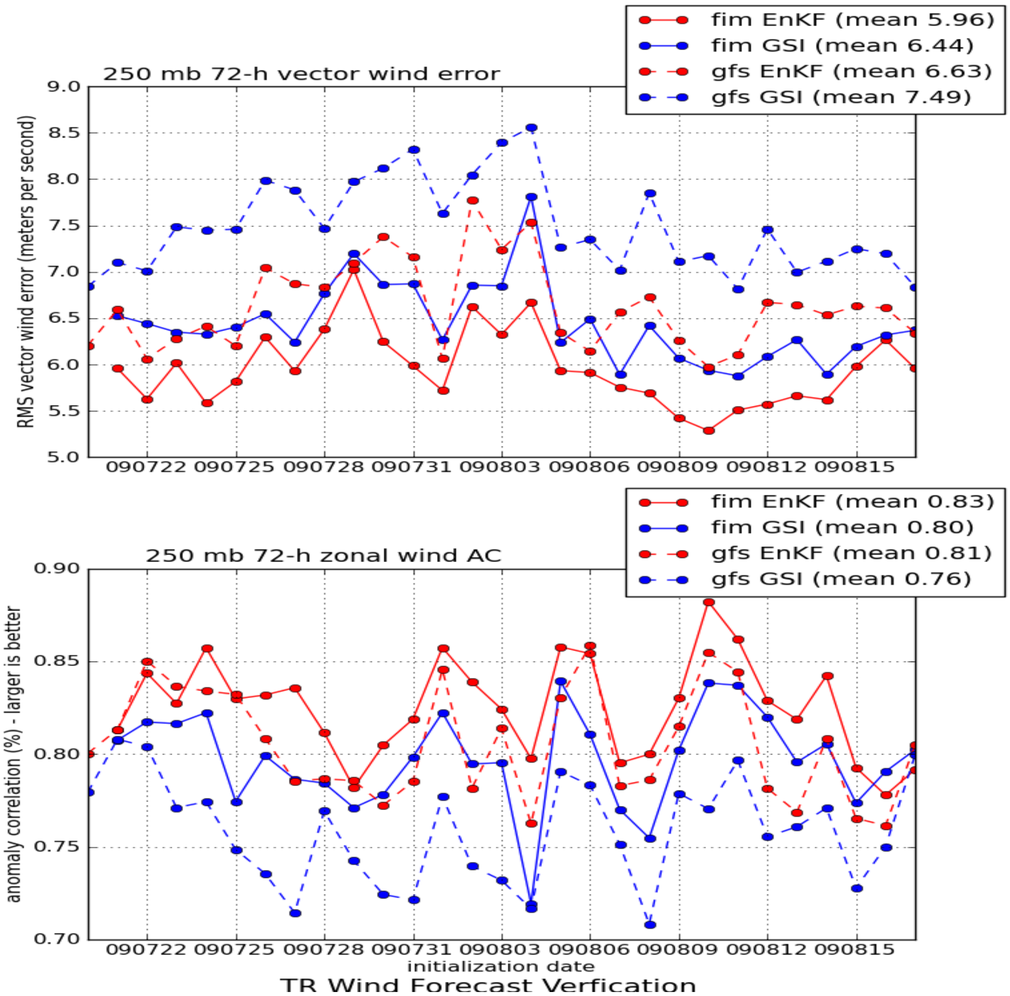
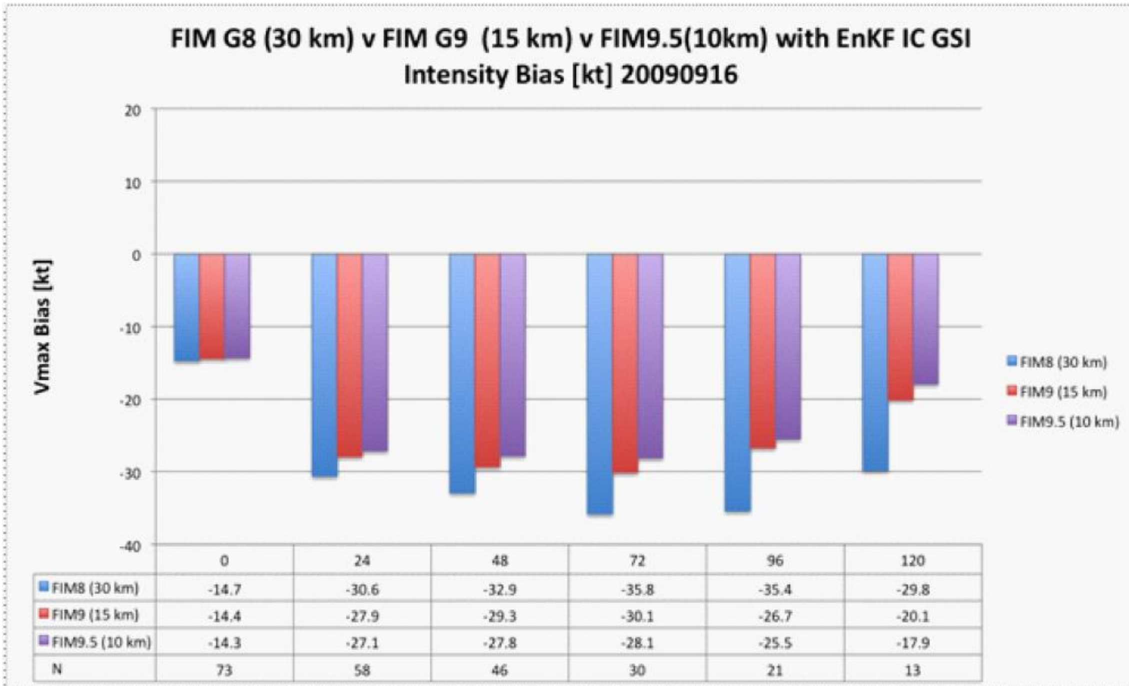


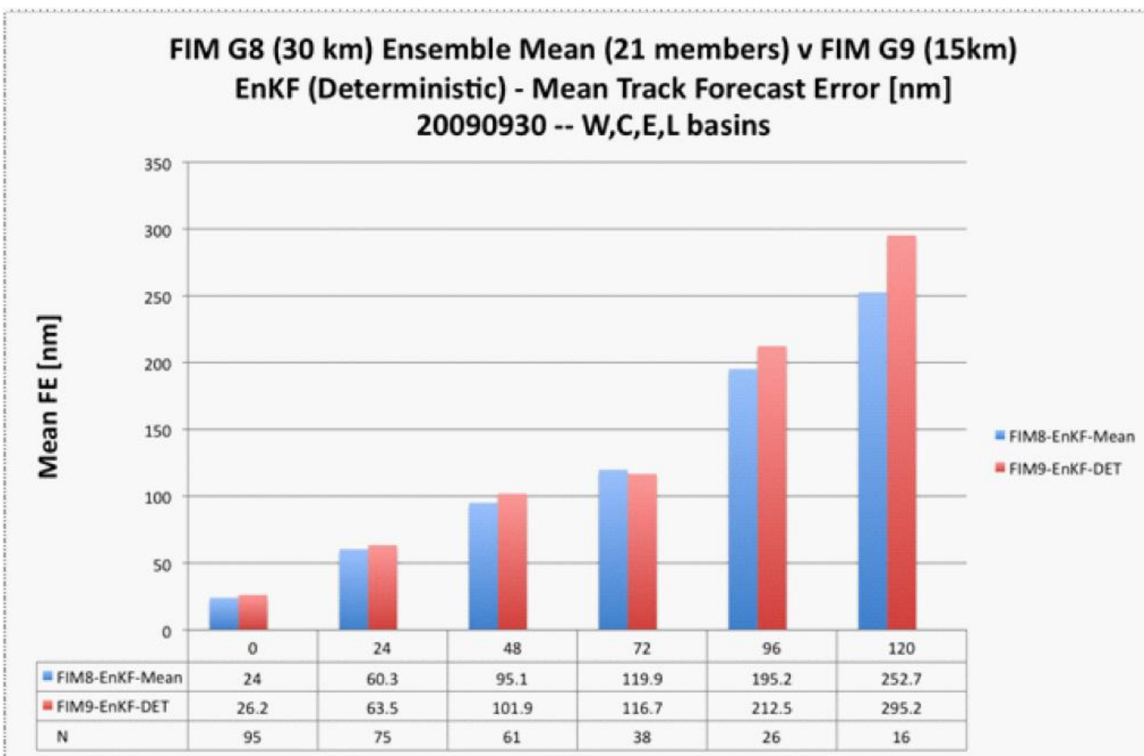
Figure 13. Verification statistics for forecasts run from initial conditions specified by the GSI operational data assimilation system and the EnKF experimental data assimilation system. The statistics for a 72 h forecast of winds in the tropics at 250 mb are shown for the Global Forecast System (GFS) operational model run at 30 km and the experimental FIM run at 15 km each using both sets of initial conditions. The legend on the upper right of each figure indicates the initialization system and model used for each curve. The number in parentheses in the legend shows the average skill for each configuration. Date is indicated along the horizontal axis using yymmdd at 0000 UTC. Shown are the RMS vector wind error (m/s, top panel) and Anomaly Correlation (%, lower panel). In the top panel, lower is better and in the bottom panel higher is better.

Figure 14 shows the impact of resolution in the FIM global model on intensity forecast bias. All three models were initialized with the 30 km EnKF system noted above, and use a vortex relocation process to initialize the vortex at the correct location so all start out with the same initial bias in part because the vortex is not well resolved in the 30 km model. With time the higher resolution models improve relative to the 30 km run with the 10 km model performing the best. The smaller biases at the longest lead times are partly because the weaker storms have decayed by then leaving the stronger and more easily forecast storms. There are also fewer cases in the statistics at the longer lead times which may bias the result. The increasing negative bias out to 72 h is because there are many weak storms in the 2009 sample, and these were poorly forecast but the sample size is very small.



	0	24	48	72	96	120
FIM8 (30 km)	-14.7	-30.6	-32.9	-35.8	-35.4	-29.8
FIM9 (15 km)	-14.4	-27.9	-29.3	-30.1	-26.7	-20.1
FIM9.5 (10 km)	-14.3	-27.1	-27.8	-28.1	-25.5	-17.9
N	73	58	46	30	21	13

Figure 14. Intensity bias in the FIM global model for various lead times (hr) and resolutions of 30 km (blue), 15 km (red) and 10 km (purple). The numbers labeled N along the bottom indicate the number of cases contained in the statistics.



	0	24	48	72	96	120
FIM8-EnKF-Mean	24	60.3	95.1	119.9	195.2	252.7
FIM9-EnKF-DET	26.2	63.5	101.9	116.7	212.5	295.2
N	95	75	61	38	26	16

Figure 15. Comparison of track errors for the 15 km FIM global model (red bars) and the 30 km FIM global ensemble (blue bars) for all basins (Atlantic, East Pacific, Central Pacific and West pacific) for 2009.

Figure 15 compares the track accuracy for the mean from the 20 member, 30 km ensemble with the error of the 15 km deterministic global model. Though the number of cases is still rather low, some promising improvements are suggested using the ensemble, especially at the longer lead times (however, note the small sample sizes for these lead times). HFIP plans to run the FIM global ensemble at higher resolution (15 km) where we expect to see still further improvement.

8. Encouraging Results from FY09

Table 6 gives a high level overview of some encouraging results from the HFIP work during FY09.

Table 6 Encouraging Results from FY09 Demonstration System

1. ***Advanced data assimilation systems (EnKF, 4DVAR) appear to improve global forecasts over the current operational data assimilation system.***
2. ***High resolution global and regional ensemble systems are showing promise but require further testing and evaluation.***
3. ***High resolution global ensembles (30 km, 20 members) can be run in real time on available computing resources meaning higher resolution is definitely possible.***
4. ***The multi-model regional ensemble showed promise.***
5. ***There is some preliminary evidence that airborne radar data used to initialize the hurricane vortex in regional models can improve forecasts of track and intensity.***
6. ***Various physics parameterization schemes were examined using HWRFX and COAMPS. Great sensitivities of storm structure to variations in the model physics were noted, strongly suggesting possible routes for improving the physics package in the operational HWRF model.***
7. ***The use of AXBT data assimilation also showed promise.***

Advanced data assimilation systems (EnKF, 4DVAR) appear to improve global forecasts over the current operational data assimilation system

This was a major result for FY09. The computing resources that HFIP was able to locate outside NOAA allowed an extensive test of the EnKF data assimilation system and a comparison with the current GSI data assimilation system. Figure 13 shows one example of this comparison. It shows comparisons of the EnKF system used in two of the models described briefly above, the FIM and the GFS, compared to the same models initialized with the GSI, the same system as is being used operationally. Shown are statistics for forecasts of the 250 mb tropical winds. The largest impacts were found in the tropics though positive impacts were noted elsewhere and for other quantities. For the GFS, the improvement in the forecast of the high-level tropical winds was about 10%.

Figure 16 compares performance of the GFS model initialized both with the GSI and EnKF with the emphasis on hurricane track. Out to 4 days in this comparison there is a steady improvement with forecast lead time with about a 20% improvement at 4 days. In this figure the same bogus vortex was used in the GSI as used in the EnKF. Day 5 was left off the figure because there were very few cases (6) at that lead time for the comparison.

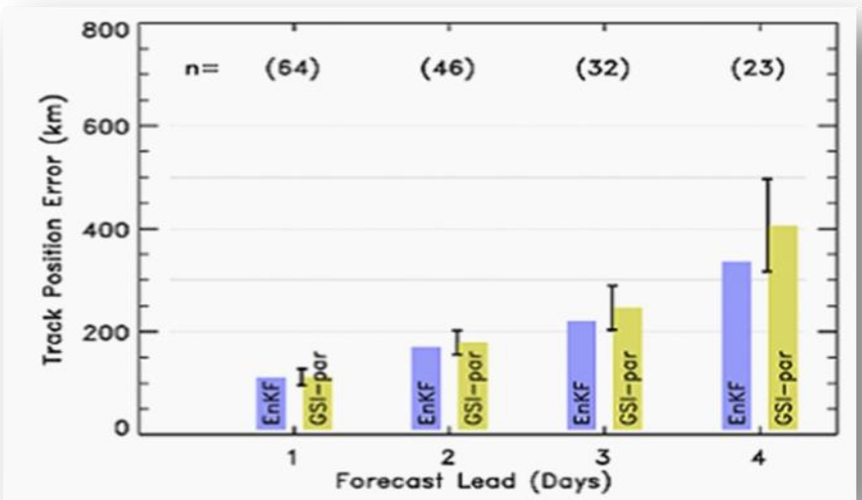


Figure 16. Comparison of hurricane track forecast errors in the GFS when EnKF and GSI are used to initialize the model. Data are from 2009 for all basins.

Figure 17 compares the NAVDAS 3DVAR system with the Navy's new 4DVAR system. As with the EnKF system, 4DVAR shows improvement in track forecasts for all lead times and by Day 5 the improvement amounts to 18%.

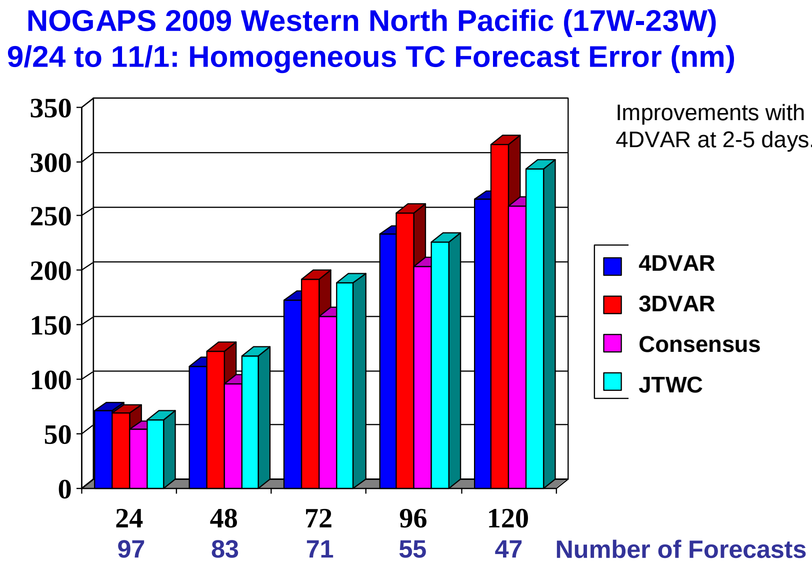


Figure 17. Comparison of NOGAPS initialized with 4DVAR and 3DVAR and with the official JTWC forecast. Consensus is the mean of the 3DVAR and 4DVAR forecast

Finally Figure 18 compares the performance of an ensemble forecast from the ECMWF with the performance of an ensemble forecast using the GFS at much higher resolution than is currently used in the GFS operational ensemble (at the same resolution as the deterministic GFS operational model). Here the EnKF data assimilation systems used to initialize the GFS ensemble. The ECMWF model is globally considered to be the current leader for global model forecasts. Using both higher resolution in the GFS and a more advanced data assimilation system (EnKF in this case) the skill of the GFS system can be made to match that of the ECMWF. It has not been established how much of this ability to match skills comes from the data assimilation system and how much comes from improved resolution though both are likely important. It is also not known which of the data assimilation systems, 4DVAR or EnKF, will prove better in the end but the results shown in Figure 18 suggest that either provide a significant improvement in hurricane track forecast skill. A hybrid that uses both may be the optimum approach to data assimilation although this hypothesis still needs to be tested.

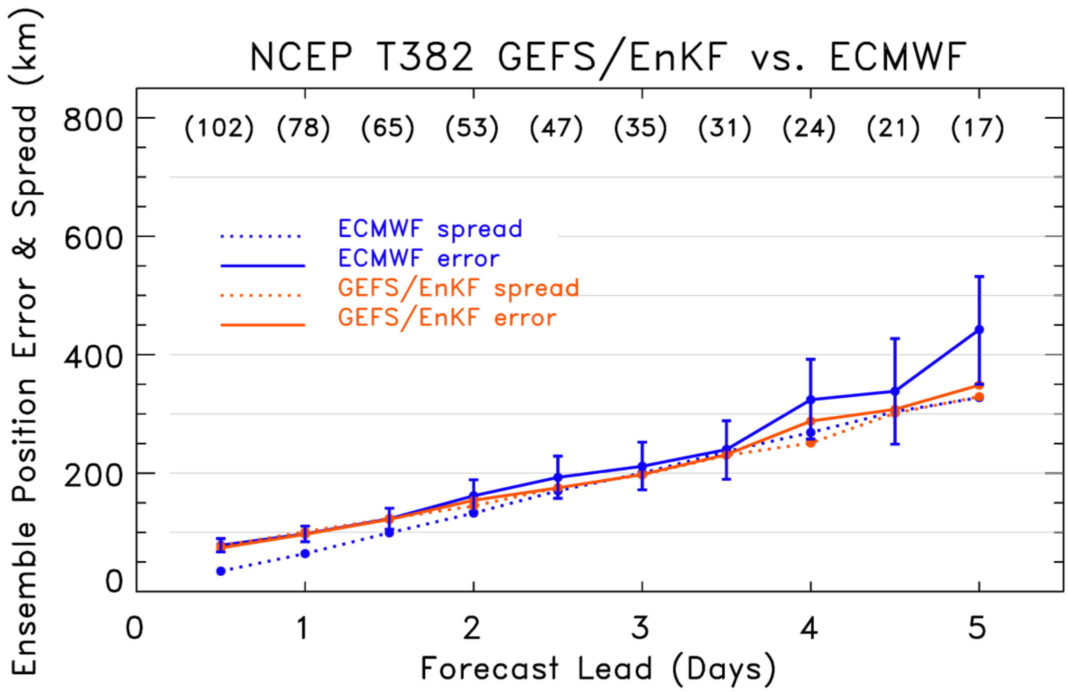


Figure 18. The NCEP GEFS (Global Ensemble Forecast System) run at T 382 initialized with an EnKF system compared to the ECMWF ensemble at T399 initialized with 4D-VAR

A regional EnKF (developed at CIRA) has been employed to produce situation-dependent background (SDB) error covariances for future use for anisotropic error covariance modeling at NCEP. The system employs 32 ensembles, and is used with HWRF (moving nest resolution 6.6 km) and with NCEP operational observations. The system directly employs HWRF and GSI infrastructure components through specially developed interface scripts. In this preliminary stage, mostly devoted towards developing and testing the new EnKF-HWRF system, the system was applied to Hurricane Gustav (2008). There are several conclusions that can be drawn: (i) moving nest can be successfully initialized from outer domain ensembles, (ii) EnKF-HWRF system creates error covariances that are smooth and dynamically consistent, and (iii) there is a great need for cloudy radiance assimilation. The problem of initializing the moving nest exists because the ensemble perturbations from the previous cycle cannot be used in the current cycle, as geographical domain of the nest has changed. We adopted a strategy to reinitialize the moving nest ensembles from outer domain ensembles (done automatically in HWRF), and our results confirm that this is a viable approach to EnKF assimilation of inner domain observations. In Figure 19 we illustrate the error covariance structure in the moving nest.

(a) (b)

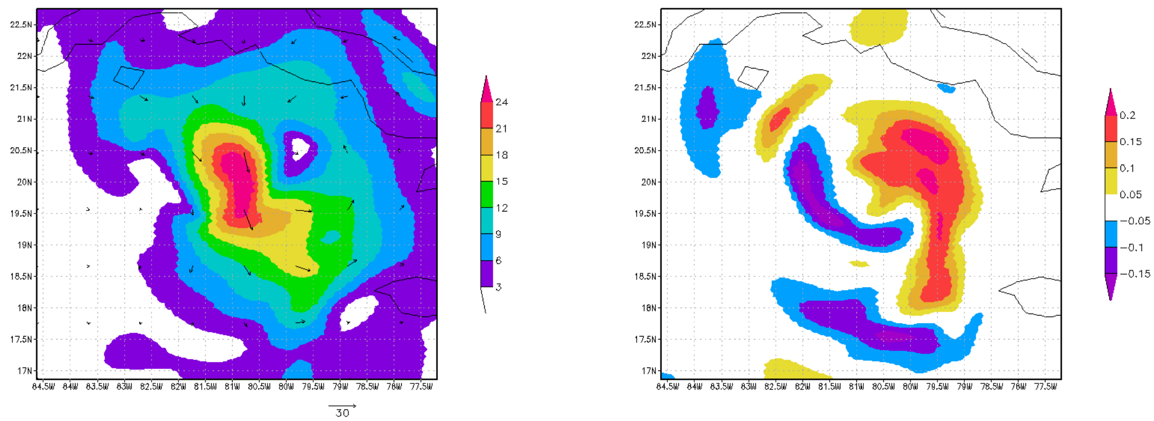


Figure 19. Error covariances from the first ensemble member at 850 hPa, valid 12UTC on 31 August 2008 for: (a) wind, and (b) specific humidity. Error covariance structure is smooth and dynamically realistic.

In Figure 20 we show the root-mean-squared (RMS) errors with respect to wind and satellite observations. The figure indicates anticipated reduction of the RMS errors due to assimilation, but also shows a general unavailability of satellite observations due to cloud clearing.

(a) (b)

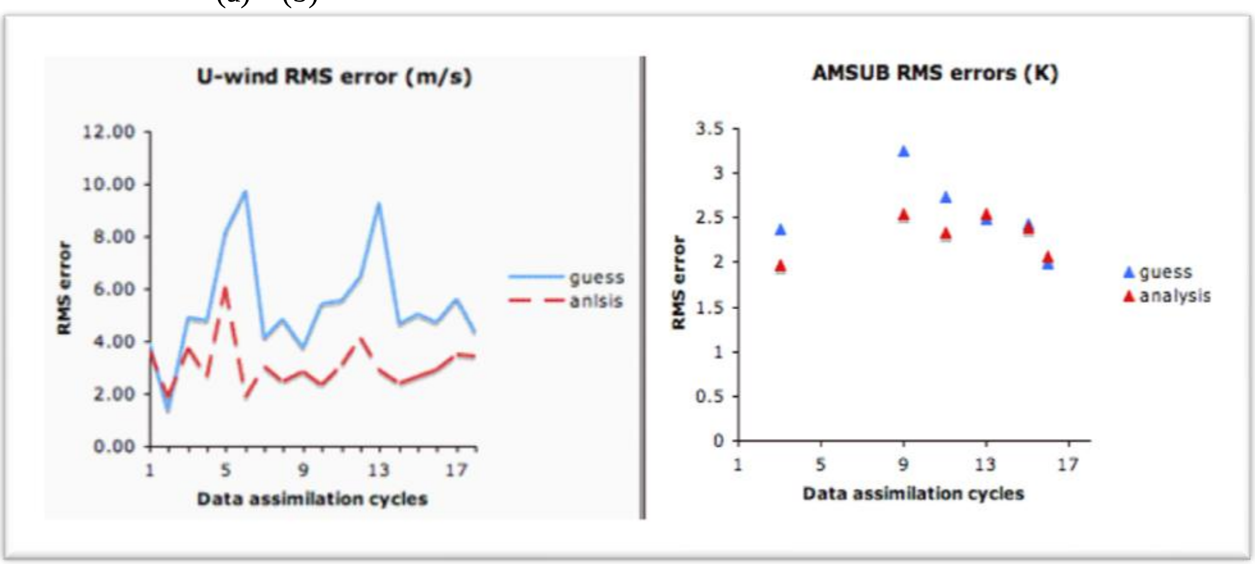


Figure 20. RMS errors with respect to observations for Hurricane Gustav: (a) east-west wind component, and (b) AMSUB satellite radiance (cloud cleared). One can notice general improvement of RMS errors due to analysis, but also that very few satellite radiance observations are available..

High resolution global and regional ensemble systems are showing promise but require further testing and evaluation

For the global models we would expect that most of the improvement in hurricane forecasts would be for track. Because the hurricane is at best only marginally resolved in the global models, (since there are limitations that computational power places on that resolution) the intensity is usually under-predicted (see Figure 14). Furthermore, it is not surprising that as the resolution increases the intensity is less under-predicted, also shown in Figure 14. Thus for the global models the emphasis should be on track for the longer lead times (4-7 days)

Figure 15 compares track forecasts of a 15 km deterministic forecast with the mean of a 21 member ensemble both using the FIM and EnKF for initialization. Even though the deterministic model has twice the resolution of the ensemble (we cannot yet run the global ensemble at higher resolution) the track forecasts of the ensemble appear to be superior to the deterministic model.

Figures 21 and 22 illustrate how regional ensembles can be used in intensity forecasting and came from HFIP work in 2009. Figure 21 shows tracks from an ensemble of 30 members using the ARW version of the WRF regional model. The color coding denotes tracks to the left (orange), middle (green) and right (blue) of the fan of tracks produced by the ensemble. Note that the observed track is to the left of the ensemble mean track (not shown but easily inferred). On the right hand side of Figure 22, traces of the maximum wind over the 4 days of the forecasts are shown and the same color coding as in Figure 21 is used to relate intensity curves to the members falling to the left, middle and right of the forecast tracks.

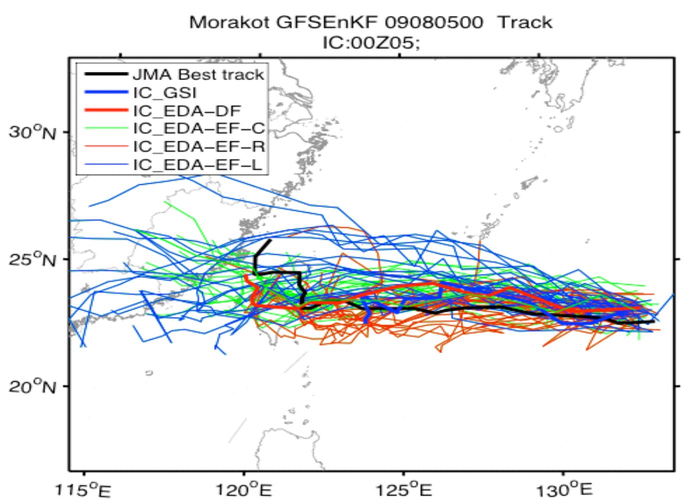


Figure 21. Track forecasts from a 30 member regional ensemble using the WRF ARW regional model for Typhoon Morokot from 6 September 2010 through 9 September 2010 in the Western Pacific. The orange color signifies the 1/3 of the storms that were near the left side of the ensemble of tracks, green the 1/3 near the middle and blue the 1.3 near the right at the 24 h forecast. The heavy orange line is a single deterministic run; the black line is the observed track.

Note from comparing Figures 21 and 22 that tracks on the left side of the ensemble fan tended to have intensities greater than storms on the right side. In fact note also that the observed track was located toward the left side of ensemble fan. We present this not as a proof of skill but rather as a way to present ensemble information to forecasters. As the ensemble will become available several (8-12) hours after the initial forecast time,

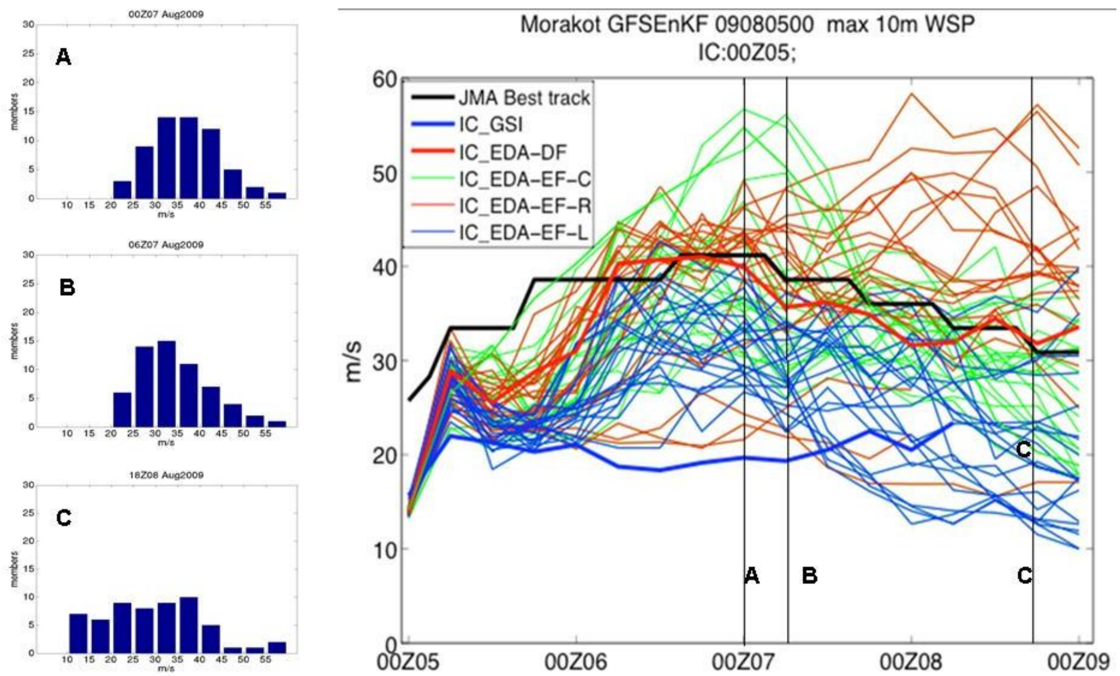


Figure 22. The same as Figure 21 except for the maximum wind over the four days of the forecast. The color coding is the same as in Figure 21 where orange are the intensity plots for tracks near the left edge of the ensemble of tracks, green are those near the center and blue those near the right. The observed wind maximum wind speeds are shown by the black line. The bar graphs on the left edge of the figure show the number of forecast maximum winds that fall in various bins of 5 m/s. They apply at the times A, B and C as noted in the right hand figure.

subsequent information may be used by the forecaster to adjust the forecast towards the more likely simulations.

To continue the spirit of demonstrating a ways to present ensemble information, the left side of figure 22 shows the frequency of forecast intensities at three different times denoted A, B and C. A and B are at the 48 and 54 hour forecast lead times, respectively, and display a typical probability distribution, when there is skill, with the most probable being near the mode of the distribution. A and B are six hours apart and show a forecast intensity that is close to the observed though somewhat weaker. A comparison at times A and B suggest that at 48 hours the ensemble is forecasting a slight weakening. Its was in fact observed beginning a long period of slow weakening.

Plot C is for much later in the forecast, after 3 days, and the flat distribution suggests that the ensemble has lost predictability in intensity by then. We note below that one of the challenges of HFIP will be to convince forecasters of the value of ensembles, how they can be used, and develop products that they will find useful. Figures 21 and 22 are presented as a way to begin the dialog between ensemble scientists and forecasters on how to present ensemble information.

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

One of the assumptions of the HFIP program is that the use of high resolution ensembles both in the regional and global models will be a key to meeting its goals of a 20% improvement in 5 years and 50% in 10 years. Of concern is the computing capability necessary to run the high resolution models in ensembles, particularly the global models, using a more advanced data assimilation scheme such as 4DVAR or EnKF. An objective of HFIP is to eventually be able to run a global ensemble of at least 15 km with 20 members. During the summer of 2009 summer we were able to run a 30 km ensemble using an EnKF data assimilation system. The 20 member ensemble was run once per day each day for 2 months. The data assimilation system ran continuously for the 2-month period 4 times per day. Both systems were run in real time on the Texas Advanced Computer Center computers.

While the 30 km resolution is still half of our ultimate goal for the global ensemble, it still is one of the highest resolution ensemble runs anywhere in the world in FY09. Thus with the computational resources available we were able to prove that at least a 30 km ensemble is possible though the currently available operational computing will not allow it. This provides a guide on the capabilities of future operational computing.

There is some preliminary evidence that airborne radar data used to initialize the hurricane vortex in regional models can improve forecasts of track and intensity.

Radar data from the NOAA P3 aircraft tail radar have been gathered in hurricanes for many years and used for understanding the structure and dynamics of the inner core of the hurricane. Only recently have there been attempts to incorporate that data into the initialization of the hurricane vortex in models. As we will note below, one of the lessons learned from the FY09 hurricane season was that the current methods of initialization of regional models was likely a reason for poor model forecasts during the 2009 season. A characteristic of the FY09 hurricane season was that many of the storms experienced considerable shear during their life cycle. When the storms were initialized at a time shear was present, the intensity forecasts were typically poor and often continued to intensify in the model even though the actual storm decayed under the shear. While there are a number of possible reasons for this problem such as a problem with the physics parameterizations, the convective parameterization and initialization of the vortex are particularly likely causes. Many current initialization methods introduce a vortex that is mostly vertical whereas in sheared storms the vortex becomes tilted.

One solution to the initialization of the vortex is to use high resolution data taken within the hurricane core such as radar data taken from aircraft. Figure 23 shows some tests of incorporating the radial wind from the aircraft radar using an EnKF approach to initialize the model. The example shown in Figure 23 suffers from there being a small number of cases available with radar data but it does indicate promise that the radar data can improve the regional model forecasts of intensity out to 48 h and track beyond.

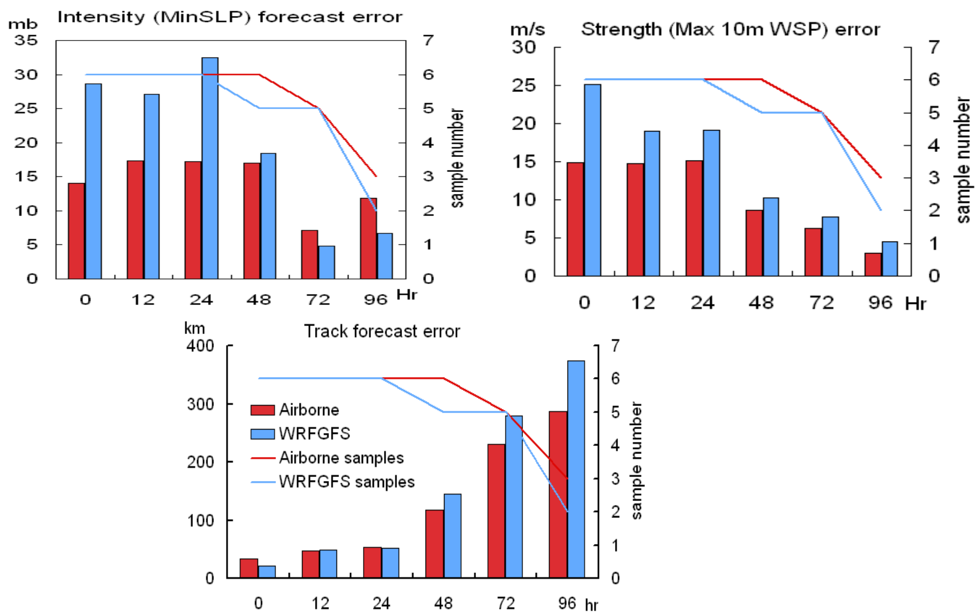
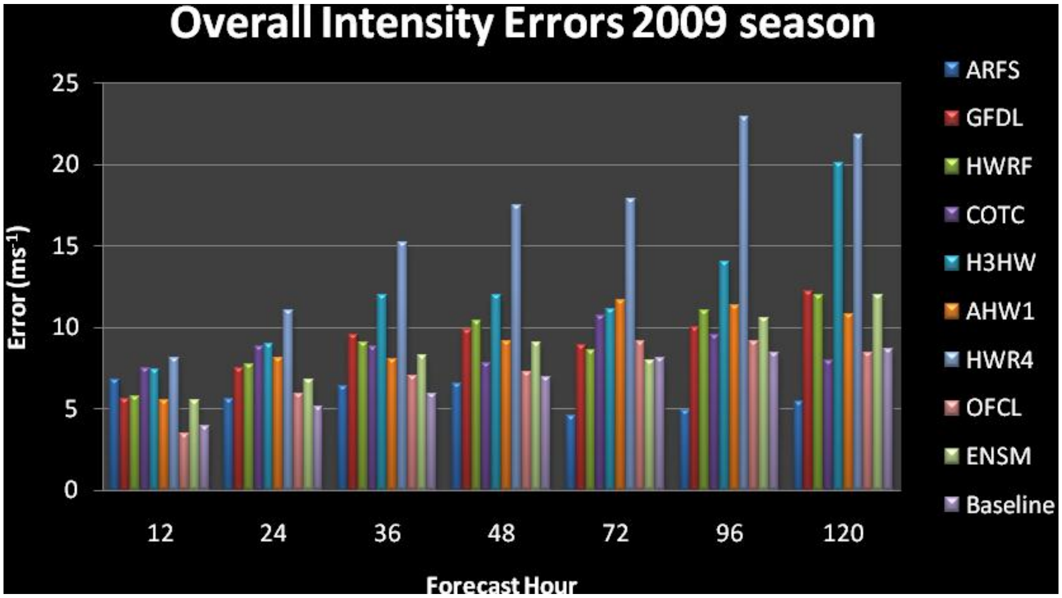


Figure 23. Comparison of regional model forecasts using the Penn State WRF system with an inner nest of 4.5 km and an EnKF data assimilation to include airborne radar data. The blue bars show intensity and track forecasts without the radial winds from the radar data but using all other available data, and the red bars show the results when the radial winds from the radar data are added. The colored lines show the number of samples used in the comparison.



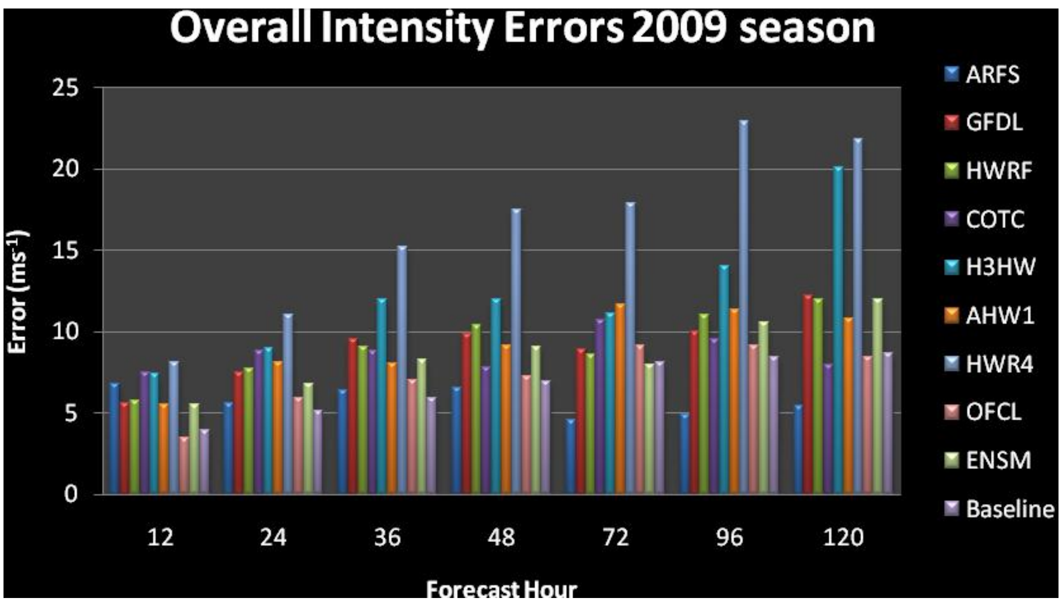


Figure 24. The above figure shows the overall track and intensity errors from the Multi-Model Ensemble (Table 5) for the 2009 Atlantic season. The ARFS model (WRF/ARW/FSU in Table 5) performed reasonably well for both the track and intensity. The ensemble mean (ENSM) performed quite well for track where it ranked among second or third for most of the forecast hours among the component models. For the intensity errors the ensemble mean was not that good as compared to the track forecasts. It suffered from the fact that H3HW (HWRF-X) and HWR4 performed badly during the season. Among the models the GFDL, HWRF and COTC (COAMPS-TC, see Table 5) performed better compared to their counterparts. AHW1 is WRF/ARW/NCAR in Table 5. OFCL is the official forecast error for the 2009 season for comparison.

The Multi-Model regional ensemble run this summer showed promise.

In Table 5 the regional multi-model ensemble was outlined as part of the 2009 Demo system. Figure 24 shows the raw results from that experiment and Figure 28, the bias corrected results. The figures indicate that the ensemble provides useful statistics even in the raw data. For example the ensemble mean provides estimates of position error that are comparable to the official forecast. The same is true for the intensity. Note that the poorest performing model, the 4 km HWRF, was put together very quickly just before the hurricane season and had several, now known, serious problems. It probably should not be included in the statistics of the ensemble mean but we have included it for completeness. The FSU ARW model was only run on a few cases (28 compared to 58 overall cases) in 2009 and is probably strongly biased toward better forecasts than the other models as a result. Still, the results show promise that, with adequate preparation of the component models and better statistical post processing, it can provide useful forecast guidance.

Various physics parameterization schemes were examined using HWRFX and COAMPS. Great sensitivities of storm structure to variations in the model physics were noted, strongly suggesting possible routes for improving the physics package in the operational HWRF model.

To help understand the response of the operational HWRF model to various physics permutations available in the WRF modeling framework, as well as to explore how to improve the operational HWRF model physics package systematically, an idealized initialization package has been implemented in the experimental version of the HWRF (HWRFX) model during FY2009. This package allows users of the HWRFX model to initialize the model with a weak axisymmetric vortex disturbance in various idealized tropical environment conditions that are favorable for the vortex disturbance to develop into a hurricane.

Sixteen physics sensitivity experiments with various permutations of physics options are available within the WRF modeling framework. All of the experiments are run with a 9:3 km nested grid configuration. They are all initialized on an *f*-plane located at 12.5°N with a prescribed axisymmetric vortex of the maximum surface tangential wind of 15 m/s and the radius of surface maximum wind of 90 km. The quiescent environment with the widely used Jordan sounding and a constant sea surface temperature of 29°C is prescribed for obtaining the initial mass and wind fields by solving the nonlinear balance equation for the prescribed vortex.

Table 7 indicates the permutations of physics options used in the 6 highlighted experiments that are shown in Figures 1-2 for the 3-km grid. In Table 7, note that the physics configuration used in Experiment 2 is very similar to that in the operational HWRF model. Also note that the SAS scheme is turned on for the 3 km grid in order for the model to produce the storm intensity comparable with the theoretical prediction.

Table 7. Physics permutations used in 6 of the 16 sensitivity experiments. MYJ refers to the Mellor-Yamada-Janjic boundary layer scheme, while GFS refers to the GFS boundary layer scheme. SAS refers to the Simplified Arakawa-Schubert convective parameterization scheme. Option 1 for the radiation scheme refers to both the Dudhia shortwave scheme and the RRTM longwave scheme. The option 5 microphysics scheme refers to the Ferrier scheme. D1 and D2 refers to 9 km and 3 km grids respectively.

Experiment #	Boundary Layer Scheme	Convective parameterization scheme (D1/D2)	Radiation Scheme	Microphysics scheme	Other
1	MYJ	SAS/SAS	1	5	NA
2	GFS	SAS/SAS	1	5	NA
3	GFS	SAS/SAS	1	5	momentum and thermal diffusivity increase in ABL
4	GFS	SAS/SAS	1	5	Momentum diffusivity increase in ABL only
5	GFS	SAS/SAS	1	5	momentum and thermal diffusivity decrease in ABL
6	GFS	SAS/SAS	1	5	Sea-spray modified surface drag

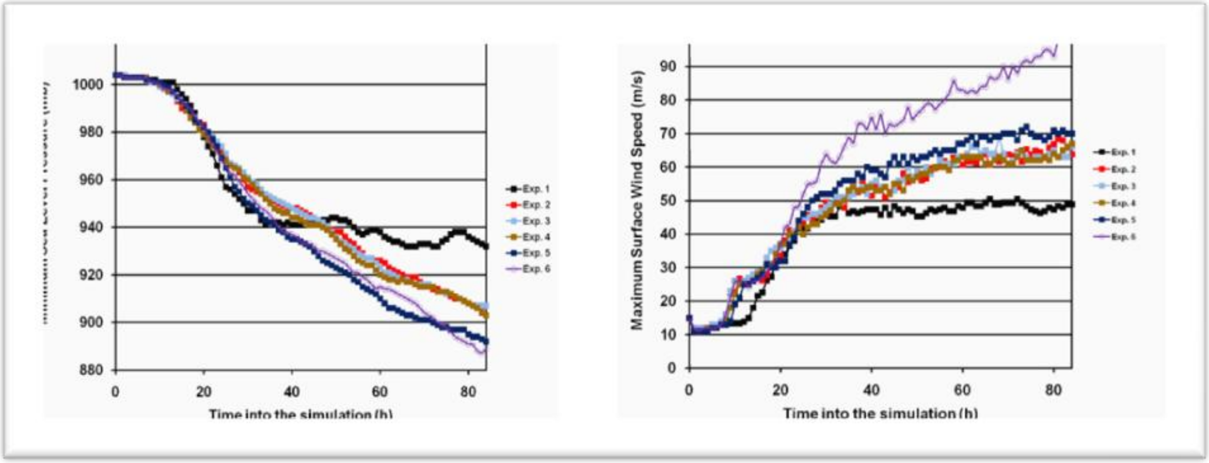


Figure 25: Left panel is the maximum surface wind speed in ms^{-1} and the right panel is the minimum sea level pressure in mb for 6 of the 16 sensitivity experiments highlighted in Table 7.

Figure 25 shows that while the maximum surface wind speed in Experiments 1, 2, 3, 4, and 5 levels off after 60 h into the forecast, the minimum sea-level pressure continues to decrease, confirming the well-known problem of the wind-pressure relationship in the HWRF model. This problem is significantly improved when a different surface drag parameterization associated with the sea spray physics is used (Experiment 6). This illustrates that the surface drag is a key parameter that controls the wind-pressure relationship in an intensifying storm. The structural differences in the time-mean azimuthally averaged secondary circulation shown in Figure 29 from Experiments 1, 2, 3, 4, and 5 indicate that in the HWRFX model, the ABL diffusivities above the surface layer control the size of the storm.

The use of AXBT data assimilation also showed promise.

EMC has conducted tests for AXBT data assimilation into RTOFS, and the preliminary results show an improvement in the simulation of the upper ocean structure. For example, Figure 26 shows model profiles with and without AXBT assimilation. The project proved to be successful in testing: 1) the NCO decoder and data format to Buffer ingest system; 2) the data pipeline set up in NCO in collaboration with EMC and HRD; and 3) the data assimilation algorithms in RTOFS.

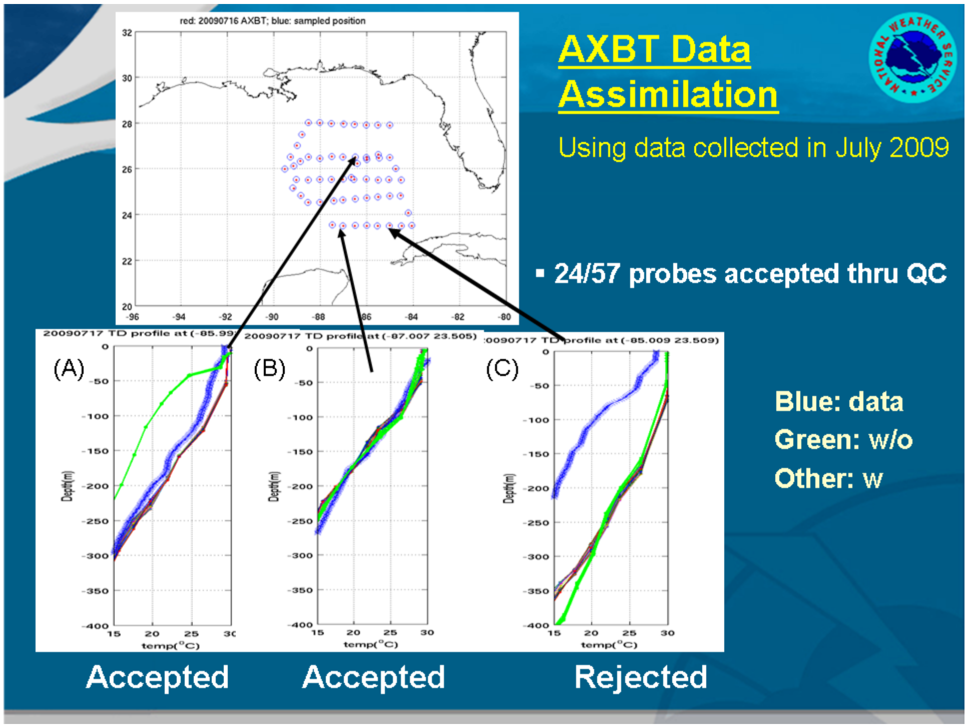


Figure 26. Sampling locations for the pre-storm AXBT survey on 17 July 2009 (top) and model profiles with (multi-color representing hourly profiles) and without (green) assimilation, superimposed on data assimilated (blue) (bottom): current quality control employed in RTOFS accepted, during 3 days assimilation, 24 out of 57 profiles for the 1st day of assimilation, followed by 18 for the 2nd day and 12 for the 3rd day.

9. *Lessons Learned from FY09 HFIP Activities*

Above we noted a number of encouraging results from the FY09 program. There were also a number of lessons learned that will help us guide the program forward. In this section we outline those lessons. In the next section we will discuss the challenges posed by these lessons and our strategy to meet the challenges. Table 8 outlines the lessons learned.

Table 8. Lessons Learned from FY09 HFIP Activities

• ***There is a major problem with initialization of regional models.***
• ***Both the regional and global models greatly over-predict genesis.***
• ***Simply increasing the resolution of the regional models alone does not lead to improvements in model guidance.***
• ***The research community is doing a poor job of conveying the value and use of ensemble information to the forecast community.***
• ***Model performance metrics for hurricanes must include more than just track and intensity metrics.***

There is a major problem with initialization of regional models.

This became particularly apparent during the FY09 season where many of the storms were initialized at a time when vertical shear was strongly affecting the hurricane. For most of the models we ran as part of the multi-model ensemble, the storms in the model continued to intensify even though the observed storm decayed often to a remnant low in the forecast period. Figure 27 is a typical illustration of a forecast of a highly sheared storm, in this case Erika on 2 September 2009. Note that most of the models, including the operational models, HWRF (blue line) and GFDL (light blue line) forecast the hurricane to grow to at least a Category 2 storm in 120 h. The storm was a remnant low at that time.

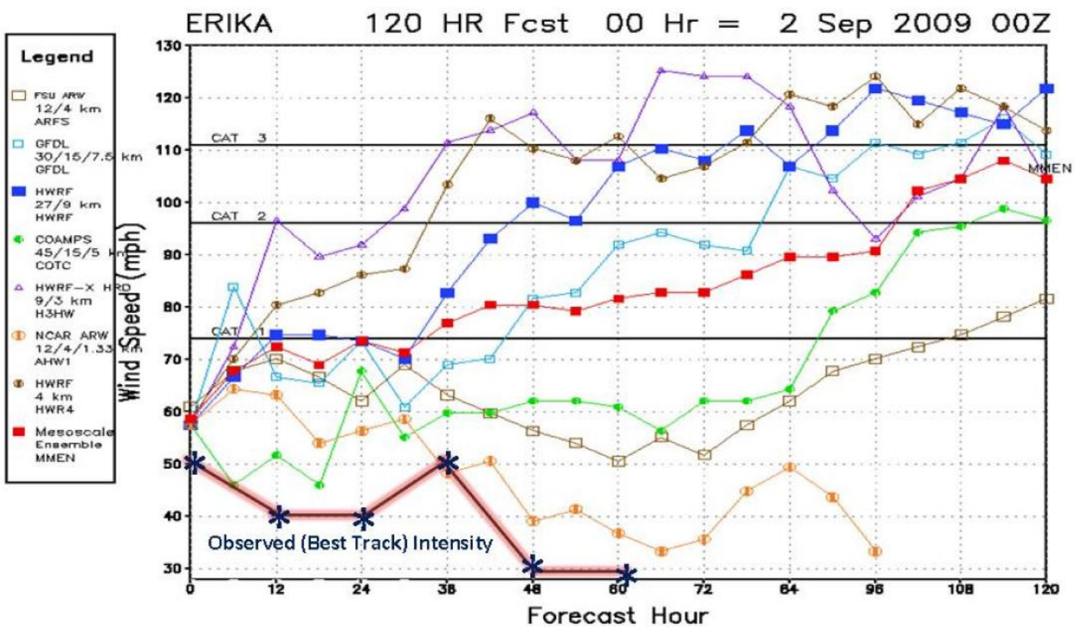


Figure 27. Forecasts of maximum wind from the various models used in the multi model ensemble color coded on the left for Hurricane Erika. Initialized 9 February 2009.

Plots like this one from the multi-model ensemble stimulated much discussion within the HFIP group, speculating on the cause of the spurious intensification. It was generally thought that one problem, especially with the sheared storms, was the initialization of the regional model. Most of the models above used a process to “move” the hurricane center and observed minimum pressure to the observed position either using a bogus vortex or moving the storm from the previous forecast to the observed position. Both methods generally result in a storm that is mostly vertical. In shear the storms start to tilt, sometimes strongly, and this is not handled well with current initialization systems. Of the one model that used a data assimilation system in the group shown above, the NCAR ARW model provided the best forecast in this case. In other cases it did not do as well.

For the case shown in Figure 27, most of the models had a very rapid intensification after start up. In other cases the models took some time to spin up. This is also an indication of inaccurate initial conditions.

In additional to problems with the initial conditions, it was also thought that the convection parameterization may also be a factor in the spurious intensification. As parameterized, most convection in the models is vertical extending through deep layers. This orientation in association with a vortex that is erroneously vertical rather than tilted can result in the intensification observed in the models. We will outline our plans to address this and other lessons learned in the section on “Challenges”.

Both the regional and global models greatly over-predict genesis.

The over-prediction of genesis in the FIM global model was noted in Figure 12 for the Eastern Pacific. In FY09 we did not have tracker software available to identify and keep track of all storms generated by the various models. Our sense of over-genesis is based on observing forecast maps. A tracker is under development and we will quantify the genesis allowing us to compute the probability of detection (POD) and the false alarm rate (FAR) of genesis. We suspect that the POD is near 1 meaning we detect most cases of observed genesis, however, we also suspect we have a very high FAR. This result is generally expected for all current global and regional models.

The problem with over-genesis is particularly important for forecasts that go out 5 to 7 days. Many storms last less than that and so in order to predict accurately we need to be certain that the storms that are being forecast to form are going to form. In addition the presence of other spurious storms relatively near the actual storms in the model (such as the string of storms in Figure 12) can affect the track and intensity of the actual storm in the model through vortex/vortex interaction.

The cause of the over-prediction of genesis is thought to be related to problems with the physics packages used in the models. There have been some preliminary indications that the shallow convection parameterization may play a role in the over-genesis but it could also be related to surface fluxes and PBL parameterization.

Simply increasing the resolution of the regional models alone does not lead to improvements in model guidance.

In FY09, HFIP funded a project conducted by the DTC to examine whether simply increasing resolution in regional models would lead to a significant increase in the skill of the models. This was known as the High Resolution Hurricane (HRH) test. In these tests, 5 groups around the country ran their models at two different resolutions (roughly 10 km and 5 km but that varied with model) over a total of 69 cases (a case is a single forecast for a single storm) from a variety of storms selected by forecasters at the NHC. Table 9 summarizes the report filed by the DTC for the HRH test.

Table 9. HFIP High-Resolution Hurricane Test

Evaluation by the Developmental Testbed Center (DTC) Ligia Bernardet, Louisa Nance et al:
• Runs for up to 69 cases at two or more horizontal grid spacings were submitted for evaluation of impact of resolution on track and intensity forecasts. • Increased resolution did not substantially improve forecasts for any model. • Modest improvement (a few lead times) were seen for HWRF-X (9 and 3 km) and AHW (13.5 and 1.5 km) in track and/or intensity. GFDL (9 and 6) showed no difference and COAMPS-TC (9 and 3) and UW-NM had some degraded tracks. • May need better physics and/or initialization to realize benefits of higher resolution. • Final Report is at: http://www.dtcenter.org/plots/hrh_test/HRH_Report_30Sept.pdf

The main conclusion was that simply increasing the regional model resolution does not significantly improve the skill of the forecasts. Many say that this should not have been a surprise but it clearly indicates that we need to address other aspects of the models before or in parallel with increasing resolution. This includes both improving the initialization, as noted above, and carefully tuning the physics packages for the target resolution.

The research community is doing a poor job of conveying the value and use of ensemble information to the forecast community.

The research community is generally convinced that use of ensembles has been proven to provide better forecasts than a single forecast by a model, known as a deterministic forecast. In fact a deterministic forecast can be regarded as a member of a virtual ensemble and the fan of tracks shown for a typical ensemble forecast in Figures 10, 11 and 21, for example, suggests that the deterministic model forecast could be anywhere within the fan. Using statistics from the ensemble there are ways to predict which of the members of the ensemble has the highest priority of being what will be observed.

Still the forecasters doubt the value of information in an ensemble beyond the consensus of the various deterministic models that are currently used. The two communities, research and operations, need to work together to develop the products from the ensemble that will be most useful in preparing forecasts. This will include developing products beyond the ensemble mean, for which groups of single-model ensembles may prove beneficial.

Model performance metrics for hurricanes must include more than just track and intensity metrics.

Comparing the storm intensity and structure among the 6 experiments shown in Figures 25 and 29 indicates that storms with different sizes and structural characteristics may share very similar intensities in terms of maximum surface wind speed and minimum sea-level pressure. Therefore we need metrics beyond the usual track and intensity metrics in order to diagnose reasons for poor model performance. These statistics should include measures of storm size and structure.

To further highlight the issue of dependence of storm size and structure on physics, Figure 28 compares Hovemoller diagrams of the axisymmetric mean tangential winds from an idealized case using the Experiment 2 physics configuration and a run in which the surface roughness formulation is replaced with the conventional Channock formulation. It is obvious that the run using the Experiment 2 physics configuration shows a stronger inner core than the counterpart produced in the run using the same physics configuration but with the conventional Channock formulation.

Of course the storm size and structure are also strongly dependent on grid resolution. Figure 30 compares the forecasted warm core development in an idealized run using the Experiment 2 physics configuration with that from a 9 km, single grid run using the same physics. As depicted by the temperature perturbation and θ_e (in K and contours indicated by black color) in Figure 30, moisture and heat are carried to the middle and upper troposphere by thermal plumes in both the 3 km and 9 km runs, but the overall structural characteristics of the two runs are significantly different. All these results strongly suggest the need to verify the model storm structure for real-time event forecasts. More importantly, they suggest that track and intensity evaluation metrics are not sufficient for fundamental improvement of the operational HWRF model.

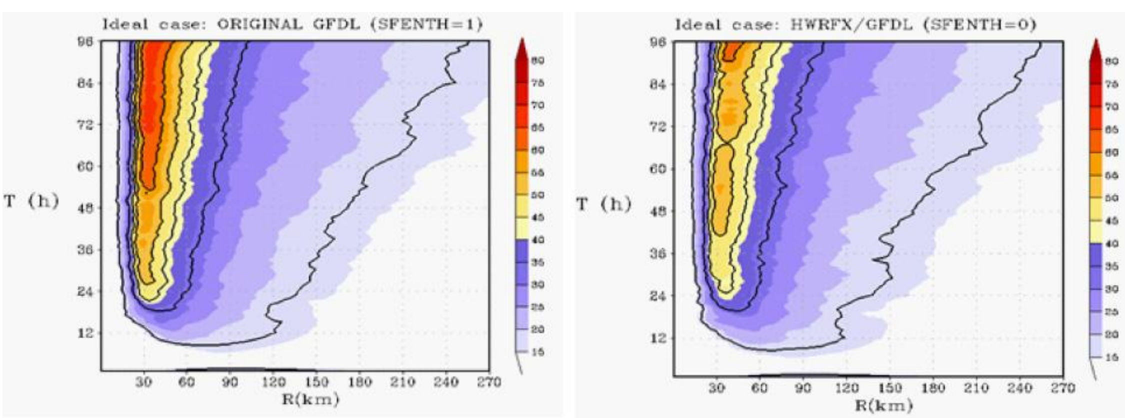
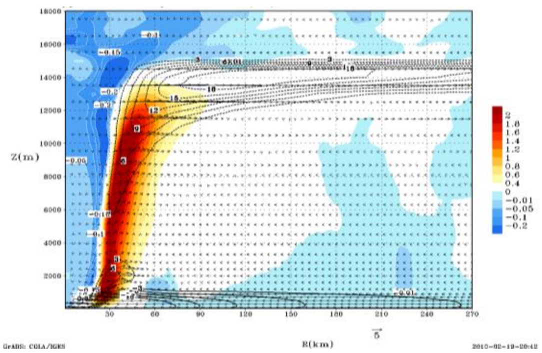
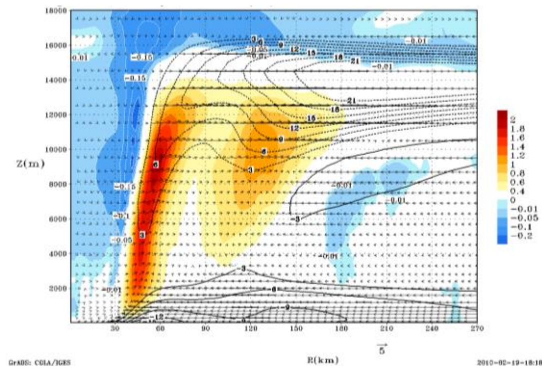


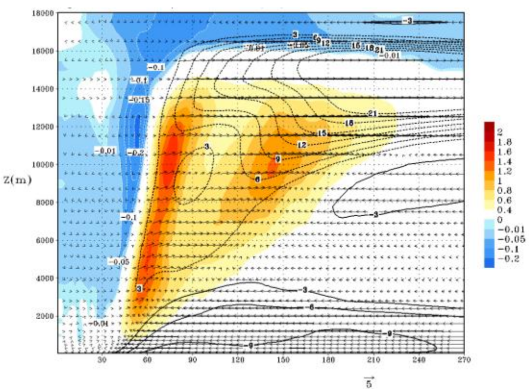
Figure 28: Hovemoller diagrams of the axisymmetric mean tangential winds from an idealized run using the Experiment 2 physics configuration (left panel) and a run (right panel) using the same physics configuration except that the surface roughness formulation is replaced with the conventional Channock formulation. The contour lines from the outer to the inner core region are respectively, 17.2 ms^{-1}, 33 ms^{-1}, 43 ms^{-1}, 50 ms^{-1} and 59 ms^{-1} representing the radius of gale forced winds and core of hurricane winds starting from minimal category of 1 to minimal category of 4.



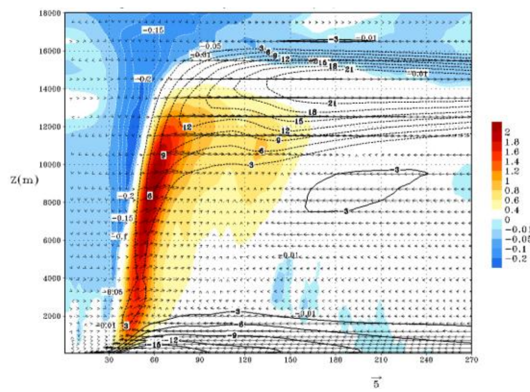
Experiment 1



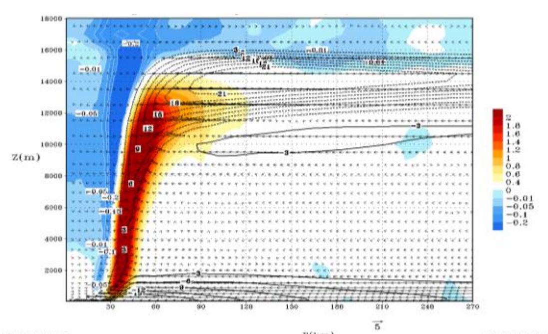
Experiment 2



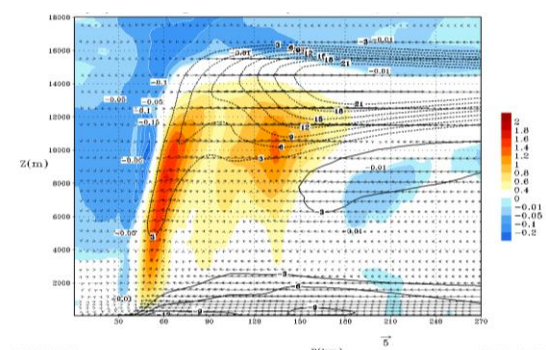
Experiment 3



Experiment 4



Experiment 5



Experiment 6

Figure 29: The 60-72 h mean azimuthally averaged secondary circulation (arrows), vertical velocity (color shaded contours) and radial winds (black contours mb for 6 of the 16 sensitivity experiments highlighted in Table 7.

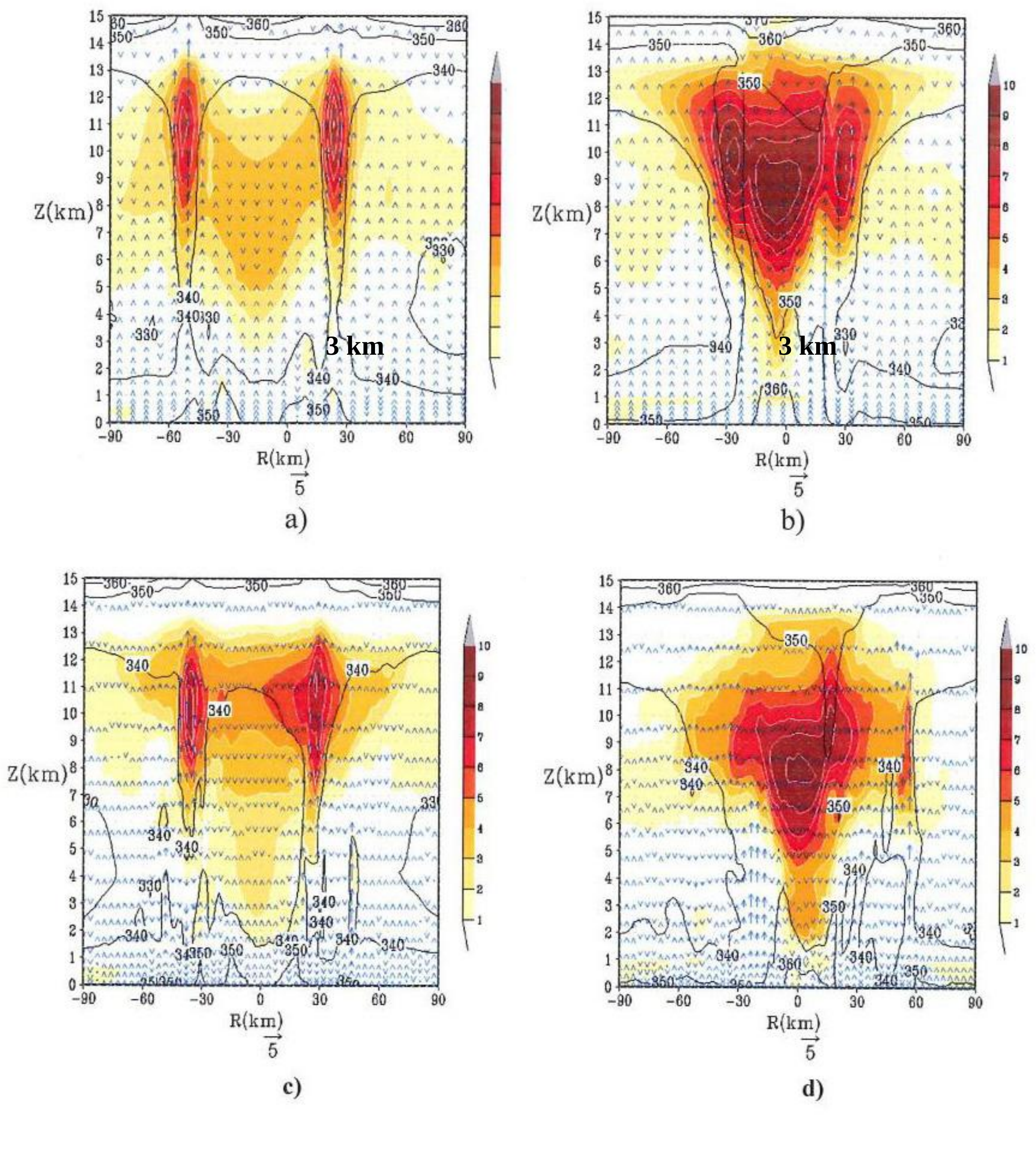


Figure 30. East-west cross-sections of instantaneous temperature perturbations (white contours and color shaded) across the eye-wall region to depict the warm core development in an idealized run using the Experiment 2 physics configuration and a 9 km, single grid run using the same physics. Panels (a) and (b) are for the 9 km run valid at 9 h and 24 h into the run, and panels (c) and (d) are for Experiment 2 valid at the same times for the 3 km run. The vector indicates the direction of the vertical motion and the black contours are equivalent potential temperature.

10. Challenges for HFIP beyond FY09

The lessons learned that have been outlined in the previous section suggest a number of challenges that must be addressed before HFIP can meet the goals it has set out for itself. These are discussed below.

A vast majority of model forecasts will be initialized for storms for which there is no aircraft data available.

Above we noted that a major problem identified in FY09 was initialization of the regional models. The most promising way to improve the initialization would be to use all available data in the vicinity of the hurricane, especially satellite data and use it within an advanced data assimilation system to provide a more accurate initialization of the vortex. For many years there has been a focus on using aircraft data. Unfortunately most of the times when we will want to initialize a regional model, there won't be any aircraft data available. This is true for any storm in the Western Pacific, most of the storms in the Eastern Pacific, and storms in the Middle or Eastern Atlantic. In fact most of the observations by aircraft will be in storms close to the Atlantic and Gulf coasts of the U.S. Therefore, a strong emphasis on utilizing satellite data is needed because it will be the primary source of information on the storm inner core for the non-aircraft cases, and a major contributor to the analysis of the storm environment over the ocean in all cases.

- We are engaging the JCSDA, AOML, CIRA to develop use of new satellite data sources for hurricane initialization.

- NRL and EMC are considering alternatives for simplified hurricane vortex initialization.

For storms with aircraft data (particularly radar data) more use of that data is crucial for hurricane initialization especially for hurricanes near the coast.

- A number of ongoing activities at EMC, AOML, PSU and other organizations are being funded by HFIP

Development and tuning of physics packages for hurricane models at high resolution is critical.

This was noted above related to the poor performance of the regional models in the sheared storms of 2009. Development, testing and evaluation of physics packages for hurricane models at high resolution is critical.

- GFS physics seems to work well down to 10-15 km but the operational version greatly over-predicted genesis in 2009
  - Next GFS package (T574+physics) looks to be a significant improvement.
  - ESRL is experimenting with other physics packages for their FIM.
- We will continue to work with operational physics packages to
  - Develop appropriate sensitivity to vertical shear

  - Apply improved understanding from carefully designed high resolution regional experiments
  - Tune them at higher resolution and include a research effort to better understand the physical processes in hurricanes.
- Plan a workshop to define a community wide effort on improving the operational physics packages—emphasis on 4-10 km resolutions.

Advanced DA systems in both regional and global models appear to lead to significant improvements.

- HFIP, EMC, AOML and ESRL will focus on developing a hybrid technique that combines advantages of 4DVAR and EnKF
- Can the transition to operations be accelerated?

For a limited sample, high resolution global and regional ensemble systems are showing promise but require further testing and evaluation

- Main challenge will be to find adequate computing for running these operationally.
- HFIP is funding several groups to further develop ensemble systems including improving initialization and model diversity.

We need to develop better products to convey ensemble information to forecasters.

- HFIP will make this a priority this year.
- A workshop is planned to define the community effort to improve ensemble products.

We need to fully engage the whole hurricane science community in improving the operational HWRF.

- The HFIP program plan is tied closely to the EMC set of priorities for HWRF.
- This will increase the resources that can be focused on improving the operational model.
- HWRF has a focus of the expanding DTC capability.
- Research and operations need to use the same HWRF code base in order to facilitate transition of new developments to NCEP. For that, additional tests on the community HWRF configuration need to be conducted and, as soon as the community HWRF model has demonstrated skill, it needs to be implemented at NCEP.

We need to emphasize coordination between the HFIP modeling, observations and evaluation components to determine observational requirements for the improvement of model physics packages.

Figure 2 shows that in the experiment where the physics configuration is similar to that used in the operational HWRF (Experiment 2), the depth of the ABL inflow is greater than in the experiment where the MYJ scheme is used (Experiment 1). Unfortunately, there are not sufficient observations to help judge whether or not this greater ABL inflow depth is realistic. We need to fully coordinate the HFIP modeling, observations and evaluation programs to determine observational requirements for the improvement of the physics package in hurricane models.

11. HFIP Progress Toward its Performance Goals.

Tables 2 and 3 define the baseline performance of model guidance that HFIP will use to measure progress toward its goals of increasing the accuracy of model guidance by 20% in 5 years starting in 2009. In Figure 31 we show the performance of the GFS system run as an ensemble at T382, the current resolution of the operational GFS deterministic model and using the EnKF data assimilation system instead of the operational GSI system. The EnKF system was described earlier and other results shown from it.

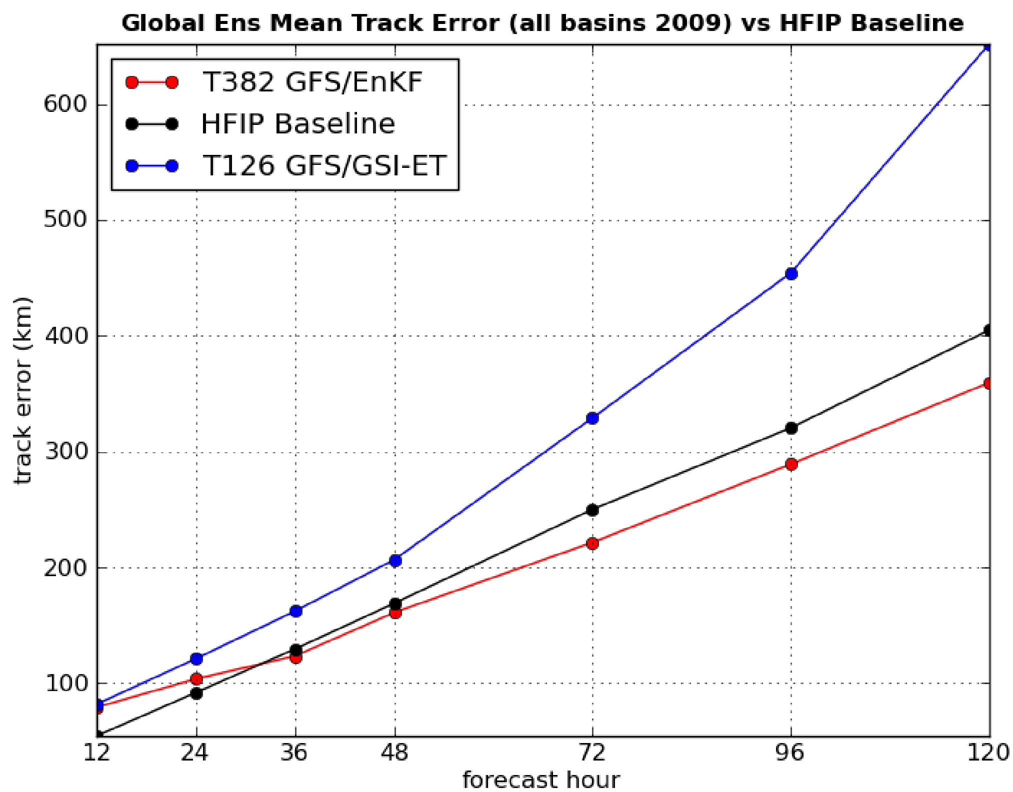
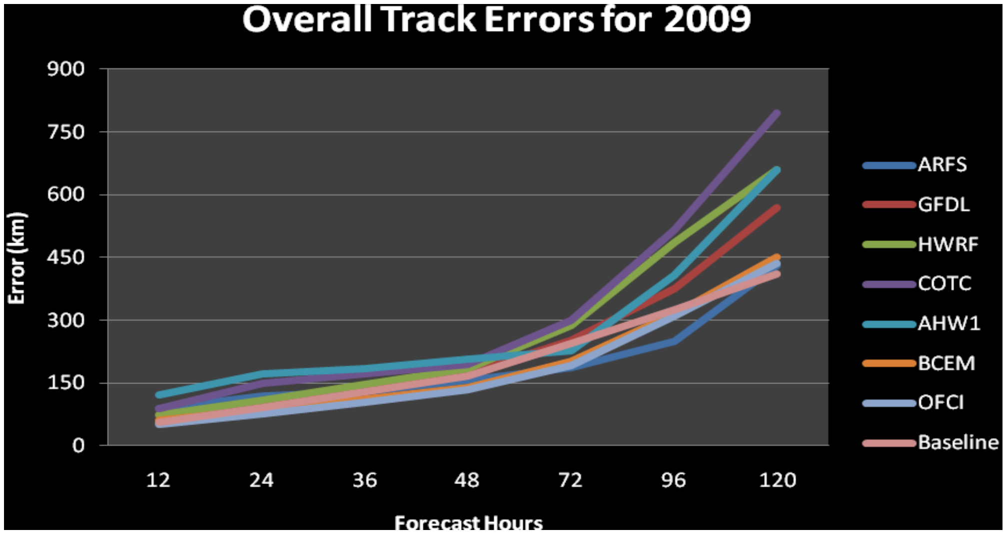


Figure 31. Forecast hurricane track errors for the GFS deterministic operational model (T382) run as a 20 member ensemble (ensemble mean shown), red, plotted against the HFIP baseline for track that was defined in Table 2, black. The blue line is the error of the ensemble mean from the GFS T126 operational ensemble. The data include all storms for all basins in 2009.



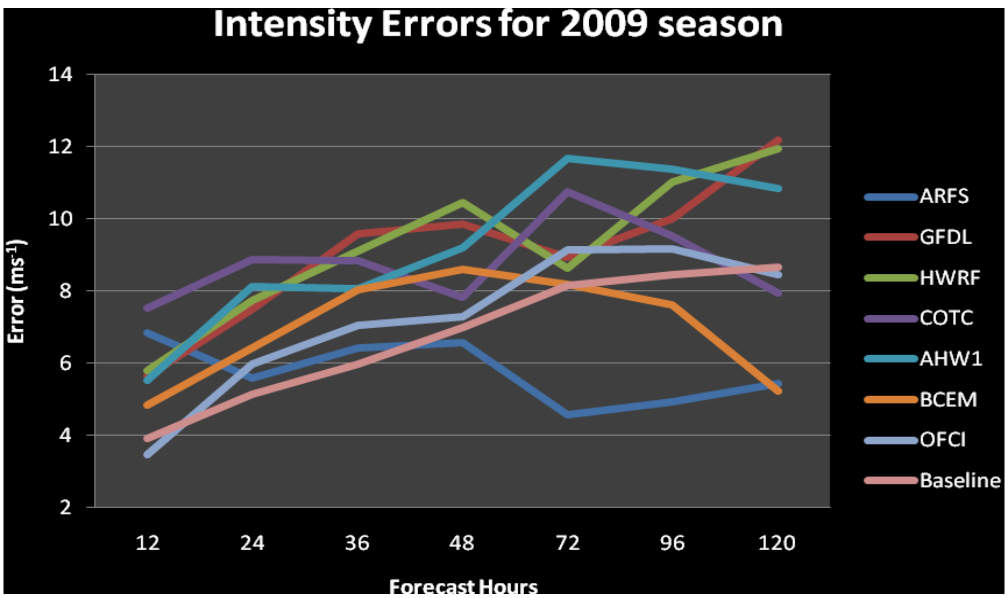


Figure 32. Same as Figure 24 but showing line graphs of the bias ensemble corrected ensemble mean (BCEM) track (upper panel) and intensity (lower panel) errors for the multi-model ensemble. Note that the errors for the individual models are not bias corrected. For this figure, only the 5 best (defined by overall error performance) models are included in the ensemble mean and are shown on the graphs.

The blue line in Figure 31 shows the performance of the operational ensemble mean (GFS at T126 using the GSI data assimilation system. Note that the GFS at the higher resolution and using the advanced data assimilation system (EnKF) is better than the baseline beyond 3 days by about 10%. Thus using the current operational system we are able to reach about half of the stated HFIP goal for track at the longer lead times in one year. We don't show a similar comparison with the intensity baseline for the global models since at 30 km the global models under-predict intensity.

The results for the regional models were summarized in Figure 24 and again in Figure 32. For the purposes of the discussion in this section just focus the data for the ensemble means (raw mean for Figure 24, bias corrected mean for 32) and on the baseline. In Figure 24 those are the last two columns in each forecast hour group in the figure. The second column from the right hand side of each group in Figure 24 is the ensemble mean from the multi-model ensemble and the right column is the baseline defined in Tables 1 and 2.

Figures 24 and 32 indicate that we have a ways to go with the regional models to reach the HFIP goals intensity and track. At the shorter lead times the bias corrected ensemble mean track error is comparable to the baseline but does not improve upon it. At the longer lead times it is worse. Figure 31 indicates that the reverse is true for the global models—they are better at longer lead times and worse at the shorter times. This provides some justification for earlier statements in this report that the focus of the regional models will be on the shorter lead times and the global models on the longer lead times.

Intensity errors are shown in the lower panel of Figures 24 and 32 and again just focus on the two rightmost columns of each forecast hour lead time. At all forecast lead times the ensemble mean errors are worse that the baseline so this indicates that we have made no progress in meeting the HFIP intensity goals this year. Note that the ARFS model was run on only half the number of cases as the other models in the ensemble and it is believed therefore that the ARFS model. is biased toward better forecasts (e.g., applied to easier to forecast storms). However, we point out that it gave forecast errors that were considerably lower than the baseline.

12. List of HFIP Supported Publications and Presentations.

This list is quite long and is available at:

http://www.hfip.org/documents/presentations_publications.php

13. Appendix A: List of HFIP Teams

Bold face type denotes team leads.

1. Global Model/Physics Development Team

Stan Benjamin (ESRL)
John Brown (ESRL)
Kevin Yeh (AOML)
Melinda Peng (NRL)
Shian-Jiann Lin (GFDL)
Steve Lord (EMC)
Jim Ridout (NRL)
Jian-Wen Bao (ESRL)
John M. Ward (EMC)

2. Regional Model/Physics Development Team

Morris Bender (GFDL)
Young Kwon (EMC)
Steve Lord (EMC)
Sundararaman.G.Gopalakrishnan (AOML)
Rich Hodur (NRL)
Shaowu Bao (ESRL)
Isaac Ginis (URI)
Jim Doyle (NRL)
Robert Rogers (AOML)
Jian-Wen Bao (ESRL)
Bob Tuleya (ODU)
Ligia Bernardet (ESRL)
Chris Davis (NESL)

3. Ensemble Systems Development Team

Zoltan Toth (ESRL)
Carolyn Reynolds (NRL)
Sim Aberson (HRD)
Tom Hamill (ESRL)
Jeff Whitaker (ESRL)
Fuqing Zhang (PSU)
Yuejian Zhu (EMC)
Jun Du (EMC)
Mike Brennan (NHC)
Mrinal K Biswas (FSU)
T. Krishnamurti (FSU)
Teddy Holt (NRL)

4. Data Assimilation/Vortex Initialization Team)

Jeff Whitaker (ESRL)
Bill Lapenta (EMC)
Steve Lord (EMC)
Jim Doyle (NRL)
Altug Aksoy (AOML)
Yuanfu Xie (ESRL)
Milija Zupanski (CIRA)
Liyan Liu (EMC)
Fuqing Zhang (PSU)
Tomislava Vukicevic (AOML)

5. Verification Team

Tim Marchok (GFDL)
Barb Brown (RAL)
Jim Goerss (NRL)
Mark DeMaria (NESDIS)
Robert Rogers (AOML)
James Franklin (NHC)
Vijay Tallapragada (EMC)
Michael Fiorino (ESRL)
Hao Jin (NRL)
Ligia Bernardet (ESRL)
Louisa Nance (RAL)
Tony Eckel (OST)

6. Applications Development/Diagnostics Team

Mark DeMaria (NESDIS/STAR)
Ed Rappaport (NHC)
Vijay Tallapragada (EMC)
Yi Jin (NRL)
Buck Sampson (NRL)
Robert Rogers (HRD)
Barb Brown (RAL)
Richard Pasch (NHC)
Michael Fiorino (ESRL)
Louisa Nance (RAL)
Bob Tuleya (Old Dominion Univ)
Jim Hansen (NRL)
John Knaff (NESDIS)
Sundararaman.G.Gopalakrishnan (AOML)
T. Krishnamurti (FSU)
Tony Eckel (OST)

7. Hurricane Observations Team

Sim Aberson (AOML)
Nick Shay (RSMAS)
Jack Beven (NHC)
Naomi Surgi (EMC)
Mark DeMaria (NESDIS)
Chris Fairall (ESRL)
Isaac Ginis (URI)
Peter Black (NRL)
Paul Chang (NESDIS)
Jim McFadden (AOC)
Tara Jensen (RAL)
Thiago Quirino (AOML)
Bob Atlas (AOML)
Yuanfu Xei (ESRL)
George Halliwell (AOML)

8. Coupled Ocean/Wave model Team

Hendrik Tolman (EMC)
George Halliwell (AOML)
Isaac Ginis (URI)
Shaowu Bao (ESRL)
Sue Chen (NRL)
Jian-Wen Bao **(**ESRL**)**
Chris Fairall (ESRL)
Nick Shay (RSMAS)
Daniel Melendez (OAR)

14. Appendix B: Organization Acronyms

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