

# HFIP Verification Team: FY11 Review

---

*HFIP Team Meeting  
Miami, FL  
08 November 2011*

- *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
- *NWS / OST*
  - Tony Eckel

# Verification Team Milestones 2011

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

# HWRF: Testing and Evaluation

## Testing and evaluation of operational models

- Pre-implementation testing of HWRF
  - Test plan (*EMC, DTC, NHC*) Stream 1 5.1.1 June 1, 2011
  - Report on testing activities and results (*EMC, DTC, NHC*) Stream 1 5.1.2 Sept 30, 2011

## Created (w/ EMC) and conducting DTC HWRF Test Plan

- January: Reference Configuration (benchmark of 27/9 community code using 2011 baseline configuration)
- July: Benchmark of 27/9 code using 2011 operational configuration (Findings led to crisis-RFC fixes)
- November: HWRF Reference Configuration (benchmark of 27/9 community code using 2011 operational configuration)
- December: Benchmark of 27/9/3 code
- December: Test of various cumulus parameterizations in HWRF



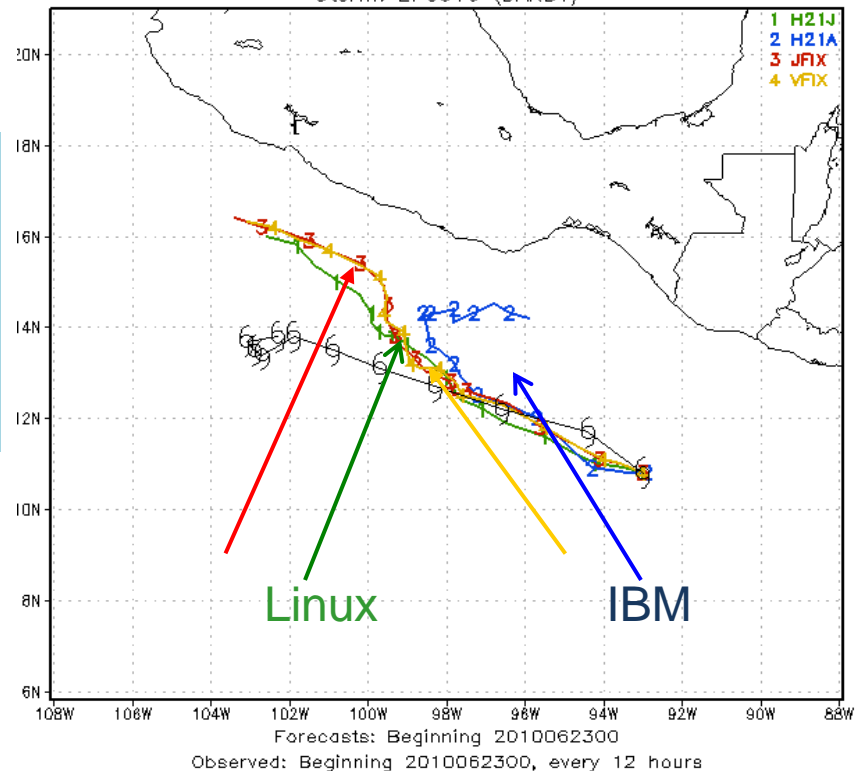
# Tests of 2011 Operational HWRF Model

## 2011 Operational HWRF model

- Testing and evaluation
- Uncovered SAS bug → crisis RFC

- Same code with bug produces different forecast on Linux and IBM.
- Fixed code produces similar forecast on Linux and IBM.

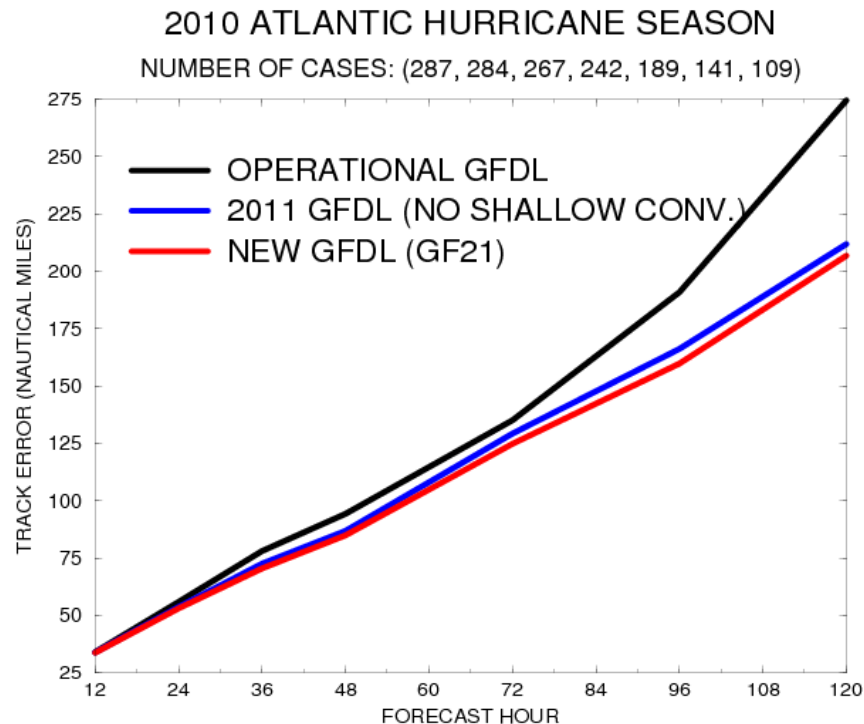
2010 Tropical Cyclone Tracks  
Storm: EP0510 (DARBY)



# GFDL: Testing and Evaluation

## 1. Testing and evaluation of first major upgrade of operational GFDL model since 2006

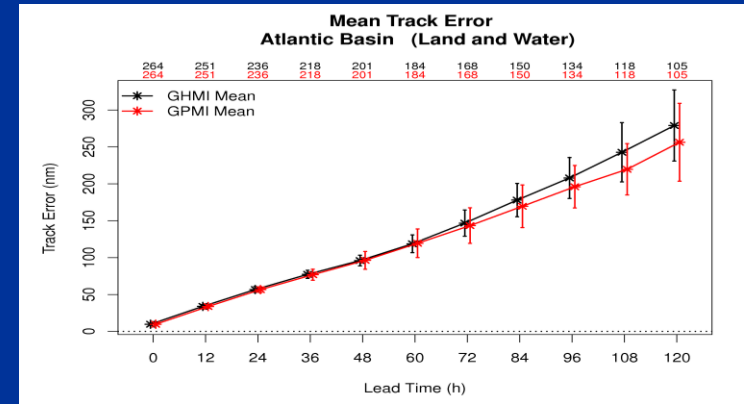
- January: New model configuration frozen.
- February: Retrospective runs from 2008-2010 completed



## 2. Testing and evaluation of GFDL ensemble for potential use as a Stream 1.5 model

# 2011 Retrospective Planning, Testing, and Evaluation

- Verification team and TCMT coordinated 2011 Stream 1.5 Retrospective Evaluation with NHC and modeling groups
- Evaluation activities coordinated with NHC
  - Use of interpolator
  - This year included performance relative to consensus, top-flight models
  - Powerful comparisons with paired tests
  - Identification of “practically” important differences
- Many types of models and groups involved – much more than 2010
- More on this Wed (Nance presentation)

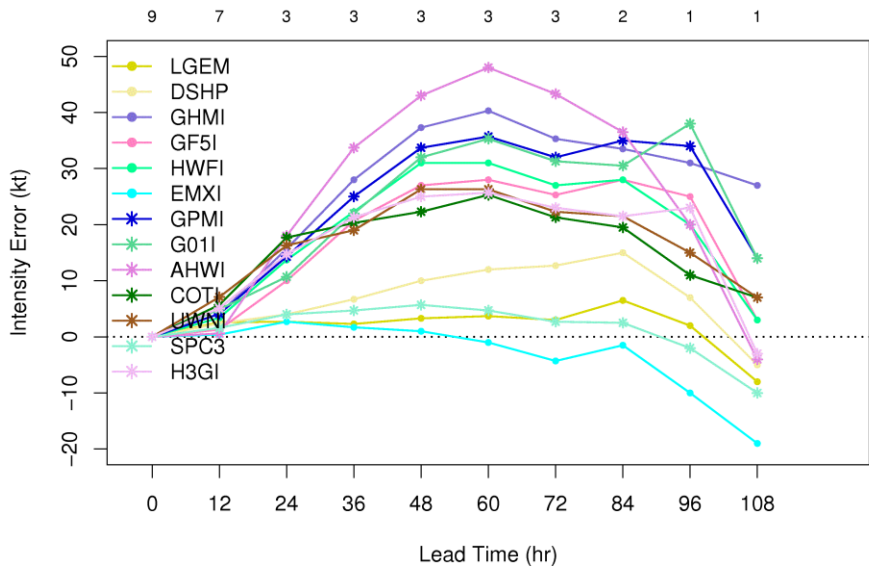


	0	12	24	36	48	60	72	84	96	108	120
0.0		5.6	10.9	16.6	15.2	19.2	14.7	20.2	1.1	12.8	15.6
0%		16%	17%	20%	15%	16%	11%	13%	1%	6%	6%
-		0.999	0.965	0.932	0.769	0.786	0.689	0.528	0.020	0.164	0.087
0.0		0.1	2.6	4.4	5.3	6.9	5.9	4.8	4.0	2.4	-0.4
0%		1%	16%	19%	19%	22%	18%	14%	12%	7%	-1%
-		0.132	0.995	0.998	0.97	0.98	0.987	0.966	0.883	0.602	0.095
0.0		2.8	5.8	16.3	11.2	9.8	11.0	1.8	17.8	109.2	74.7
0%		9%	11%	20%	12%	9%	9%	1%	9%	35%	23%
-		0.836	0.850	0.995	0.911	0.752	0.661	0.080	0.45	0.951	-
0.0		0.8	0.1	4.0	4.6	9.5	17.0	15.9	2.0	1.7	-17.3
0%		3%	0%	6%	5%	9%	13%	10%	1%	1%	-8%
-		0.343	0.025	0.531	0.481	0.477	0.625	0.481	0.051	0.050	0.419
0.0		-1.3	-1.4	-2.9	-6.4	-7.6	-9.9	-11.9	-12.9	-13.0	-10.5
0%		-16%	-11%	-18%	-39%	-43%	-54%	-66%	-72%	-64%	-47%
-		0.893	0.952	0.972	0.999	0.998	0.984	0.968	0.931	0.864	0.891
0.0		-1.7	-2.2	-4.1	-6.6	-8.9	-12.5	-14.8	-16.8	-17.7	-18.2
0%		-23%	-19%	-28%	-42%	-56%	-82%	-101%	-123%	-123%	-118%
-		0.999	0.993	0.999	0.999	0.999	0.999	0.998	0.992	0.979	0.968

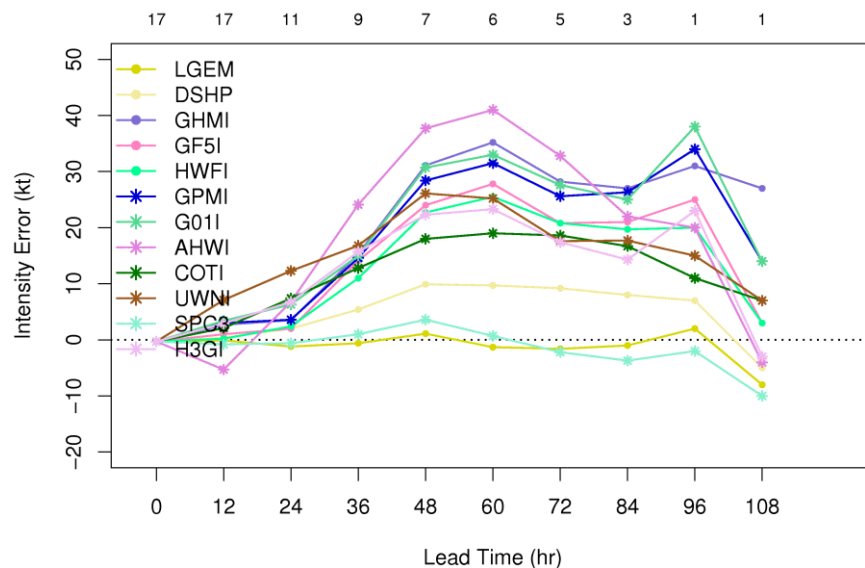


# Homogeneous Comparison of Stream 1.5 Models - Maria

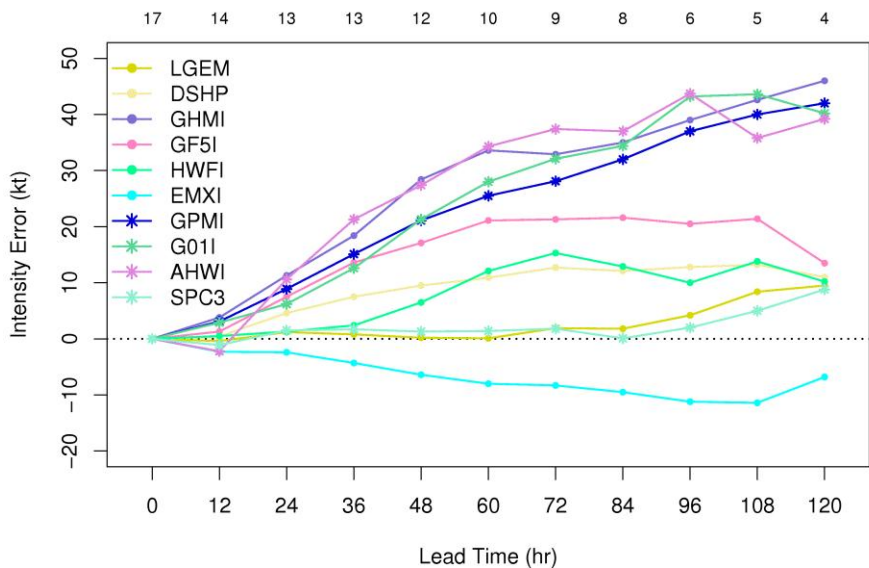
Mean Intensity Error  
Atlantic Basin (Water Only)



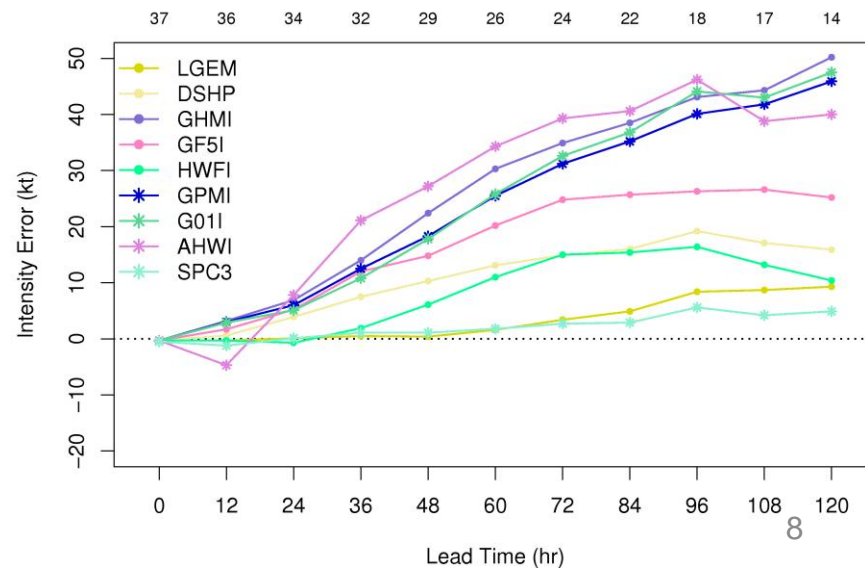
Mean Intensity Error  
Atlantic Basin (Water Only)



Mean Intensity Error  
Atlantic Basin (Water Only)

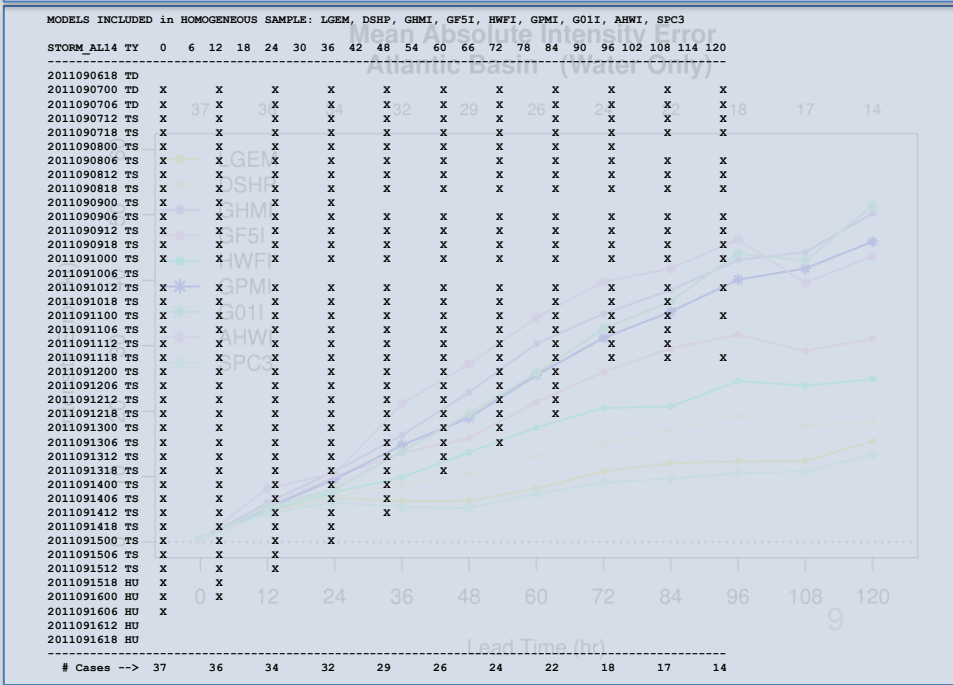
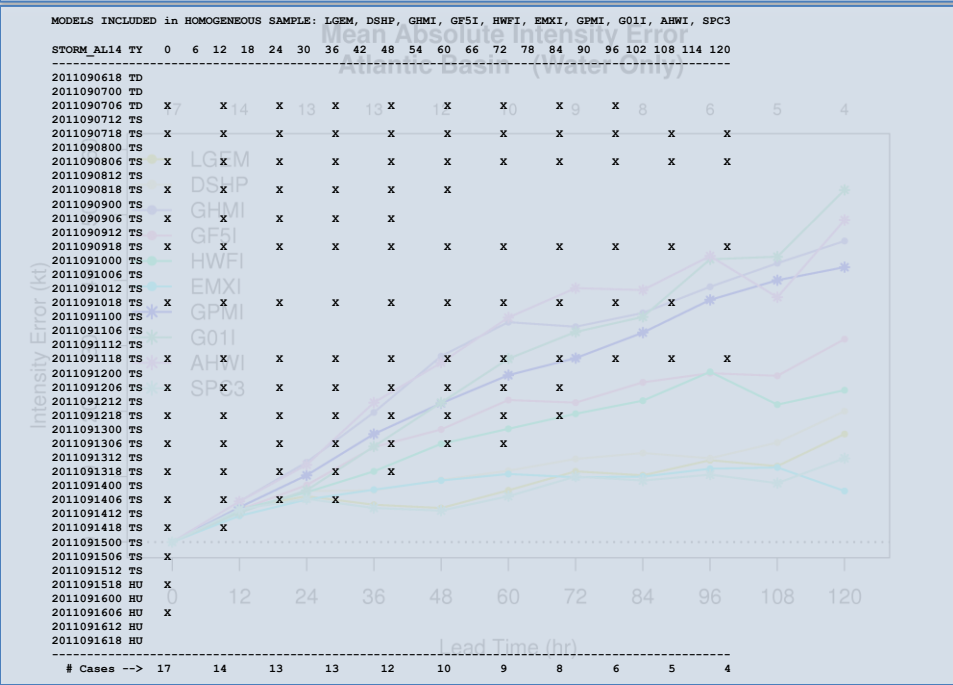
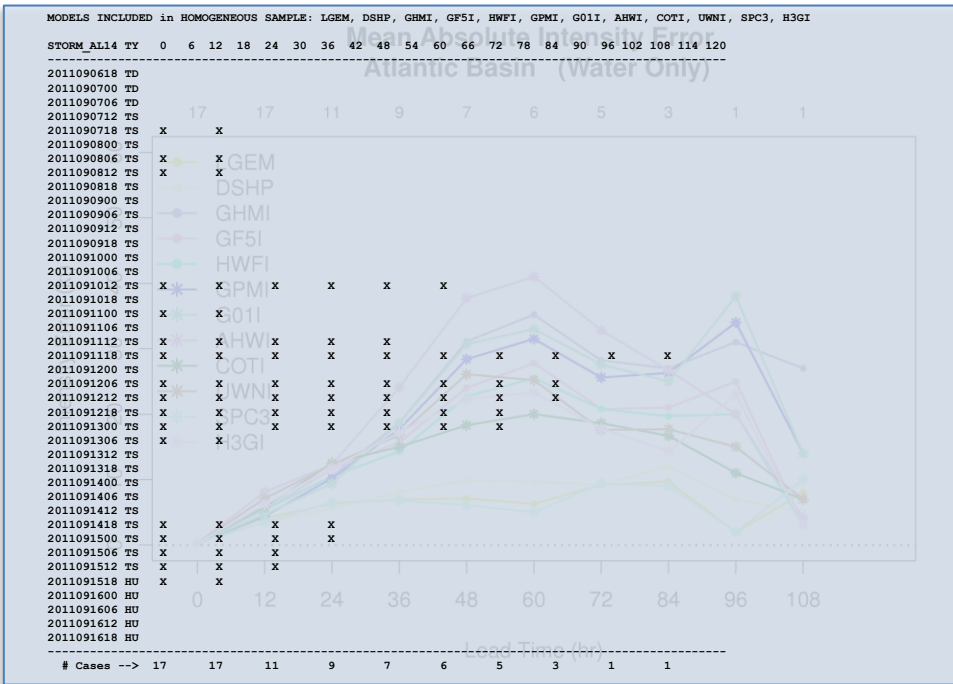
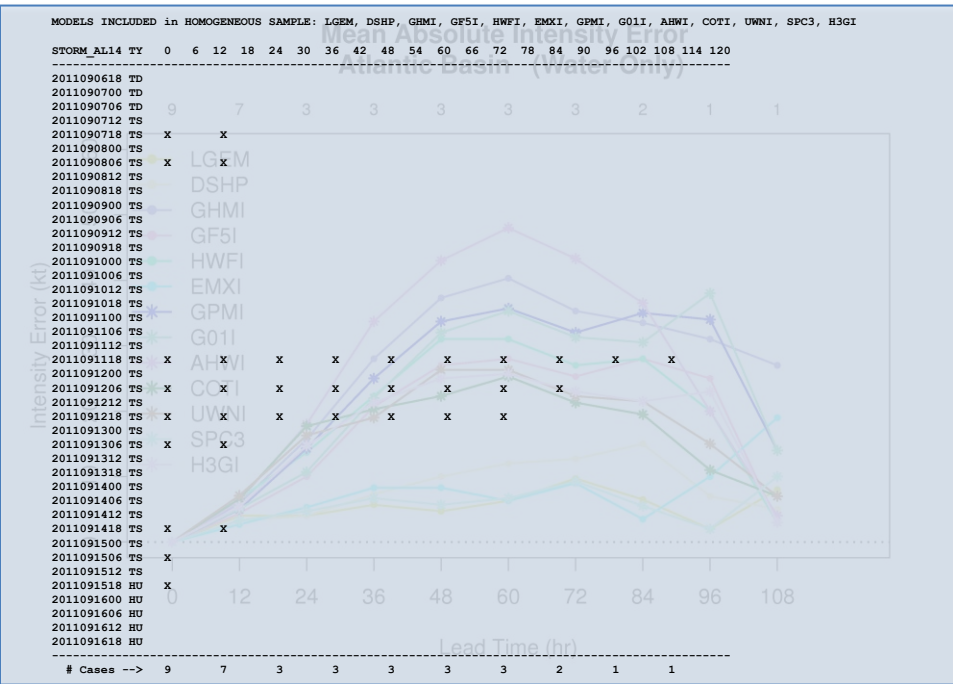


Mean Intensity Error  
Atlantic Basin (Water Only)



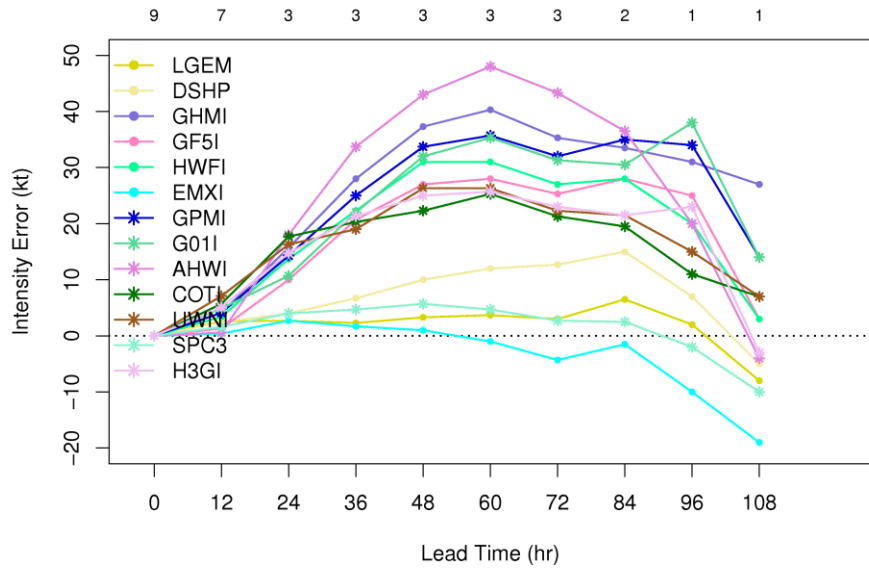


# Homogeneous Comparison of Stream 1.5 Models - Maria

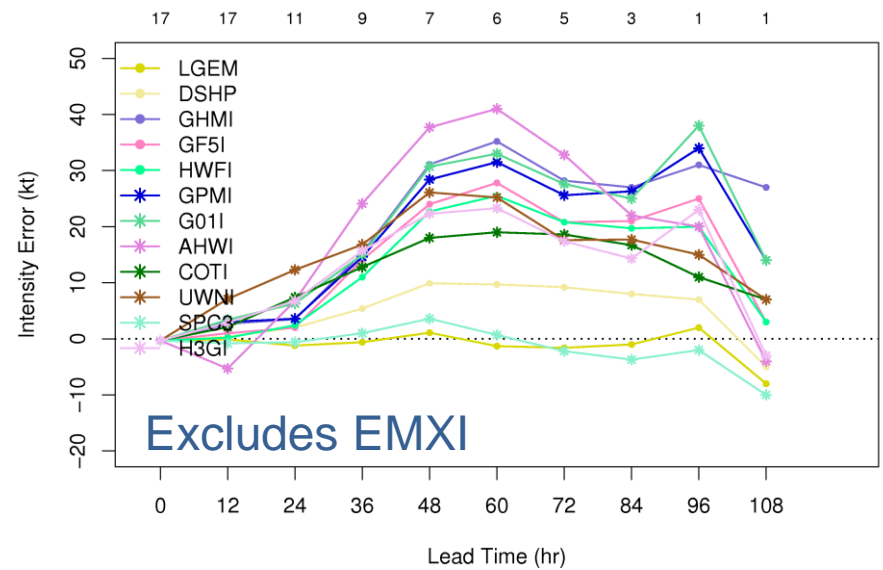


# Homogeneous Comparison of Stream 1.5 Models - Maria

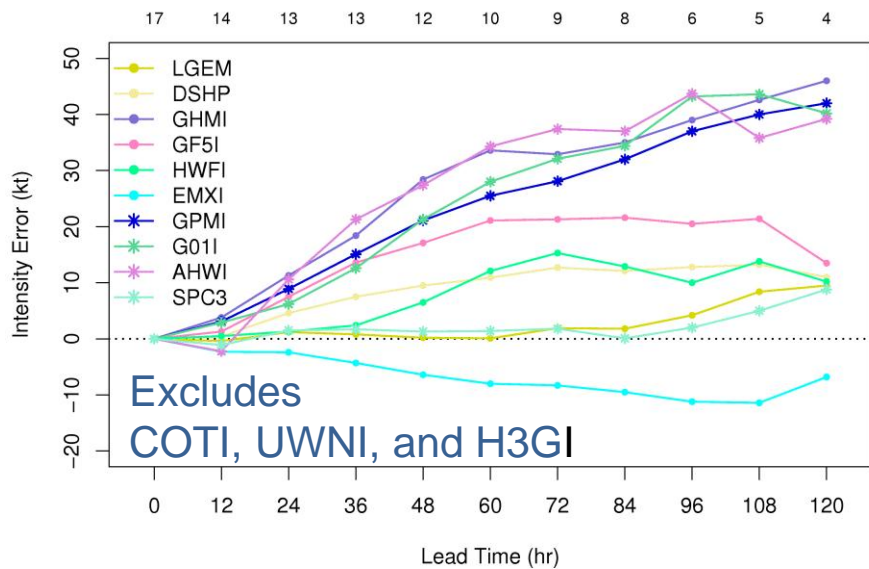
Mean Intensity Error  
Atlantic Basin (Water Only)



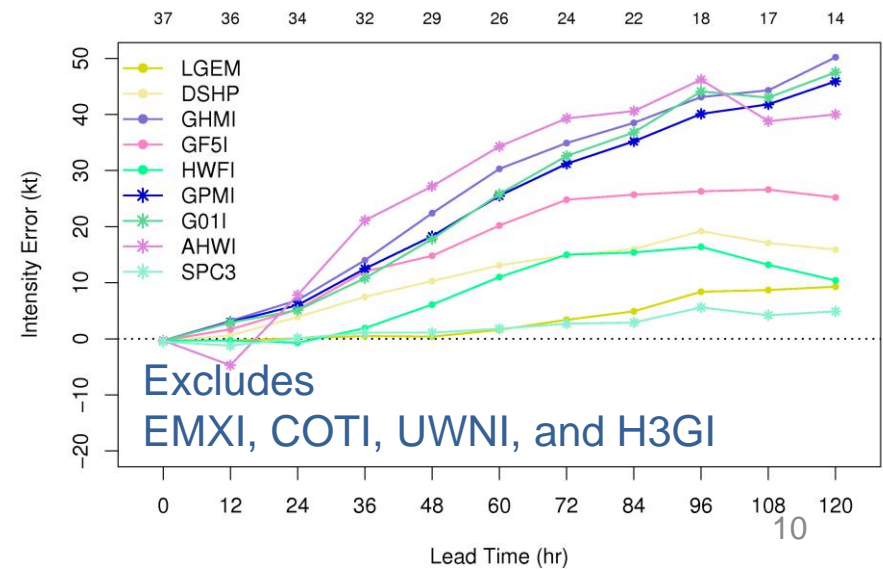
Mean Intensity Error  
Atlantic Basin (Water Only)



Mean Intensity Error  
Atlantic Basin (Water Only)

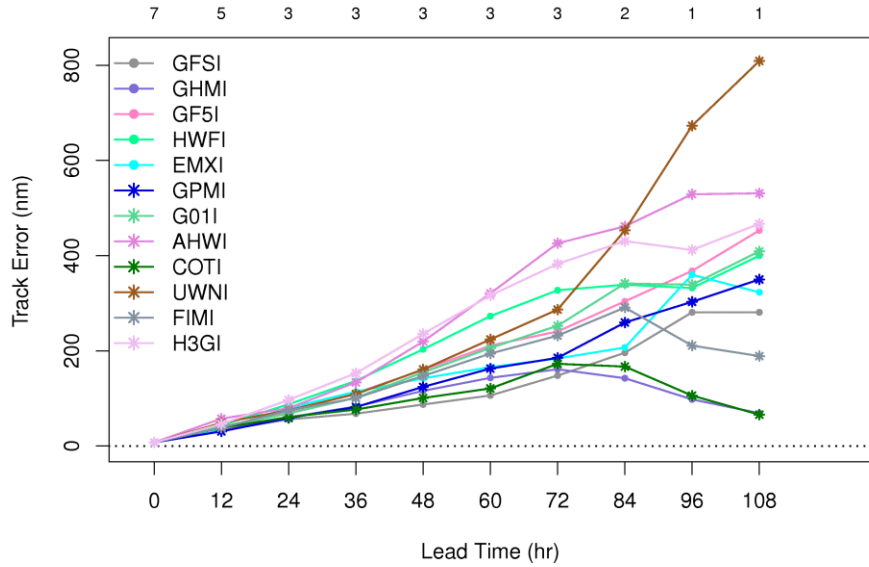


Mean Intensity Error  
Atlantic Basin (Water Only)

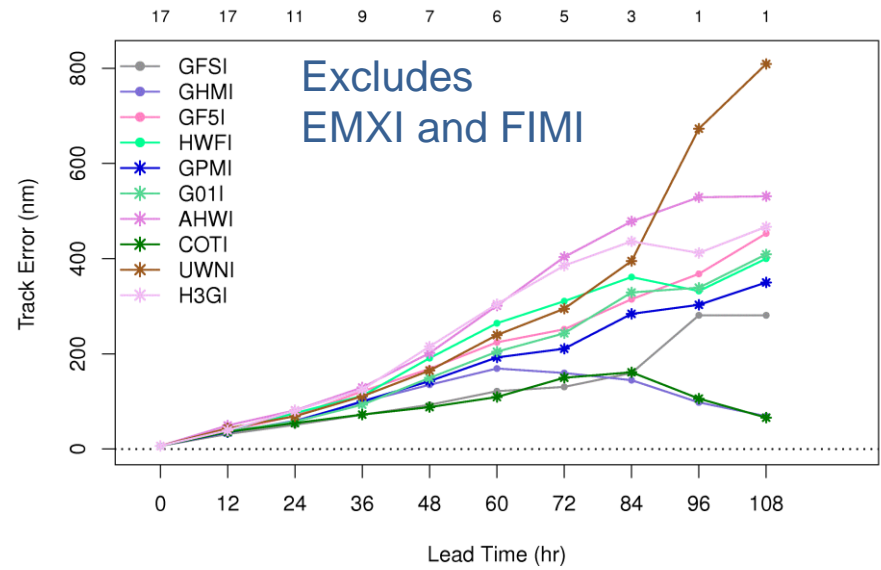


# Homogeneous Comparison of Stream 1.5 Models - Maria

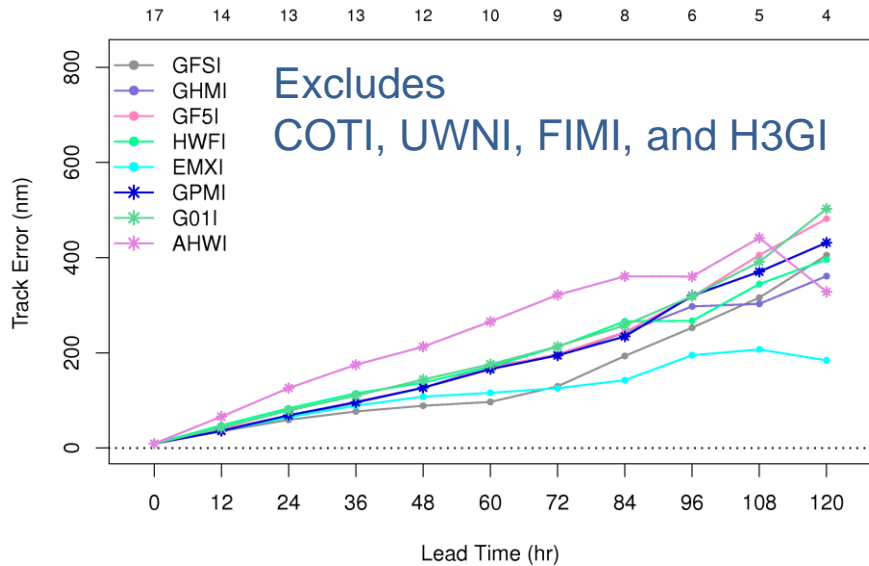
Mean Track Error  
Atlantic Basin (Water Only)



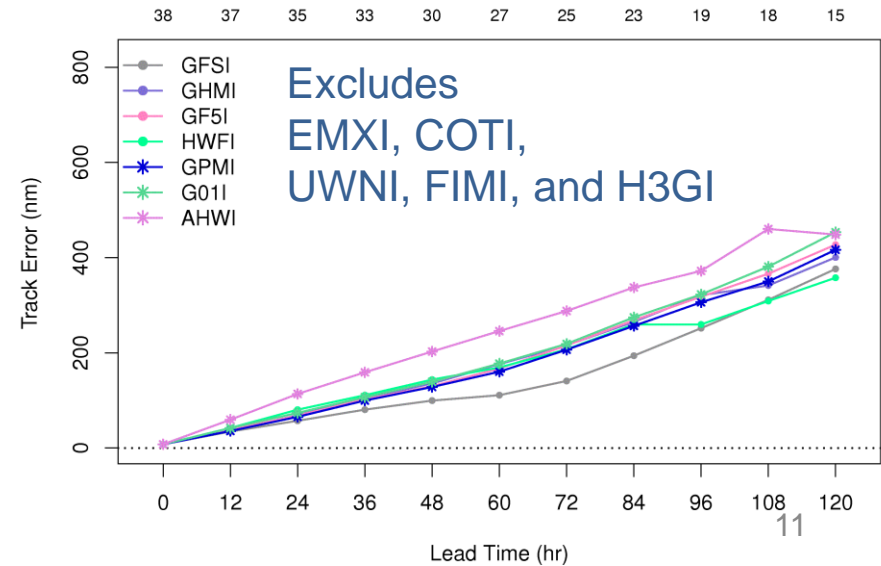
Mean Track Error  
Atlantic Basin (Water Only)



Mean Track Error  
Atlantic Basin (Water Only)



Mean Track Error  
Atlantic Basin (Water Only)



# Verification tools and methods

- Community verification tools
  - Development of community tools to replicate and extend NHC capability is in progress
  - Requirements document ready for review
  - Implementation and testing this winter by DTC
- Interpolator
  - Interpolator code implemented by TCMT
  - Eventually will make available to community
- Verification methods document
  - Summarize current capabilities for track and intensity
  - In progress



Collaboration with WMO Joint Working Group on Forecast Verification Research

# Verification tools and methods

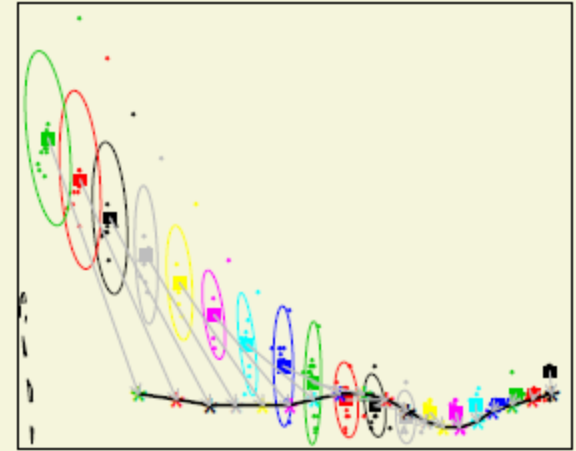
## Additional topics

- Ensemble method investigation
- Sample size investigation
- Genesis probabilities
- Tracker implementation in community code

# Ensemble verification methods

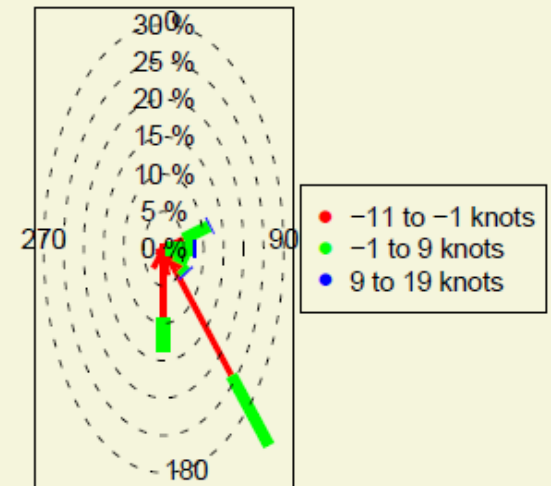
- Exploring methods to jointly evaluate ensemble track and intensity forecasts
  - **Minimum Spanning tree** (analogous to rank histogram for multivariate forecasts)
  - **Energy score** (analogous to CRPS for multivariate forecasts)

Storm 02, initialized at 081200 2009

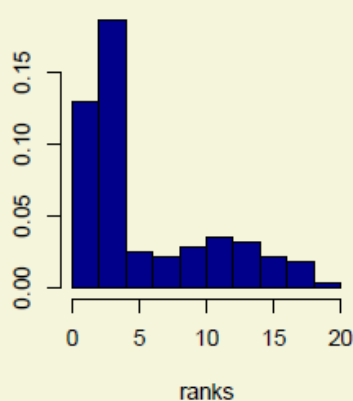


From Wilks (2004)

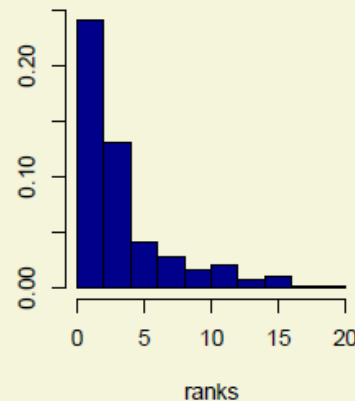
Wind Speed Error



12-h lead time



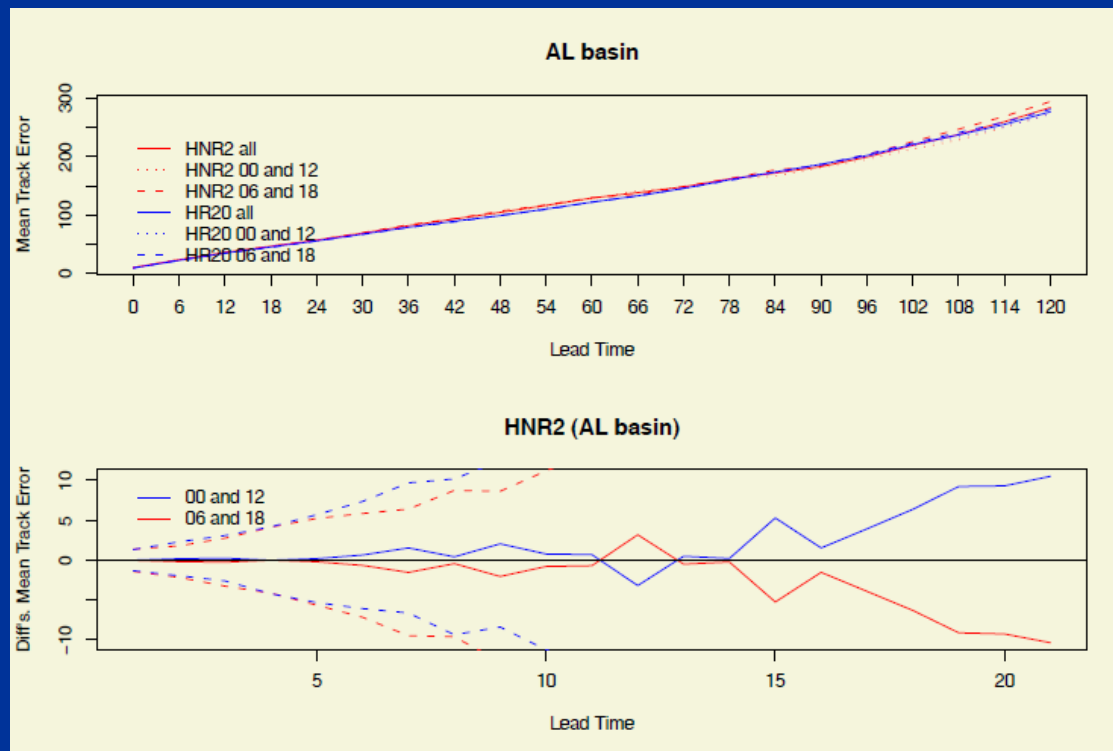
12-h lead time (bias corrected/scaled)



# Sample size investigations

Exploration of how many samples are needed to obtain stable verification results

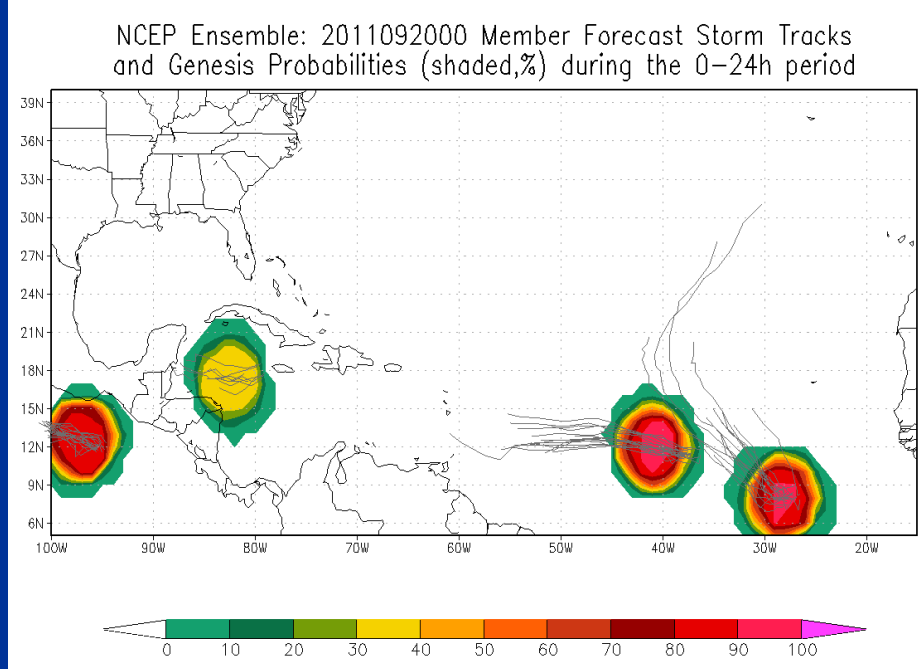
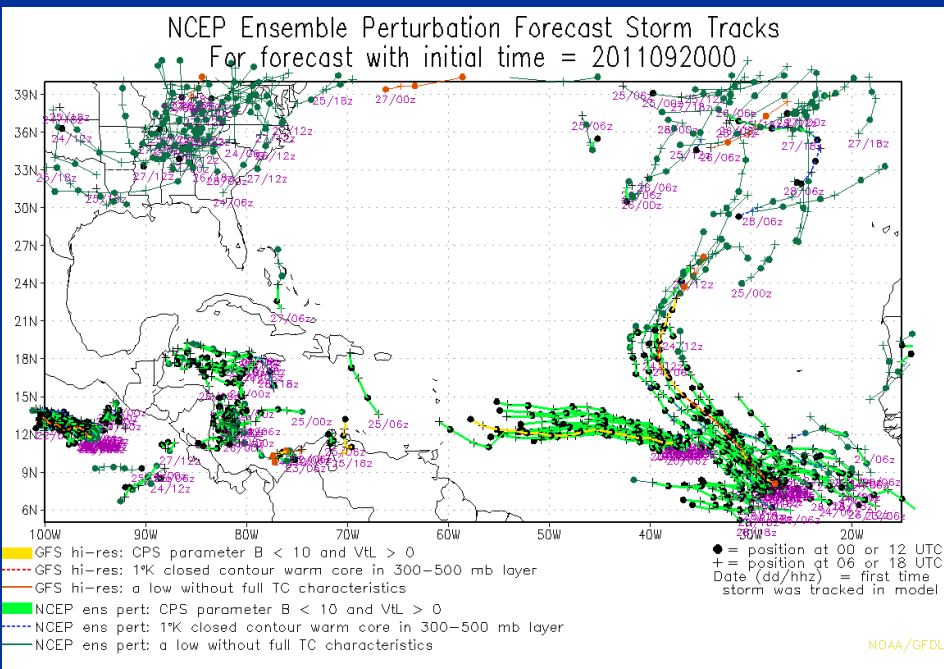
Ex: What happens if some issue times not included?  
What are the impacts of autocorrelation on verification analyses?



# Verification tools and methods

- Development of model-based tools for forecasting genesis; Perform verification of model genesis forecasts (5.3.3, 5.4.1).

Ensemble products sub-group has begun work to use a combination of statistical / diagnostic methods (Majumdar) as well as tracker-based methods (below) to create a genesis forecast product using global ensembles.





## Upgrades for tracker

- Effort currently underway involving GFDL, EMC, ESRL, DTC to unify versions of the tracker with most recent version. Targeting Spring, 2012 for implementation at NCEP (5.4.1).
- Tracker upgrade at NCEP to include cyclone phase detection, tracking on moveable nests, as well as tracking for additional ensembles such as SREF, FNMOC, NAEFS, and 12Z ECMWF.

# Release of tracker to community

Continue to improve tracker, release latest version to the Community (*GFDL*, DTC)

Stream 2 5.4.2 Sept 30, 2011

## GFDL tracker releases

- August – component of HWRF v3.3a release
- November– stand alone version (v3.3b) for models other than HWRF

## GFDL tracker upgrades

- Ongoing - Upgraded tracker for all NCEP applications will be transitioned to DTC when ready



# Challenges and Issues

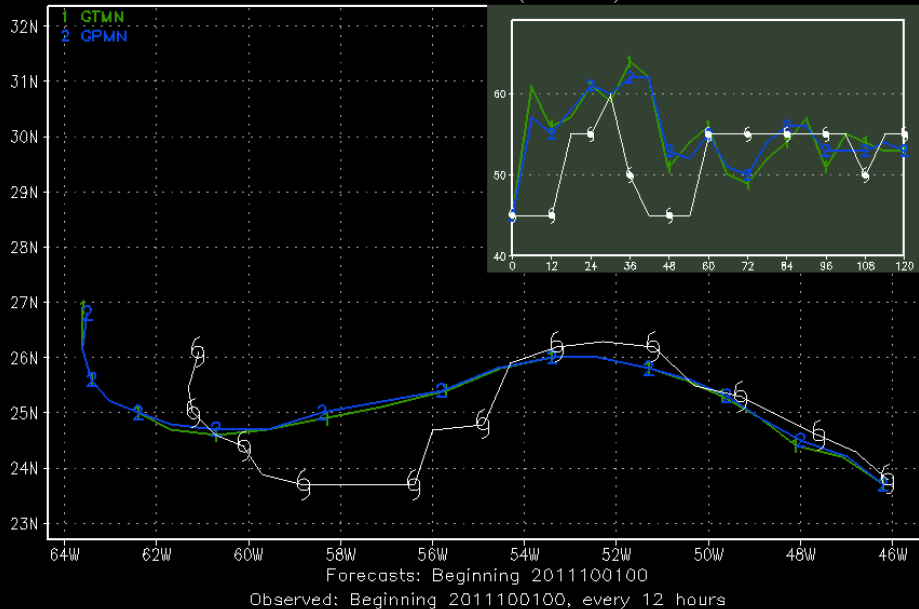
- Use of common tracker
- Estimation of forecast intensity
- Need for central verification activities for consistent model evaluations
- Stratification of results
  - What are appropriate subsamples?
- Evaluation of additional storm attributes
  - What additional storm characteristics should be considered?  
Precipitation? Storm surge? Storm structure variables? Others?

# Challenges and Issues

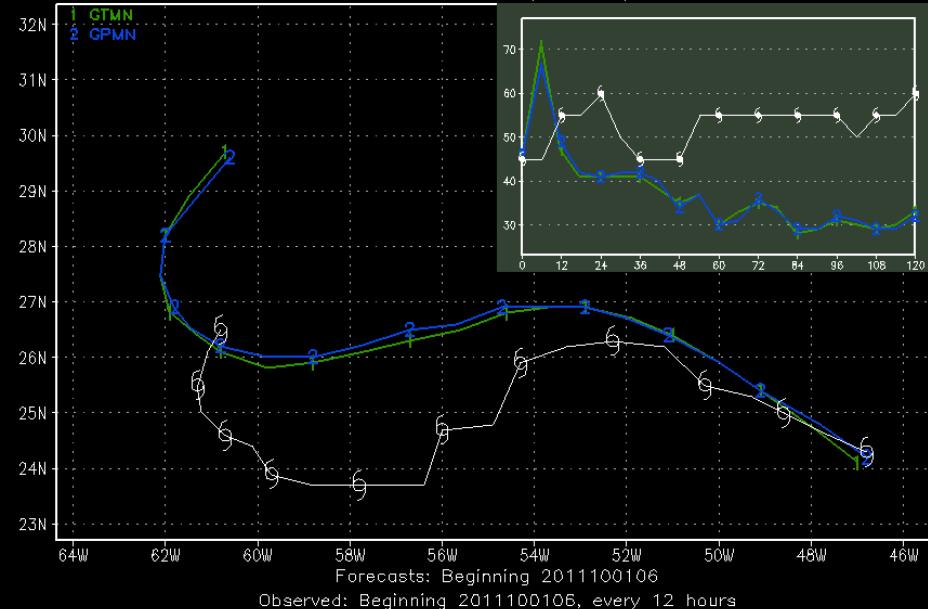
## Use of common tracker / Estimation of forecast intensity

Two cases of Philippe from GFDL ensemble: **Green** track uses external tracker (no time averaging), **Blue** track uses internal model tracker (3-h time averaging).

2011 Tropical Cyclone Tracks  
Storm: AL1711 (PHILIPPE)



2011 Tropical Cyclone Tracks  
Storm: AL1711 (PHILIPPE)



Individual ensemble members all showed similar variability, with mean intensity differences on the order of 1% - 4% at each lead time over the full season (n=350).

# GFDL Tracker vs. COAMPS-TC Built-in Tracker

- GFDL tracker has been implemented in COAMPS-TC and used for the 2011 real-time forecasts.
- Comparison of GFDL tracker with COAMPS-TC built-in tracker shows that they are very close.

