HFIP Verification Team: FY12 Review

28 November 2012

- *OAR / GFDL*
  - Tim Marchok
- *OAR / AOML / HRD*
  - Rob Rogers
- *OAR / ESRL*
  - Mike Fiorino
- *SUNY-Albany*
  - Ryan Torn
- *NRL*
  - Jim Goerss
  - Hao Jin
- *TCMT and DTC*
  - Barb Brown
  - Louisa Nance
  - Ligia Berndardet (DTC)
  - Paul Kucera
- *NESDIS*
  - Mark DeMaria
- *NCEP / NHC*
  - James Franklin
- *NCEP / EMC*
  - Vijay Tallapragada

Verification Team Milestones 2012

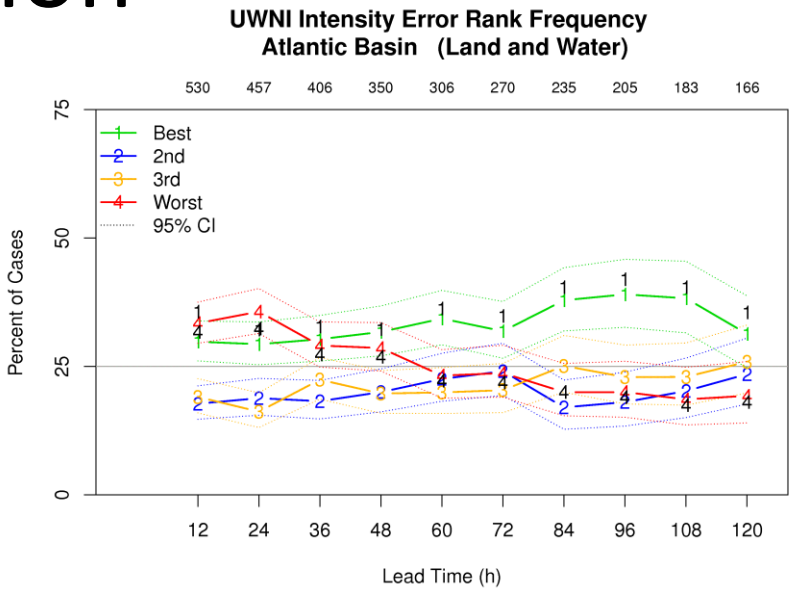
- Testing and evaluation of operational models
- Planning and coordination of HFIP hurricane model evaluations for retrospective and demonstration tests
- Ongoing model verification analyses for retrospective and demonstration exercises
- Development, testing, and implementation of new tools for evaluation of hurricane forecasts.

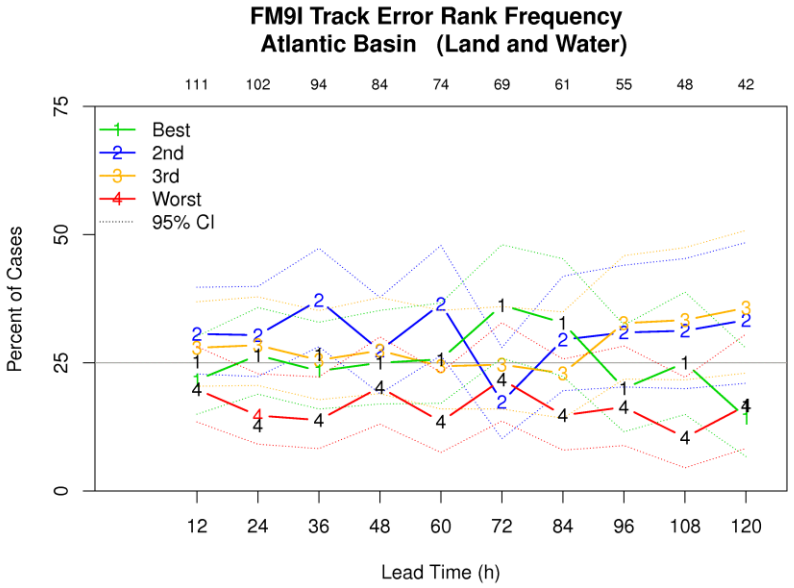
5.2.1, 5.3.1

2012 Retro Planning, Testing & Evaluation

- 8 participants
- Comparisons
  - Top-flight models
    - 1-to-1
    - Rank frequencies
  - Consensus (1-to-1)
    - Add to operational consensus

    and/or
    - Direct comparison
- Additional analysis
  - Performance of Stream 1.5 consensus
  - Impact of PSU w/out radar on Stream 1.5 consensus
  - Direct comparison of PSU w/ & w/out radar
  - SPICE -vs- fixed consensus





5.2.1,
5.3.1

2012 Retrospective Participants

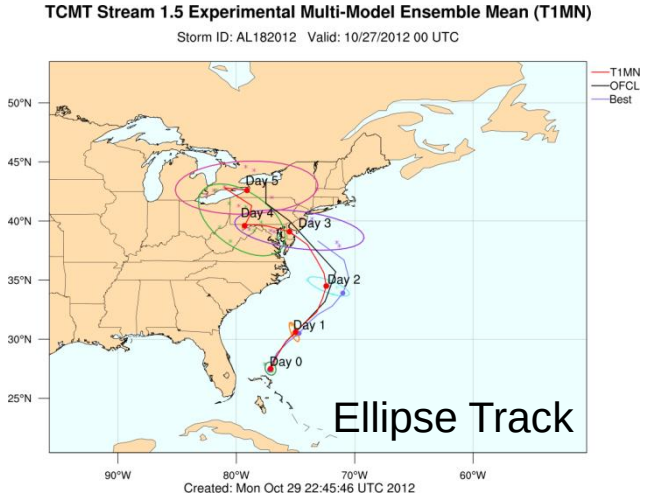
Organization	Model	Type	Form of guidance
NCAR/MMM and SUNY-Albany	AHW	Regional-deterministic	TV15, IV15
NRL	COAMPS-TC	Regional-deterministic	IV15
GFDL	GFDL hurricane model	Regional-ensemble	Ensemble mean and unbogused member - explict track and intensity
PSU	ARW	Regional-deterministic	TV15, explicit intensity, IV15
UW-Madison	UW-NMS	Regional-deterministic	IV15
ESRL	FIM	Global-deterministic	TV15
FSU	MMEN	Correlation-based consensus	None
NESDIS/STAR and CIRA	SPICE	Statistical-dynamical-consensus	Explicit intensity

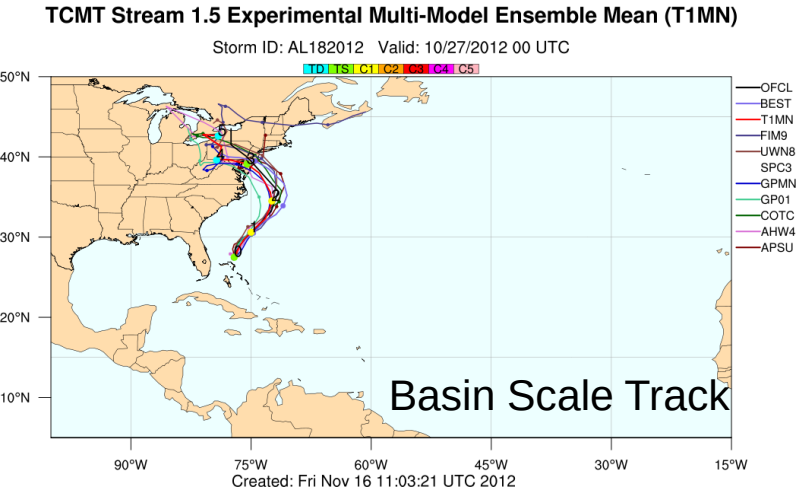
See presentation by L. Nance during 5 December telecon

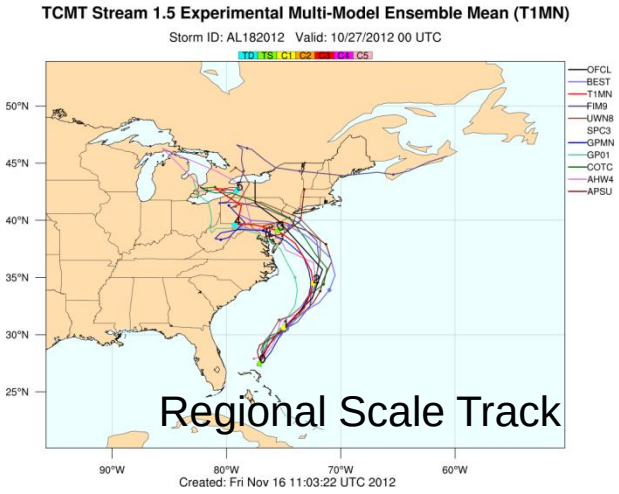
5.3.2

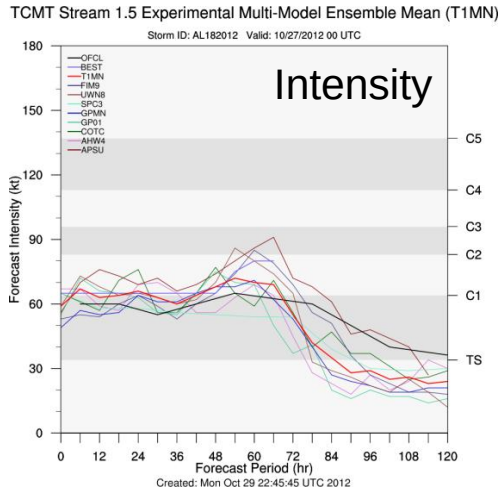
HFIP Real-time Demonstration Display

- Supported the real-time display of experimental and operational forecasts of track and intensity for the HFIP Demonstration
- Created multi-model ensemble mean forecasts
- Products available at the TCMT and HFIP Websites:
  - http://www.ral.ucar.edu/projects/hfip/d2012/forecasts/
  - http://www.hfip.org





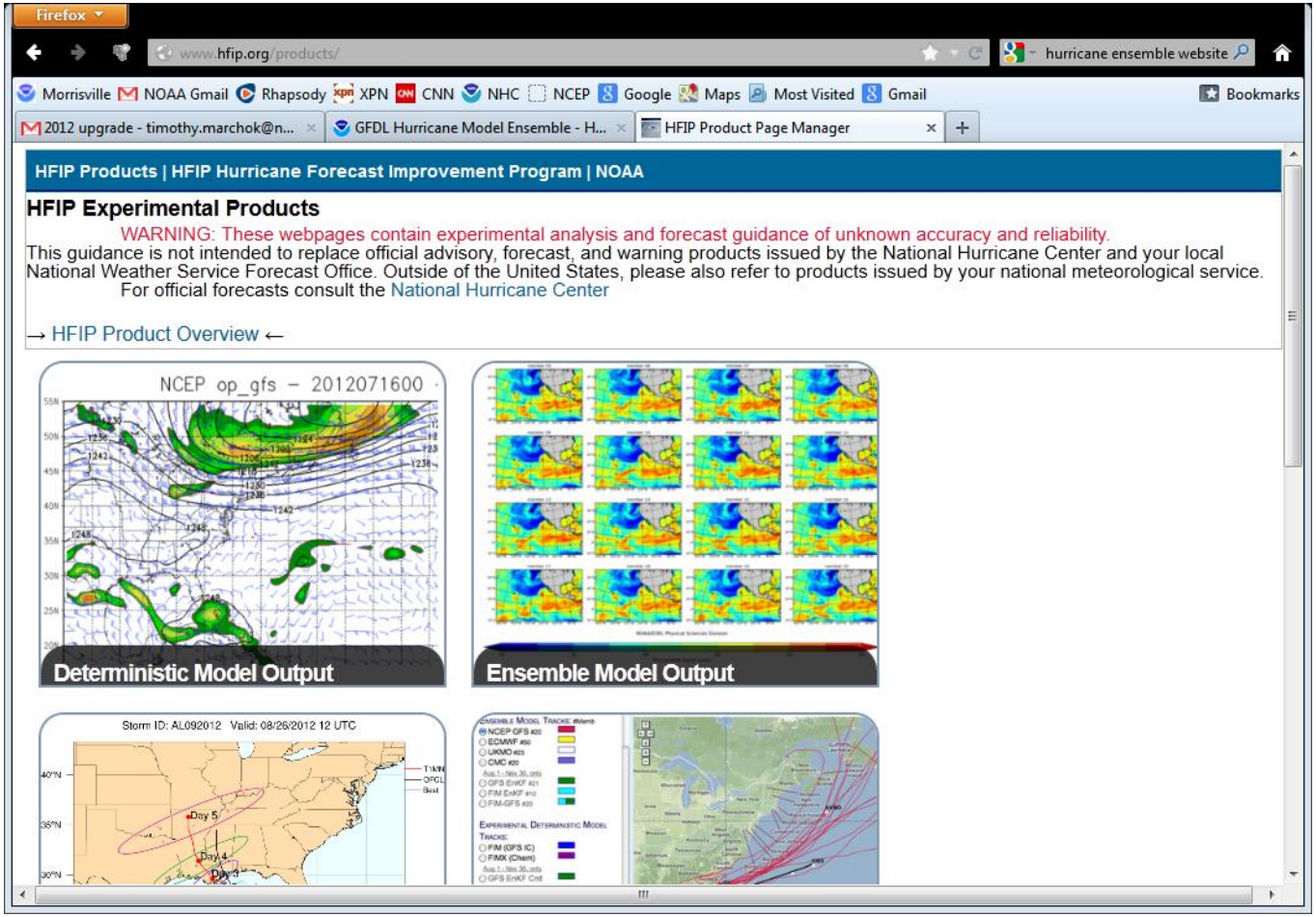




5.3.2 Provide graphics of Demo experiment Stream 1.5 forecasts and ongoing verification analyses in real time.

http://www.hfip.org/products/

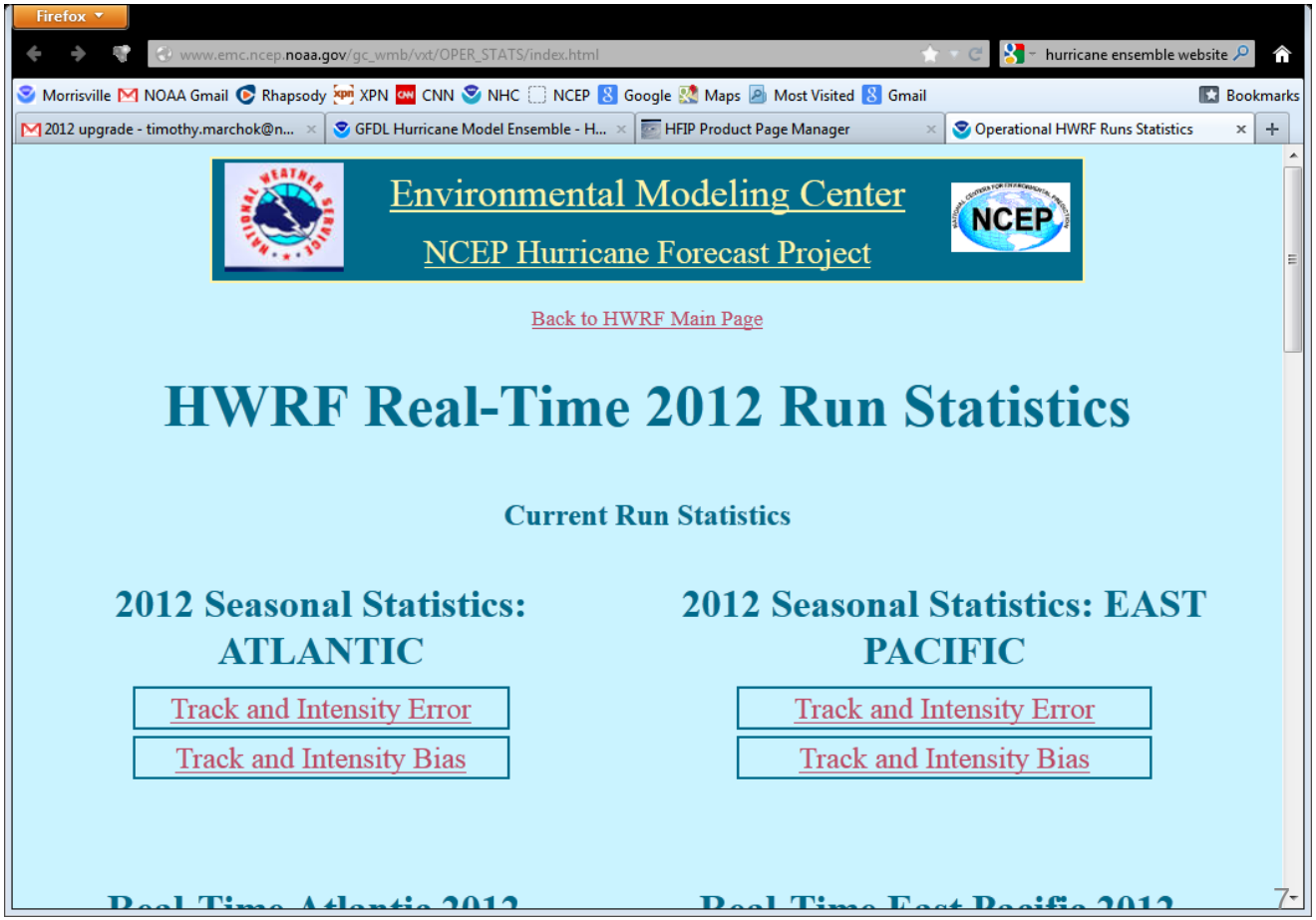
HFIP website
(Paula McCaslin)



5.3.2 Provide graphics of Demo experiment Stream 1.5 forecasts and ongoing verification analyses in real time.

Storm-by-storm verification in real-time can be found here (password protected):
http://www.emc.ncep.noaa.gov/gc_wmb/vxt/OPER_STATS/index.html

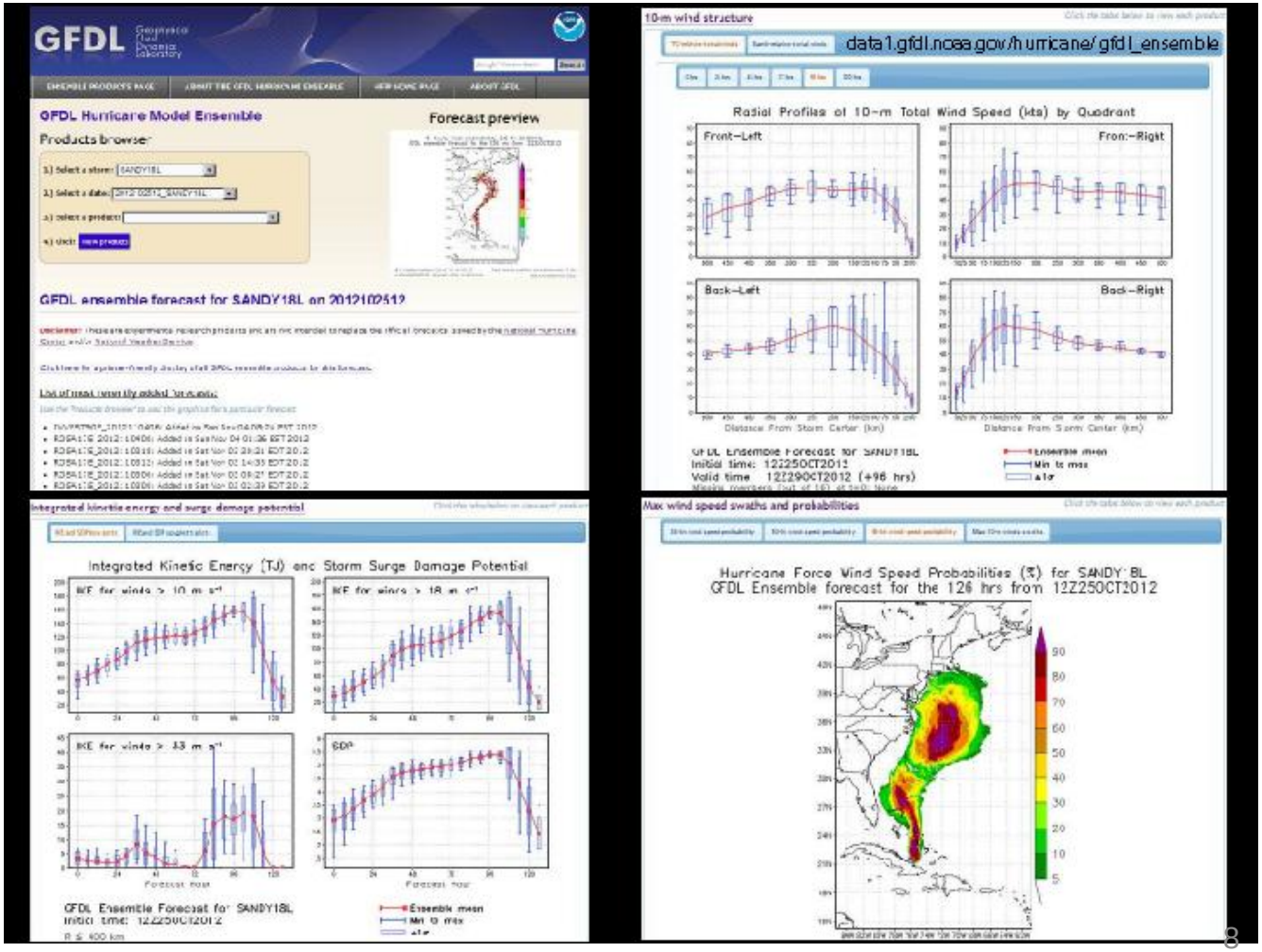
EMC/HWRF website (Vijay, Chanh Kieu)



5.3.2 Provide graphics of Demo experiment Stream 1.5 forecasts and ongoing verification analyses in real time.

http://data1.gfdl.noaa.gov/hurricane/gfdl_ensemble/

GFDL Ensemble website (Matt Morin)



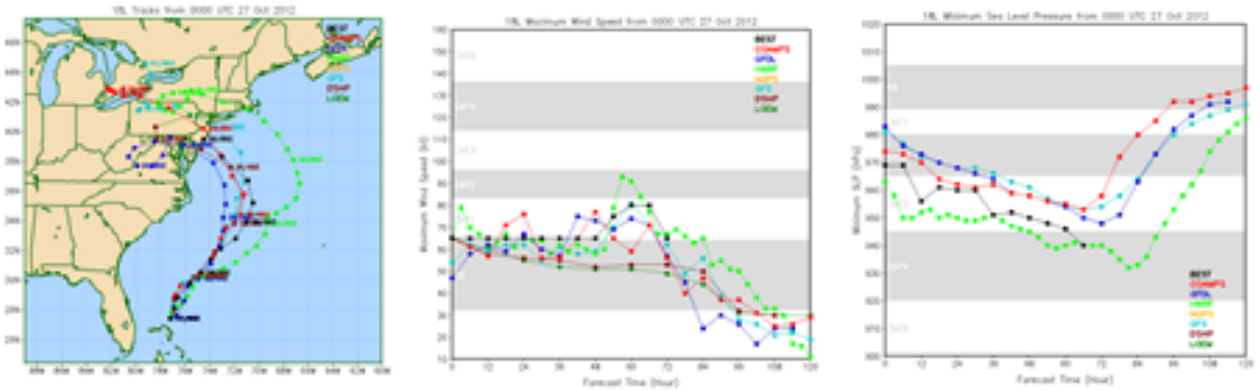
5.3.2 Provide graphics of Demo experiment Stream 1.5 forecasts and ongoing verification analyses in real time.

NRL website
(Hao Jin)

Real-Time Multi-model Verification

http://www.nrlmry.navy.mil/coamps-web/web/tc

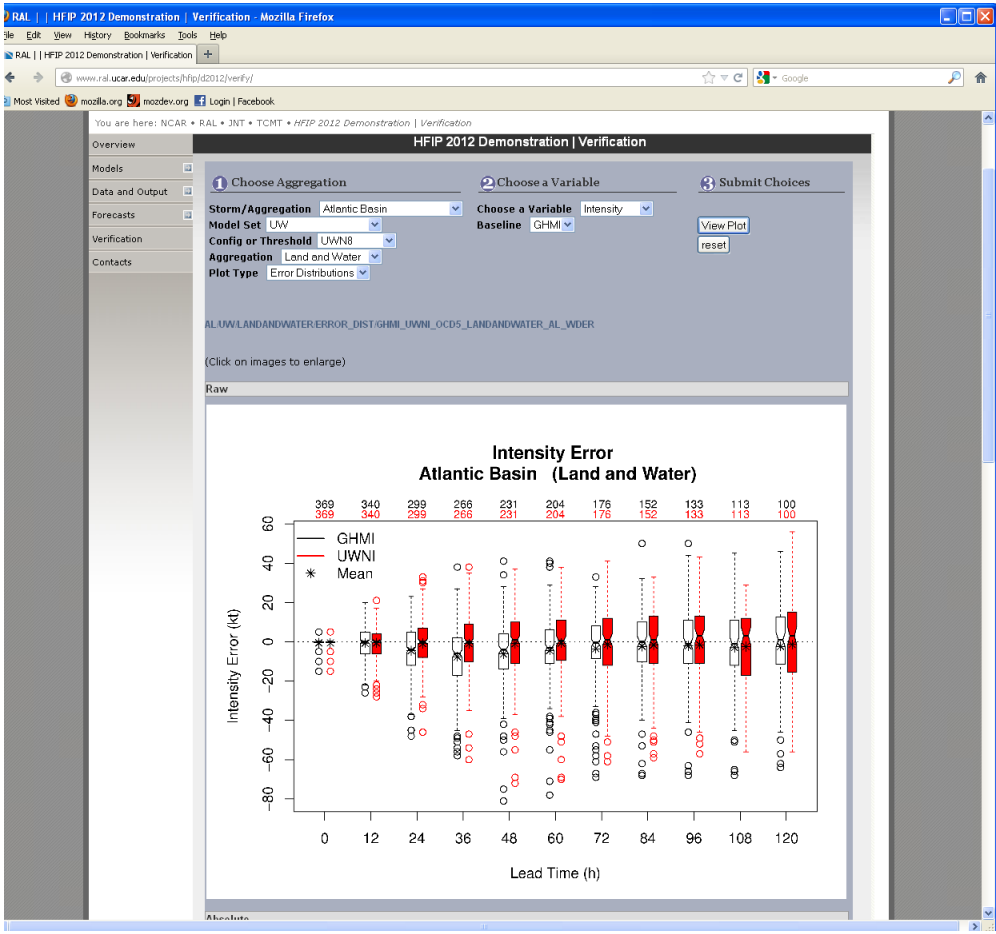
- Improved the NRL tropical cyclone (TC) web site using the active storm list to provide the more robust multi-model verification of TC track and intensity in real-time.
- Improved the web site performance to provide the hourly update.
- Added the statistics models to the multi-model verification.



5.3.2

Online Access to HFIP Demonstration Evaluation Results

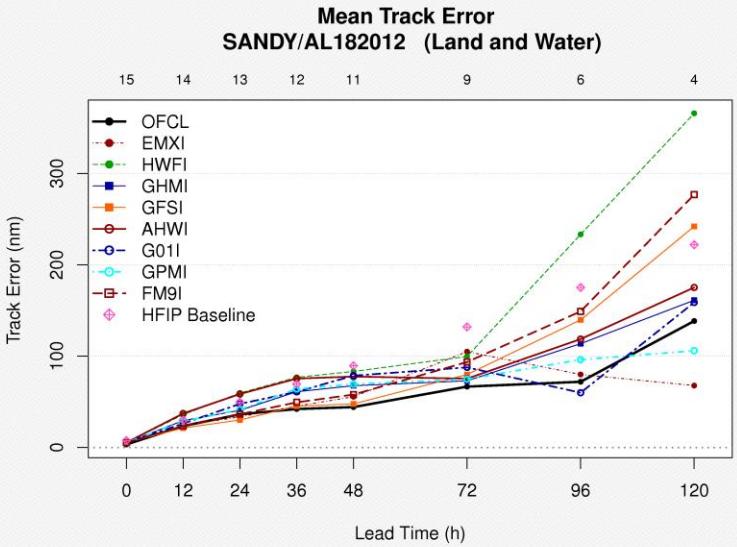
- Evaluation graphics are available on the TCMT website:
  - http://www.ral.ucar.edu/projects/hfip/d2012/verify/
- Wide variety of evaluation statistics are available:
  - Aggregated by basin or storm
  - Aggregated by land/water, or water only
  - Different plot types: error distributions, line plots, rank histogram, Demo vs. Retro
  - A variety of variables and baselines to evaluate

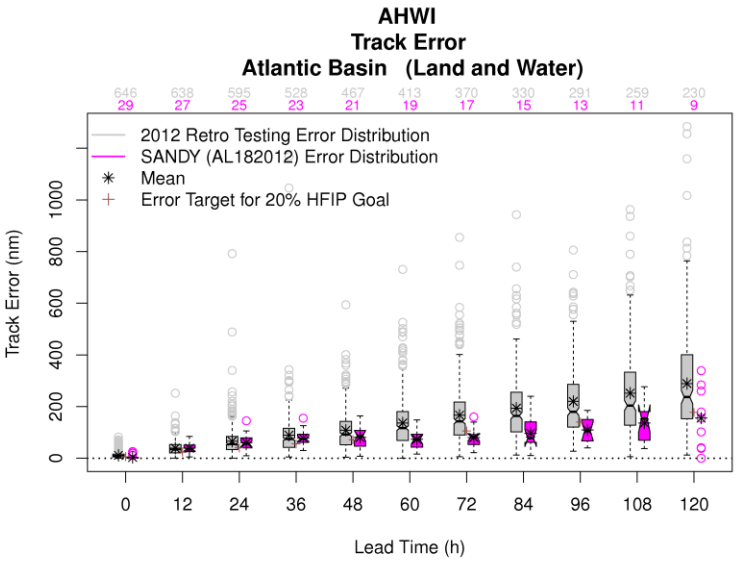


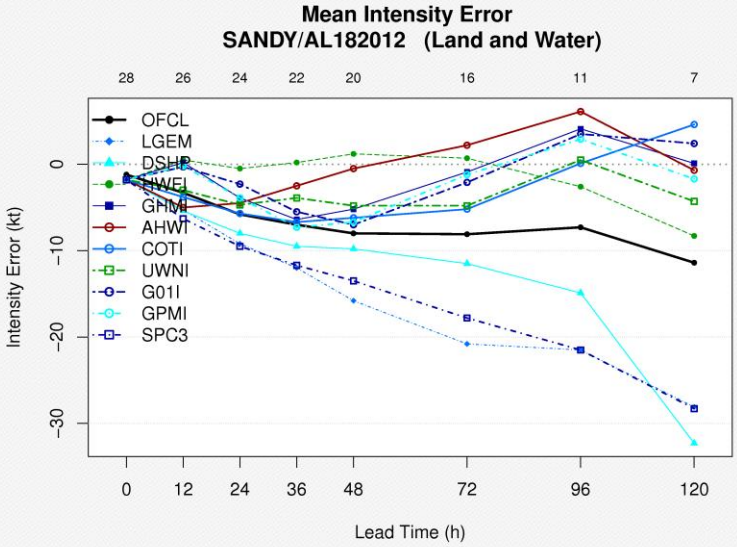
5.3.2

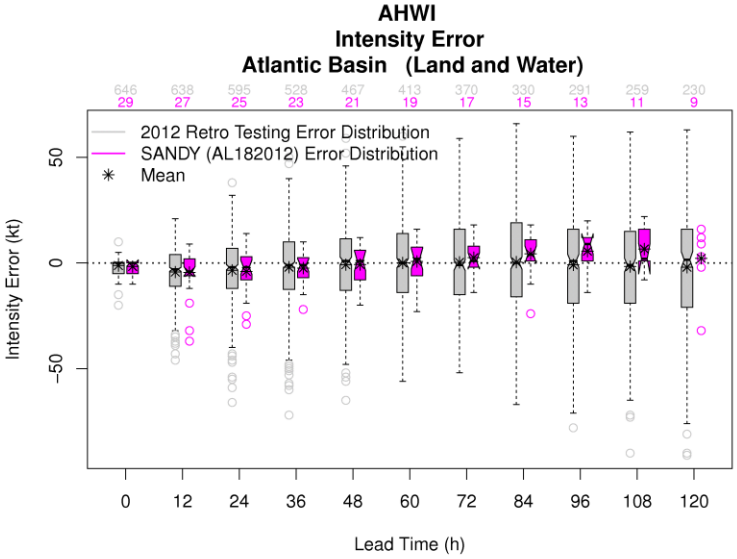
Case Study Analysis

Hurricane Sandy Evaluation





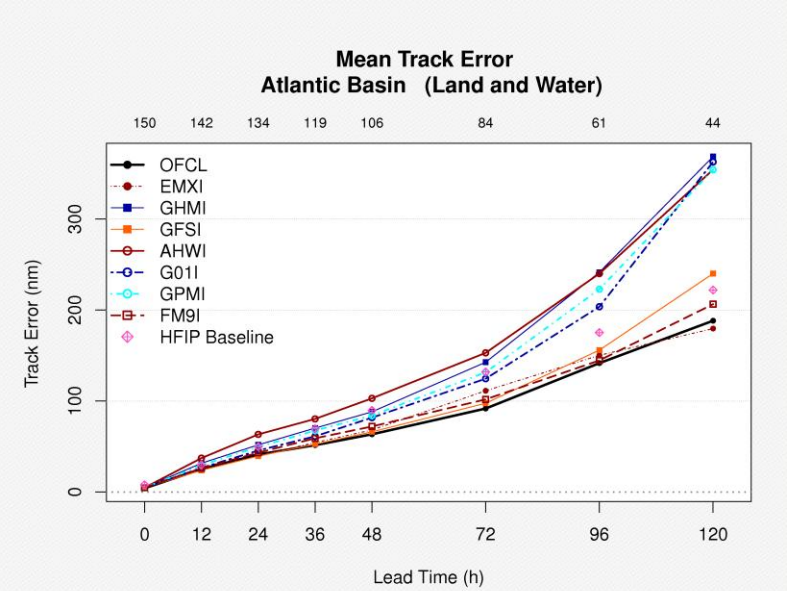


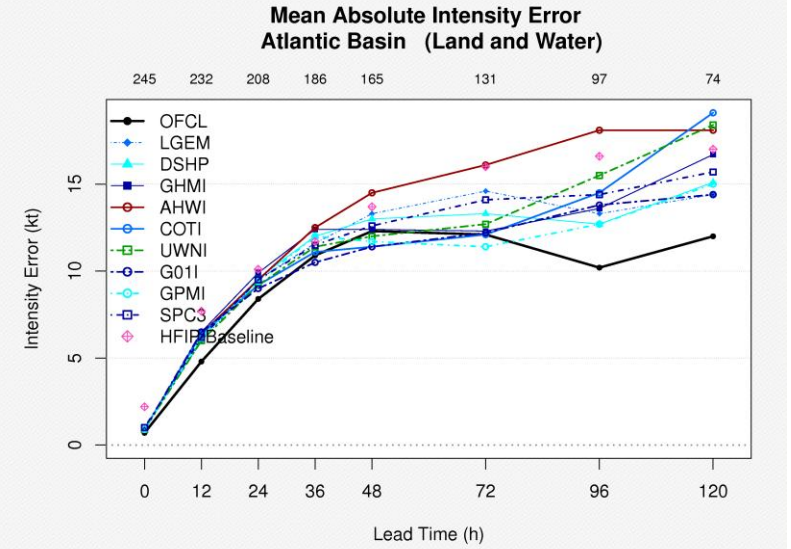


5.3.2

Demonstration Evaluation

- Stream 1.5, 2.0 and operational models were evaluated for the 2012 HFIP Demonstration
- Models were evaluated with a homogeneous sample
- A variety of evaluations were conducted following the methodology of the Retrospective evaluation
- Mean track and intensity errors are presented on the right

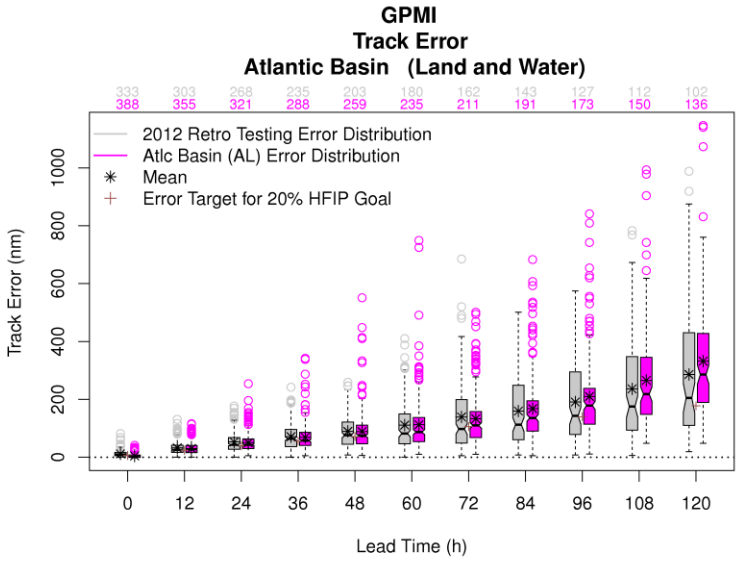


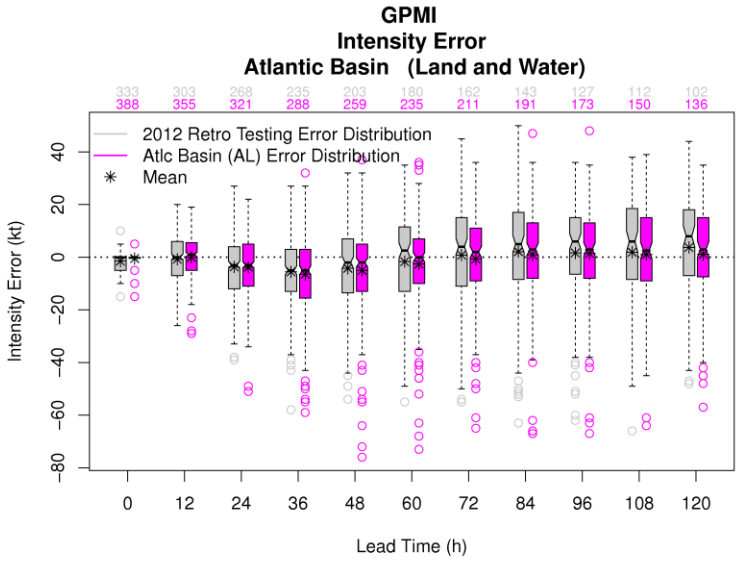


Retrospective vs. Demonstration Evaluation

Comparison of track and intensity error distributions

- Retrospective (gray) vs. Demonstration (magenta) evaluations
- All stream 1.5 candidates were evaluated
- Example is for GPMI: error distributions have similar characteristics





5.3.2

Retrospective vs. Demonstration Evaluation – SS Evaluation

Example – Consensus w/AHWI – 2012 Retrospective

	Forecast Hour	0	12	24	36	48	60	72	84	96	108	120
Atlantic Basin	TVCA	0.0	0.0	0.8	1.4	2.0	1.8	1.9	3.7	8.0	11.2	14.3
	Track	0%	0%	2%	2%	3%	2%	2%	3%	5%	6%	6%
	(Land and Water)	–	0.000	0.954	0.980	0.999	0.954	0.915	0.995	0.999	0.999	0.999
	ICON	0.0	0.1	0.4	0.7	0.8	0.8	0.7	0.2	0.1	0.2	0.4
	Intensity	0%	1%	4%	5%	6%	5%	4%	1%	1%	1%	2%
	(Land and Water)	–	0.682	0.999	0.999	0.999	0.999	0.999	0.495	0.261	0.382	0.575
	ICON	0.0	0.1	0.4	0.7	0.8	0.7	0.5	0.1	0.0	0.1	0.2
	Intensity	0%	1%	4%	5%	6%	5%	3%	1%	0%	1%	1%
	(Water Only)	–	0.682	0.999	0.999	0.999	0.999	0.987	0.261	0.000	0.197	0.310
Eastern North Pacific Basin	TVCE	0.0	0.3	0.9	1.3	1.5	2.5	2.9	5.4	5.4	6.3	6.8
	Track	0%	1%	2%	2%	2%	3%	3%	4%	3%	3%	3%
	(Land and Water)	–	0.682	0.927	0.935	0.715	0.810	0.791	0.992	0.952	0.890	0.830
	ICON	0.0	-0.1	0.0	0.2	0.5	0.6	0.9	1.2	1.5	2.2	2.5
	Intensity	0%	-2%	0%	2%	4%	4%	5%	6%	7%	10%	11%
	(Land and Water)	–	0.682	0.000	0.682	0.987	0.953	0.997	0.911	0.824	0.929	0.973
	ICON	0.0	-0.1	0.0	0.3	0.5	0.4	0.6	0.8	1.0	1.8	2.0
	Intensity	0%	-2%	0%	2%	3%	2%	3%	4%	5%	9%	10%
	(Water Only)	–	0.682	0.000	0.865	0.987	0.816	0.952	0.680	0.634	0.861	0.994

5.3.2

Retrospective Vs. Demonstration Evaluation – SS Evaluation

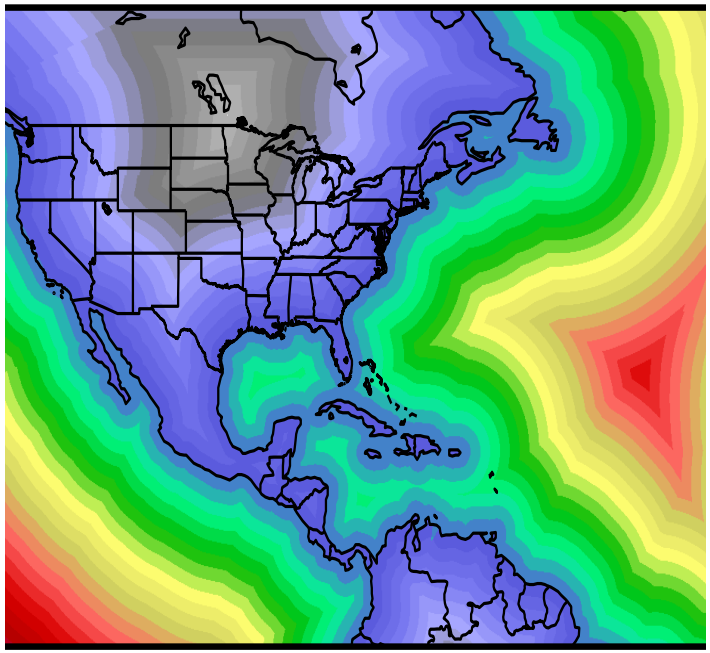
Example – Consensus w/AHWI – 2012 Demonstration

	Forecast Hour	0	12	24	36	48	60	72	84	96	108	120
Atlantic Basin	TVCA	0.0	-0.2	-0.5	0.2	0.1	0.1	-0.7	-0.9	-2.4	-0.1	1.6
	Track	0%	-1%	-1%	0%	0%	0%	-1%	-1%	-2%	0%	1%
	(Land and Water)	–	0.495	0.595	0.176	0.061	0.047	0.212	0.130	0.283	0.008	0.158
	ICON	0.0	0.1	0.2	0.2	0.0	0.1	0.4	0.2	0.2	0.7	0.5
	Intensity	0%	2%	2%	2%	0%	1%	4%	2%	2%	6%	4%
	(Land and Water)	–	0.682	0.953	0.953	0.000	0.197	0.816	0.494	0.310	0.979	0.786
	ICON	0.0	0.1	0.2	0.1	-0.1	-0.1	0.2	0.3	0.6	0.5	0.0
	Intensity	0%	2%	2%	1%	-1%	-1%	2%	3%	5%	4%	0%
	(Water Only)	–	0.682	0.953	0.382	0.261	0.158	0.310	0.450	0.863	0.785	0.000
Eastern North Pacific Basin	TVCE	0.0	-0.4	-2.1	-0.6	-1.5	0.8	3.5	8.7	25.2	50.9	-34.8
	Track	0%	-2%	-6%	-1%	-2%	1%	4%	8%	19%	33%	-40%
	(Land and Water)	–	0.575	0.781	0.224	0.340	0.180	0.496	0.453	0.922	0.993	–
	ICON	0.0	-0.1	-0.4	-0.2	0.0	-0.2	-0.1	-0.6	-1.1	-0.2	3.8
	Intensity	0%	-2%	-6%	-2%	0%	-3%	-1%	-8%	-11%	-2%	22%
	(Land and Water)	–	0.681	0.952	0.680	0.000	0.493	0.260	0.947	0.918	0.193	–
	ICON	0.0	-0.1	-0.4	-0.2	0.0	-0.2	-0.1	-0.6	-1.1	-0.2	3.8
	Intensity	0%	-2%	-6%	-2%	0%	-3%	-1%	-8%	-11%	-2%	22%
	(Water Only)	–	0.681	0.952	0.680	0.000	0.493	0.260	0.947	0.918	0.193	–

Hurricane Verification Toolkit

- The hurricane verification toolkit (MET-TC) has been developed to replicate the functionality of the current NHC verification software
- Utilizes capabilities from Model Evaluation Tools (MET) software
- The MET-TC Code consists of three tools:
  - **tc_dland**: computes gridded field consisting of distance to land
  - **tc_pairs**: compares ADECK and BDECK tracks, computes pair statistics
  - **tc_stat**: reads output from tc_pairs, applies user-selected filtering

es summary statistics

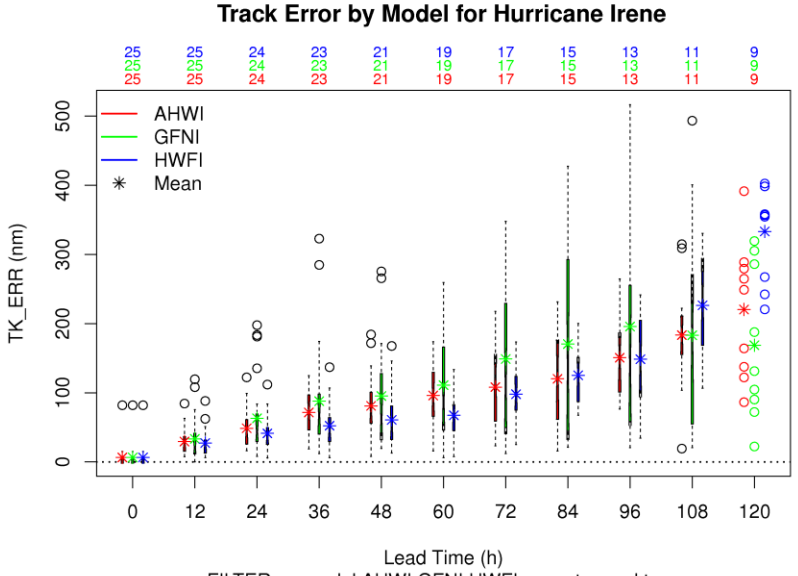


Plot showing the distance to land

Hurricane Verification Toolkit

- Pair statistics can be generated on independent model data or user-specified consensus forecasts
- Filtering options replicate those in NHC VX code
  - watch/warning in effect, over water only, hours prior to landfall, wind threshold, etc…
- Computes basic error statistics as well as frequency of superior performance and serial correlations

- Graphical capabilities included with release
- Planned official release with METv4.1 *(Jan 2013)*



HWRF Testing

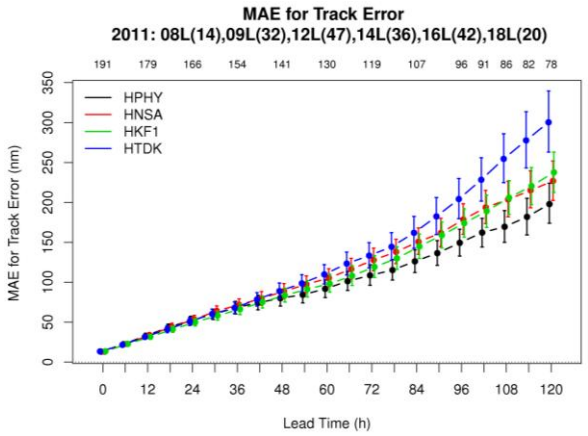
- 5.1.1 Develop operational HWRF Test plan (EMC, NHC)

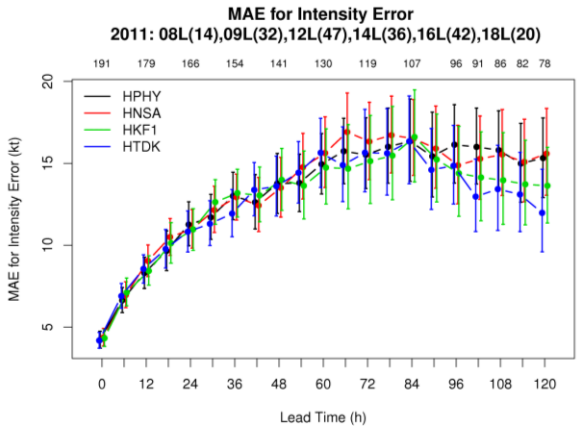
	Baseline (H130)	Physics upgrades					Combined (H213)
		PBL2 (H131)	Meso-SAS (H132)	RRTMG (H133)	MP (H134)	Ocean (H135)	
Description	Revised init/GSI New nest parent interpolations Radiation bug fix Revised nest movement Increased frequency of Physics calls	Variable Ric	Meso SAS	Radiation	2 way feedback of MP species	Removal of flux truncation MPI-POM?	Baseline+ physics
Person	Qingu, In-Hyuk Sam Trahan Mingjing, Young	Young	Qingfu	Chanh	Sam	Zhan/DTC/ URI/HRD	All
Cases	Whole 2012 storms	Priority cases	Priority cases	Priority cases	Priority cases	Priority cases	2010+2011+ 2012 all storms
Due date	Jan. 15	Jan. 15	Jan. 15	Jan. 15	Jan. 15	Jan. 15	March 15.
Platform	Jet/Zeus/WCOSS*	Jet	Jet	Zeus	Zeus	Jet	Jet/WCOSS*

- 5.1.2 Report on HWRF testing activities and results (EMC, NHC, DTC)

Test of HWRF sensitivity to cumulus schemes

- Test stemmed from discussions in the 2011 HFIP Reg Model Phys Workshop
- Test plan developed in collaboration with EMC





Tested HWRF SAS, new SAS, Tiedtke, Kain-Fritsh

HWRF SAS performs best for track; differences in intensity have little statistical significance

Track

	12	24	36	48	60	72	84	96	108	120
HNSA										
HKF1										
HTDK										

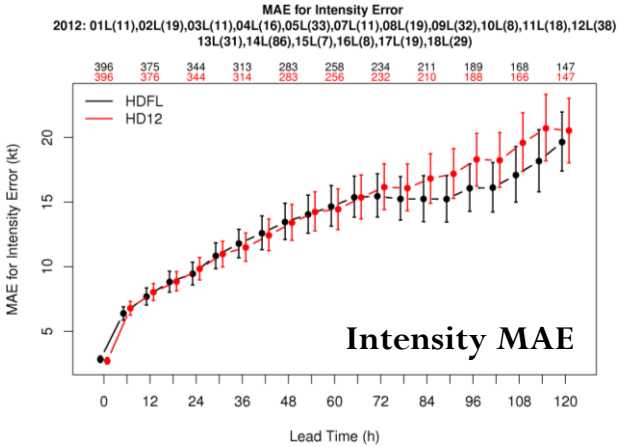
Intens

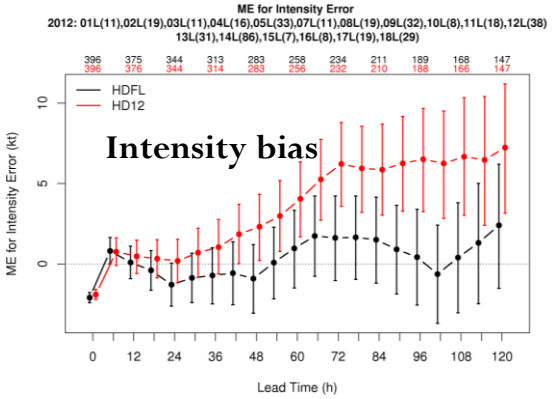
	12	24	36	48	60	72	84	96	108	120
HNSA										
HKF1										
HTDK										

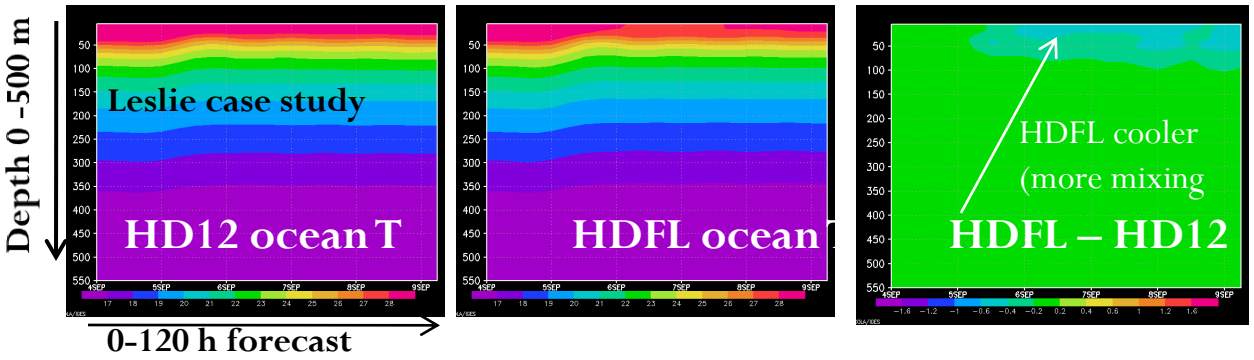
Statistical Significance 95%
Green= HWRF SAS better
Red = HPHY SAS worse

Test of HWRF momentum flux transfer atmos -> ocean

- Test stemmed from diagnostics of HWRF ocean response by HRD
- Test plan developed in collaboration with EMC, URI, and HRD







Artificial momentum flux reduction present in operational HWRF removed for test

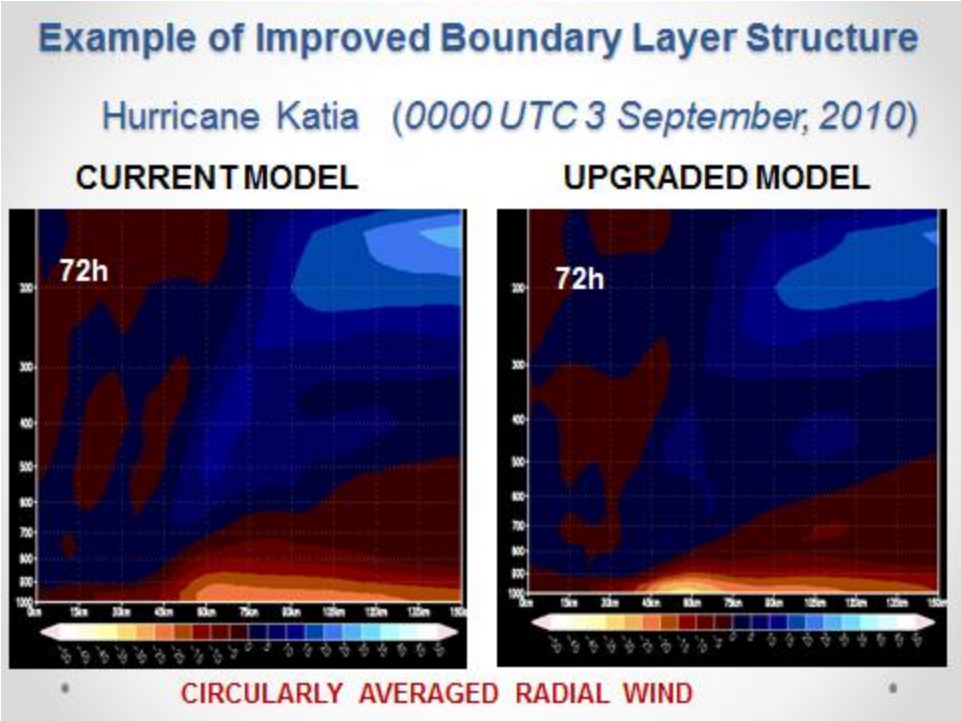
DTC tested entire 2012 season, AL and EP

Results indicate substantial decrease of positive bias in AL

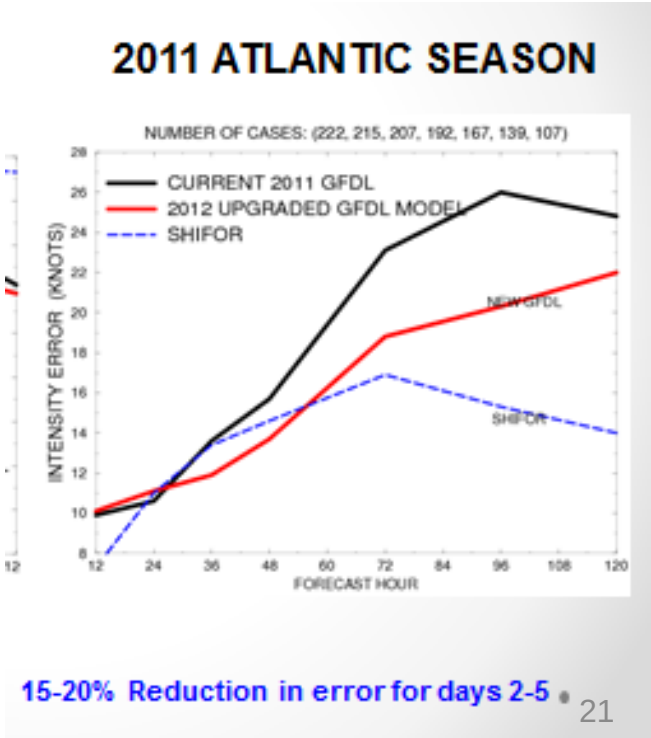
Results will be taken into account for 2013 configuration

Non-linear effects make interpretation complex: additional flux makes ocean cooler, which reduces storm intensity, which leads to reduced fluxes…

- 5.1.3 Pre-implementation testing and evaluation of GFDL model (GFDL, NHC)

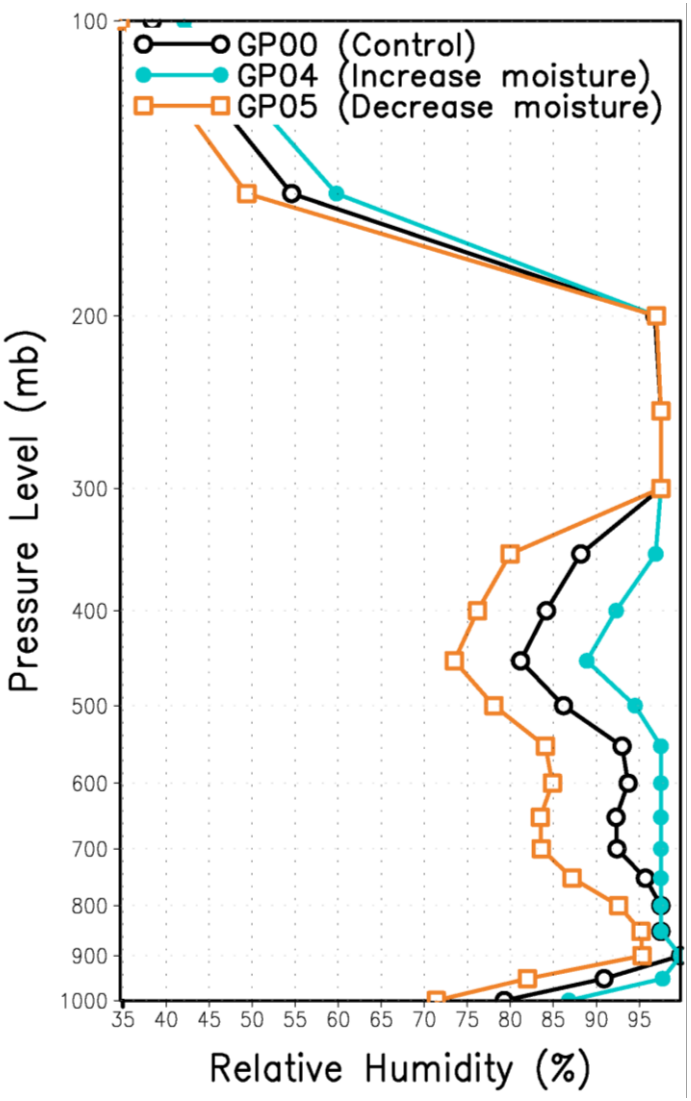


- Improved PBL structure; improved formulation of surface exchange coefficients (ch, cd); implementation of GFS shallow convection scheme; a number of bug fixes.

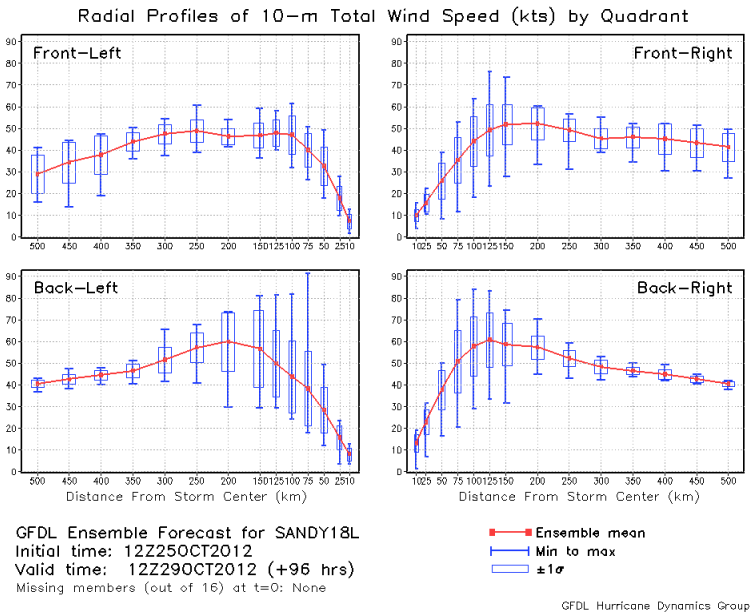


- Upgrades led to large reductions in intensity forecast error, and smaller reductions in track forecast error.
- Implemented into NCEP ops in May

- 5.1.4 Test and evaluate GFDL ensemble for possible inclusion as a Stream 1.5 model for the 2012 Demo (GFDL, TCMT, NHC)



- Perturbations include modifications to storm intensity and structure; near-storm moisture; and near-storm SST.
- Half of the perturbations run with GFS background field, half with GEFS background field.
- Run as a Stream 1.5 system in 2012.

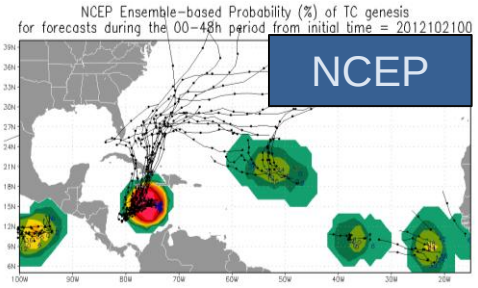


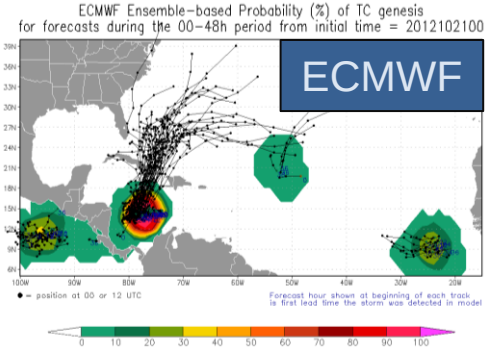
5.3.4 Run new multi-ensemble based genesis products in real time during 2012 season

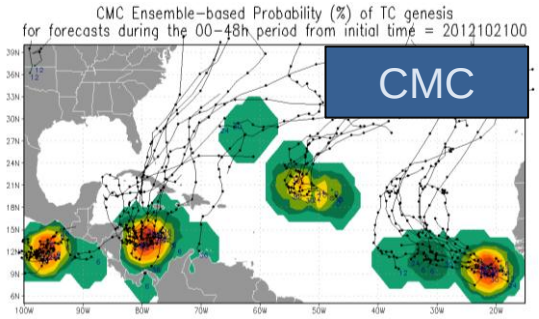
http://www.emc.ncep.noaa.gov/gmb/tpm/emchurr/tcgen/

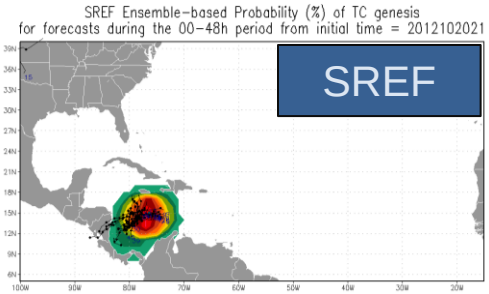
- Probabilities based on:
  (1) Global ensembles (NCEP, FNMOC, CMC, ECMWF)
  (2) Regional ensembles (SREF)
  (3) Consensus of global ensembles
  (4) Consensus of global deterministic models (GFS, NOGAPS, CMC, ECMWF)

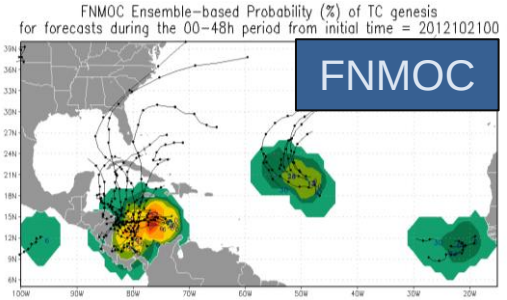
Example: 2012102100 (~36h prior to genesis of Sandy & Tony)

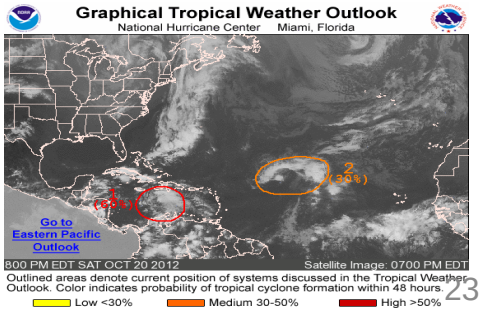






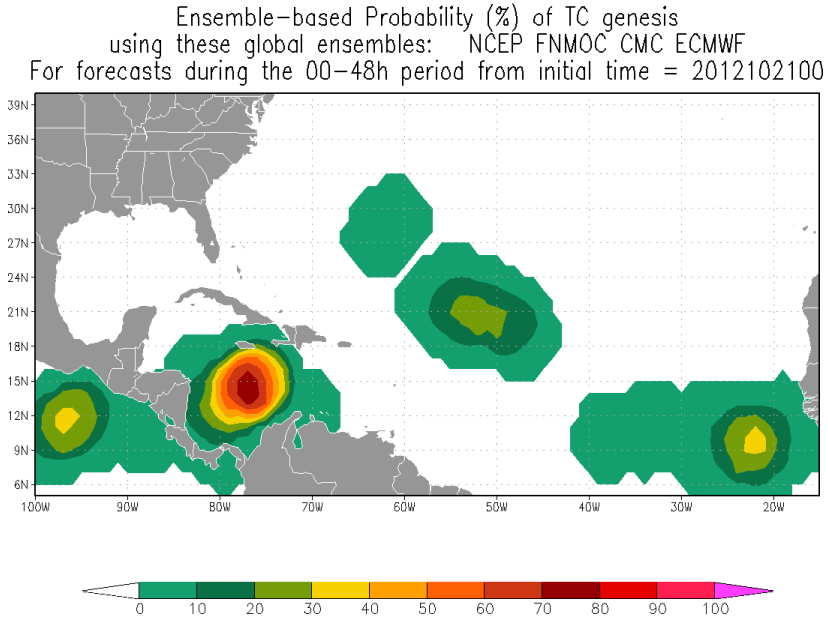




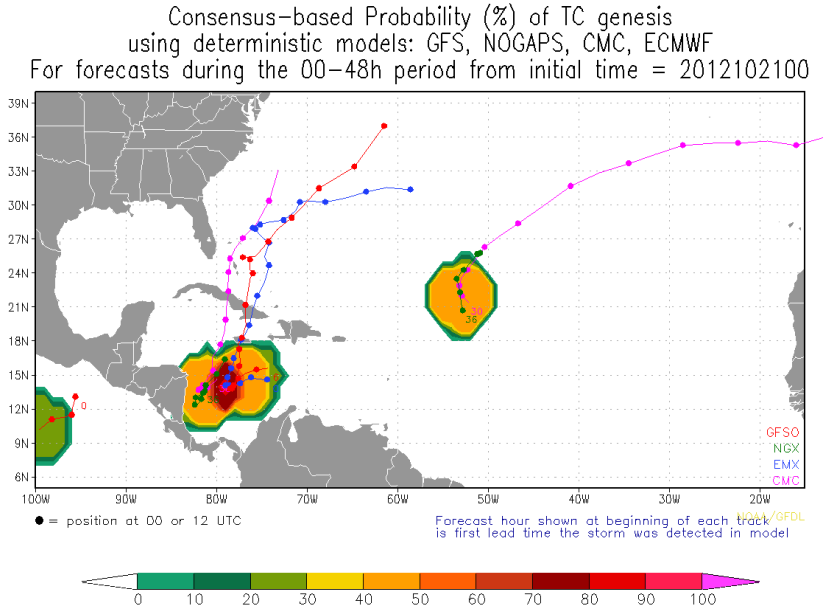


5.3.4 Run new multi-ensemble based genesis products in real time during 2012 season

Example: 2012102100 (~36h prior to genesis of Sandy & Tony)

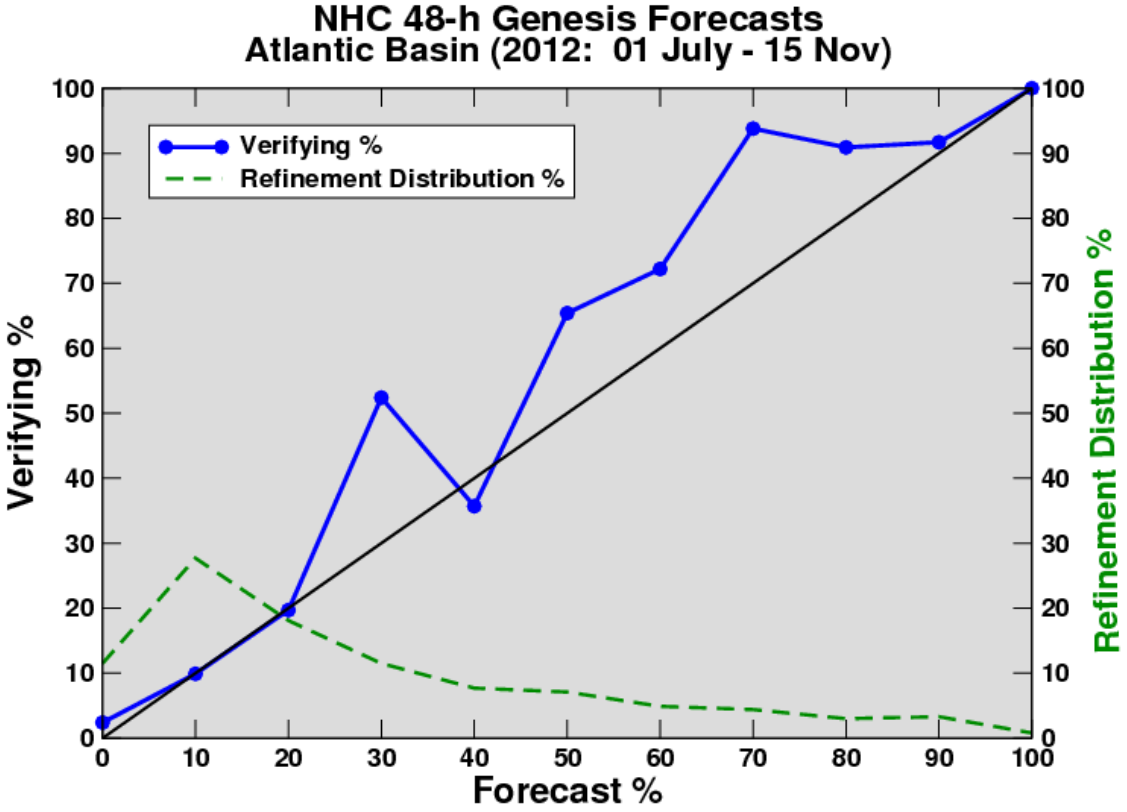


Probabilities based on a consensus of the global ensembles



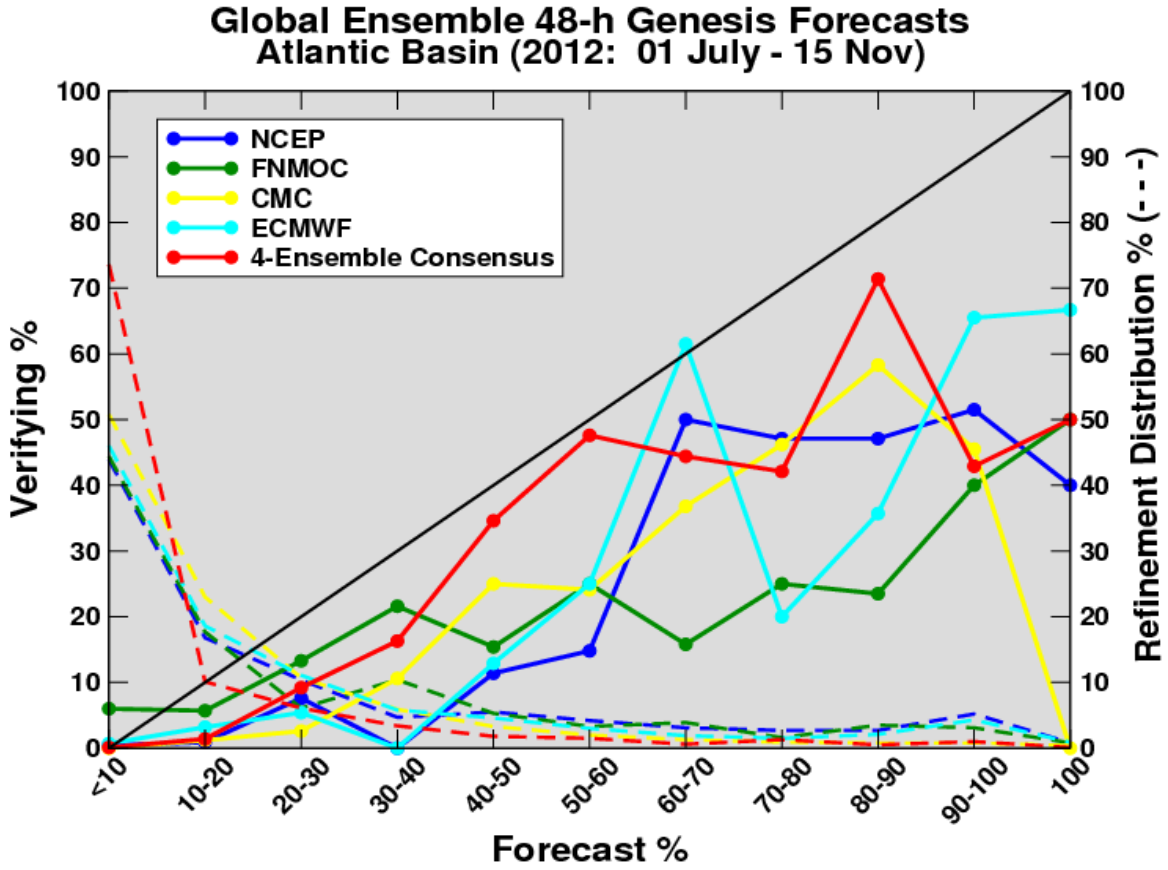
Probabilities based on a consensus of the global deterministic models

5.3.5 Perform verification of model genesis for operational global models



- Reliability of NHC official forecasts: For 2012, reliable for low probability forecasts, but then a slight bias towards underforecasting at higher forecast probabilities.

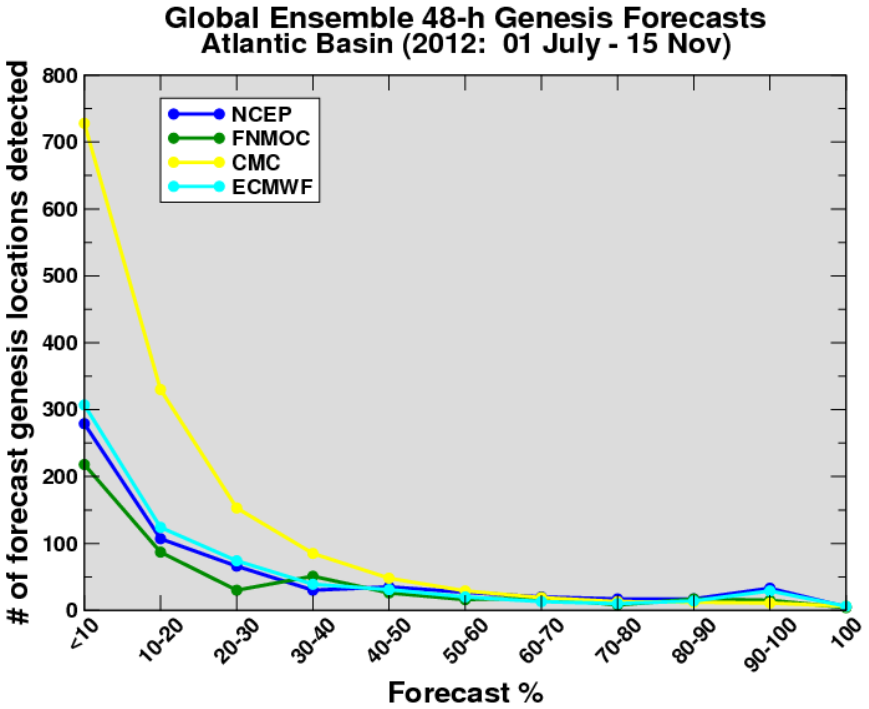
5.3.5 Perform verification of model genesis for operational global models



- All models have a bias towards over-prediction, caused by both false alarms as well as genesis occurring in the forecast long (>>48h) before observed genesis.
- 4-ensemble consensus close to reliable up through 50-60%.
- Reliability diagram fails to convey the barrage of low-probability false alarms from the CMC ensemble.

5.3.5 Perform verification of model genesis for operational global models

- CMC has an issue with spinning up a huge number of false alarms.
- NCEP & ECMWF ensembles were very similar in the Atlantic in 2012 in their climatology of producing storms.

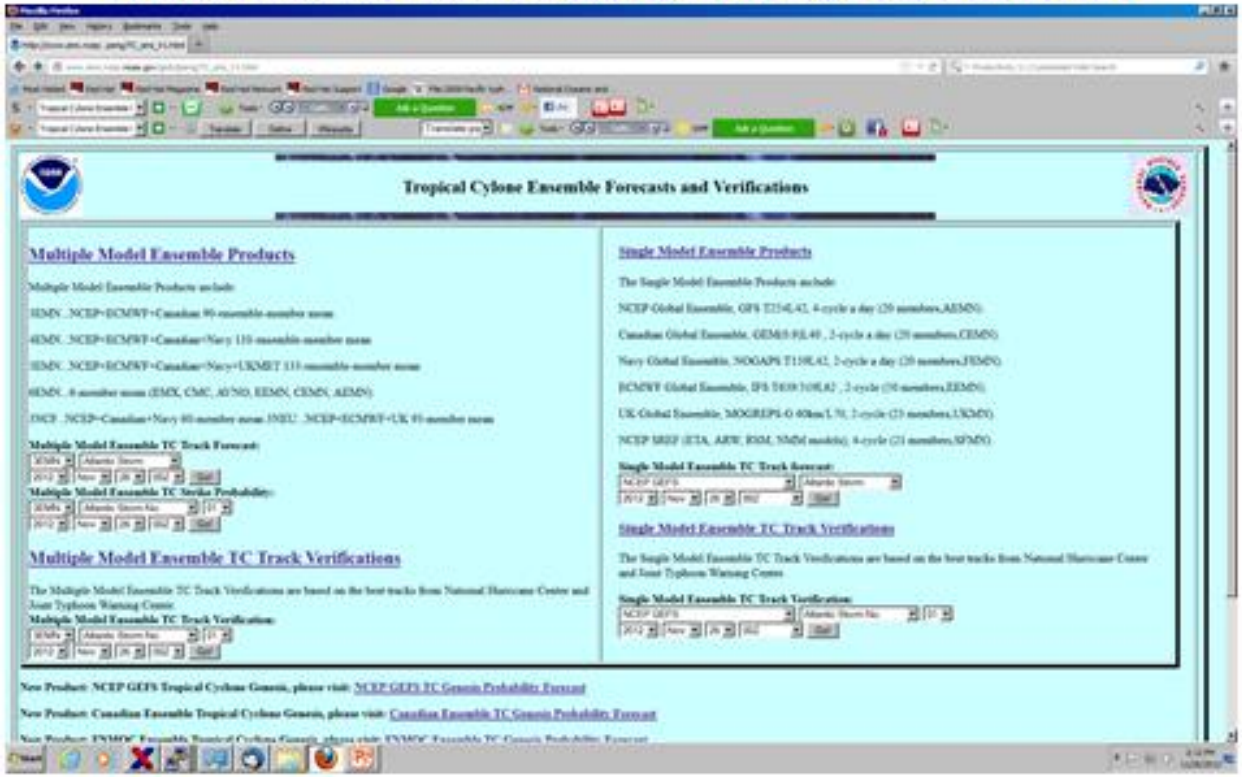


- There were often issues with all models with tracks & probabilities being erroneously triggered in the SW Caribbean Sea near Panama. This was especially the case with the FNMOC ensemble. Tracker adjustments will be tested to help alleviate this issue.
- Current forecast genesis determination is made via a combination of CPS Parameter B and low-level CPS warm core values. The season has been rerun using the additional upper-level CPS warm core criterion, and also using a simple non-CPS warm-core check alone. Analysis of these results will follow.

5.4.1 Run latest version of tracker in parallel with upgrades that include thermodynamic phase determination, tracking for SREF, FNMOC, NAEFS and 12Z ECMWF Ensembles.

http://www.emc.ncep.noaa.gov/gmb/jpeng/TC_ens_V1.html

- Jiayi Peng (EMC) is producing single-model ensemble track output, as well as multiple-ensemble track output (see website at URL to the right).
- Forecast tracks used in real time by JTWC in 2012.

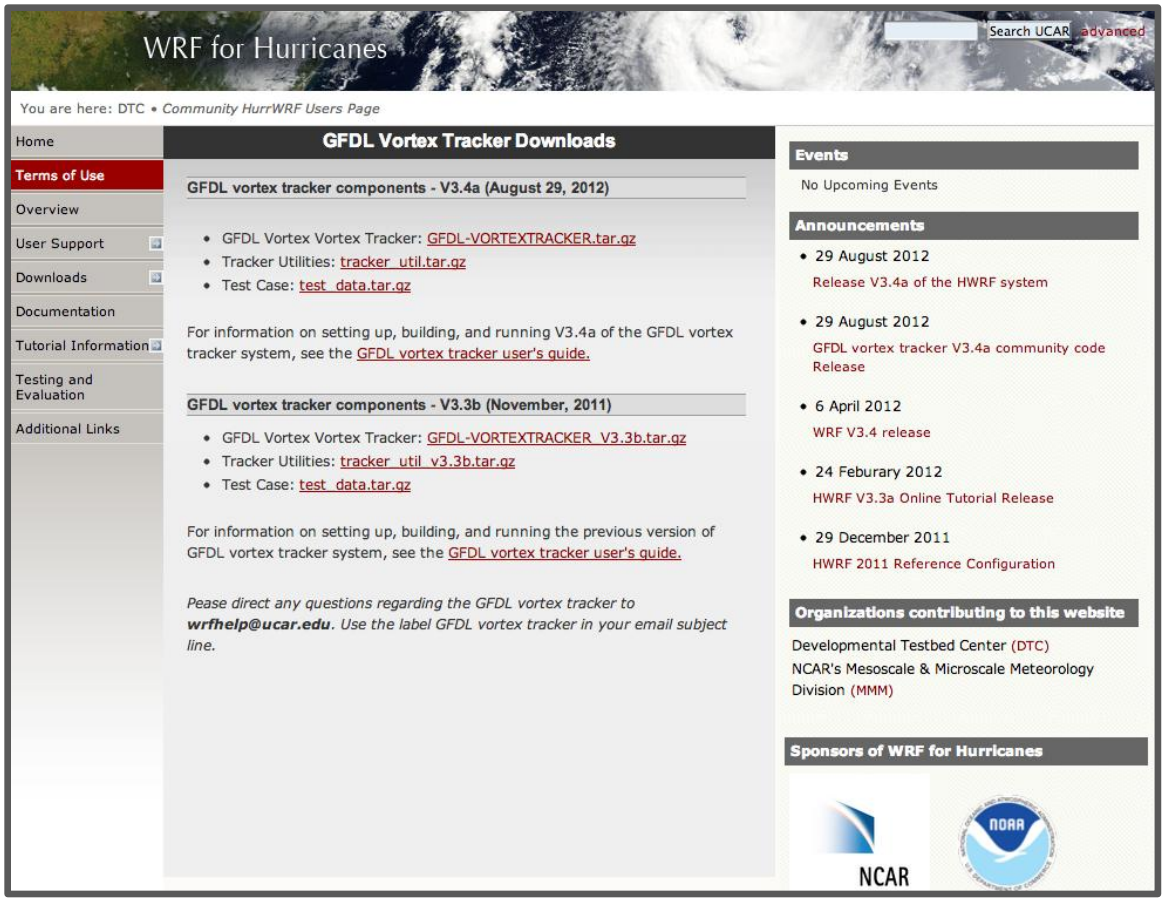


Real-time tropical cyclone genesis and track ensemble forecasts

Data also available at: ftp://ftp.emc.ncep.noaa.gov/gc_wmb/jpeng/

www.dtcenter.org/HurrWRF/users

GFDL vortex tracker community release



GFDL Vortex Tracker v3.4a was released in August 2012

It is part of the HWRF release and is also released as stand alone (can be used by other models)

User support, test datasets available

New supported capabilities include diagnosis of cyclone thermodynamic phase as well as use of the tracker in genesis detection & tracking mode.

TC Verification Team Report – Contributions from EMC

Stream 1 (5.1.1/5.1.2)

New verification capabilities include:

1. Verification for RMW and PMIN, 6-h intensity change, P-W relationship

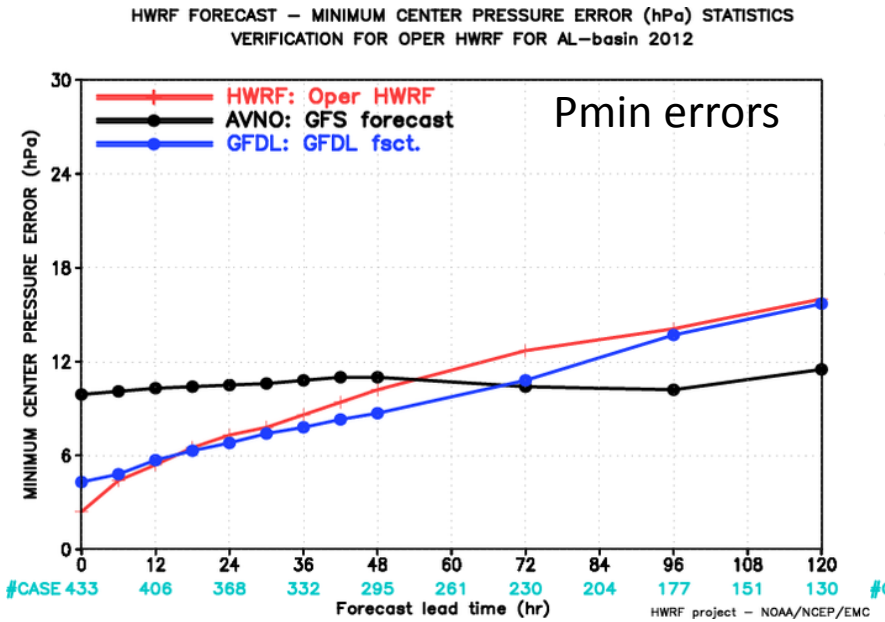
2. Included confidence intervals for all the track/intensity/radii plots

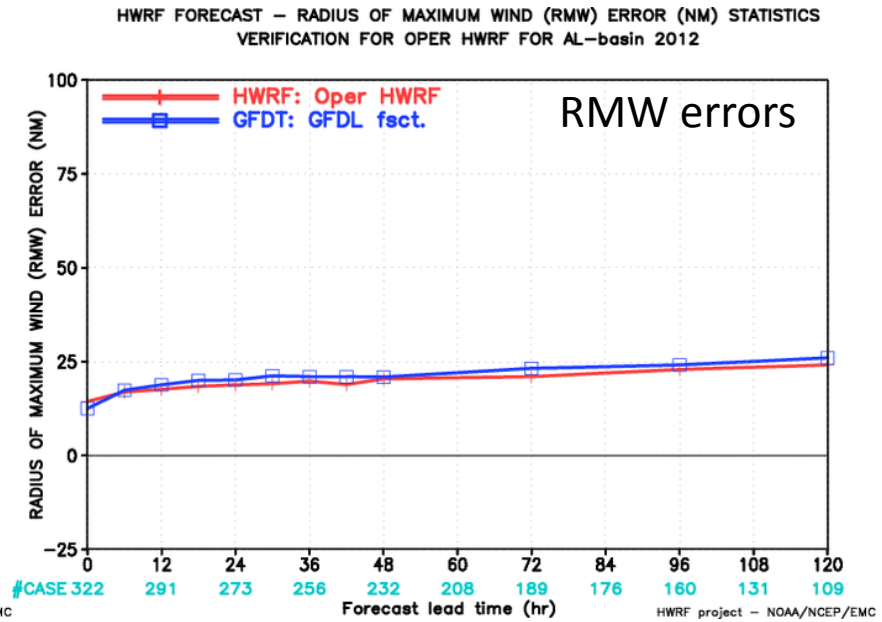
3. Stratified verification with respect to strong versus weak storms, land or ocean points

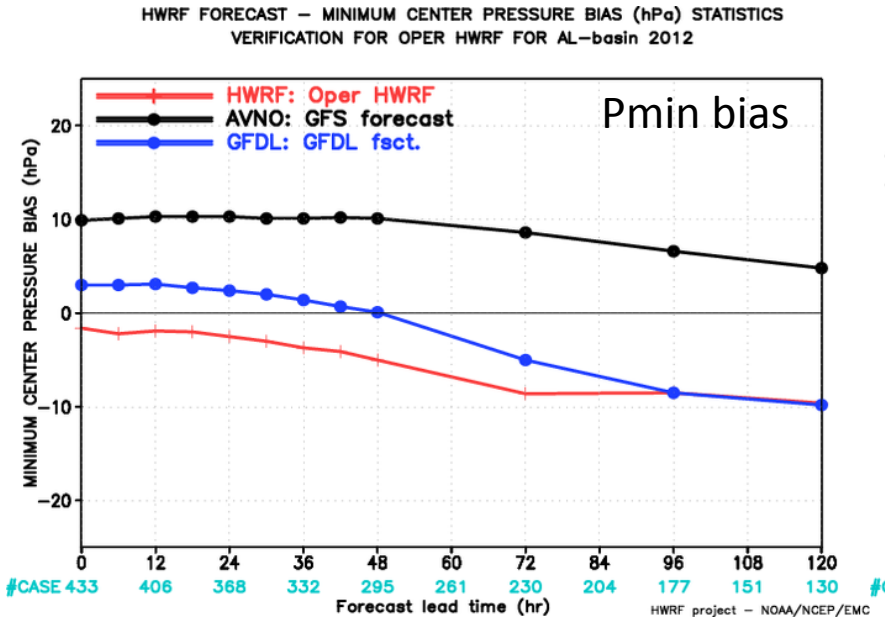
4. Added graphic capability for 34, 50, and 64-kt radii verification and along/across track verification

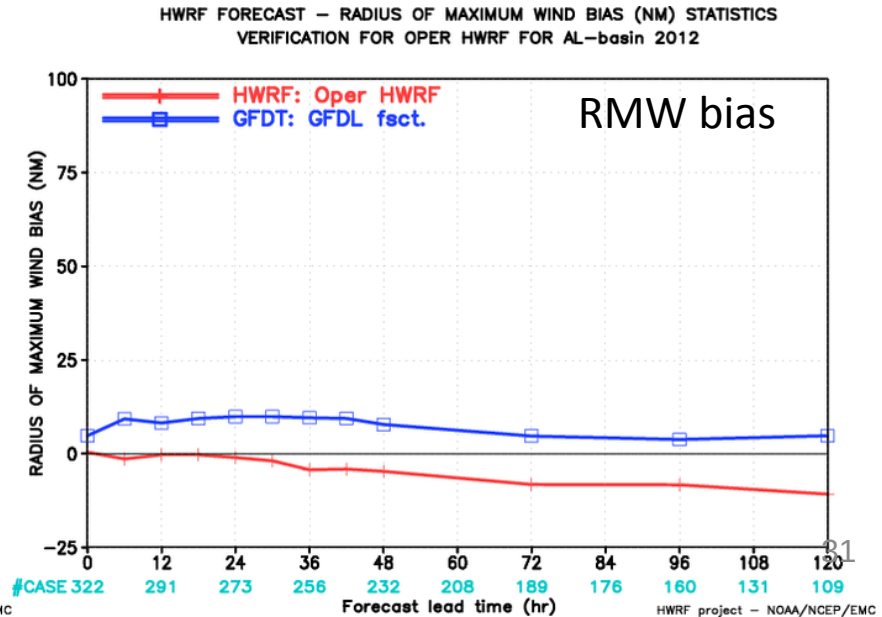
5. Added capability to verify Western Pacific and Indian Ocean storms

RMW/PMIN verification for 2012

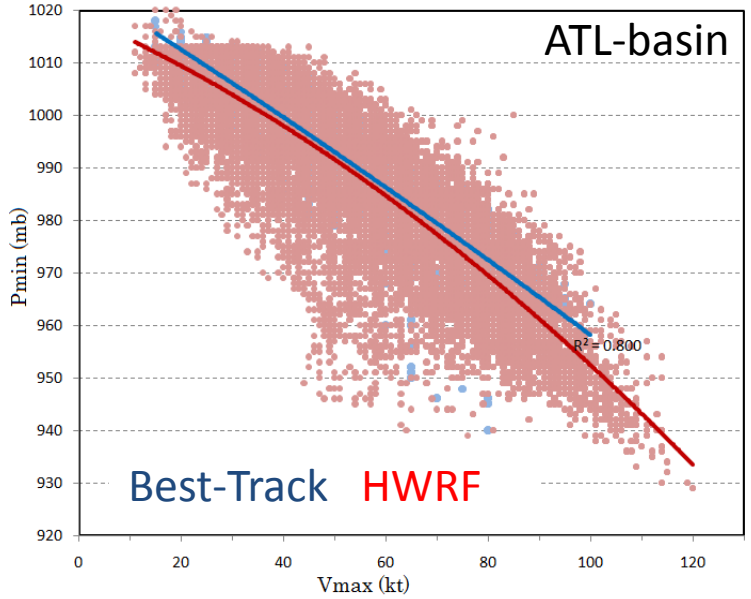


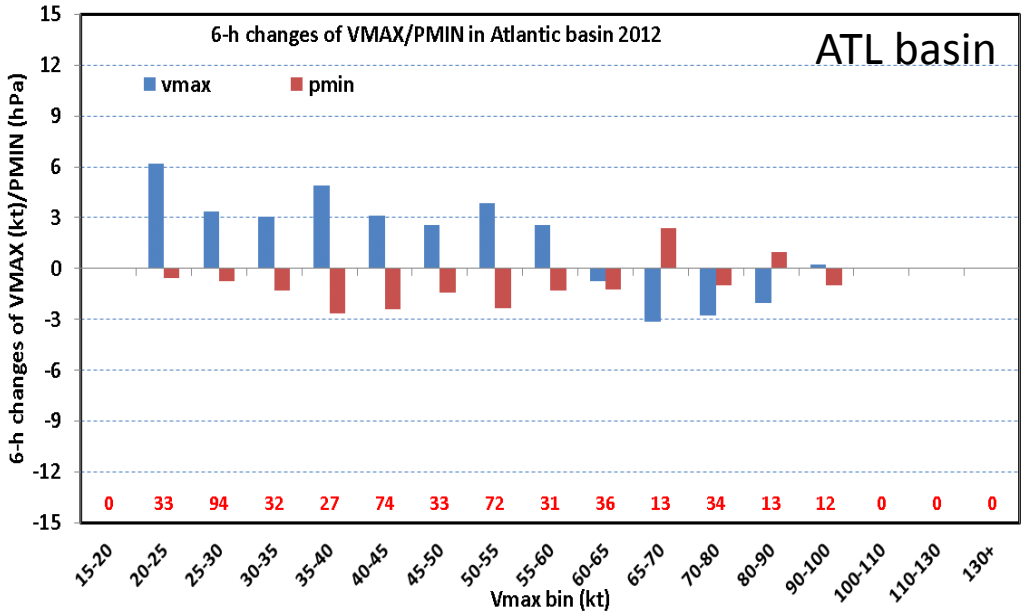


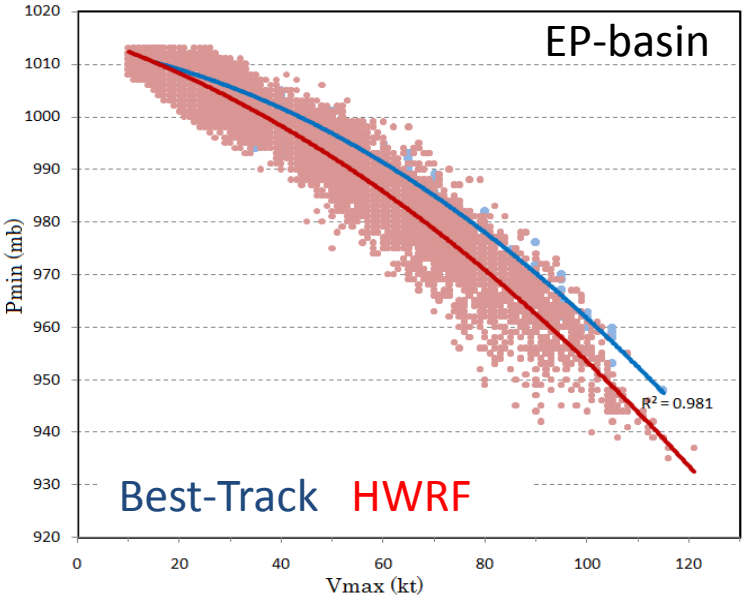


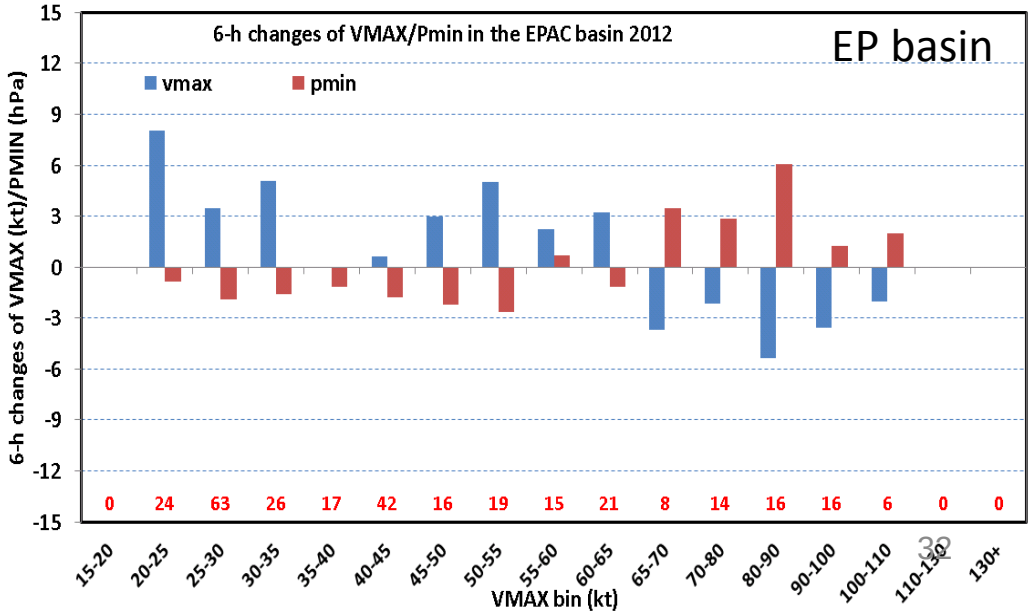


Wind-Pressure relationship, 6-h intensity change





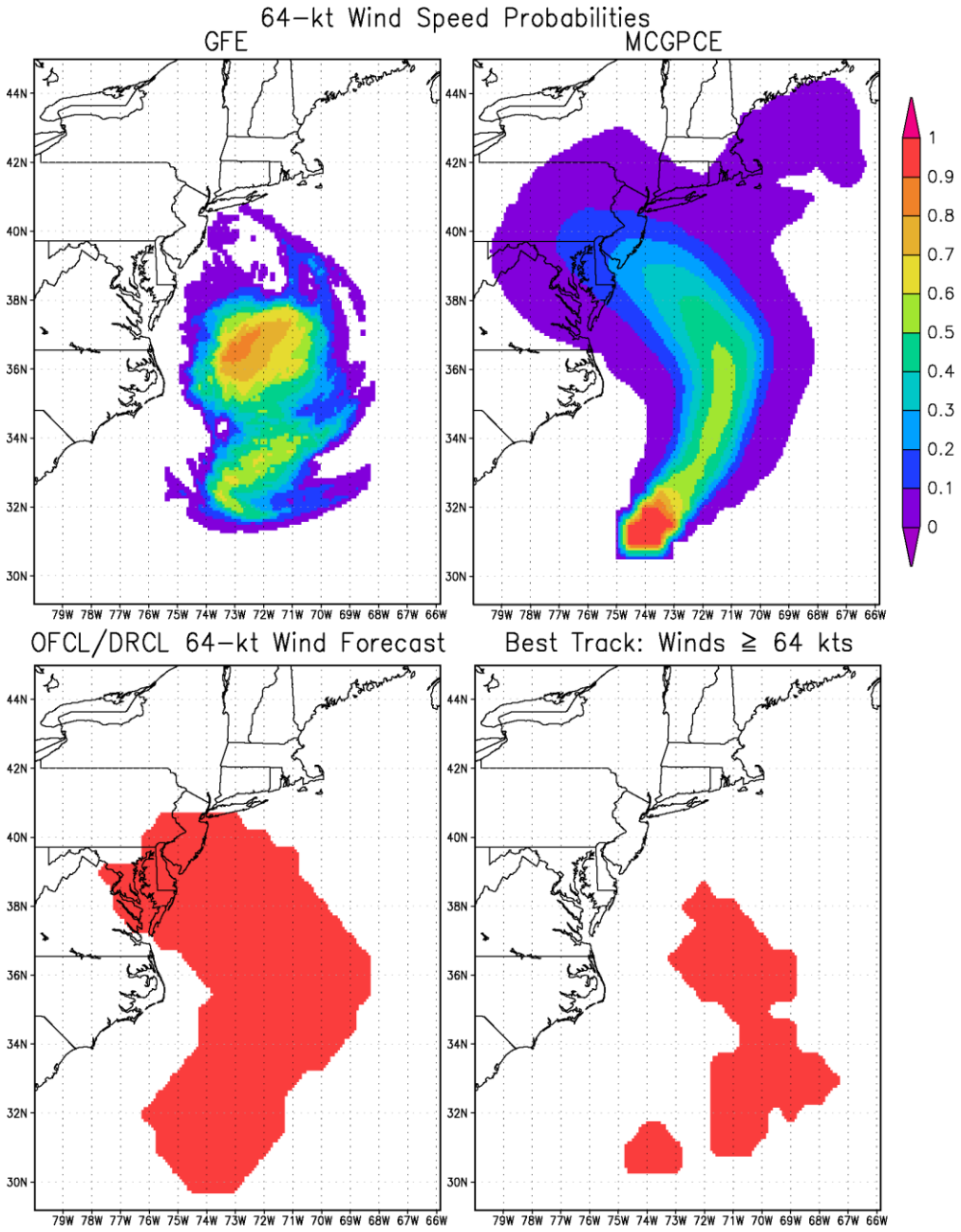




Additional verification work

- **Verification of wind speed probabilities** generated with output from a regional ensemble model (Matt Morin / GFDL. Thanks to Mark DeMaria, Andrea Schumacher and John Knaff for their collaboration and assistance).

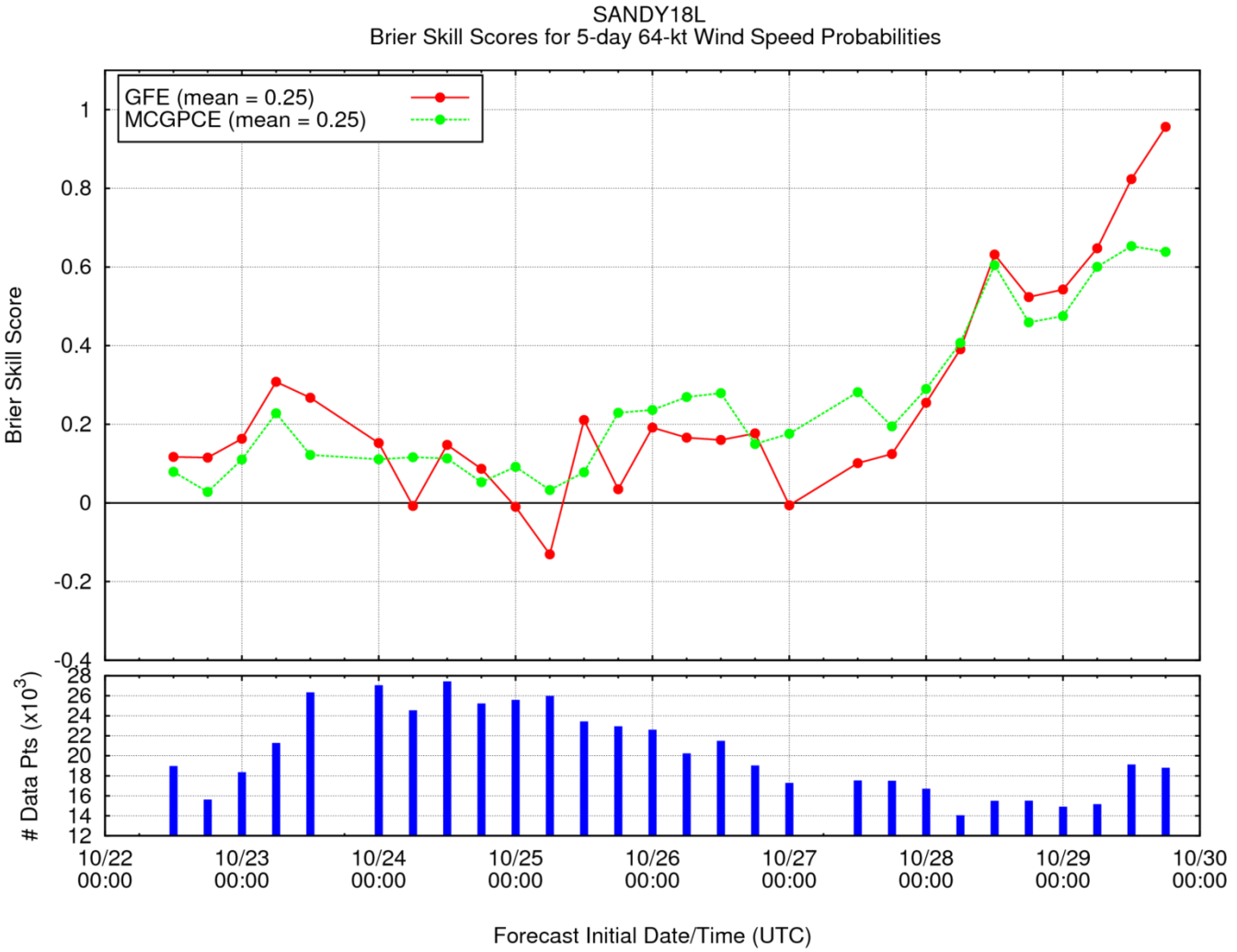
- Follows same techniques as, and offers comparisons with, methods used to verify forecasts from DeMaria & Knaff's Monte Carlo probability model.



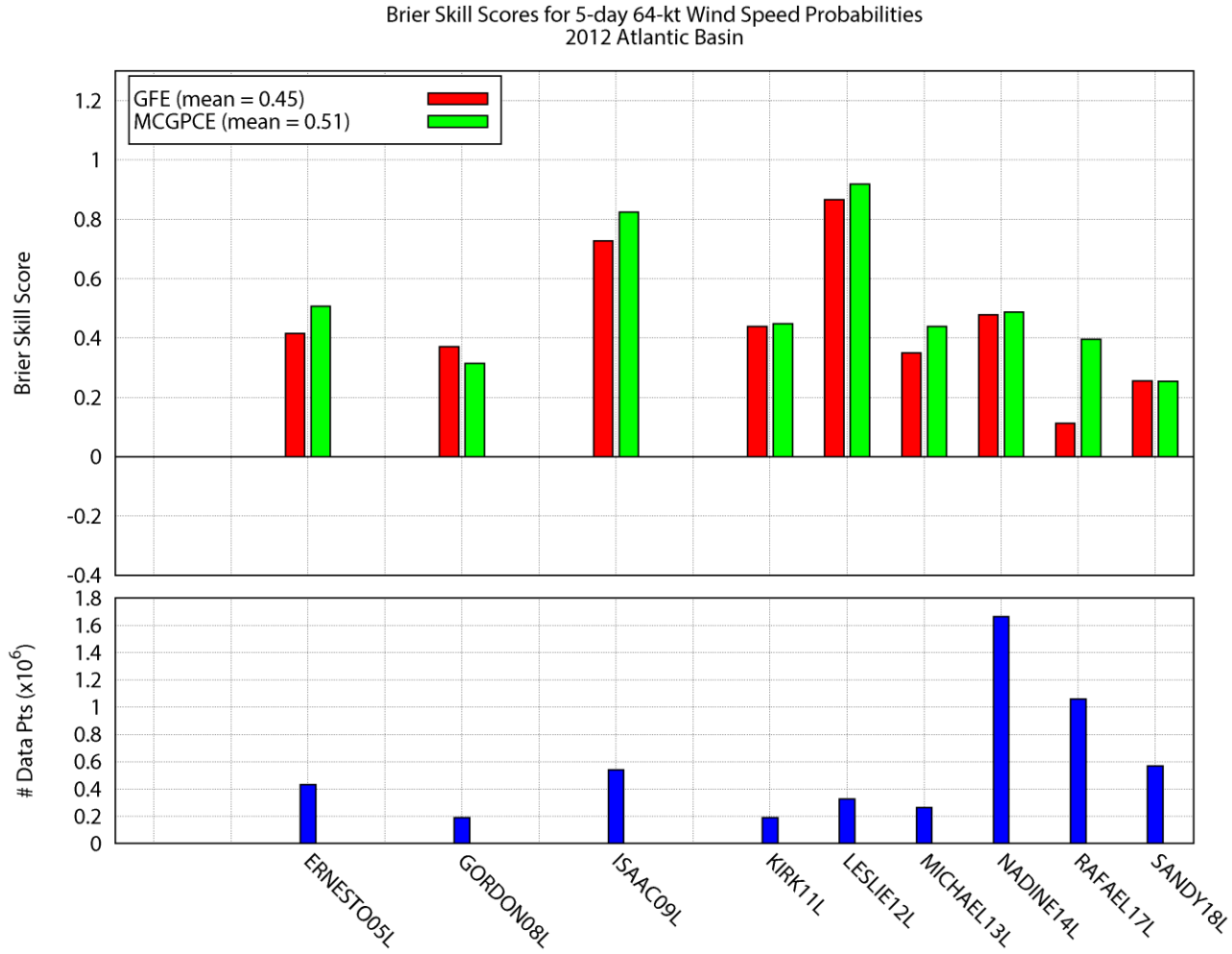
Sandy (18L) Init: 2012102812

Additional verification work

- For Sandy, Brier Skill Scores for 64-kt wind speed probabilities are comparable between the GFDL ensemble and the Monte Carlo model.



Additional verification work



- For 2012, looking only at hurricanes, Brier Skill Scores for 64-kt wind speed probabilities are lower for the GFDL ensemble (0.45) than the Monte Carlo model (0.51), but they do show enough skill to encourage the utility of this type of wind speed probability product based on dynamical ensemble model output.

Challenges and Issues

- Use of common tracker
  - Facilitates easy comparison among multiple model results
- Estimation of forecast intensity
- Stratification of results
  - What are appropriate subsamples?
- Need for central verification activities for consistent model evaluations
- Work towards comprehensive verif suite:
  - Track & intensity first, then radii (+more 2D & 3D structure), pressure, genesis, rainfall, surge, … others?