



Hurricane Forecast Improvement Project (HFIP)

Some Results and Preliminary Project Plan



Bob Gall

HFIP Development Manager

November 9, 2009



FY09 Hurricane Season Demo system



HFIP 2009

Hurricane Season Activities



Global Models

- **30 km EnKF Data assimilation System run during August 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) - started August 15.**
- **Global Ensembles:**
 - **30 km FIM (initialized with EnKF) 20 members.**
 - **~45 km GFS (Initialized with GSI-3DVAR) 5 members.**
 - **55 km NOGAPS (3DVAR then 4DVAR), 9 members**



HFIP 2009

Hurricane Season Activities

Regional models:

Multi Model Ensemble (various initialization schemes):

- (run for all storms—not all models present for all run times)
 - HWRF 9km
 - HWRF 4km
 - GFDL 7.5km
 - HWRF-x 3km
 - WRF/ARW/NCAR 1.3km
 - WRF/ARW/FSU 4km
 - TC-COAMPS 5km
- Single model Ensemble (run for most storms) – Separate allocation from TACC
 - WRF/ARW/PSU 4.5 km 30 members
 - a) Initialized with an EnKF system
 - b) Initialized with P3 radar data when available



2009 Hurricane Season Operational and Demonstration System



Operational Model Suite	Demonstration System
<ul style="list-style-type: none">• Global Model: GFS T382 (~35km), 64 layers• Global Ensemble: GFS T126 (100km), 28 layers, 20 members (40 with NAEFS); FIM will be tested at 60km when ported• Regional Model: No upgrade for 2009 season. HWRF work on physics package upgrade and coupled ocean continues. HWRF 9km, 69 levels, coupled ocean (POM)• Regional Ensemble: 35-45km, 28-51 layers, multi-member ensemble (ARW, NMM, RSM, ETA) 21 members (non-hurricane model fixed domain)• NHC Consensus Ensemble: Track (GFSI, EGRI, NGPI, GFMI, HWFI); Intensity (DHSP, LGEM, GHMI, HWFI, GFNI)• Data assimilation, Global: GSI-3DVAR• Data assimilation regional: GSI-3DVAR	<ul style="list-style-type: none">• Global Model: FIM 10km, 15km, 30km, 64 layers• Global Ensemble: FIM 30km, 20 members; GFS T574 (~27km), 5 members; NOGAPS 55km 9 mem.• Regional Model Ensemble: ARW, 4.5km, 30 members• Multi-Model Regional Ensemble: 1.3km - 9 km, 36-43 layers, multi-member ensemble (ARW (two versions), MMM, HWRF, HWRF-X, COAMPS-TC, GFDL, HWRF 4km)• Data Assimilation Global: GSI-3DVAR, EnKF, 4DVAR• Data Assimilation Regional: EnsKF (with Tail Doppler data when available)



Preliminary Results



First Look Observations FY 2009 Hurricane Season



Global Models

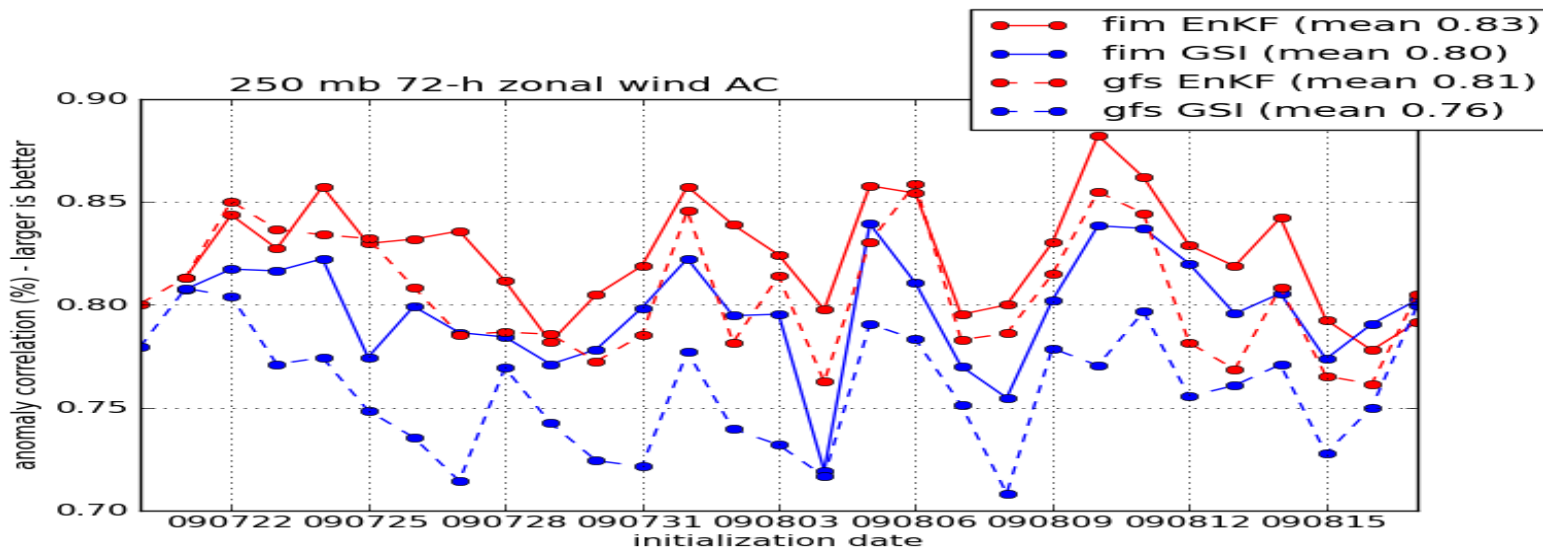
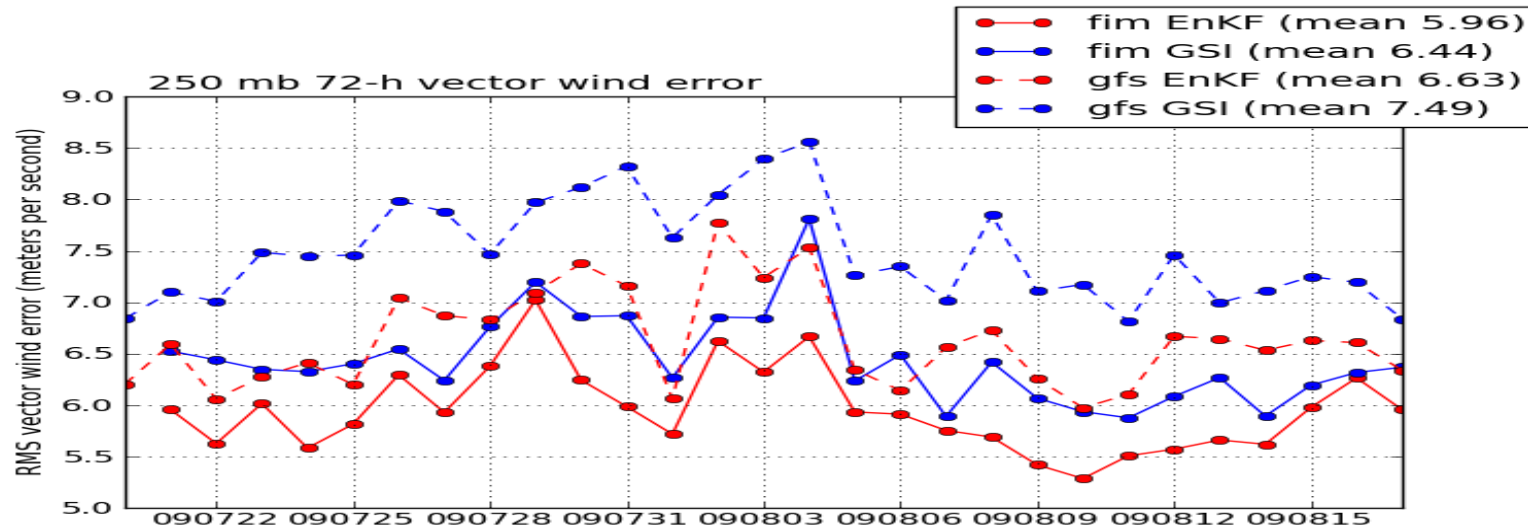
The biggest impact on skill has come from using the EnKF approach as opposed to GSI-3DVAR:

- This is true for both GFS and FIM. (See figure 1 and 2.) (The runs this summer used GFS as the model to use with EnKF.)
- Definite impact of increasing resolution on intensity forecast skill - See figure 3a.
- Significant impact on intensity bias seen with higher resolution – See figure 4.

Global Model Ensembles

- Definite positive impact on both track and intensity using ensemble mean (see figure 4).
- NOAA Ensemble beat the UKMet ensemble (see figure 5) and was comparable to the UKMET ensemble (see figure 6).
- Joint FIM- GFS ensembles not yet evaluated

Figure 1



TR Wind Forecast Verification

Figure 2

Deterministic 15-km FIM GSI vs. EnKF Track Errors
20090715 to 20091004

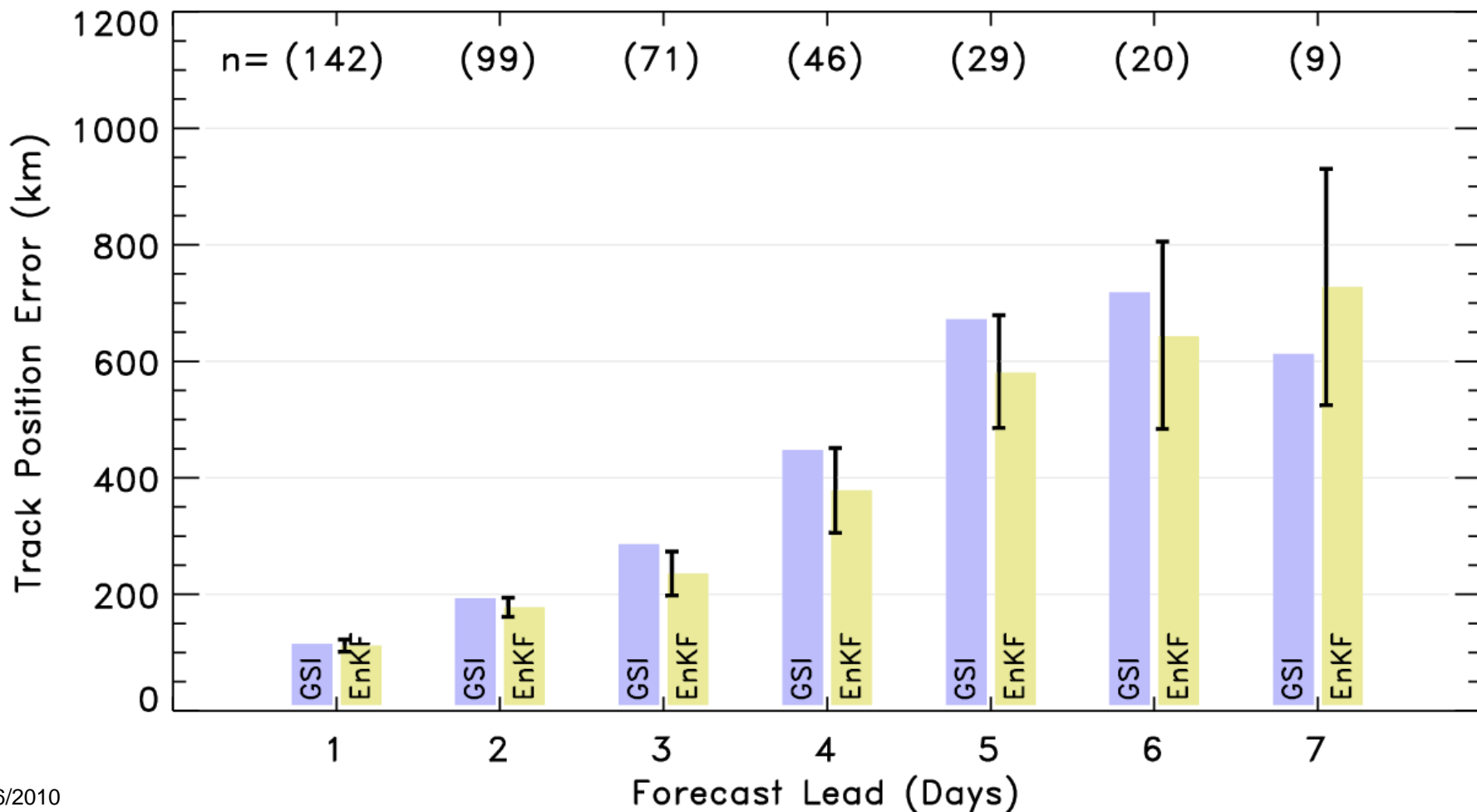
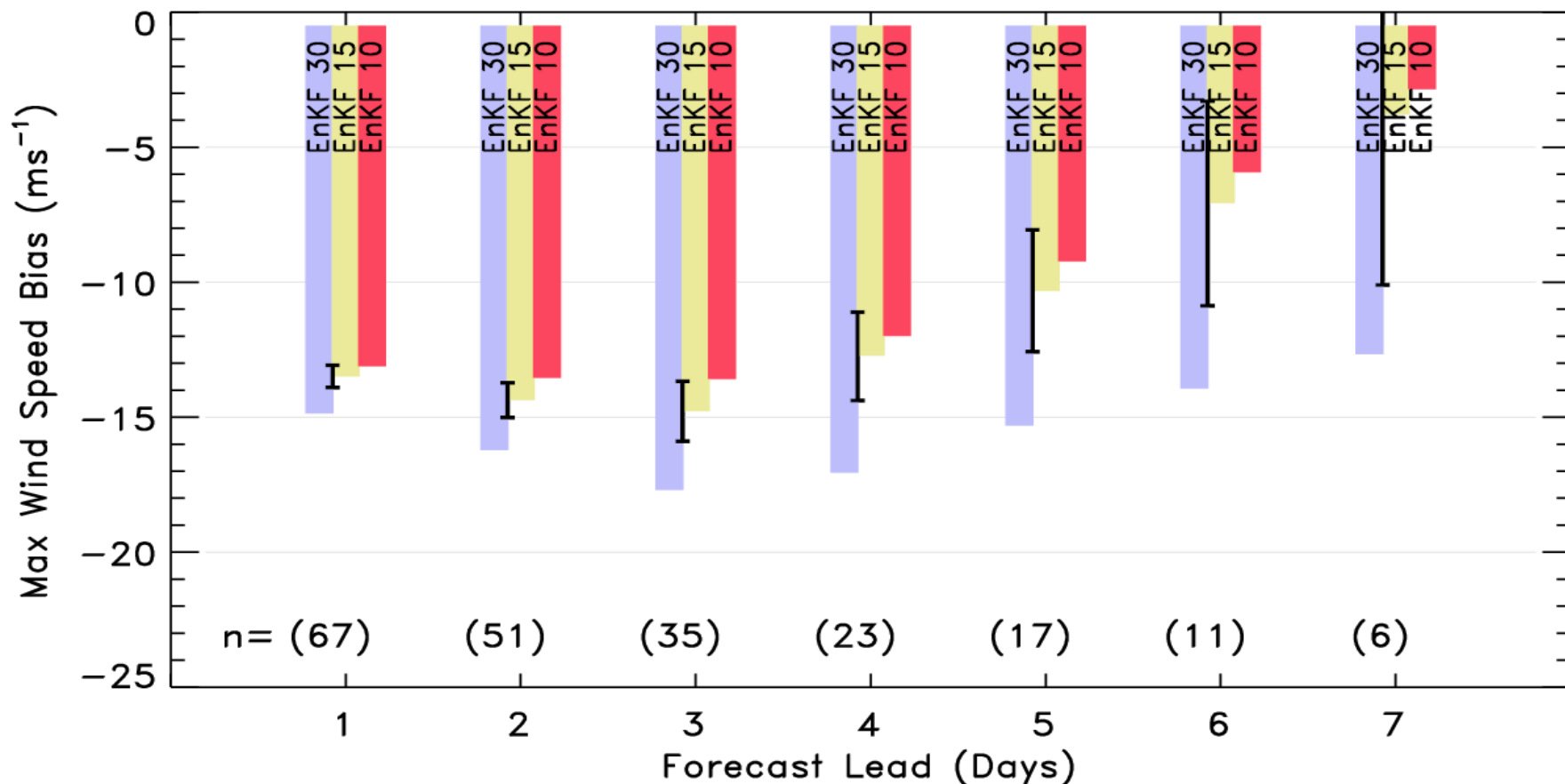


Figure 3a

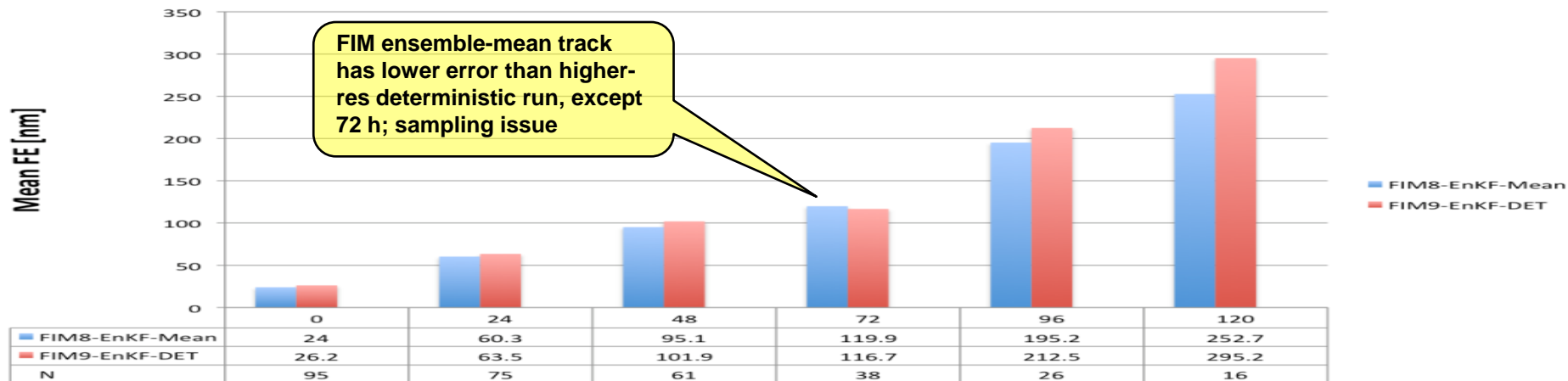
Deterministic 30 vs. 15 km EnKF Max Wind Speed Bias
20090715 to 20091004



- Strong signal that GFS-EnKF FIM9 (15 km) deterministic forecast are superior to GFS-GSI-3DVAR FIM9
- GFS-EnKF uses same model and observations as GFS GSI-3DVAR, except for TC obs based on TCvitals.
- Bug in EnKF assimilated only the central pressure from the TC obs.

Figure 4

FIM G8 (30 km) Ensemble Mean (21 members) v FIM G9 (15km) EnKF (Deterministic) - Mean Track Forecast Error [nm] 20090930 -- W,C,E,L basins



FIM G9 (15 km) GSI v EnKF v FIM9.5(10km) EnKF v GFS(~45km) GSI Intensity Bias [kt] 20090930 W,C,E,L basins

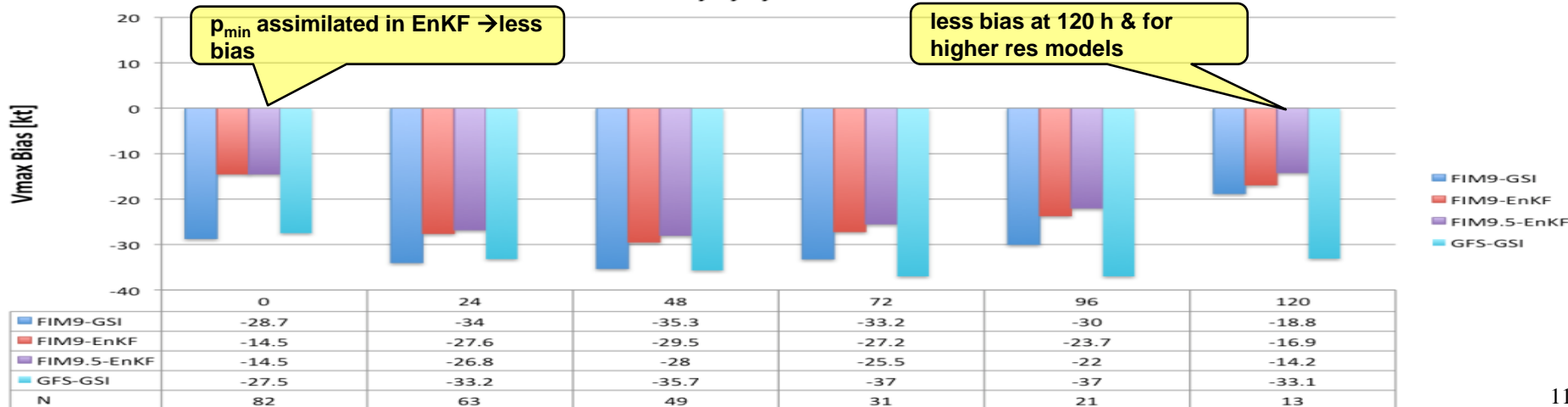
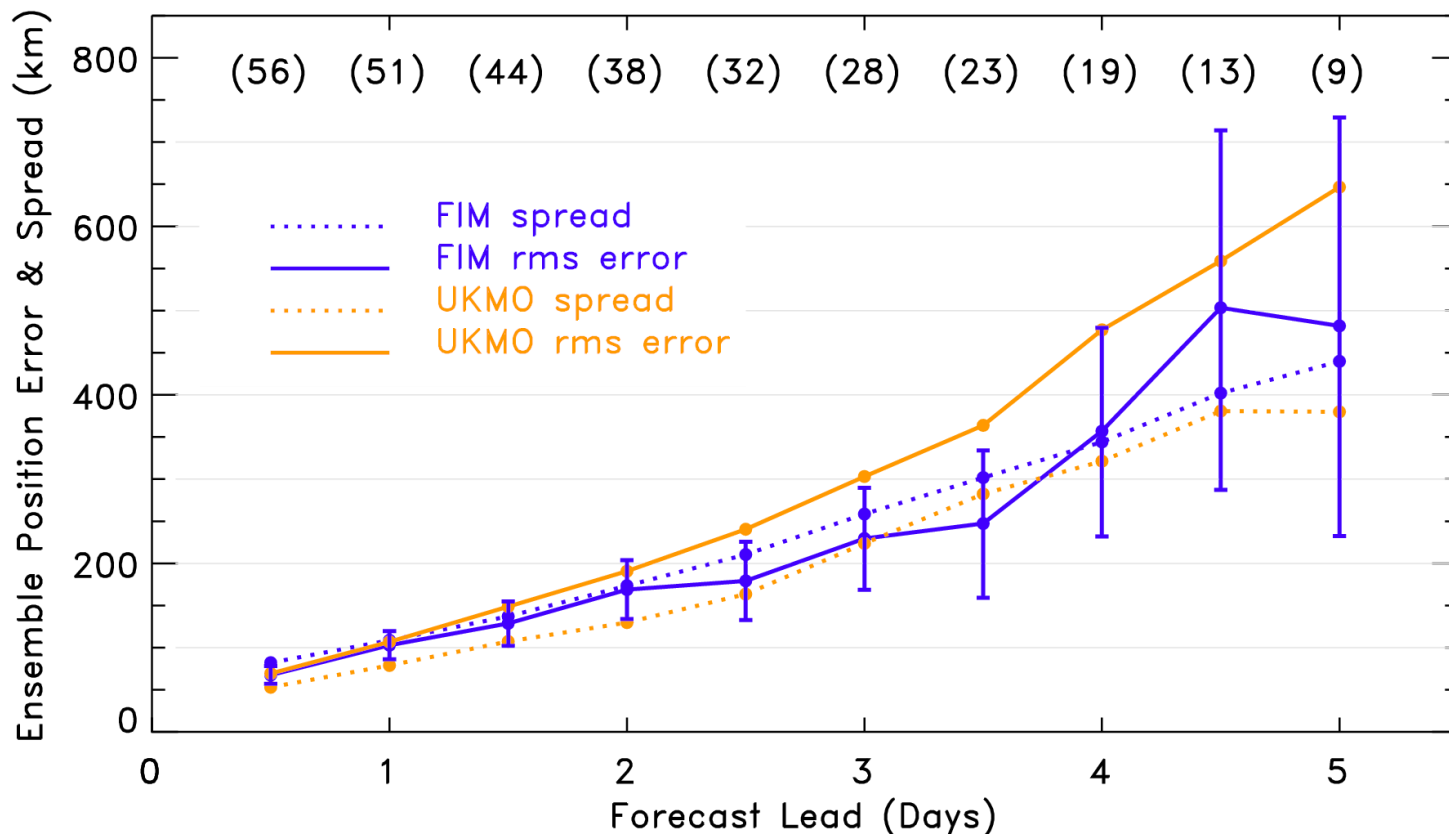


Figure 5

FIM G8/EnKF vs. UK Met Office

UK Met Office EPS vs. FIM G8/EnKF Track Error & Spread
20090715 to 20091009

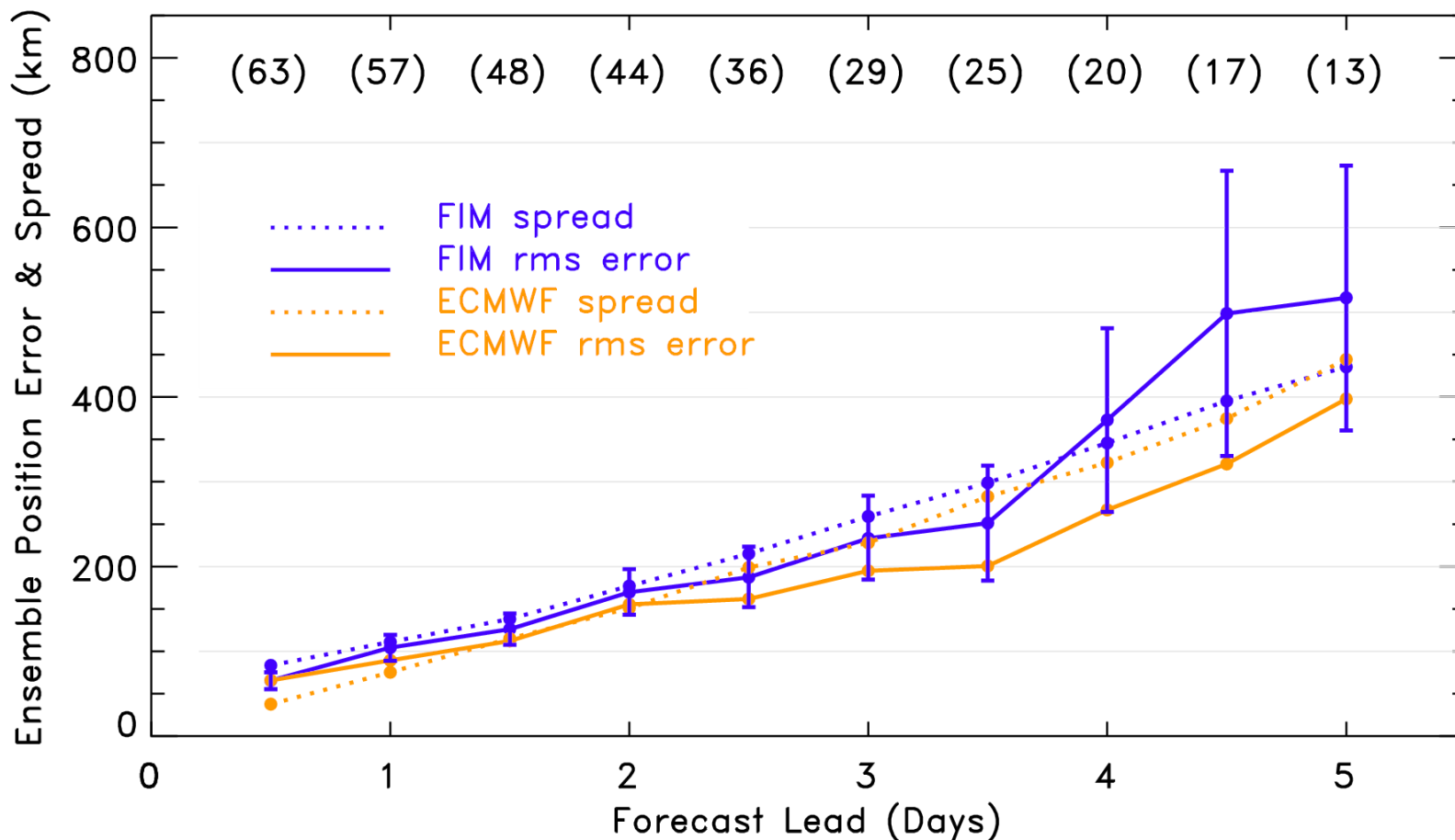


Error bars are 5th and 95th percentiles from paired block bootstrap.
Numbers in parentheses are the sample size at this lead.

Figure 6

FIM G8/EnKF vs. ECMWF

ECMWF EPS vs. FIM G8/EnKF Track Error & Spread
20090715 to 20091009





First Look Observations FY 2009 Hurricane Season



Regional Models

- In general regional model performance was poor—all over the map (figure 7).
- High resolution (10km to 1 km) – alone - did not produce desired impact (figure 8).

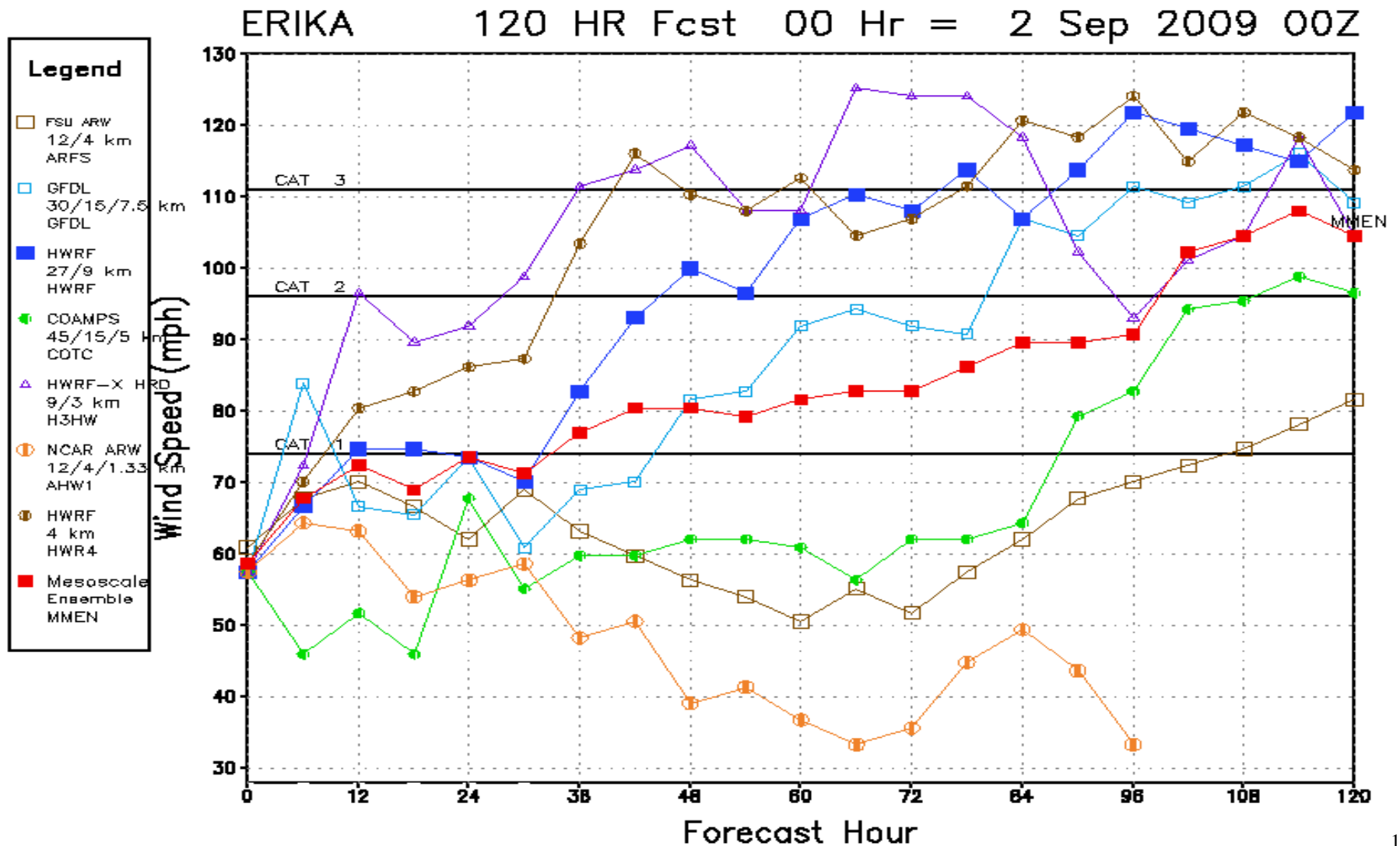
Regional Model Ensembles

- Single model ensemble captured the spread shown by figure 7 and perhaps there is improved skill from the ensemble mean.
- But we must address reasons for very wide spread in intensity forecasts.

A lesson from Regional models in FY09

- Models do very poorly in highly sheared storms.
 - Appears related to model initialization and convection parameterization.

Figure 7



Another Example Hurricane Rick

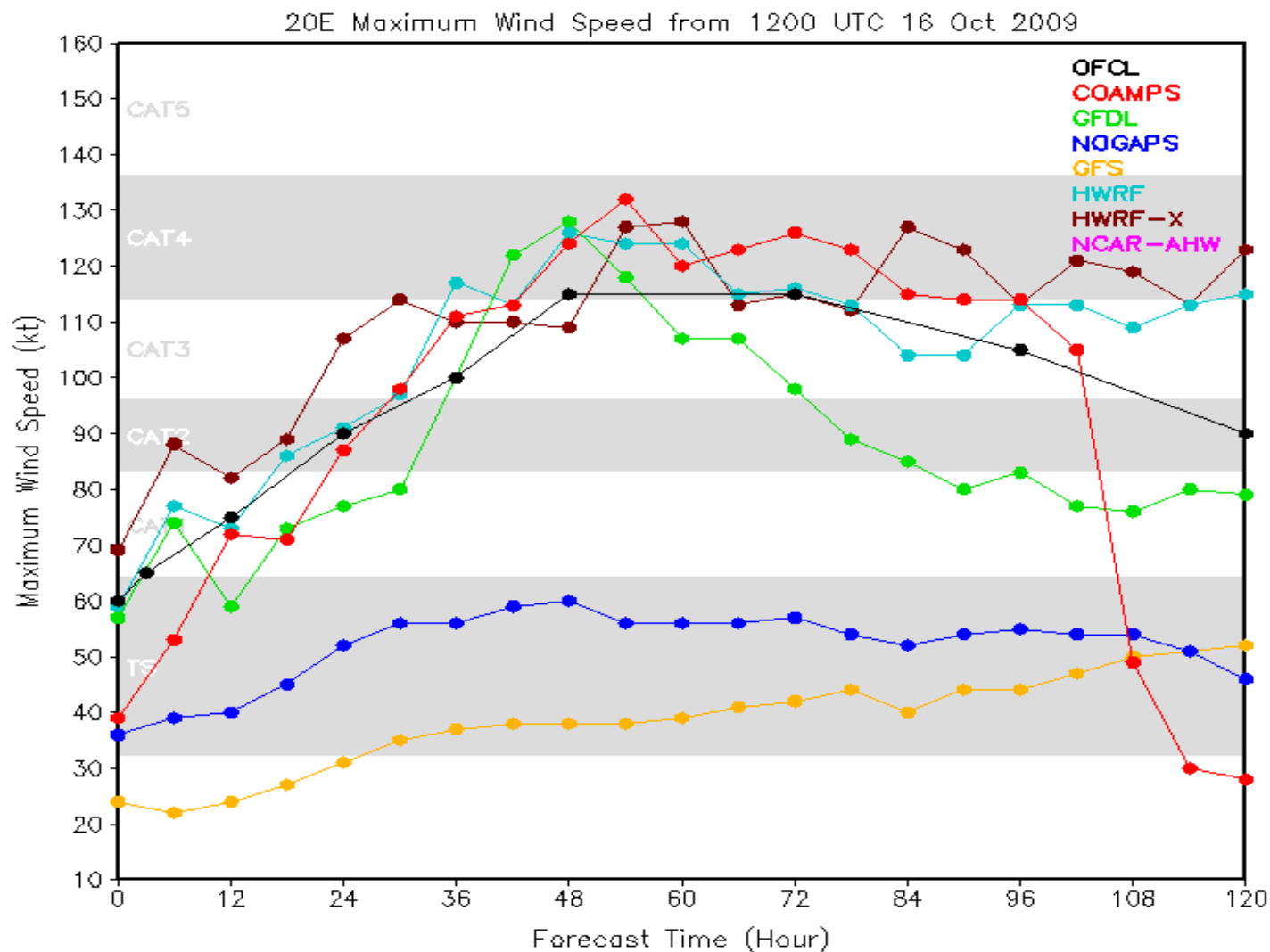




Figure 8

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

Figure 9

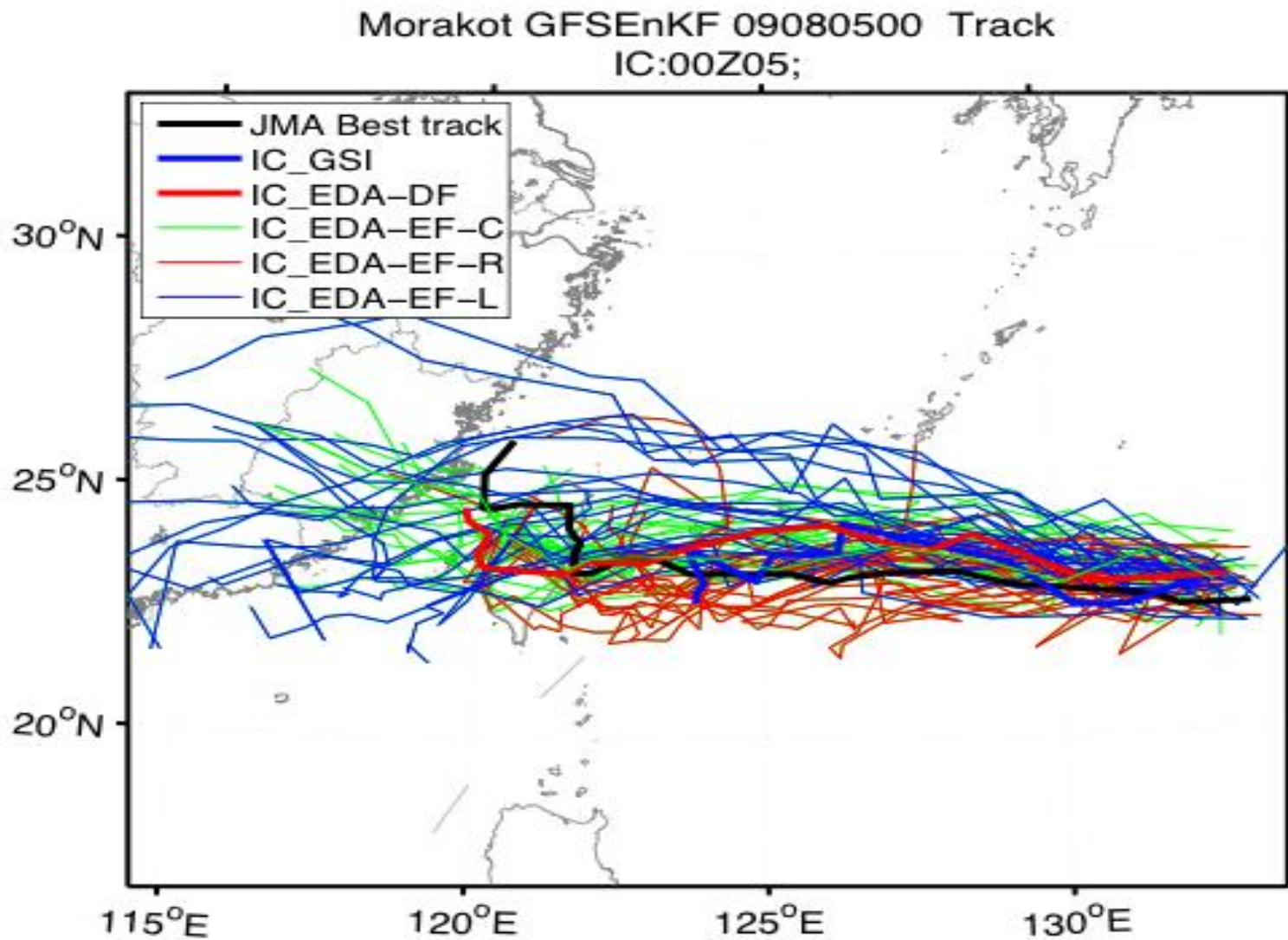


Figure 9b

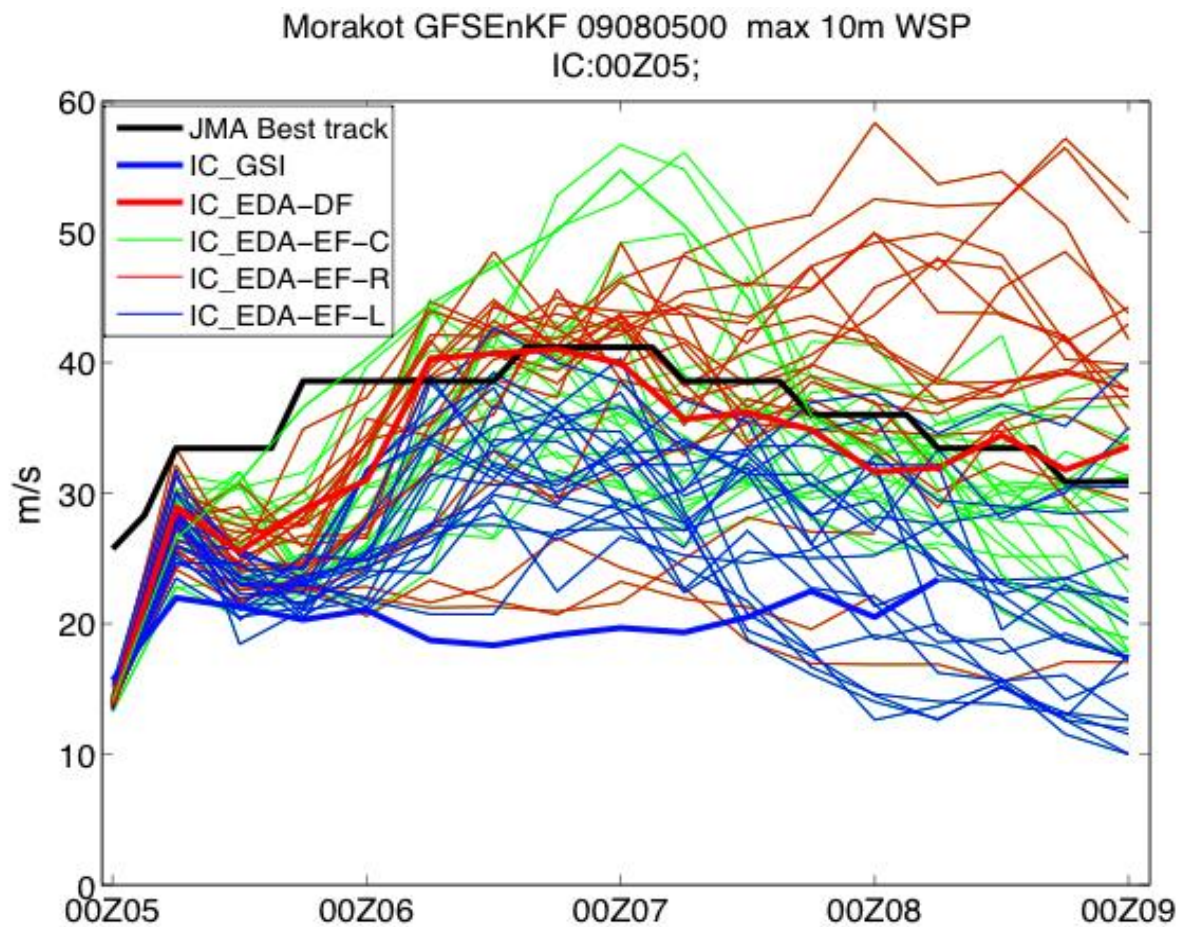
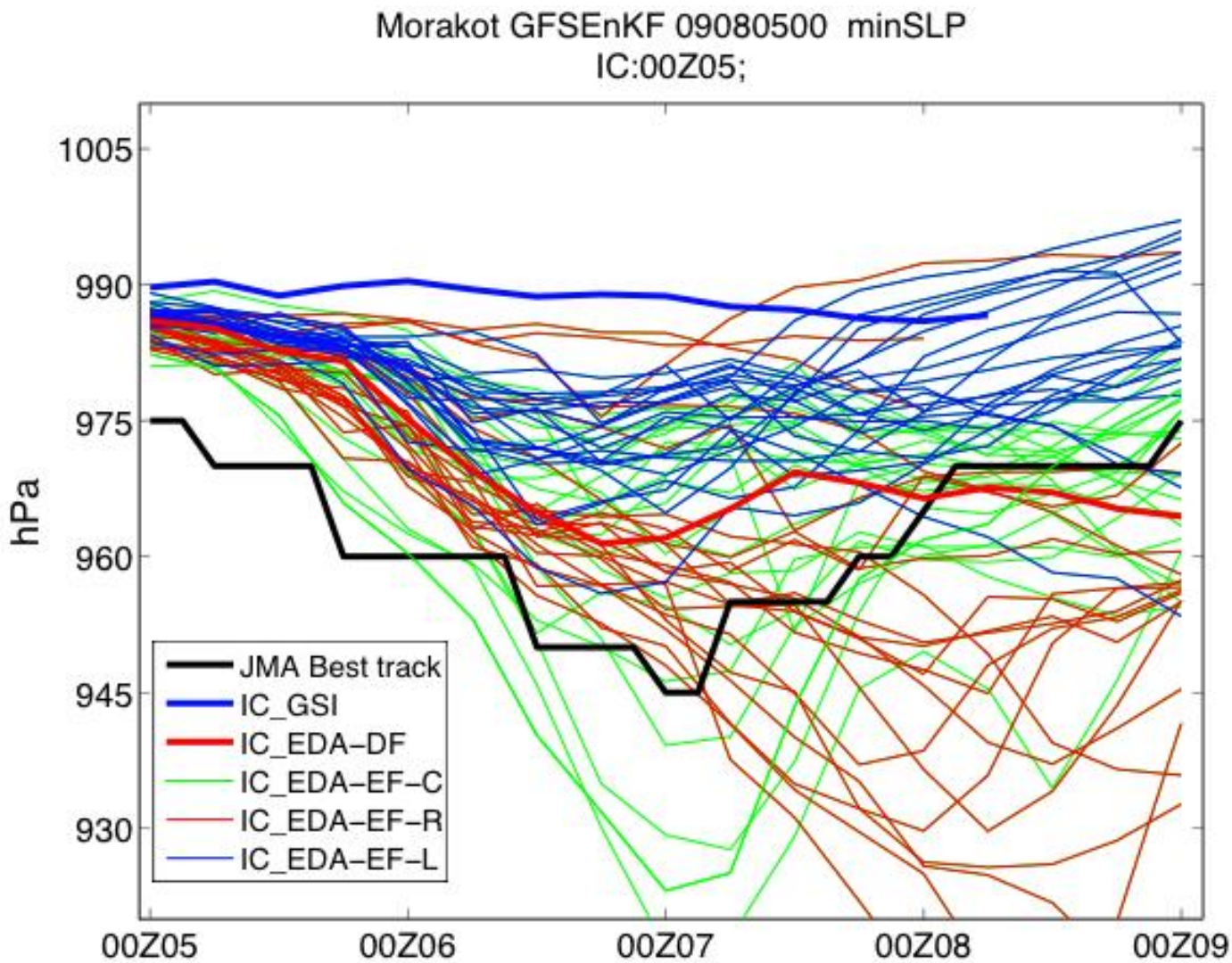


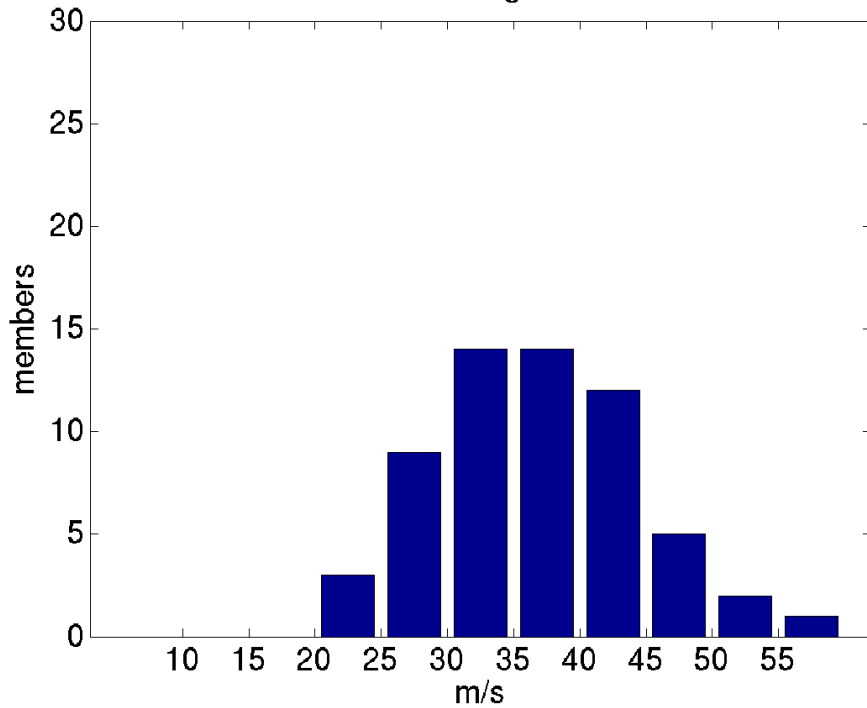
Figure 9b



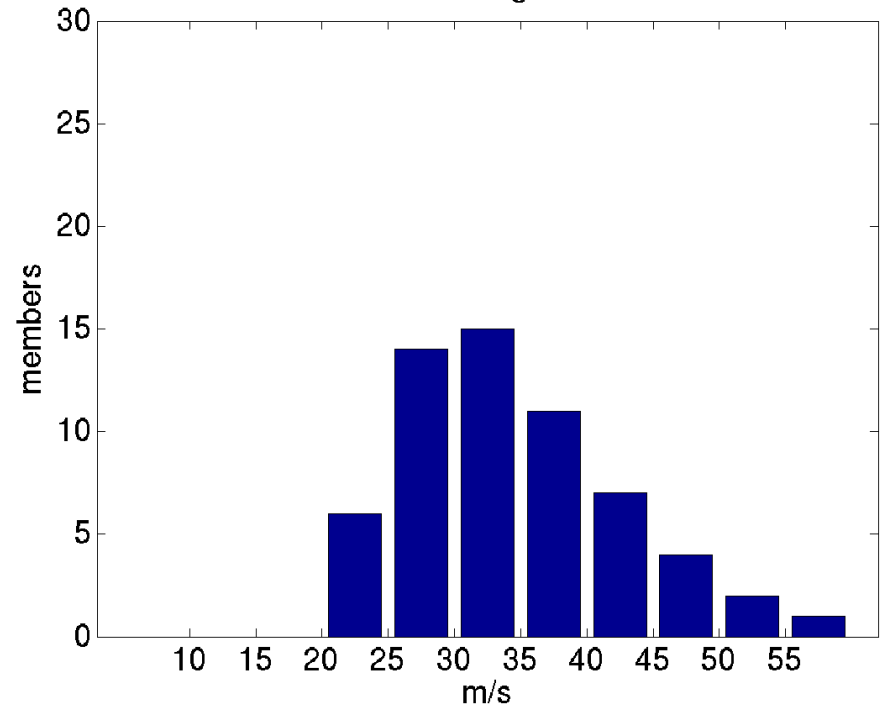


Morakot intensity PDFs

00Z07 Aug2009

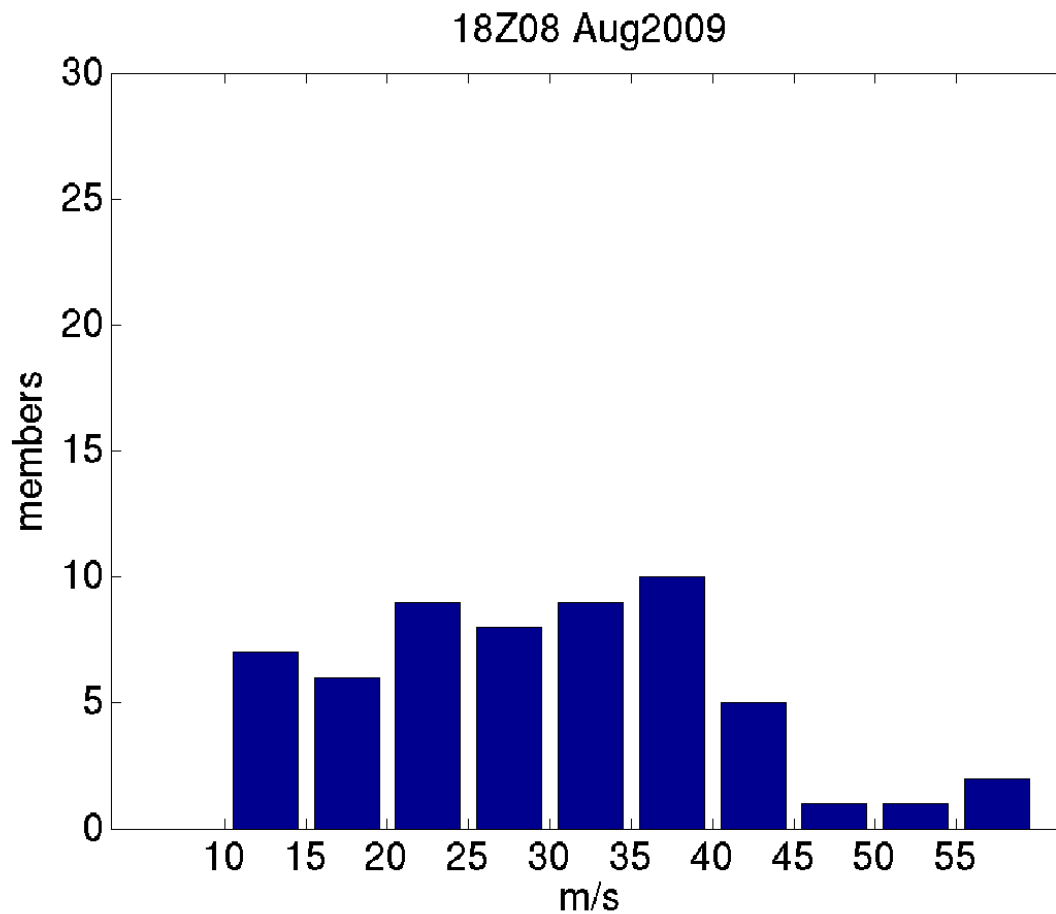


06Z07 Aug2009





Morakot intensity PDFs





Preliminary Program Plan



Preliminary Recommendations From FY09 Demo system



- **An advanced data assimilation technique (beyond GSI-3DVAR) gives additional benefits.**
 - **Recommendation:** Operational and research groups should define and execute the most expeditious and practical path to implement an advanced, global, hybrid data assimilation system at NCEP.
- **High resolution global ensembles are showing clear promise.**
 - **Recommendation 1:** Continue to evaluate high resolution ensembles with target resolutions of 10-15 km.
 - **Recommendation 2:** Improve ensemble member composition and the benefits therein, given available operational computing (global and regional).
- **Regional models still need a lot of work.**
 - Errors appear to be related to initialization of the vortex and to physics, particularly convection.
 - **Recommendation 1:** Institute and focus a community wide effort to improve the initialization, physics issues in the regional model.
 - **Recommendation 2:** Target resolution should be 1-5 km. This may eliminate much of the convection parameterization issues.



HFIP Development Planned/Anticipated FY10 HPCC Resources



System	Available Processors for HFIP	Processor Hours Available	Available Date	Notes
Boulder	~3000	Dedicated 24-7	Now	Available - For HFIP Dedicated Use.
Gaithersburg	160	Dedicated	Now	Available - For HFIP Dedicated Use.
TACC	16,000 max	1) ~4M thru 12/31/09 2) 20M Proposed for FY10	1) ~5M Now 2) 20M 1/1/10	1) Available – For HFIP Use. 2) Planned: Proposal submitted for additional 20M processor hours starting on January 1, 2010.
Boulder	~2000	Dedicated 24/7	8/1/10	Planned upgrade of nJet with new FY10 HFIP Funds.
DOE		36M	1/1/10	Planned: Proposal submitted, Questions Answered. Potential – pursuing participation in DOE-NOAA Petaflop computing partnership at Argonne NL



Planned FY10 HFIP HPCC Resource Usage

System	Planned Usage
TACC	High-Resolution Global Models and Global Dual-Model Ensembles; Data Assimilation (EnKF); Regional Ensembles (PSU ARW); Testing and Evaluation
Boulder	High-Resolution Deterministic Regional Models, High Resolution Nested Deterministic Models, Testing and Evaluation.
DOE	High-Resolution Global Models; High-Resolution Two-Way Nested Global Models.
Gaithersburg	Final Implementation Testing of HWRF/Testing of HWRF 4 km



HFIP FY10 Program Plan

General Strategy

HFIP will continue to emphasize a Two-Suite Parallel process:

- 1) Operational Model Suite (EMC Managed Testing & Implementation):
 - Assumes operations computing power availability.
 - Development is focused on HWRF. HFIP will focus on alternatives to improve HWRF prediction (initial conditions & physics). (Widen the funnel.)
2. Demonstration Model Suite:
 - Primary Purpose: To prove that the enhanced computing resource will lead to improved guidance accuracy. This will hopefully lead to increased operational resources.
 - Need to continue to seek computer resources outside of NOAA.



HFIP FY10 Program Plan Strategy

Strategy 1: Develop and test the model for and products from high resolution global ensembles.

- Beyond ~3 days, the forecast problem requires a global model.
- Ensembles will provide the most probable forecasts (especially true for longer lead times).
- Ensembles need to be at high resolution in the 5-10 km range (such that the feedback from the hurricane to its steering flow can be adequately represented).
- Need to focus on computer resources outside of NOAA.

Strategy 1 Implementation:

- A Proposal for 36M processor-hours was submitted to DOE (INCITE). Also submitted a proposal for 20M additional processor-hours on the TACC. With these resources and with ~4M processor hours on TACC still remaining, we plan to:
 - Complete analysis of the 2009 results including rerunning those cases that were not run in quasi-real-time.
 - 1) Generation of genesis statistics for the GFS and FIM global models.
 - 2) Analysis of the global ensemble runs, examining error statistics for track and intensity from the ensemble mean and mode. Comparison with similar error statistics for the various deterministic models.
 - 3) Comparison of the error statistics for various resolutions (30 km, 15 km and 10 km) for the deterministic models.
 - Run all cases for 2008 for the months of August and September. Includes the deterministic global models and the 30 km ensemble. Requires running the EnKF system for Aug-Sept, 2008. Results will be added to the statistics for 2009.
 - Run a global ensemble at 15 Km with as many members as possible. Nominal number of members will be 10. Compare error statistics with the 30 km ensemble.



HFIP FY10 Program Plan Strategy (continued)

Strategy 2: Develop and test very high resolution regional models including ensembles of the regional model(s).

- Regional model resolutions of 1-3 km will be necessary to resolve the inner core dynamics of the hurricane to accurately forecast intensity.
- Ensembles (both multi-member and multi-model) will provide the most probable forecasts (especially true for longer lead times). Ensembles tend to cancel out initial condition error, and in case of multi-model ensembles, inherent model error.

Strategy 2 Implementation:

- Implement much of the regional model development on the n-jet computer, saving the TACC and DOE facilities for the global models with an exception noted below.
 - Complete analysis of 2009 results for multi-model ensemble that included 7 models from across the HFIP community. Includes rerunning cases that were not run in quasi-real-time.
 - Calculation of bias corrections for each component model.
 - Analysis of the multi-model ensemble runs examining error statistics for track and intensity from the ensemble mean.
 - Development of products that can be derived from the ensemble output.
 - Prepare to run the multi-model ensemble during the 2010 hurricane season.
- Complete analysis of the Penn State ensemble runs for 2009.
 - Calculation of bias corrections for each component model.
 - Analysis of the multi-model ensemble runs examining error statistics for track and intensity from the ensemble mean.
 - Development of products that can be derived from the ensemble output.
 - Prepare to run ensemble during the 2010 hurricane season.
- Run the Penn State system on additional cases in 2008 and for the east pacific.



HFIP FY10 Program Plan Strategy (concluded)

Strategy 3: Organize and focus components of the HFIP community on examining ways to improve the accuracy of HWRF.

Strategy 3 Implementation:

EMC not only to lead, but to assume the bulk of the effort in developing and implementing upgrades to HWRF. EMC does not have adequate resources to handle all aspects of the investigation. Hence, HFIP will organize a number of organizations outside EMC to work with EMC. The focus will be on physics and initialization development:

- AOML to take the lead in developing new or alternative methods for initializing regional models with emphasis on ultimately using these new initialization techniques in HWRF. They will collaborate closely with the initialization work being done at Penn State, NCAR and NRL.
 - Modifications to the current HWRF initialization process.
 - Explore alternative initialization procedures such as EnKF and adjustment of the initial vortex to match NHC observations of the vortex..
 - Inclusion of new data sets particularly aircraft radar data and novel use of satellite information when aircraft data are not available.
- GFDL, NRL, and NCAR will focus on evaluation and testing of the physics package that will be part of the HWRF system.
- NCAR will coordinate this community wide effort to in physics and initialization of regional models.



HFIP FY10 Development Objectives and Milestones by Strategy (Draft)



Strategy	Milestones
<p>Strategy 1: Develop and test the model for and products from high resolution global ensembles.</p>	<ol style="list-style-type: none"> 1) Complete analysis of FY09 demo results. (Q2) 2) Complete FY08 runs of demo system and analysis of results. (Q4) 3) Complete 15 km 20 member ensemble and analysis of results. (Q4—provisional depending on availability of computing)
<p>Strategy 2: Develop and test very high resolution regional models including ensembles of the regional model(s).</p>	<ol style="list-style-type: none"> 1) Complete analysis of FY09 multi model ensemble. (Q3) 2) Complete analysis of FY09 Penn State model ensemble. (Q3) 3) Run Penn State system over 2008 season and complete analysis. (Q4)
<p>Strategy 3: Organize and focus components of the HFIP community on examining ways to improve the accuracy of HWRF.</p>	<ol style="list-style-type: none"> 1) NCAR complete a plan for managing community physics/initialization effort. (Q2) 2) Complete prototype regional physics/initialization package for testing in FY10 demo system. (Q3)



HFIP FY10 Development Objectives and Milestones by Organization (Draft)



Org.	Overall Objective	Milestones/Deliverables
ESRL	<ol style="list-style-type: none"> 1) Develop and test the FIM global model at as high a resolution as possible on available computing platforms. 2) Develop and test high resolution Global Ensemble (target resolution 10-15 km). 3) Work with NRL in developing an optimum physics package for a global model at 10-15km. 4) Continue to develop and evaluate the EnKF system. 	<ol style="list-style-type: none"> 1) Run 15 km global ensemble for 2008. 20 members if possible. (Q2) 2) New physics package for use in global models. (Q3) 3) Evaluate various improvements to the global model. (Q4) 4) Evaluate EnKF relative to other DA system with emphasis on 4DVAR. (Q4)
AOML	<ol style="list-style-type: none"> 1) Develop and evaluate new methods to incorporate satellite data for initialization of the hurricane vortex in regional models. 2) Work with NCAR to develop physics packages suitable for regional models with resolutions of 105 km. 3) Working closely with EMC Develop and evaluate high resolution HWRF system for 4 km implementation. 	<ol style="list-style-type: none"> 1) Develop a preliminary new initialization method making best use of satellite data. (Q3) 2) New physics package for use in regional models. (Q3) 3) Provide components of the HWRF system to improve operational model. (Q2)
EMC	<ol style="list-style-type: none"> 1) Test and evaluate improvements to HWRF to insure that there is a steady annual increase in HWRF performance Skill. 2) Work with HRD, NCAR, NRL to improve regional model physics. 	<ol style="list-style-type: none"> 1) Implement in operations an HWRF system with improved skill over 2008. 2) New physics package for use in regional models. (Q3)



HFIP FY10 Development Objectives and Milestones by Organization (Draft) (continued)



Org.	Overall Objective	Milestones/Deliverables
GFDL	<ol style="list-style-type: none"> 1) Develop and evaluate the cubed sphere model for hurricane applications. 2) Work with AOML and NCAR to develop physics packages for 1-5 km regional models. 	<ol style="list-style-type: none"> 1) Compare hurricane forecast skill between the cubed sphere model and GFS for the 2008 and/or 2009 hurricane season. 2) New physics package for use in regional models. (Q3)
NRL	<ol style="list-style-type: none"> 1) Compare relative skill of 3DVAR and 4DVAR in global models. 2) Work with ESRL in developing an optimum physics package for a global model at 10-15km. 3) Develop and evaluate TC-COAMPS for 4 km resolution including new physics packages and initialization methods. 	<ol style="list-style-type: none"> 1) Comparison of 3DVAR with 4DVAR for skill in hurricane forecasting using Navy DA systems and NOGAPS. 2) New physics package for use in regional models. (Q3) 3) New physics package for use in global models. (Q3)
NCAR	<ol style="list-style-type: none"> 1) Maintain the hurricane codes for community access. 2) Develop and maintain hurricane verification and model diagnostics systems. 3) maintain HFIP demo model results repository. 4) Lead effort to improve regional model physics and initialization. 	<ol style="list-style-type: none"> 1) Complete repository for FY09 demo system results. (Q2) 2) Release HWRF to the community. (Q2) 3) Release hurricane related components of MET to the community. (Q4) 4) Release plan for improving physics and initialization in regional models. (Q2)



HFIP FY10 Development Objectives and Milestones by Organization (Draft) (concluded)



Org.	Overall Objective	Milestones/Deliverables
NESDIS	1) Work with DTC to and model groups to develop methods to diagnose model deficiencies. 2) Work with DTC to develop new verification systems for hurricane model guidance.	1) Release new verification and diagnostic techniques in the DTC MET system. (Q4) 2) Analysis of HWRF, ARW(MMM), and GFDL for 2009. (Q2)
NHC	1) Work with EMC, NESDIS and DTC to test and evaluate HWRF with a focus on model diagnostics. 2. Work with ESRL, and others to develop ensemble products. 2) Develop forecaster useful products from model guidance.	1) Develop new model guidance products for forecaster use. (Q3) 2) Analysis of HWRF, GFDL performance for 2009. (Q2)



Managed Regional Ensemble?



- **In FY09 the multi model ensemble proved we could run and distribute results from a multi-model ensemble**
 - We would like to formalize that ensemble and begin to develop new products
 - System will be run on N-JET
- **Some notes**
 - With Multi model ensembles as models are added, for the first two the improvement in skill (eg: mean) is dramatic but as more are added additional improvement usually diminishes
 - Optimal number may be on the order of 3-4.
 - Adding poor models eventually decreases skill
 - There is a computational limit to the number of models we can run



Managed Regional Ensemble?



- **Assembling the Ensemble**

- Any model regional model is a potential candidate
- Each potential model will be run over a standard set of test cases
 - The HRH suite?
 - The 2008/2009 NHC selected cases?
 - May be an important option for those models with their own DA system
- The top three or four will become members of the formal biased corrected ensemble for the following hurricane season
 - Shall we run more than one member per model?

- **Some notes**

- Make up of the formal ensemble can change from year to year
 - As the models change they need to be evaluated against the standard set each year
 - Some models may fall out of the top four while others are elevated
 - Since all component models will have been evaluated against the standard set, we will also have the data necessary to compute the overall ensemble skill.



Managed Regional Ensemble?



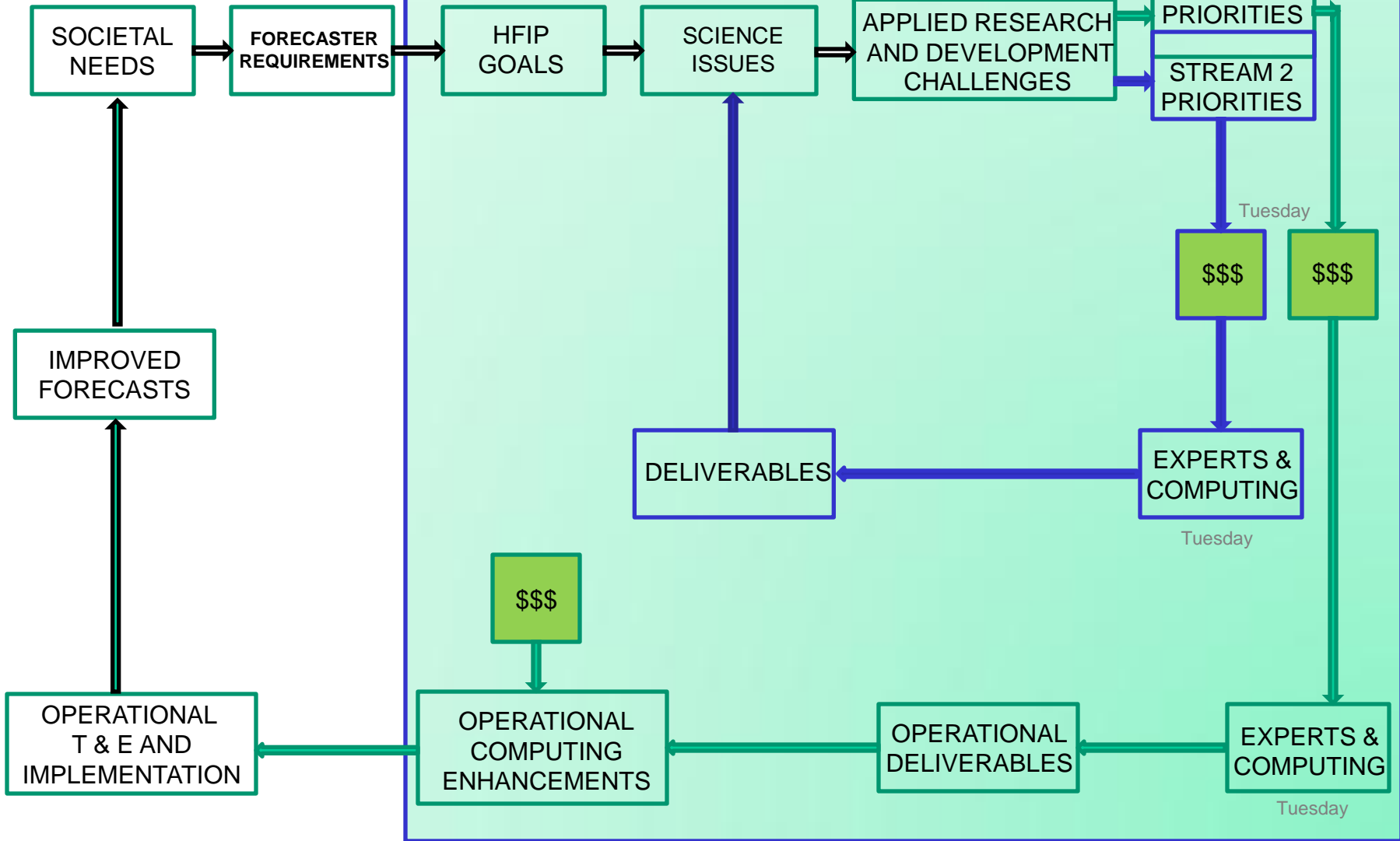
- **Output from the “managed” ensemble**
 - Traditional (mean, mode, spread)
 - Non traditional? (who will build)?
 - Track dependency of intensity
 - How many members do we need for this
 - PDFs
 - Other
- **Can/will it be used by the forecasters**
 - How can we get products to NHC?
 - What level of verification is necessary/desirable

HURRICANE FORECAST IMPROVEMENT PROJECT

Monday Team Reports

Monday afternoon

Monday afternoon





Backup Slides



HFIP Performance Measure Baseline Development

- Draft Recommendations by James Franklin in “A Proposal for HFIP Performance Baselines”, dated 23 April 2009:
 - Track and intensity goals: use a consensus (equally-weighted average) of operational guidance models, evaluated for the Atlantic basin over the period 2006-2008.
 - The track baseline recommendation is for a consensus of GFSI, GFDI, UKMI, NGPI, HWFI, GFNI, and EMXI.
 - The intensity baseline recommendation is for a consensus of GHMI, HWFI, DSHP, and LGEM.
 - Goal for 7 Day Lead time, Accuracy at least as good as the 5-day official forecast in 2003.
 - For RI: use GFDL model (frozen) averaged over the period, 2006-2008.



HFIP Performance Metrics

Draft



Item / Baseline Derivation	Baseline	Goal	Status
Track Error	Day 1 – 49.8 nm Day 2 – 89.6 nm Day 3 – 132.0 nm Day 4 – 175.2 nm Day 5 – 221.9 nm	<ul style="list-style-type: none"> Reduce average track error by 50% for Days 1 through 5 	
Intensity Error	Day 1 – 10.1 kts Day 2 – 13.7 kts Day 3 – 16.0 kts Day 4 – 16.6 kts Day 5 – 17.0 kts	<ul style="list-style-type: none"> Reduce average intensity error by 50% for Days 1 through 5. 	
Rapid Intensity Changes Baseline	POD: Day 1 – 30% Day 2 – 10% Day 5 – 0% FAR: Day 1 – 82% Day 2 – 85% Day 5 – 100%	<ul style="list-style-type: none"> POD: Day 1 – 90% (linear 60% at day 5) Day 2 – 82% Day 5 - 60% FAR: Day 1 – 10% (linear to 30% at day 5) Day 2 – 15% Day 5 – 30% 	