Evaluation of Experimental Models for Tropical Cyclone Forecasting in Support of the NOAA Hurricane Forecast Improvement Project (HFIP)

Barbara Brown, Paul Kucera,
Louisa Nance, and Christopher Williams

Joint Numerical Testbed
Tropical Cyclone Modeling Testbed
National Center for Atmospheric Research
(bgb@ucar.edu)

Interdepartmental Hurricane Conference, March 2012

Tropical Cyclone Modeling Testbed (TCMT): HFIP Model evaluation activities

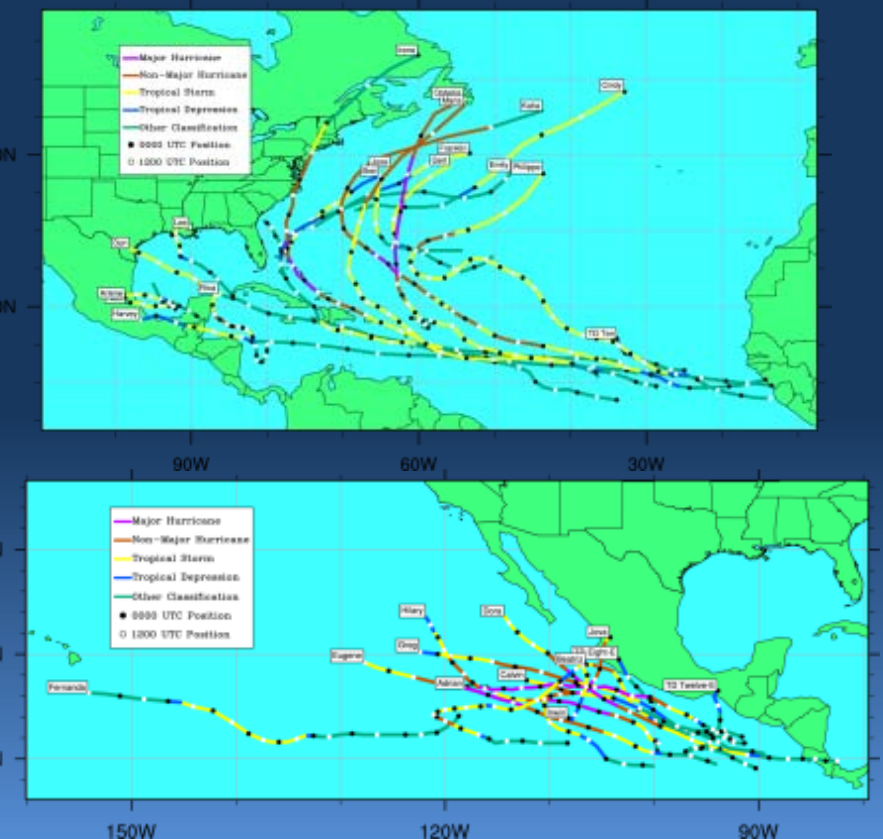
- *Main focus*: Independent diagnostic evaluation of HFIP models
- Planning and evaluation of HFIP Retrospective evaluations
  - Implementation of meaningful diagnostic verification approaches: *Focus on NHC requirements*
  - Evaluation of Stream 1.5 candidates; report to NHC and HFIP
- Real-time demonstration during Demo period (Aug-Nov)
- Evaluation of Demo models
  - Intensive evaluations of specific storms
  - Overall evaluation of forecasts for all storms
- Development of verification methods and tools

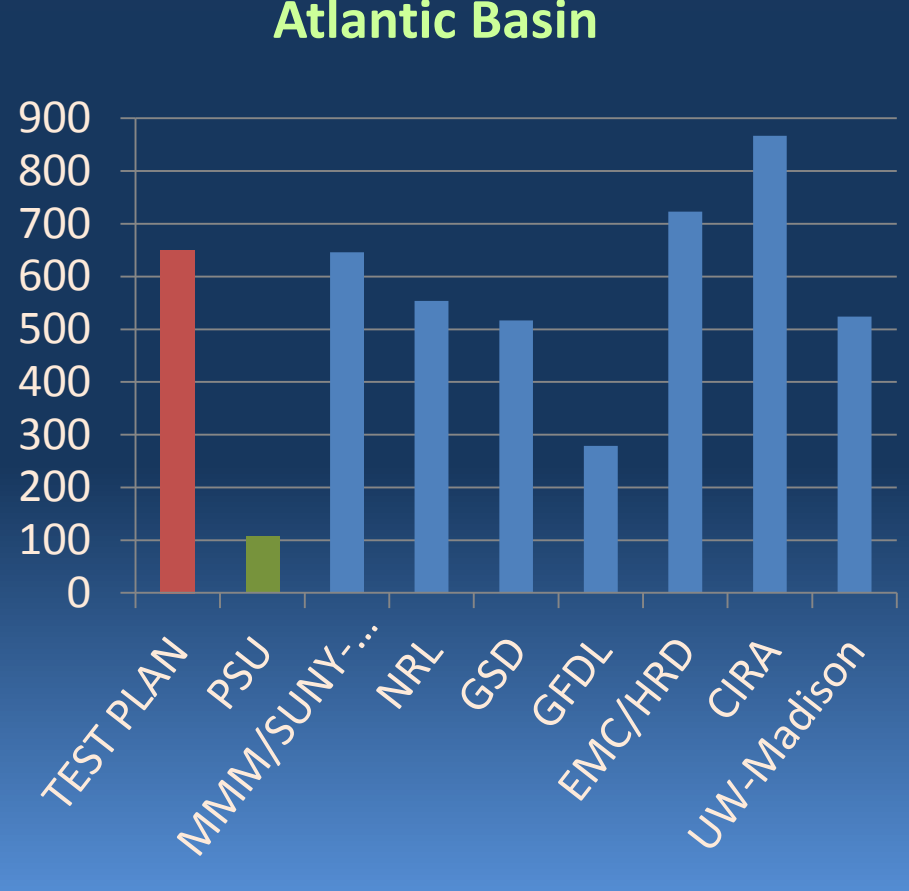
HFIP 2011 Retrospective Cases

Goal: Select Stream 1.5 models for HFIP Demo exercise

Modeling groups ran retrospective cases for more than 600 cases from 2008-2010

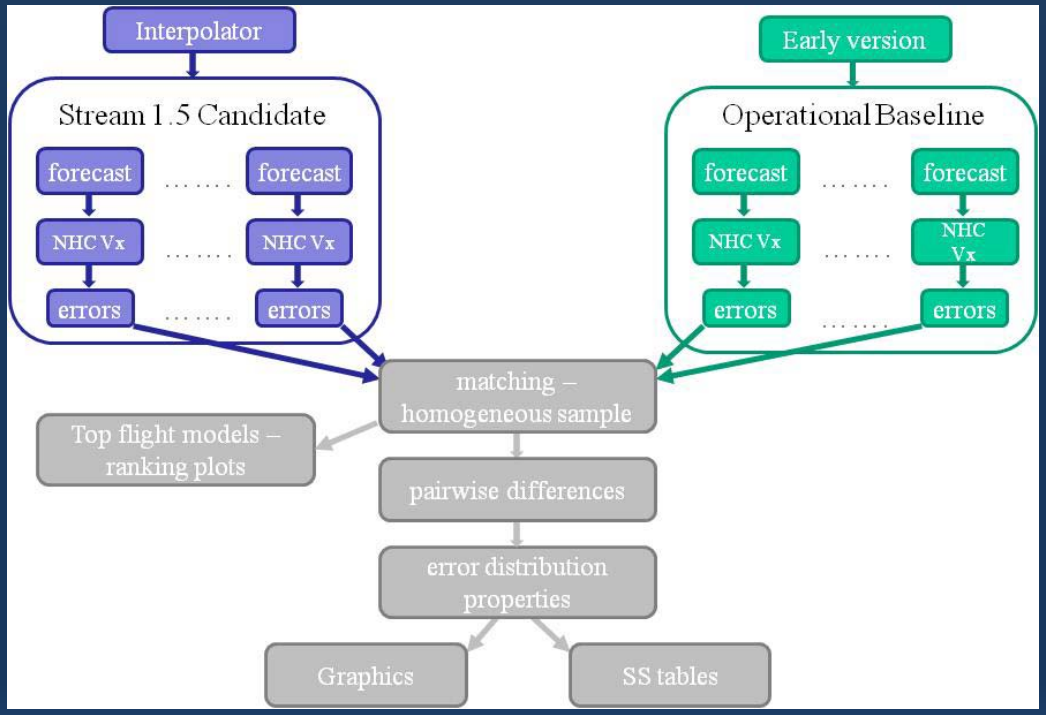
Collaboration with NHC





Retrospective evaluations

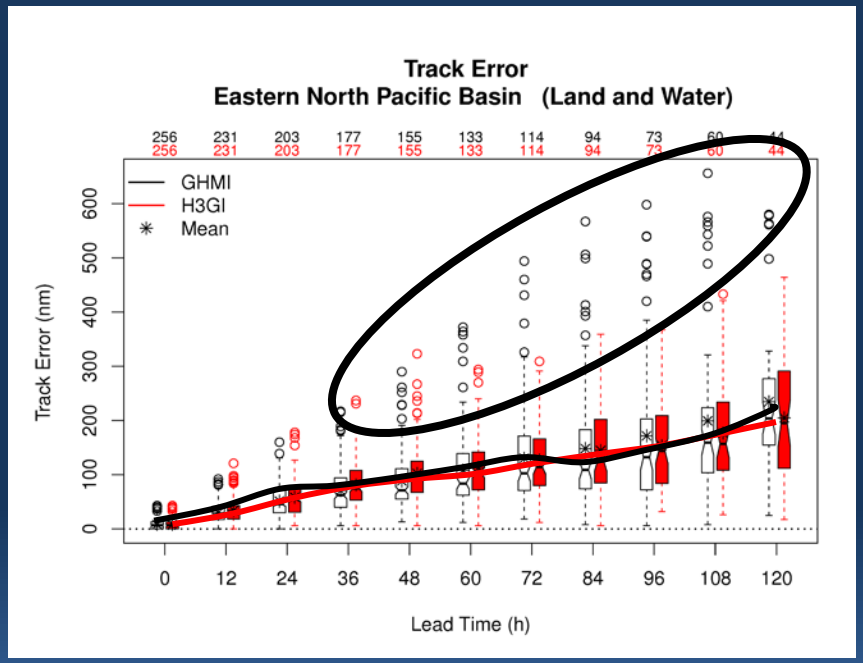
- Comparisons with baseline models
  - Significant and "practical" differences
  - Frequencies of large error differences
- Contributions to consensus forecasts
- Performance relative to "top flight" models

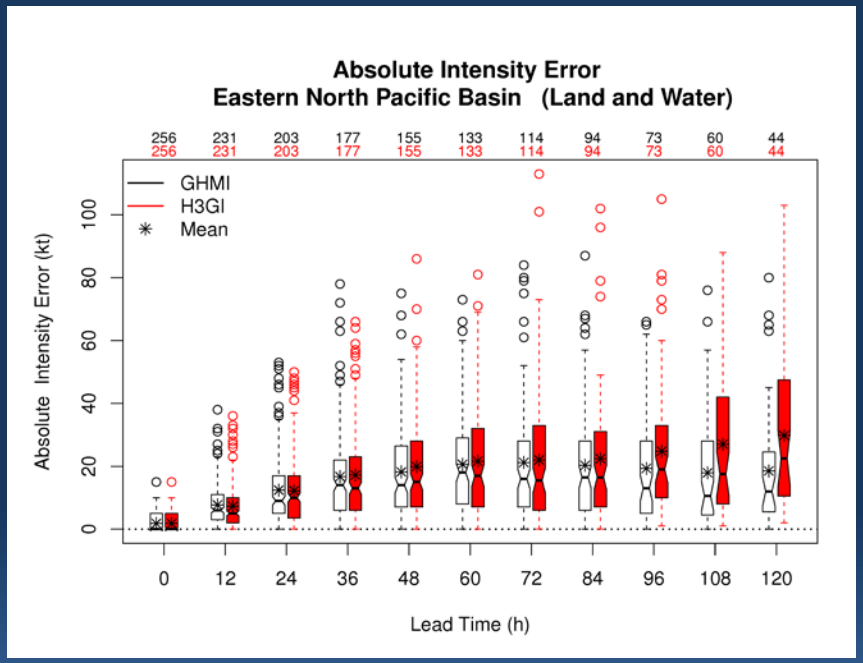


2011 Retrospective Evaluation

Example: 3 km HWRF

Error Distribution Evaluation



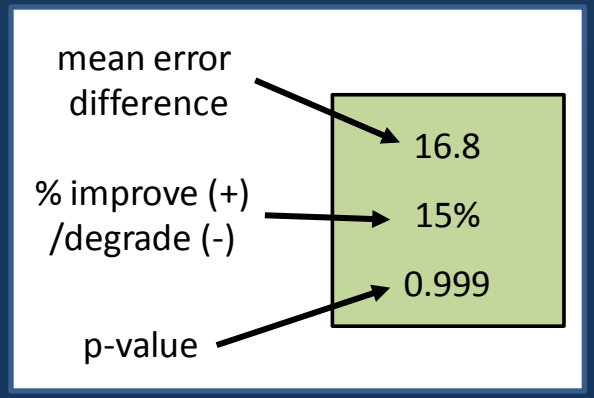


Stream 1.5 model: Fewer cases w/ large track errors
Similar or larger errors for intensity

2011 Hurricane Retrospective Evaluation - Results

Example: 3 km HWRF

Practical Significance Evaluation

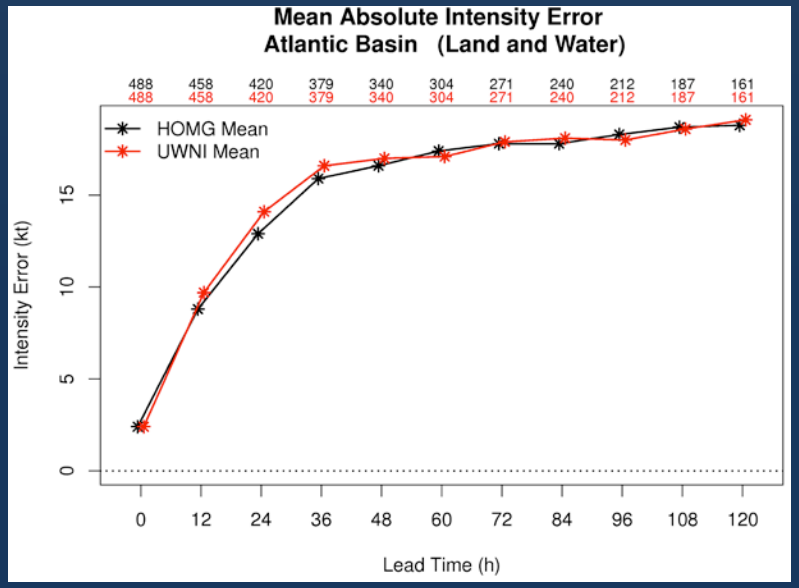


		Track	Intensity
SS differences		Δ < -20	Δ < -2
		-20 < Δ < -10	-2 < Δ < -1
		-10 < Δ< 0	-1 < Δ< 0
		0 < Δ< 10	0 < Δ < 1
		10 < Δ< 20	1 < Δ < 2
		Δ > 20	Δ> 2
Not SS		Δ< 0	Δ < 0
		Δ> 0	Δ> 0

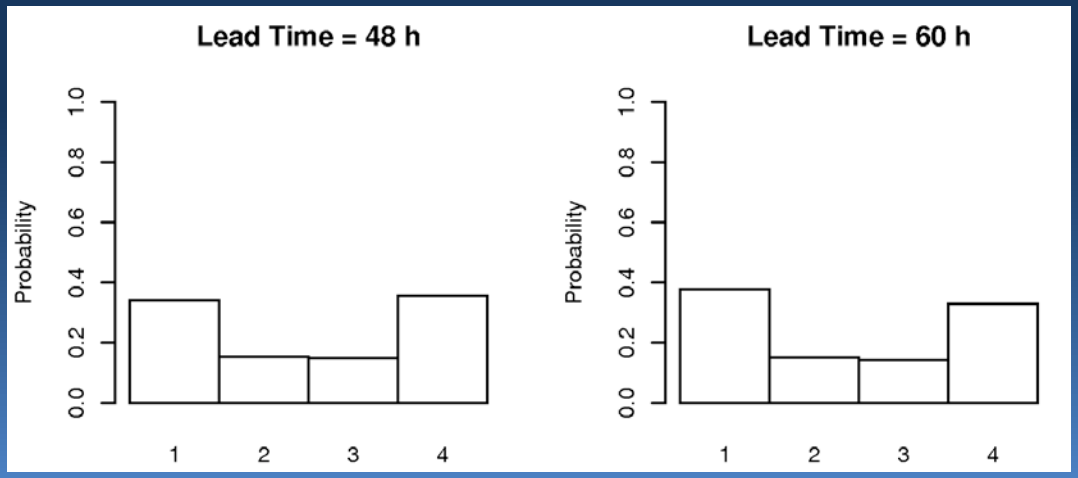
	Forecast Hour	0	12	24	36	48	60	72	84	96	108	120
Atlantic Basin	GFSI	0.0	-0.7	-4.0	-5.4	-9.1	-13.2	-16.5	-16.1	-20.5	-28.3	-44.4
	Track	0%	-2%	-8%	-8%	-10%	-12%	-12%	-10%	-11%	-13%	-19%
	(Land and Water)	–	0.618	0.987	0.986	0.961	0.936	0.819	0.837	0.876	0.929	0.969
	GHMI	0.0	1.4	0.9	1.7	-3.1	-6.2	-5.7	2.5	2.9	-1.6	-7.2
	Track	0%	4%	2%	2%	-3%	-5%	-4%	1%	1%	-1%	-3%
	(Land and Water)	–	0.880	0.426	0.540	0.682	0.667	0.420	0.201	0.185	0.083	0.307
	GHMI	0.0	0.5	1.0	2.1	1.9	1.7	0.8	0.0	-0.5	-0.6	-0.2
	Intensity	0%	6%	8%	13%	11%	9%	4%	0%	-3%	-3%	-1%
	(Land and Water)	–	0.987	0.987	0.999	0.998	0.984	0.532	0.000	0.279	0.292	0.094
	GHMI	0.0	0.5	1.3	2.4	2.1	2.4	2.8	2.9	2.0	0.7	0.5
	Intensity	0%	5%	9%	14%	11%	12%	14%	15%	10%	4%	3%
	(Water Only)	–	0.904	0.999	0.999	0.963	0.970	0.998	0.973	0.816	0.338	0.188
	LGEM	0.0	0.1	-0.2	-0.2	-1.1	-1.4	-1.5	-1.4	-1.8	-1.5	-0.9
	Intensity	0%	1%	-2%	-1%	-7%	-9%	-9%	-9%	-11%	-9%	-5%
	(Land and Water)	–	0.261	0.311	0.261	0.728	0.796	0.751	0.649	0.769	0.682	0.451
	LGEM	0.0	-0.1	-0.3	-0.4	-1.0	-0.7	-0.6	-0.6	-1.0	-0.7	-1.1
	Intensity	0%	-1%	-2%	-3%	-7%	-4%	-4%	-4%	-6%	-4%	-6%
	(Water Only)	–	0.261	0.451	0.311	0.558	0.359	0.276	0.248	0.336	0.229	0.327
Eastern North Pacific Basin	GFSI	0.0	-2.1	-3.6	-6.8	-12.9	-10.8	-12.4	-11.1	1.3	13.0	7.5
	Track	0%	-7%	-7%	-9%	-14%	-11%	-11%	-8%	1%	7%	3%
	(Land and Water)	–	0.618	0.519	0.681	0.882	0.694	0.738	0.693	0.071	0.331	0.123
	GHMI	0.0	-4.0	-7.3	-10.4	-13.9	-4.5	3.3	3.5	18.2	22.7	29.9
	Track	0%	-13%	-14%	-15%	-16%	-4%	3%	2%	11%	11%	13%
	(Land and Water)	–	0.995	0.962	0.889	0.880	0.238	0.129	0.100	0.342	0.405	0.570
	GHMI	0.0	0.4	0.1	-0.5	-1.7	-0.9	-0.8	-2.1	-5.5	-9.1	-11.3
	Intensity	0%	5%	1%	-3%	-9%	-4%	-4%	-10%	-28%	-51%	-61%
	(Land and Water)	–	0.816	0.132	0.350	0.841	0.270	0.154	0.399	0.869	0.999	0.999
	GHMI	0.0	0.3	0.2	-0.4	-2.4	-3.6	-4.7	-6.2	-9.6	-9.3	-8.4
	Intensity	0%	4%	1%	-2%	-12%	-18%	-25%	-34%	-56%	-55%	-43%
	(Water Only)	–	0.382	0.225	0.224	0.911	0.939	0.956	0.985	0.999	0.999	0.999
	LGEM	0.0	-0.1	-1.0	-2.4	-3.2	-4.7	-6.0	-6.9	-9.6	-11.9	-15.2
	Intensity	0%	-1%	-9%	-16%	-19%	-28%	-38%	-45%	-63%	-78%	-96%
	(Land and Water)	–	0.197	0.682	0.912	0.953	0.893	0.864	0.821	0.868	0.944	0.994
	LGEM	0.0	-0.1	-1.1	-2.5	-3.0	-4.5	-6.3	-7.3	-8.8	-8.7	-11.4
	Intensity	0%	-1%	-10%	-17%	-17%	-26%	-39%	-46%	-53%	-49%	-66%
	(Water Only)	–	0.197	0.727	0.903	0.920	0.838	0.845	0.804	0.780	0.774	0.968

Top Flight Model Comparison

UW-NMS Example



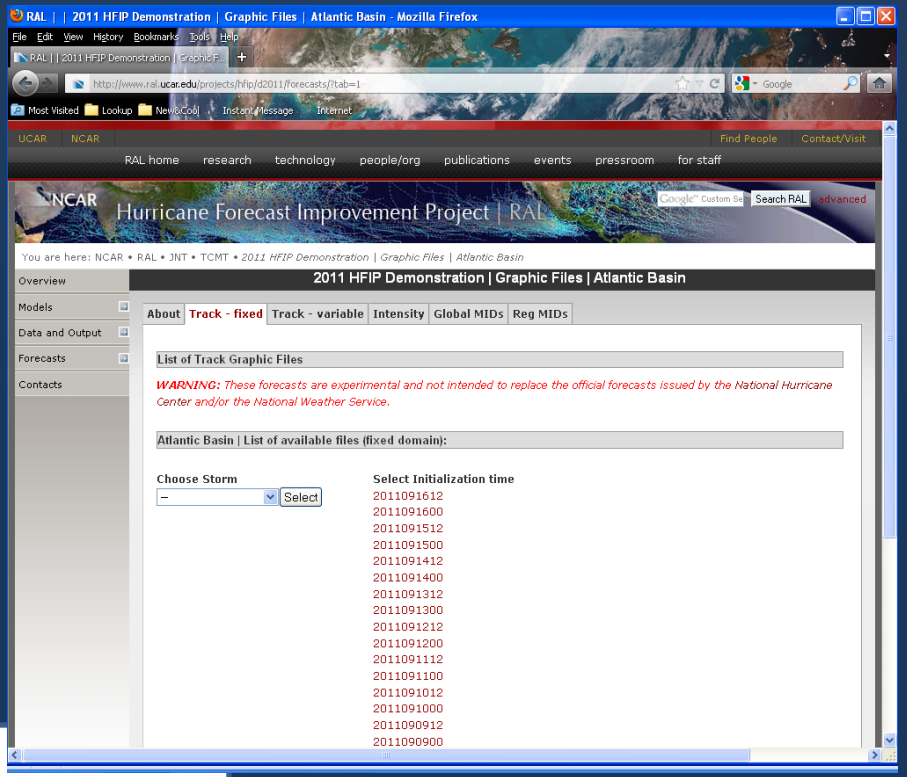
Average Errors

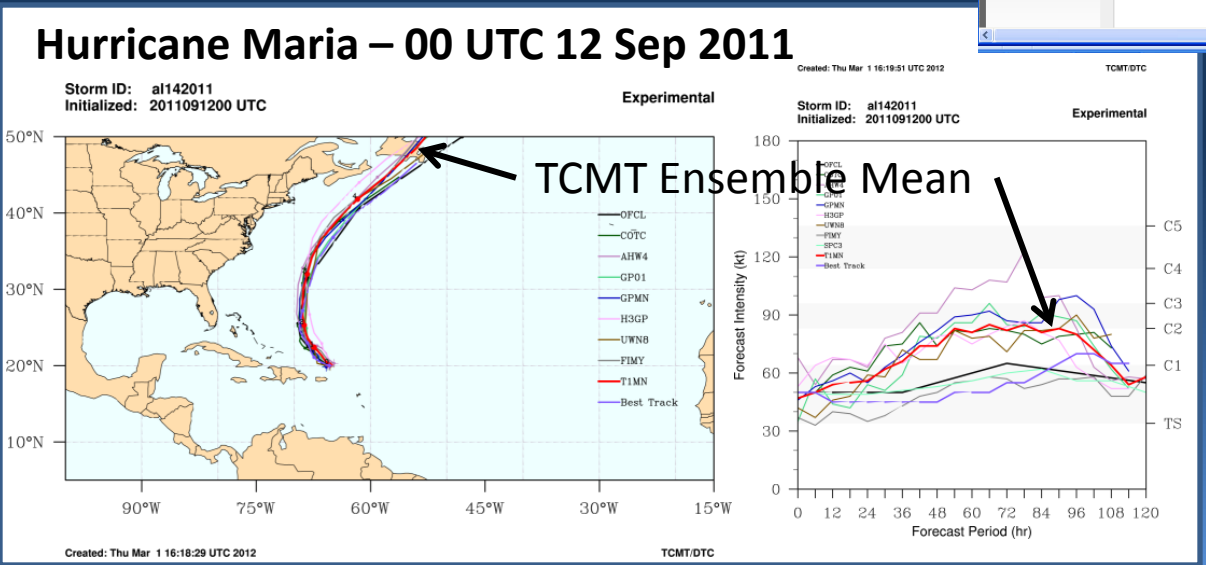


Rank

Real-Time Demonstration and Evaluation

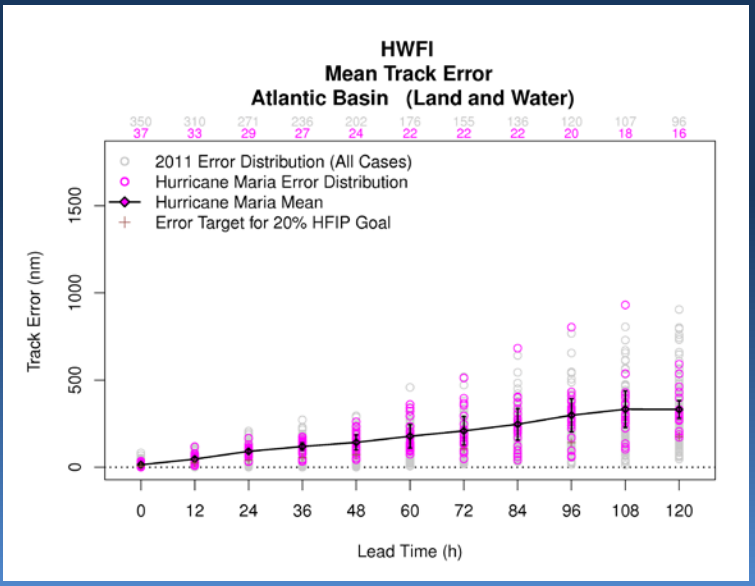
- Operational and experimental models provided in real-time to TCMT website in support of the HFIP Demonstration
- Multi-model ensemble mean computed and displayed using the HFIP experimental models
- Track plots and verification results available on the HFIP web site
- Near real time diagnostic evaluation of individual storms (Irene, Maria)



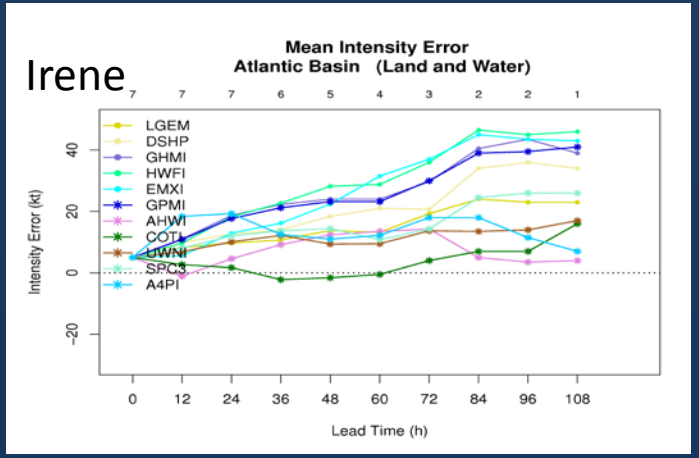


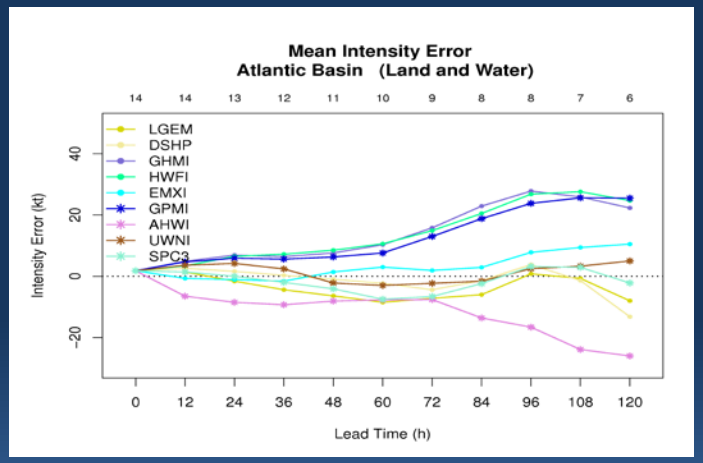
Real-Time Demonstration and Evaluation

- Operational and experimental models provided in real-time to TCMT website in support of the HFIP Demonstration
- Multi-model ensemble mean computed and displayed using the HFIP experimental models
- Track plots and verification results available on the HFIP web site
- Near real time diagnostic evaluation of individual storms (Irene, Maria)



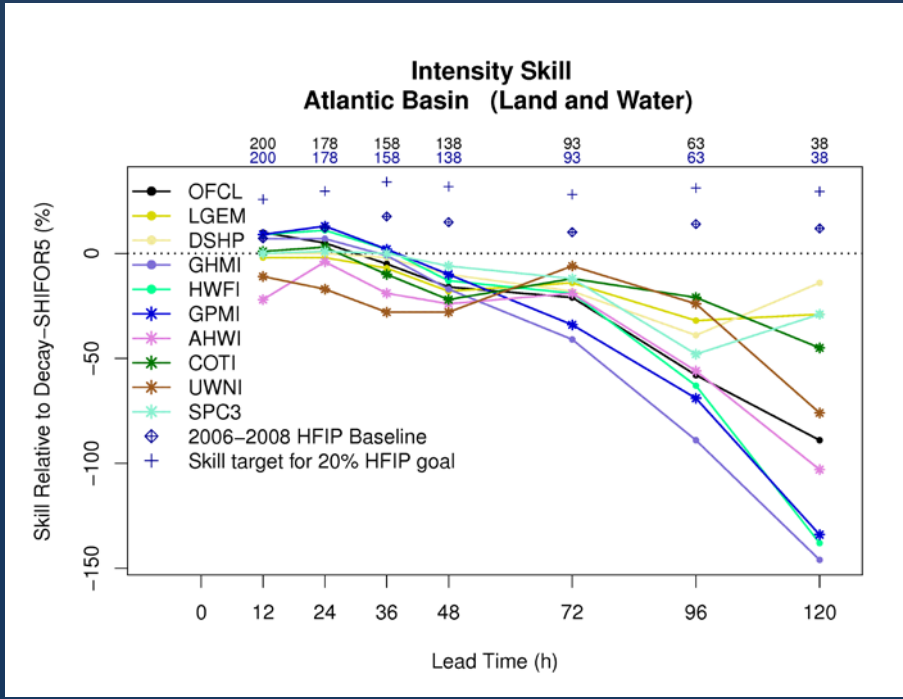
Sample Size Sensitivity

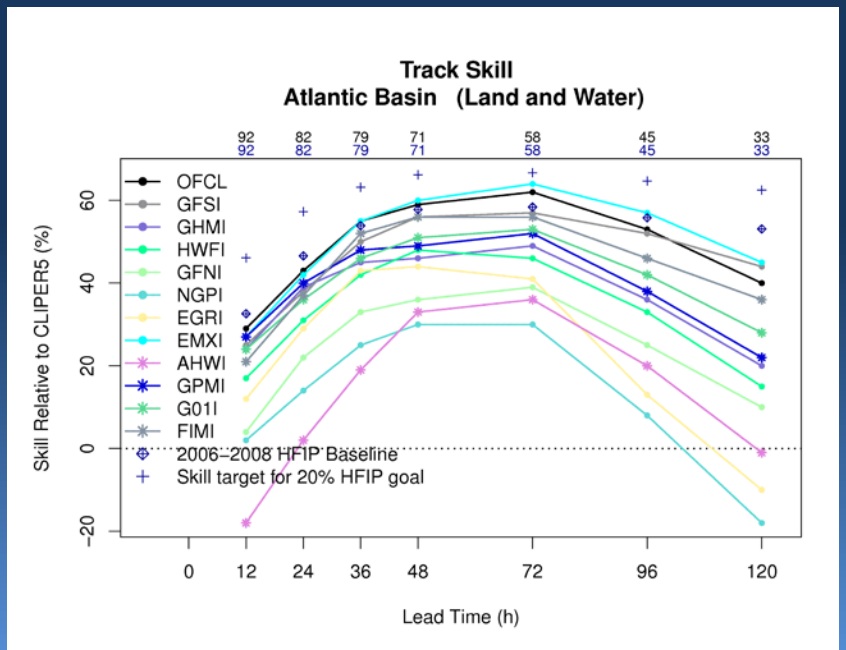




Real-Time Demonstration and Evaluation

- Operational and experimental models provided in real-time to TCMT website in support of the HFIP Demonstration
- Multi-model ensemble mean computed and displayed using the HFIP experimental models
- Track plots and verification results available on the HFIP web site
- Near real time diagnostic evaluation of individual storms (Irene, Maria)



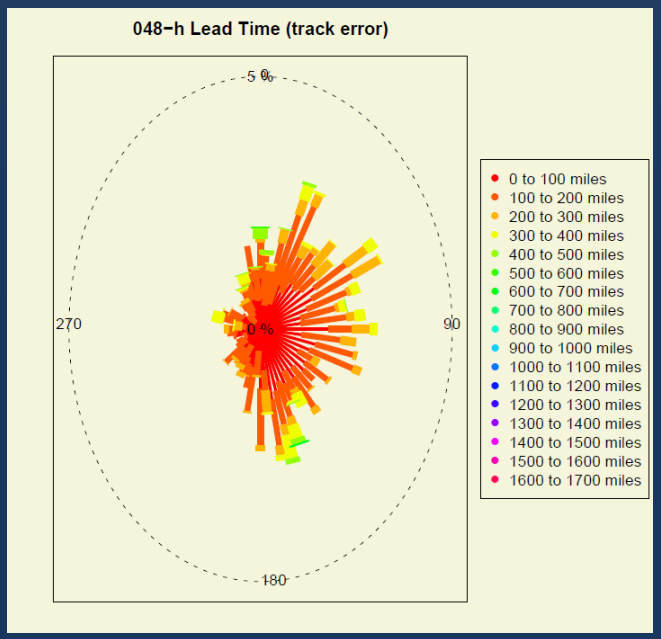


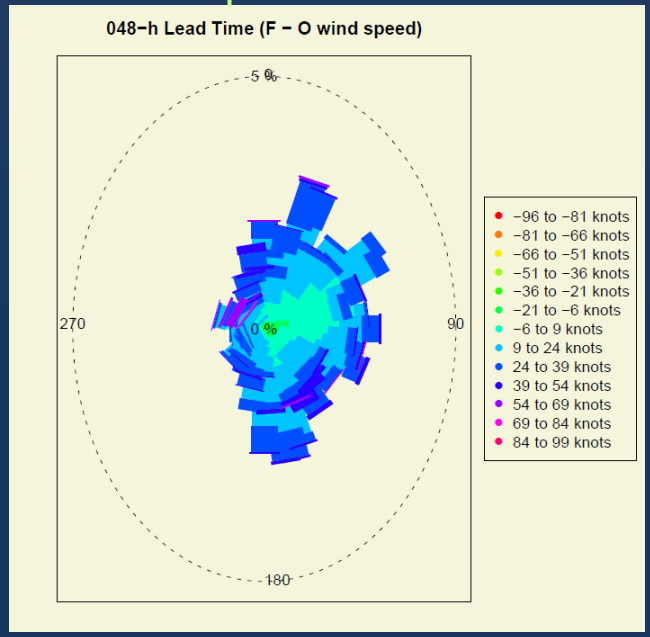
New Methods for Evaluation of TC Forecasts

Developing new methods to evaluate ensemble and multi-model forecasts

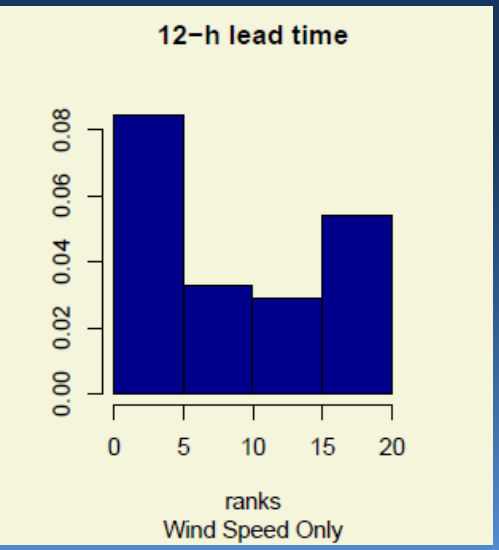
Example: GFDL 18-member Ensemble

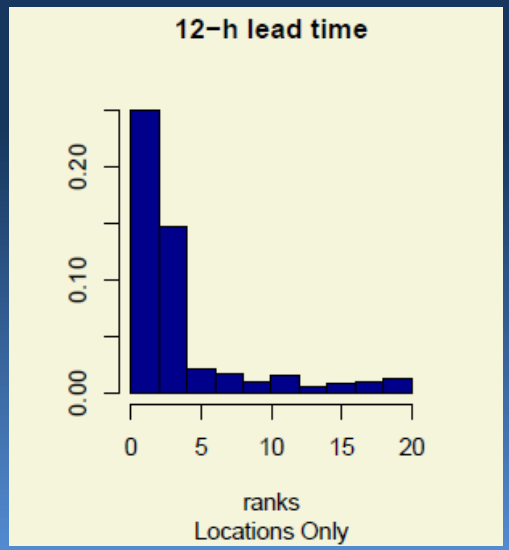
"Wind Rose" Plots for Track and Wind Speed Errors



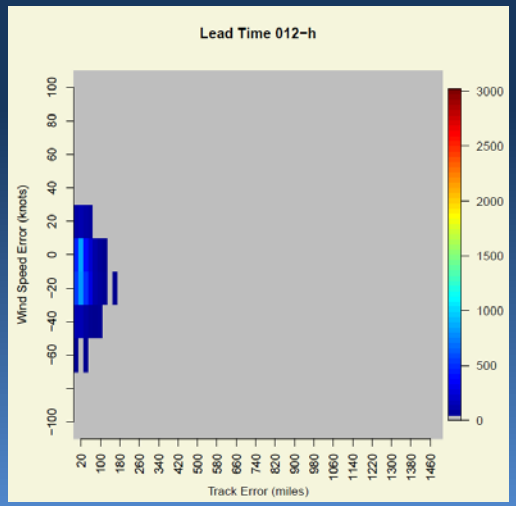


Rank Histograms





Wind Speed/ Track Error Diagrams



Summary

- TCMT supports HFIP efforts by providing independent, diagnostic, forecast verification of experimental model forecasts
- New approaches for evaluation
  - To meet needs of NHC for selecting promising models for Stream 1.5
  - To evaluate new kinds of forecasts (e.g., ensembles)
- Results available for 2009-2012 at http://www.rap.ucar.edu/jnt/tcmt/
- Currently gearing up for 2012 retrospective (Stream 1.5) evaluation