

# Verification of RI Forecasts in 2021

James L. Franklin and Matt Onderlinde

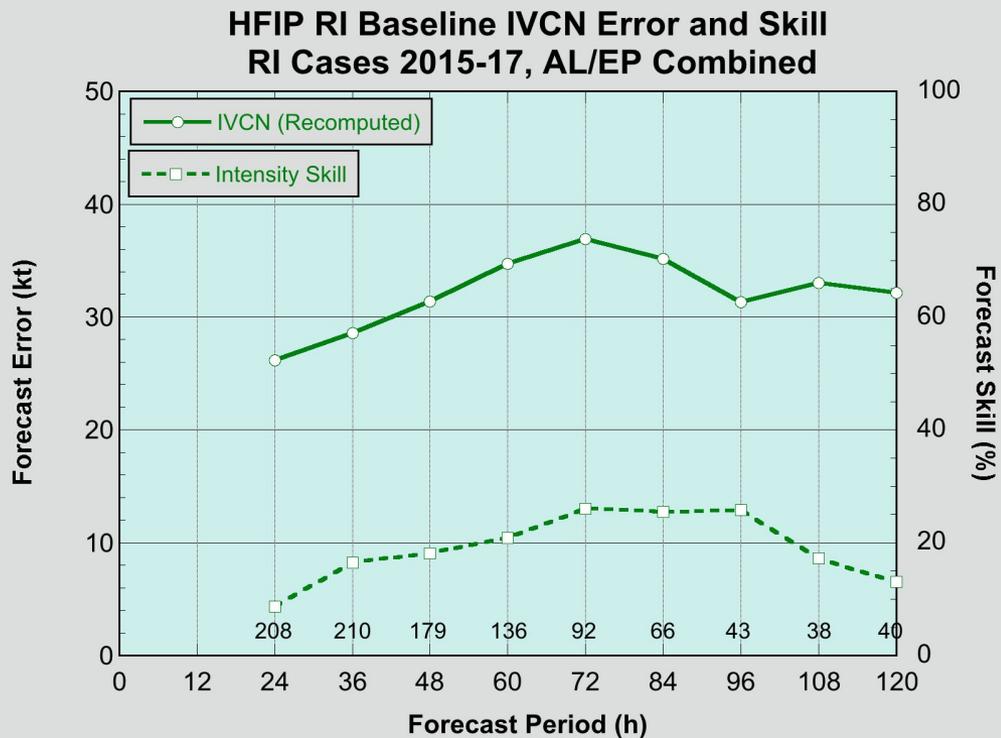
HFIP Annual Meeting  
November 2021



# HFIP RI Goal and Baseline

- Goal: Reduce intensity forecast guidance errors by 50% for RI events.
- Baseline: IVCN MAE, Atlantic and eastern Pacific combined, 2015-17, at 24, 36, 48, 72, 96, and 120 h, evaluated at only those times when RI (30 kt over the preceding 24 h) was either ongoing or forecast by one or more of the IVCN member models.
  - Standard verification rules apply.
  - Baseline IVCN comprises all the models used in operational IVCN at any time from 2015-17: DSHP, LGEM, GHMI, HWFI, CTCI. Future composition of IVCN will vary.
- IVCN (2018-21): DSHP, LGEM, HWFI, CTCI, HMNI

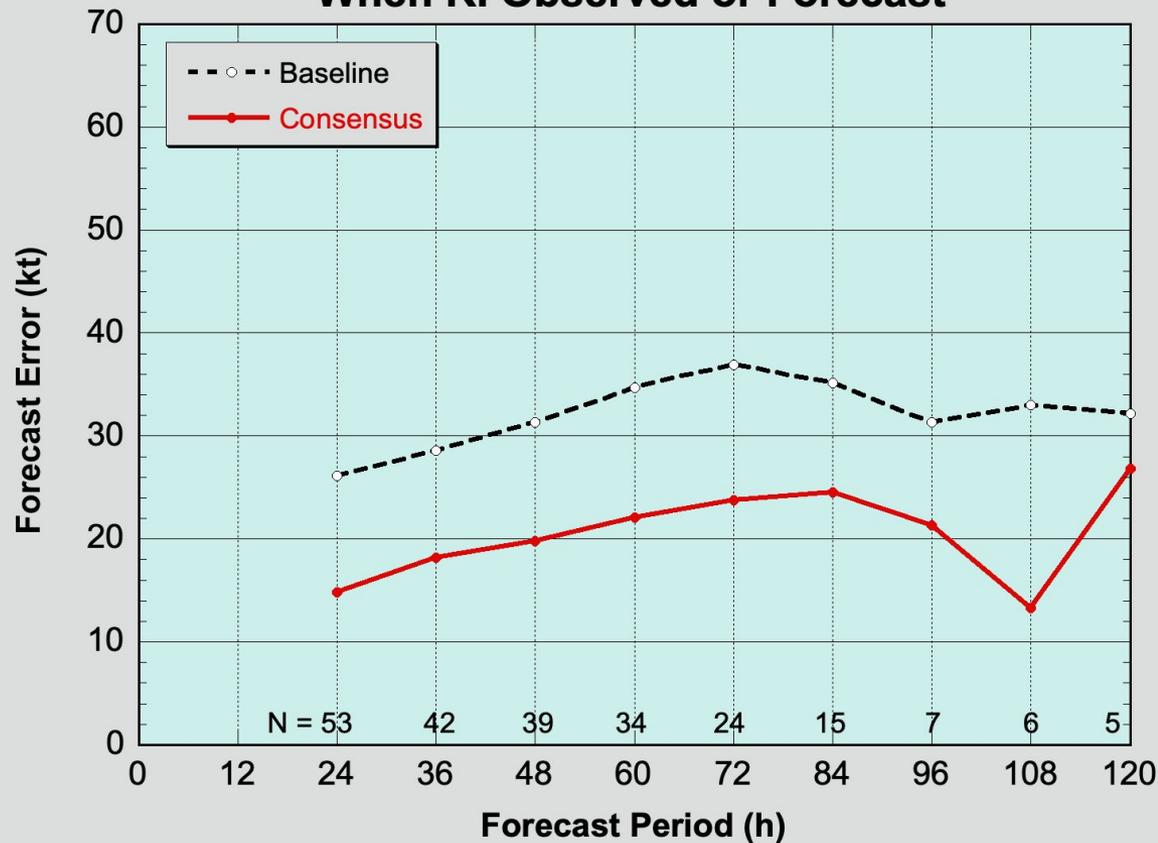
# RI Intensity Baseline



Vt (h)	Baseline	Target
24	26.1	13.1
36	28.6	14.3
48	31.4	15.7
72	36.9	18.5
96	31.3	15.6
120	32.1	16.1

# 2021 Verification: Consensus

**2021 Preliminary AL/EP Errors: Consensus  
(DSHP, LGEM, HWFI, CTCI, HMNI)  
When RI Observed or Forecast**

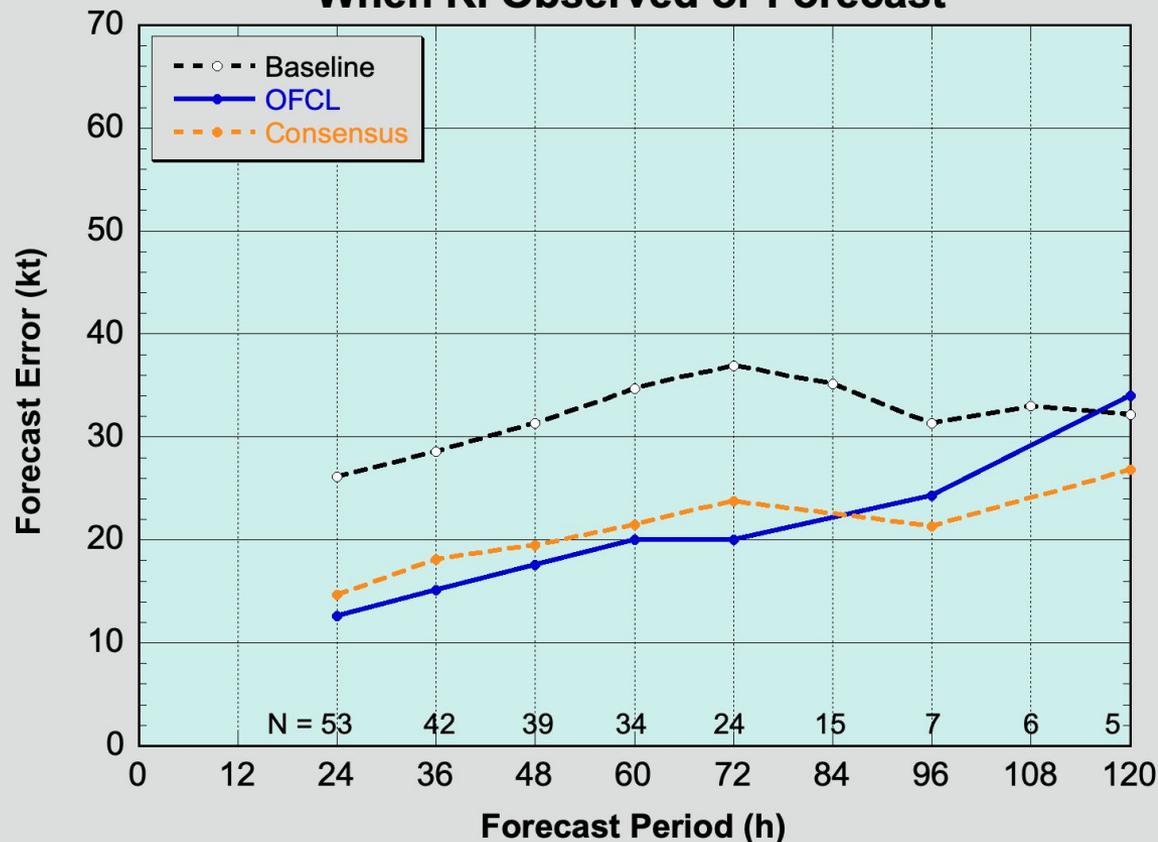


Verifications that follow are through AL Tropical Storm Victor and EP Hurricane Pamela.

Consensus errors through 72 h were well below the HFIP baseline in 2021. Very few cases at 96-120 h, though.

# 2021 Verification: OFCL

**2021 Preliminary AL/EP Errors: OFCL v Consensus (DSHP, LGEM, HWFI, CTCI, HMNI) When RI Observed or Forecast**



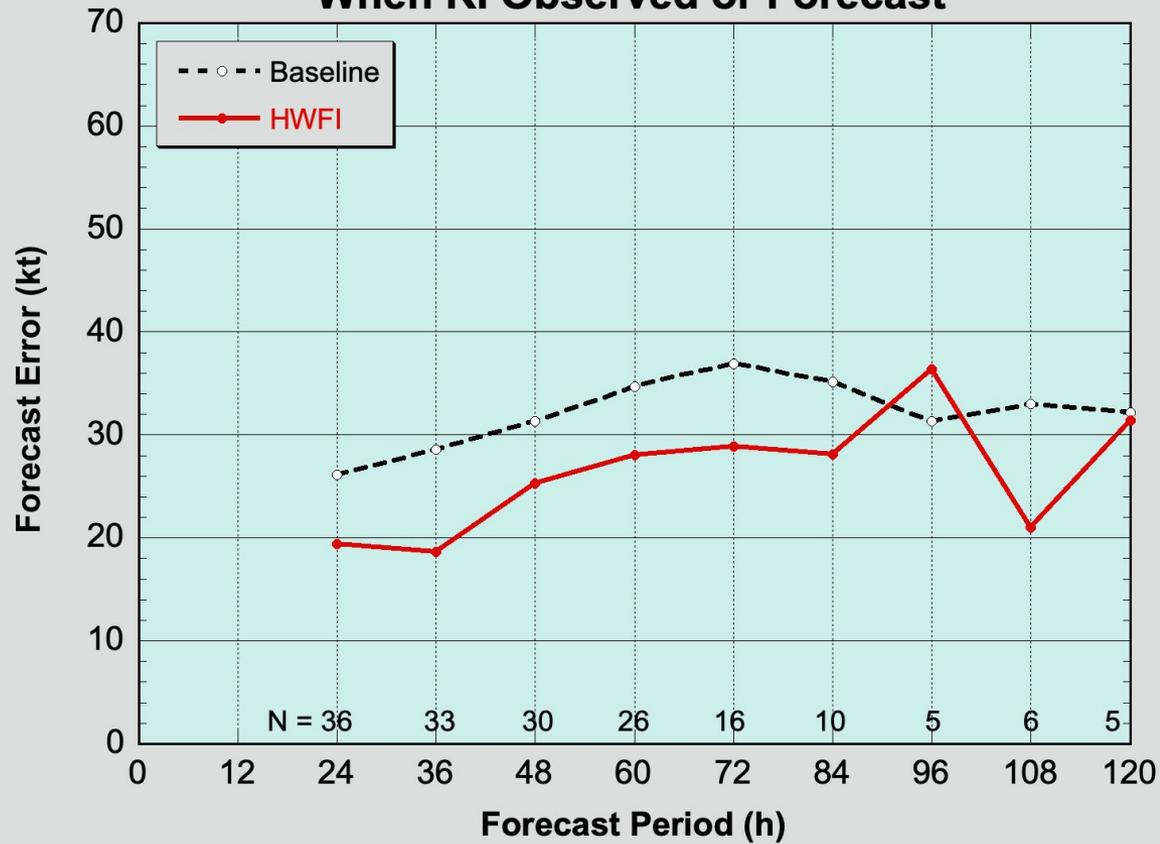
OFCL outperformed the consensus through 72 h when sample sizes were decent.

How did the forecasters beat the deterministic guidance?

Perhaps the probabilistic guidance played a role...

# 2021 Verification: HWFI

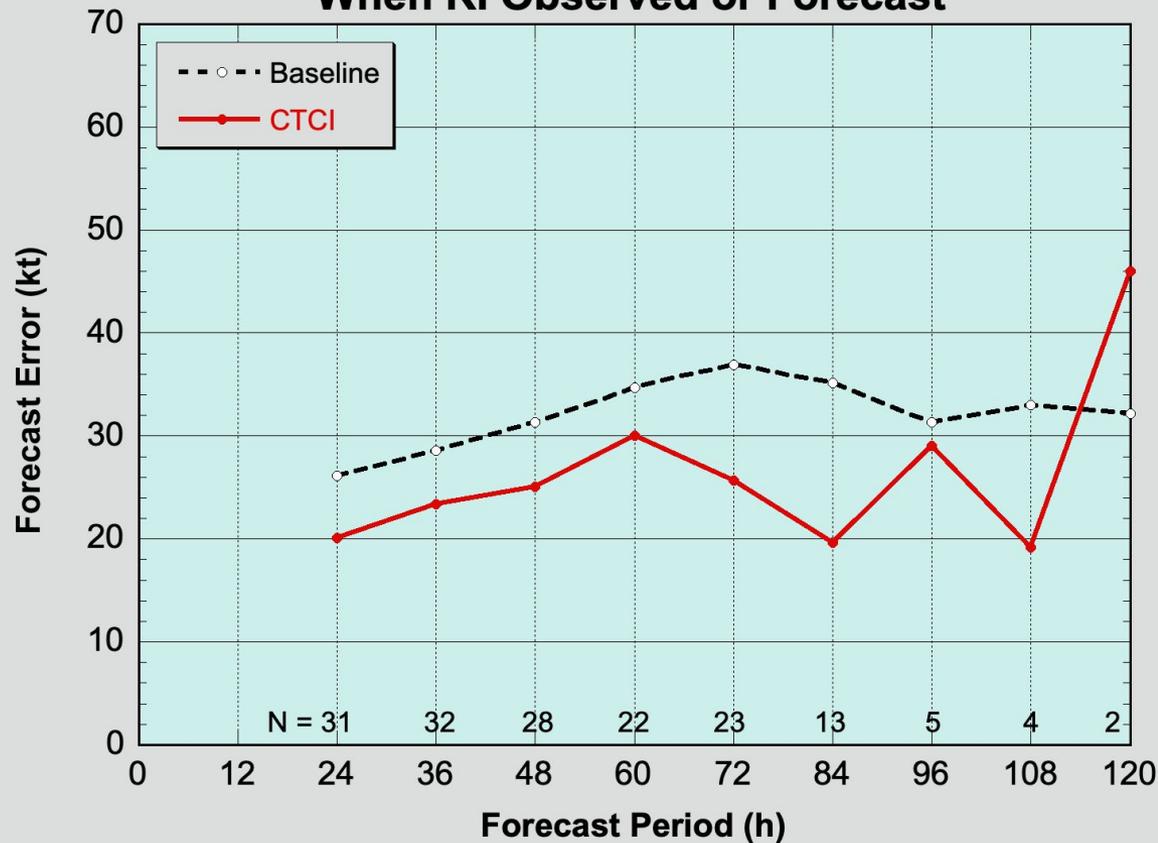
2021 Preliminary AL/EP Errors: HWFI  
When RI Observed or Forecast



HWFI did very well through 72 h, but did not improve on the baseline after ~72 h.

# 2021 Verification: CTCI

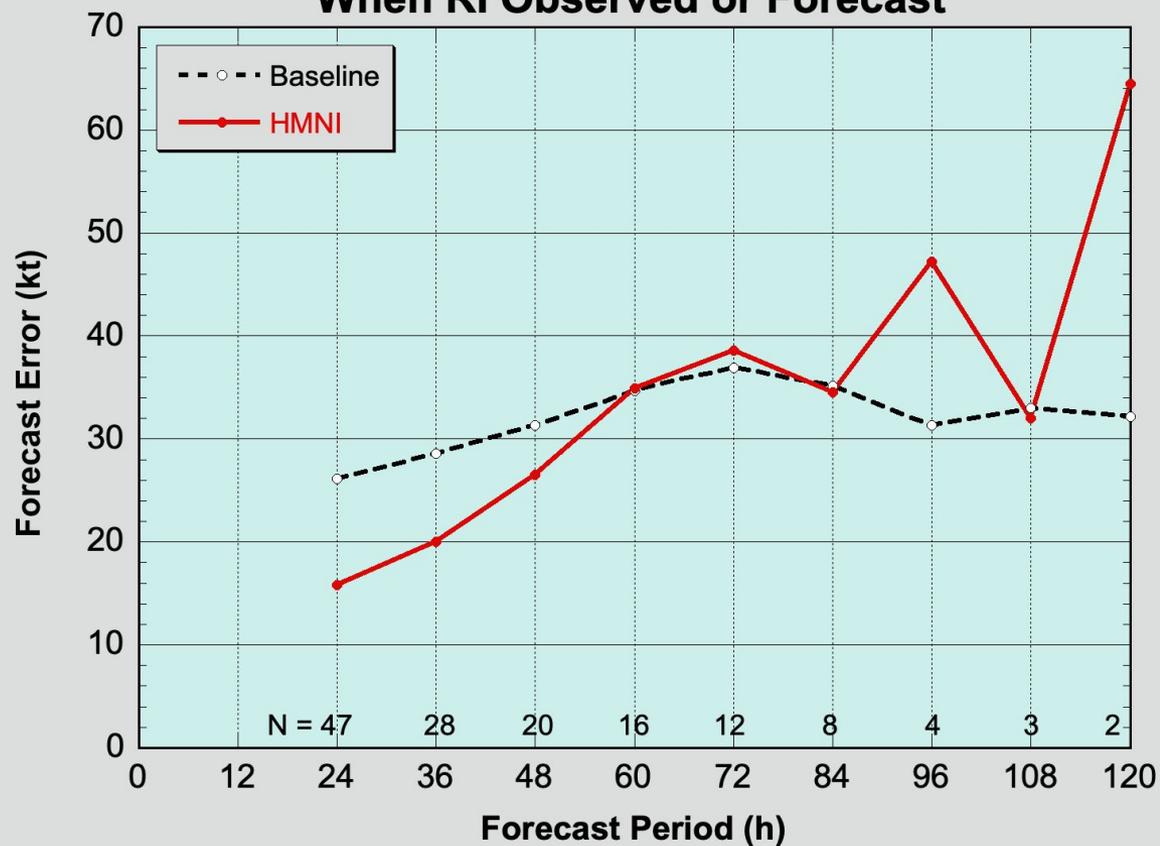
**2021 Preliminary AL/EP Errors: CTCI  
When RI Observed or Forecast**



CTCI performance beat the baseline and showed improvement over 2020.

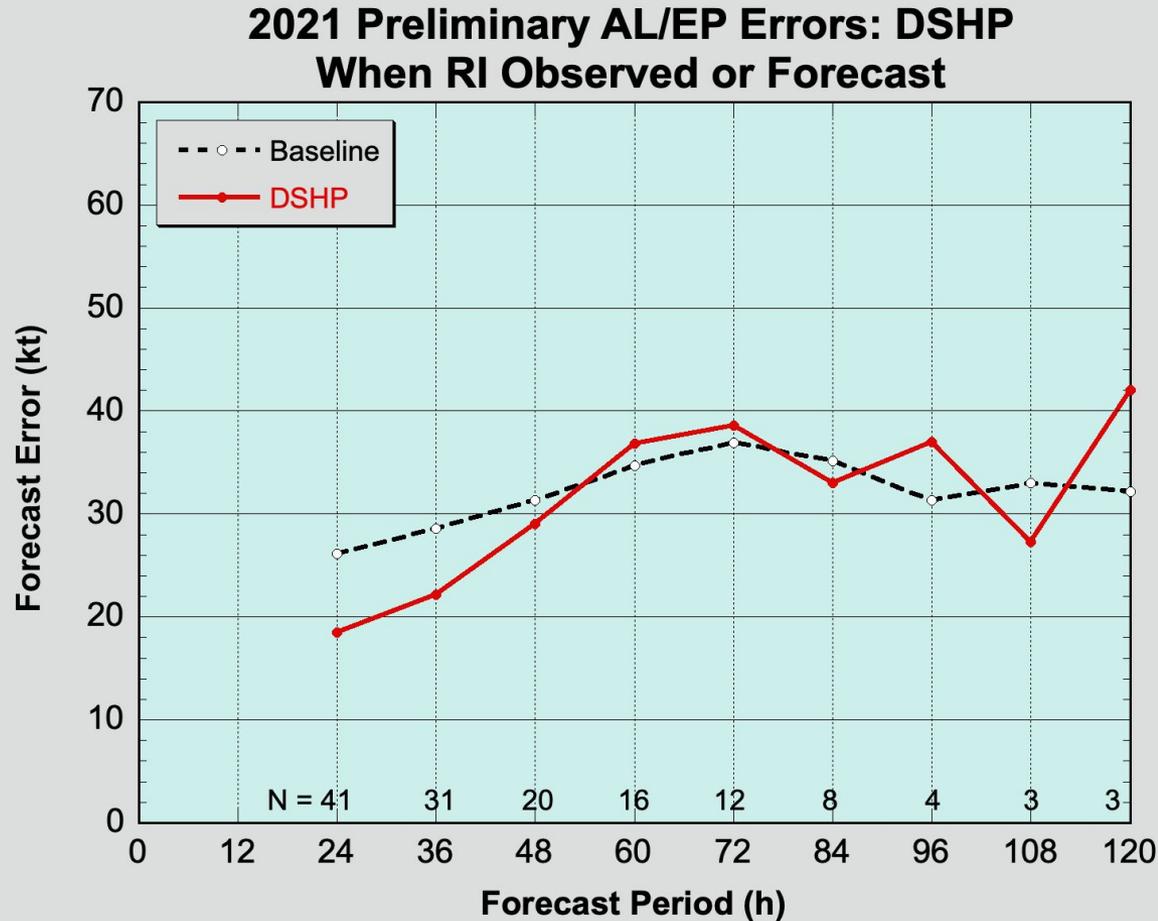
# 2021 Verification: HMNI

**2021 Preliminary AL/EP Errors: HMNI  
When RI Observed or Forecast**



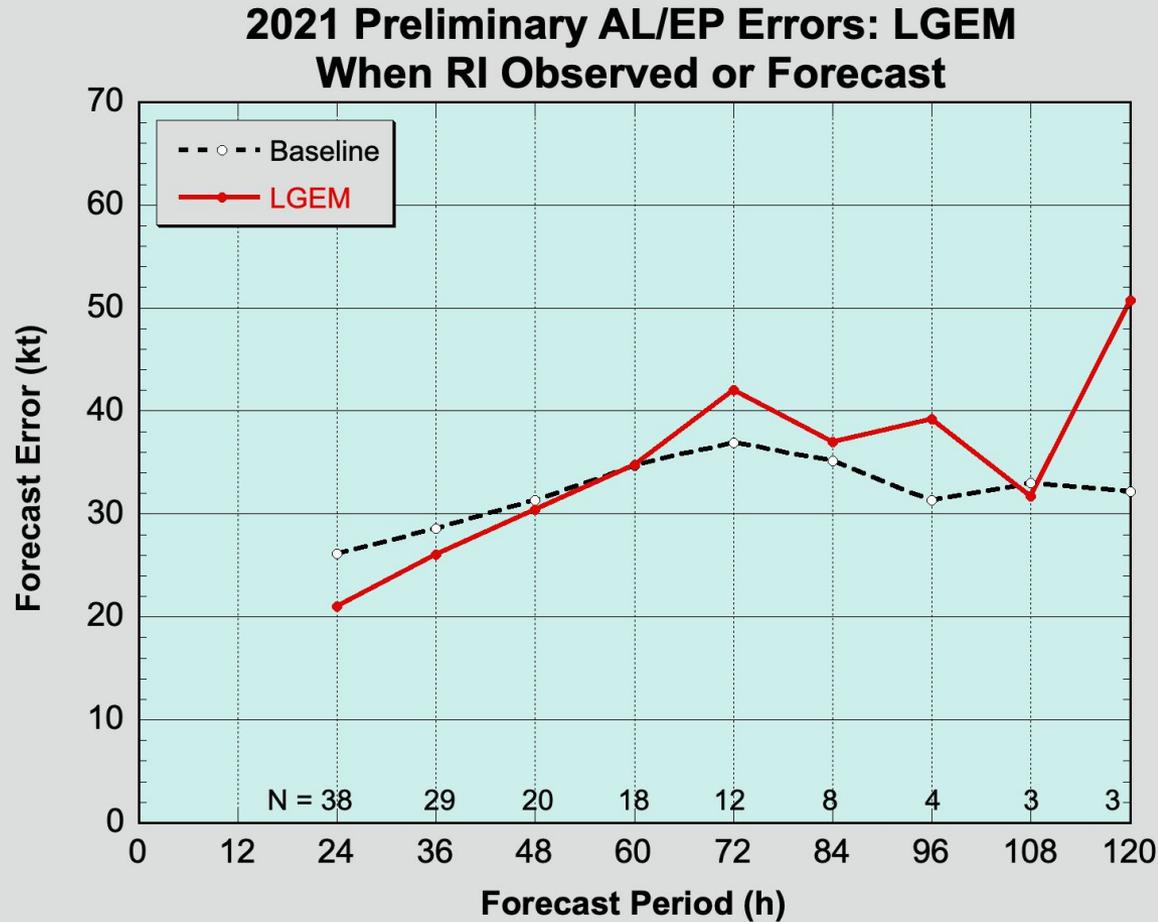
HMNI not consistently better than the baseline (and it wasn't in 2020 either).

# 2021 Verification: DSHP



After a surprisingly good 2020, behavior was about what one would expect in 2021.

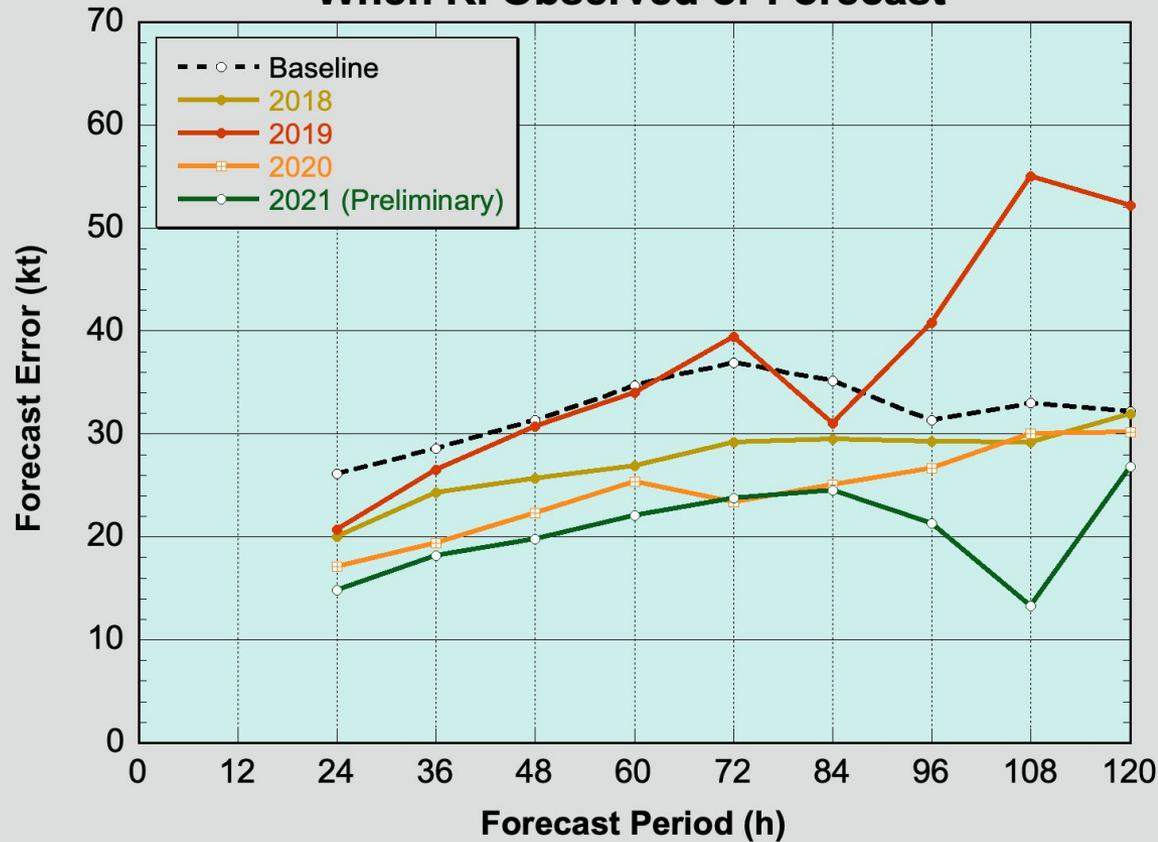
# 2021 Verification: LGEM



Not as bad as last year, but, unsurprisingly, still not contributing much.

# 2018-21 Verification: Consensus

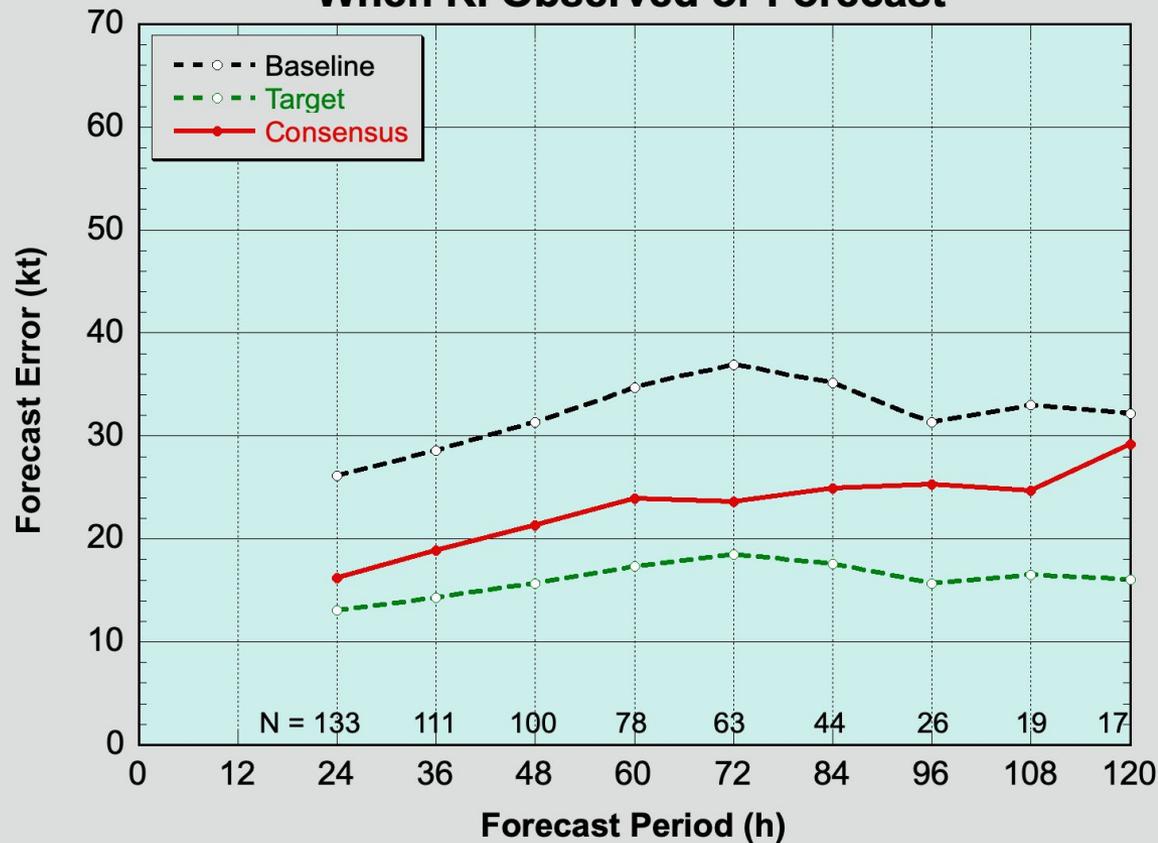
**AL/EP Errors: Consensus  
(DSHP, LGEM, HWFI, CTCI, HMNI)  
When RI Observed or Forecast**



Past two years  
have been very  
good!

# 2020-21 Verification: Consensus

**2020-21 Preliminary AL/EP Errors: Consensus  
(DSHP, LGEM, HWFI, CTCI, HMNI)  
When RI Observed or Forecast**

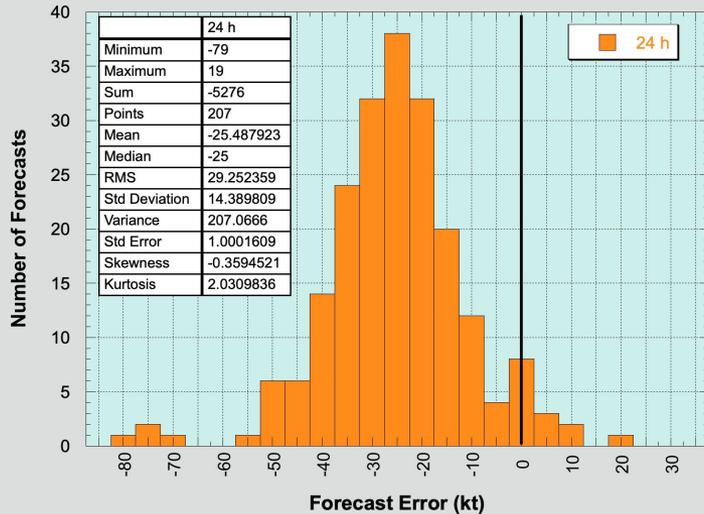


Taking the last two years together, we've almost reached the target RI error at 24 h, and are 2/3 to 3/4 of the way there through 72 h.

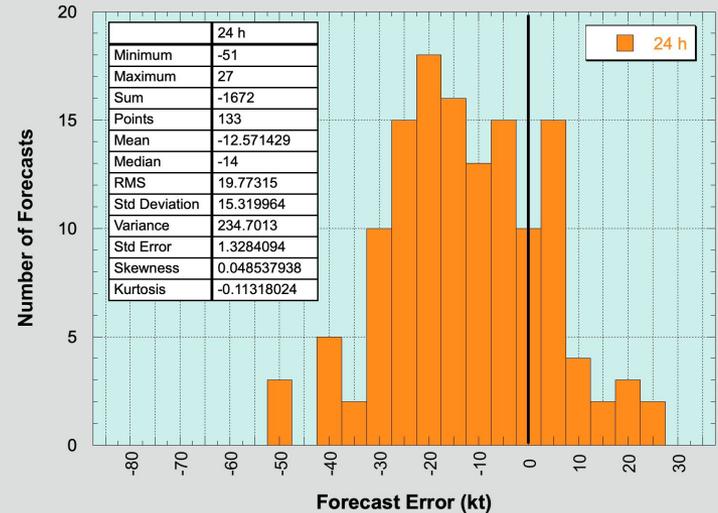
72-hr progress is particularly nice to see, since that's far enough out to make a difference to EMs and other decision makers.

# Consensus Error Distribution For RI Cases

**HFIP RI Baseline Consensus  
2015-17 AL/EP Combined  
When RI Observed or Forecast**



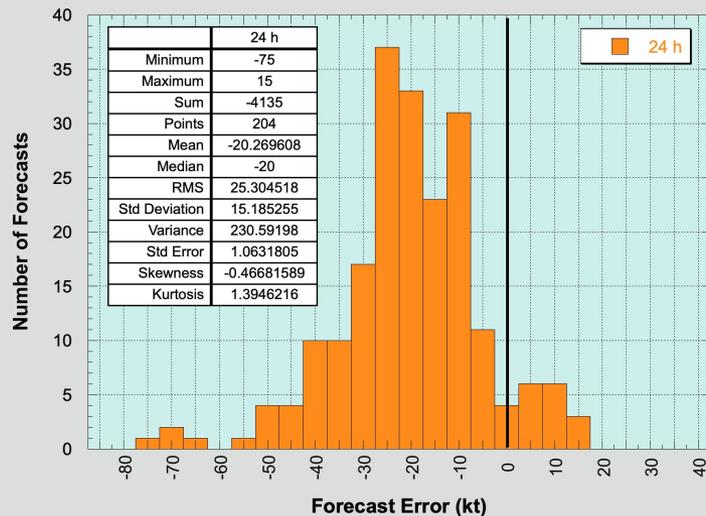
**IVCN Consensus  
2020-21 AL/EP Combined  
When RI Observed or Forecast**



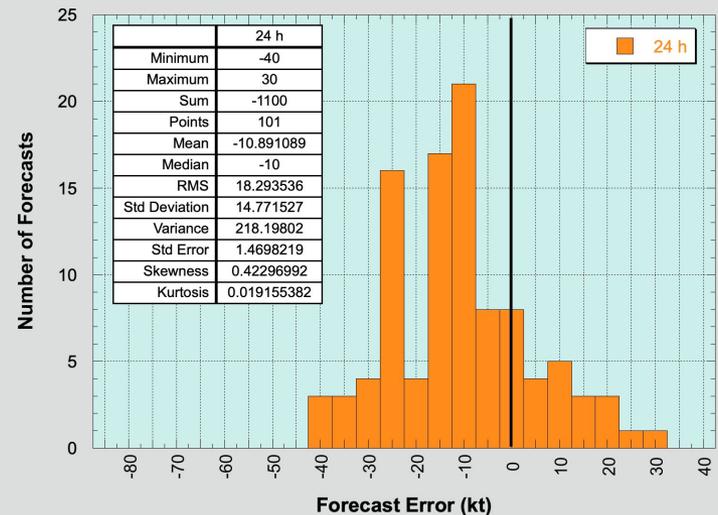
Error distribution of the consensus has shifted dramatically since the baseline was established. Median error has improved from -25 to -14. During the baseline period, IVCN virtually never got RI correct whereas now it's a common occurrence.

# NHC OFCL Error Distribution For RI Cases

**NHC Official Forecast  
2015-17 AL/EP Combined  
When RI Observed or Forecast**



**NHC Official Forecast  
2020-21 AL/EP Combined  
When RI Observed or Forecast**



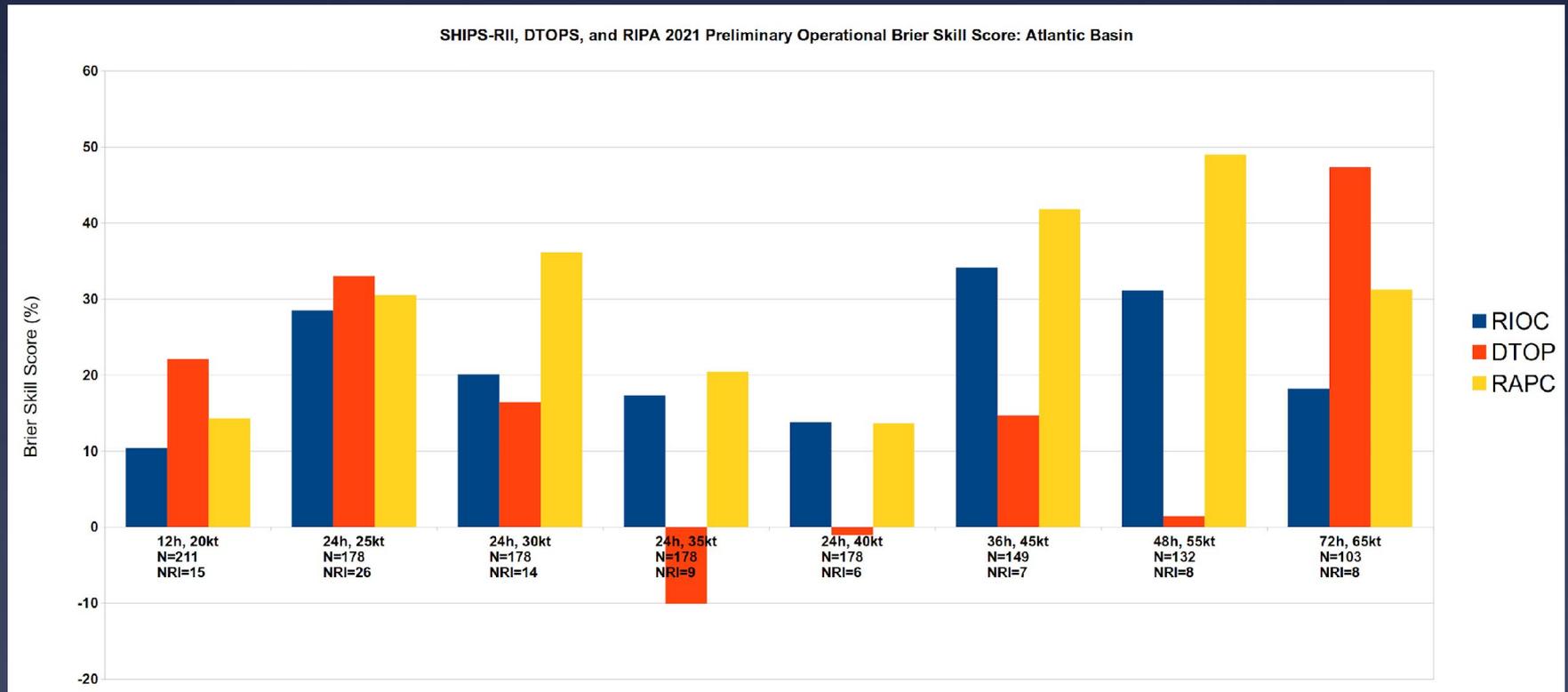
Error distribution of OFCL has shifted dramatically since the baseline period. Median error has improved from -20 to -10. Very large negative errors are gone, and median/mode are now within the margin of error for estimating TC intensity.

Has the RI forecasting problem been largely solved?

# Probabilistic RI Model Definitions

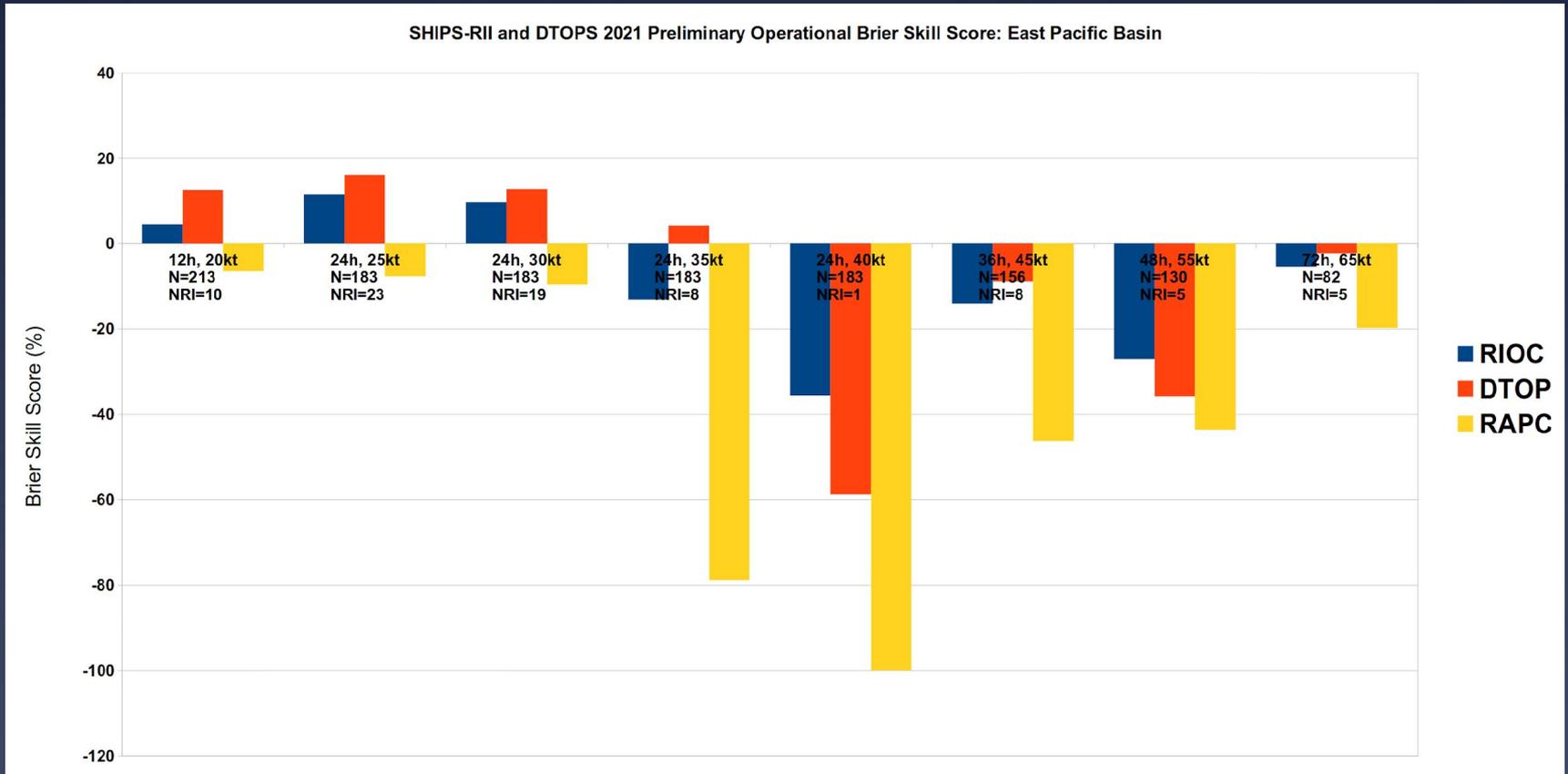
- SHIPS-RII (Kaplan et al. 2015) uses storm environment predictors such as vertical shear, ocean heat content, and parameters from infrared satellite imagery as input to several model formulations:
  1. Bayesian method (RIOB)
  2. Discriminant analysis algorithm (RIOD)
  3. Logistic regression from SHIPS environmental predictors (RIOL)
  4. Consensus probabilities from the three methods above are also produced (RIOC).
- DTOPS (DeMaria et al. 2021): Logistic regression from deterministic model forecasts (HWRF, GFS, ECMWF, SHIPS, LGEM)
- RIPA (Knaff et al. 2020): LDA (linear discriminant analysis), LOGR (logistic regression), RAPC (consensus of LDA and LOGR)
- Probabilistic RI guidance was skillful in 2020.

# 2021 RI Verification: Probabilistic Models



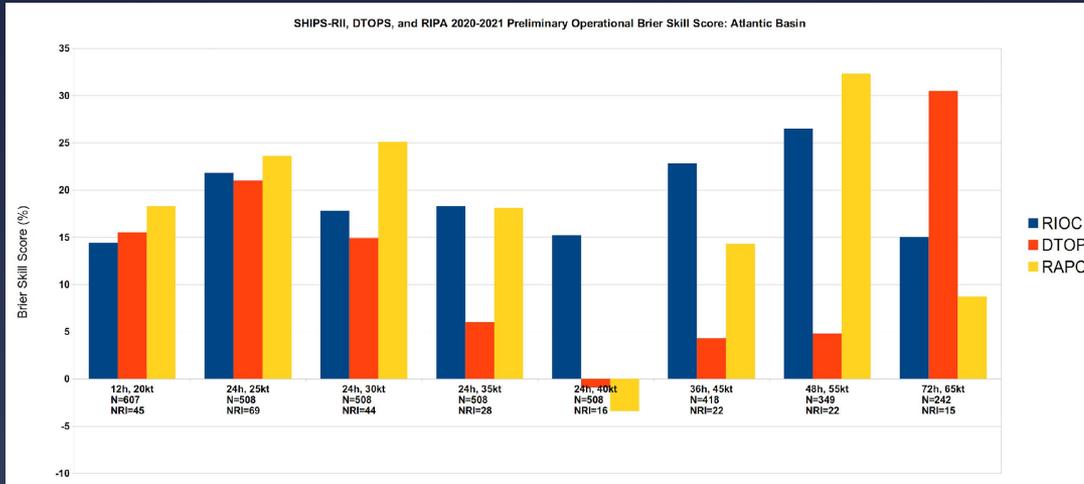
Probabilistic models were skillful again in the Atlantic, although DTOPS performance was inconsistent across the various RI thresholds/intervals.

# 2021 RI Verification: Probabilistic Models

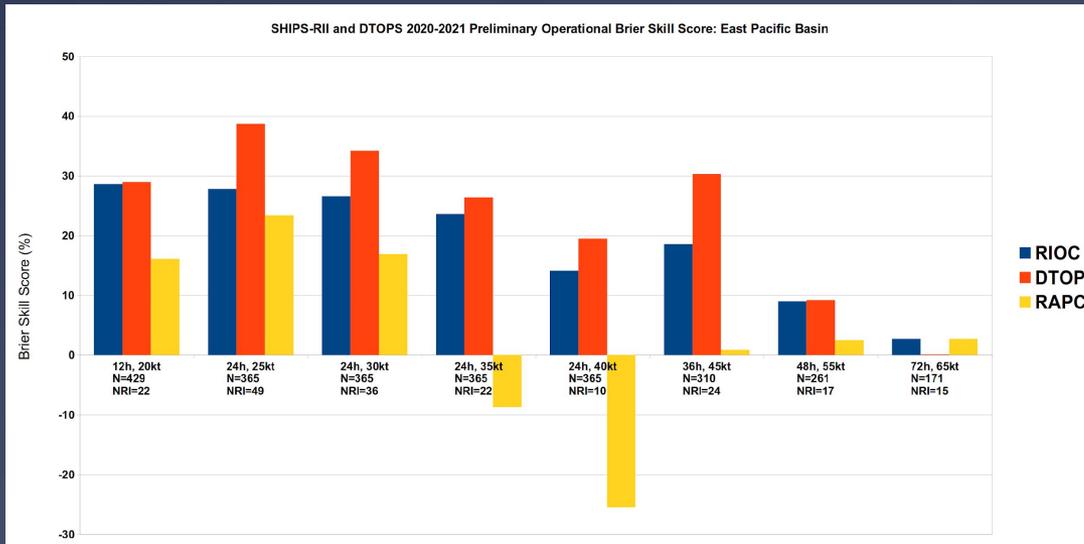


Rough year for the probabilistic models in the eastern Pacific.

# 2020-21 RI Verification: Probabilistic Models

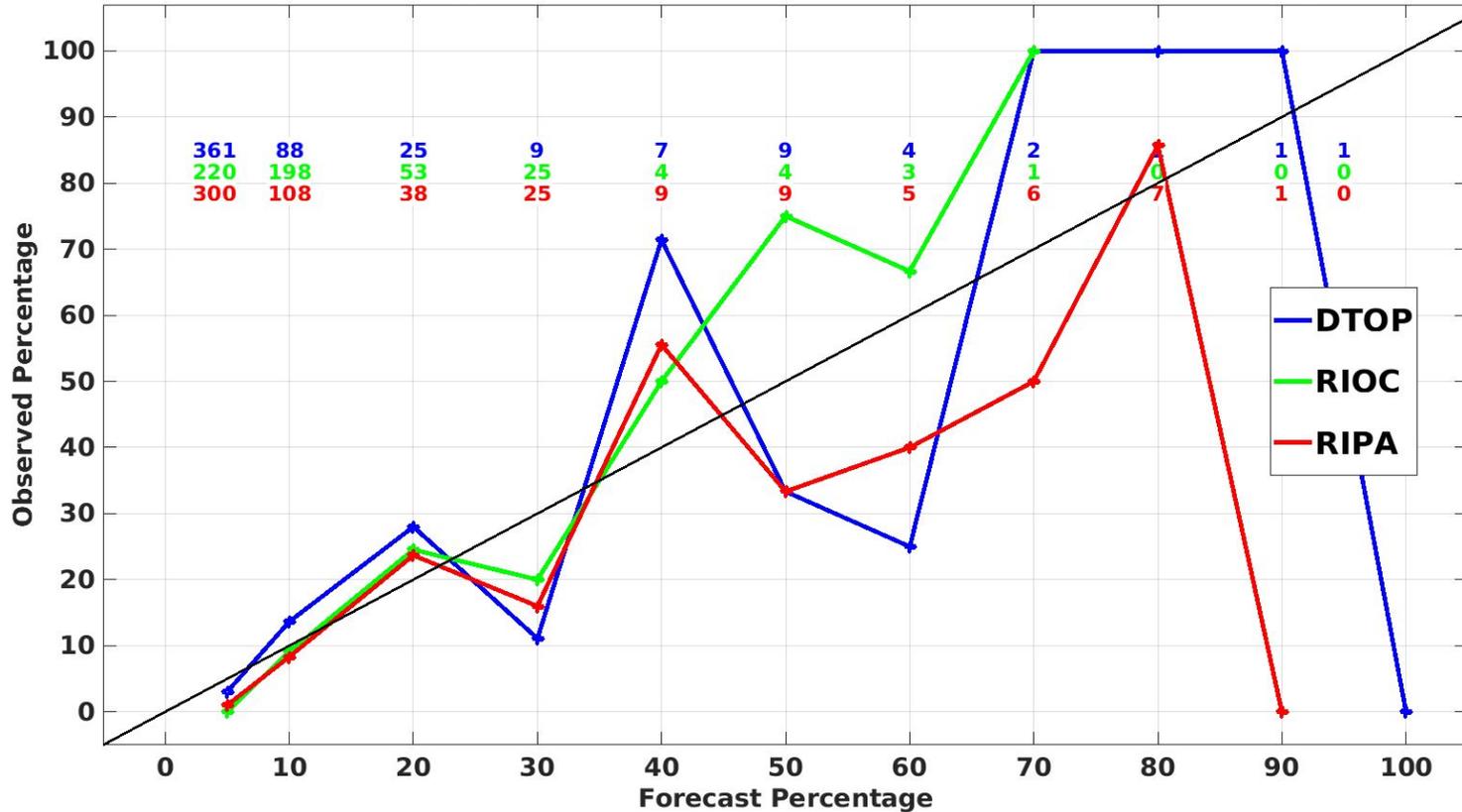


Two-year sample looks a little more promising.



# 2020-21 RI Verification: Probabilistic Models

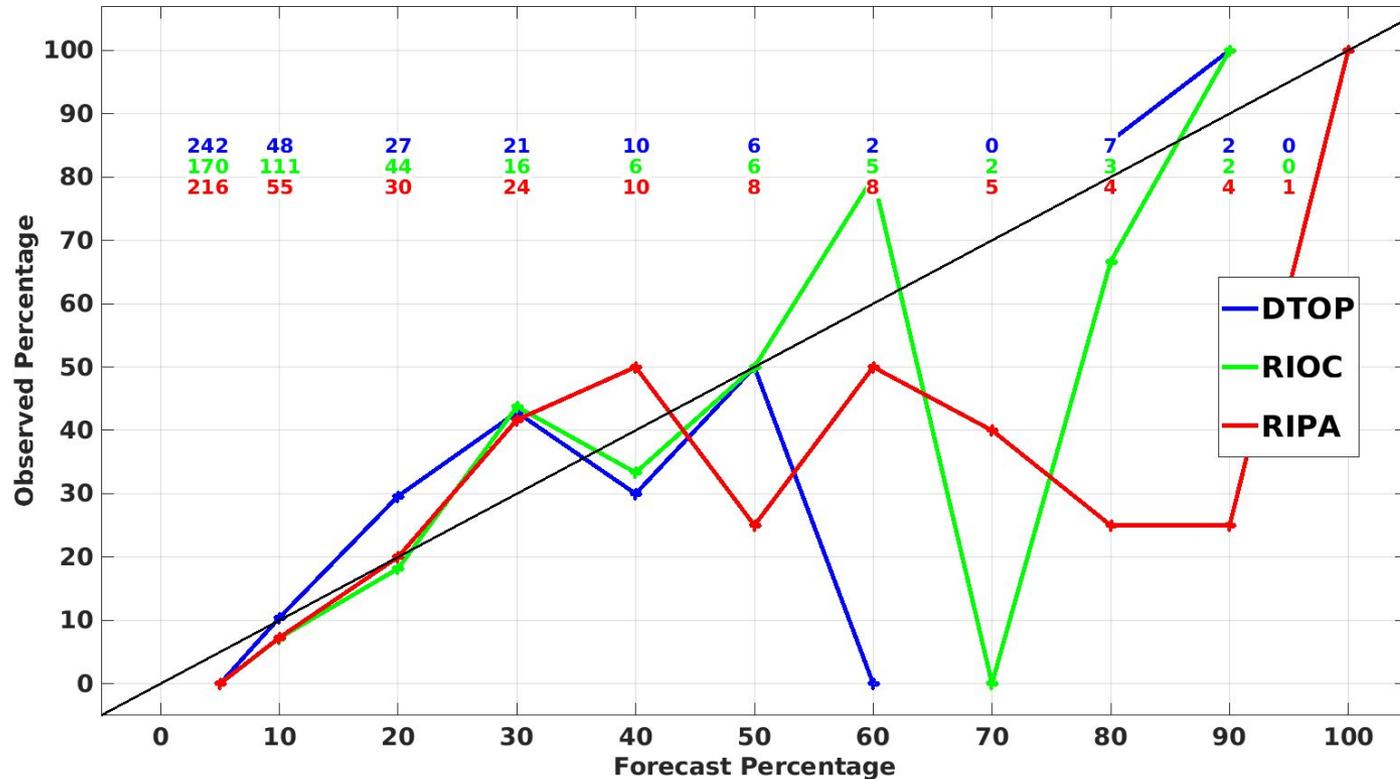
2020-2021 Atlantic Homogeneous SHIPS-RII, DTOPS, and RIPA 30 kt, 24h Reliability



Good reliability, especially for SHIPS-RIOC.

# 2020-21 RI Verification: Probabilistic Models

2020-2021 East Pac Homogeneous SHIPS-RII, DTOPS, and RIPA 30 kt, 24h Reliability



Eastern Pacific seems to be harder once you get to the mid-range percentages.

# Conclusions

- We are making excellent progress toward meeting the HFIP RI performance targets. Over the past two seasons, RI metric errors were 2/3 to 3/4 of the way towards meeting the targets through 72 h.
- NHC forecasters beat the deterministic guidance again in 2021 (through 72 h).
- Probabilistic guidance in 2021 was not as skillful or consistent as in 2020. SHIPS-RIOC seems to have good reliability.