Evaluation of the 2015 GFDL Hurricane Ensemble Forecast System

2015 HFIP "Demo"

Matt Morin (Engility Holdings, Inc. / GFDL)
Tim Marchok (NOAA / GFDL)
Morris Bender (NOAA / GFDL)

HFIP Workshop on Effective Use of Hurricane Ensembles
November 17, 2015

Outline

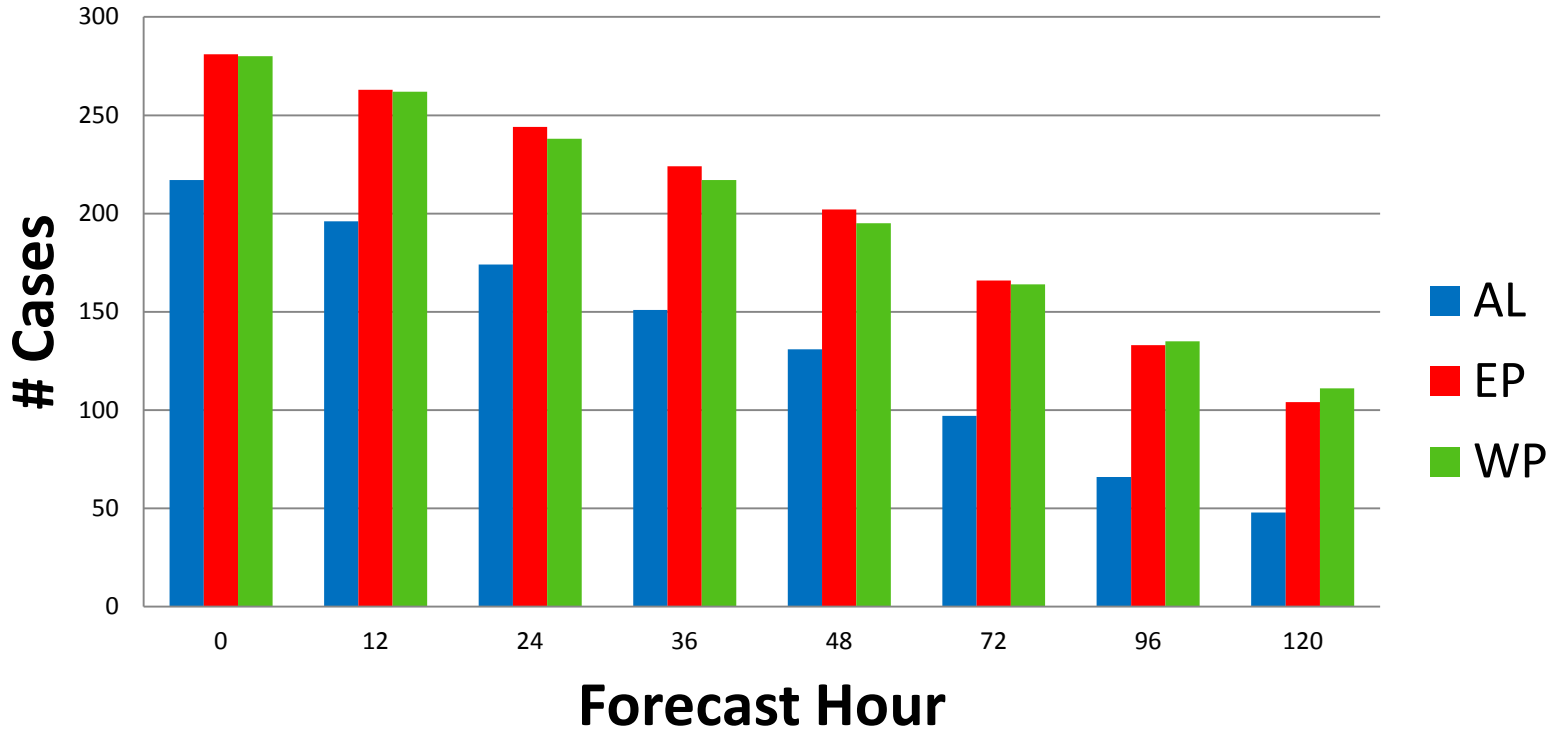
- 2015 GFDL ensemble overview
- 2015 HFIP “Demo” verifications and comparisons with operational guidance
  - AL, EP, and WP
- GFDL ensemble probability products
- Summary
- Future work

2015 GFDL Ensemble Overview

- 12 members (11 perturbations + 1 control)
  - 2 more perturbations than in 2013 and 2014
- Designed to produce large and realistic intensity spread and lower average intensity errors than our control model
- Able to run up to four simultaneous storms worldwide in real-time under Jet reservation
- Tested first formulation of new bias-corrected GFDL ensemble mean

2015 Demo Data Set

Verified GFDL Ensemble Mean Forecasts (Late guidance)



- EP and WP incomplete; will perform updated verification after running missing cases
- Switching inputs from GFS spectral files to grib will greatly improve GFDL forecast reliability (planned for 2016)

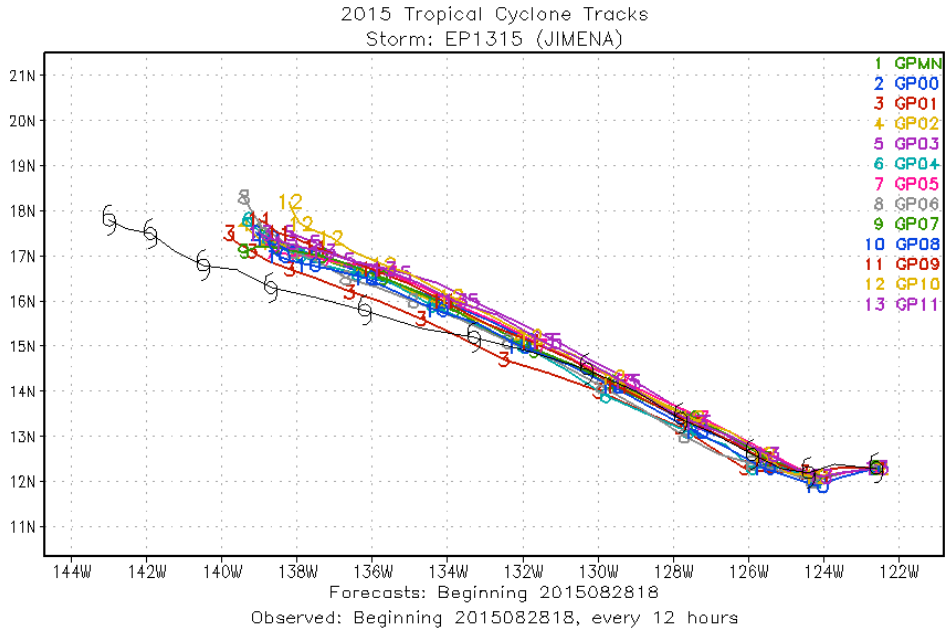
New GFDL Bias-corrected Ensemble Mean

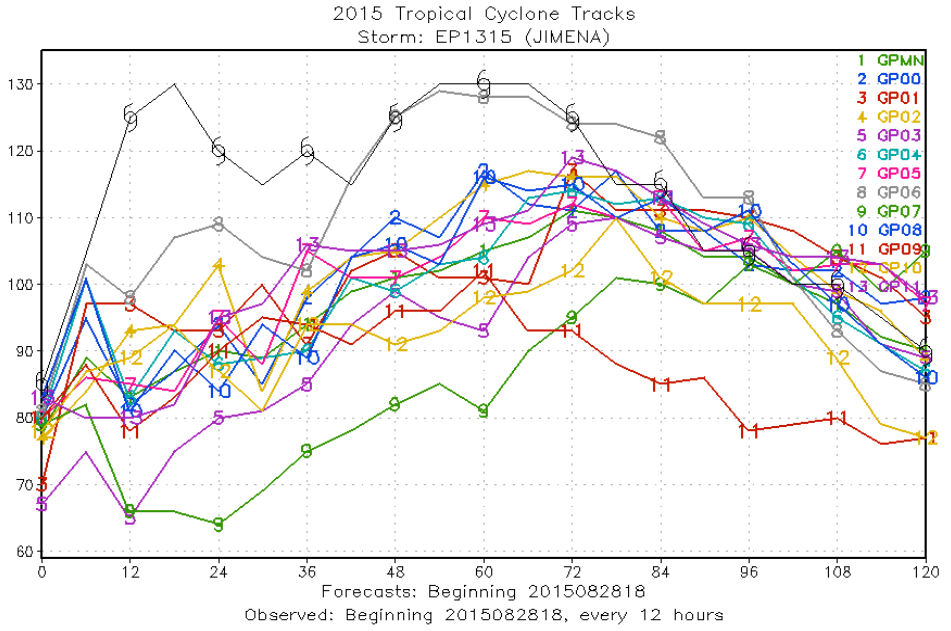
- Motivation: Lingering large biases in ensemble mean

- Method: Compute average of bias-corrected ensemble members
  - Basin-specific linear regression equation computed at each forecast lead time for each ensemble member
    - Vmax, min SLP, center lat., and center lon.
    - 2011-2014 used as regression equation "training phase"; 2015 is not used to create bias corrections

ATCF ID	**2015 GFDL Ensemble Membership** ↑(↓) = overall effect is **increased** (**decreased**) intensity relative to the Control	
GP00	**Control** forecast (configured similar to NCEP 2015 operational GFDL)	
GP01	**Unbogussed** forecast using the 2015 GFDL control model (**bogussed** for Invests)	
GP02	**Increase** NHC-observed V_{max} 10%, R34 25%, R50 40%, ROCI 25%	↑
GP03	**Decrease** NHC-observed V_{max} 10%, R34 25%, R50 40%, ROCI 25%	↓
GP04	**Increase** inner-core moisture by a max of 10%	↑
GP05	**Decrease** inner-core moisture by a max of 10%	↓
GP06	**Increase** SSTs by a max of 3°C within the initial extent of the TC	↑
GP07	**Decrease** SSTs by a max of 3°C within the initial extent of the TC	↓
GP08	Surface physics modification: ***GFDL 2011 operational formulation*** of C_D & C_H (surface drag and enthalpy exchange coefficients)	↑
GP09	Surface physics modification: ***HWRF 2014 operational formulation*** of C_H (surface enthalpy exchange coefficient)	↓
GP10	Physics modification: Effectively **increase** mean boundary layer depth	↑
GP11	Physics modification: Effectively **decrease** mean boundary layer depth	↓
GPMN	**Bias-uncorrected** ensemble mean: Average of uncorrected members computed at each lead time where the member availability is at least 4 members (40% threshold)	
GRMN	**Bias-corrected** ensemble mean: Average of linearly regressed members computed at each lead time where the member availability is at least 4 members (40% threshold)	

2015 GFDL Ensemble Spread Example





- Ensemble still too under dispersive for track; Ideal for intensity
- High priority goal for 2016 is to finally include track-specific perturbations
  - Initial TC center position
  - Vorticity confinement
  - Stochastic physics
  - Run from different global models

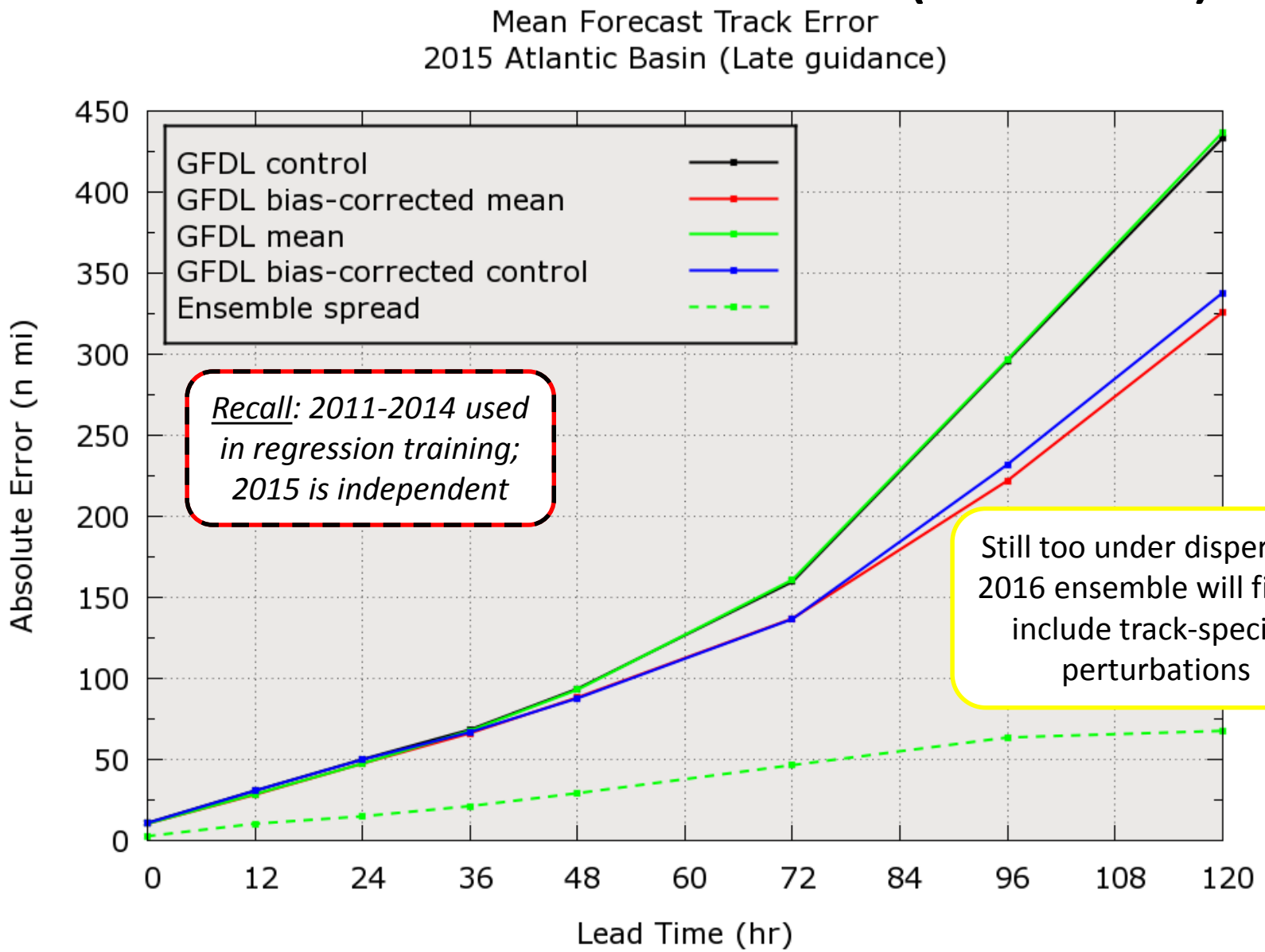
Model Verification

Internal GFDL comparisons
(verifications in error space)

Points of clarification for next series of figures:

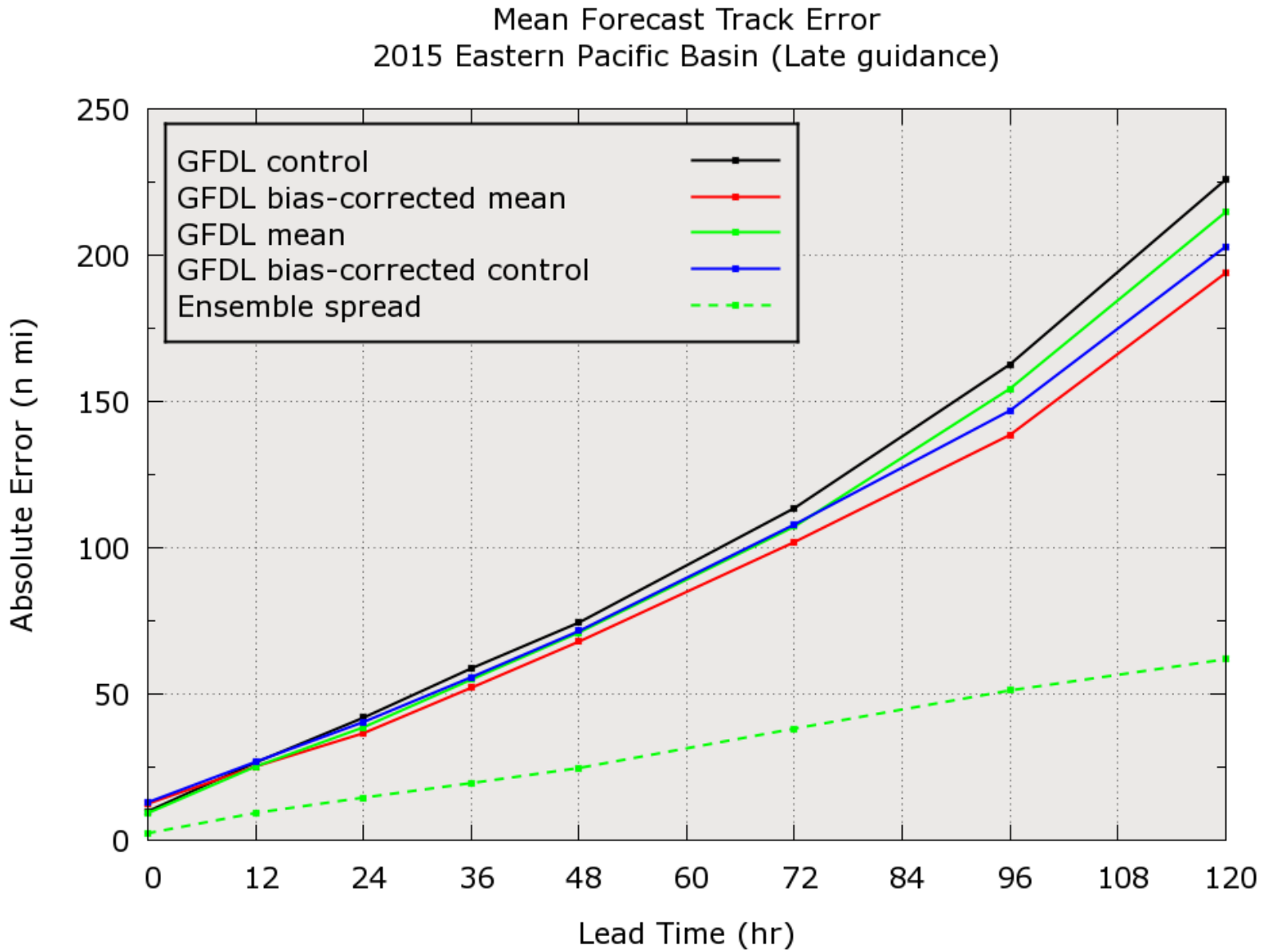
- "**Ensemble spread**" = Spread of **uncorrected** members from the 2015 GFDL ensemble
- "**IMPROV**" = **First model** improvement over the **second model**
- Statistical significance indicated in font style:
  - <95% uses gray font
  - >=95% uses black font
  - >=99% uses magenta font
- Practical significance indicated in font style:
  - For forecast difference >=1 kt or >=10 n mi, I made the font ***bold and italic***

Track Forecast Verification (Atlantic)



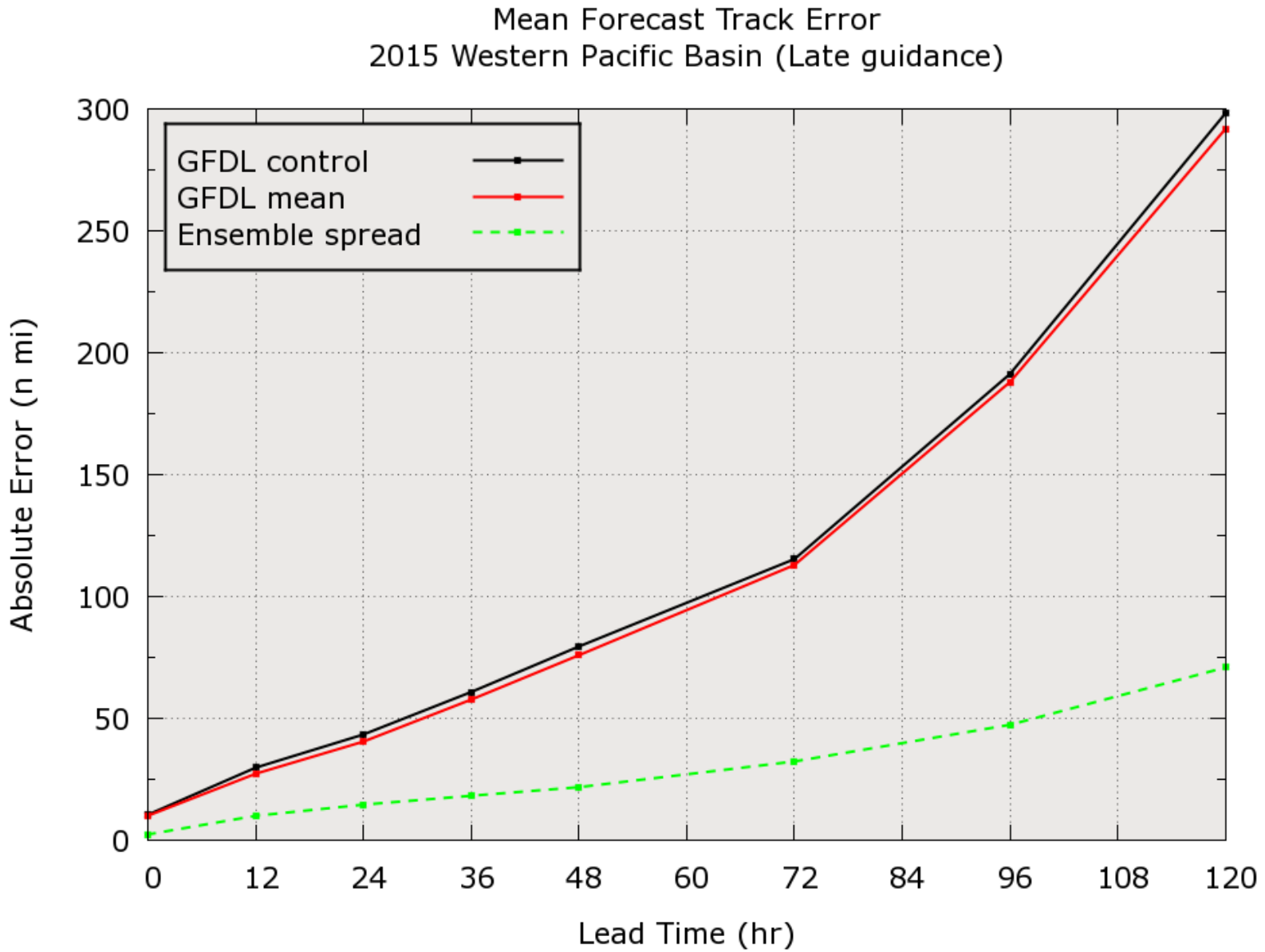
#CASES:	217	196	174	151	131	97	66	48
IMPROV:	4.5%	8.1%	5.2%	3.2%	5.8%	14.2%	25.0%	24.8%
SIGNIF:	93%	100%	100%	95%	100%	*100%*	*100%*	*100%*

Track Forecast Verification (Eastern Pacific)



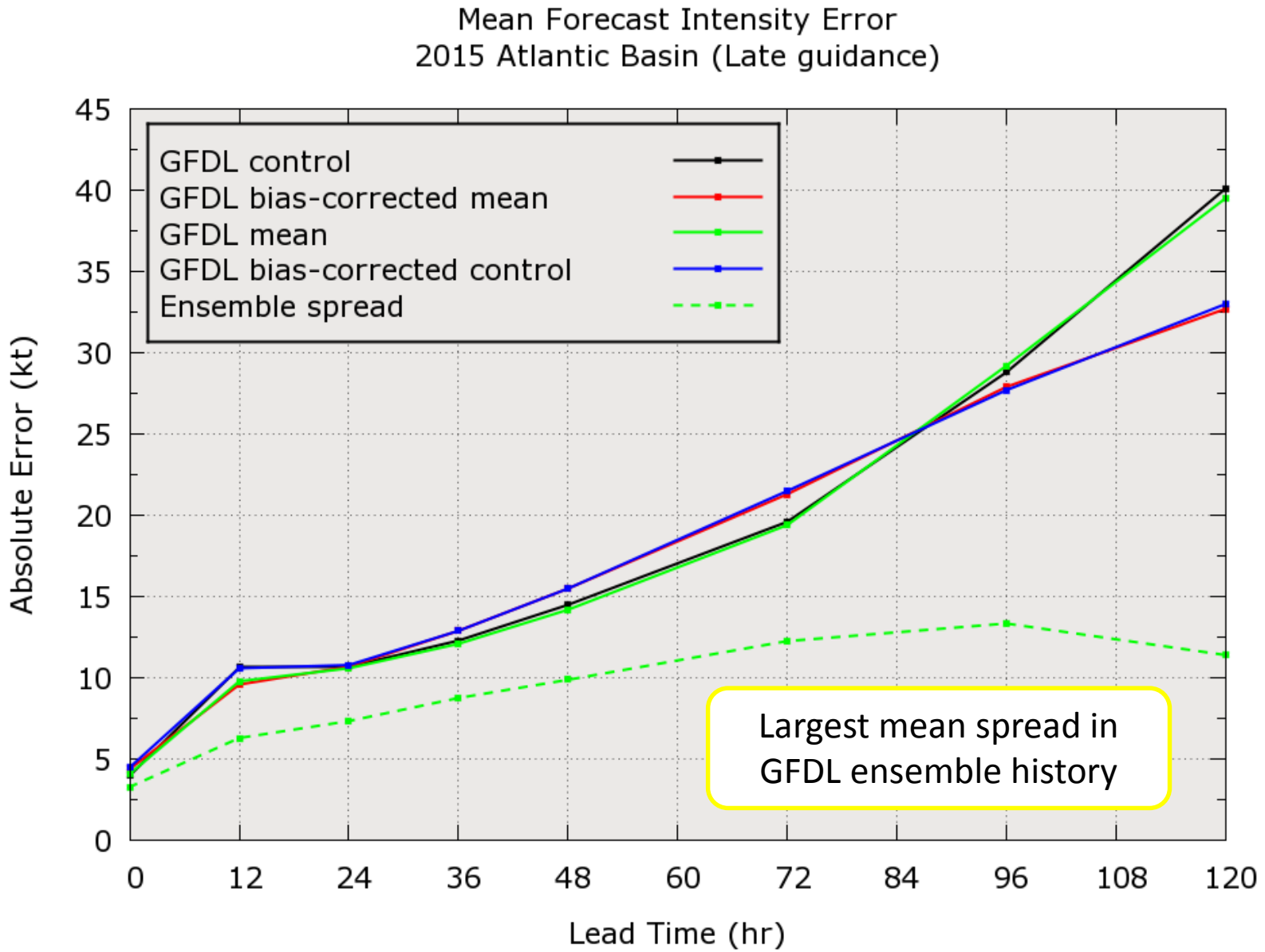
#CASES:	281	263	244	224	202	166	133	104
IMPROV:	-26.0%	5.2%	12.6%	11.2%	8.7%	10.2%	14.8%	14.1%
SIGNIF:	100%	100%	100%	100%	100%	*100%*	*100%*	*100%*

Track Forecast Verification (Western Pacific)



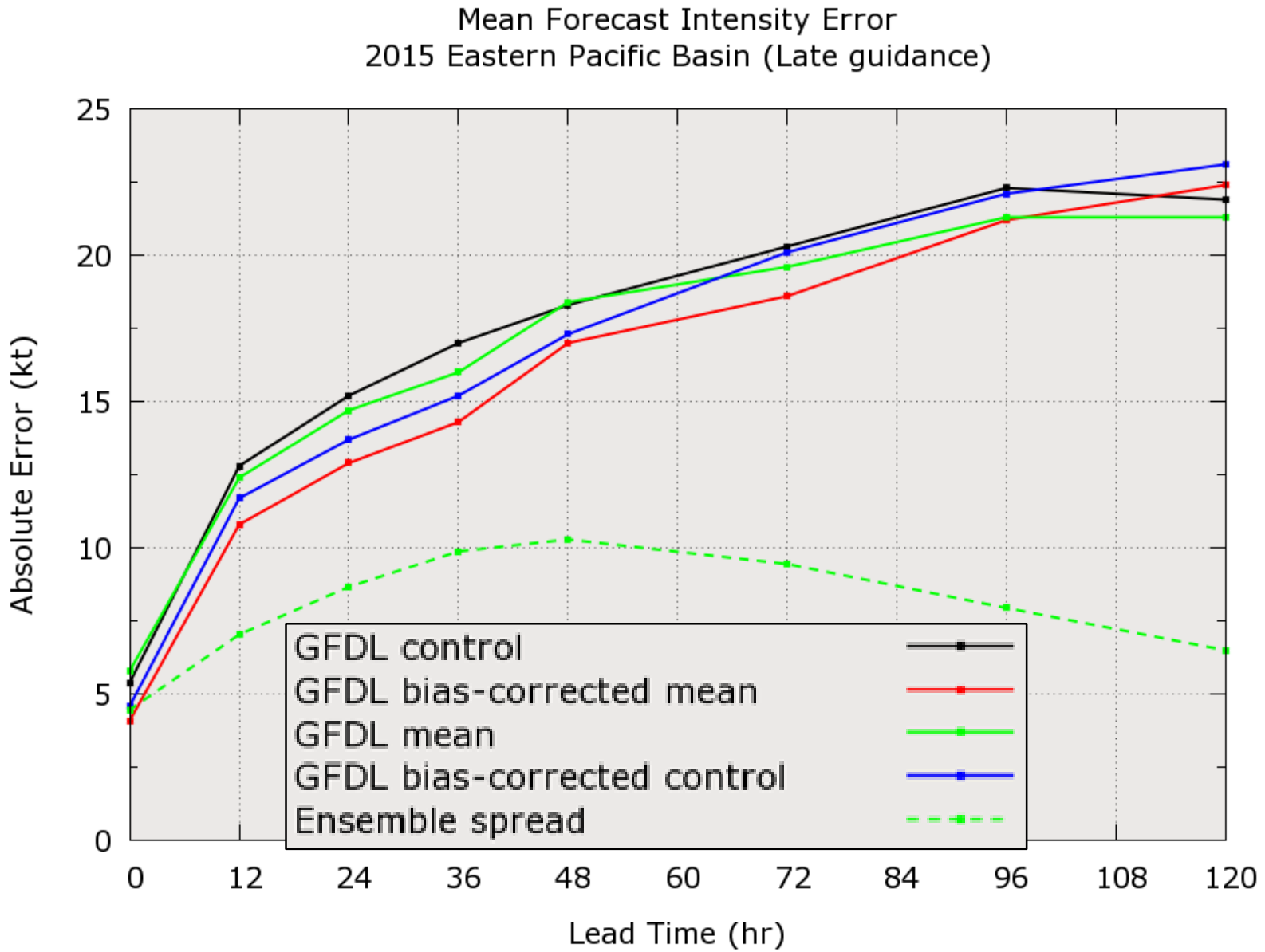
#CASES:	280	262	238	217	195	164	135	111
IMPROV:	3.8%	8.4%	6.7%	5.1%	4.5%	2.1%	1.7%	2.1%
SIGNIF:	100%	100%	100%	100%	100%	96%	94%	95%

Intensity Forecast Verification (Atlantic)



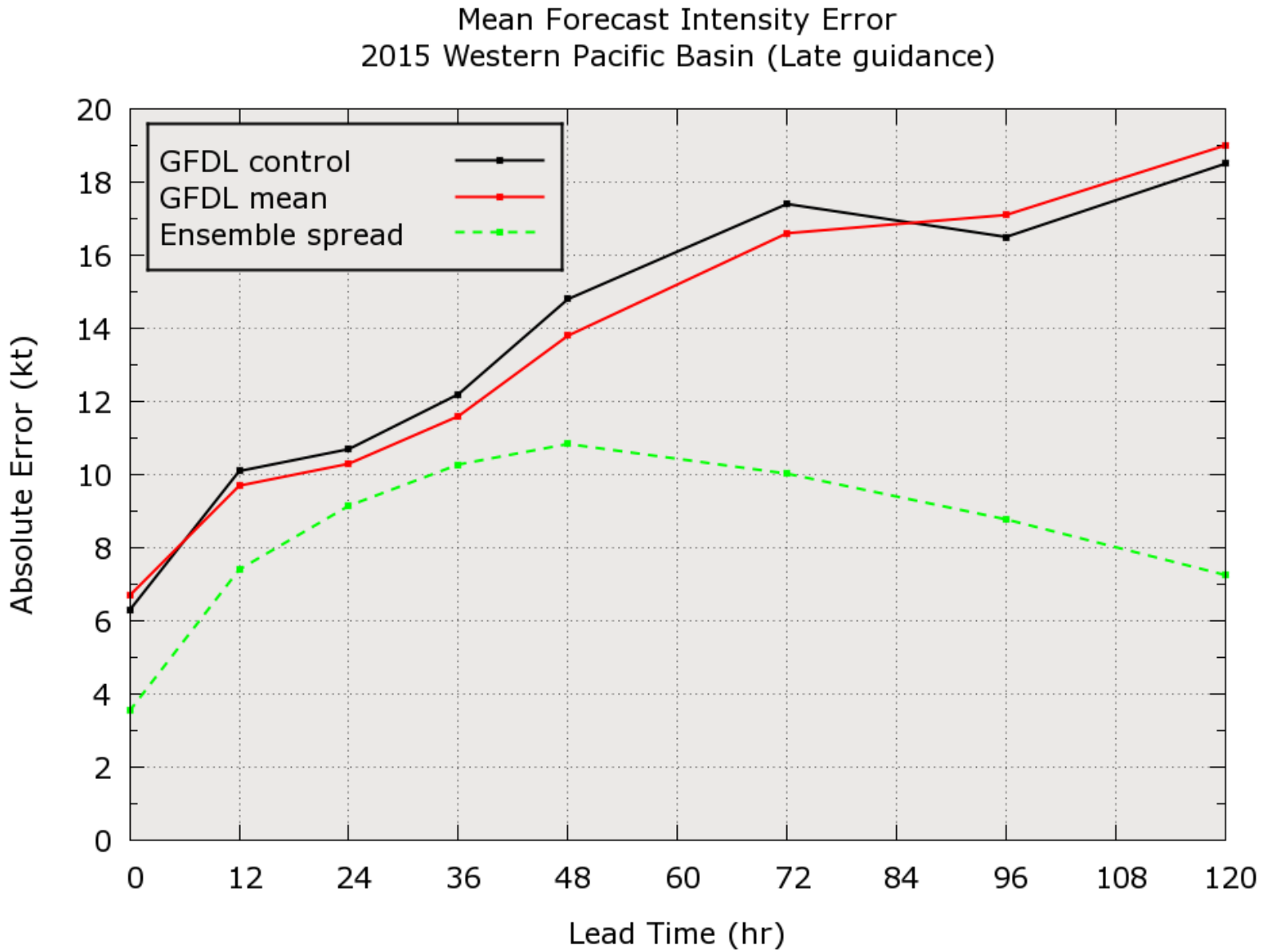
#CASES:	217	196	174	151	131	97	66	48
IMPROV:	-10.0%	10.3%	0.0%	-4.9%	-6.9%	-8.7%	3.1%	18.5%
SIGNIF:	94%	***100%***	52%	88%	*93%*	***97%***	76%	***100%***

Intensity Forecast Verification (Eastern Pacific)



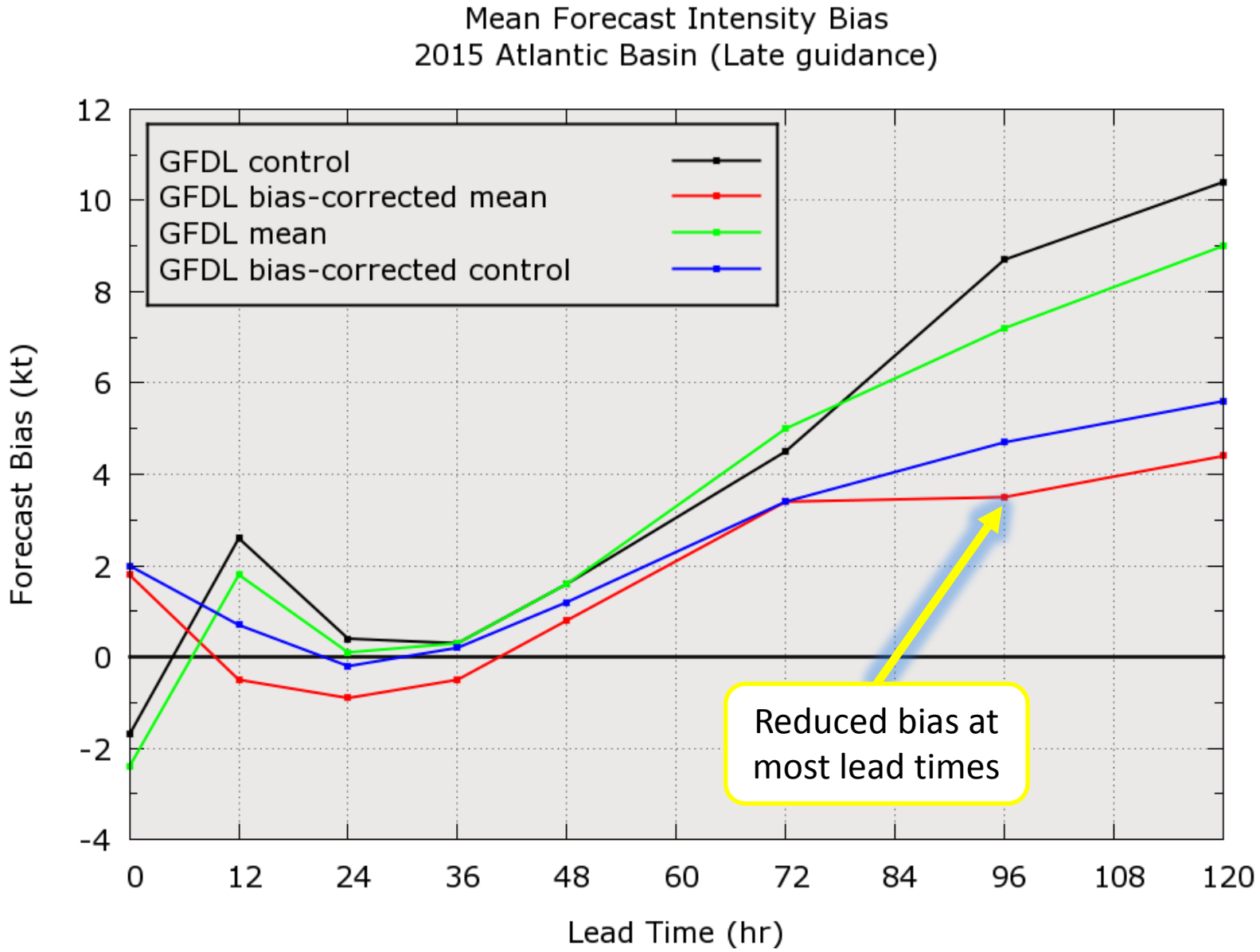
#CASES:	281	263	244	224	202	166	133	104
IMPROV:	24.1%	15.6%	15.1%	15.9%	7.1%	8.4%	4.9%	-2.3%
SIGNIF:	*100%*	*100%*	*100%*	*100%*	*99%*	*100%*	*96%*	64%

Intensity Forecast Verification (Western Pacific)

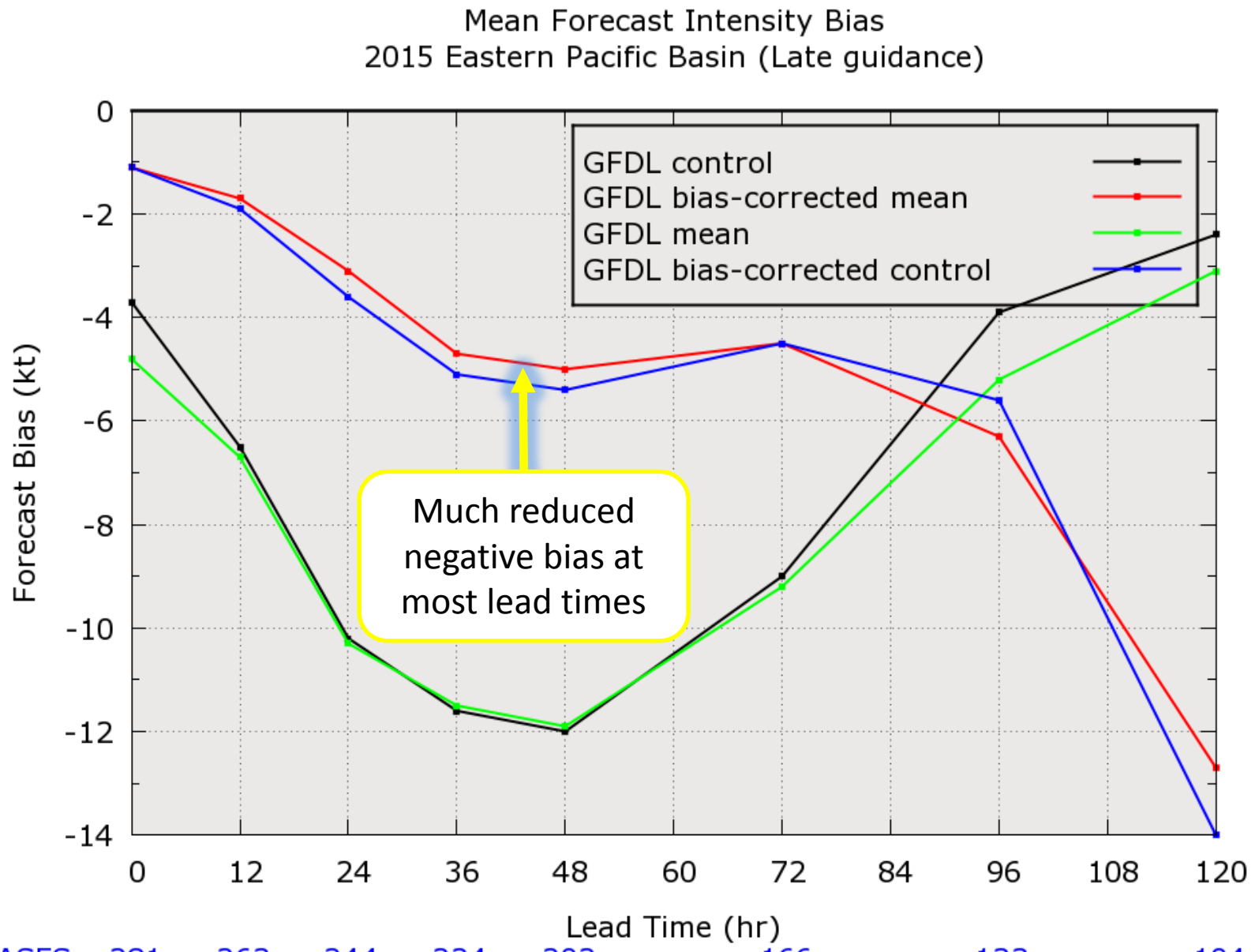


#CASES:	280	262	238	217	195	164	135	111
IMPROV:	-6.3%	4.0%	3.7%	4.9%	6.8%	4.6%	-3.6%	-2.7%
SIGNIF:	100%	91%	90%	96%	*99%*	97%	88%	84%

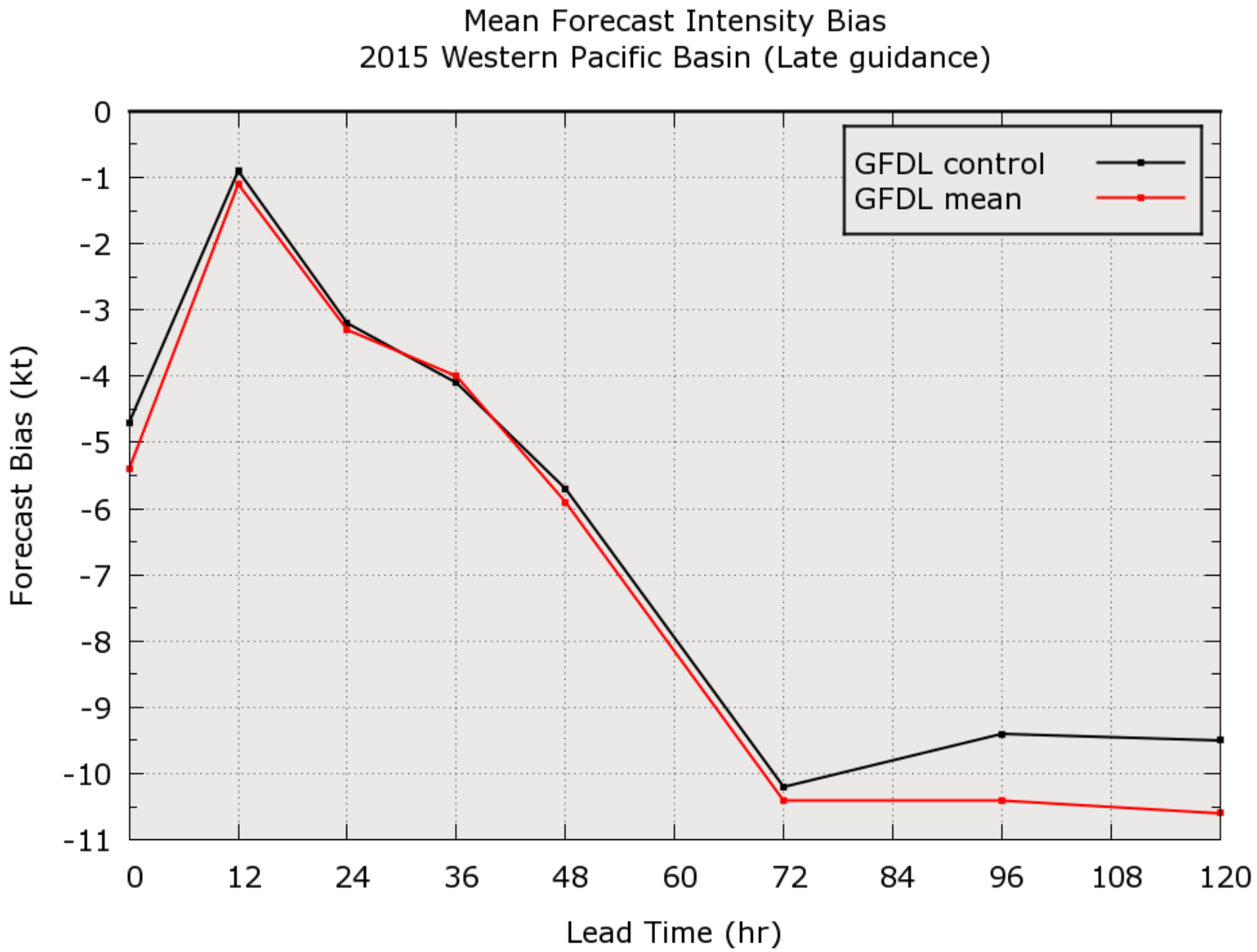
Intensity Bias (Atlantic)



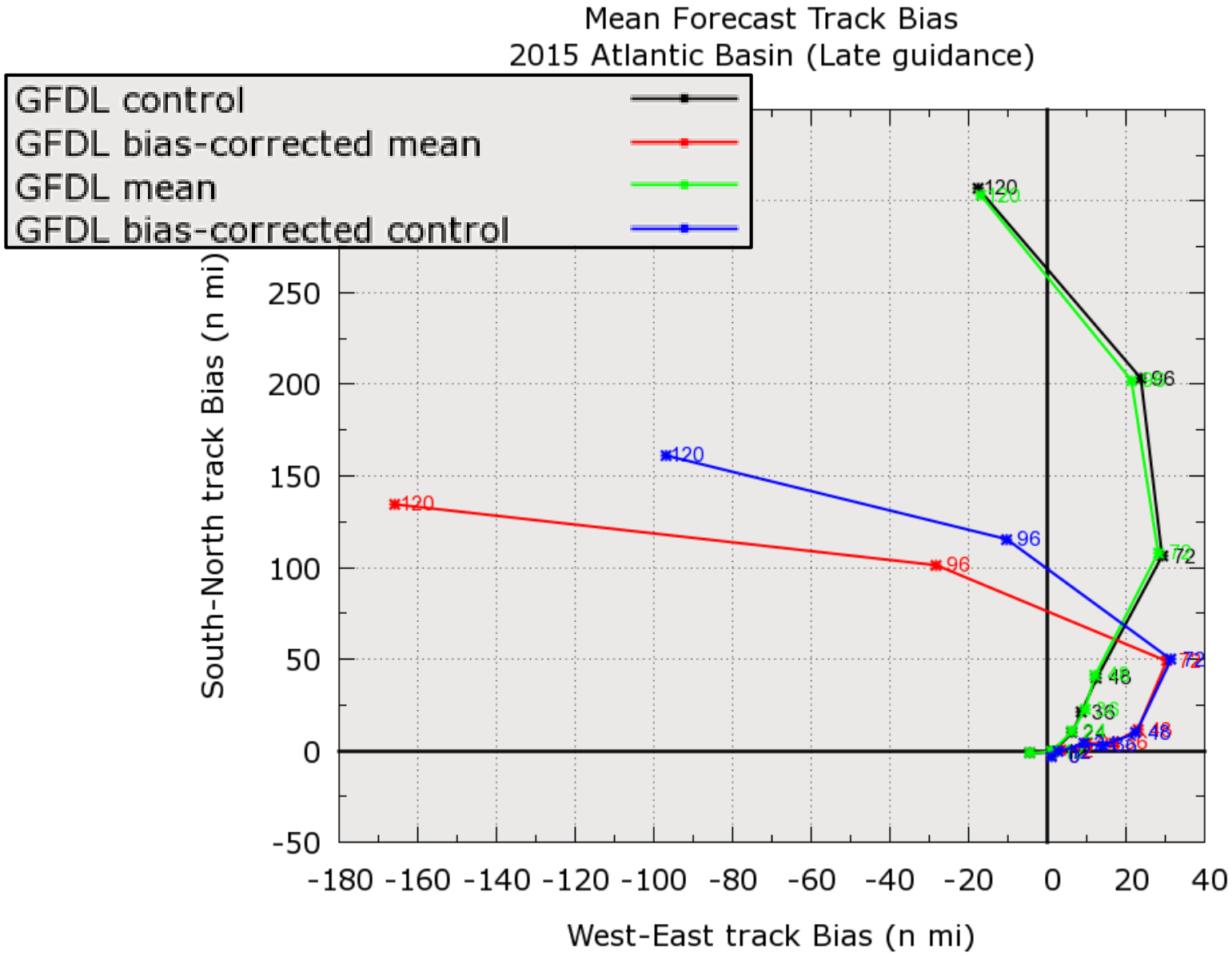
Intensity Bias (Eastern Pacific)



Intensity Bias (Western Pacific)

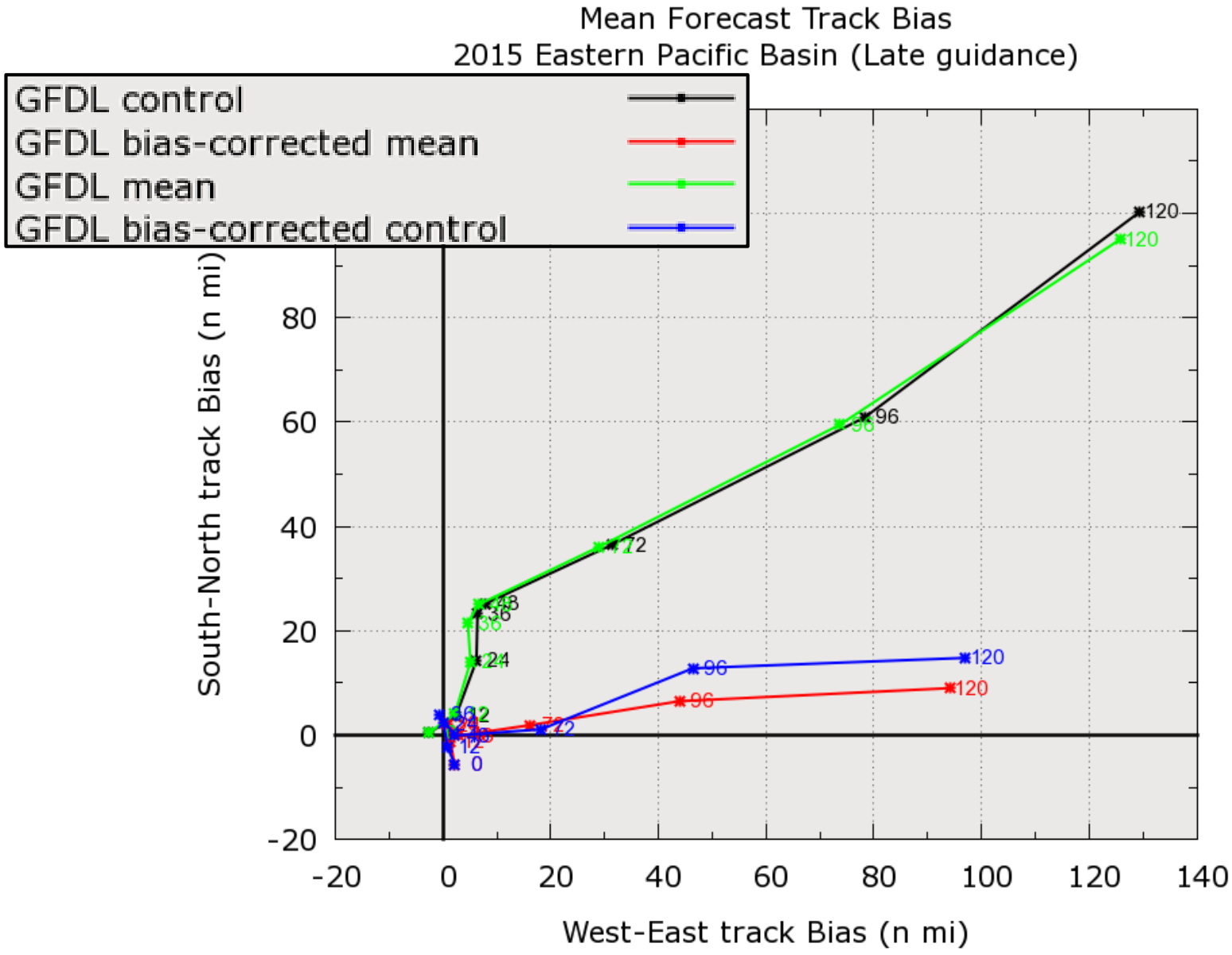


Track Bias (Atlantic)



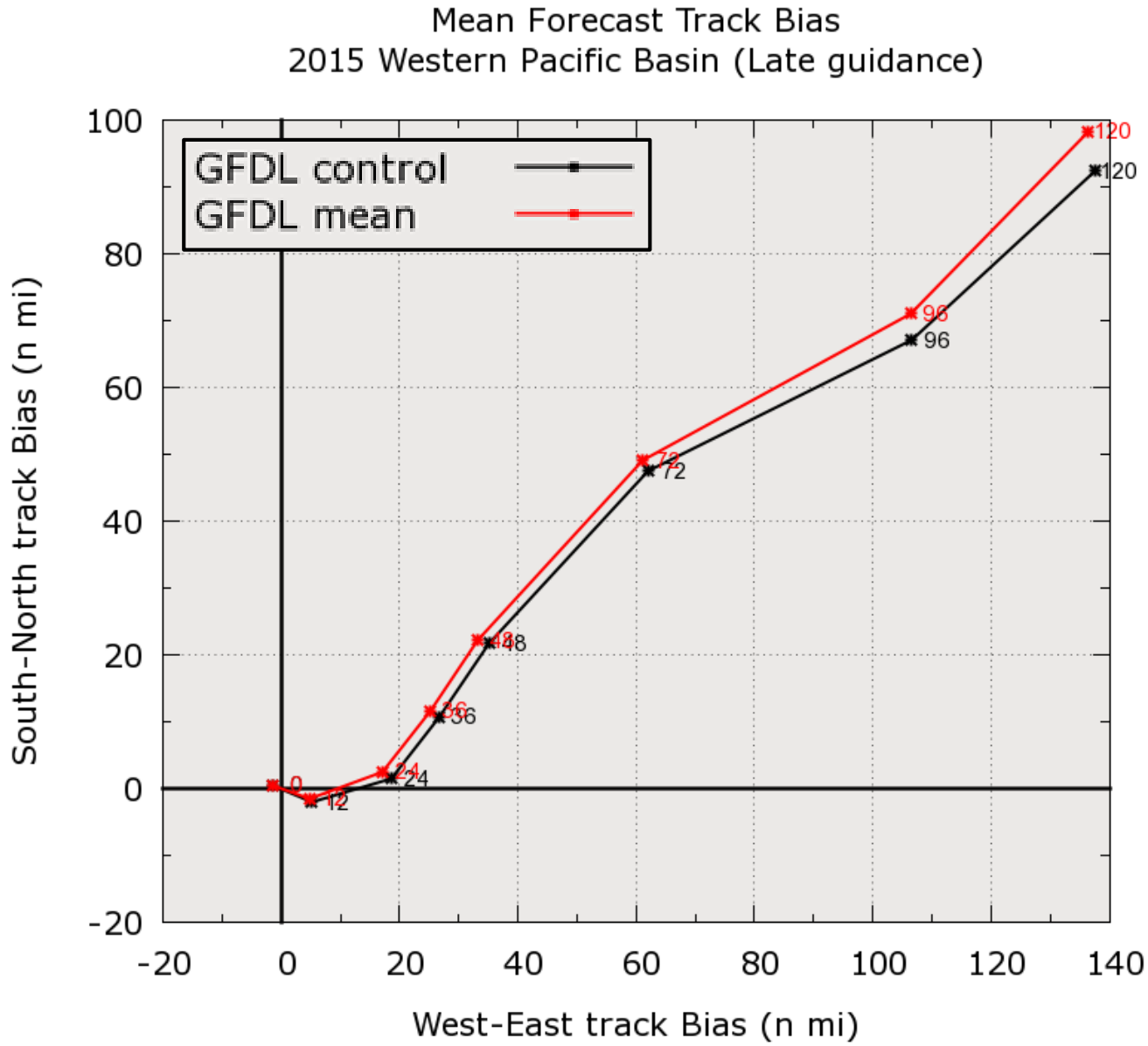
CASES at HR 0, 12, 24, 36, 48, 72, 96, 120 = 217, 196, 174, 151, 131, 97, 66, 48

Track Bias (Eastern Pacific)



CASES at HR 0, 12, 24, 36, 48, 72, 96, 120 = 281, 263, 244, 224, 202, 166, 133, 104

Track Bias (Western Pacific)



CASES at HR 0, 12, 24, 36, 48, 72, 96, 120 = 280, 262, 238, 217, 195, 164, 135, 111

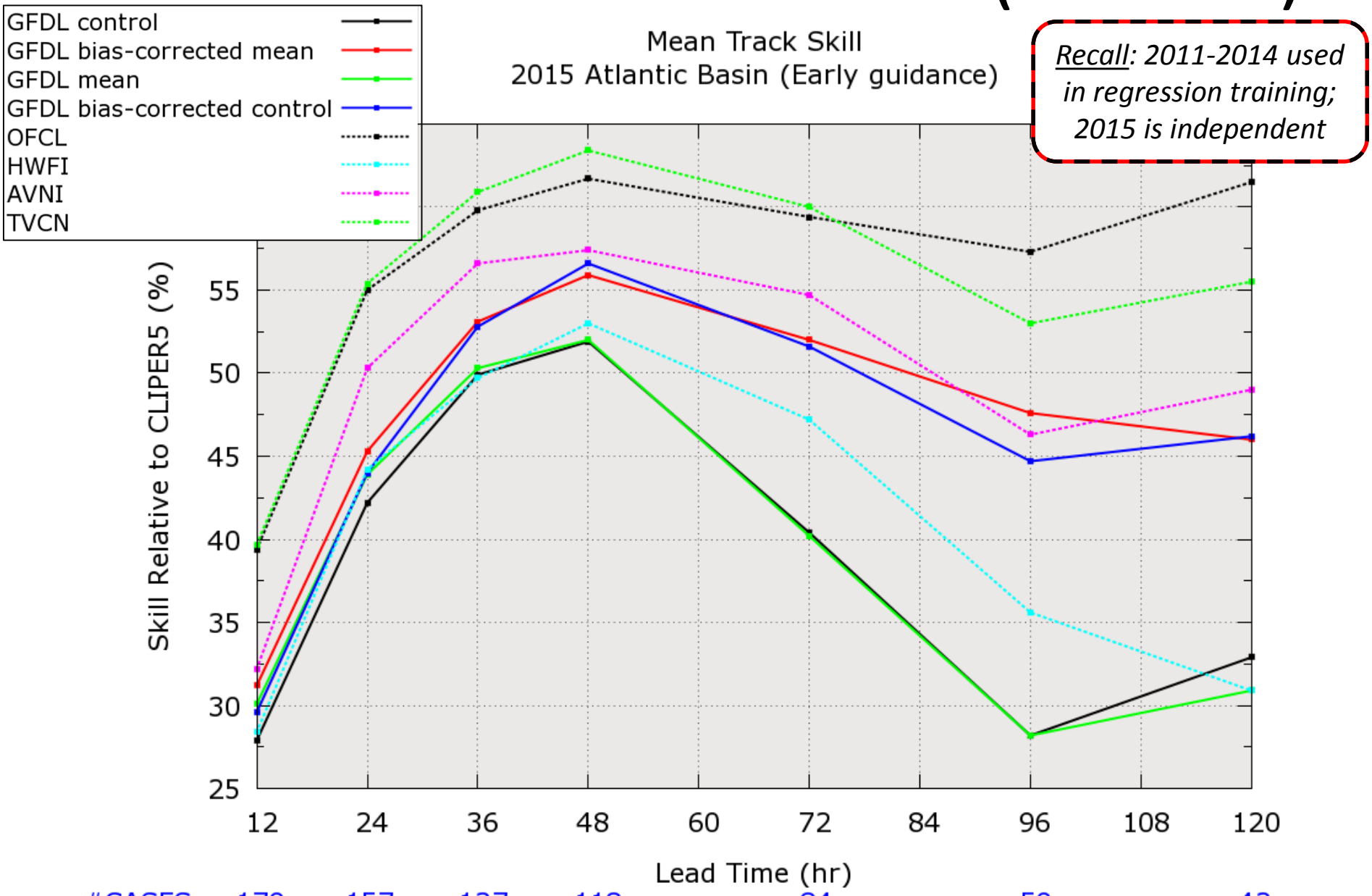
Model Verification

Comparisons with Operational Guidance *(verifications in skill space)*

Point of clarification for next series of figures:

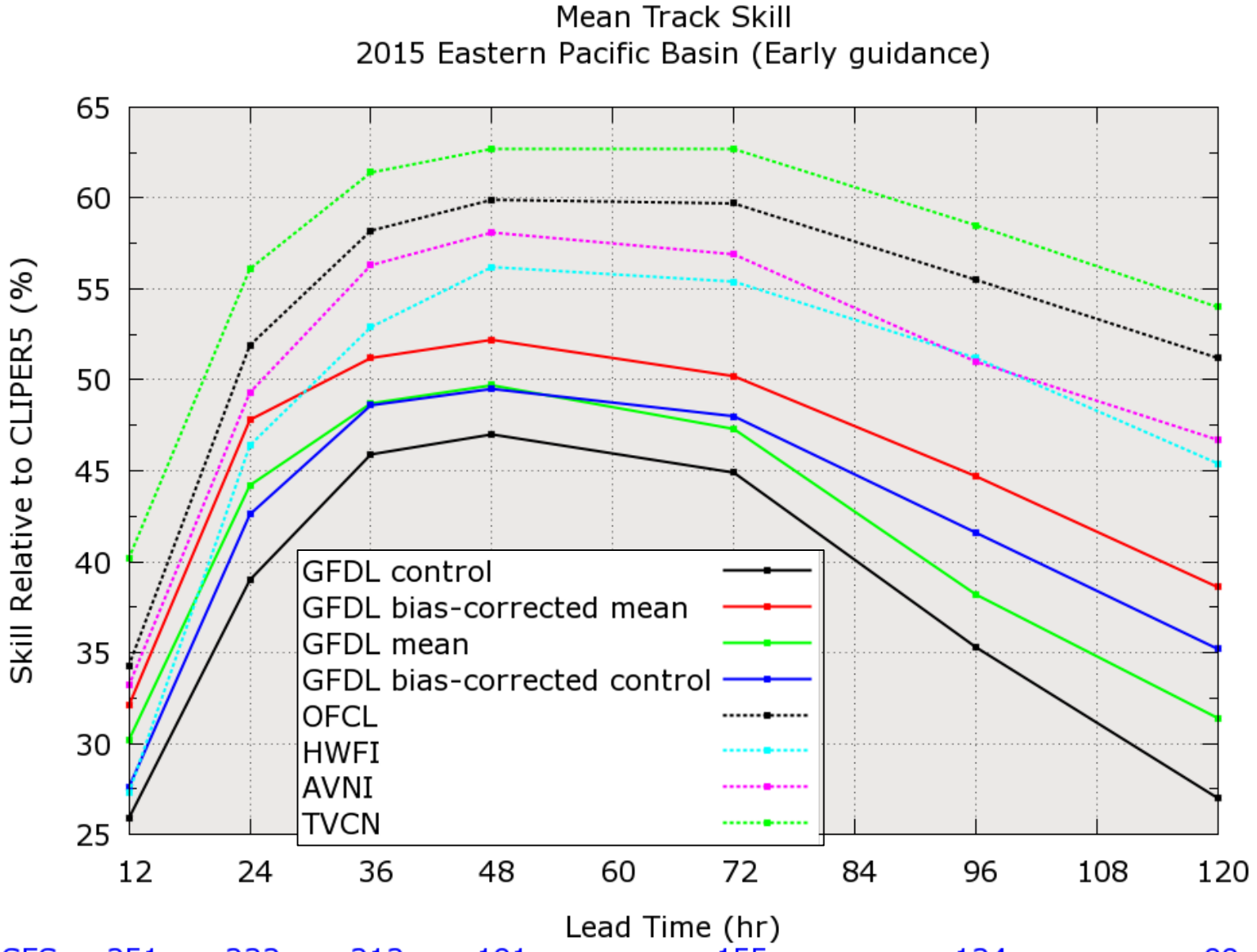
- HWFI, AVNI, and TVCN were taken from the operational a-decks and represent the guidance available in real-time
  - plotted just for reference

Track Forecast Verification (Atlantic)

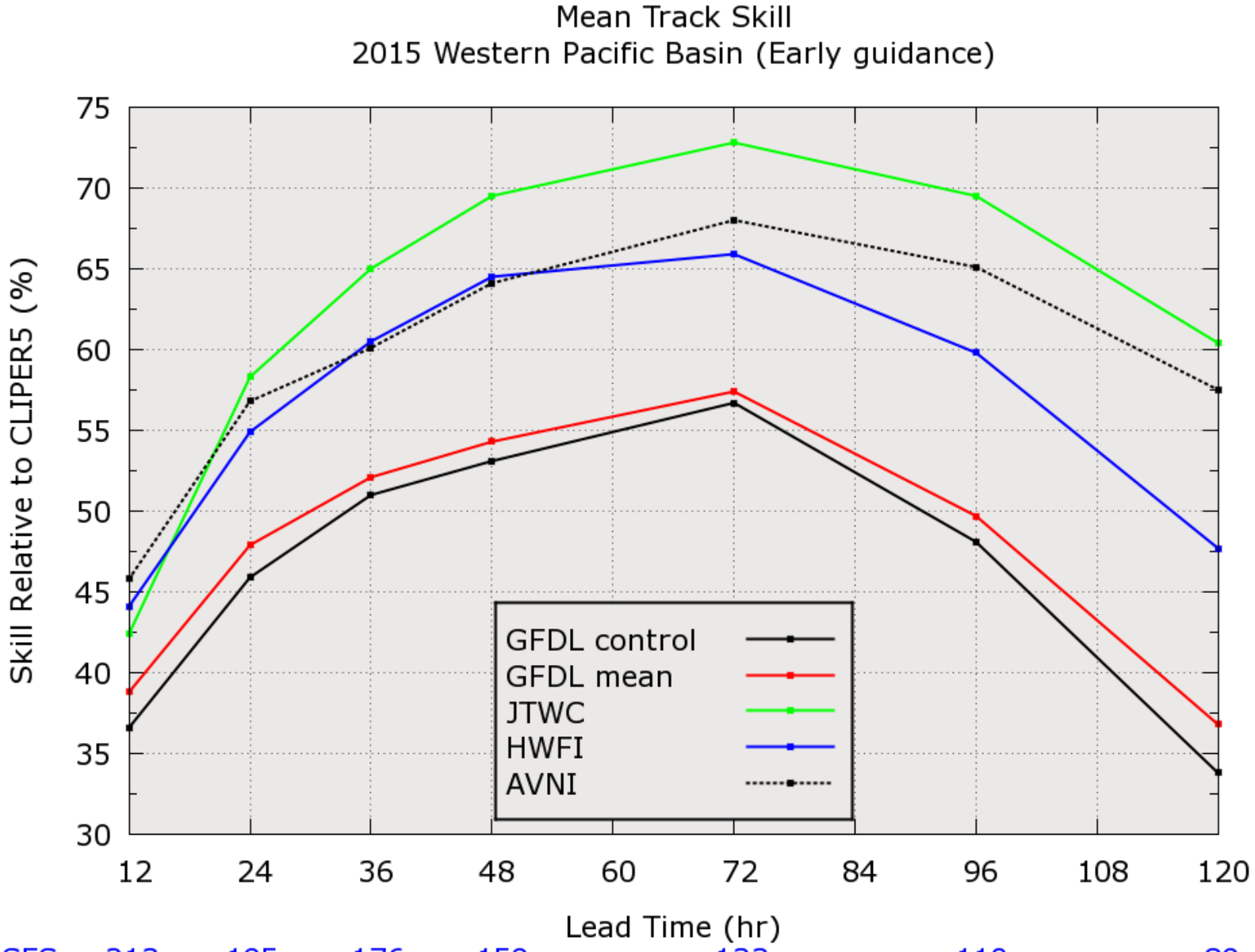


Recall: 2011-2014 used in regression training; 2015 is independent

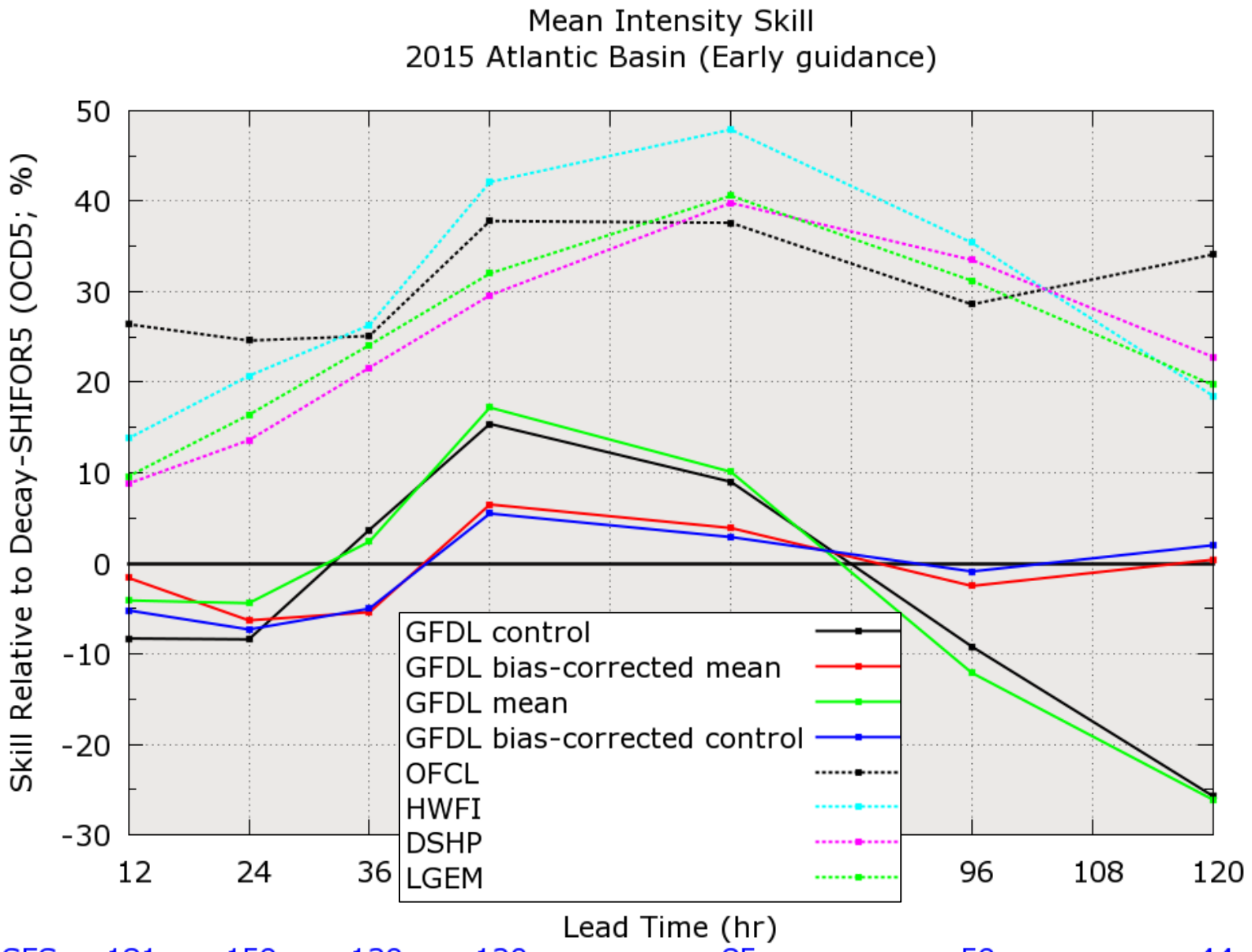
Track Forecast Verification (Eastern Pacific)



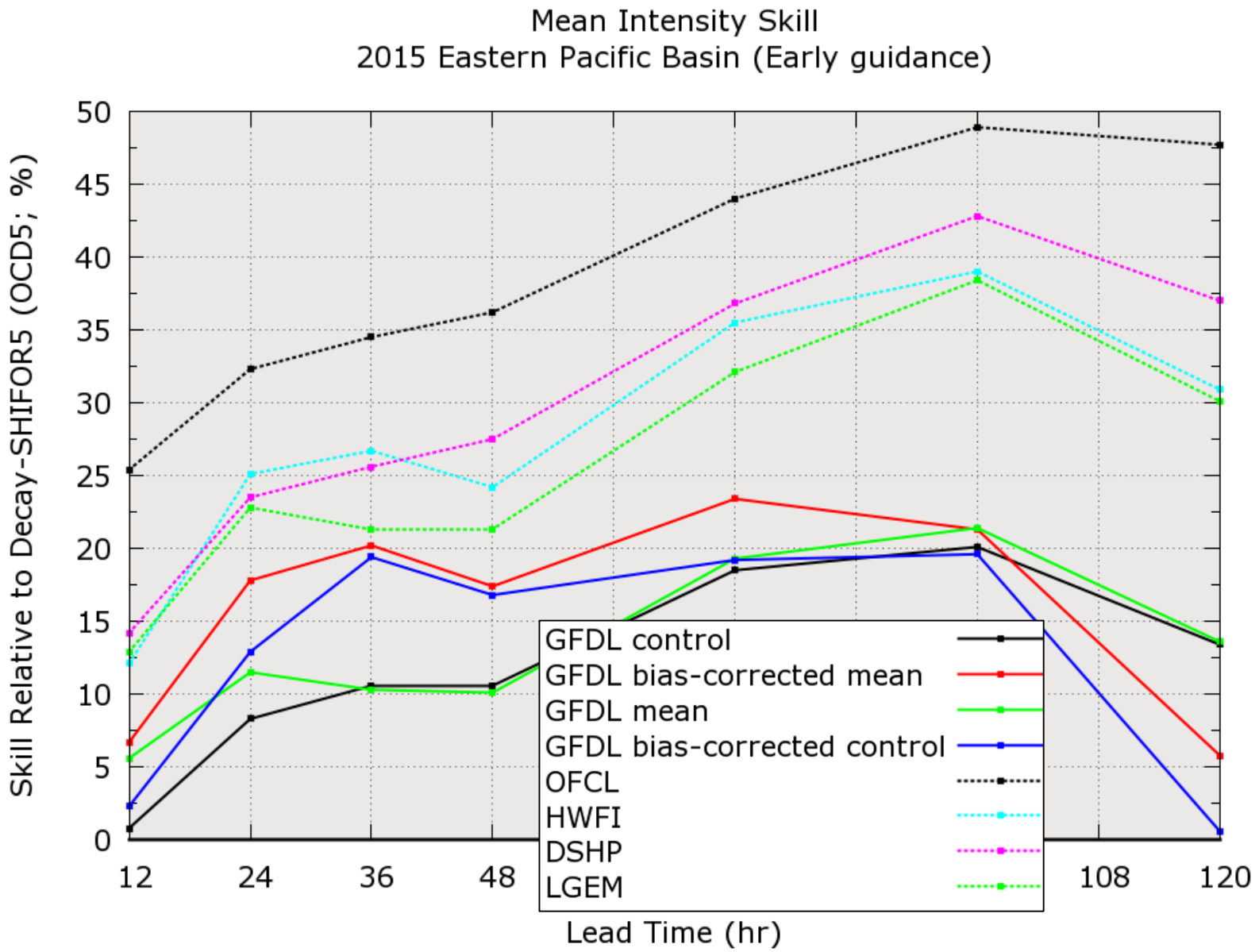
Track Forecast Verification (Western Pacific)



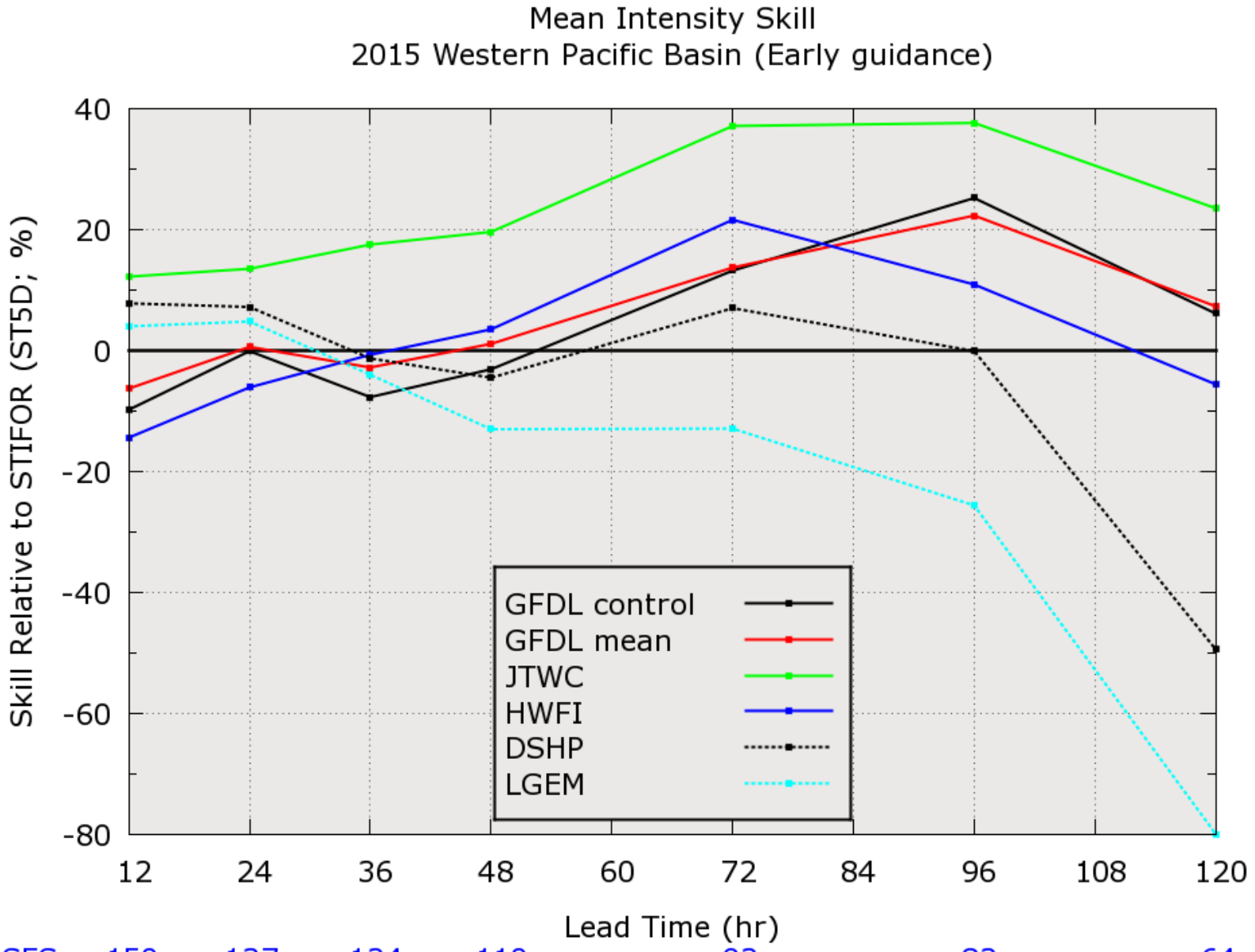
Intensity Forecast Verification (Atlantic)



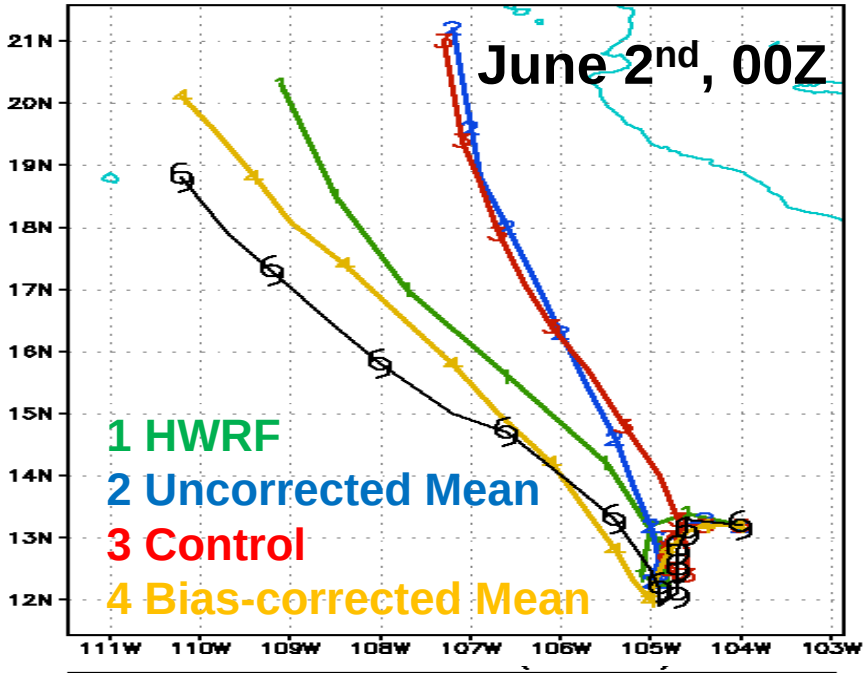
Intensity Forecast Verification (Eastern Pacific)

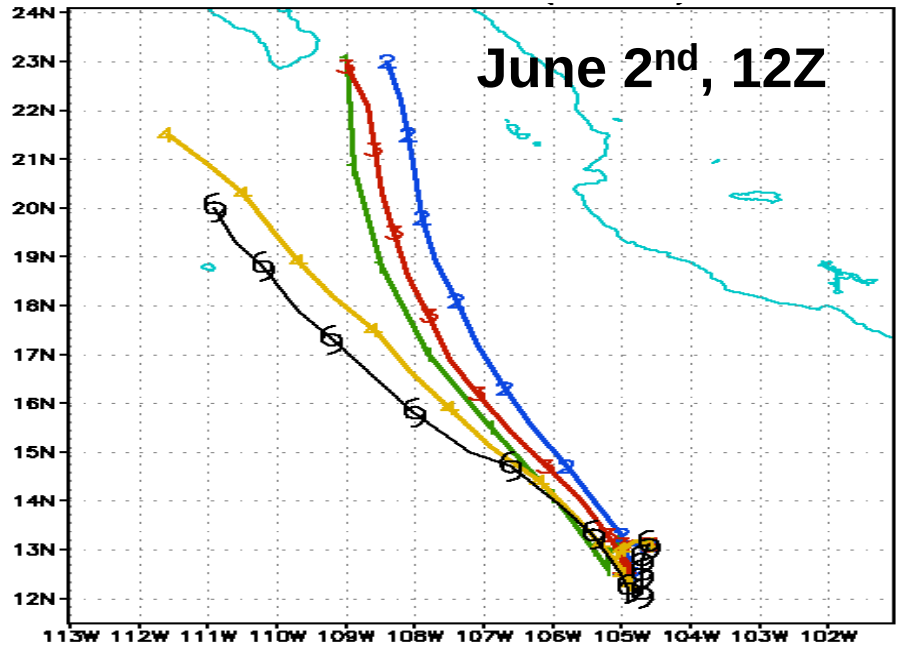


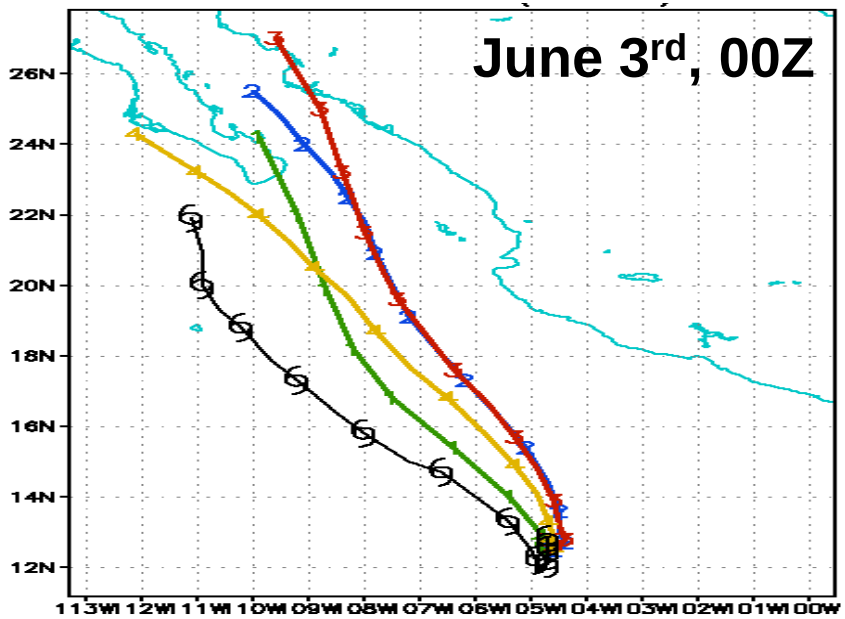
Intensity Forecast Verification (Western Pacific)

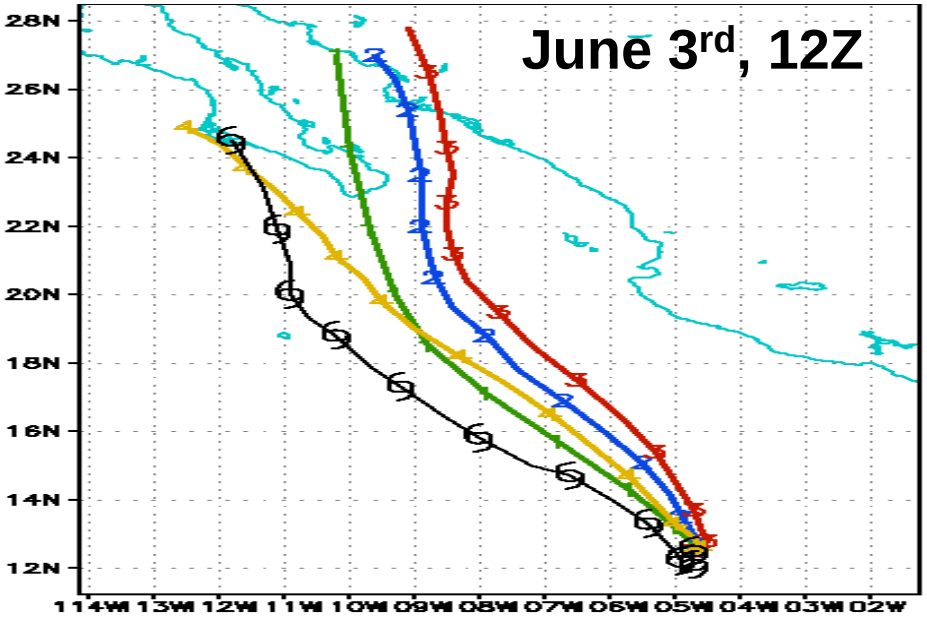


Example: Hurricane Blanca (02E 2015)

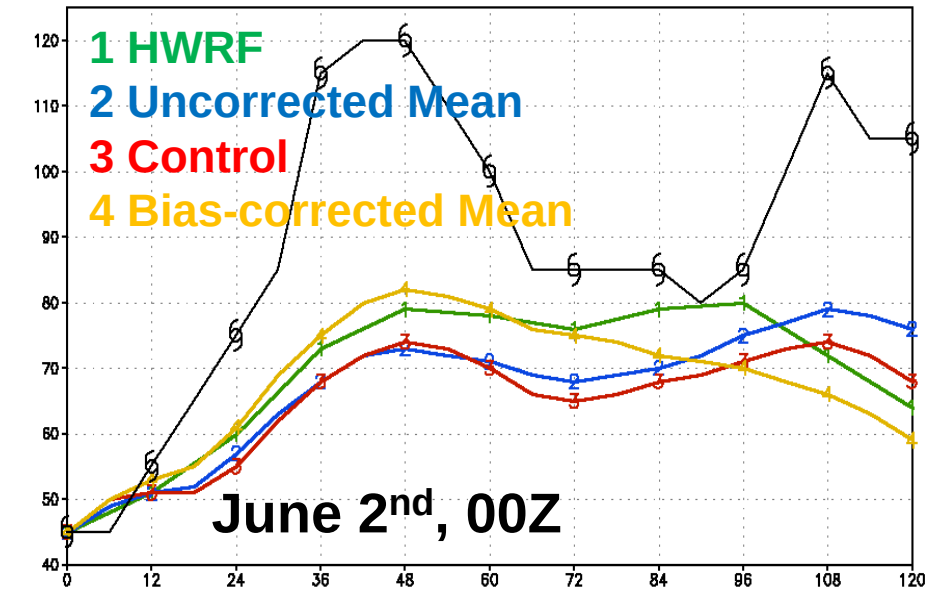


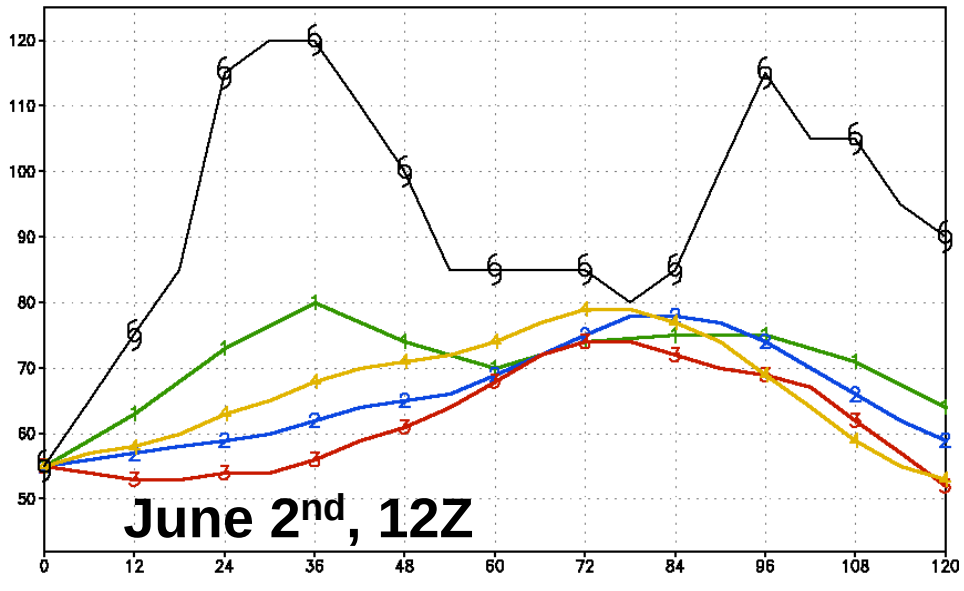


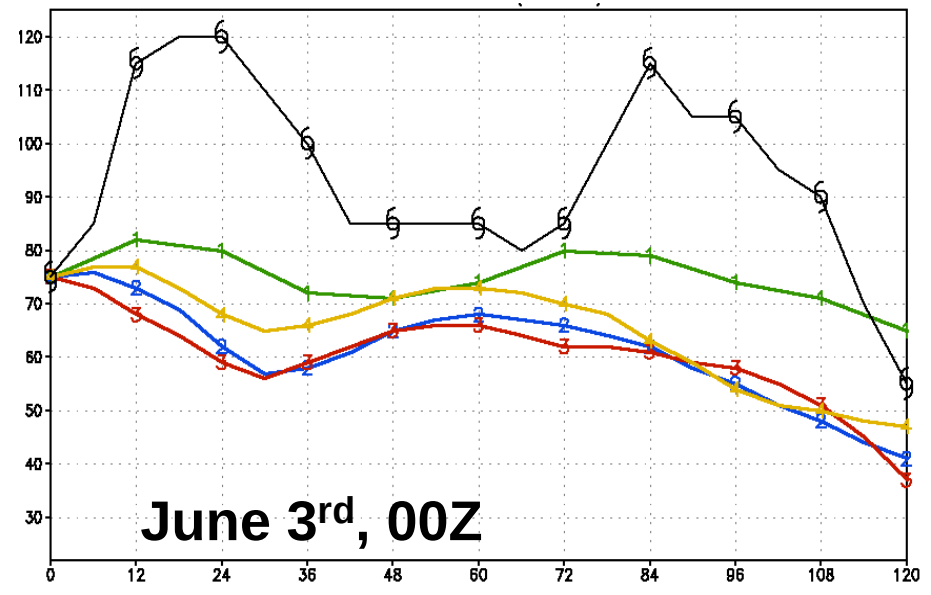


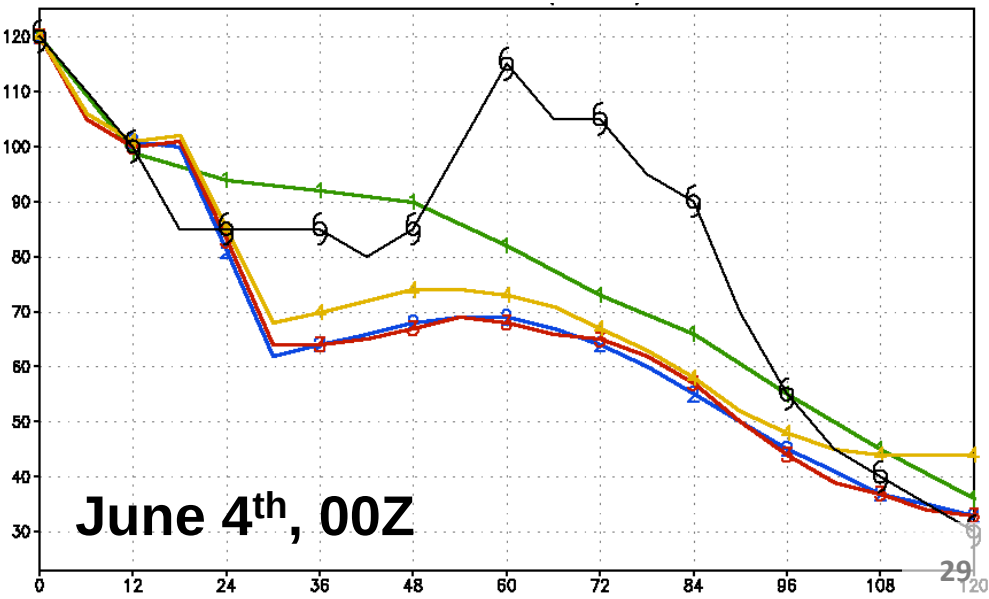


Example: Hurricane Blanca (02E 2015)

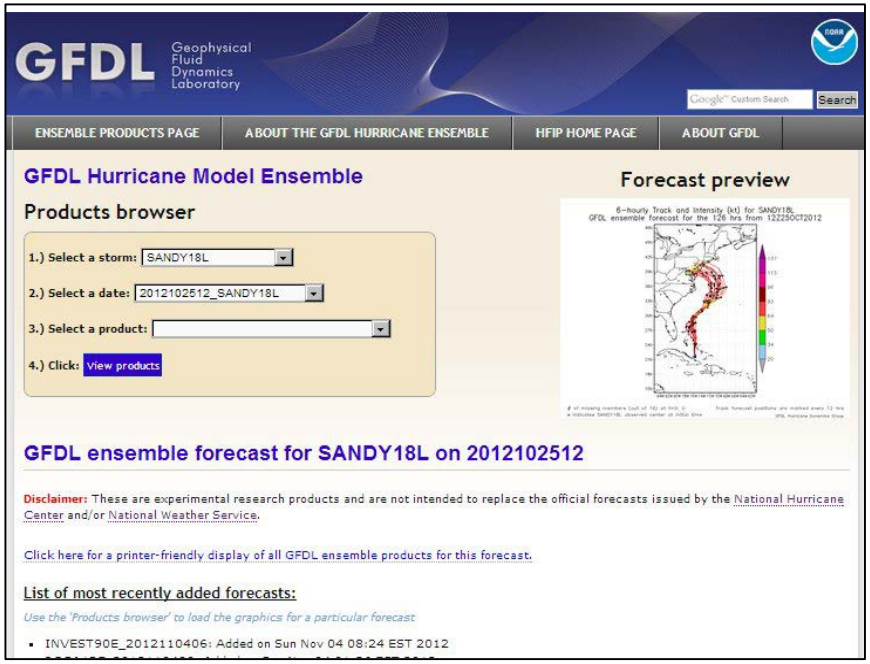


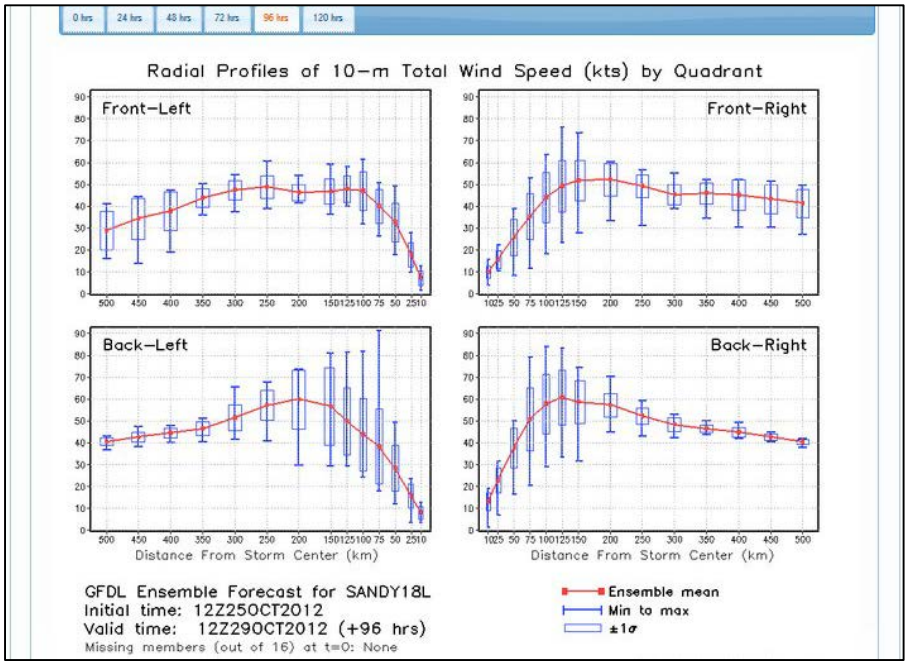


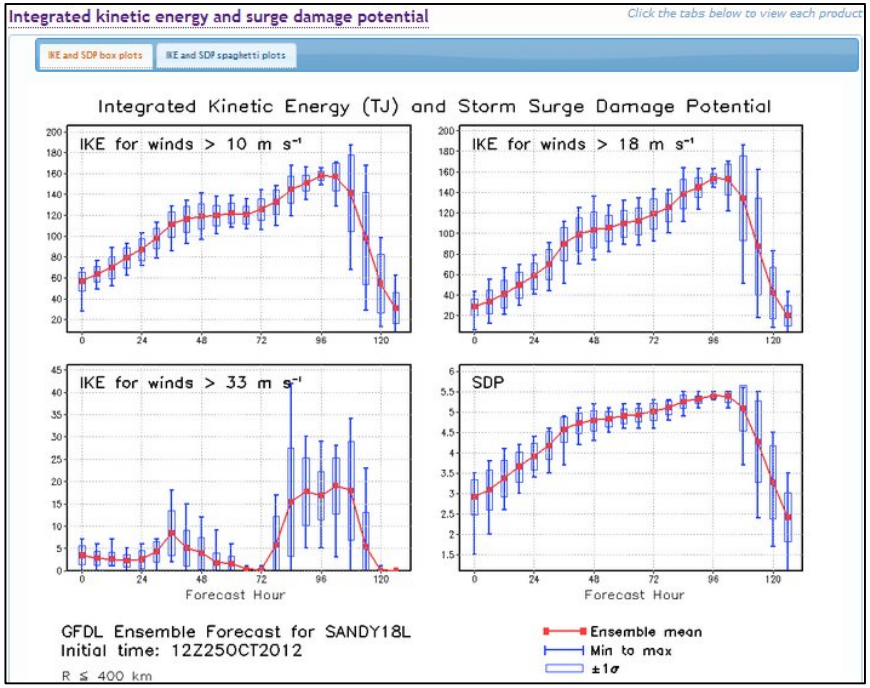


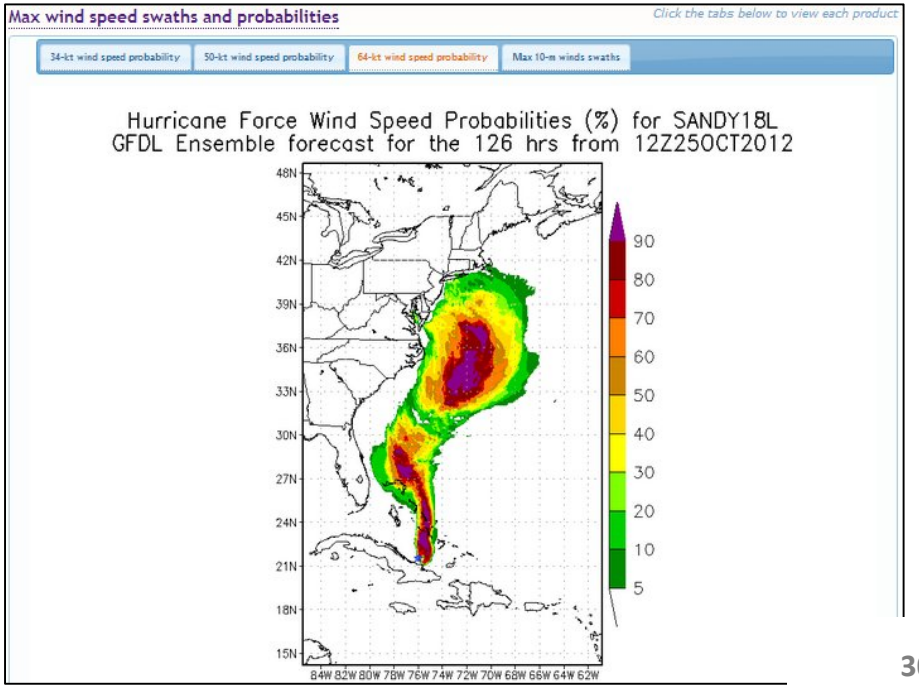


data1.gfdl.noaa.gov/hurricane/gfdl_ensemble (or search for "GFDL ensemble")

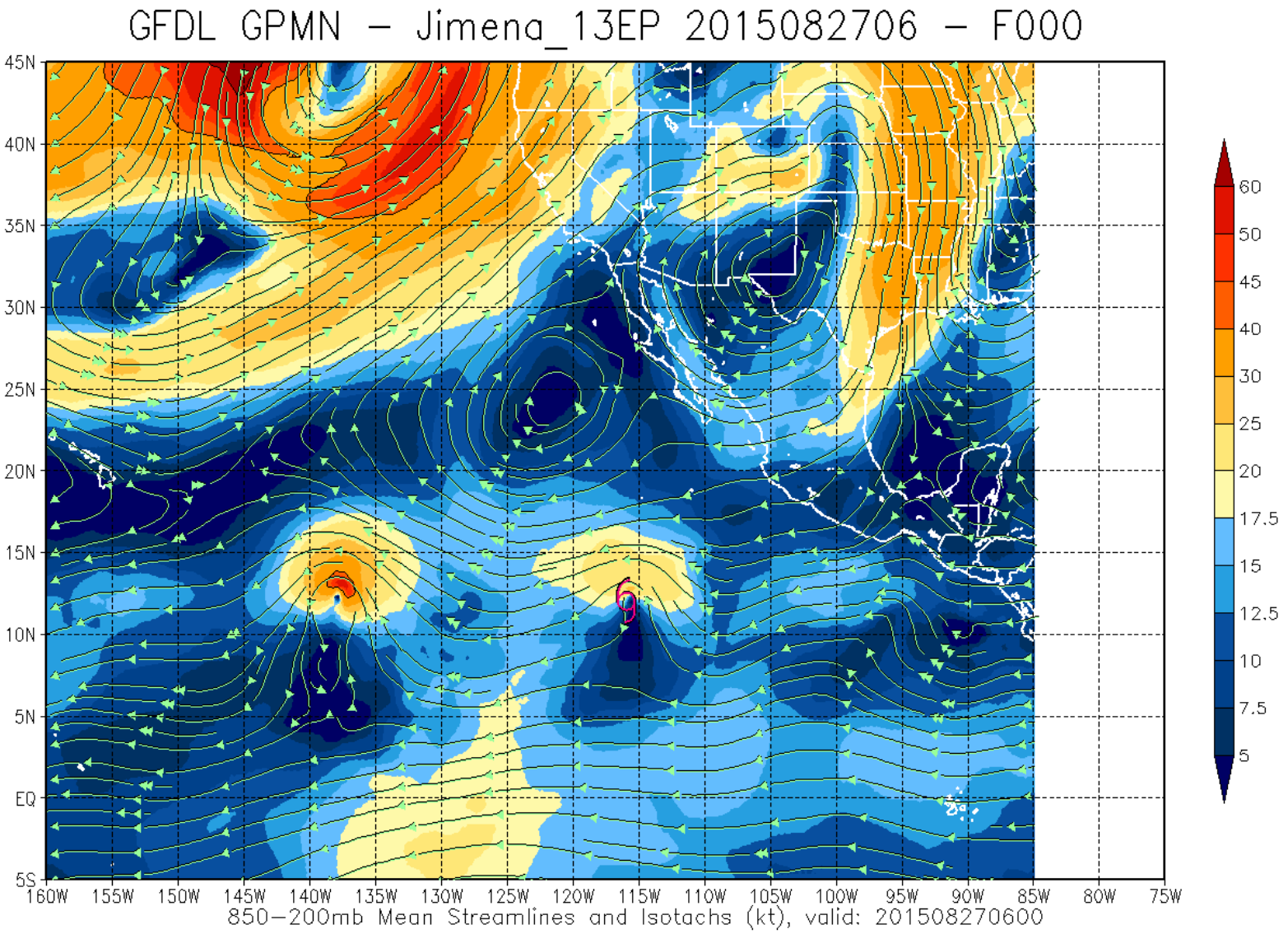




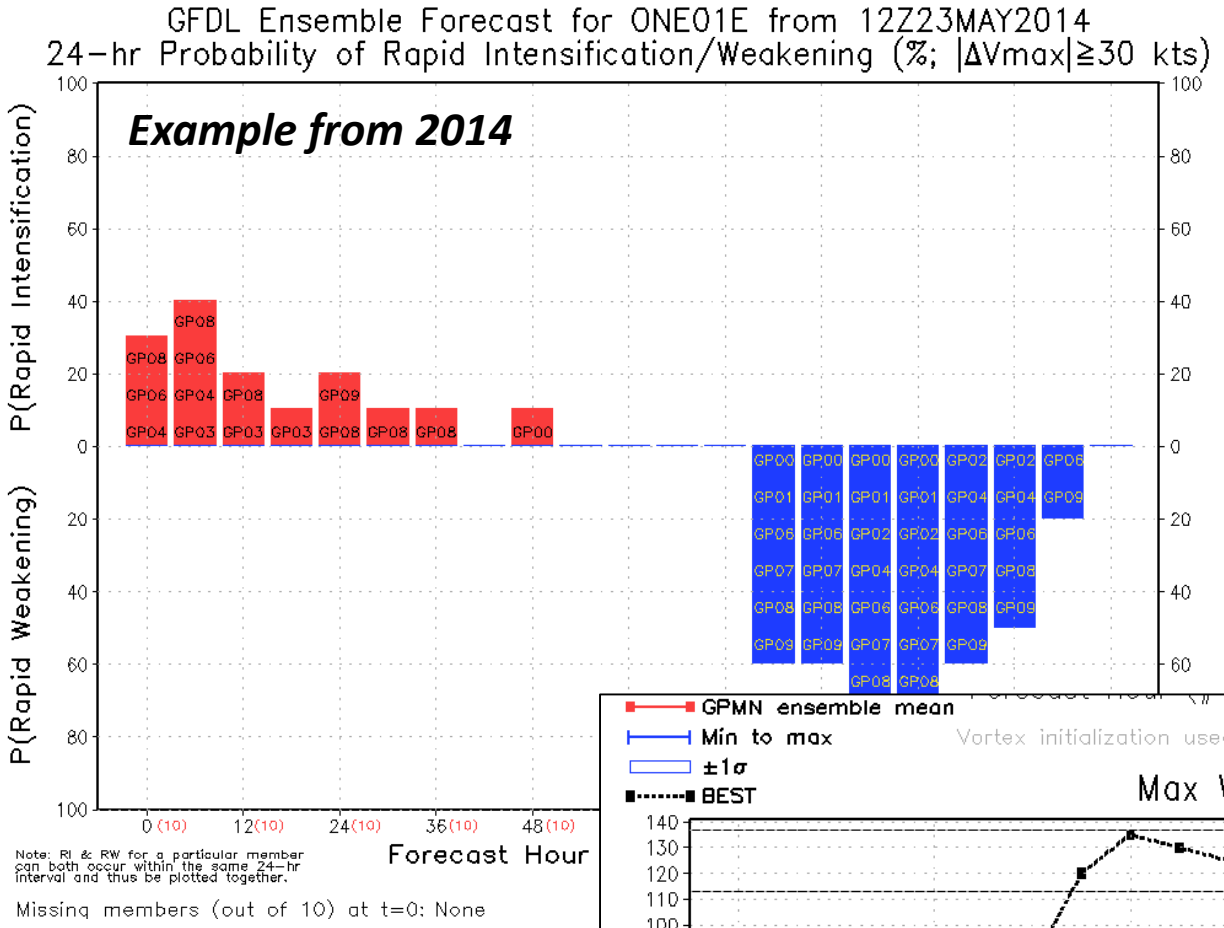


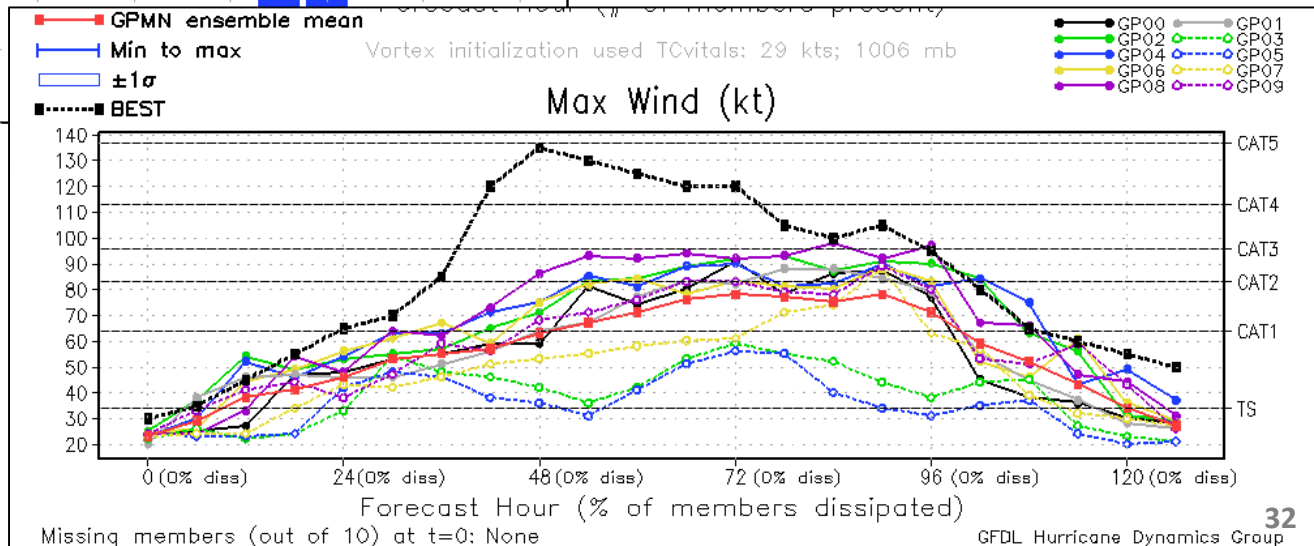


Ensemble Products: Deep-layer Mean Wind



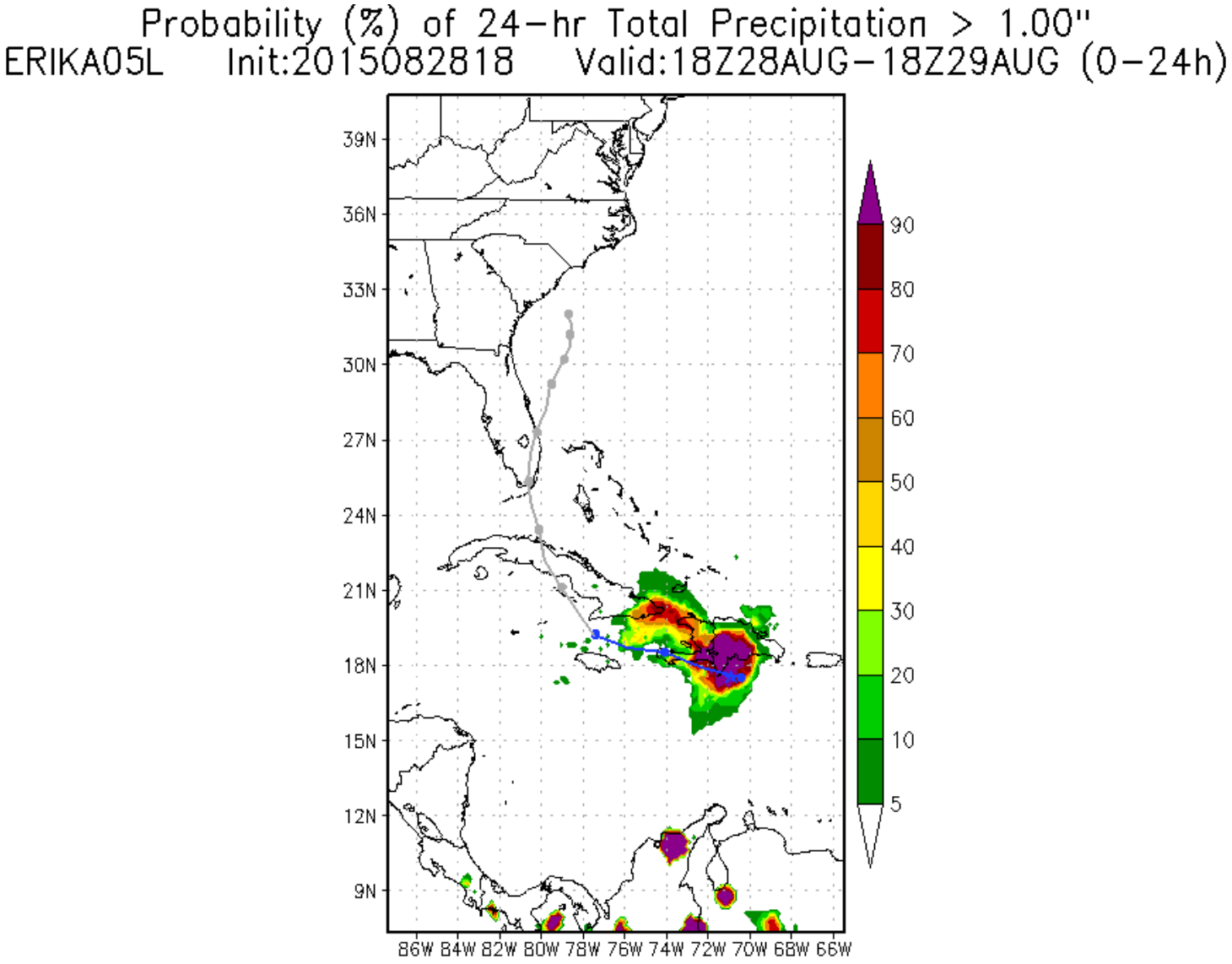
Ensemble Probability Products: Rapid Intensification





- RI not forecasted well by our ensemble this year in any basin
- Ensemble is able to reach Cat. 5 intensity

Ensemble Probability Products: Precipitation Swaths



# of missing members (out of 12) at t=0 hrs: 0

★ indicates ERIKA05L observed center at initial time

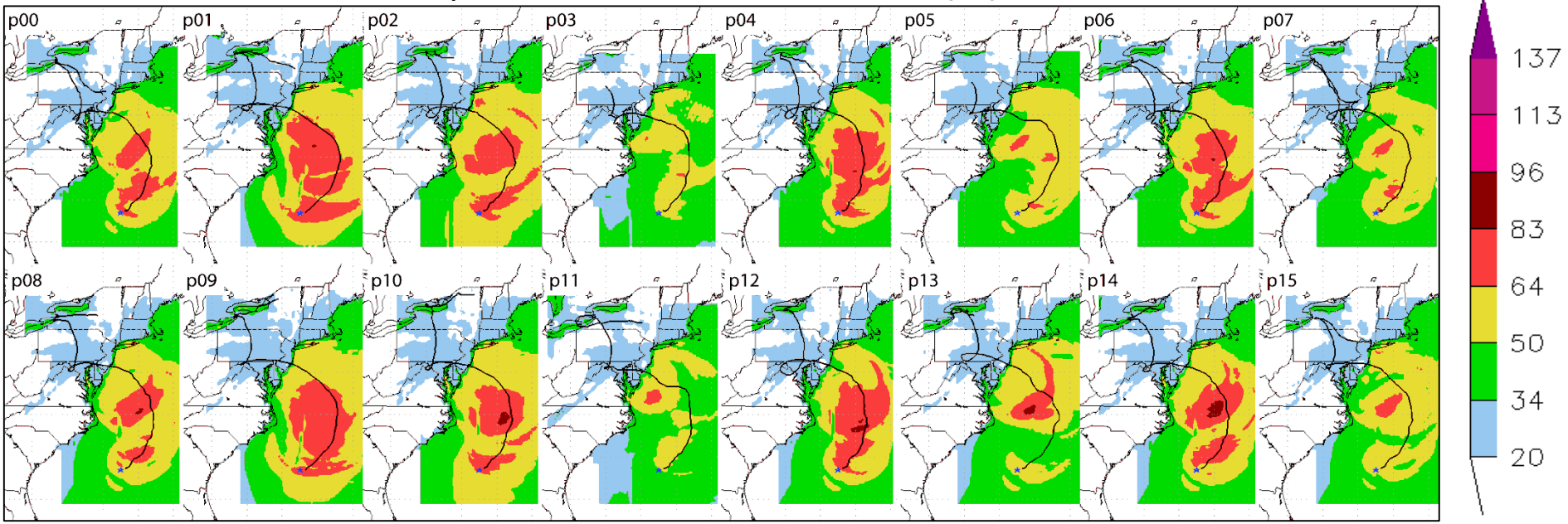
GPMN 0-24 hr forecasted track highlighted in blue

GFDL Hurricane Dynamics Group

Ensemble Probability Products: Wind Speed Swaths

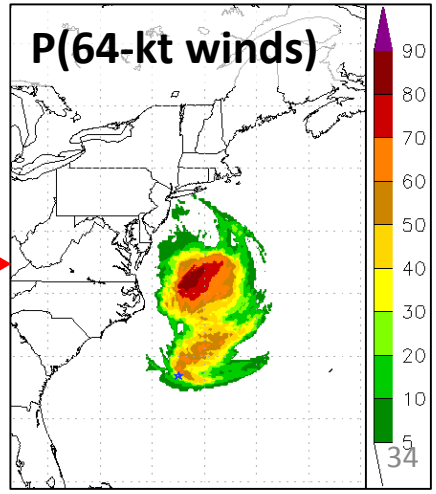
Hurricane Sandy: 2012102812

2012 GFDL ensemble example **Maximum 10-m winds (kt)**



Forecast tracks shown by black trace in each plot

In 2012, graphical products showing the probabilities of 34-, 50-, and **64-kt winds** were added to the suite of ensemble products delivered in near real-time to the GFDL ensemble website.

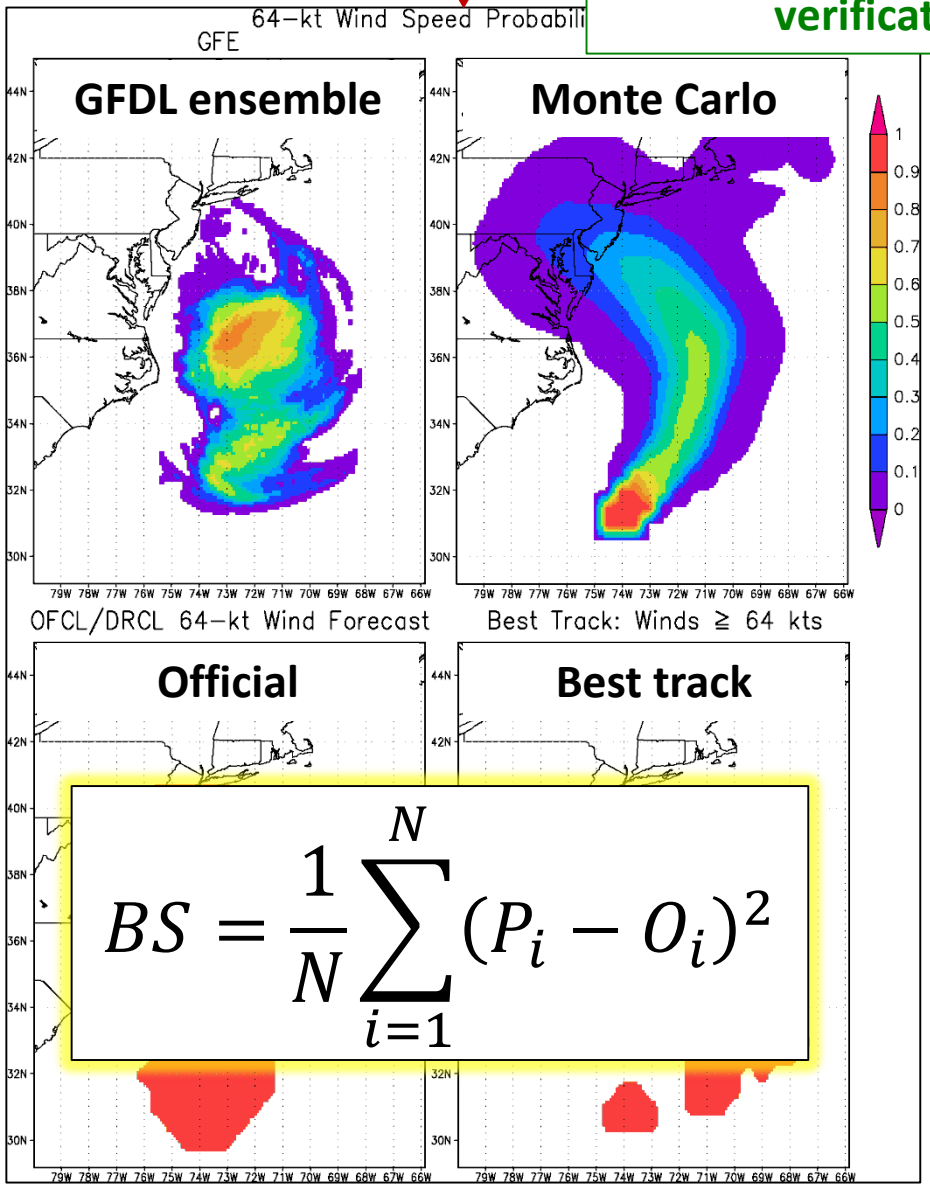


Ensemble Wind Speed Probability Verification

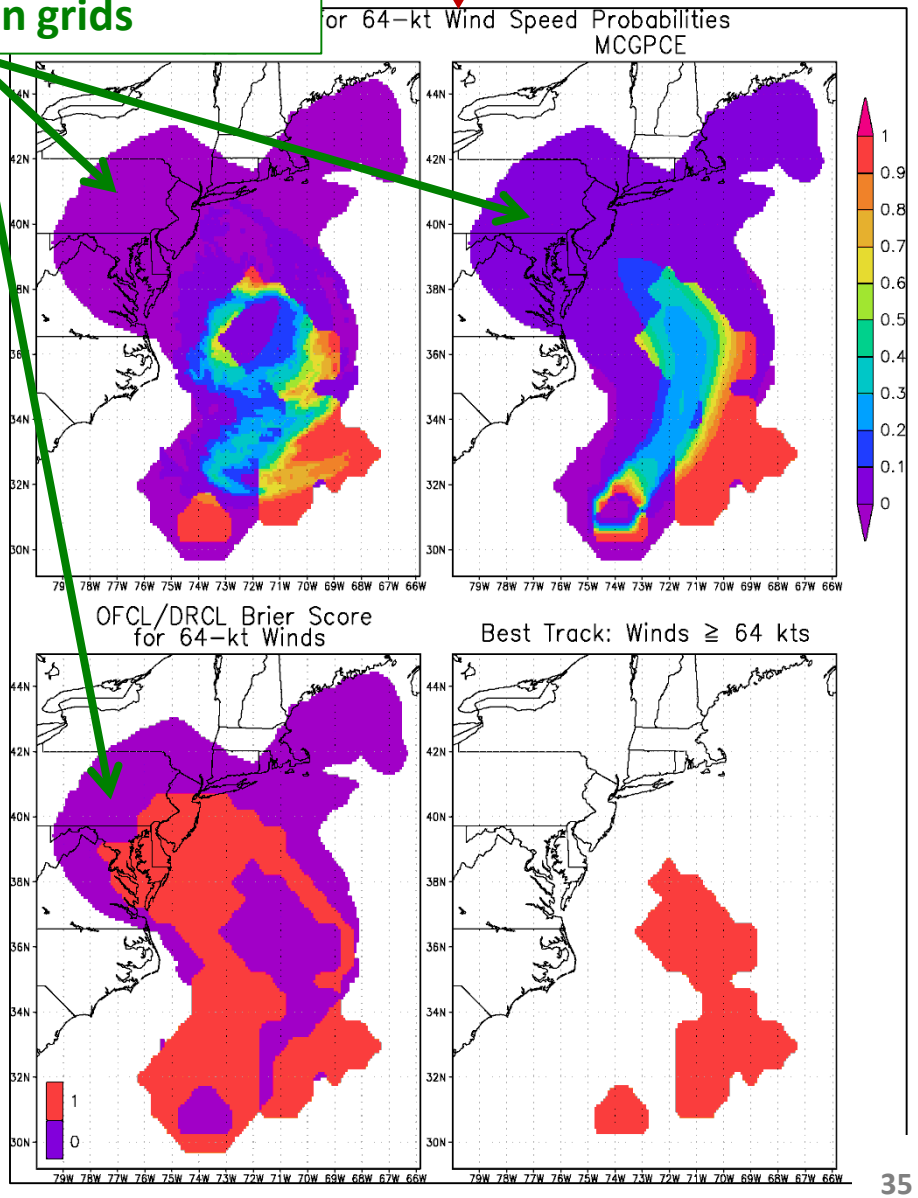
Inputs

Adjustable, equivalent area verification grids

Brier Score



$$BS = \frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2$$



Summary

- GFDL ensemble tracks are still too under dispersive, while intensity spread is the largest in the ensemble's history

- Mixed results for uncorrected ensemble mean intensity improvement over the control in the Atlantic; generally more skillful than control in EP and WP

- Bias-corrected mean intensity has shown some potential to improve over the uncorrected mean, but needs reformulation; bias-corrected track generally showed solid improvement over uncorrected mean

Future Work (*part 1 of 2*)

- Research and test new perturbations that increase track skill and spread
  - Initial TC center position
  - Vorticity confinement
  - Stochastic physics
  - Run from different global models

- Run all missing cases in Eastern and Western Pacific
  - Show results at 2016 AMS Conference on Hurricanes and Tropical Meteorology

Suggestions welcomed! matthew.morin@noaa.gov

Future Work (*part 2 of 2*)

- Improve regression method
  - Currently using a simple one that assumes a normal distribution
    - Distribution plots: ftp://ftp.gfdl.noaa.gov/pub/m1m/verif/
  - Generate multilinear regression equations on the fly when all ensemble members are not present
  - Test resampling techniques (e.g., jackknife, bootstrap, etc…)
  - Remove outliers from training phase data set
- Multilinear regression equations for ensemble in West Pac
- Test calculating and using different regression equations for:
  - main development region, Gulf of Mexico, NE Atlantic
  - weak vs. strong storms (Vmax threshold based on having an equal number of cases in each distribution)

Suggestions welcomed! *matthew.morin@noaa.gov*

Future Work (*summary*)

- Research and test new perturbations that increase track skill and spread
- Run all missing cases in Eastern and Western Pacific
- Improve regression method
- Generate multilinear regression equations for ensemble in West Pac
- Test calculating and using different regression equations for different sub-regions and storm conditions

Suggestions welcomed! *matthew.morin@noaa.gov*
http://data1.gfdl.noaa.gov/hurricane/gfdl_ensemble/

Extra Slides

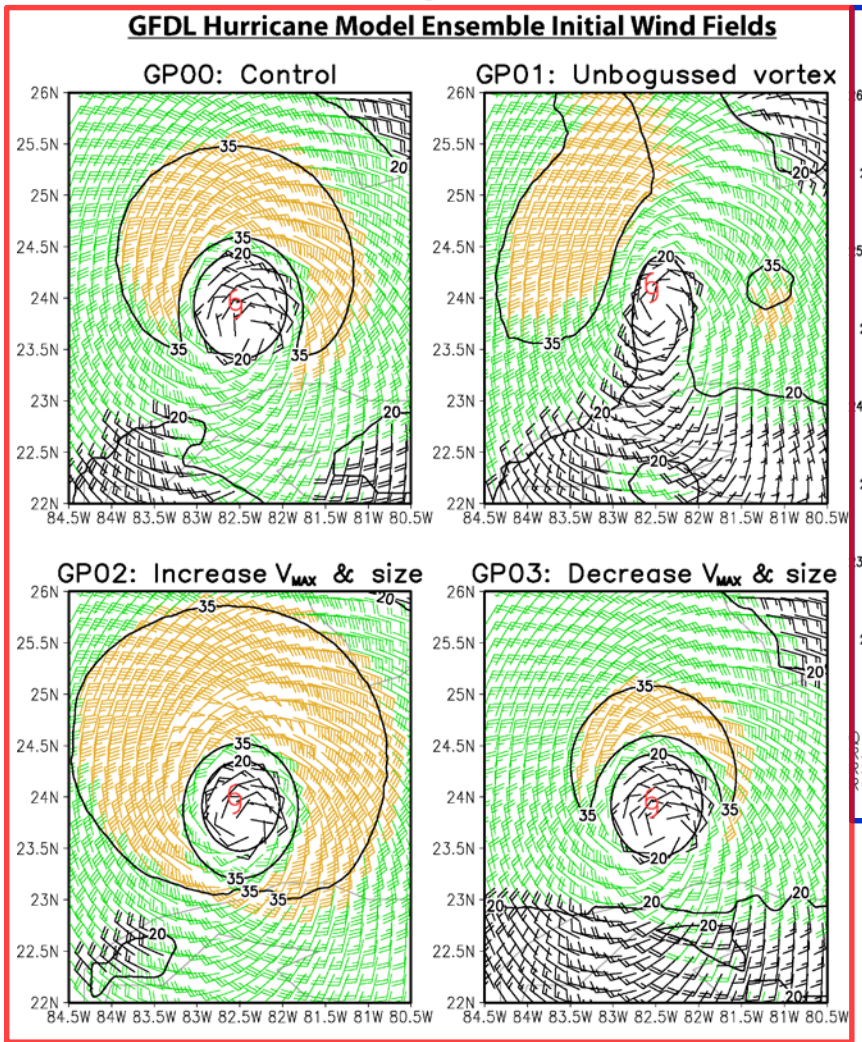
Vortex Size & Intensity Perturbations

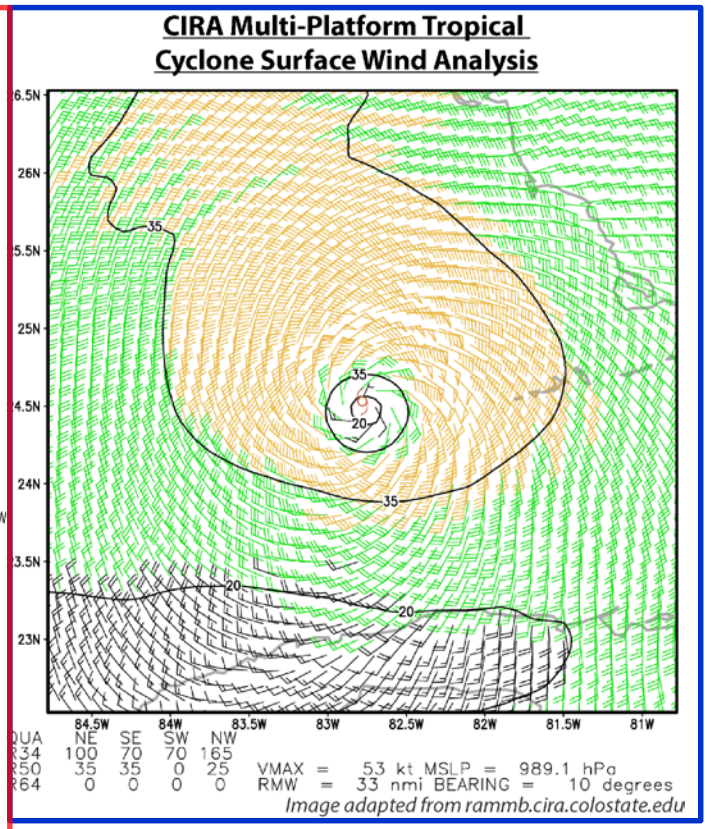
Example: **Plan view at 10 meters above ground level**

NHC-reported surface V_{MAX} (from TC vitals) ≈ 55 knots

Tropical Storm Isaac 10-m Winds (knots)

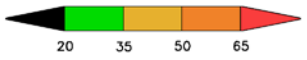
Analysis based on model's initial conditions





Analysis based on observations

00 UTC 27 August 2012



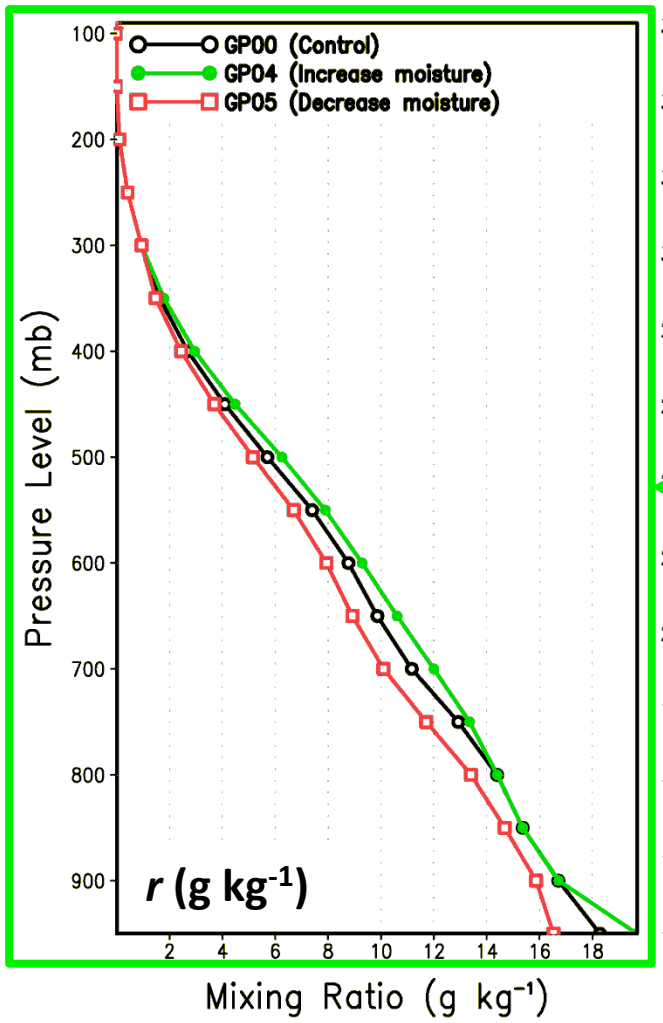
GFDL Hurricane Dynamics Group

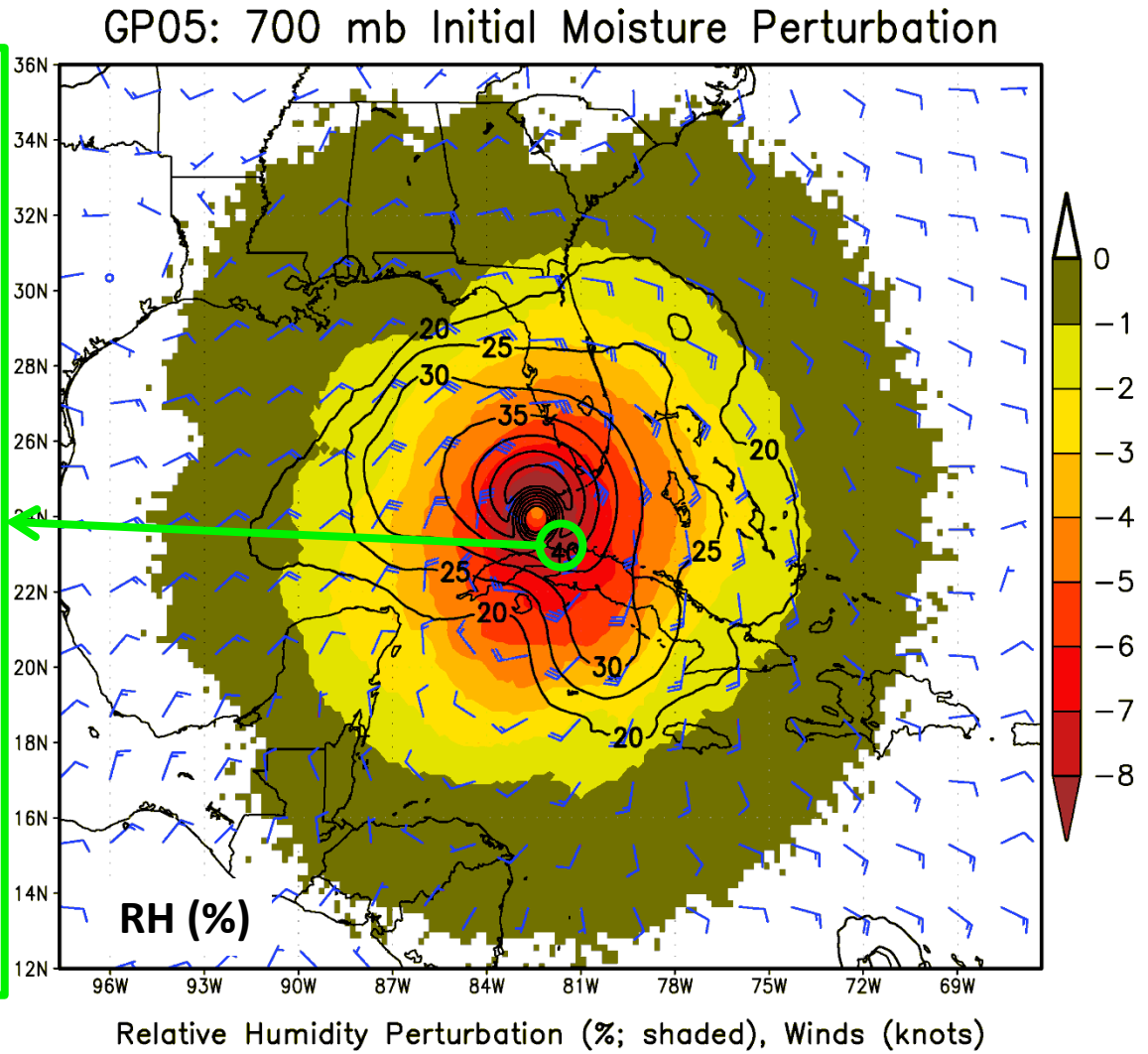
Moisture Perturbations

GFDL Hurricane Model Ensemble Initial Moisture Perturbations

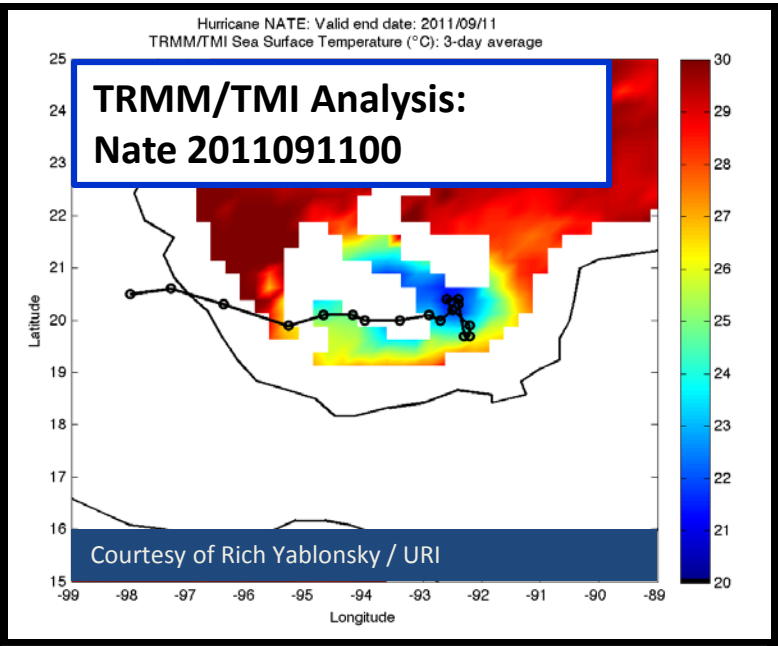
Tropical Storm Isaac: 00 UTC 27 August 2012

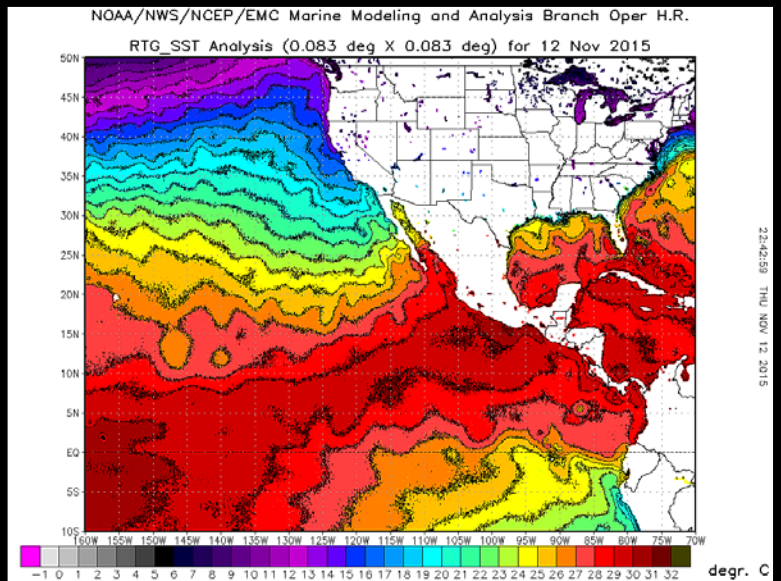
Example:

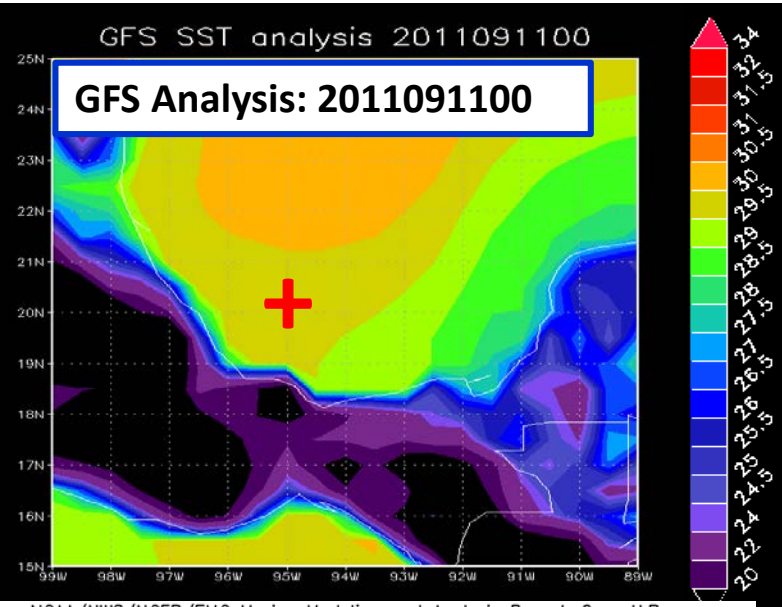


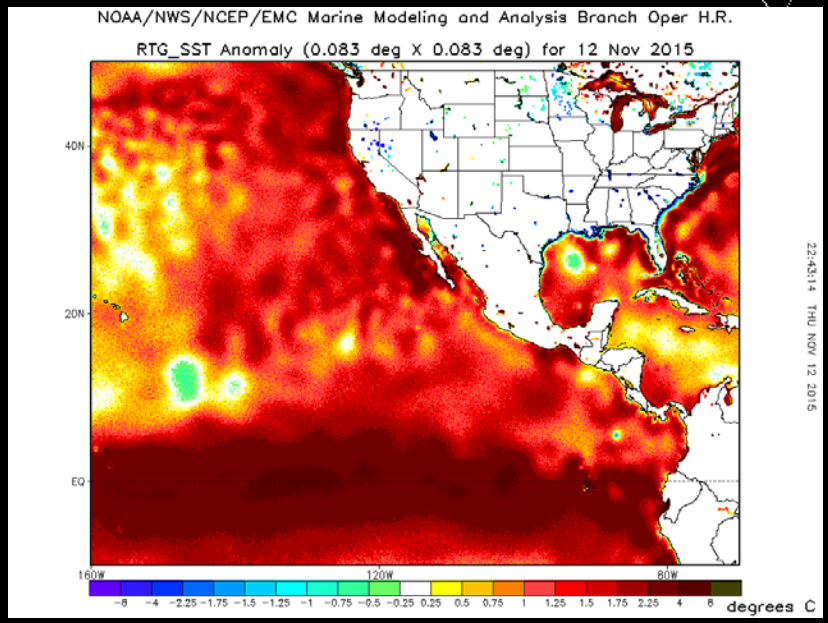


SST Perturbation Motivation



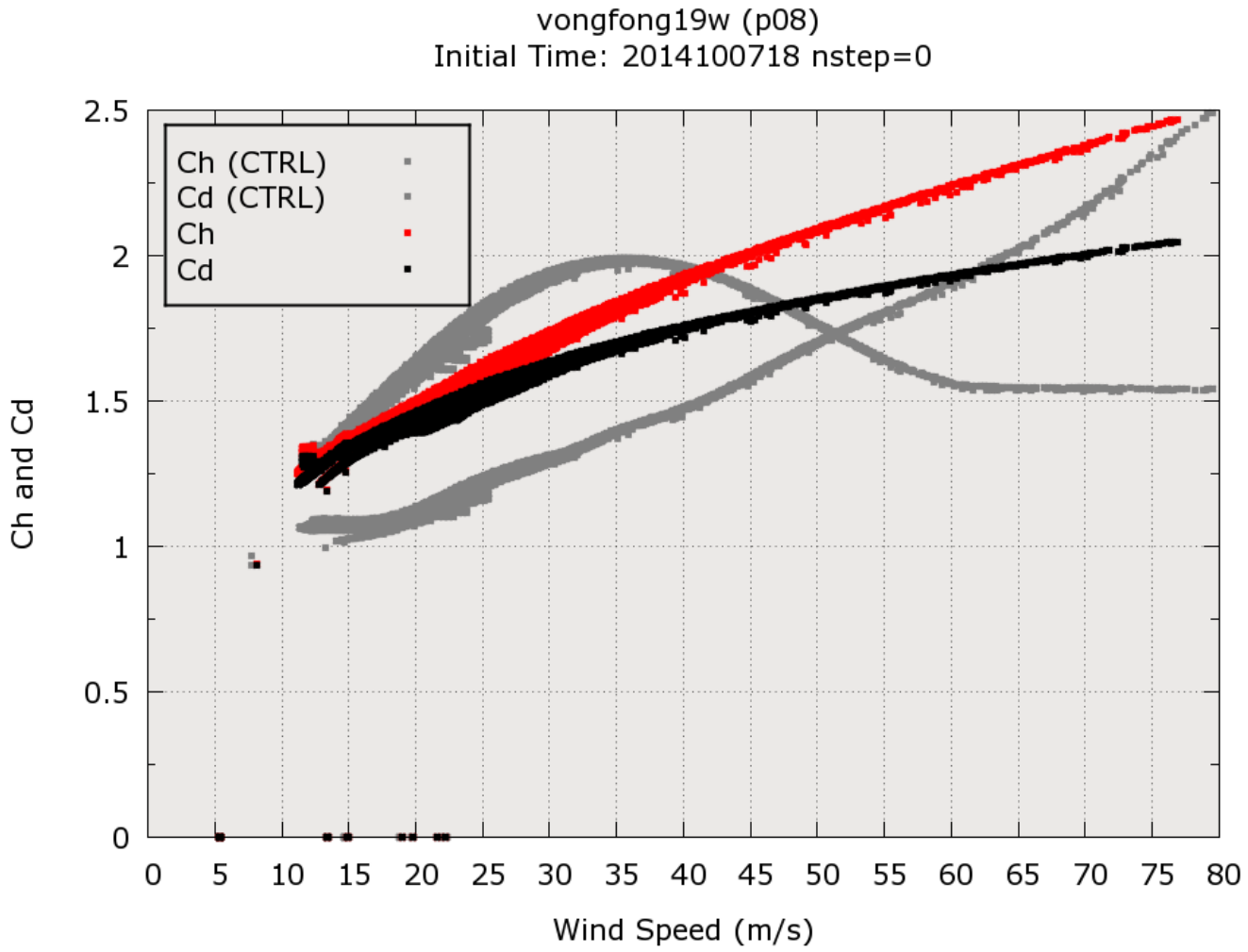






GP08 (2015 ensemble member)

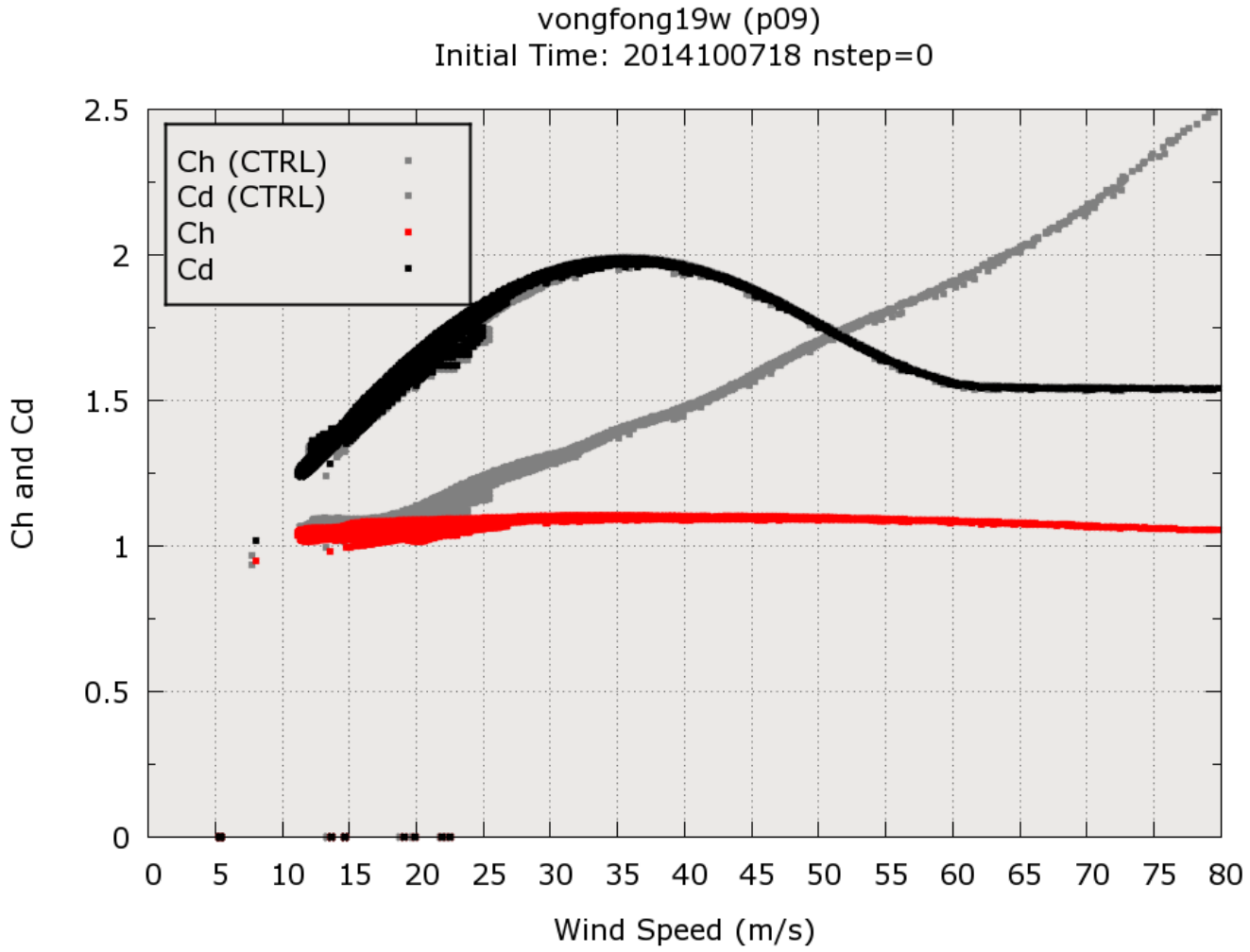
GFDL 2011 operational formulation of C_D & C_H
(surface drag and enthalpy exchange coefficients)



2015 GFDL Control shown in gray

GP09 (2015 ensemble member)

HWRF 2014 operational formulation of C_H
(surface enthalpy exchange coefficient)



2015 GFDL Control shown in gray

Results valid 2015JUL29

2015 Bias-corrected Ensemble Mean Performance

	Atlantic	Eastern Pacific
Average Intensity Error *(w.r.t. **Control**)*	**3.5%** improvement at 0-2 days **17.7%** improvement at 3-5 days	**7.7%** improvement at 0-2 days **5.2%** improvement at 3-5 days
Average Intensity Spread *(w.r.t. **2014 config.**)*	**15.7%** more 0.5-5 day spread	**19%** more 0.5-5 day spread
Average Intensity Bias *(w.r.t. **Control**)*	**Much reduced bias** at 3-5 days	**Much reduced negative bias** at all lead times
Average Track Error *(w.r.t. **Control**)*	**5.1%** improvement at 0.5-2 days **10.3%** improvement at 3-5 days	**8.7%** improvement at 0.5-2 days **12.4%** improvement at 3-5 days
Average Track spread *(w.r.t. **2014 config.**)*	**8.7%** more 0.5-5 day spread	**7%** more 0.5-5 day spread