

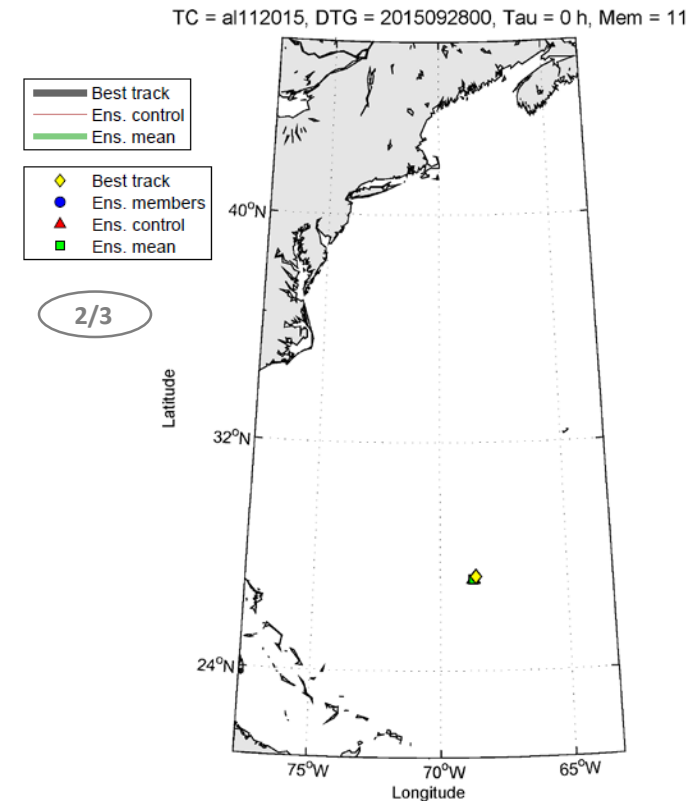
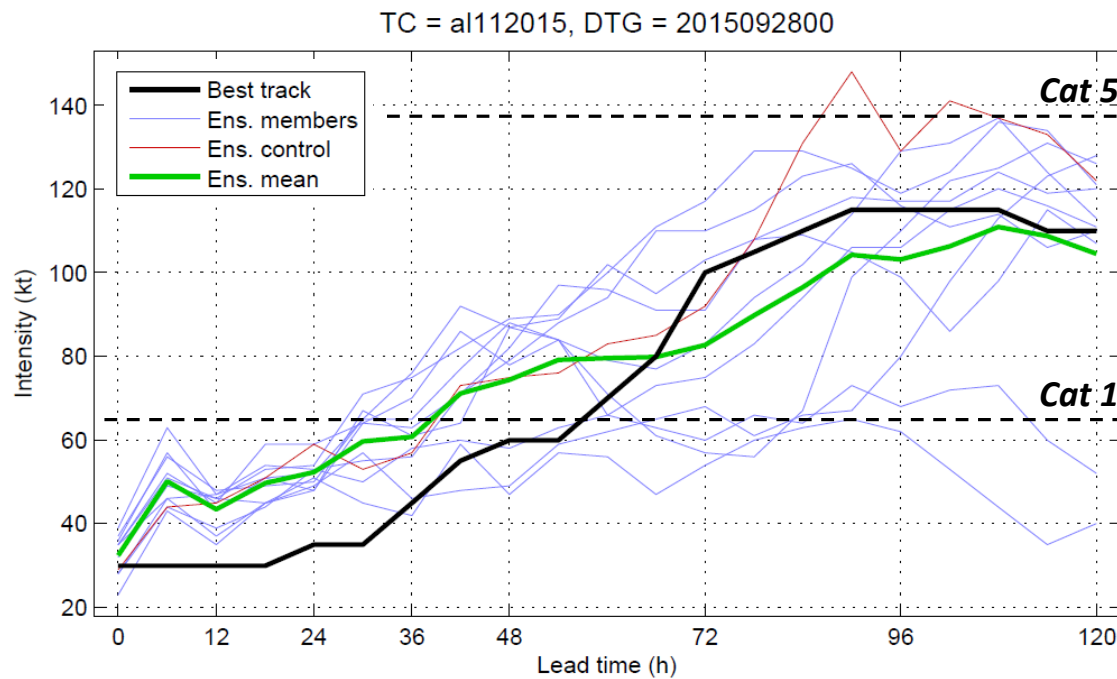
Performance of the 2015 real-time COAMPS-TC ensemble and combined COAMPS-TC/HWRF/GFDL multi-model ensemble



Jon Moskaitis, Alex Reinecke, Jim Doyle and the COAMPS-TC team
Naval Research Laboratory, Monterey, CA

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

Real-time forecast example: Hurricane Joaquin (11L)



Outline

- (1) 2015 COAMPS-TC real-time ensemble demonstration**
- (2) COAMPS-TC ensemble results & conclusions**
- (3) COAMPS-TC/HWRF/GFDL combined ensemble results & conclusions**
- (4) Future directions for COAMPS-TC ensemble**

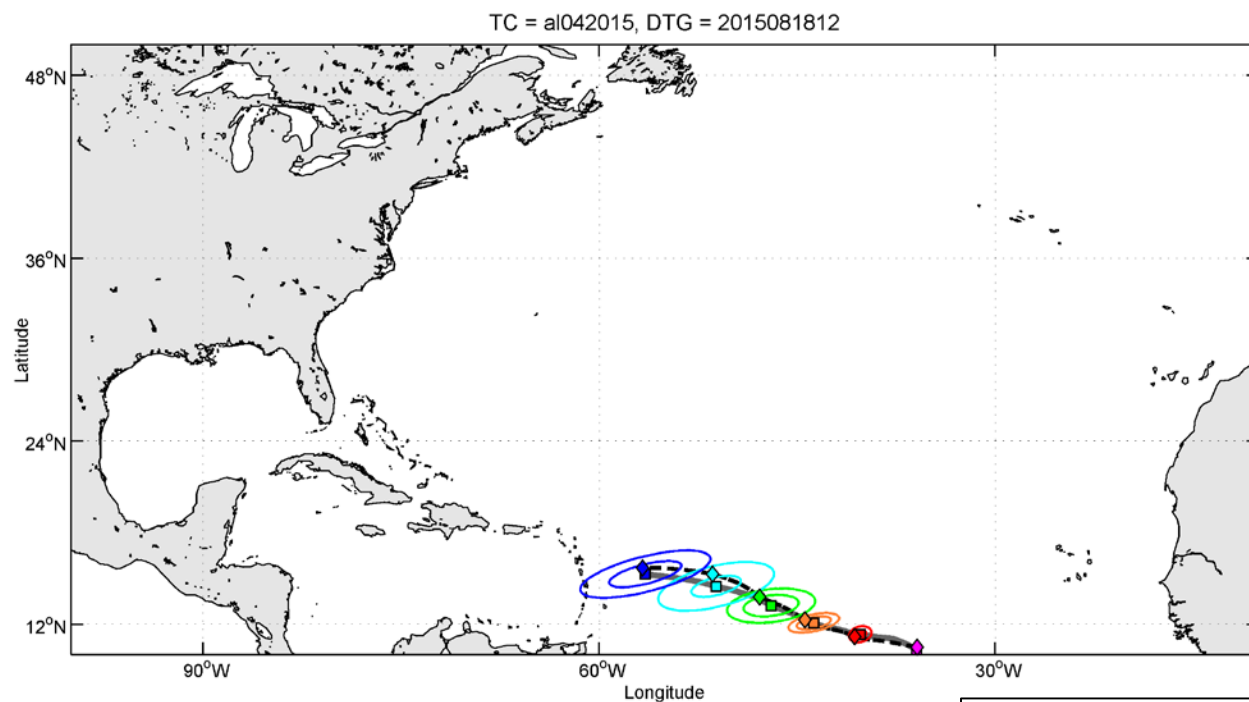
(1) 2015 COAMPS-TC real-time ensemble demonstration

Forecast sample

Number	Name	Forecasts
04L	Danny	25
06L	Fred	31
07L	Grace	16
09L	TD9	8
10L	Ida	38
11L	Joaquin	40
07E	Felicia	5
08E	TD8	10
09E	Guillermo	28
10E	Hilda	31
11E	TD11	4
12E	Ignacio	1
14E	Kevin	20
15E	Linda	20
16E	TD16	2
17E	Marty	10
19E	Olaf	26
20E	Patricia	16
12W	TD12	5
13W	Soudelor	41
14W	TD14	9
15W	Molave	24
16W	Goni	44
17W	Atsani	41

Basic ensemble configuration

- COAMPS-TC model same as ops except 27/9/3 km resolution (instead of 45/15/5 km) and GFS as parent global model (instead of NAVGEM)
- Ensemble = 1 unperturbed control + 10 perturbed members
- Perturbations to synoptic-scale initial state and TC vortex initial state



— Best track
— Ens. mean

◇ Best track
□ Ens. mean

● 0 h (11)
● 24 h (10)
● 48 h (10)
● 72 h (9)
● 96 h (9)
● 120 h (9)

Track plot for every 4th
Atlantic basin
COAMPS-TC ensemble
forecast

158 Atlantic cases

173 EastPac cases

164 WestPac cases

495 Total forecasts

(1) 2015 COAMPS-TC real-time ensemble demonstration

Ensemble configuration details

- Control forecast:
 - Initialized from the GFS analysis
 - Vortex initialized with a Rankine vortex based on TC vitals
- 10 ensemble members ICs perturbed about the control:
 - Synoptic-scale perturbations drawn from WRFVAR cv3 static covariance
 - Perturb the synoptic-scale initial state and lateral boundary conditions
 - Vortex ICs based on perturbed TC vitals
- No data assimilation, no physics perturbations

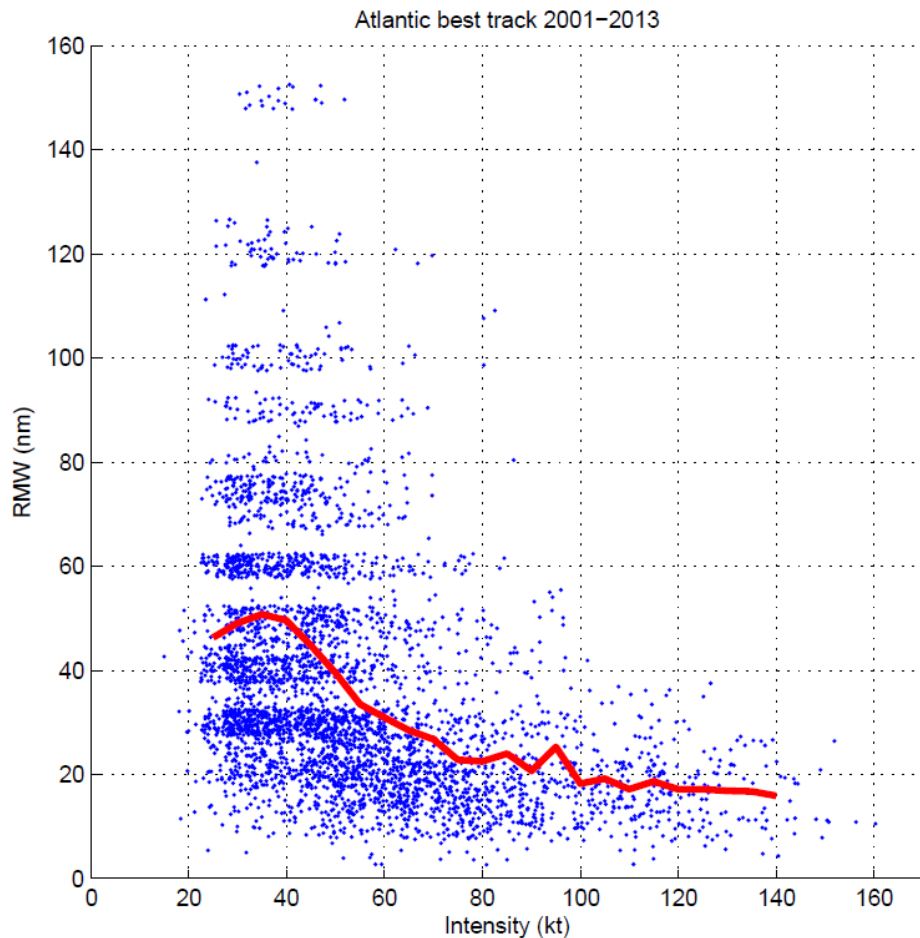
Changes in ensemble configuration w.r.t 2014 real-time demonstration

- 2015 version of COAMPS-TC model (new Cd is primary difference)
- Inner nest size matches ops COAMPS-TC (smaller nests in 2014)

(1) 2015 COAMPS-TC real-time ensemble demonstration

Vortex-scale perturbation details

- Vortex position, max wind, and RMW.
- Perturbation variance from:
 - Torn and Snyder 2012
 - Landsea and Franklin 2013
- Max wind and RMW covariance derived from 2001-2013 best track data.
- Variances and covariances depend on TC-vital max wind speed.



(2)

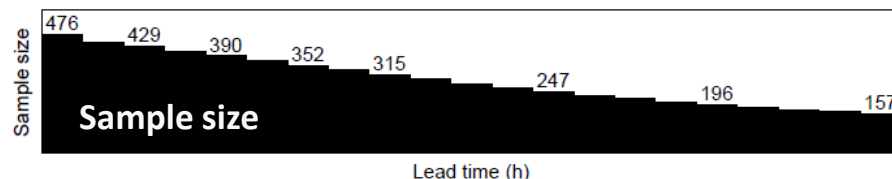
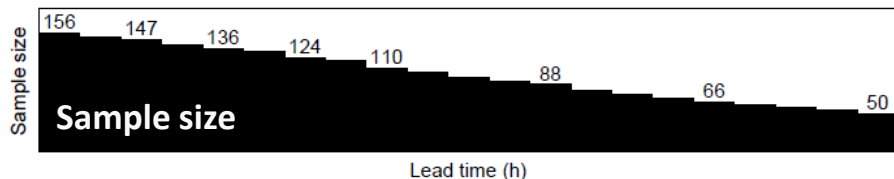
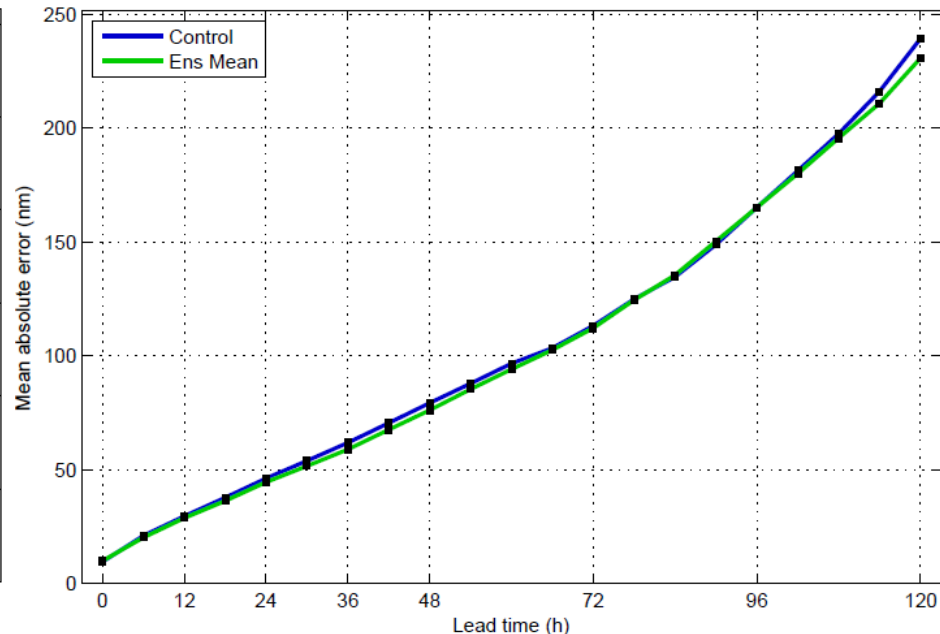
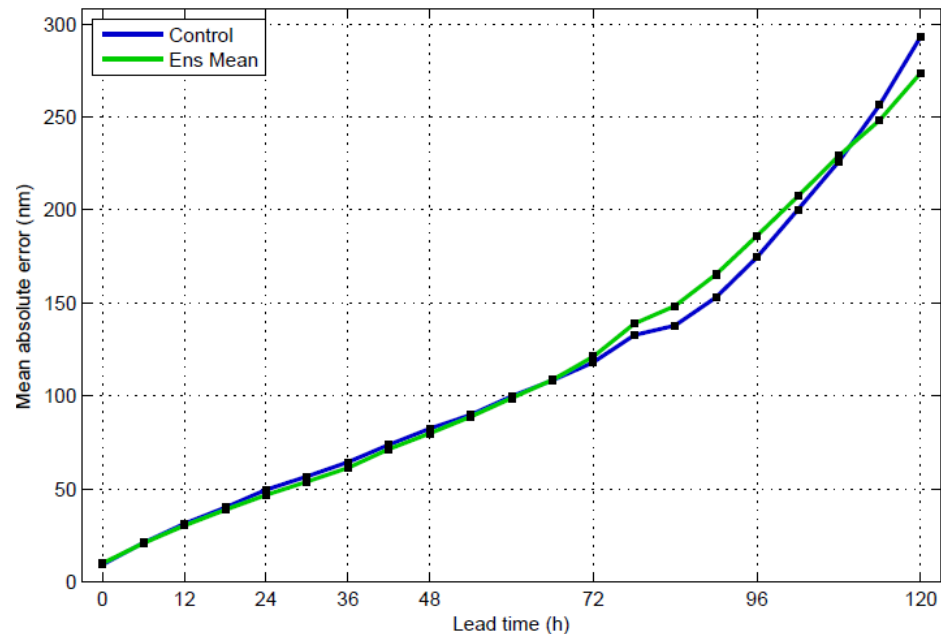
COAMPS-TC ensemble results

Deterministic verification: **Control** vs. **Ensemble mean***

Atlantic

Track MAE

All basins



Unperturbed control member and ensemble mean track forecast accuracy is similar

* Ensemble mean defined to exist if 9 of 11 members present

(2)

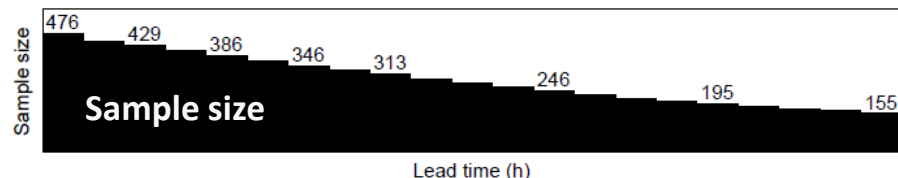
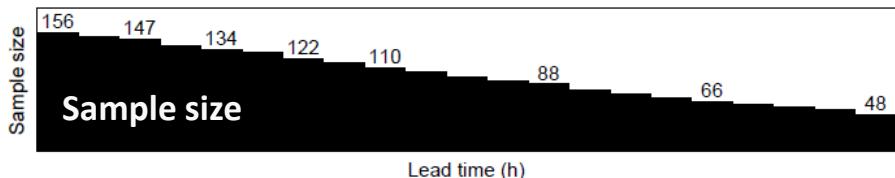
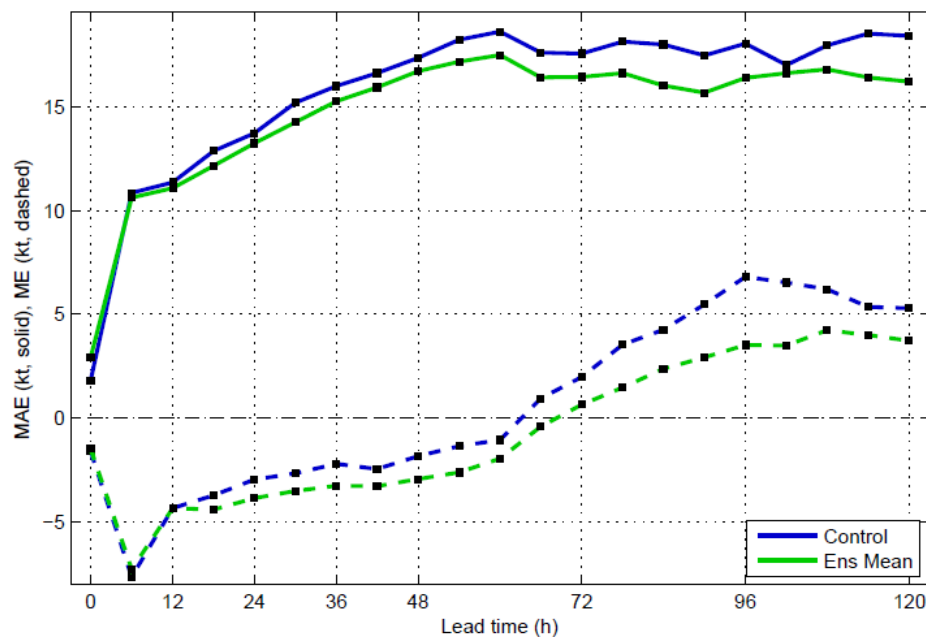
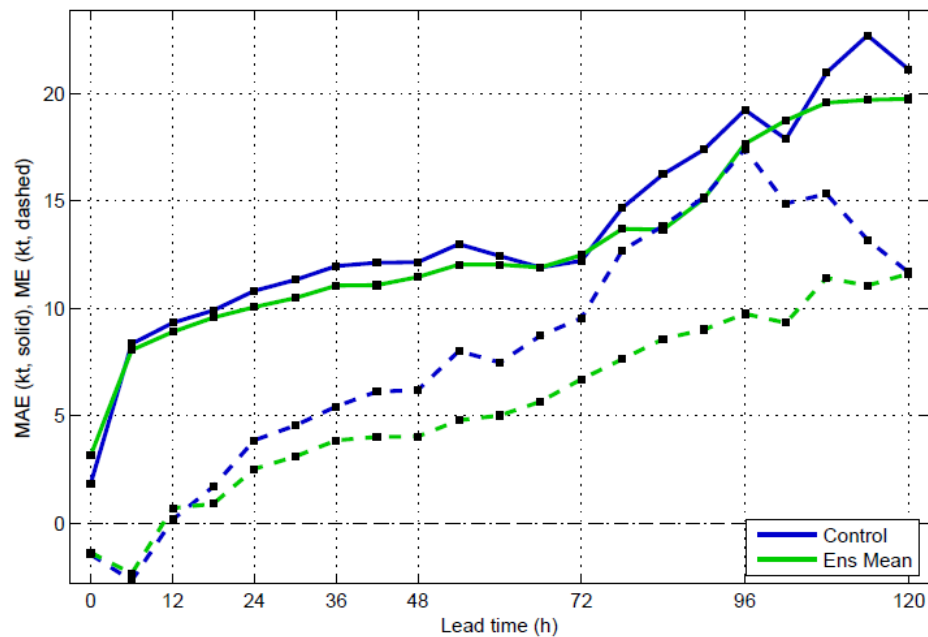
COAMPS-TC ensemble results

Deterministic verification: **Control** vs. **Ensemble mean***

Atlantic

Intensity MAE (solid) and ME (dashed)

All basins



Ensemble mean has superior intensity forecast accuracy in both the Atlantic and All basins samples. Mean error is different, reflecting lower average intensity in ensemble mean.

* Ensemble mean defined to exist if 9 of 11 members present

(2)

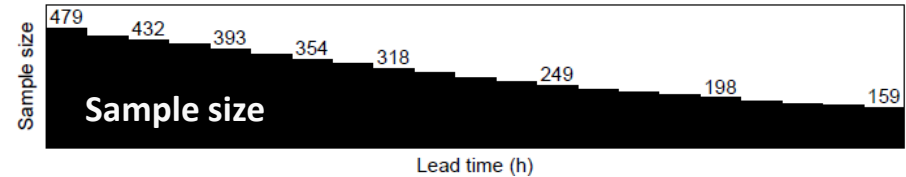
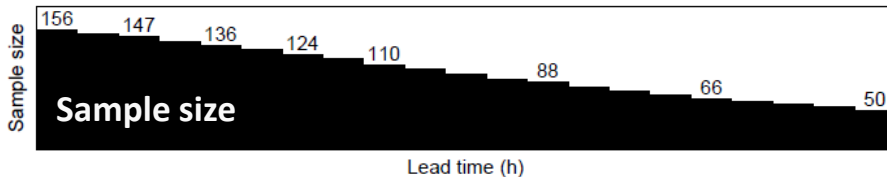
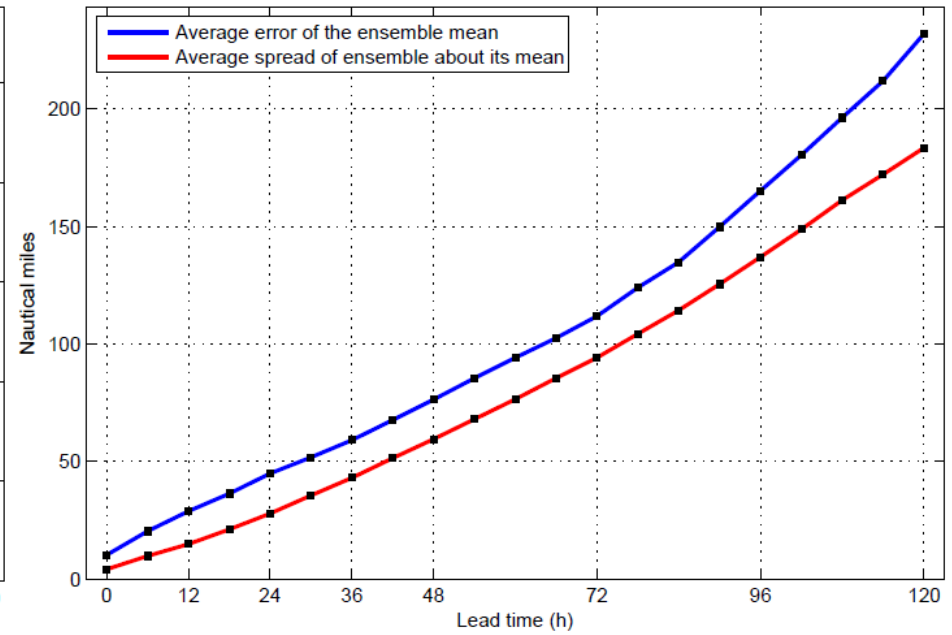
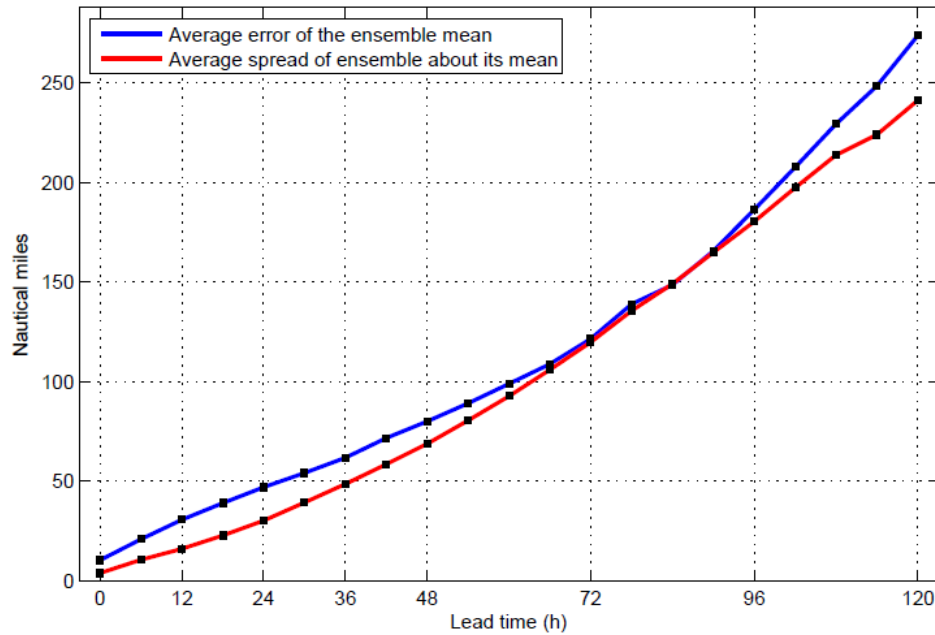
COAMPS-TC ensemble results

Probabilistic verification: Ensemble spread vs. Ensemble mean error

Atlantic

Track

All basins



For track, average ensemble spread and average ensemble mean error are comparable in the Atlantic and Western Pacific, but spread is lacking in the Eastern Pacific. Larger initial perturbations to vortex position could help increase spread at early lead times.

(2)

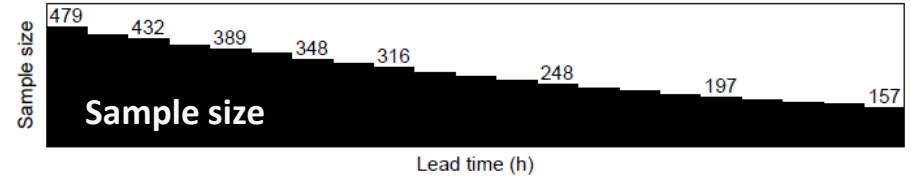
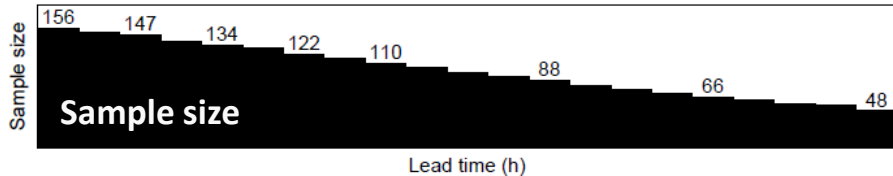
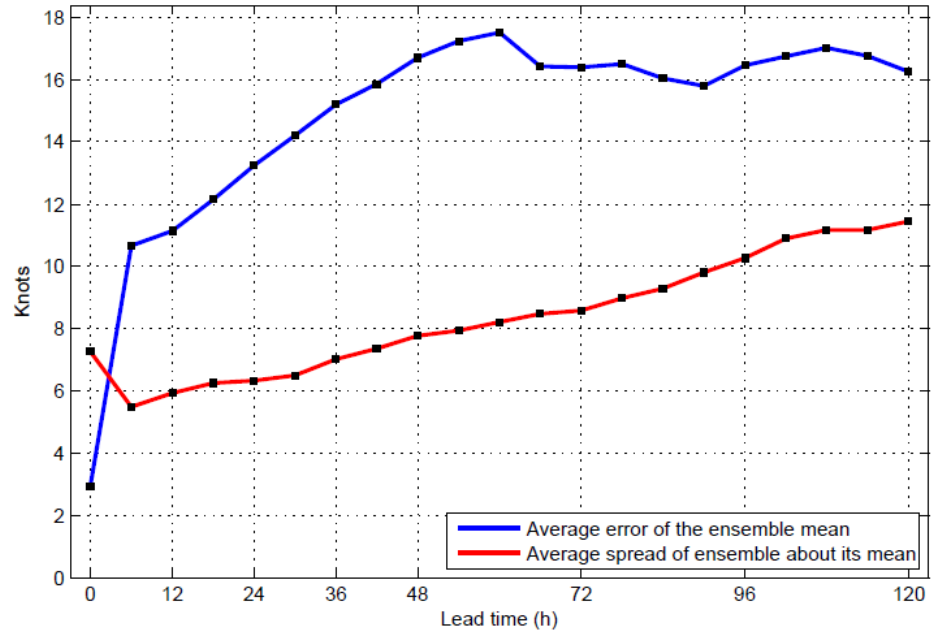
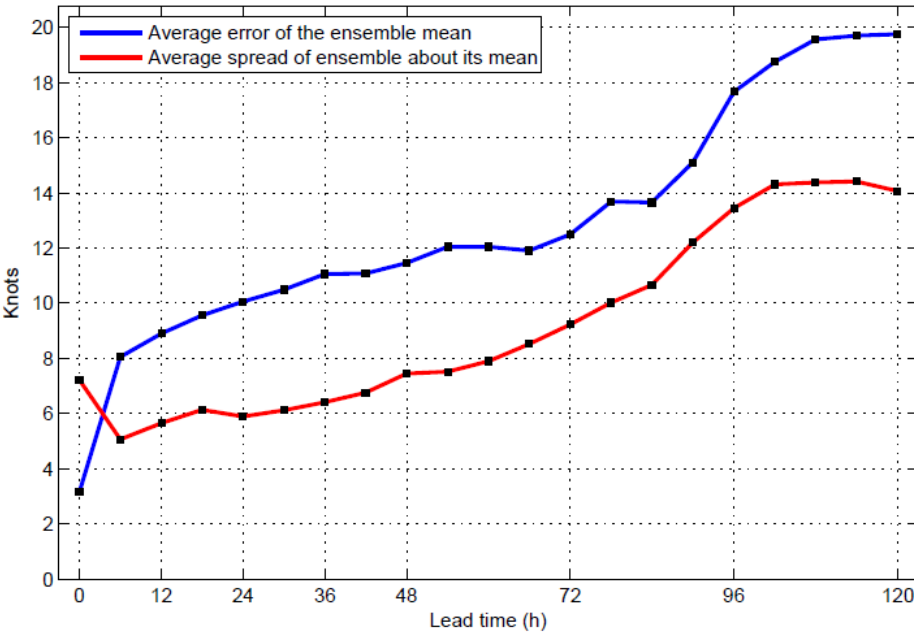
COAMPS-TC ensemble results

Probabilistic verification: Ensemble spread vs. Ensemble mean error

Atlantic

Intensity

All basins

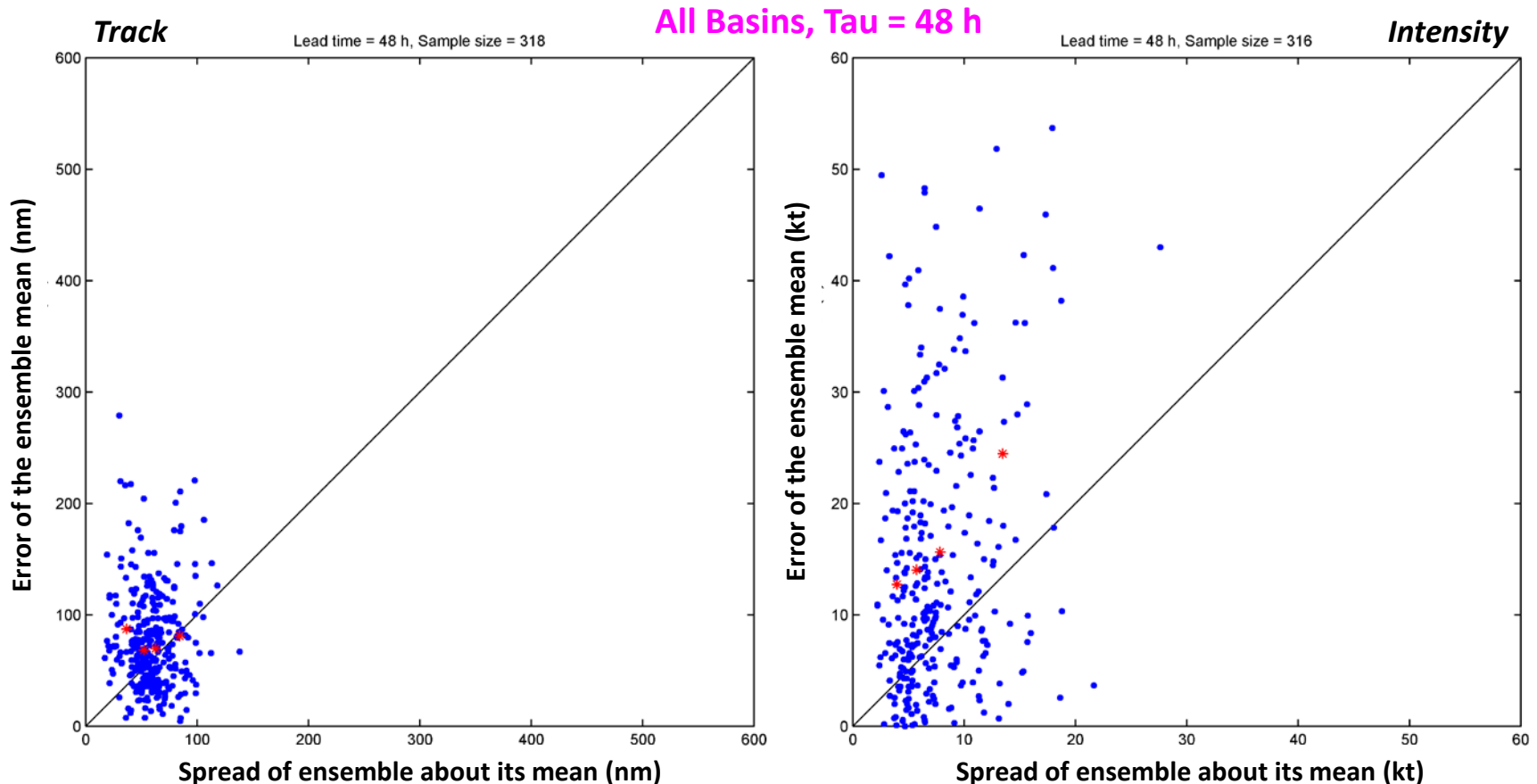


Ensemble intensity forecasts are underdispersive, especially in the Eastern Pacific and Western Pacific basins. Spread grows with lead time, but not quickly enough.

(2)

COAMPS-TC ensemble results

Probabilistic verification: Ensemble spread vs. Ensemble mean error

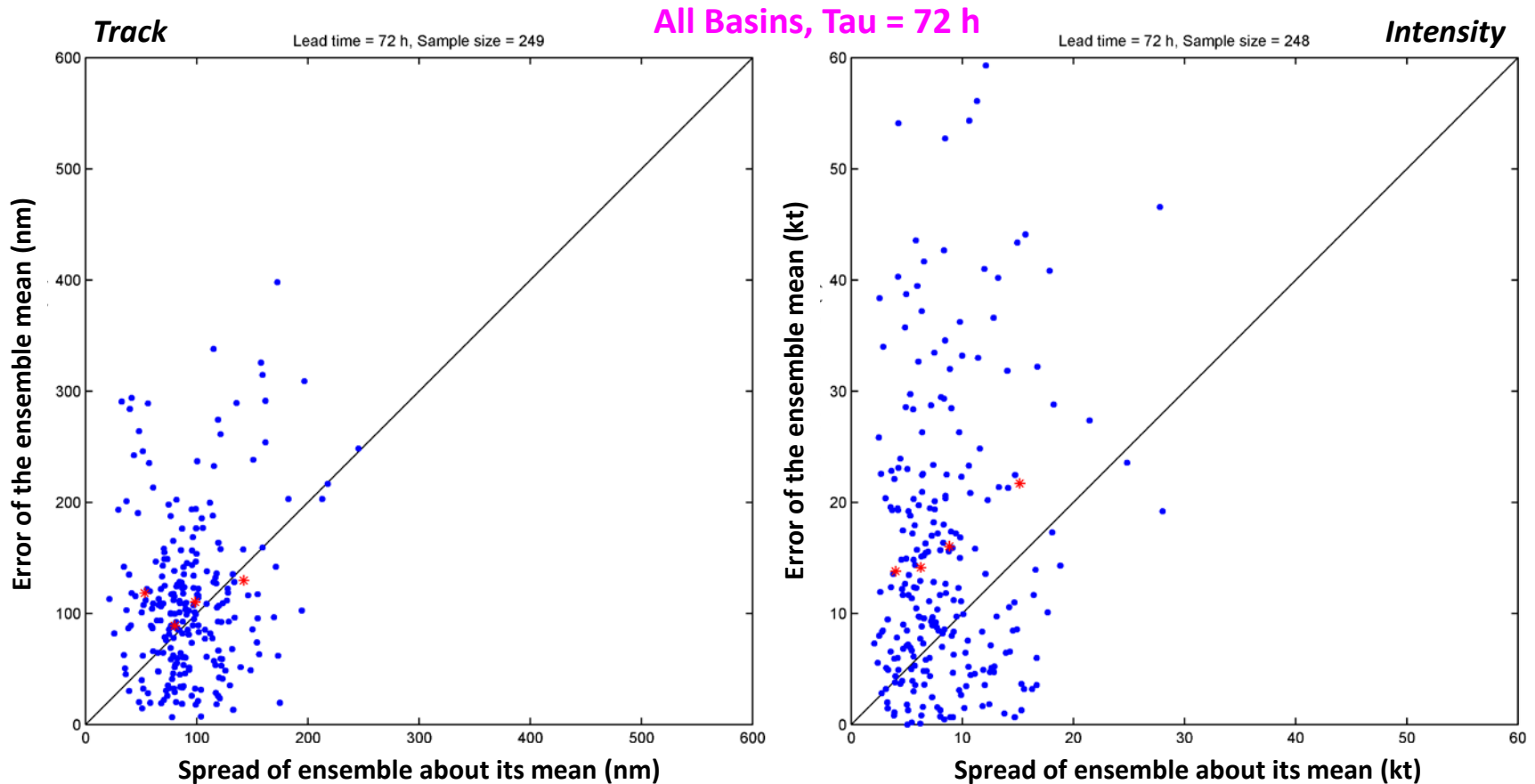


- The **blue dots** represent individual forecasts and **red stars** are bin-averages (4 equally populated bins)
- Would like to see red stars line up along diagonal and few blue dots in upper left (large error / low spread)

(2)

COAMPS-TC ensemble results

Probabilistic verification: Ensemble spread vs. Ensemble mean error

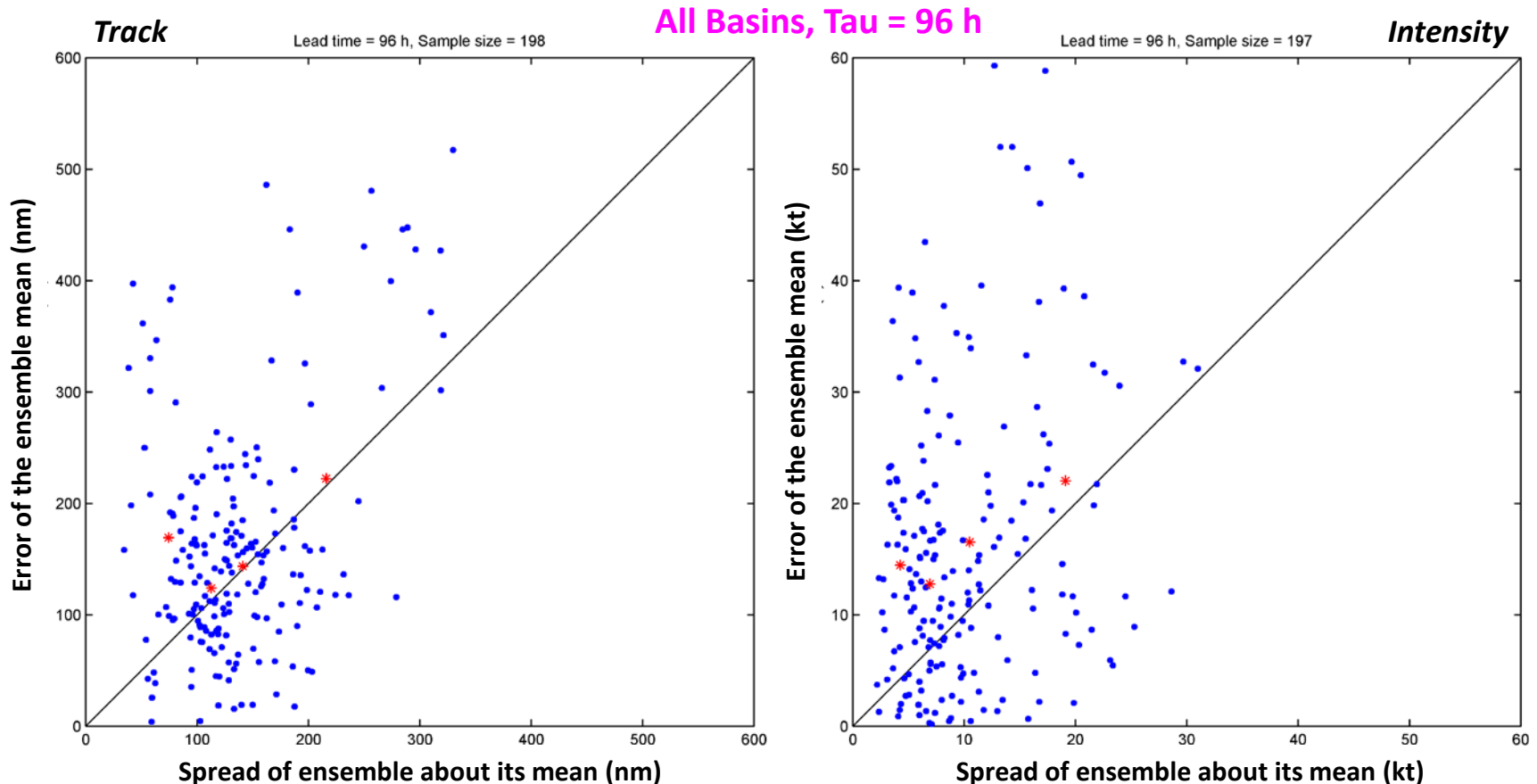


- The **blue dots** represent individual forecasts and **red stars** are bin-averages (4 equally populated bins)
- Would like to see red stars line up along diagonal and few blue dots in upper left (large error / low spread)

(2)

COAMPS-TC ensemble results

Probabilistic verification: Ensemble spread vs. Ensemble mean error

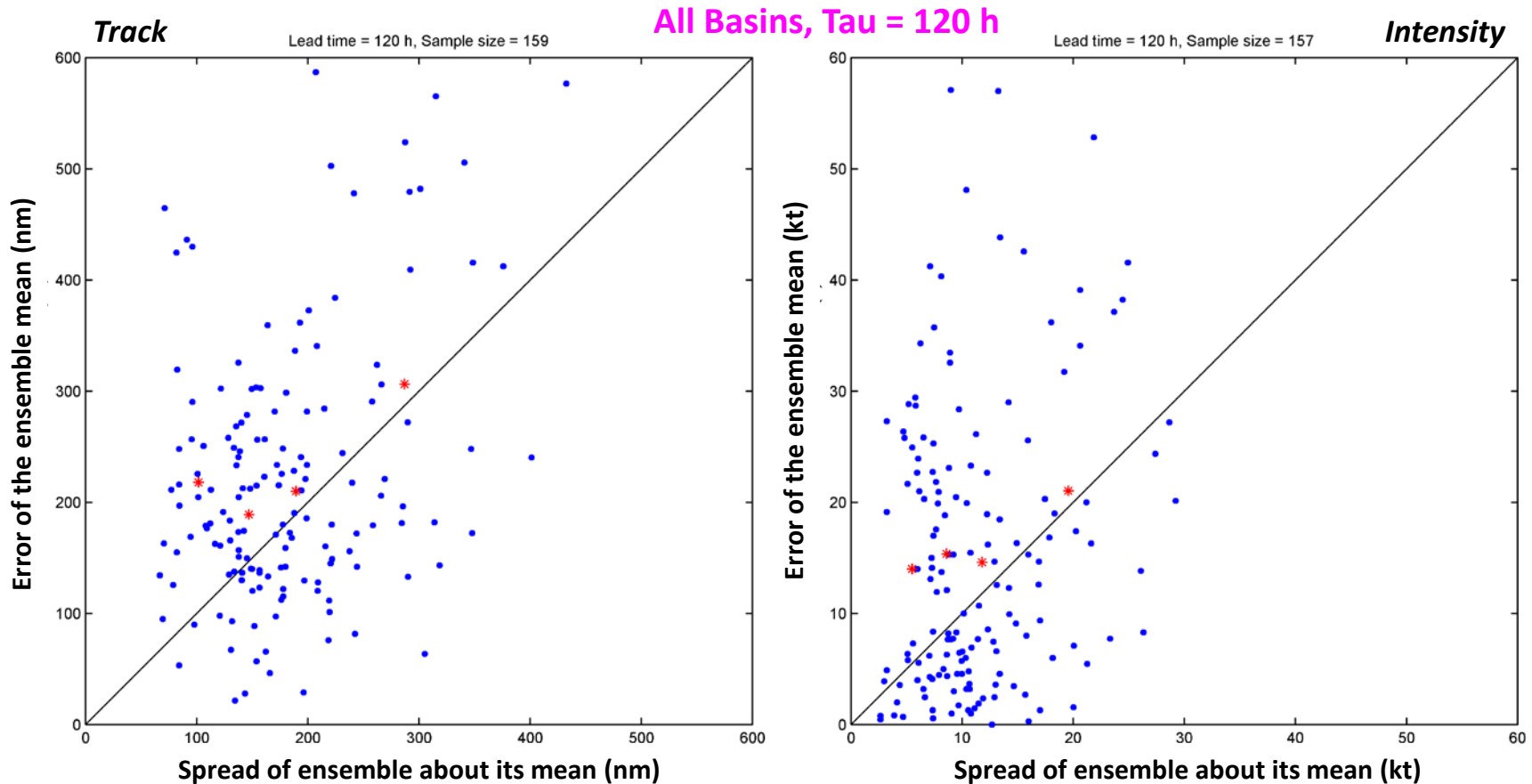


- The **blue dots** represent individual forecasts and **red stars** are bin-averages (4 equally populated bins)
- Would like to see red stars line up along diagonal and few blue dots in upper left (large error / low spread)

(2)

COAMPS-TC ensemble results

Probabilistic verification: Ensemble spread vs. Ensemble mean error



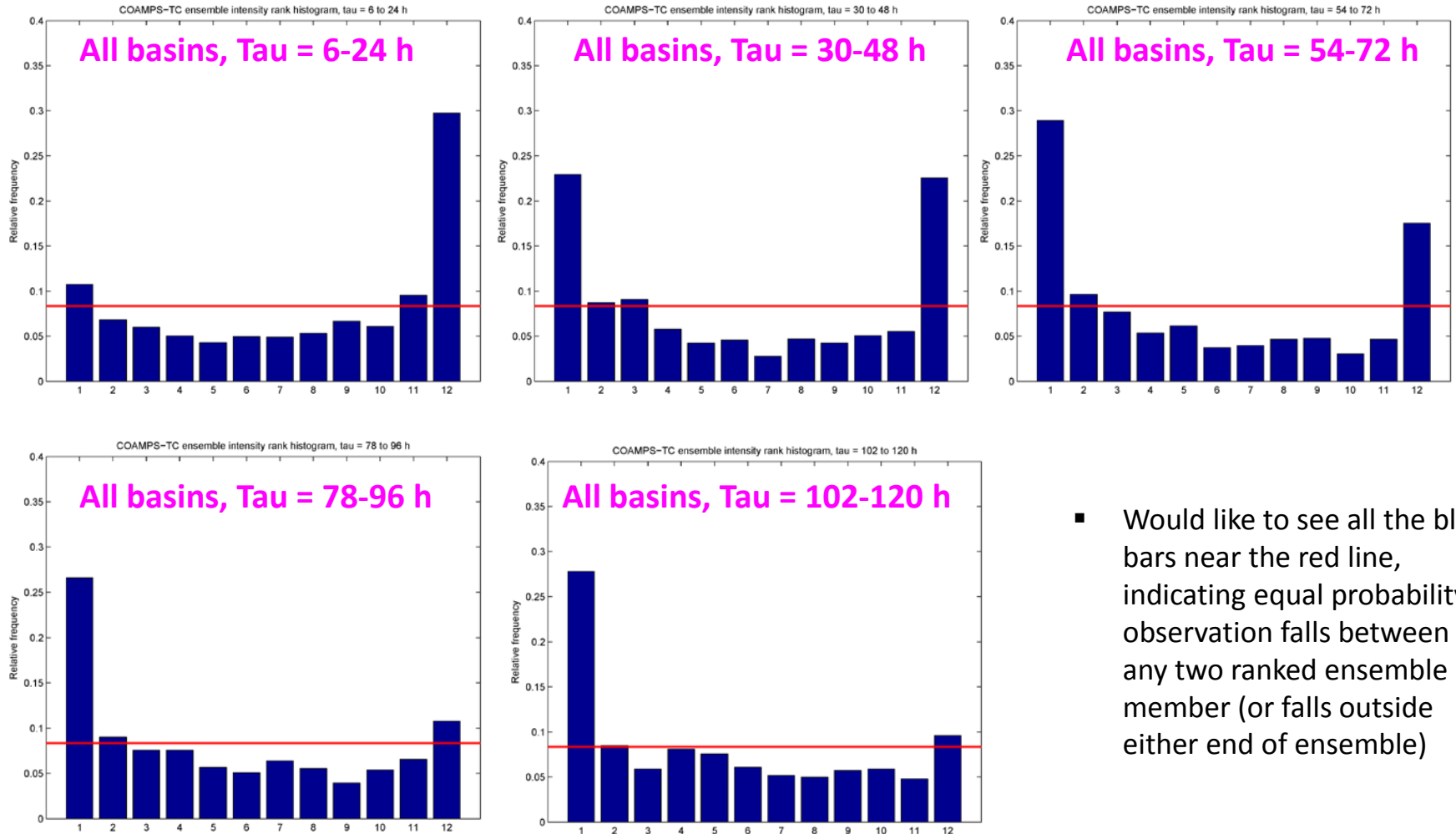
For track, larger spread generally implies larger ensemble mean error. However, too many cases of small spread and large error (Hilda!)

As for track, for intensity larger spread generally implies larger ensemble mean error. Spread is too small though, so red dots above diagonal.

(2)

COAMPS-TC ensemble results

Probabilistic verification: Intensity rank histograms



- Would like to see all the blue bars near the red line, indicating equal probability observation falls between any two ranked ensemble member (or falls outside either end of ensemble)

There is overpopulation of the end bins, indicating observation is outside the ensemble envelope about 2-3x more than would be expected from a perfectly reliable ensemble

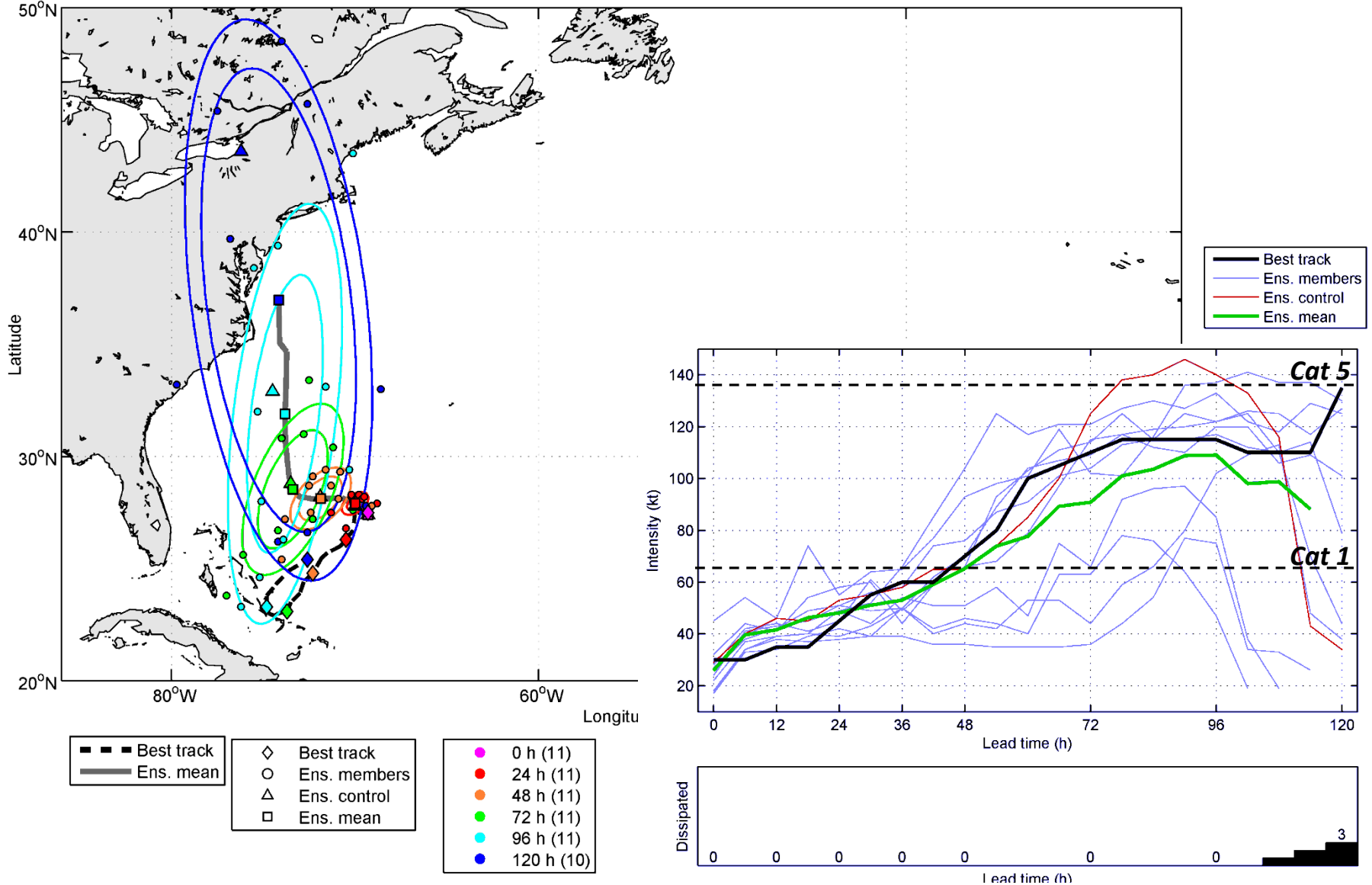
(2)

COAMPS-TC ensemble results

Joaquin (11L) example

TC = al112015, DTG = 2015092812

Track spread: 495 nm (#1 of 159)
Intensity spread: 29.3 kt (#1 of 159)



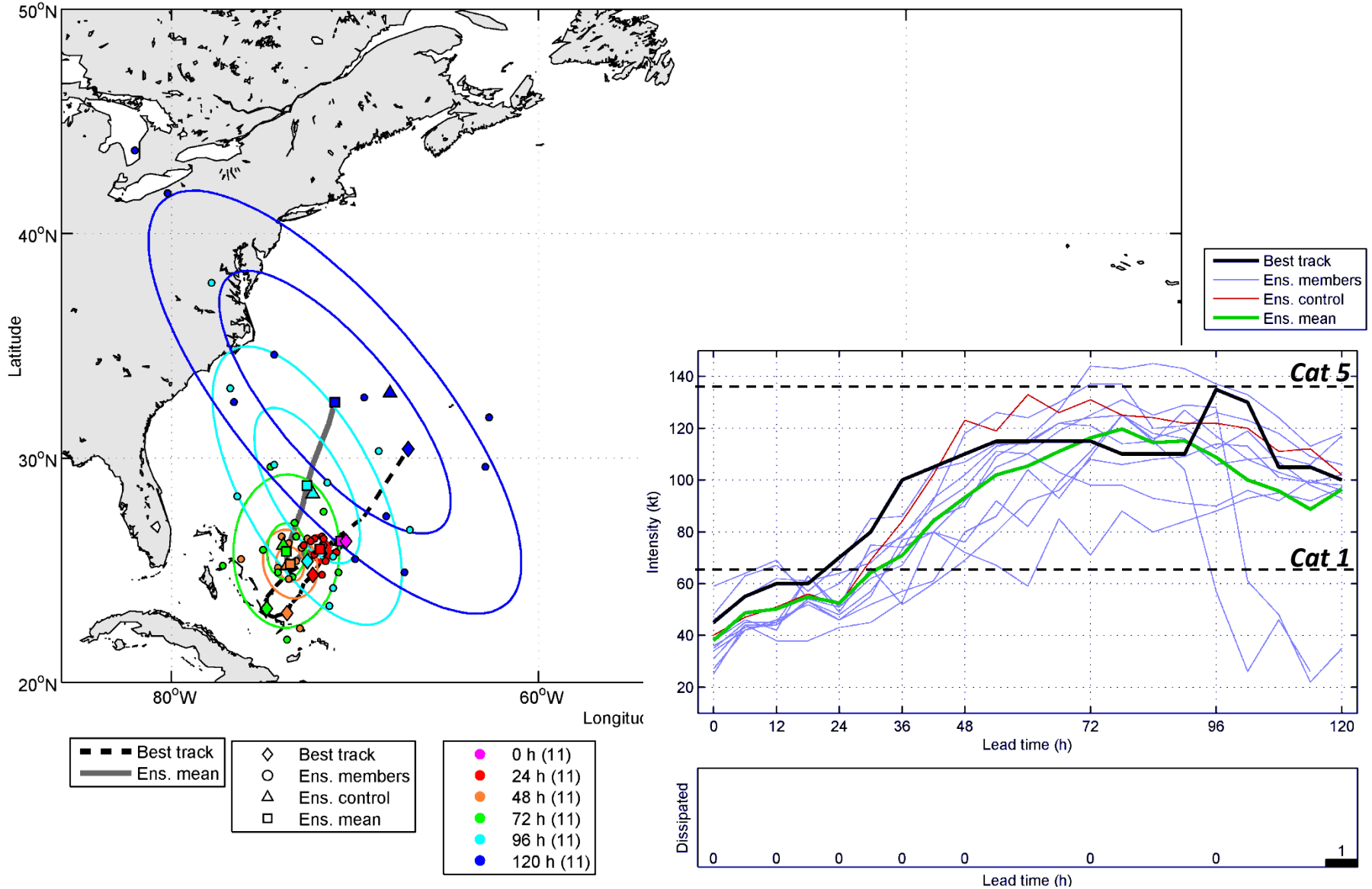
(2)

COAMPS-TC ensemble results

Joaquin (11L) example

TC = al112015, DTG = 2015092912

Track spread: 402 nm (#3 of 159)
Intensity spread: 13.0 kt



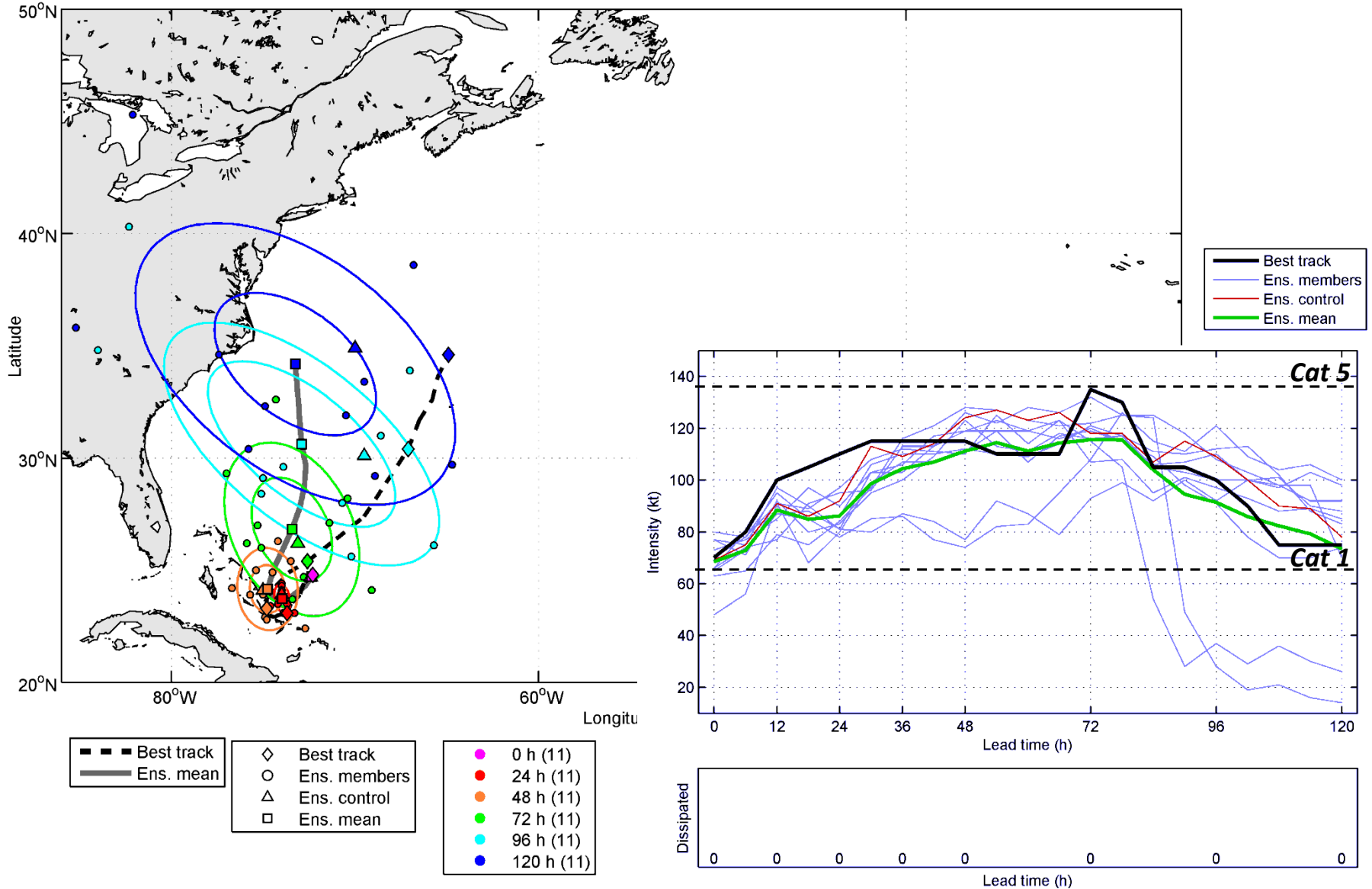
(2)

COAMPS-TC ensemble results

Joaquin (11L) example

TC = al112015, DTG = 2015093012

Track spread: 349 nm (#5 of 159)
Intensity spread: 20.1 kt (#18 of 159)



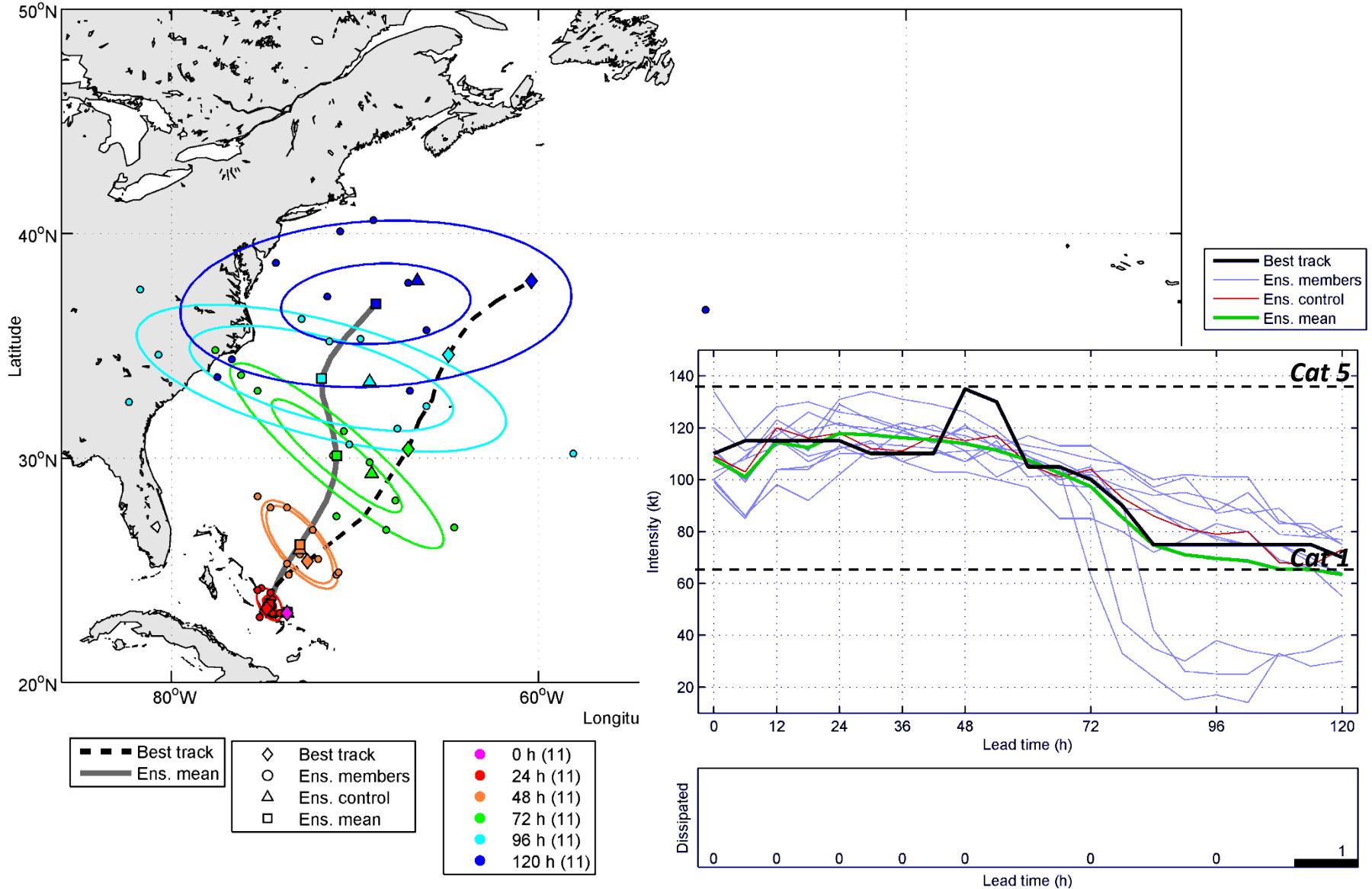
(2)

COAMPS-TC ensemble results

Joaquin (11L) example

TC = al112015, DTG = 2015100112

Track spread: 292 nm (#15 of 159)
Intensity spread: 13.1 kt



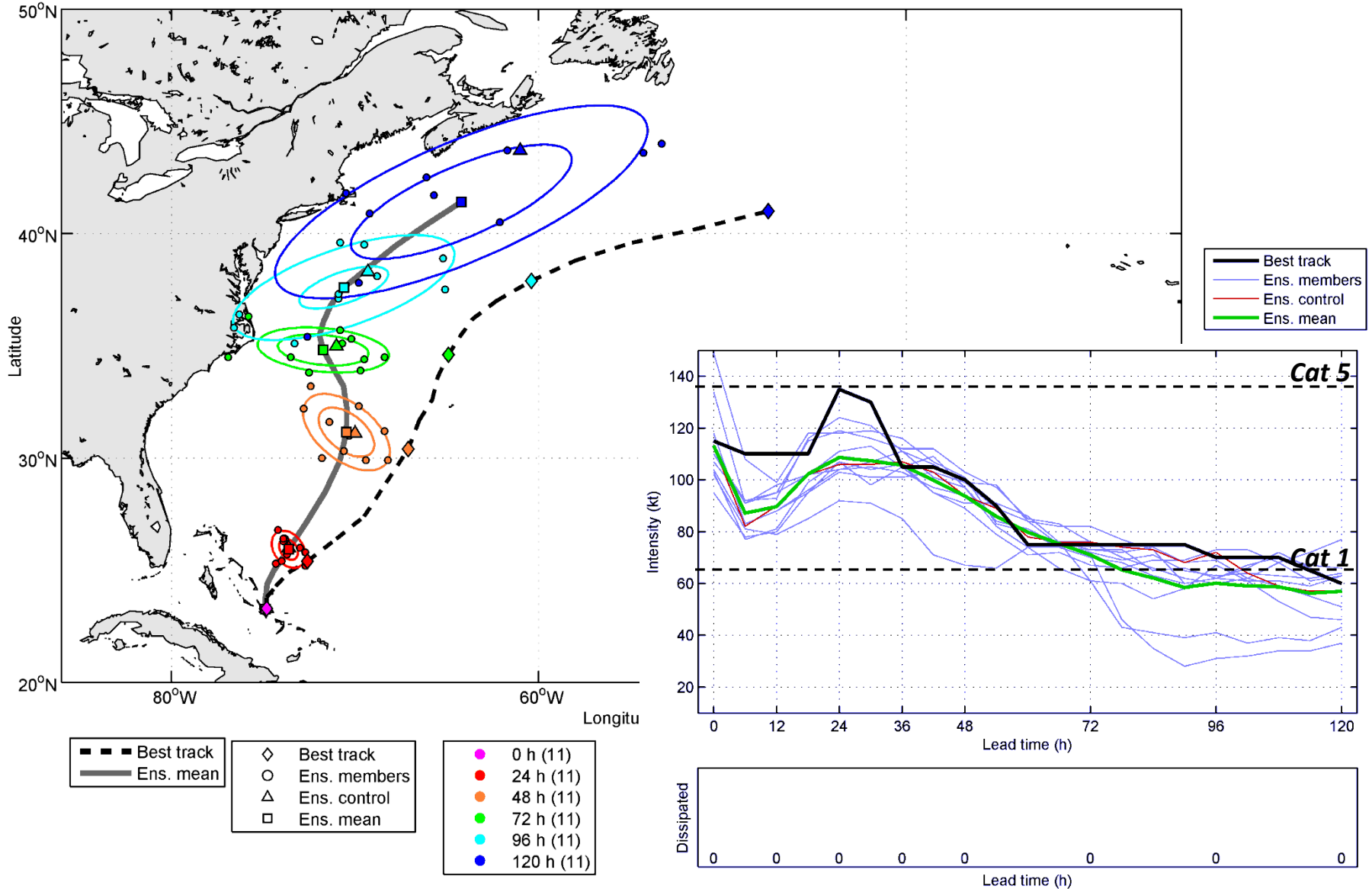
(2)

COAMPS-TC ensemble results

Joaquin (11L) example

TC = al112015, DTG = 2015100212

Track spread: 274 nm (#22 of 159)
Intensity spread: 9.3 kt

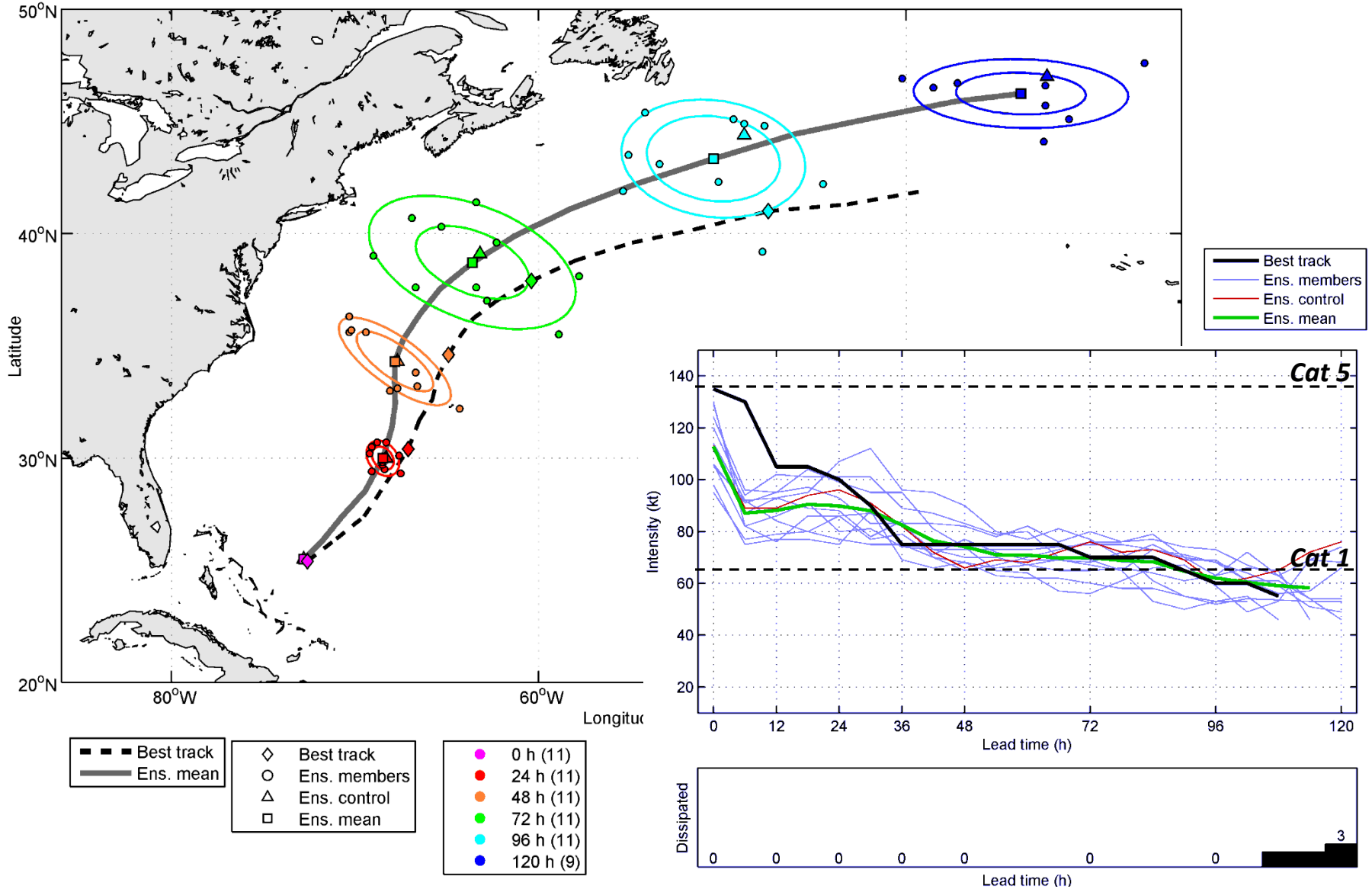


(2)

COAMPS-TC ensemble results

Joaquin (11L) example

TC = al112015, DTG = 2015100312

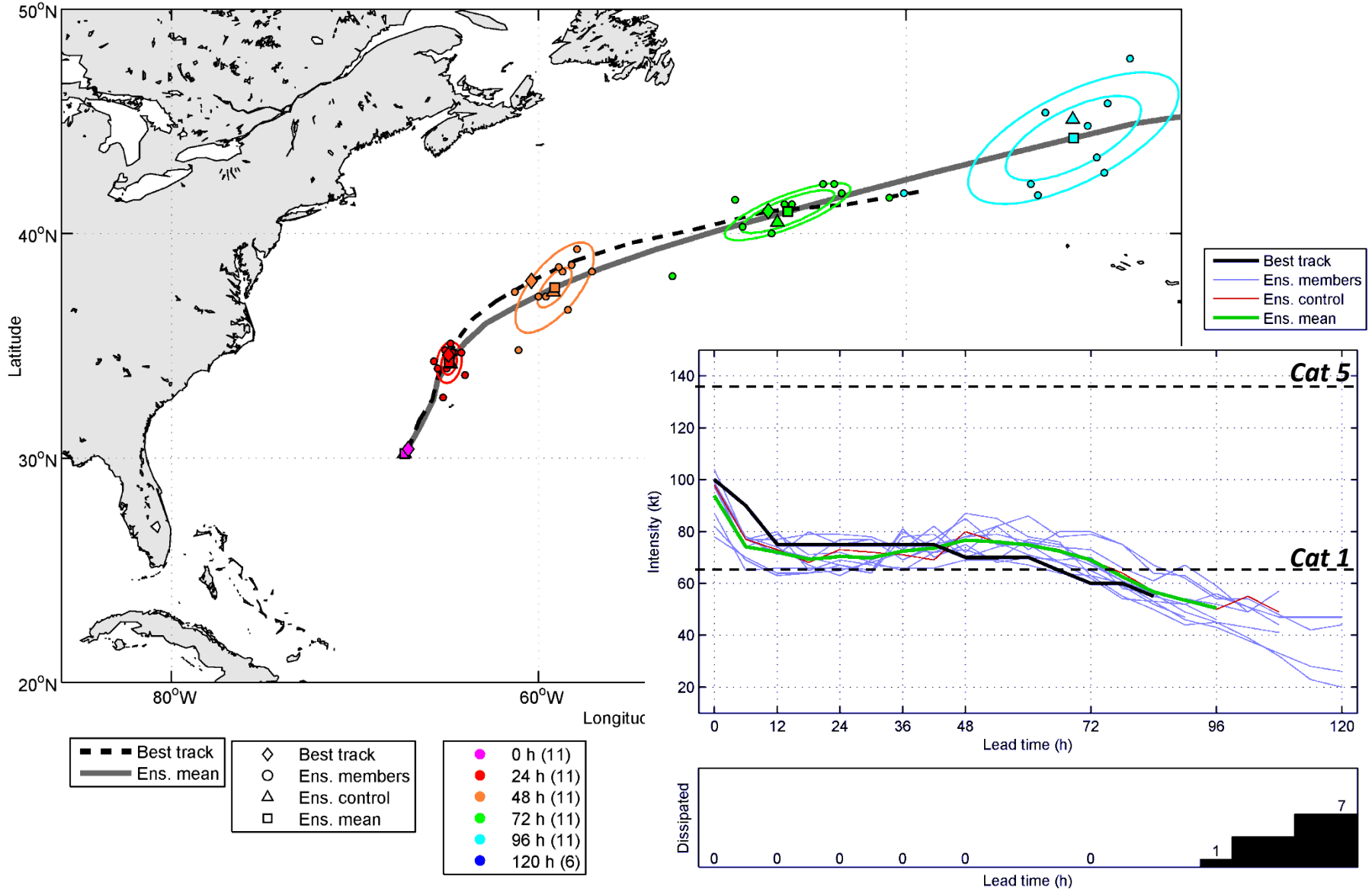


(2)

COAMPS-TC ensemble results

Joaquin (11L) example

TC = al112015, DTG = 2015100412

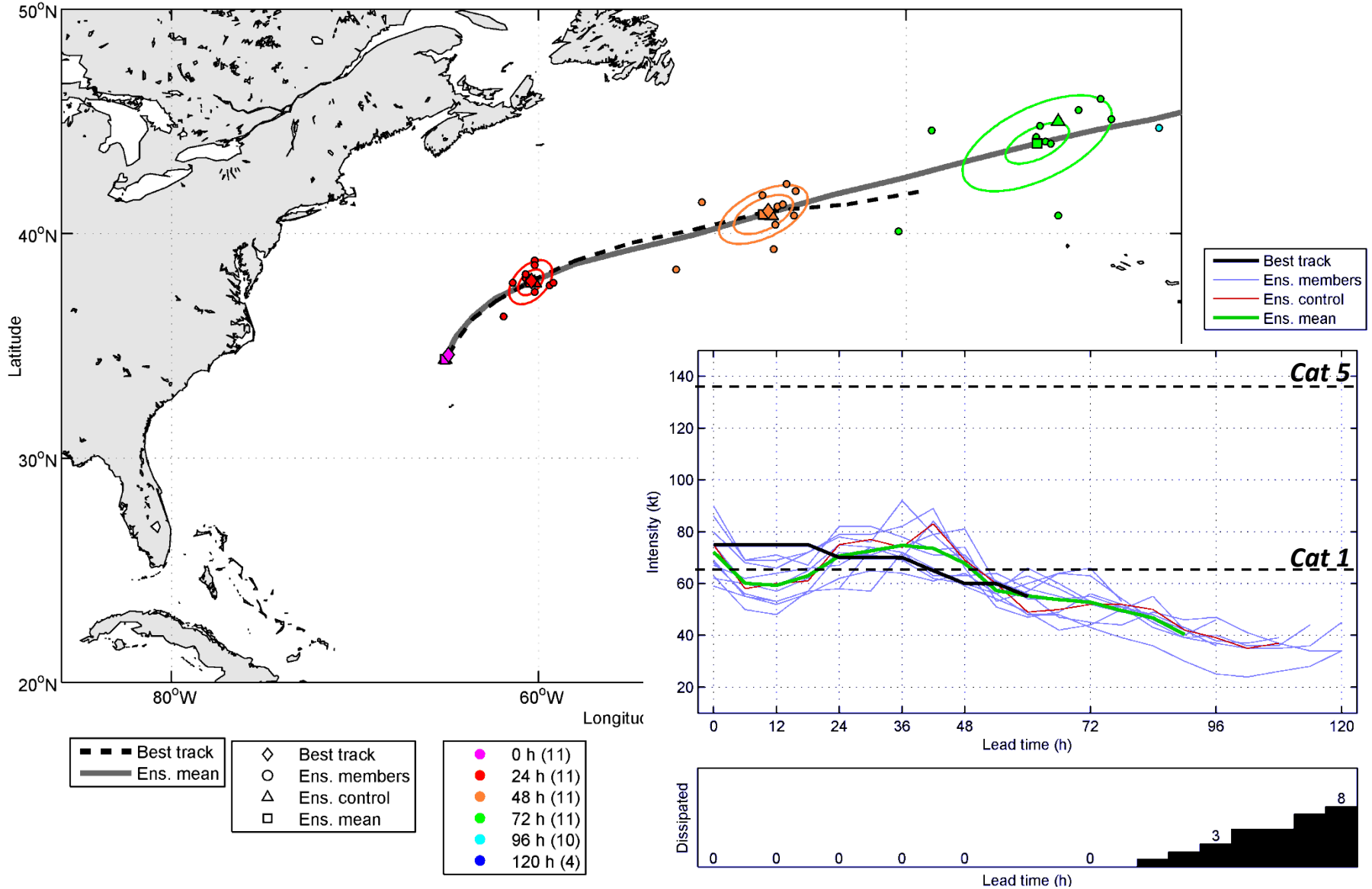


(2)

COAMPS-TC ensemble results

Joaquin (11L) example

TC = al112015, DTG = 2015100512

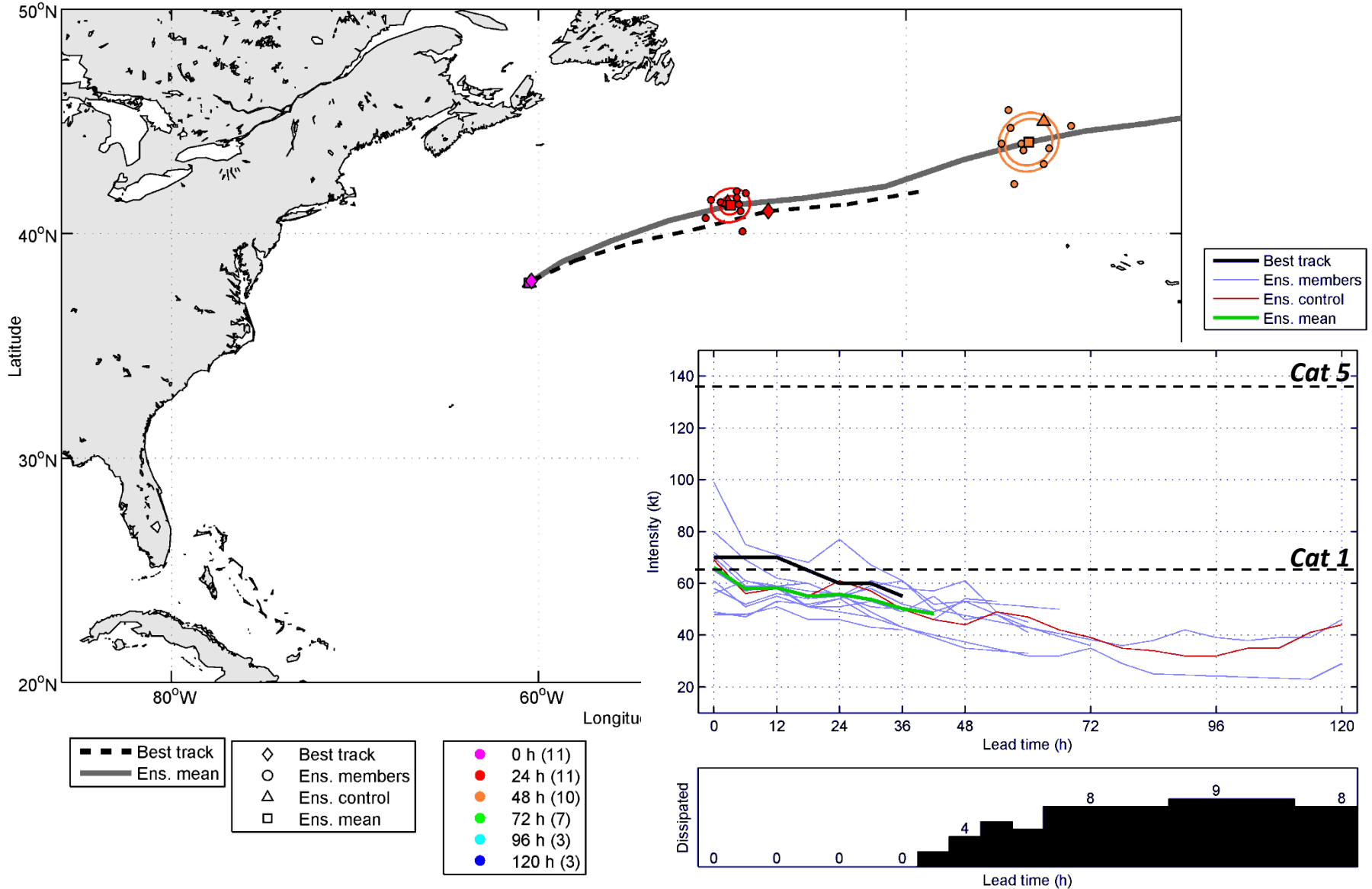


(2)

COAMPS-TC ensemble results

Joaquin (11L) example

TC = al112015, DTG = 2015100612

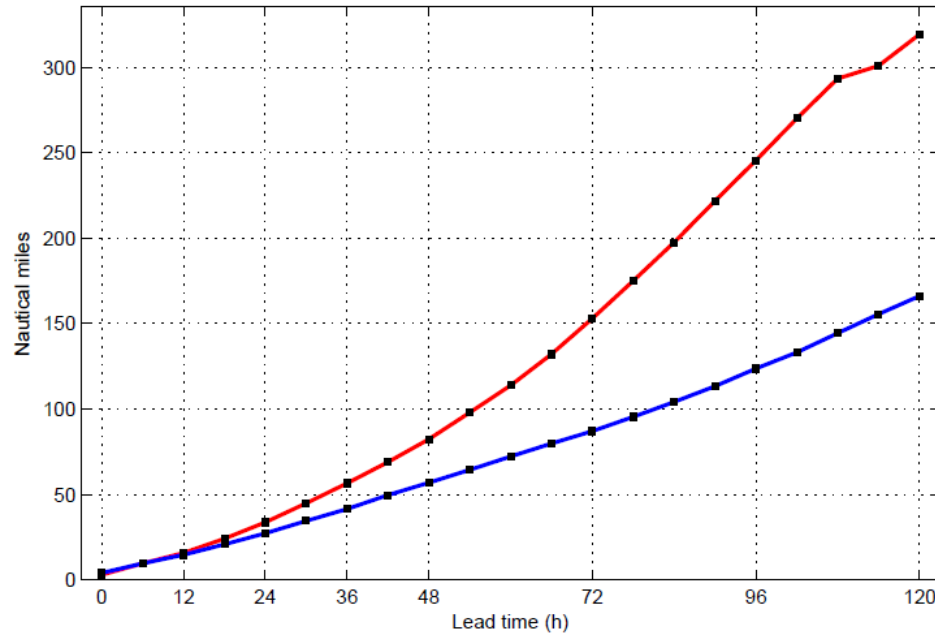


(2)

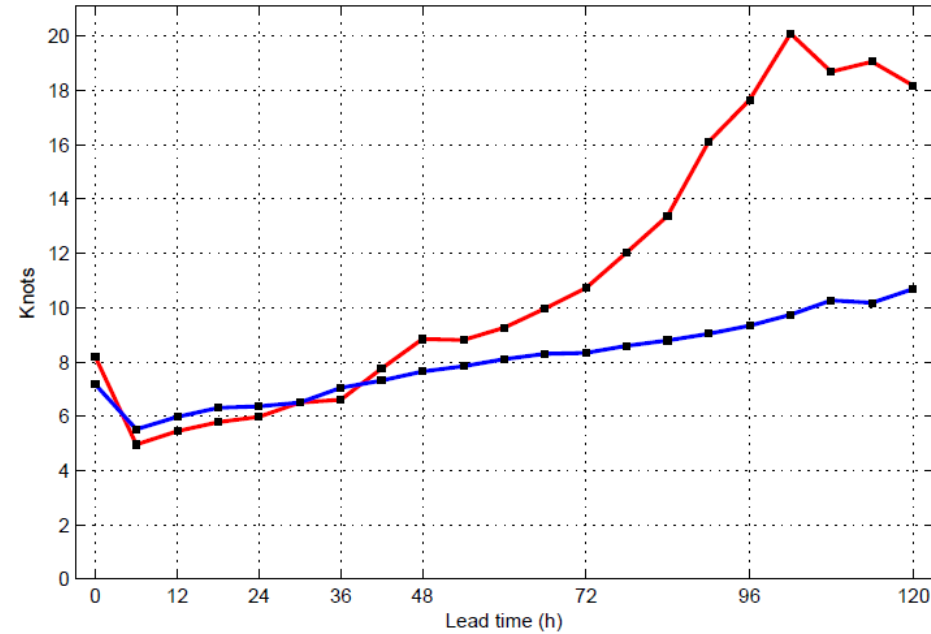
COAMPS-TC ensemble results

Average spread for Joaquin vs. Average spread for all other predicted TCs in 2015

Track spread



Intensity spread



Track spread for Joaquin was unusually large relative to the other TCs in the 2015 COAMPS-TC ensemble sample. At the later lead times, Joaquin spread is ~2x that of the other TCs.

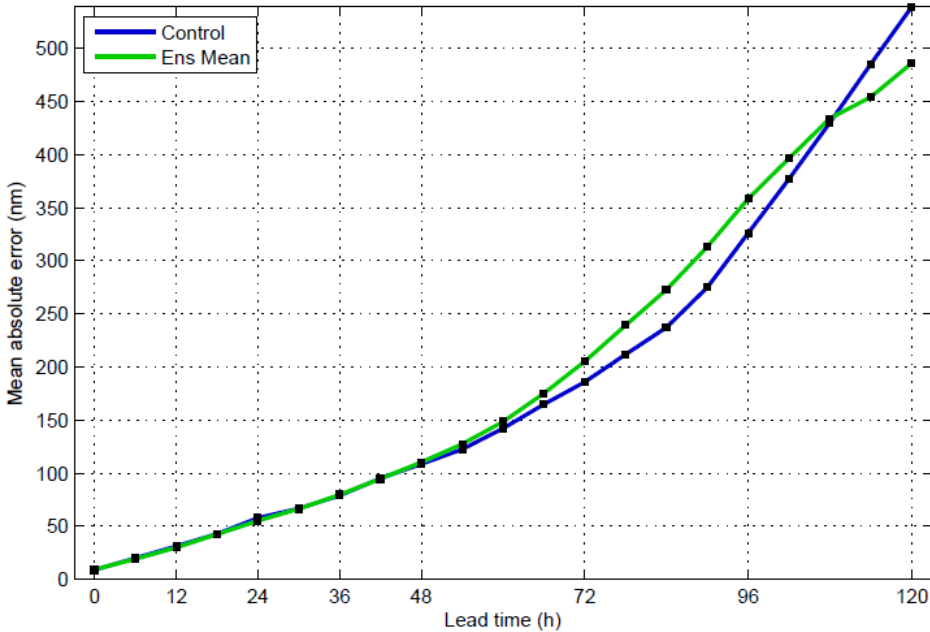
For the later lead times, intensity spread was unusually large relative to the other TCs in the 2015 COAMPS-TC ensemble sample (up to 2x as large)

(2)

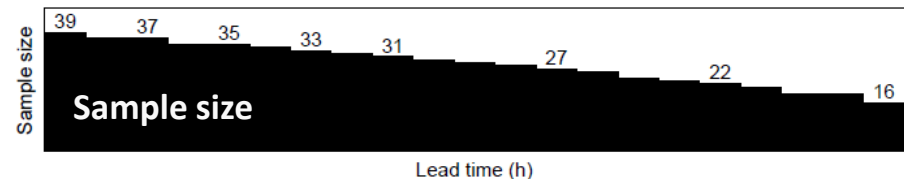
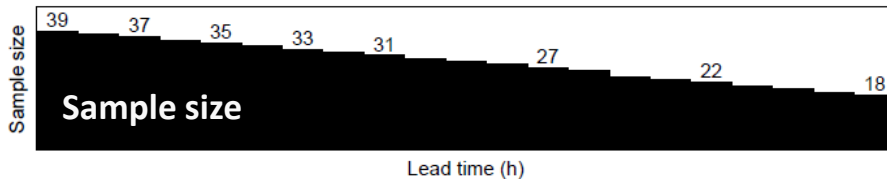
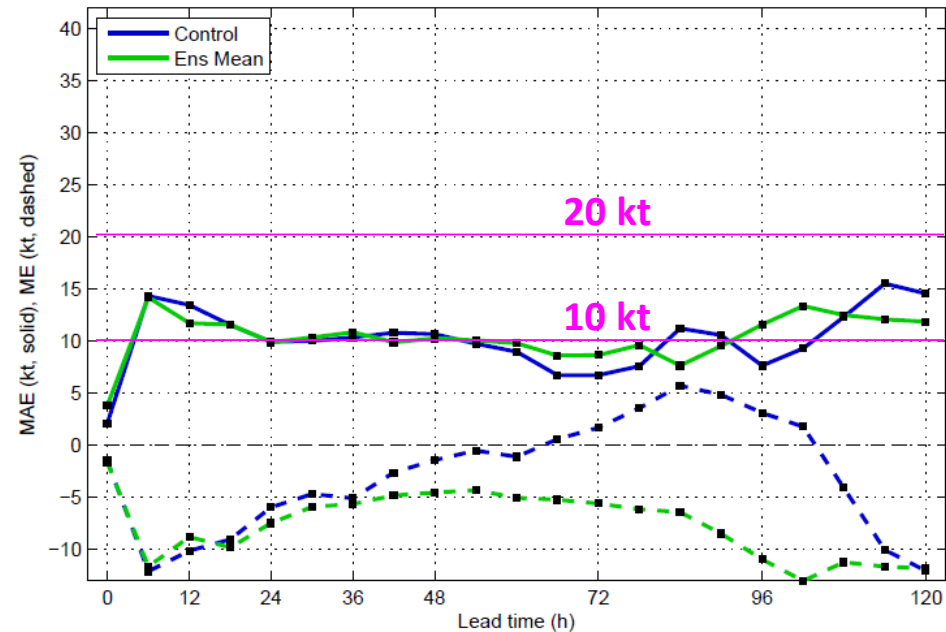
COAMPS-TC ensemble results

Joaquin (11L): Control vs. Ensemble mean

Track MAE



Intensity MAE (solid) and ME (dashed)



Control and ensemble mean track MAE is quite large, relative to that for other TCs (not shown). Ensemble indicated possibility of unusually large errors.

Control and ensemble mean intensity MAE is low, relative to that for other TCs (not shown).

(2) Conclusions for COAMPS-TC ensemble

- 495 forecasts were produced by the real-time demonstration system, for TCs in the Atlantic, Eastern Pacific, and Western Pacific basins
- For track, the accuracy of the ensemble mean is similar to that of the control member, but for intensity the ensemble mean is superior
- The average spread of the track predictions is consistent with the error of the ensemble mean, and **the ensemble can generally distinguish between high and low uncertainty forecasts**
- The intensity predictions are underdispersive, as shown by the rank histograms and spread-skill comparison. However, **the ensemble can still distinguish between high and low uncertainty forecasts**

Results are largely consistent with those of the 2014 real-time demonstration, so with a robust overall sample of ~850 cases we are confident in the capabilities of the ensemble system and plan to transition it to operations

(3) COAMPS-TC/HWRF/GFDL combined ensemble results

2015 real-time homogeneous forecast sample

Atlantic

Number	Name	Forecasts
04L	Danny	16
06L	Fred	27
07L	Grace	7
09L	TD9	5
10L	Ida	19
11L	Joaquin	39

Eastern Pacific

Number	Name	Forecasts
09E	Guillermo	1
10E	Hilda	9
15E	Linda	6
20E	Patricia	8

Total = 137 forecasts

- **COAMPS-TC ensemble: 1 control + 10 perturbed members**
- **HWRF ensemble: 1 control + 20 perturbed members**
- **GFDL ensemble*: 1 control + 11 perturbed members**

* Because of a coding error, I only used the control and first 9 GFDL ensemble members in this validation. The track and intensity accuracy of the 10-member and 12-member GFDL ensemble means is nearly identical

(3) COAMPS-TC/HWRF/GFDL combined ensemble results

Deterministic verification: 2015 real-time track

For individual model, ensemble mean has accuracy similar to control member

Control forecasts:

COAMPS-TC: C00C

HWRF: HW00

GFDL: GP00

Combo: Consensus of

C00C, HW00, and GP00

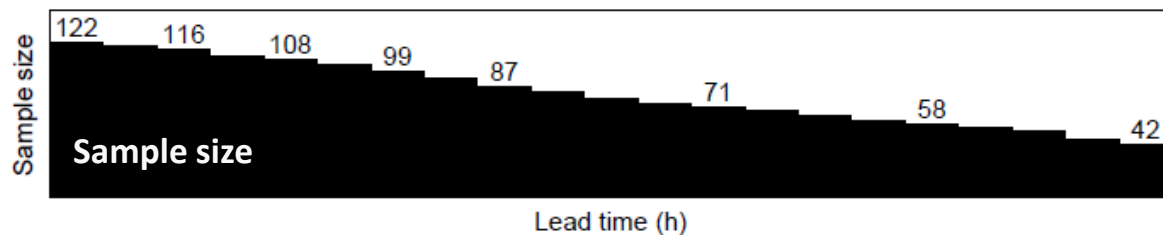
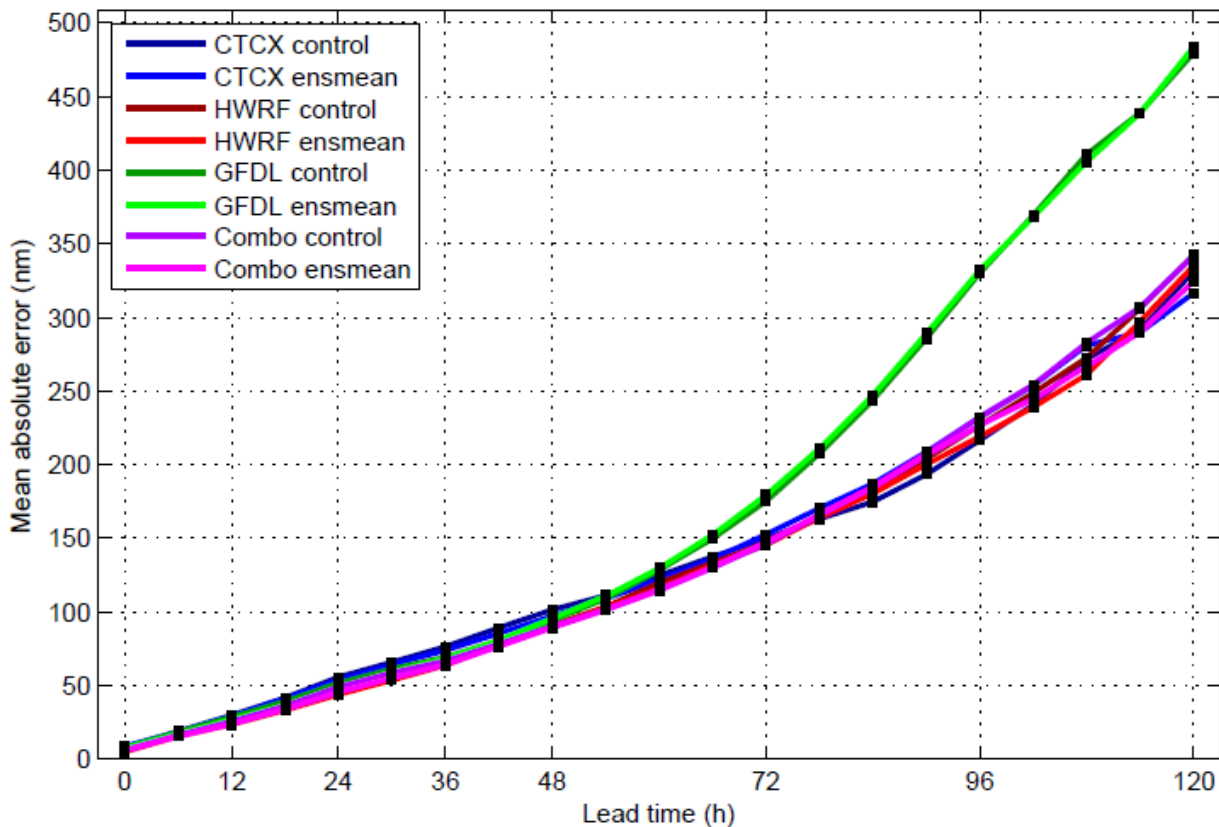
Ensemble mean requirements:

COAMPS-TC: 9 of 11 members

HWRF: 17 of 21 members

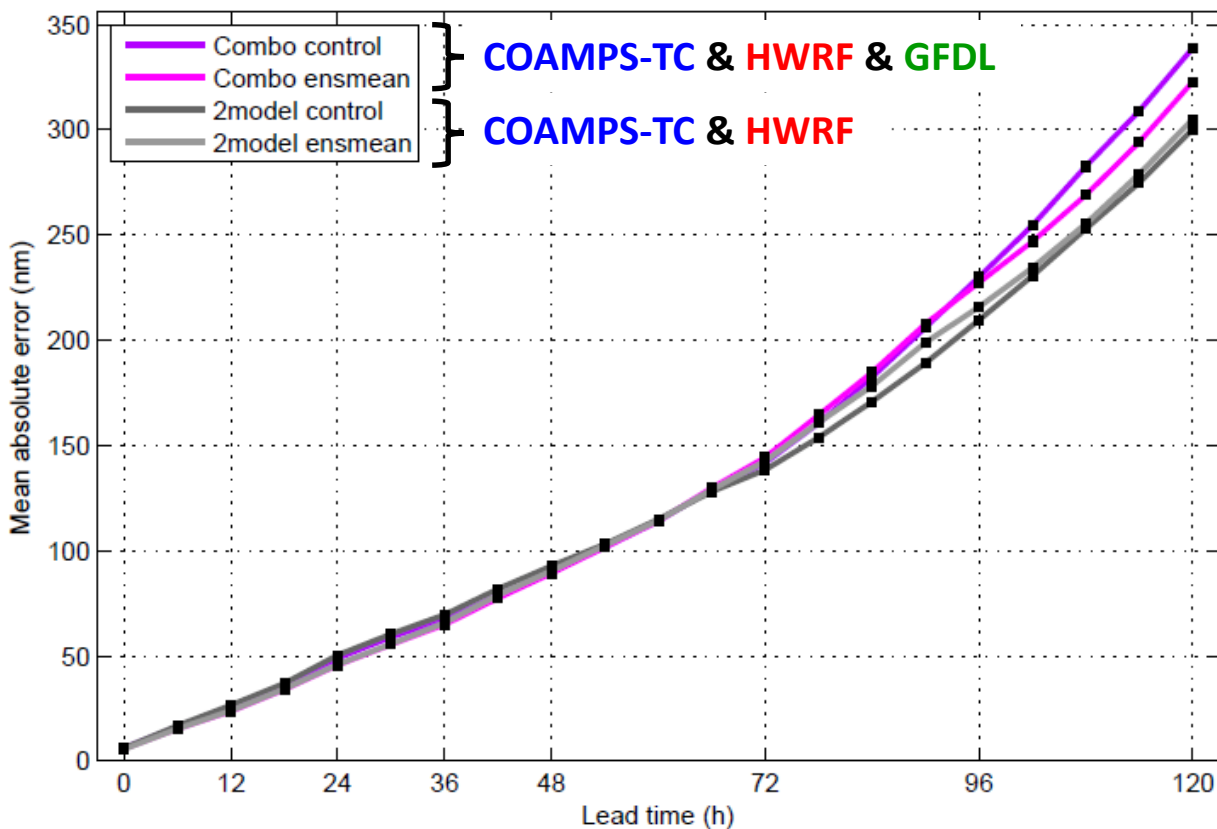
GFDL: 8 of 10 members

Combo: 34 of 42 members (from COAMPS-TC, HWRF and GFDL)



(3) COAMPS-TC/HWRF/GFDL combined ensemble results

Deterministic verification: 2015 real-time track



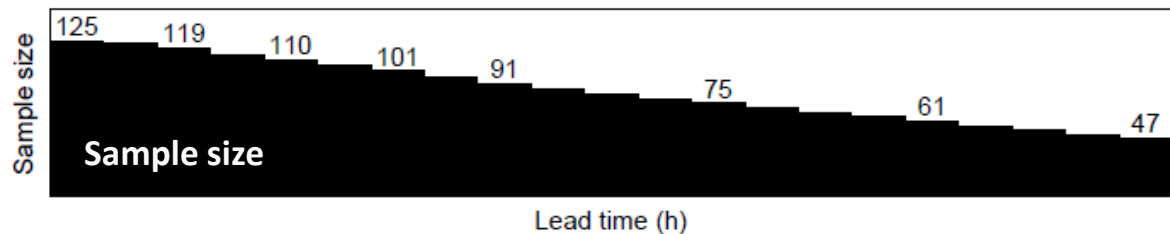
COAMPS-TC & HWRF & GFDL

COAMPS-TC & HWRF

COAMPS-TC & HWRF combination outperforms COAMPS-TC & HWRF & GFDL combination.

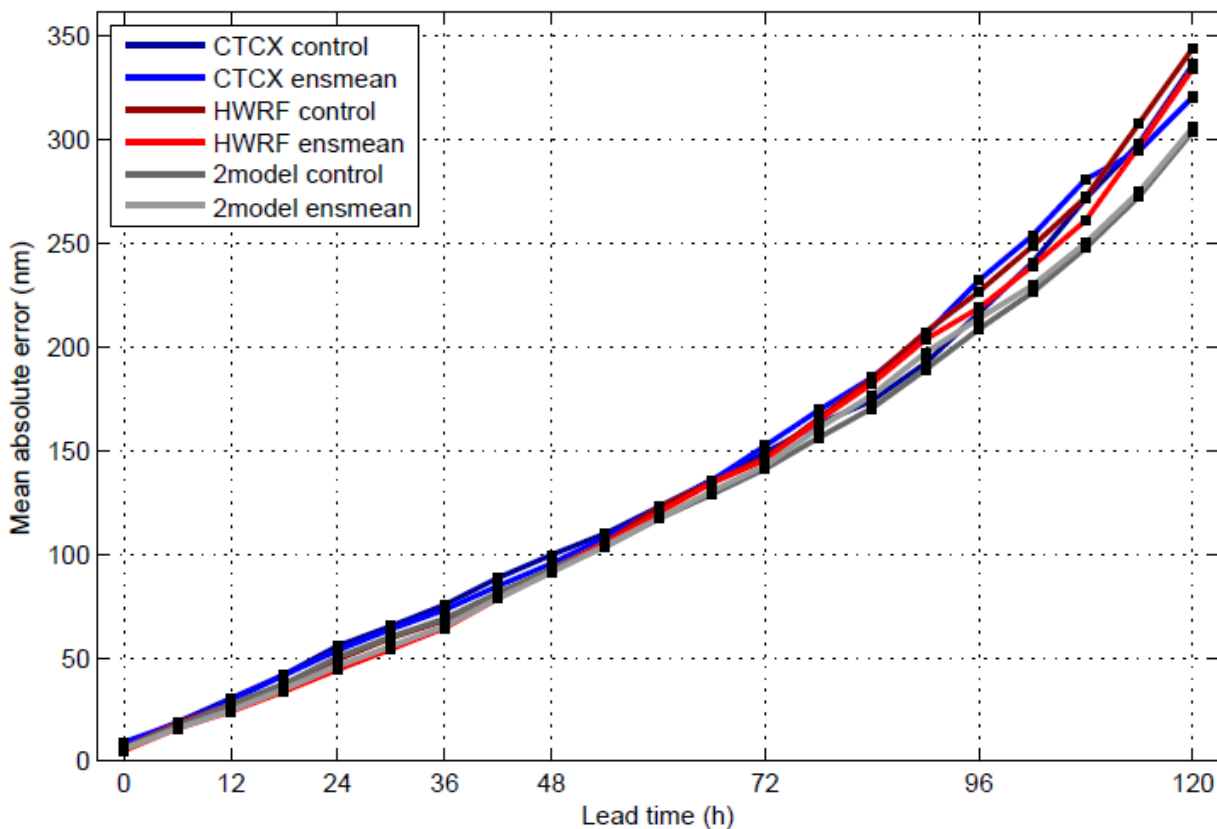
Control forecasts:
Combo: Consensus of COOC, HW00, and GP00
2model: Consensus of COOC, HW00

Ensemble mean requirements:
Combo: 34 of 42 members (from COAMPS-TC, HWRF and GFDL)
2model: 26 of 32 members (from COAMPS-TC and HWRF)



(3) COAMPS-TC/HWRF/GFDL combined ensemble results

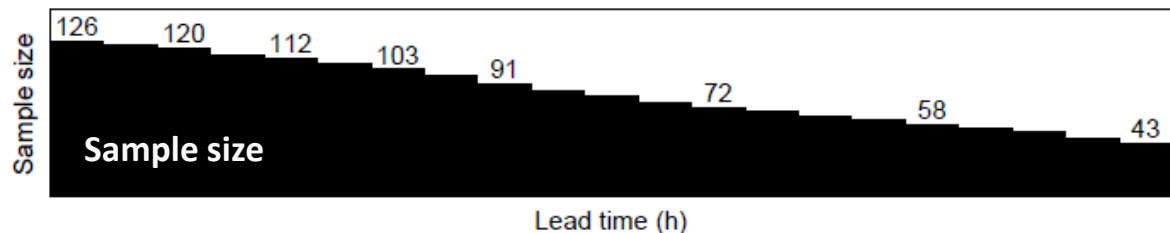
Deterministic verification: 2015 real-time track



COAMPS-TC & HWRF combination outperforms the two individual models

COAMPS-TC & HWRF ensemble mean has similar track accuracy to consensus of COAMPS-TC and HWRF control members

Control forecasts:
COAMPS-TC: COOC
HWRF: HW00
2model: Consensus of COOC, HW00

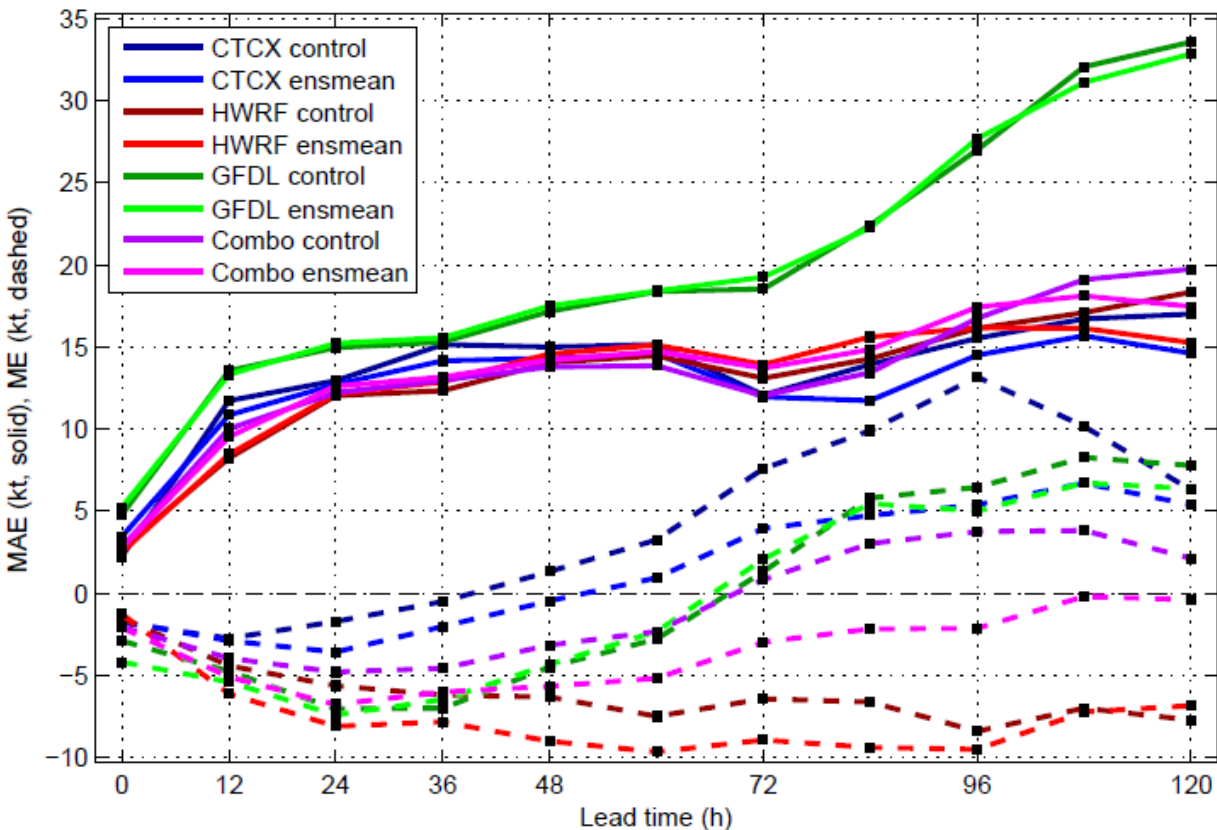


Ensemble mean requirements:
COAMPS-TC: 9 of 11 members
HWRF: 17 of 21 members
2model: 26 of 32 members (from COAMPS-TC and HWRF)

(3) COAMPS-TC/HWRF/GFDL combined ensemble results

Deterministic verification: 2015 real-time intensity

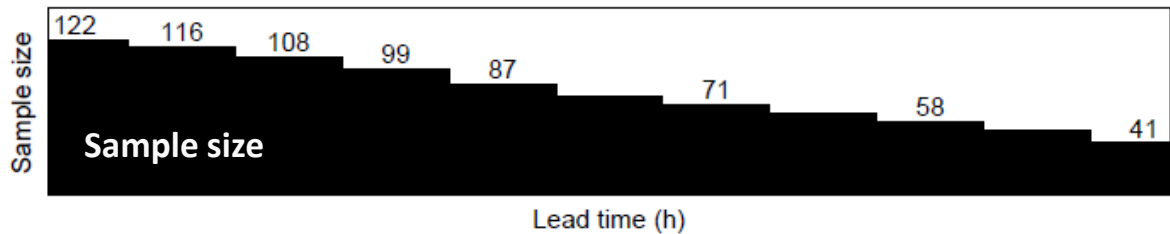
Solid: Mean absolute error Dashed: Mean error



For individual model, ensemble mean has accuracy similar to or somewhat better than control member

Control forecasts:
COAMPS-TC: C00C
HWRF: HW00
GFDL: GP00
Combo: Consensus of C00C, HW00, and GP00

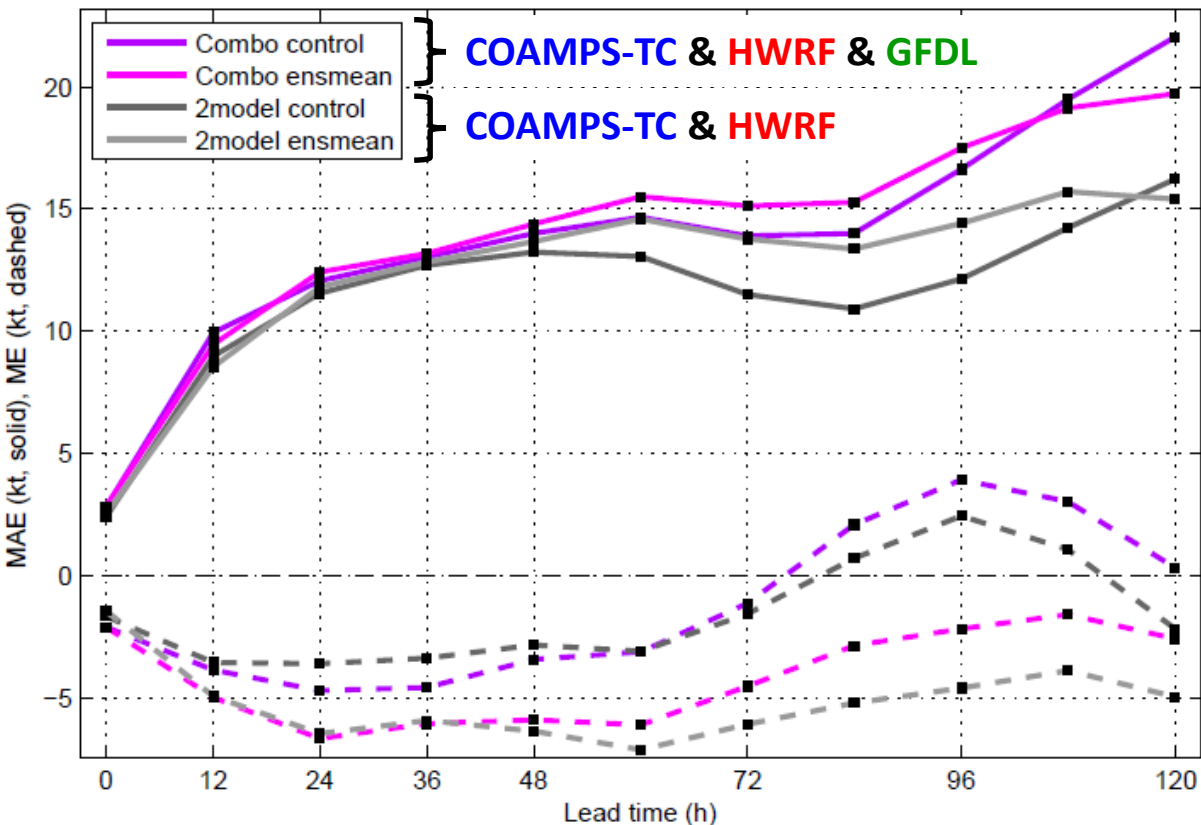
Ensemble mean requirements:
COAMPS-TC: 9 of 11 members
HWRF: 17 of 21 members
GFDL: 8 of 10 members
Combo: 34 of 42 members (from COAMPS-TC, HWRF and GFDL)



(3) COAMPS-TC/HWRF/GFDL combined ensemble results

Deterministic verification: 2015 real-time intensity

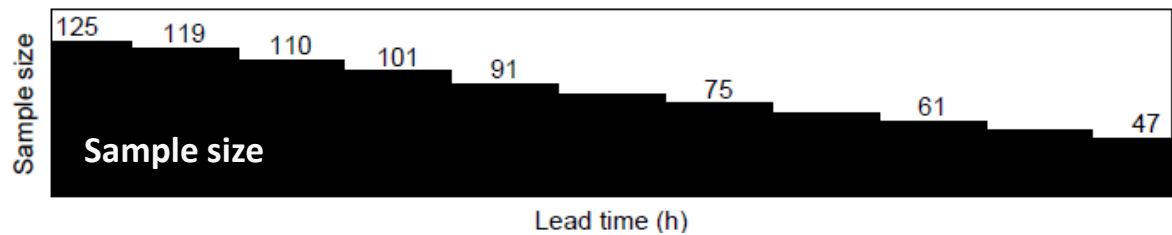
Solid: Mean absolute error Dashed: Mean error



COAMPS-TC & HWRF combination outperforms COAMPS-TC & HWRF & GFDL combination.

Control forecasts:
Combo: Consensus of COOC, HW00, and GP00
2model: Consensus of COOC, HW00

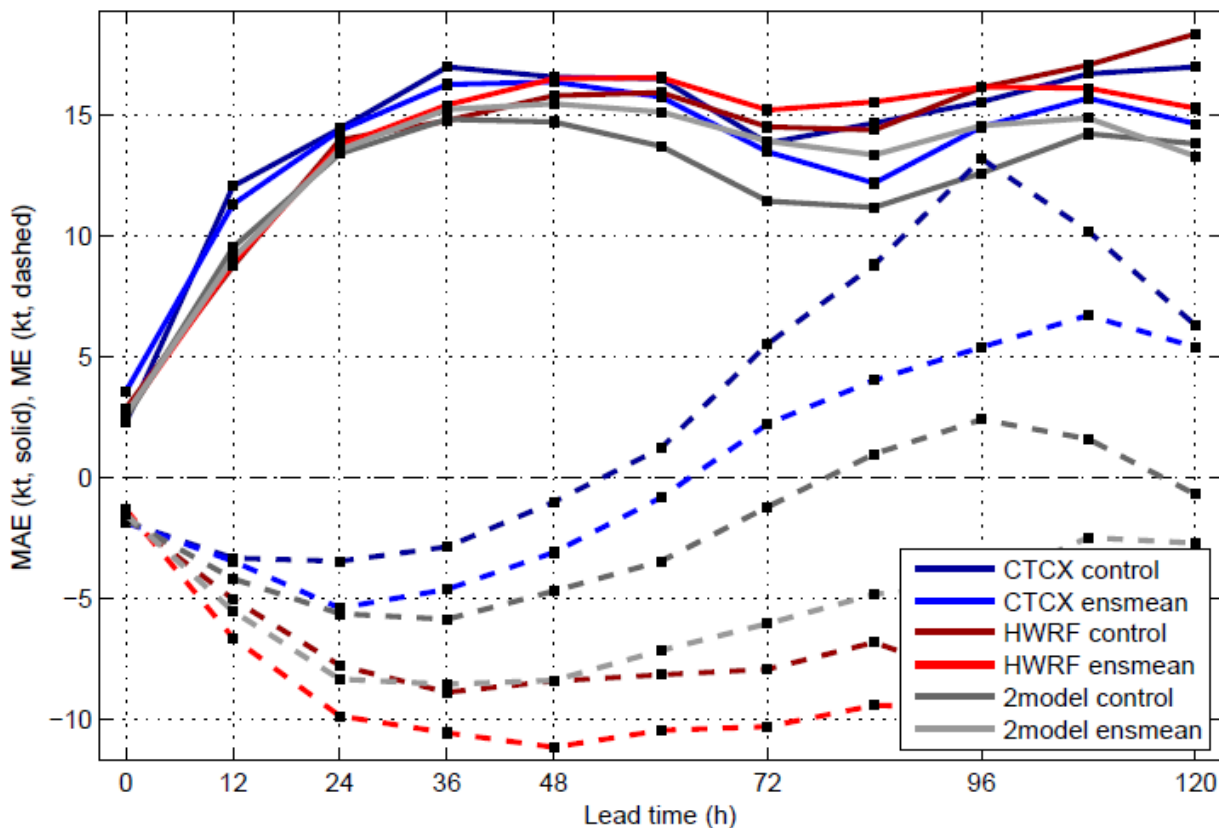
Ensemble mean requirements:
Combo: 34 of 42 members (from COAMPS-TC, HWRF and GFDL)
2model: 26 of 32 members (from COAMPS-TC and HWRF)



(3) COAMPS-TC/HWRF/GFDL combined ensemble results

Deterministic verification: 2015 real-time intensity

Solid: Mean absolute error Dashed: Mean error



COAMPS-TC & HWRF combination outperforms the two individual models

Consensus of COAMPS-TC and HWRF controls has superior accuracy and bias w.r.t COAMPS & HWRF ensemble mean

Control forecasts:

COAMPS-TC: COOC

HWRF: HW00

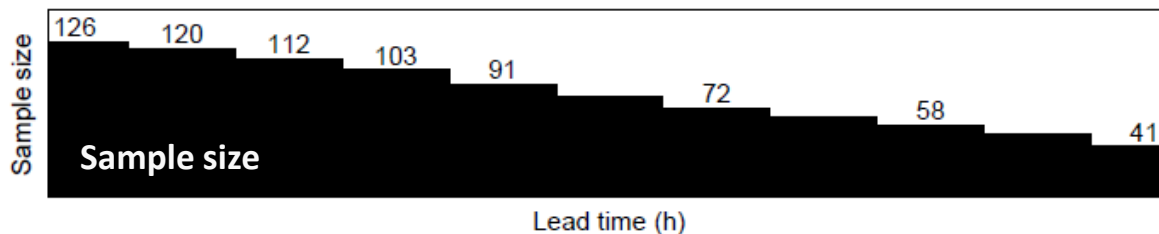
2model: Consensus of COOC, HW00

Ensemble mean requirements:

COAMPS-TC: 9 of 11 members

HWRF: 17 of 21 members

2model: 26 of 32 members (from COAMPS-TC and HWRF)



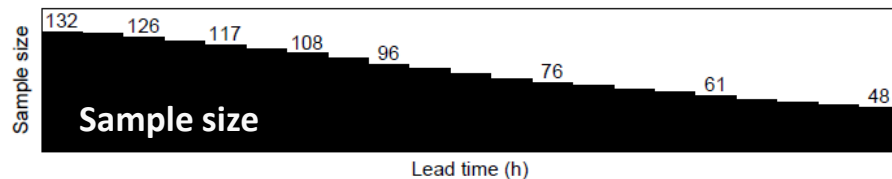
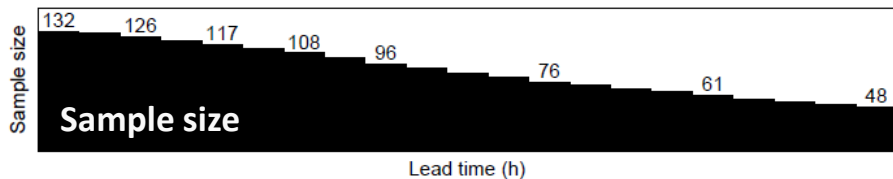
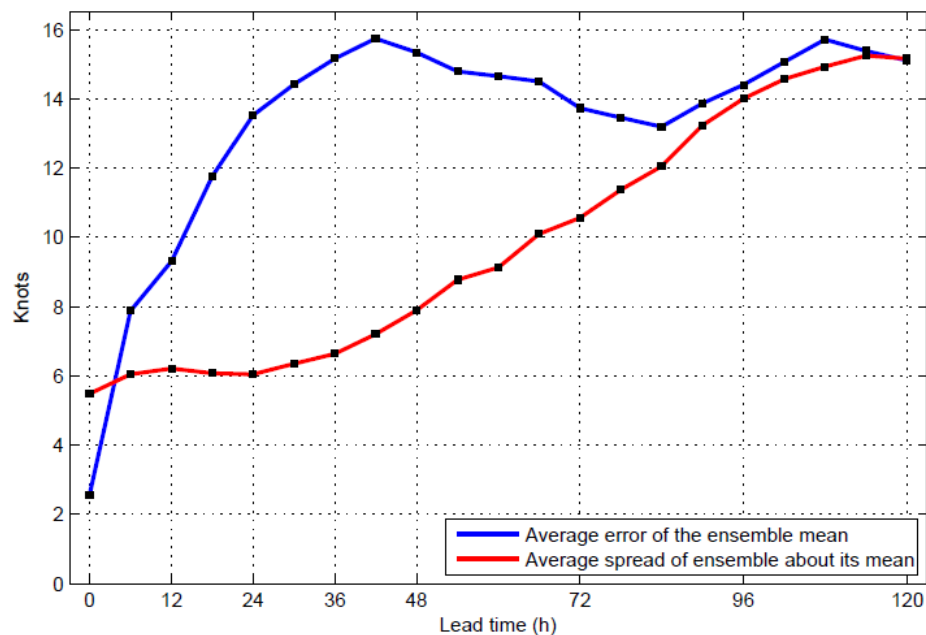
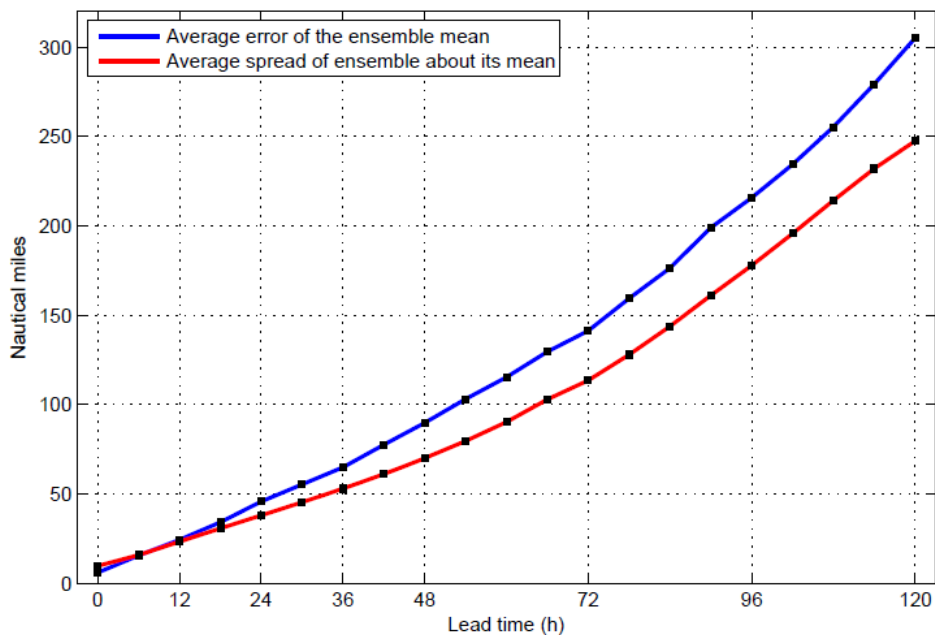
(3) COAMPS-TC/HWRF/GFDL combined ensemble results

Probabilistic verification: Ensemble spread vs. Ensemble mean error

Track

COAMPS-TC & HWRF

Intensity

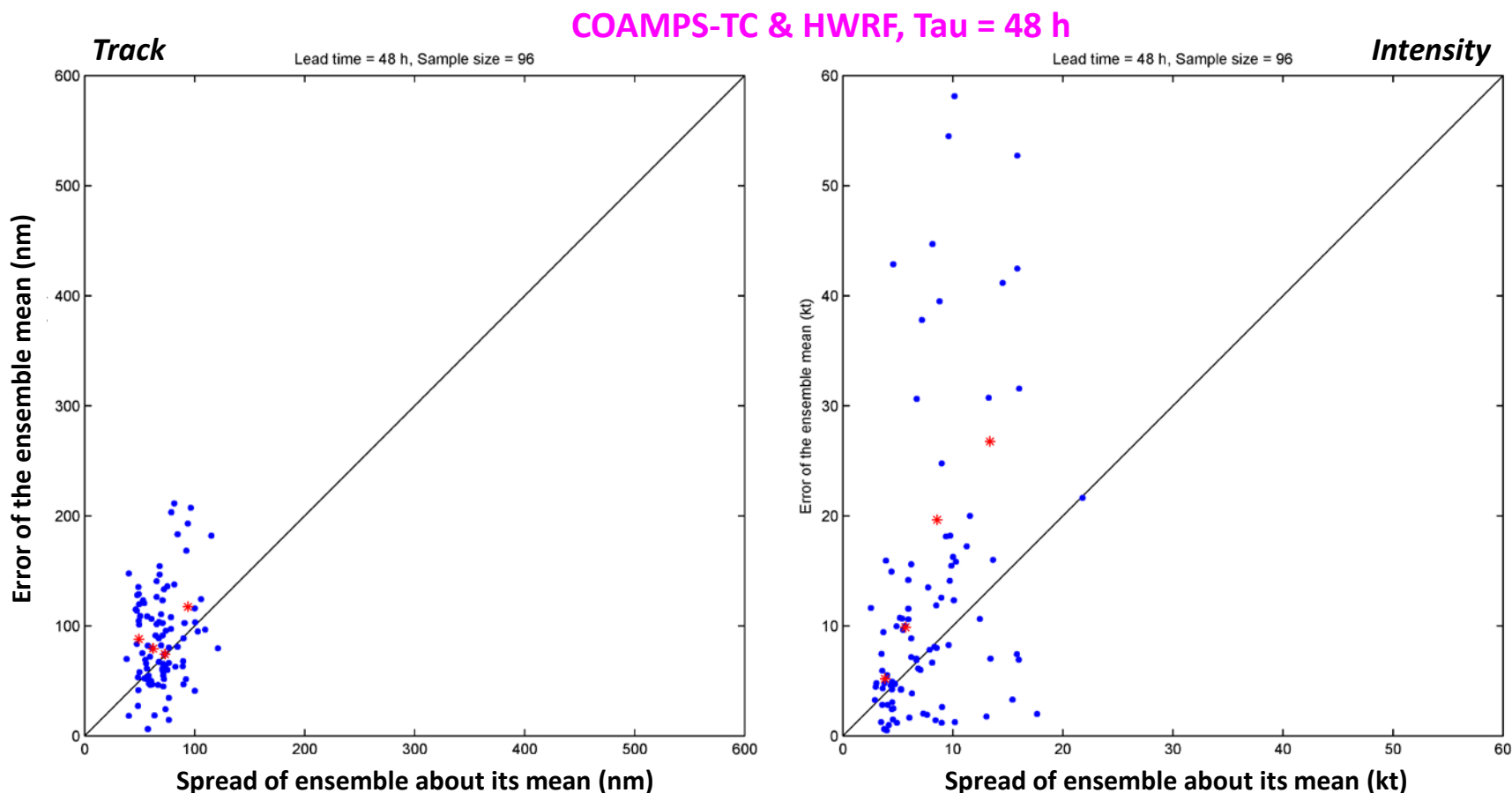


For track, ensemble is somewhat underdispersive for all but the earliest lead times

For track, ensemble needs more spread from 12 to 84 h; spread growth is too sluggish

(3) COAMPS-TC/HWRF/GFDL combined ensemble results

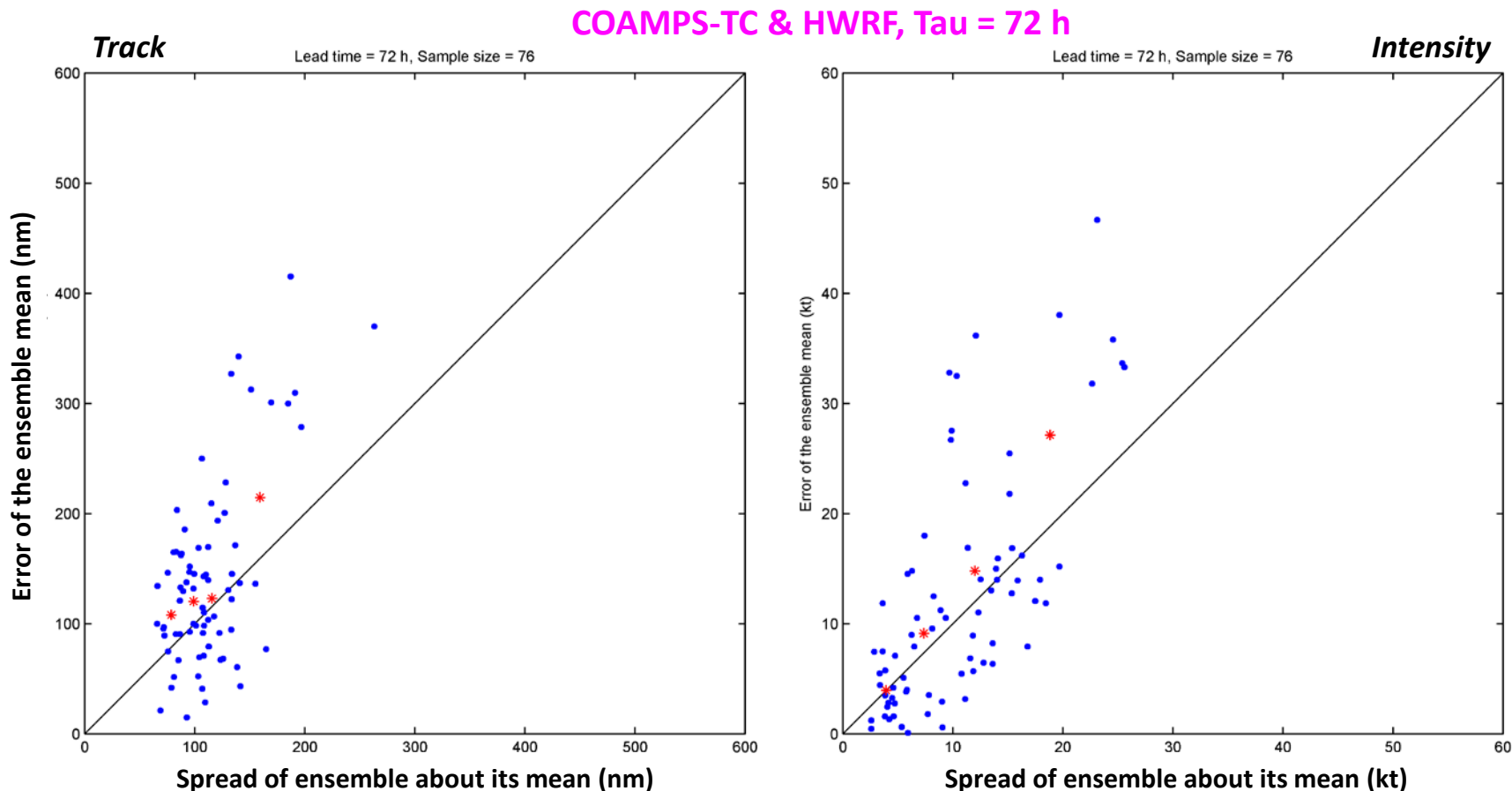
Probabilistic verification: Ensemble spread vs. Ensemble mean error



- The **blue dots** represent individual forecasts and **red stars** are bin-averages (4 equally populated bins)
- Would like to see red stars line up along diagonal and few blue dots in upper left (large error / low spread)

(3) COAMPS-TC/HWRF/GFDL combined ensemble results

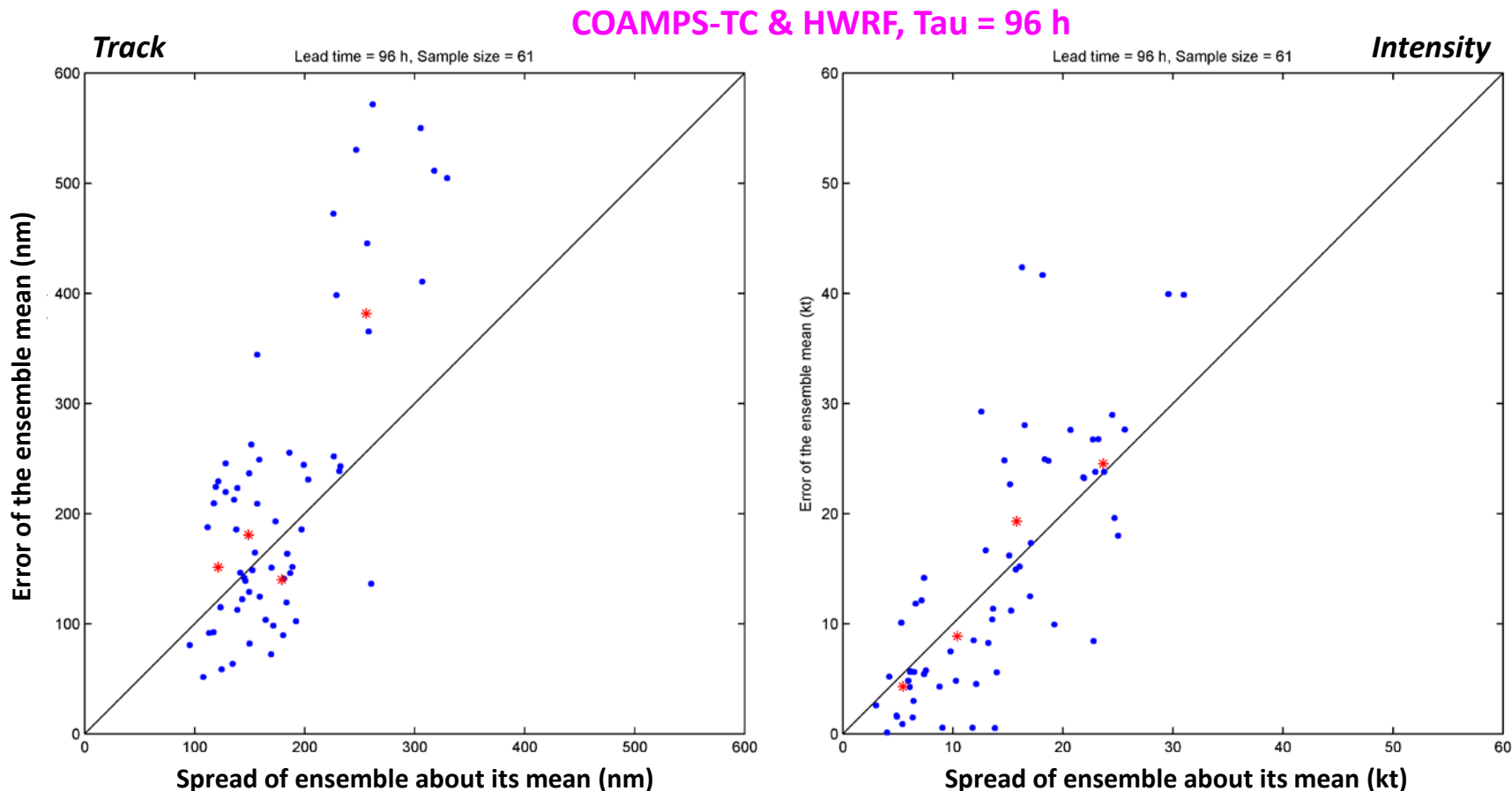
Probabilistic verification: Ensemble spread vs. Ensemble mean error



- The **blue dots** represent individual forecasts and **red stars** are bin-averages (4 equally populated bins)
- Would like to see red stars line up along diagonal and few blue dots in upper left (large error / low spread)

(3) COAMPS-TC/HWRF/GFDL combined ensemble results

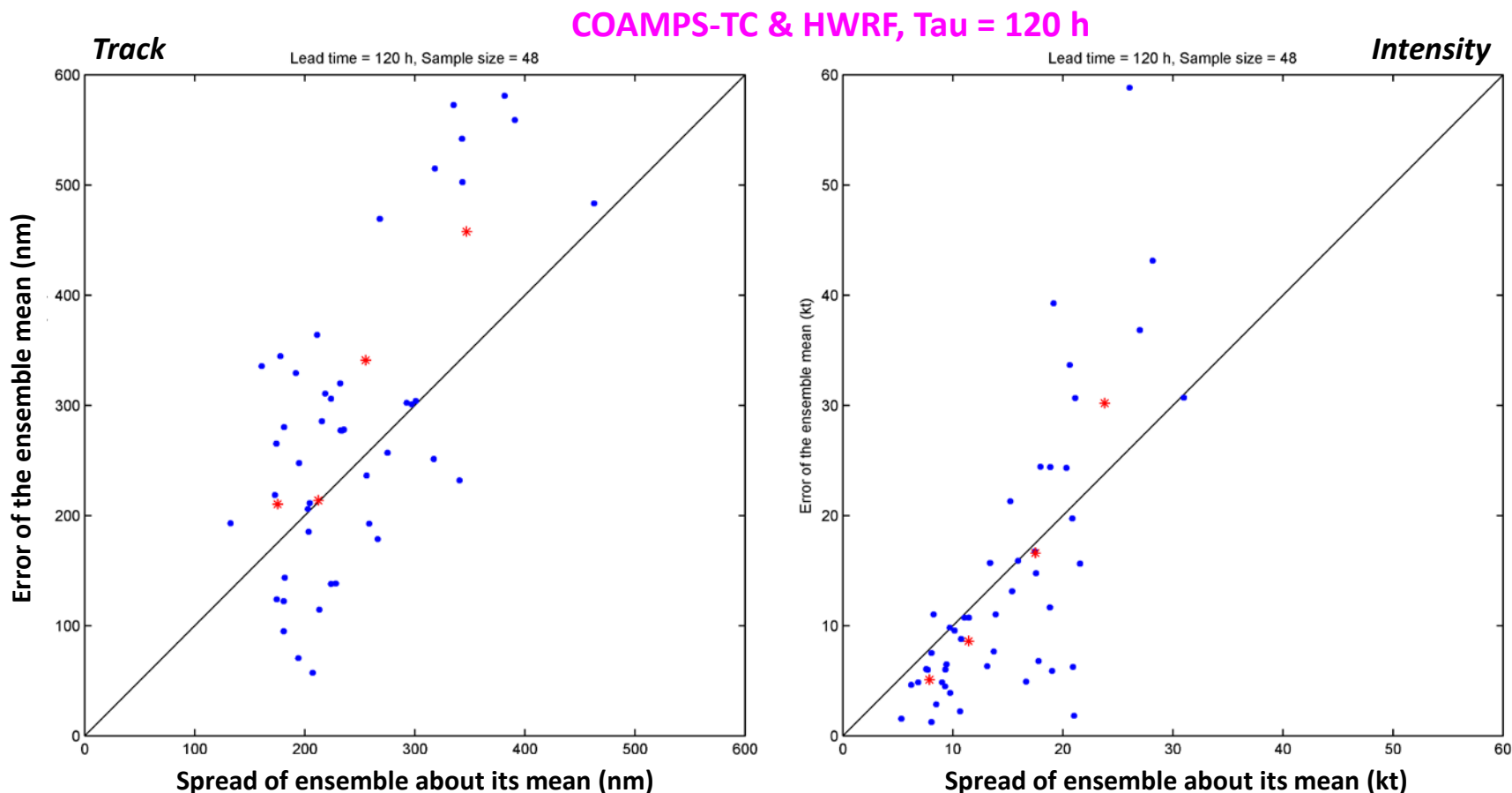
Probabilistic verification: Ensemble spread vs. Ensemble mean error



- The **blue dots** represent individual forecasts and **red stars** are bin-averages (4 equally populated bins)
- Would like to see red stars line up along diagonal and few blue dots in upper left (large error / low spread)

(3) COAMPS-TC/HWRF/GFDL combined ensemble results

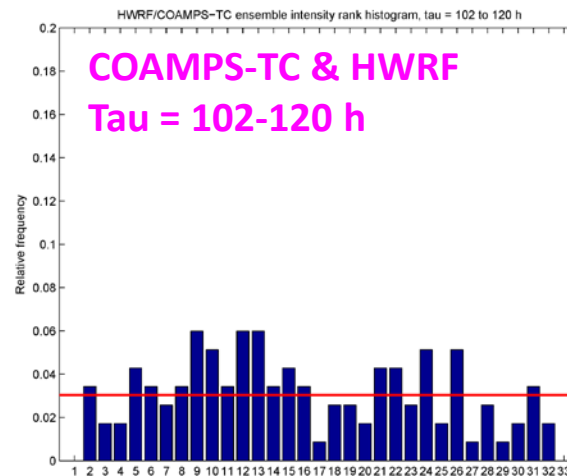
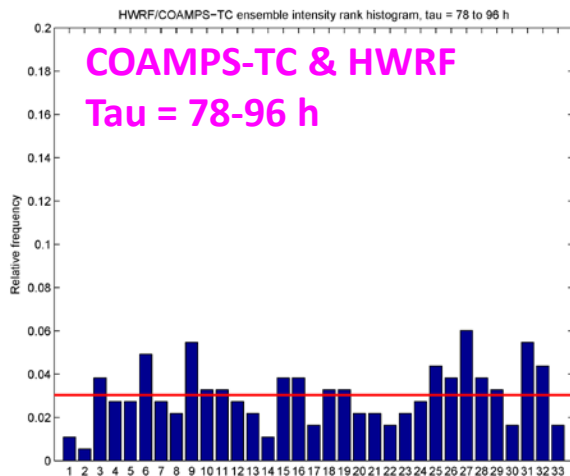
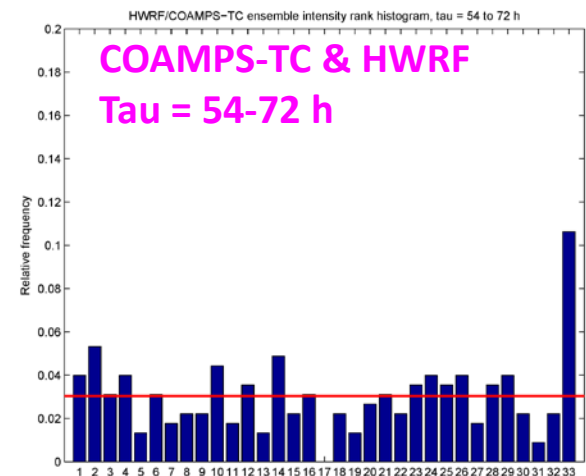
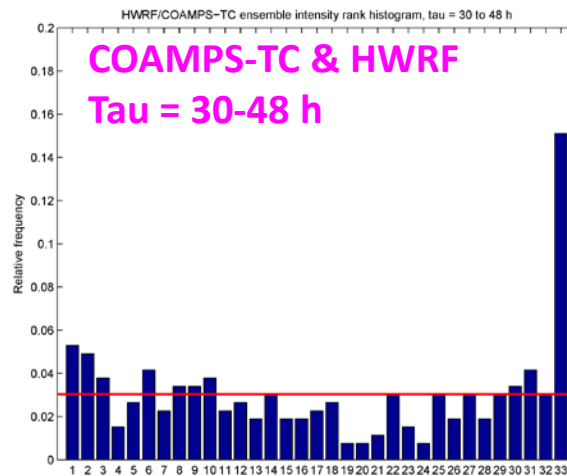
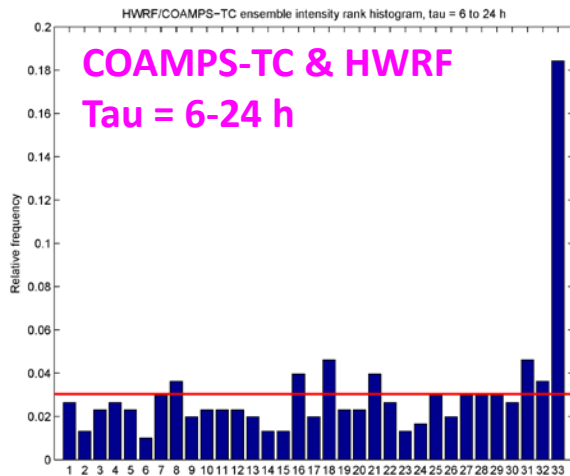
Probabilistic verification: Ensemble spread vs. Ensemble mean error



As for the COAMPS-TC-only ensemble, the COAMPS-TC & HWRF combined ensemble can distinguish between low-uncertainty and high-uncertainty cases, for both track and intensity.
Large spread = Higher chance of large ensemble mean error

(3) COAMPS-TC/HWRF/GFDL combined ensemble results

Probabilistic verification: Intensity rank histograms



- Would like to see all the blue bars near the red line, indicating equal probability observation falls between any two ranked ensemble member (or falls outside either end of ensemble)

There is overpopulation of the right-most bin (all ensemble member forecasts < observed intensity) for 6-72 h lead time, but reliability is very good at later lead times

(3)

Conclusions for combined ensemble

- Nearly 140 real-time cases were predicted by all three ensembles, in both the Atlantic and Eastern Pacific (small sample, but more cases than in 2014)
- For this particular sample, the COAMPS-TC & HWRF two model combination outperforms the COAMPS-TC & HWRF & GFDL three model combination in deterministic validation
- COAMPS-TC & HWRF control consensus and ensemble mean outperform their single-model counterparts in deterministic validation
- The combined ensemble (either two or three model) spread is not large enough, particularly for intensity at the earlier lead times. However, **the ensemble can distinguish between low-uncertainty and high-uncertainty cases, for both track and intensity**

(4)

Future directions for COAMPS-TC ensemble

New three-year Navy project to transition COAMPS-TC ensemble to operations at FNMOC

FY16: Real-time demo in Atlantic and Western Pacific

FY17: Transition basic capability into operations (10 members)

FY18: Introduce perturbed physics to account for uncertainty due to model error

COAMPS-TC ensemble development challenges

- Increase intensity spread without degrading the ability of the ensemble to distinguish between low and high uncertainty cases
- Model testing & development: 1 ensemble test run ~ 11 deterministic test runs. EPS uses COAMPS-TC model that has been optimized for deterministic prediction
- Product development & validation: Intensification rate probabilities, combining statistical and ensemble-based forecast information, input into decision aids.

(4)

Future directions for COAMPS-TC ensemble

New three-year Navy project to transition COAMPS-TC ensemble to operations at FNMOC

FY16: Real-time demo in Atlantic and Western Pacific

FY17: Transition basic capability into operations (10 members)

FY18: Introduce perturbed physics to account for uncertainty due to model error

COAMPS-TC ensemble development challenges

- Increase intensity spread without degrading the ability of the ensemble to distinguish between low and high uncertainty cases
- Model testing & development: 1 ensemble test run ~ 11 deterministic test runs. EPS uses COAMPS-TC model that has been optimized for deterministic prediction
- Product development & validation: Intensification rate probabilities, combining statistical and ensemble-based forecast information, input into decision aids.

We would like to continue to partner with NOAA/HFIP to work together on these challenges

EXTRA SLIDES

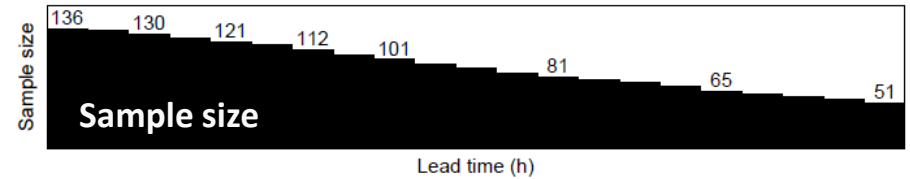
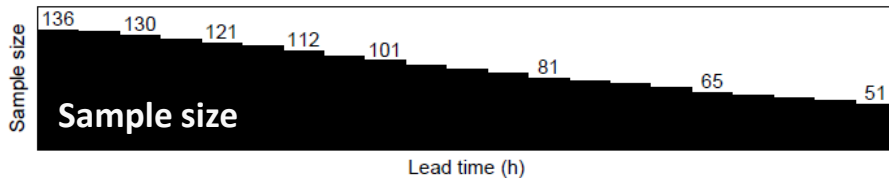
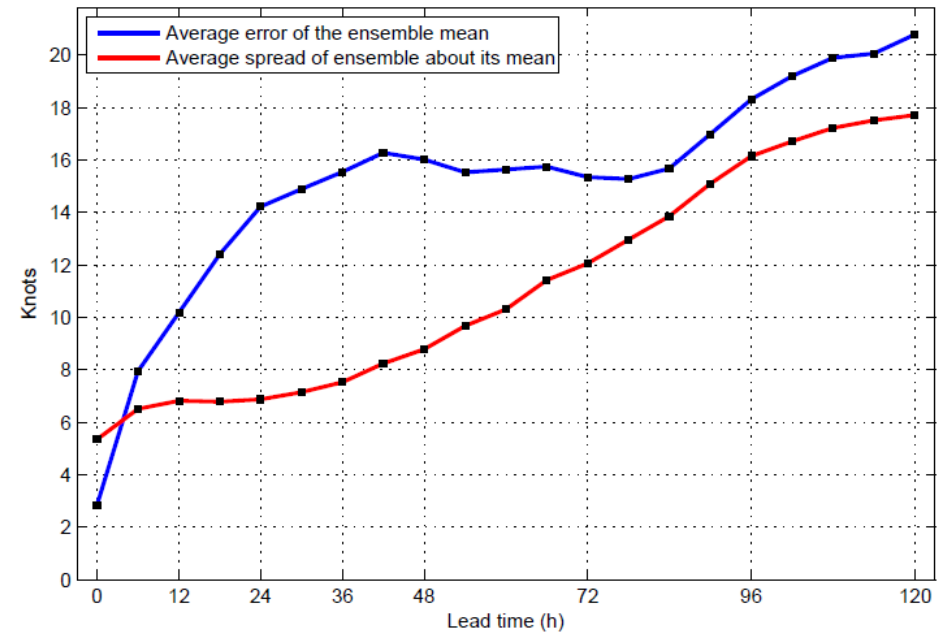
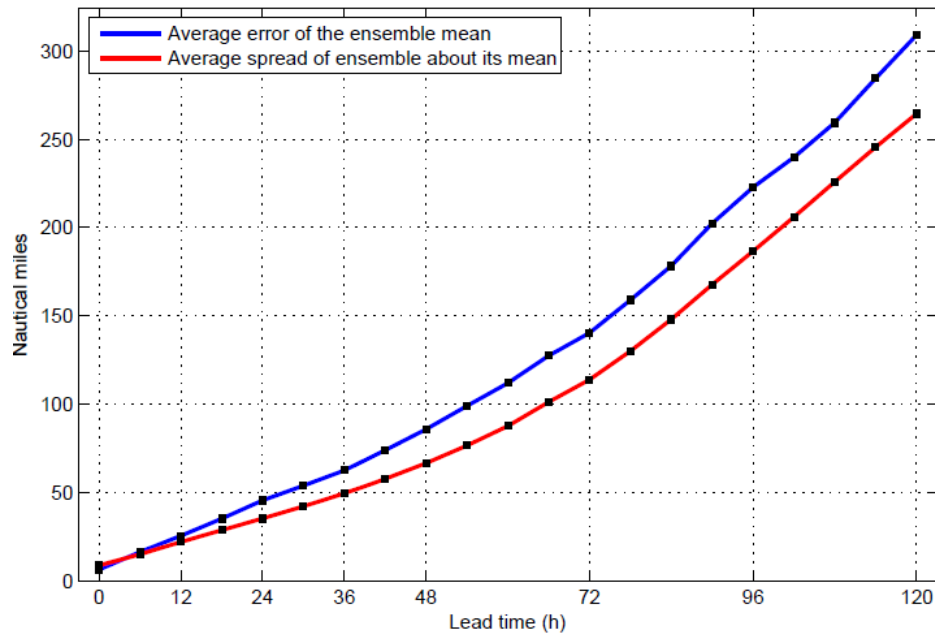
COAMPS-TC/HWRF/GFDL combined ensemble results

Probabilistic verification: Ensemble spread vs. Ensemble mean error

Track

COAMPS-TC & HWRF & GFDL

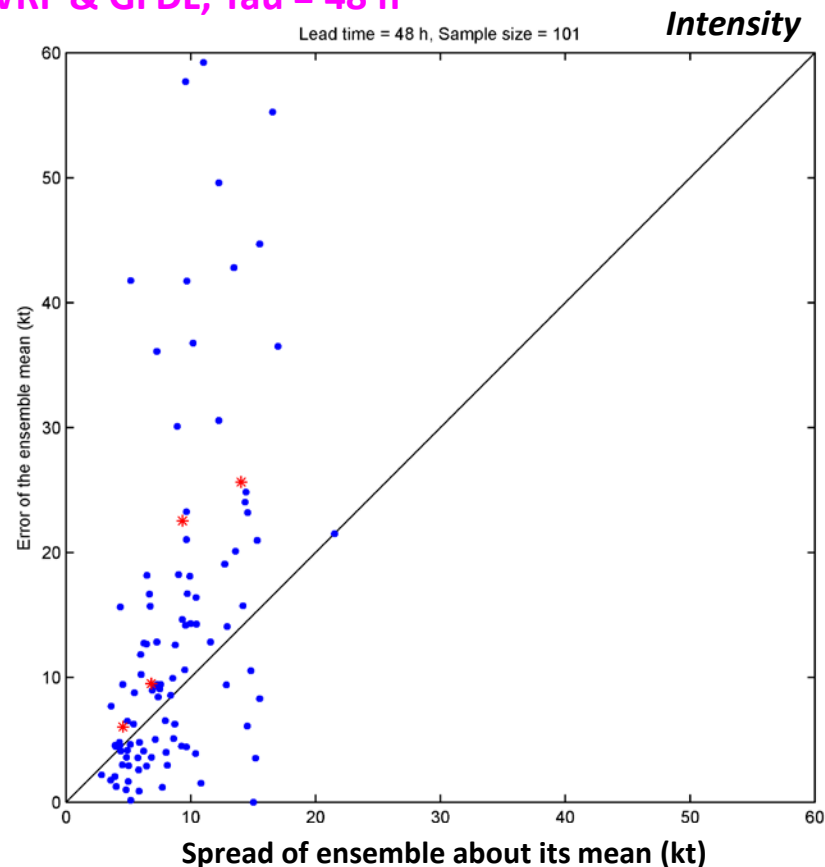
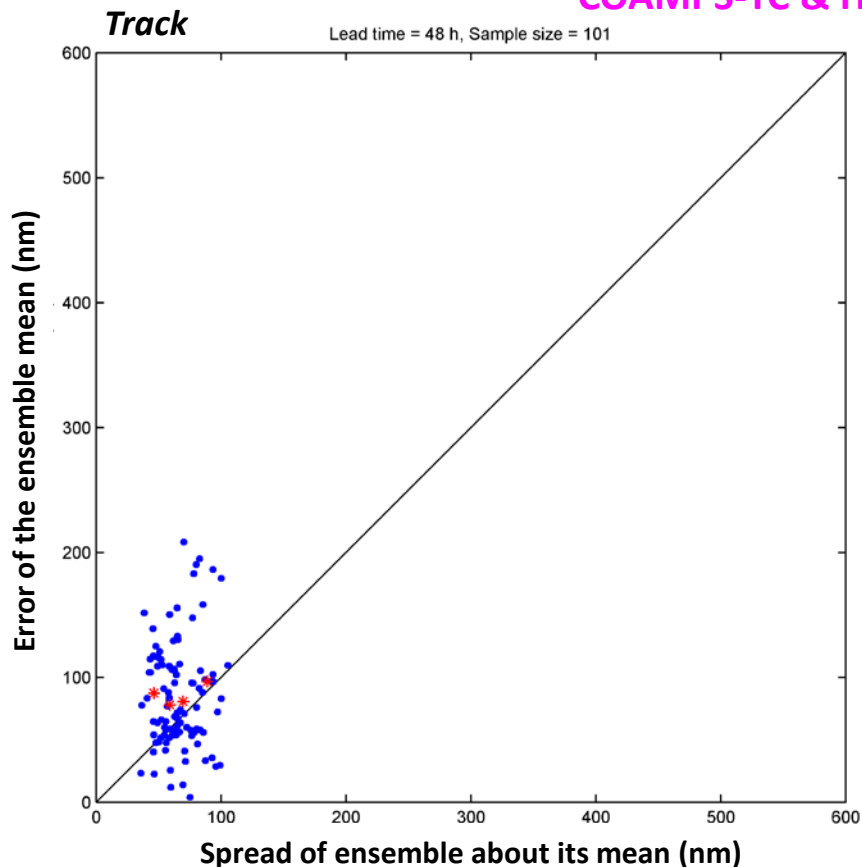
Intensity



COAMPS-TC/HWRF/GFDL combined ensemble results

Probabilistic verification: Ensemble spread vs. Ensemble mean error

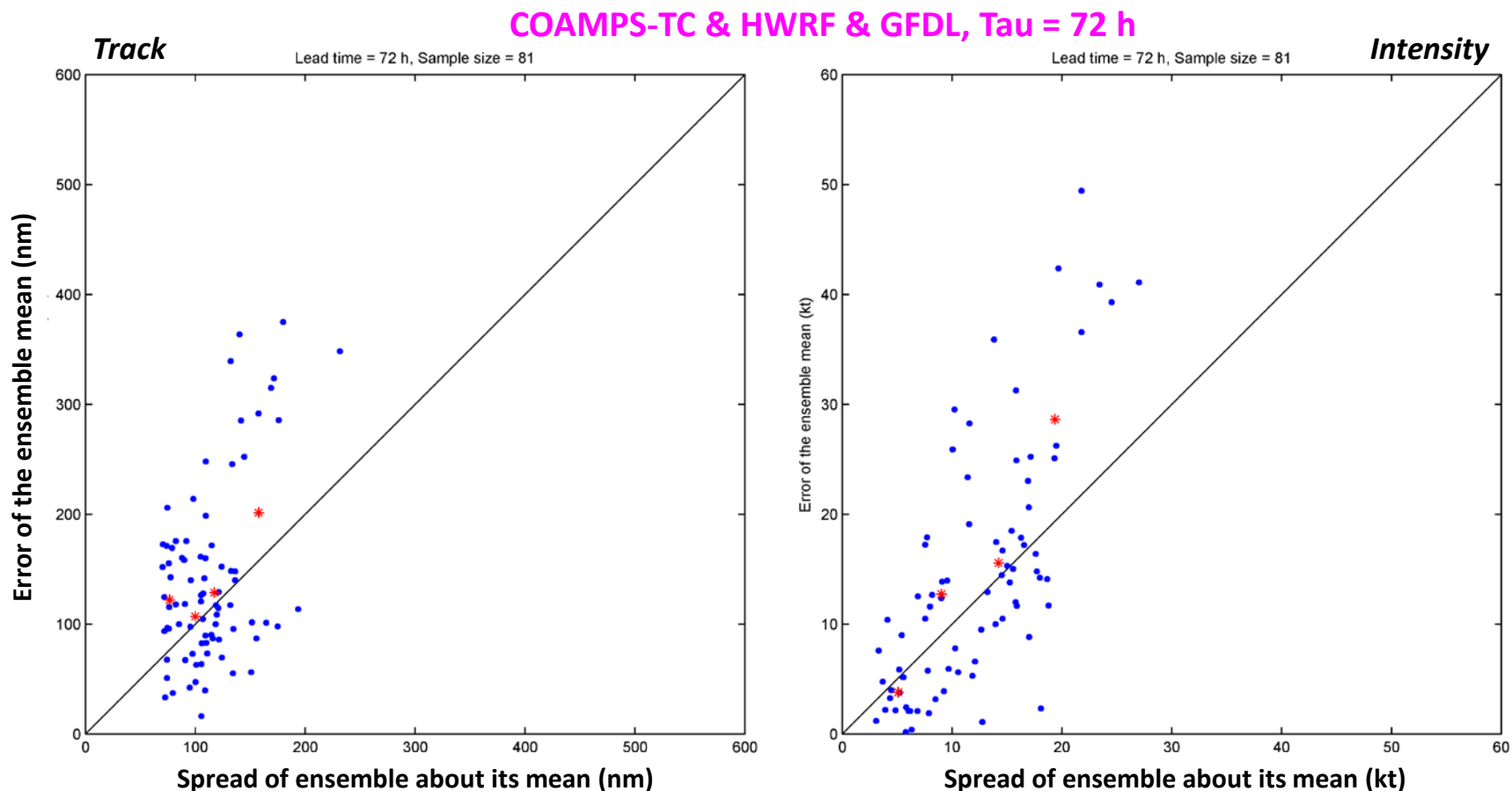
COAMPS-TC & HWRF & GFDL, Tau = 48 h



- The **blue dots** represent individual forecasts and **red stars** are bin-averages (4 equally populated bins)
- Would like to see red stars line up along diagonal and few blue dots in upper left (large error / low spread)

COAMPS-TC/HWRF/GFDL combined ensemble results

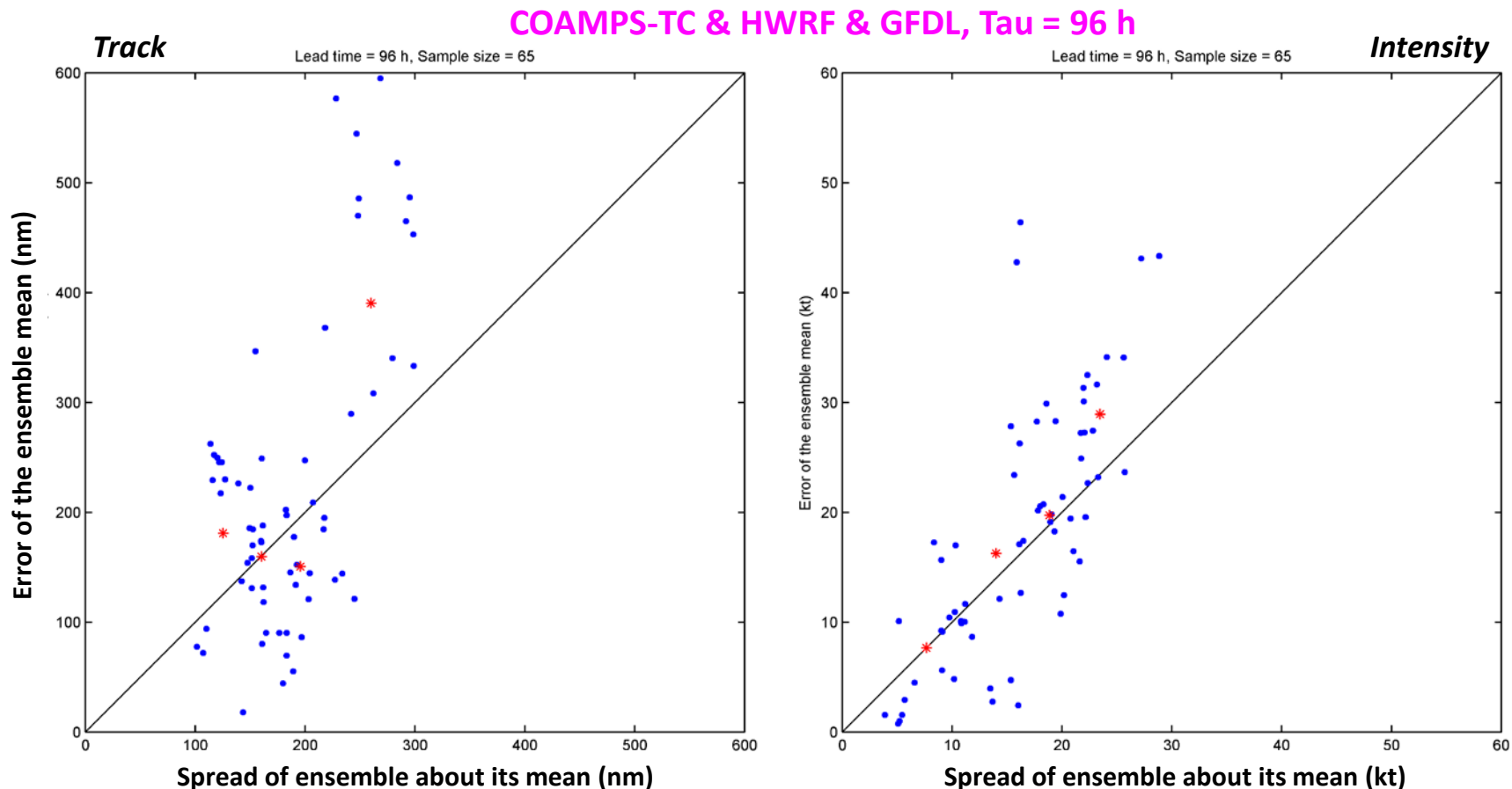
Probabilistic verification: Ensemble spread vs. Ensemble mean error



- The **blue dots** represent individual forecasts and **red stars** are bin-averages (4 equally populated bins)
- Would like to see red stars line up along diagonal and few blue dots in upper left (large error / low spread)

COAMPS-TC/HWRF/GFDL combined ensemble results

Probabilistic verification: Ensemble spread vs. Ensemble mean error

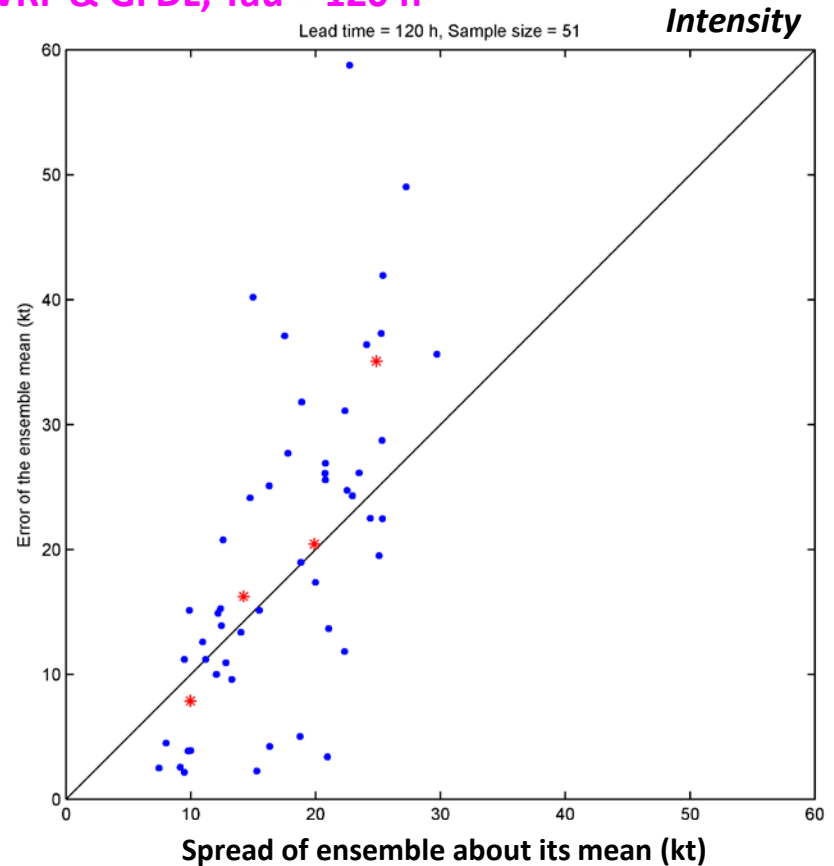
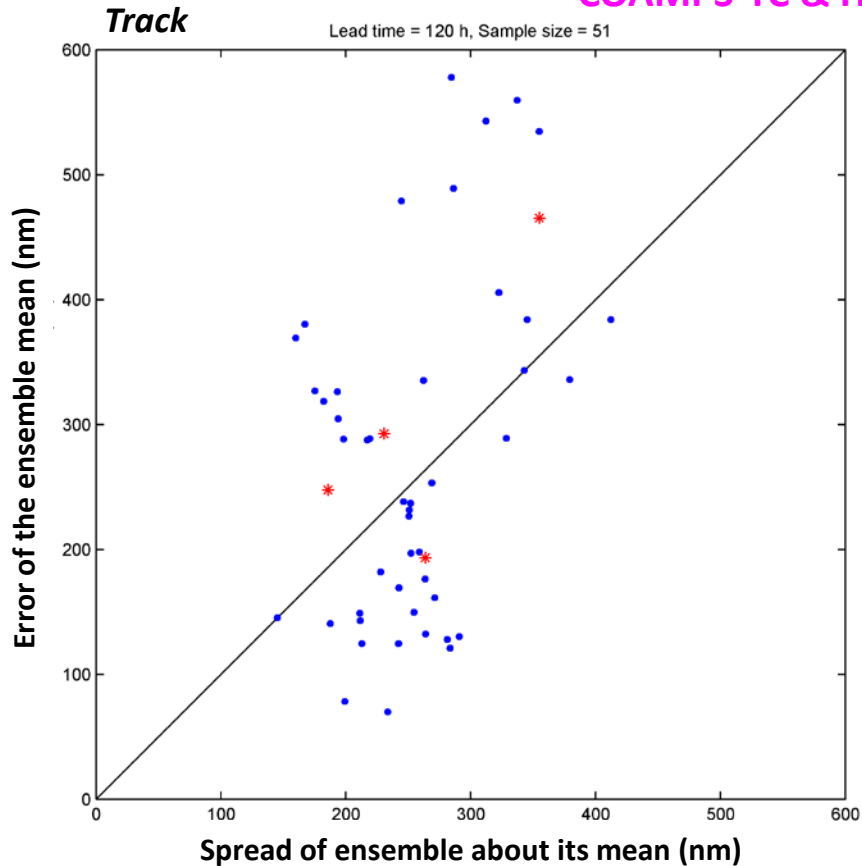


- The **blue dots** represent individual forecasts and **red stars** are bin-averages (4 equally populated bins)
- Would like to see red stars line up along diagonal and few blue dots in upper left (large error / low spread)

COAMPS-TC/HWRF/GFDL combined ensemble results

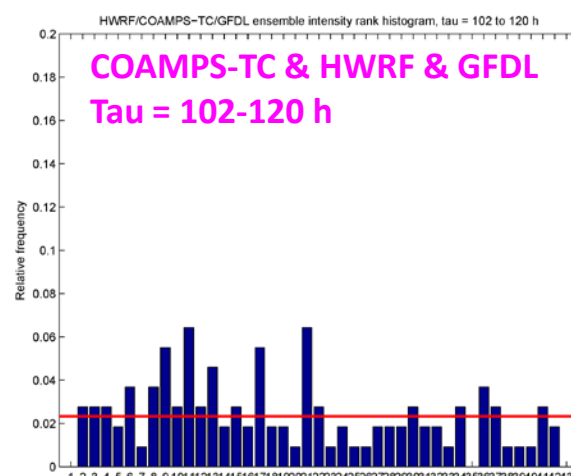
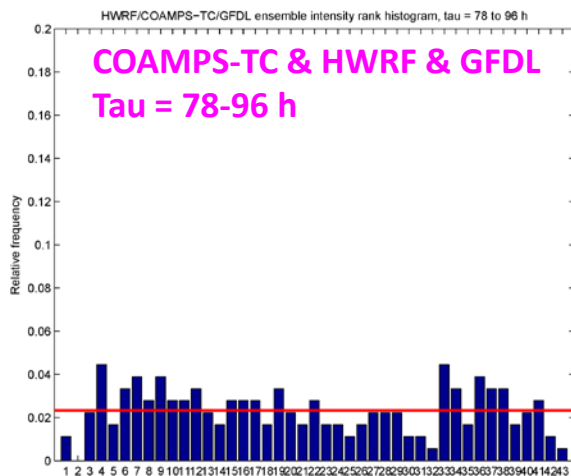
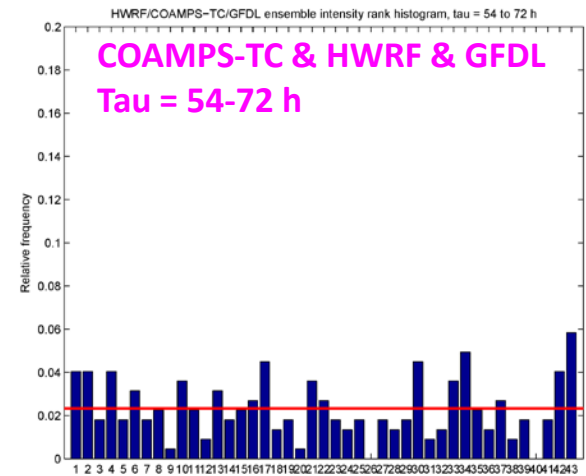
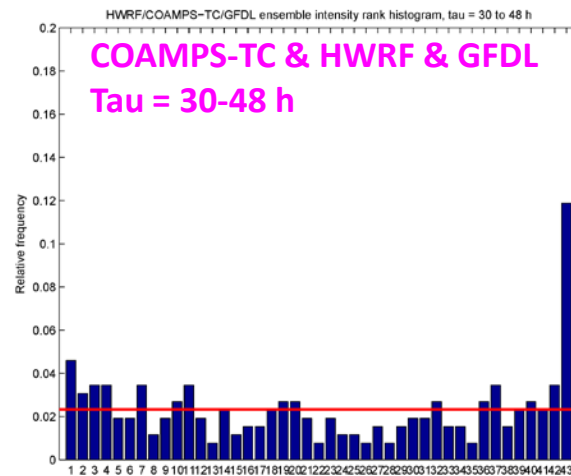
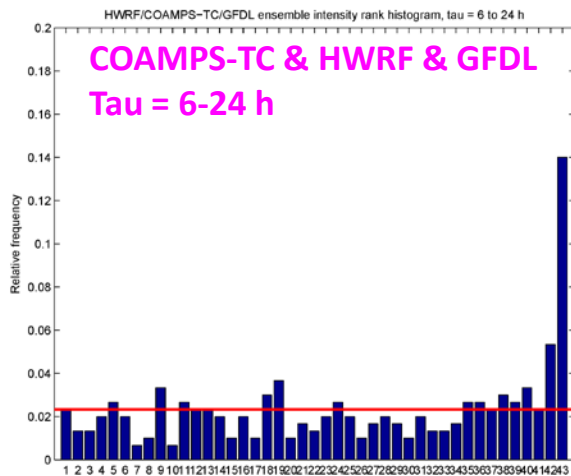
Probabilistic verification: Ensemble spread vs. Ensemble mean error

COAMPS-TC & HWRF & GFDL, Tau = 120 h



COAMPS-TC/HWRF/GFDL combined ensemble results

Probabilistic verification: Intensity rank histograms



- Would like to see all the blue bars near the red line, indicating equal probability observation falls between any two ranked ensemble member (or falls outside either end of ensemble)