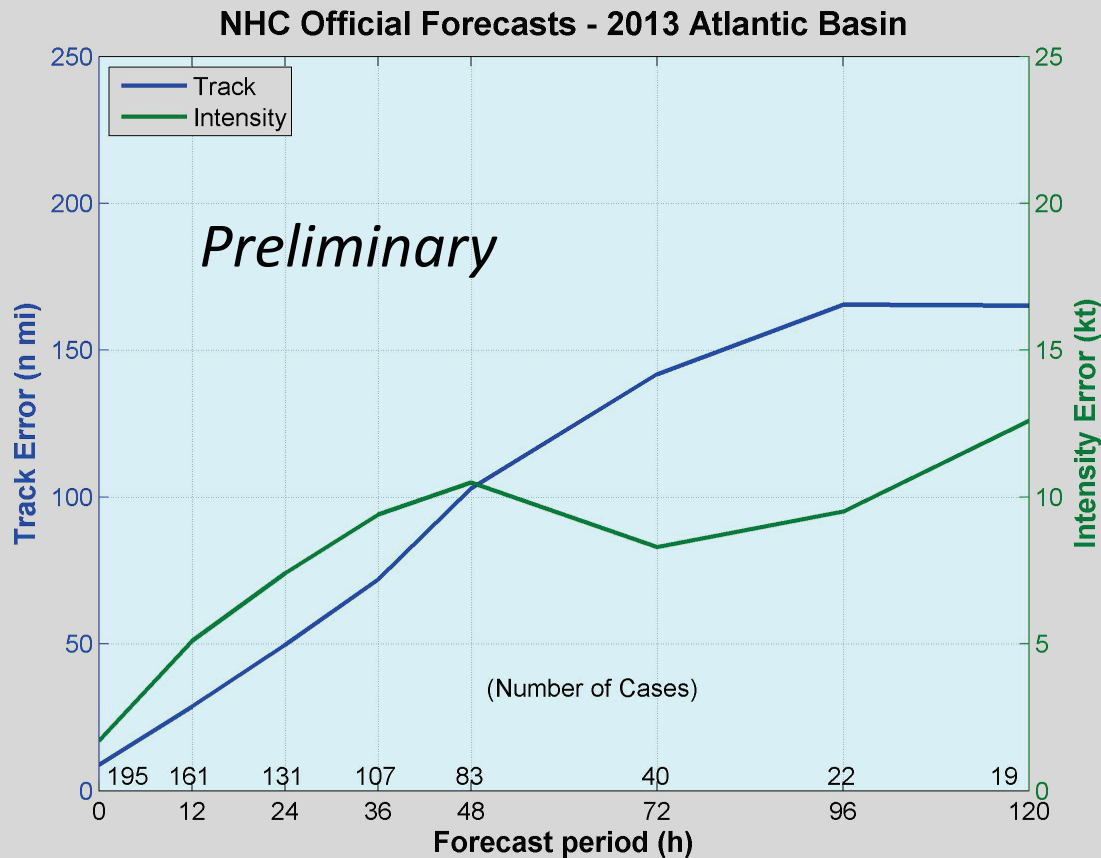


National Hurricane Center 2013 Forecast Verification Preliminary

John Cangialosi and James Franklin
Hurricane Specialist Unit
National Hurricane Center



2013 Atlantic Verification



VT (h)	NT	TRACK (n mi)	INT (kt)
000	195	8.7	1.7
012	161	28.6	5.1
024	131	49.6	7.4
036	107	71.9	9.4
048	83	102.9	10.5
072	40	141.8	8.3
096	22	165.5	9.5
120	19	165.2	12.6

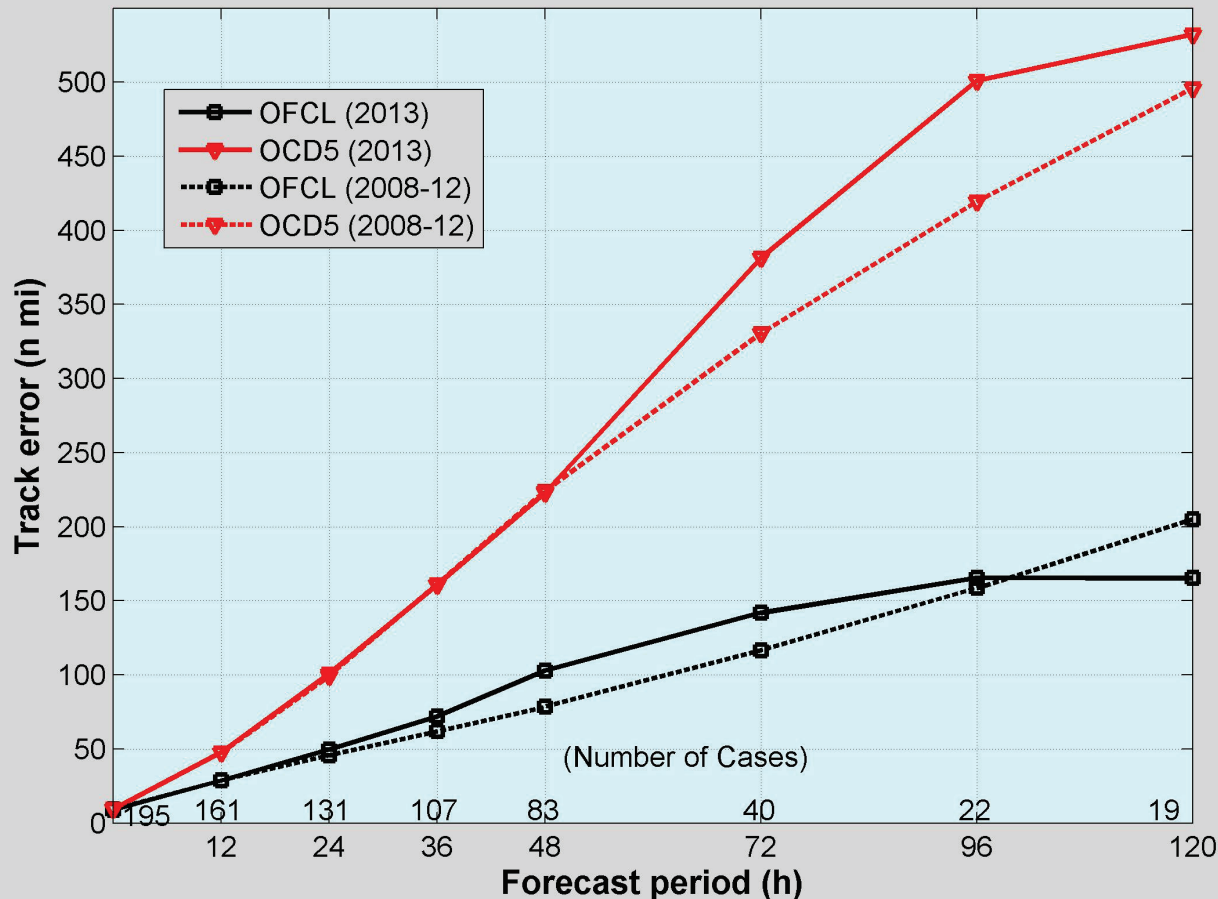
Values in green exceed all-time records.

A lot of intensity error records broken.

Sample at 48 h is 38% of normal and sample at 120 h is only 17% of normal.

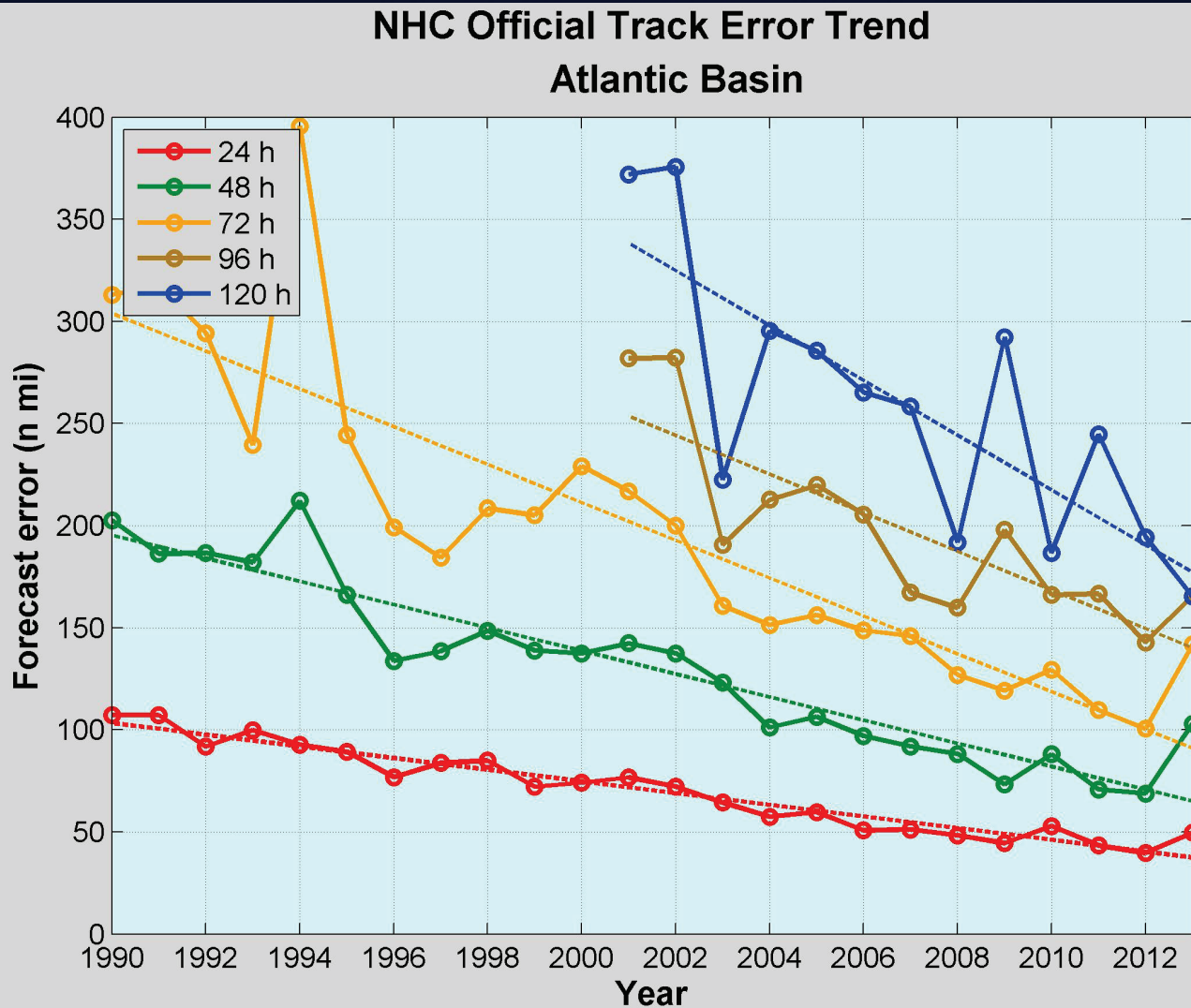
Atlantic Track Errors vs. 5-yr Mean

NHC Official vs. CLIPER5 Forecasts
Atlantic Basin



Official forecast errors were slightly above the 5-yr means, perhaps because the season's storms were a little harder to forecast than normal.

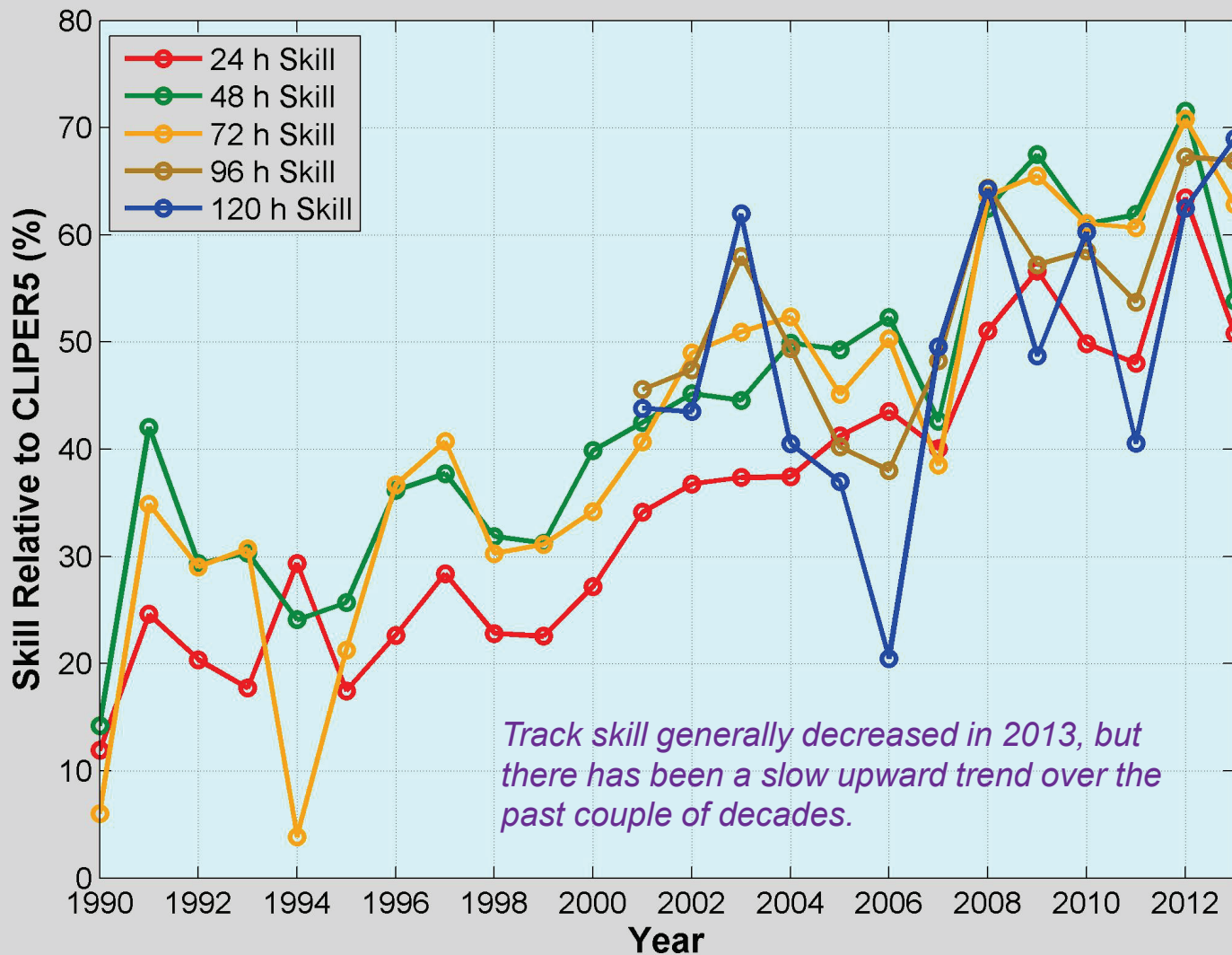
Atlantic Track Error Trends



Track errors increased in 2013 (except at 120 h), but long-term trend shows strong improvements during the past couple of decades.

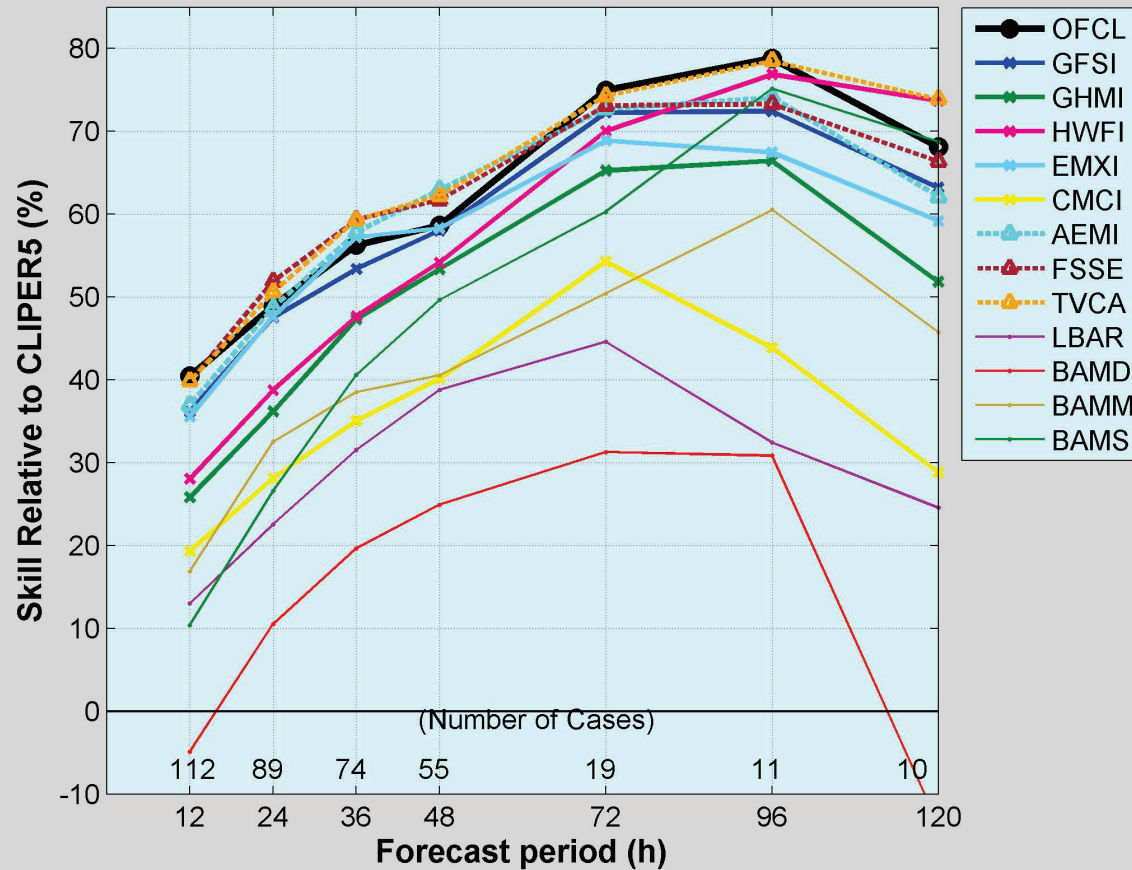
Atlantic Track Skill Trends

NHC Official Track Skill Trend Atlantic Basin



2013 Track Guidance

Track Forecast Skill (Early Models)
2013 - Atlantic Basin



Official forecasts were very skillful, near the best models.

TVCA best model, beat FSSE from 72-120 h

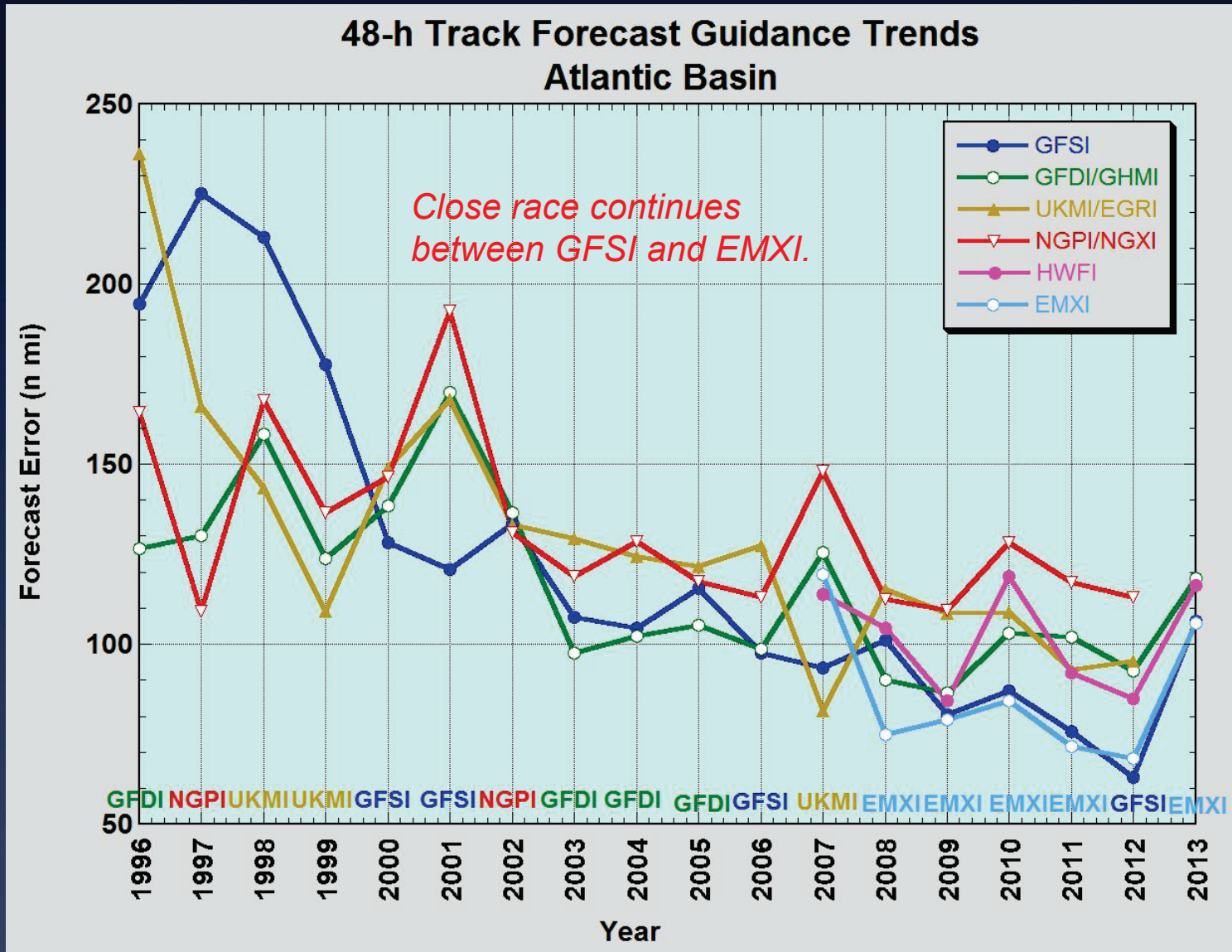
EMXI and GFSI best individual models in short-term, but HWFI was the best model at the longer range (small sample).

GFS ensemble mean excellent performer, better than the GFS.

GHMI middle of the pack, CMCI trailed.

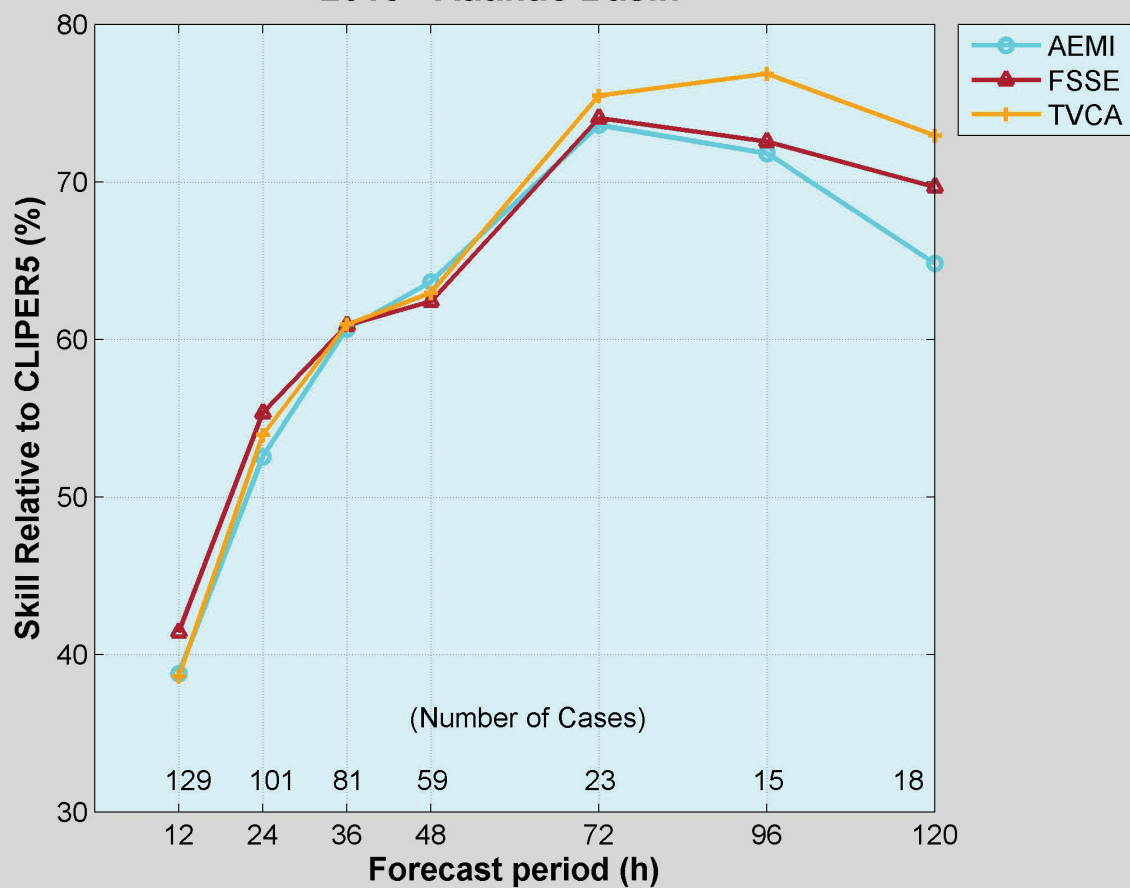
BAMS beat much of the guidance at 96 and 120 h.

Track Model Trends



2013 Consensus Guidance

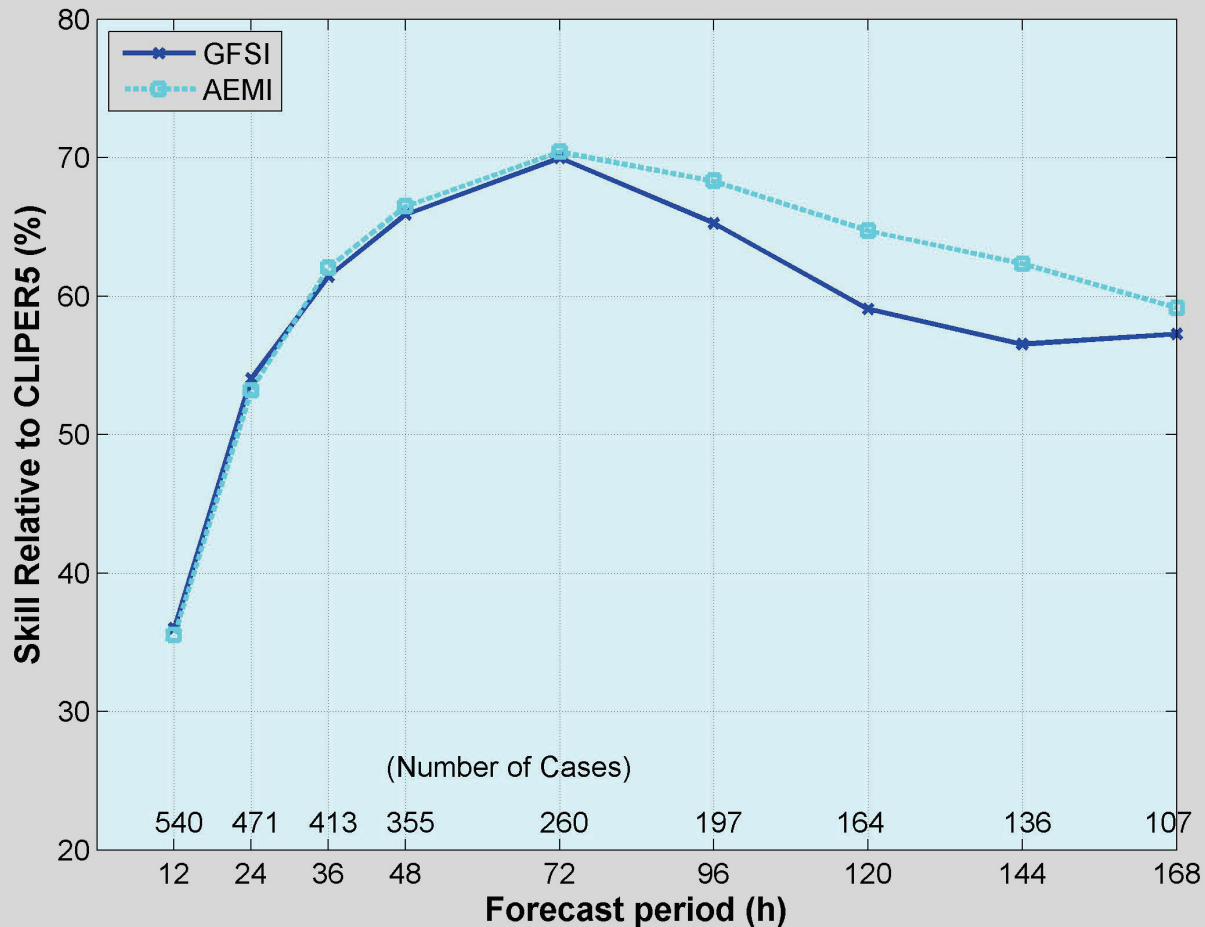
Track Forecast Skill (Consensus Models)
2013 - Atlantic Basin



Consensus models were close to each other through 48 h. After that, TVCA more skillful than FSSE and AEMI.

GFS vs. GFS ensemble mean

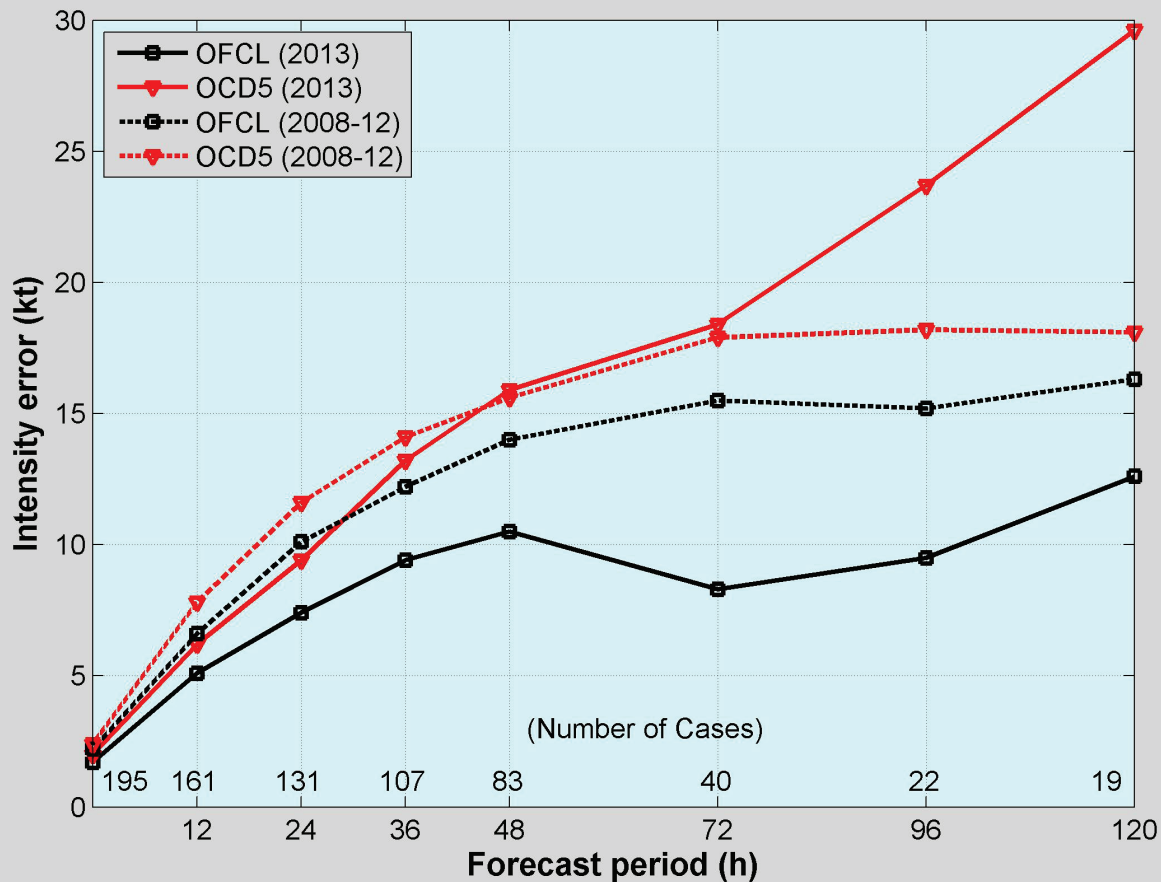
Track Forecast Skill (GFS vs GFS ensemble mean)
2012-13 - Atlantic Basin



- This is a 2-yr sample (2012-13)
- The skill of GFS and AEMI are very similar through 72 h. After that, AEMI is more skillful.

Atlantic Intensity Errors vs. 5-Year Mean

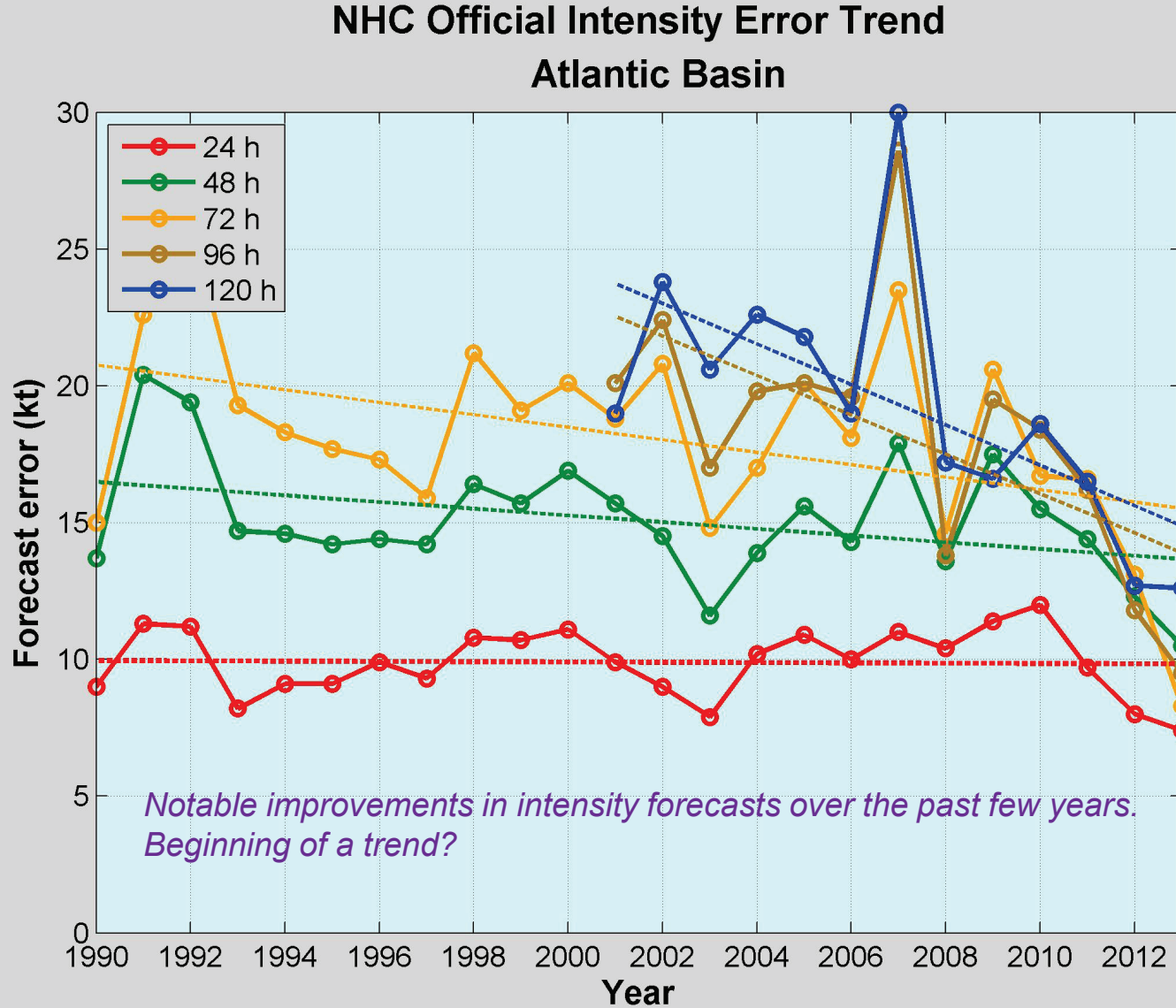
NHC Official vs. Decay-SHIFOR5 Forecasts
Atlantic Basin



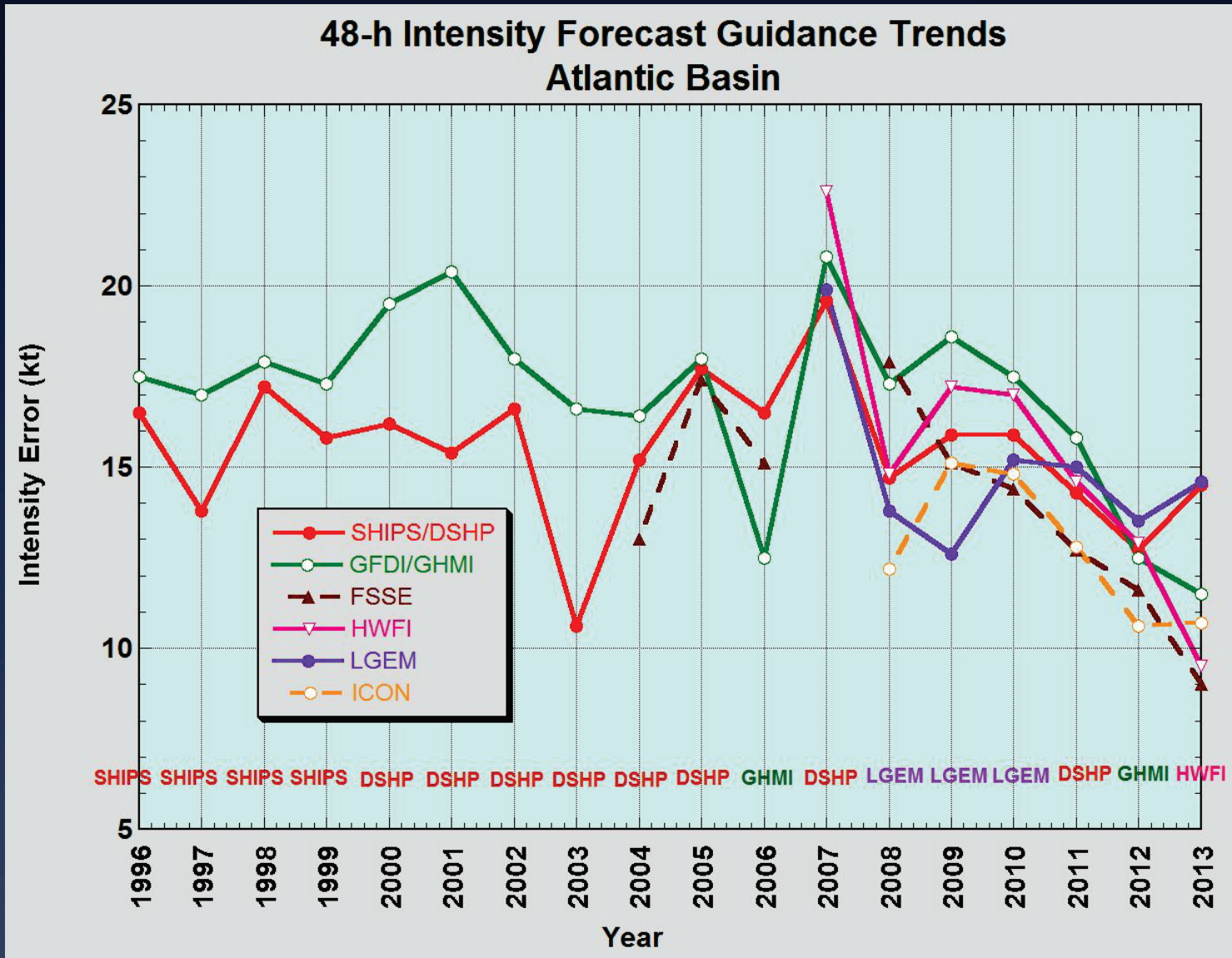
Official forecast errors were much lower than the 5-yr mean. The season's storms were easier than normal to forecast through 36 h; the same or harder after that.

Errors were likely low because the storms were weak, and to some extent we recognized that. However, official forecast error was almost exclusively a high bias.

Atlantic Intensity Error Trends



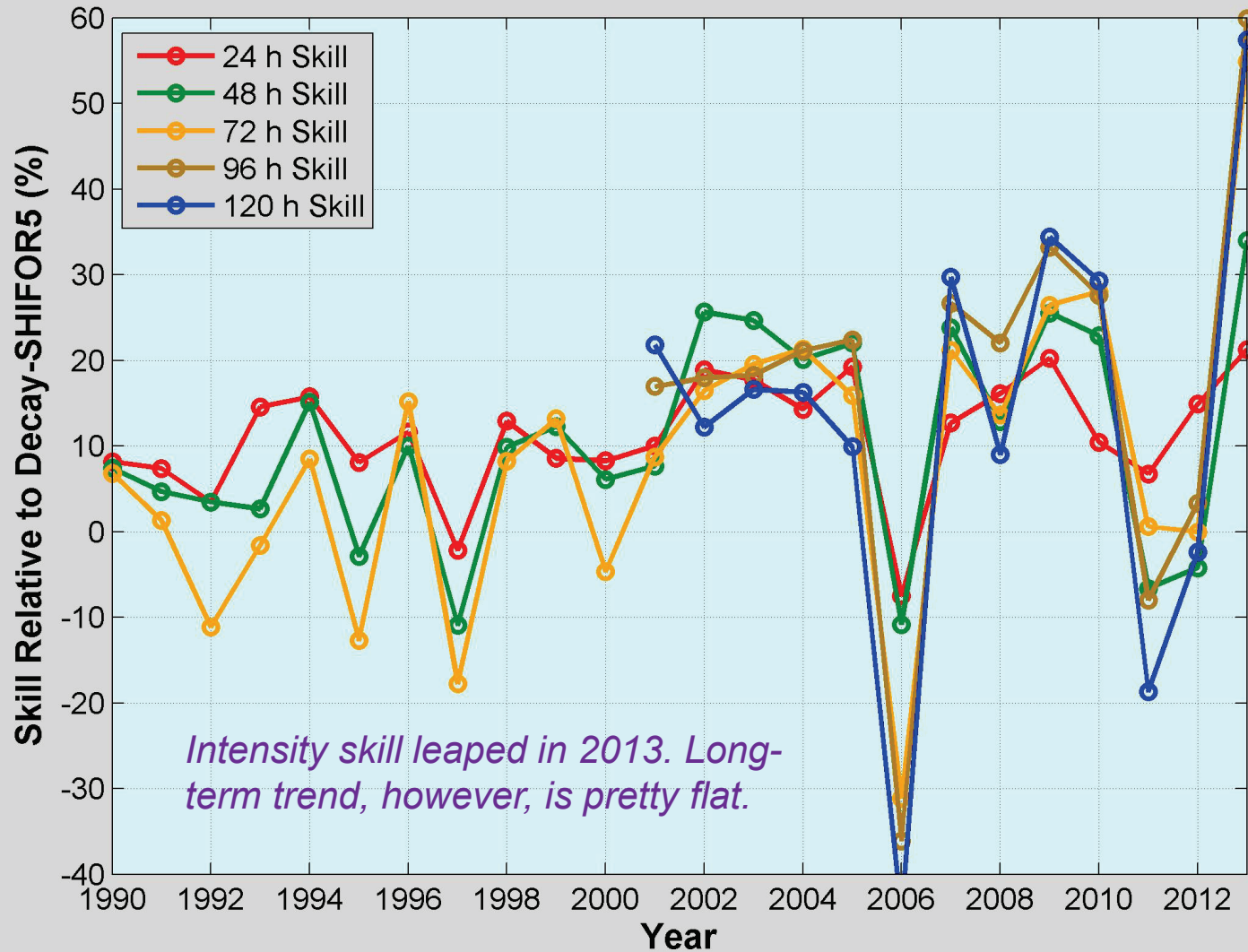
Intensity Model Error Trends



Models have been improving recently, pronounced trend in dynamical and consensus models

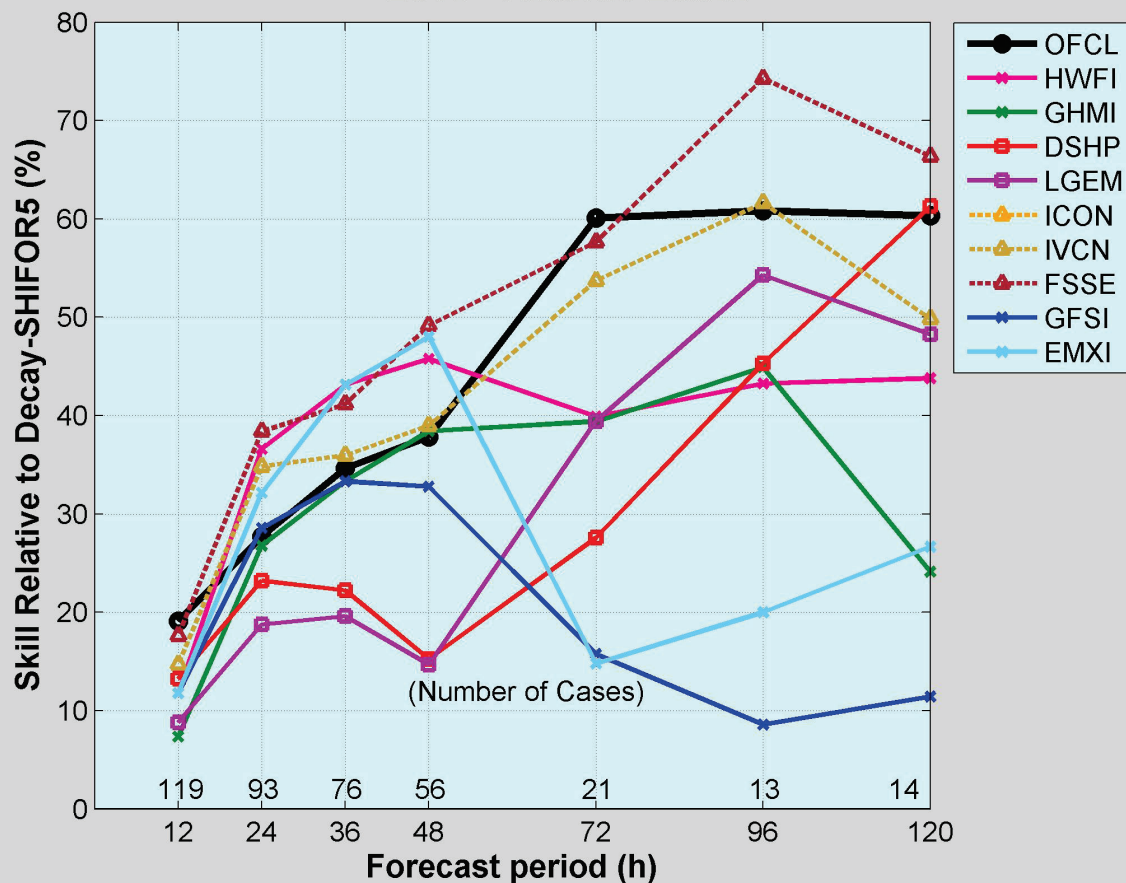
Atlantic Intensity Skill Trends

NHC Official Intensity Skill Trend Atlantic Basin



2013 Intensity Guidance

Intensity Forecast Skill
2013 - Atlantic Basin



Official forecasts most skillful at the longer ranges, and is near the best models, *FSSE* and *ICON/IVCN*.

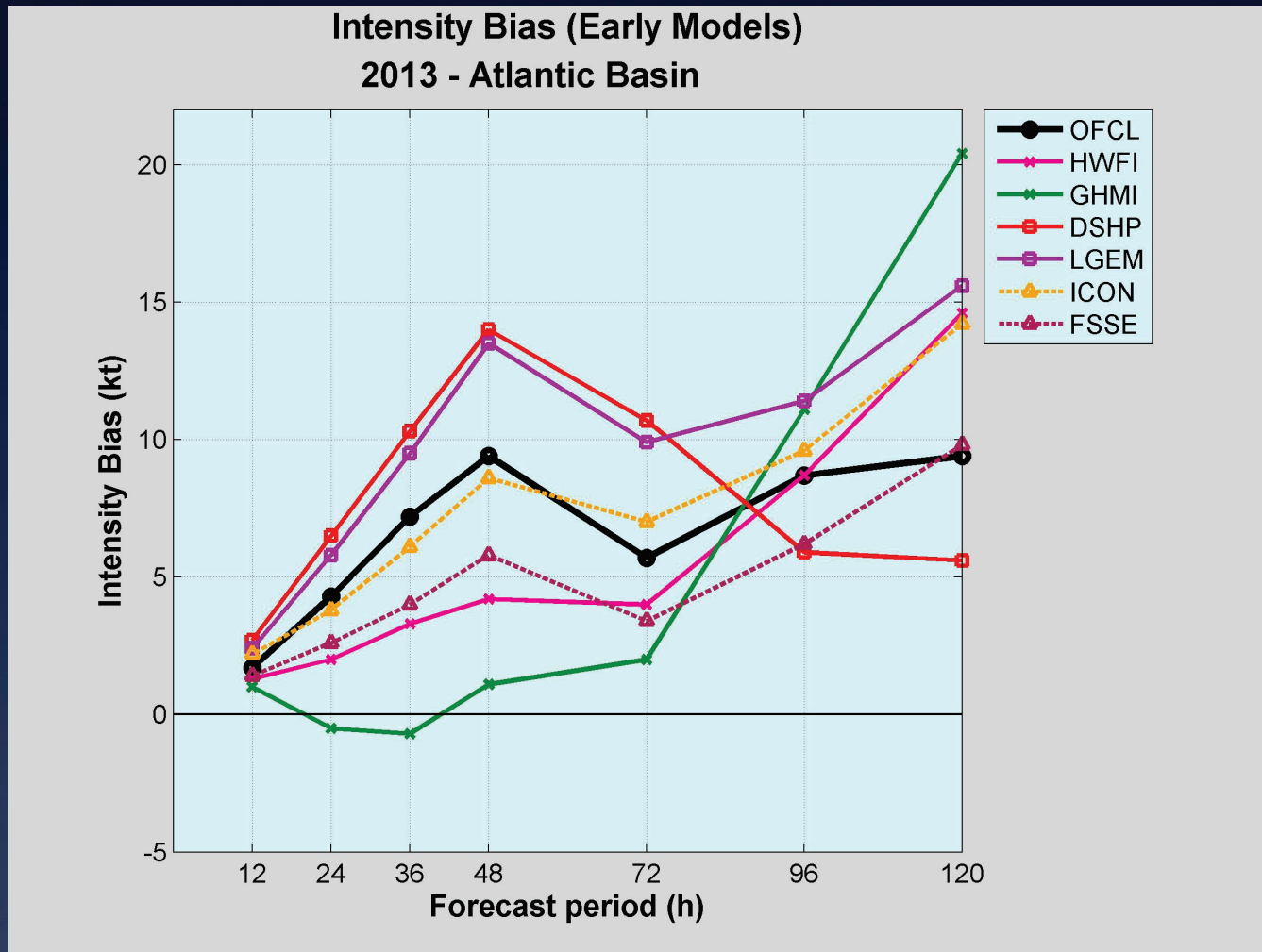
DSHP and *LGEM* were least skillful of the models in short-term, competitive with the others at longer time period.

HWRF fairly good performer, better than the statistical guidance.

GFSI and *EMXI* had notable skill early, decreased late.

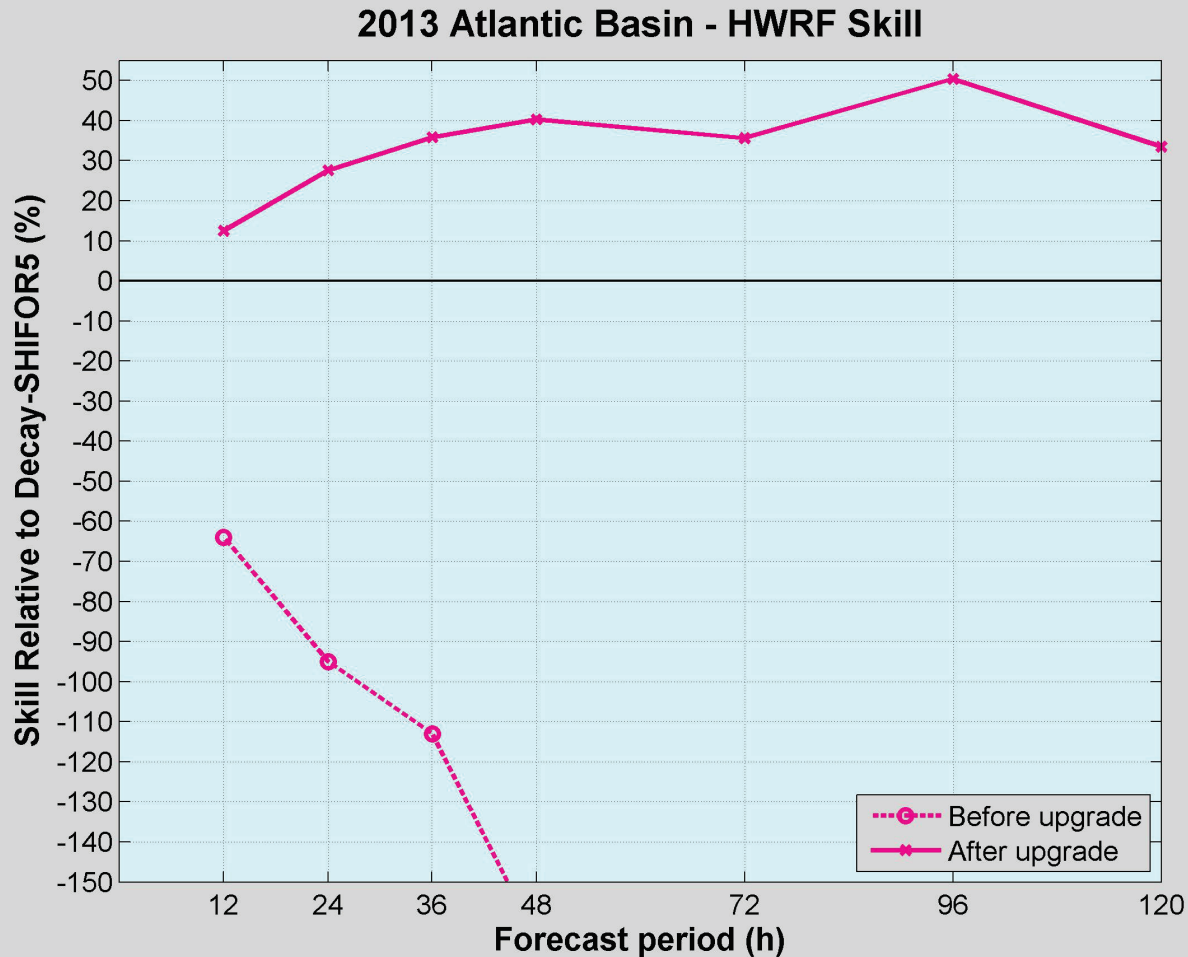
All the guidance had a high bias. *OFCL* had a smaller high bias than *ICON*.

2013 Intensity Bias



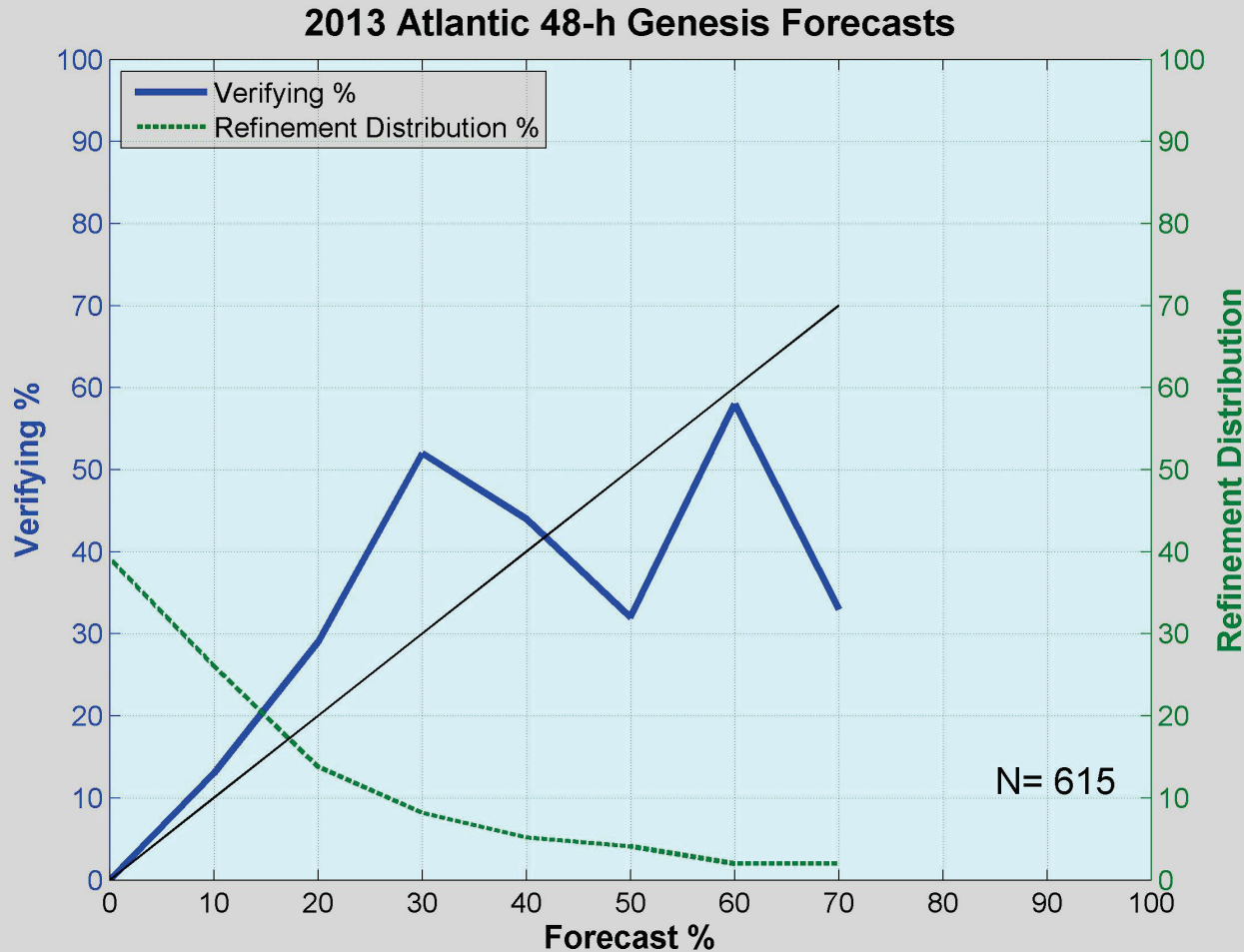
The main intensity models had a high bias in 2013.

HWRF before and after the upgrade



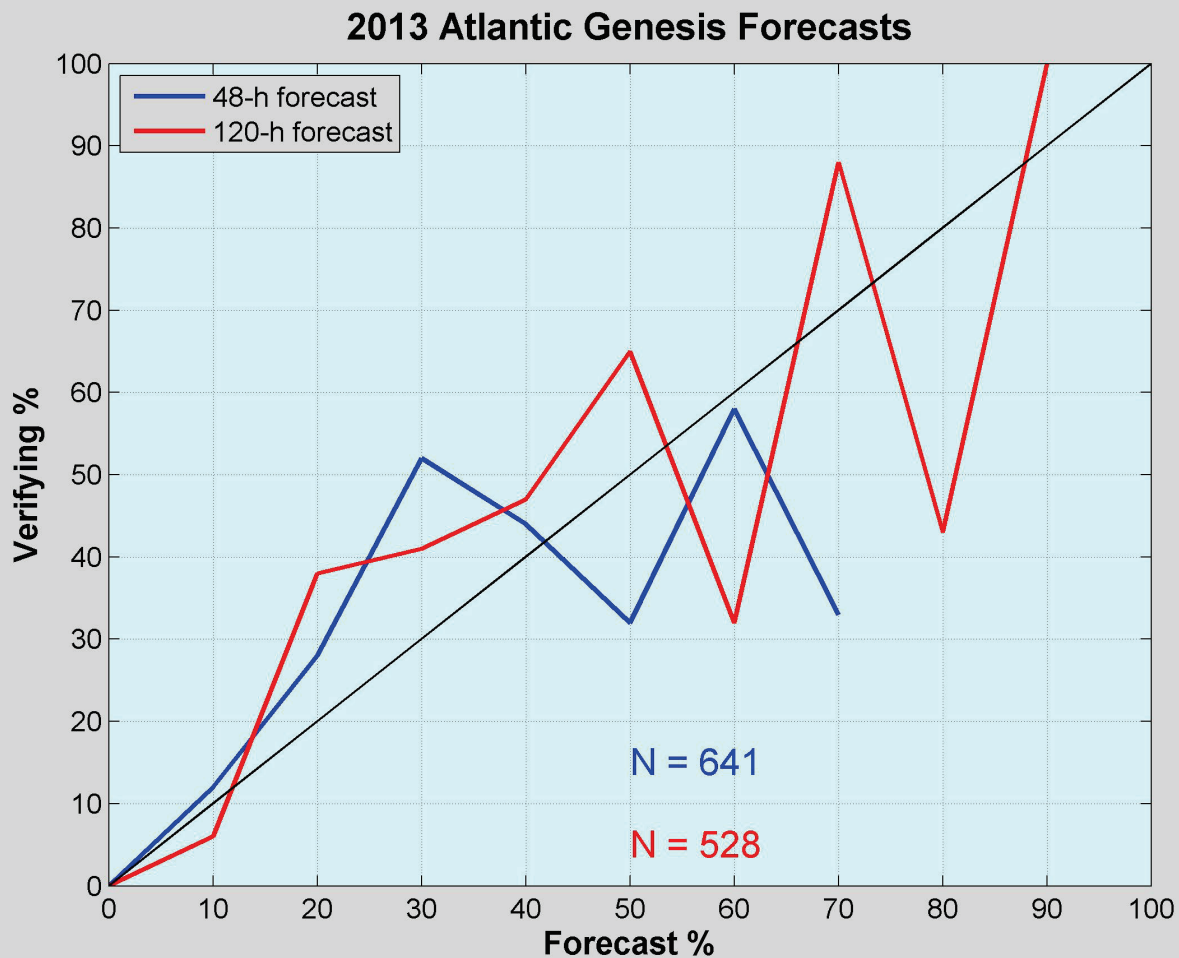
Strongly negative skill before the transition in early July due to fairly large HWRF errors and low Decay-SHIFOR errors. The opposite is true for the larger sample after the transition.

2013 48-h Genesis Forecasts



A slight under-forecast (low) bias was present at the low to medium probabilities. Sample is small at the high probabilities

2013 48-h and 120-h Genesis Forecasts

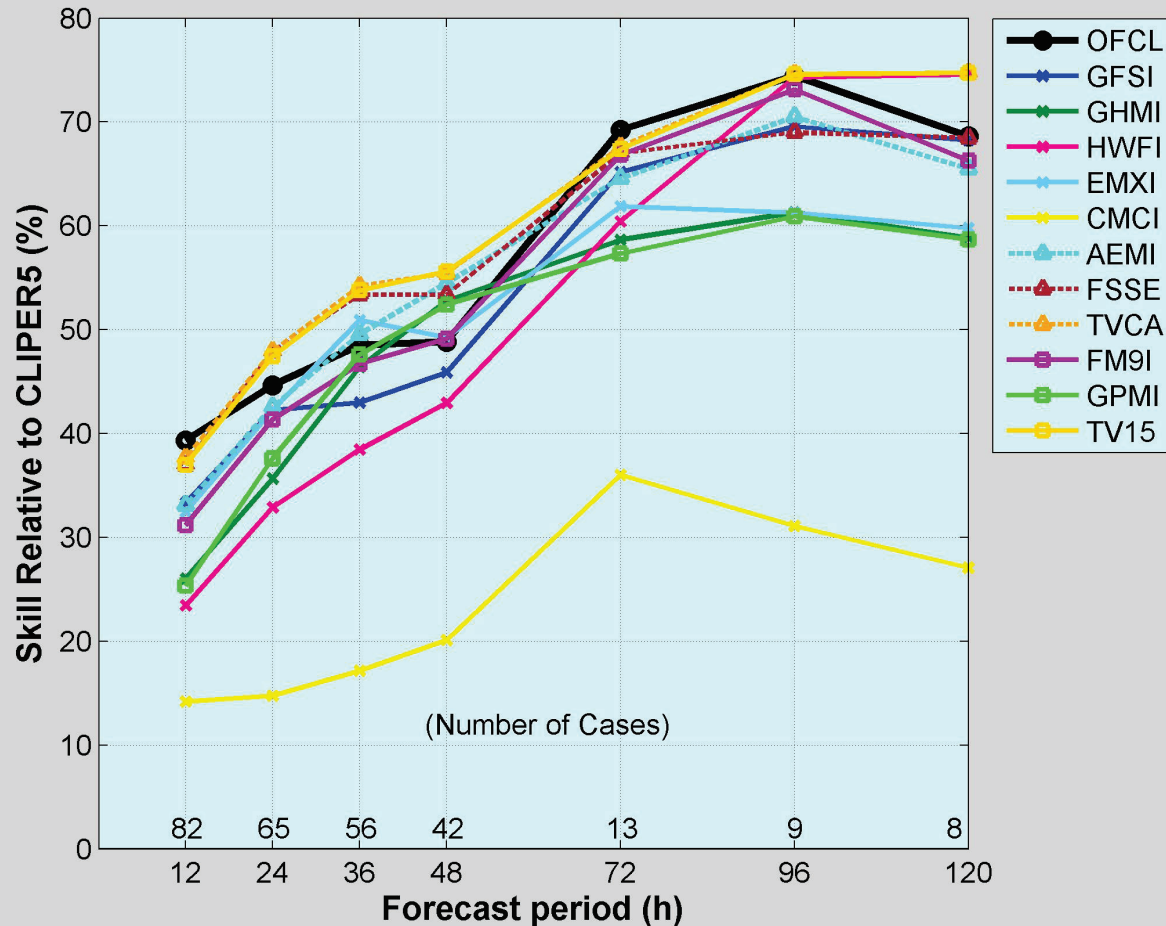


Although quite noisy, there was more reliability in the 5-day genesis forecasts.

Stream 1.5 models

2013 Track Guidance

Track Forecast Skill (Stream 1.5 Early Models)
2013 - Atlantic Basin



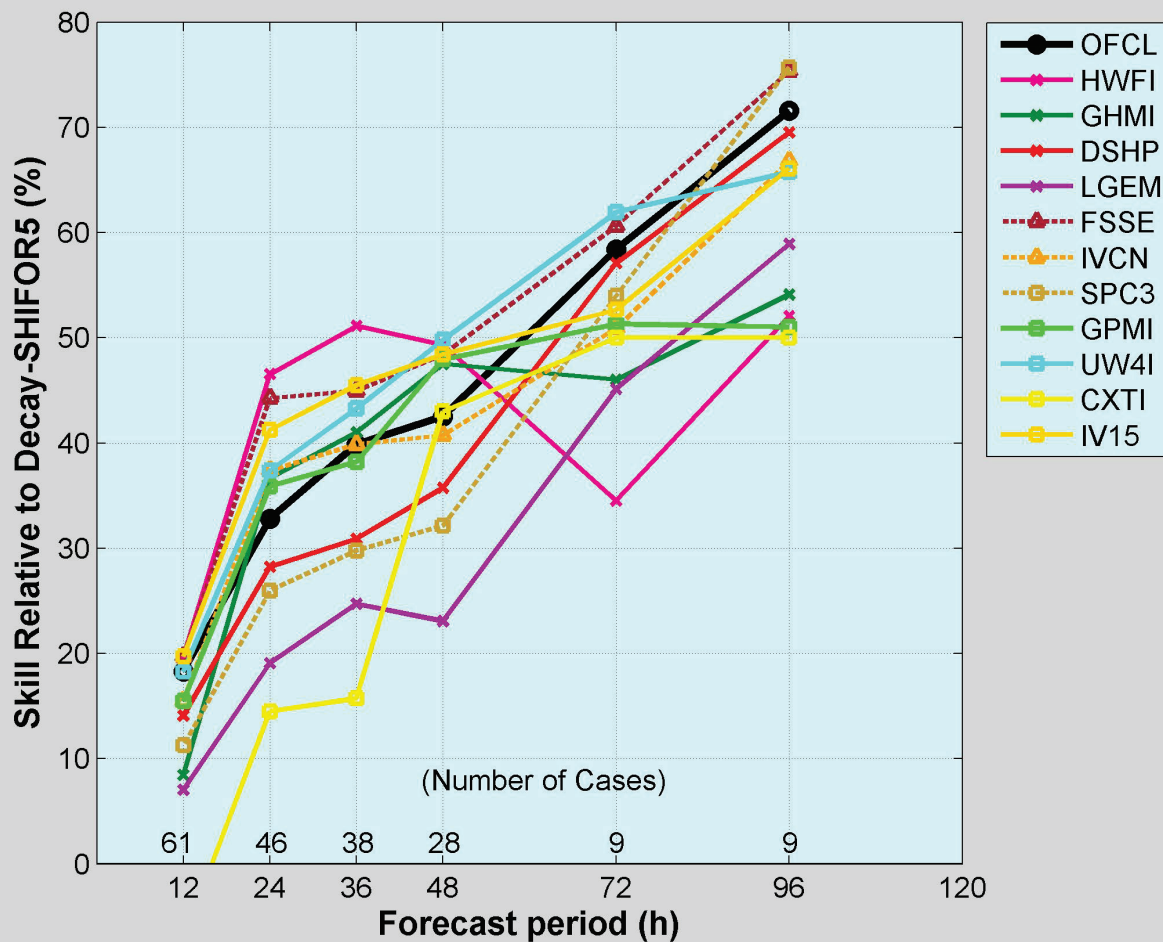
FM9I very good performer, competitive or slightly better than the best performing operational models.

GPMI close to *GHMI*, skill levels about 10-15% lower than the best models at days 4 and 5.

TV15/TVCA most skillful guidance.

2013 Intensity Guidance

Intensity Forecast Skill (Stream 1.5 Early Models)
2013 - Atlantic Basin



Nearly all of the guidance was skillful in 2013.

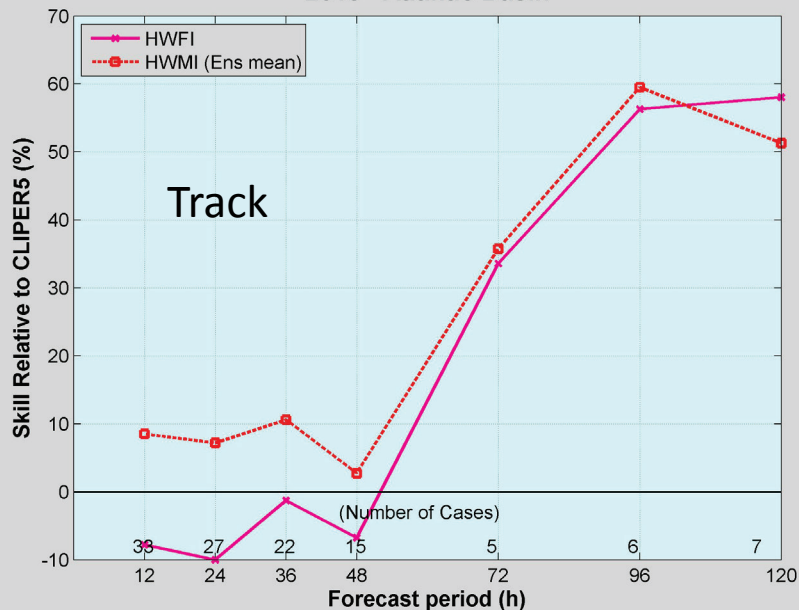
SPC3 less skill than most models early, but most skillful at 96 h (for a small sample).

UW4I very good performer throughout, near high end of the models.

CXTI low skill early, but increased beyond 36 h.

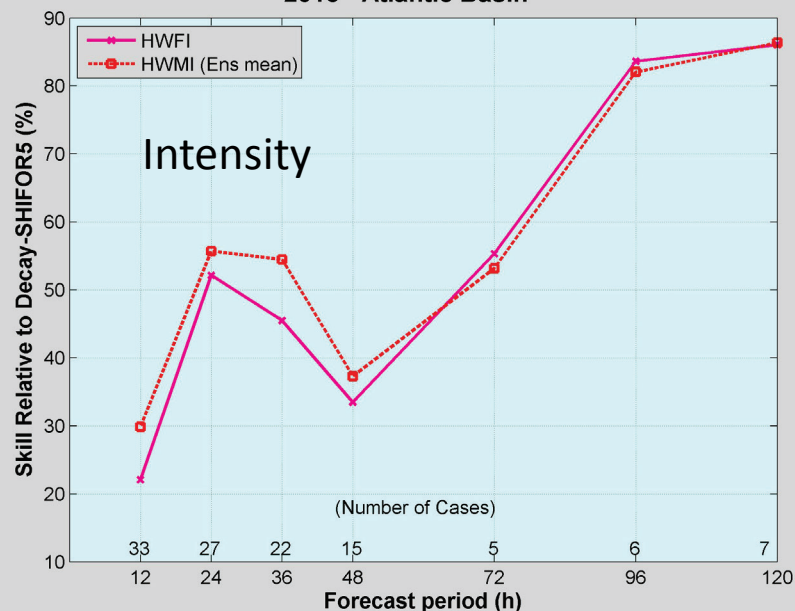
HWRF vs. HWRF ensemble mean

Track Forecast Skill (HWRF Stream 1.5)
2013 - Atlantic Basin



For track, the HWRF ensemble mean is better than the control run at nearly all times (largest improvement in the short term).

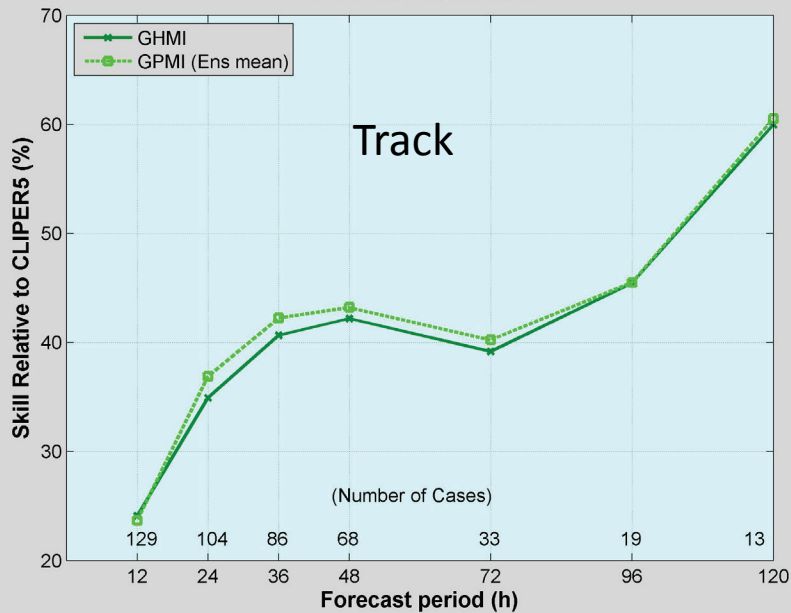
Intensity Forecast Skill (HWRF Stream 1.5)
2013 - Atlantic Basin



For intensity, the HWRF ensemble mean is better than the control run through 48 h, then similar to the control run after that time.

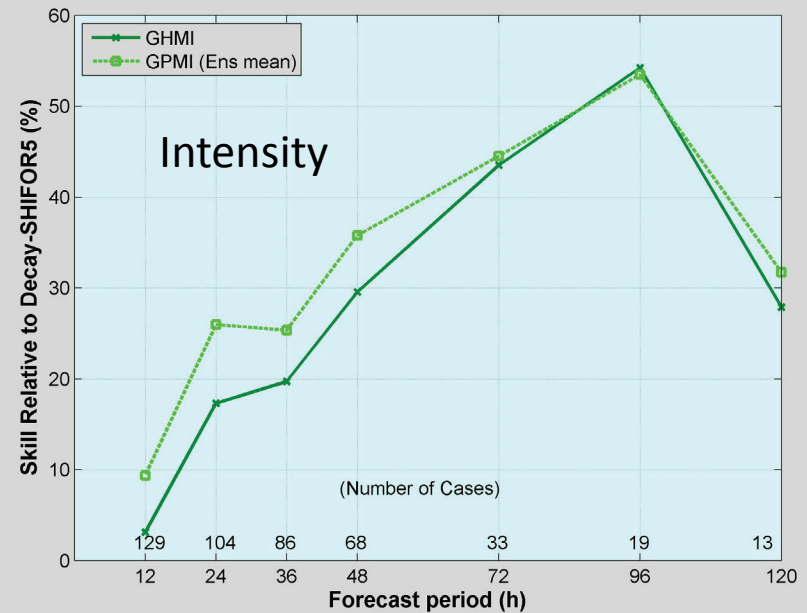
GFDL vs. GFDL ensemble mean

Track Forecast Skill (GFDL Stream 1.5)
2013 - Atlantic Basin



For track, the GFDL ensemble mean is slightly better or similar to the control run.

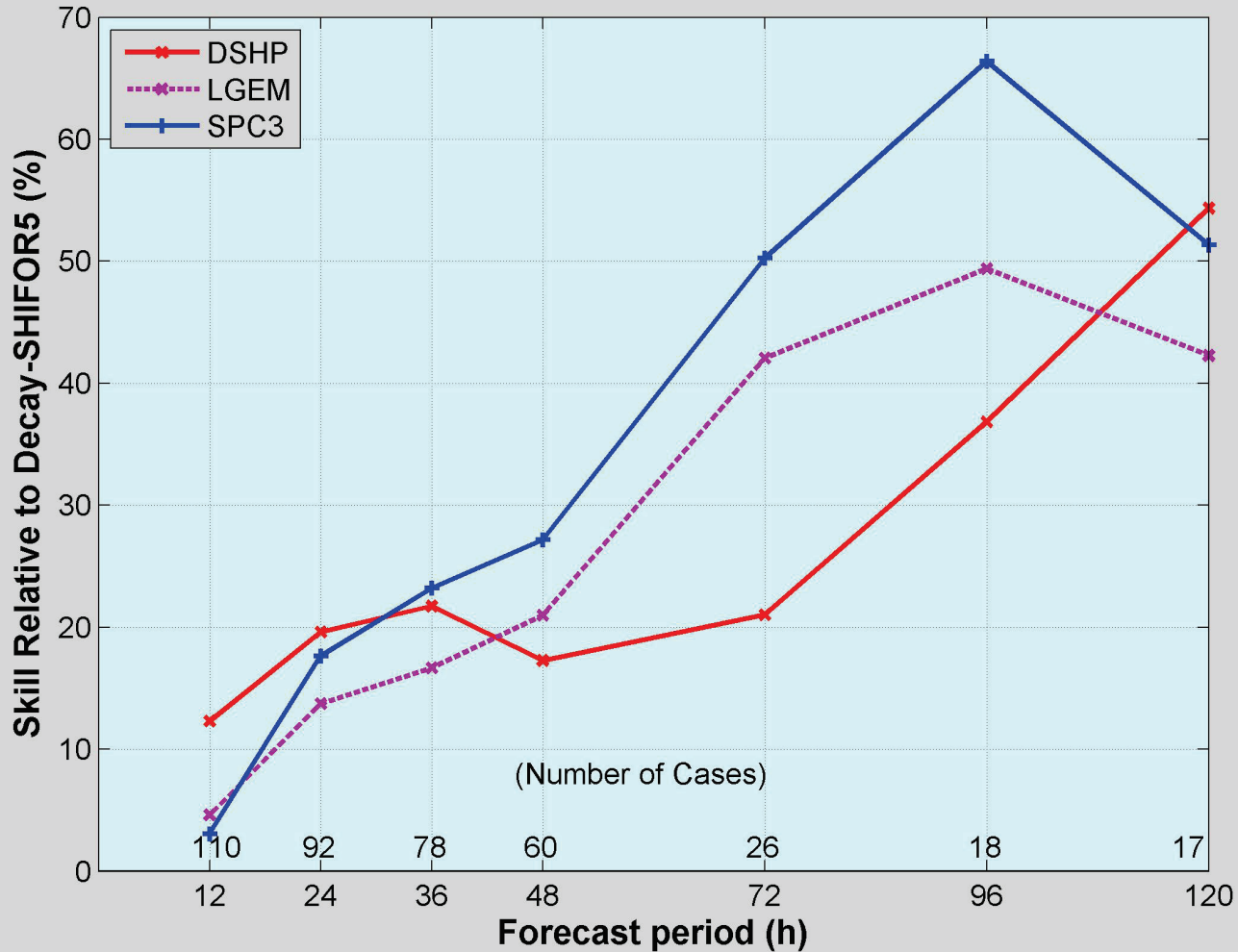
Intensity Forecast Skill (GFDL Stream 1.5)
2013 - Atlantic Basin



For intensity, the GFDL ensemble has about 5% more skill than the control run through 48 h, then similar from 72 to 120 h.

DSHP/LGEM vs. SPC3

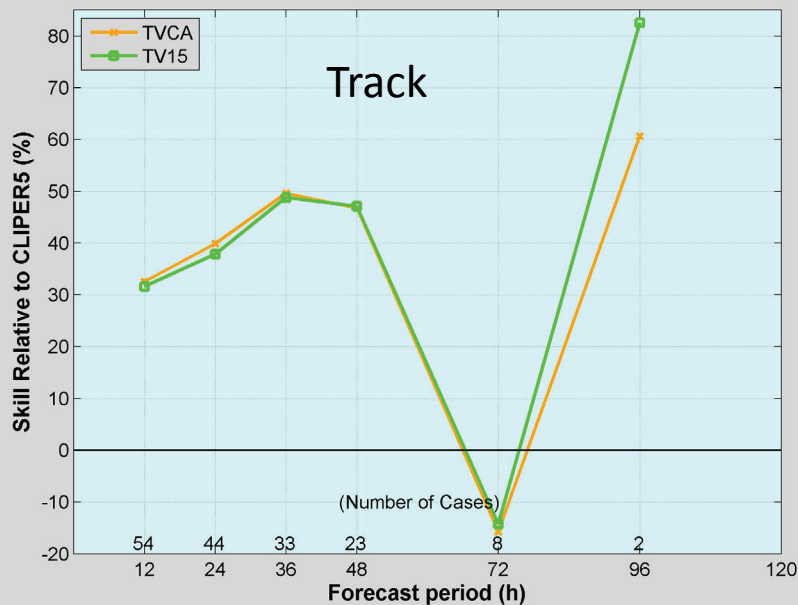
Intensity Forecast Skill (Stream 1.5)
2013 - Atlantic Basin



SPC3 more skillful than DSHP and LGEM at 36 h and beyond.

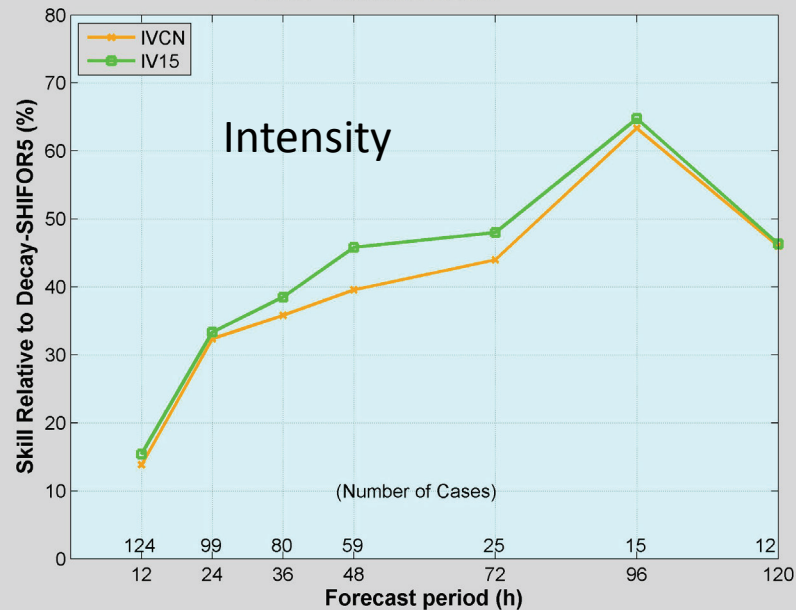
Consensus Aids

Track Forecast Skill (Stream 1.5 Consensus)
2013 - Atlantic Basin



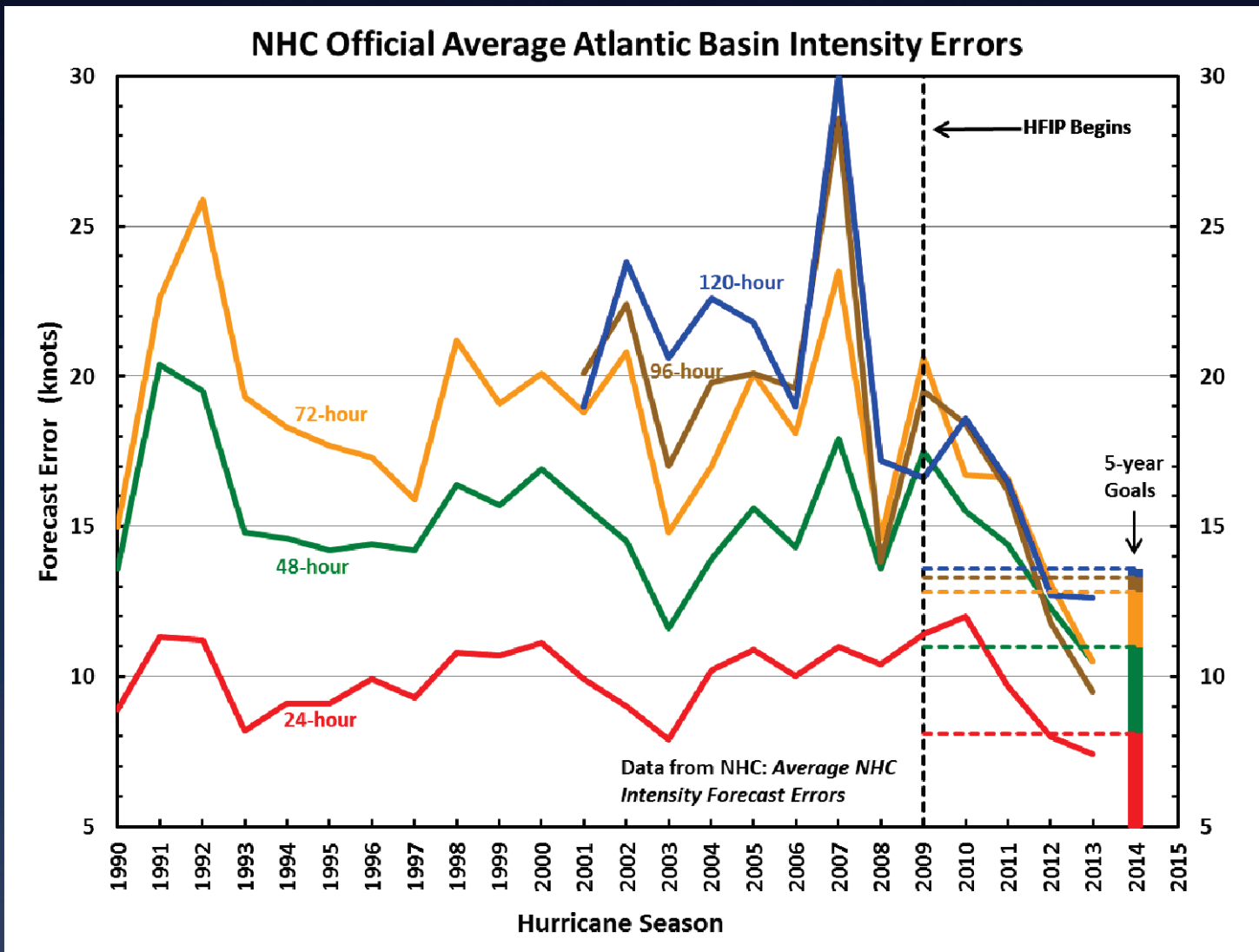
TVCA and TV15 similar, except at 96 h where TV15 is better for a tiny sample. APSI was the only Stream 1.5 track model used in TV15.

Intensity Forecast Skill (Stream 1.5 Consensus)
2013 - Atlantic Basin



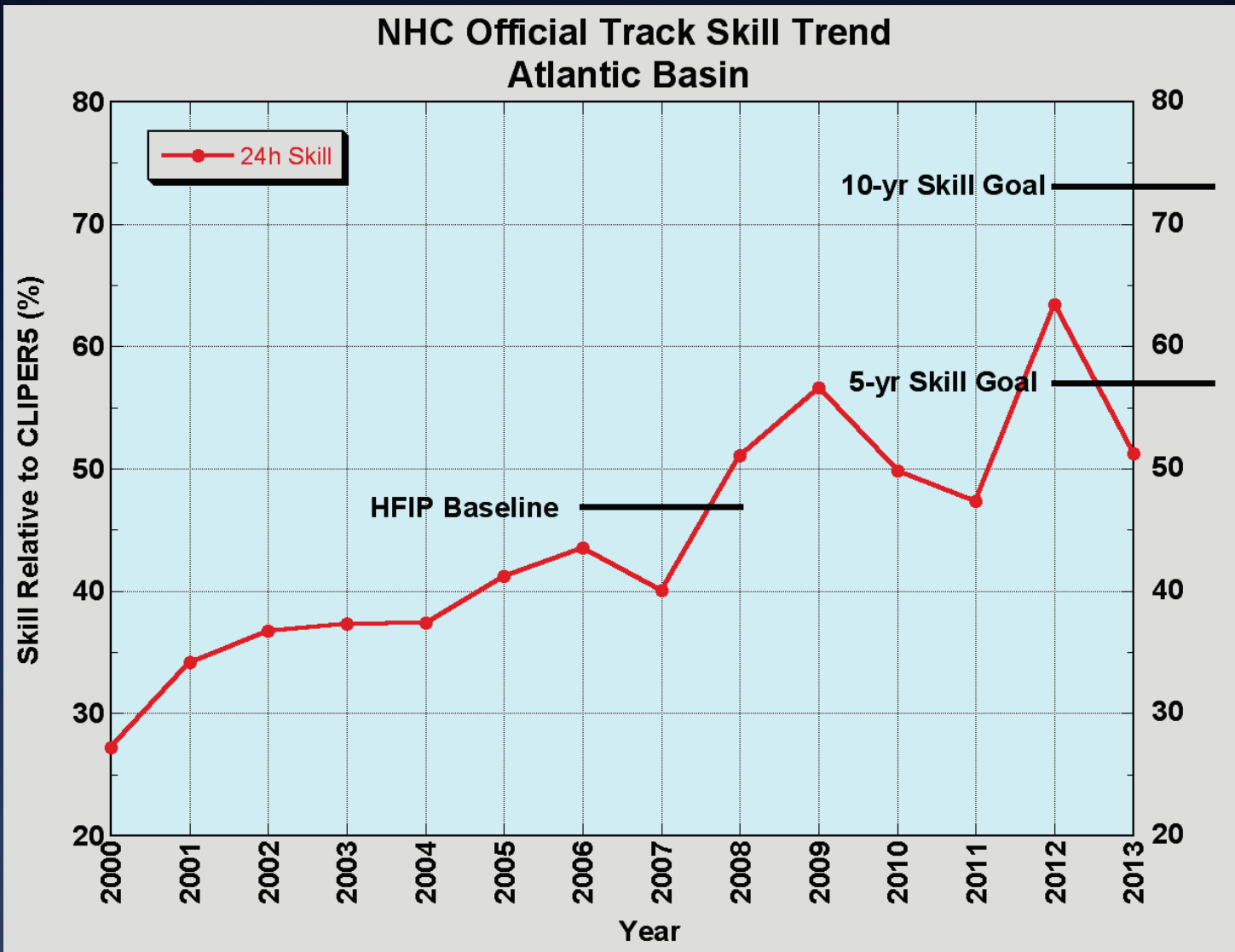
IV15 is a little better than IVCN (up to 5 % more skillful). Stream 1.5 models used in IV15 were APSI, UW4I, and CXTI.

So Where Are We?



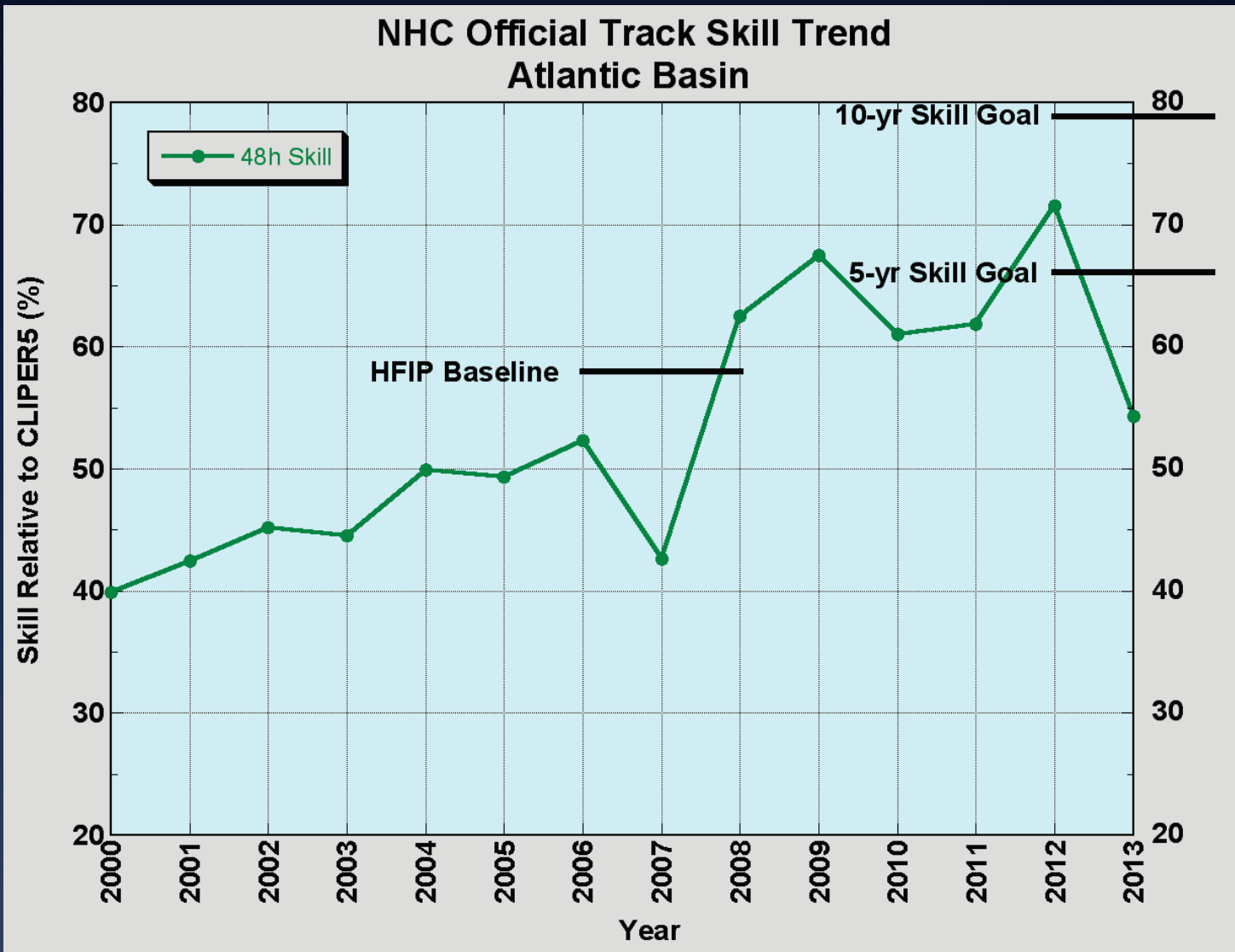
OFCL intensity errors seem to show a sharp decline beginning when HFIP began...

24-h OFCL Track Skill



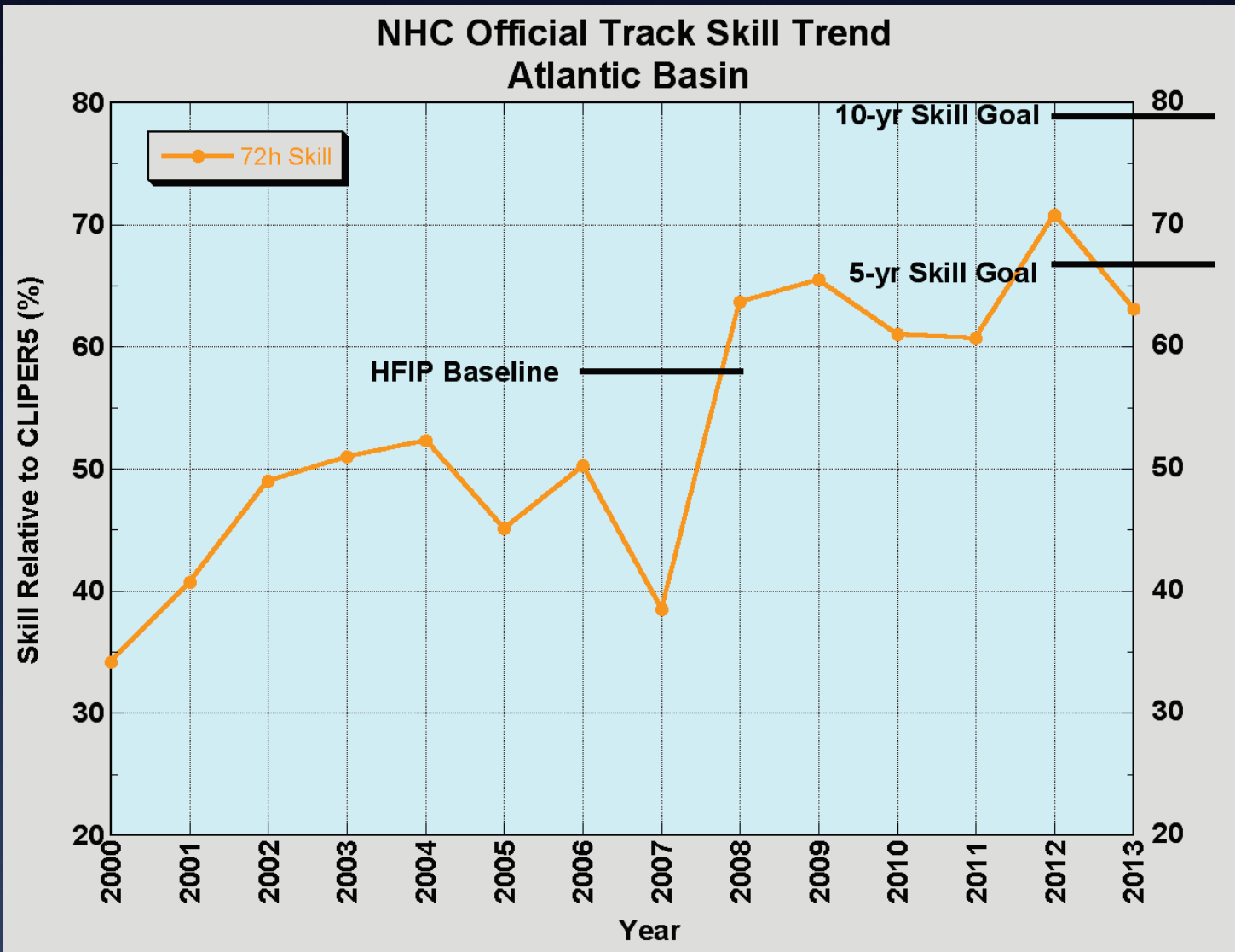
OFCL 24-h track forecasts over past two years shows track skill near the 5-yr goal

48-h OFCL Track Skill



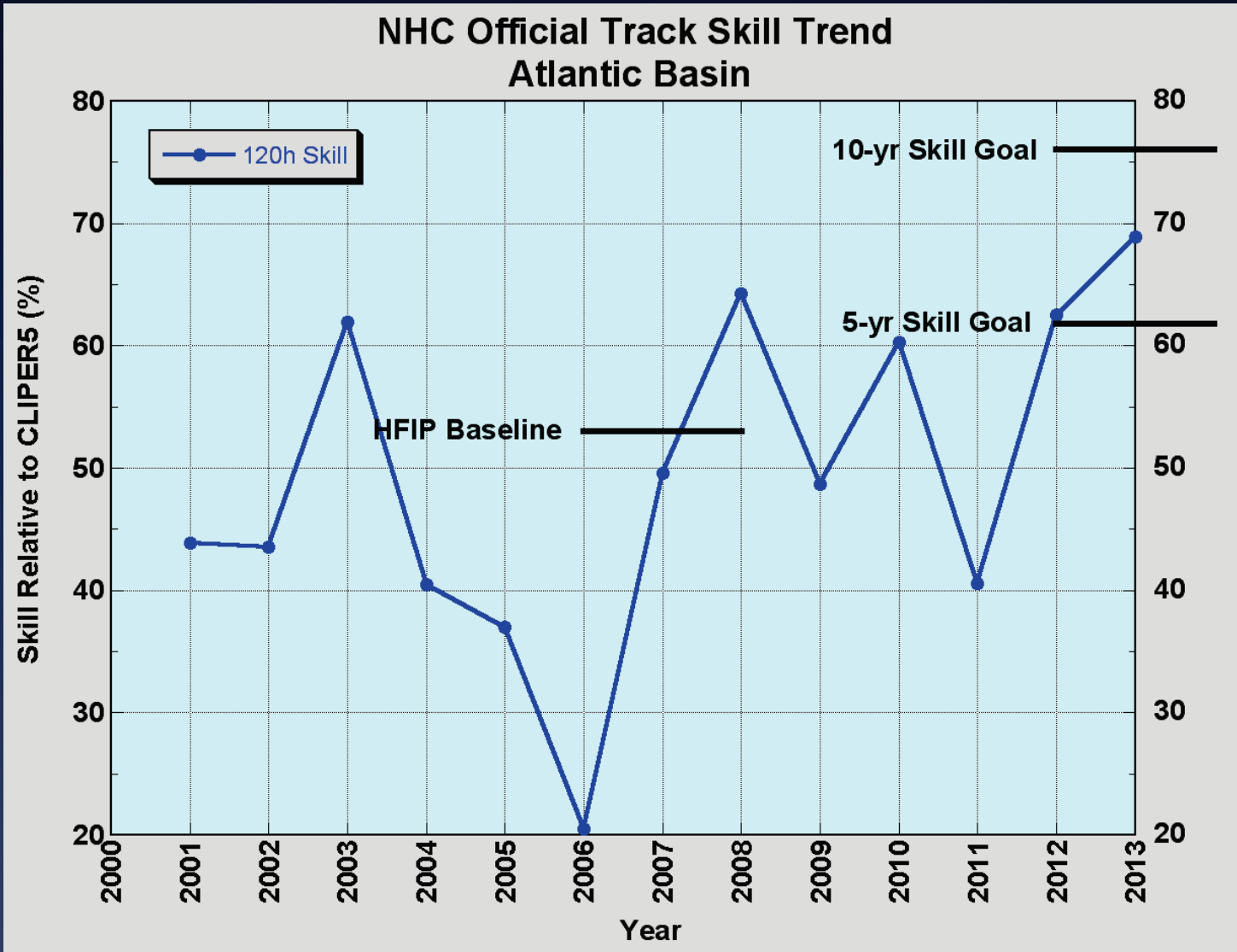
OFCL 48-h track forecasts show little trend, but might be near the 5-yr goal.

72-h OFCL Track Skill



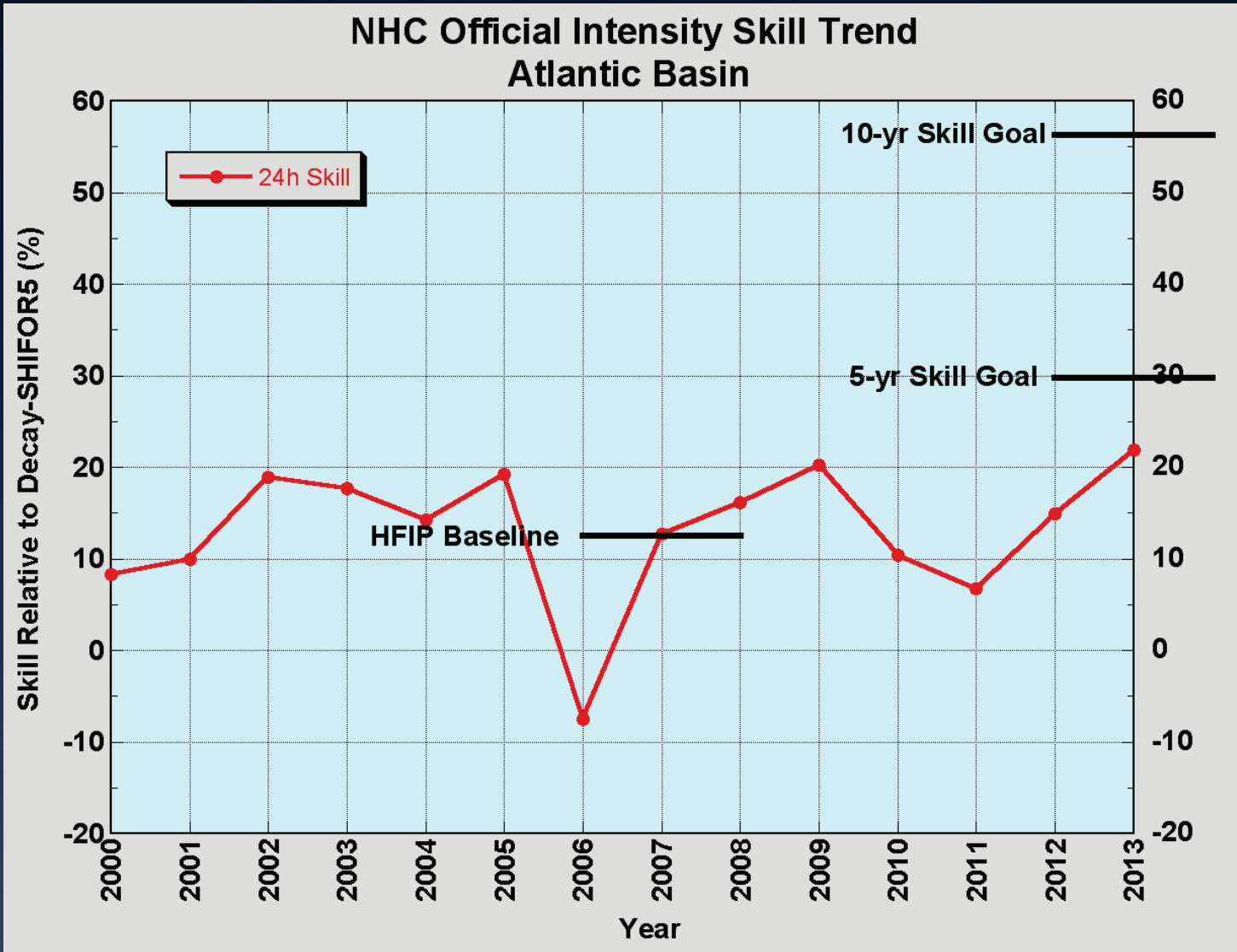
OFCL 72-h track forecasts show little trend, but might be near the 5-yr goal.

120-h OFCL Track Skill



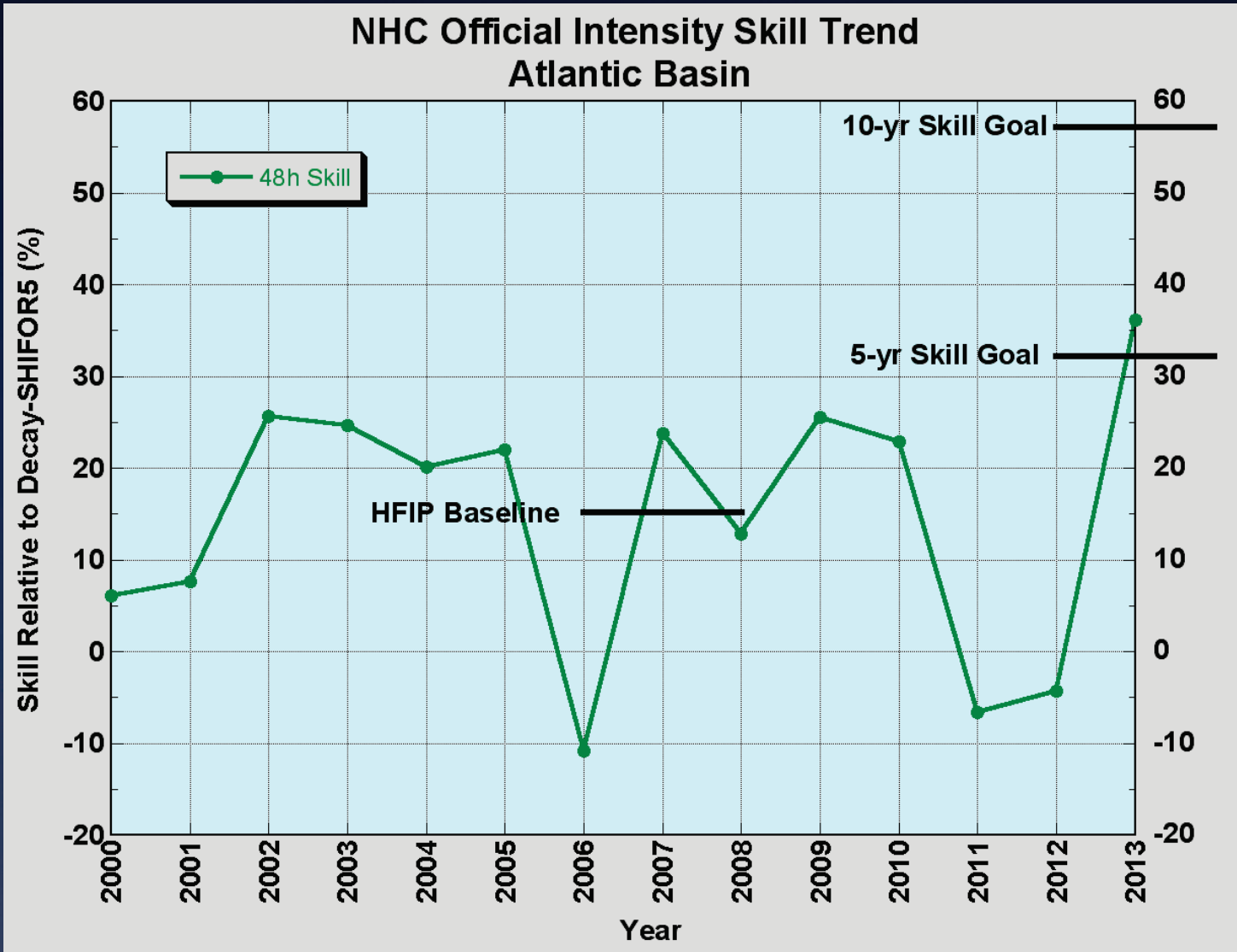
Skill in OFCL 120-h track forecasts too noisy to be sure of progress, but last two years are encouraging.

24-h OFCL Intensity Skill



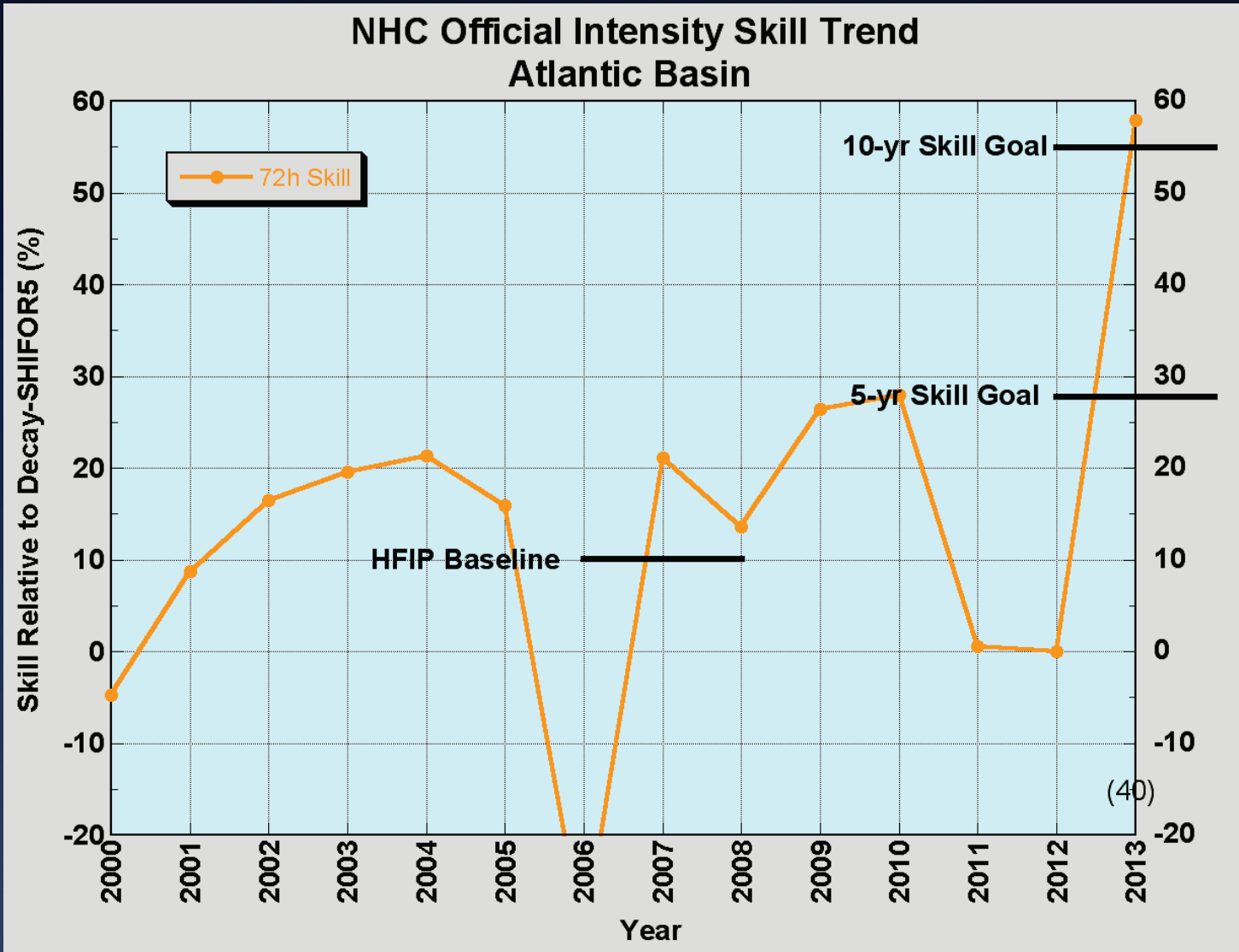
No real evidence that we're near the 5-year skill goal. 2013 sample was very small and basin conditions very unfavorable.

48-h OFCL Intensity Skill



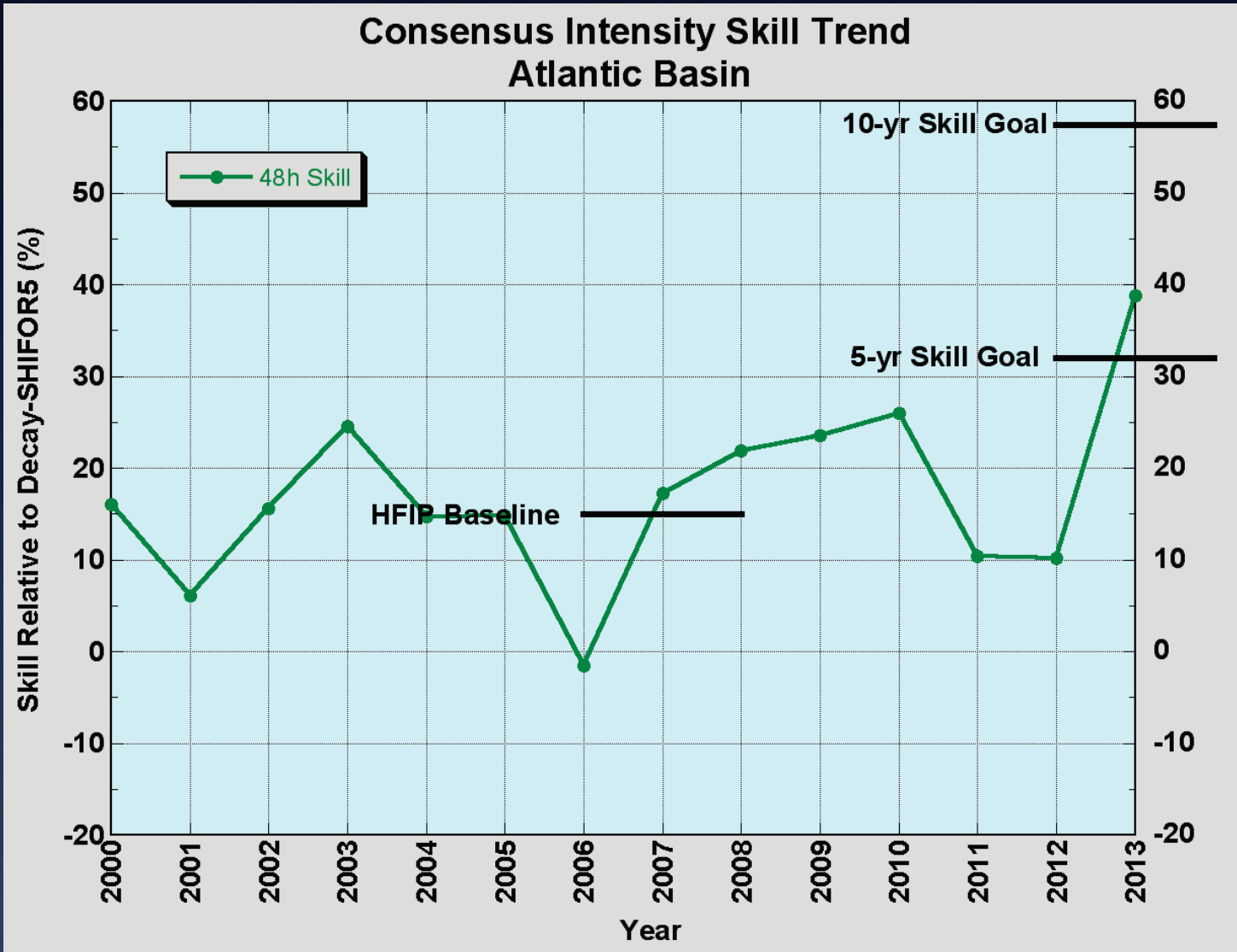
No linear improvement trend in skill space. Fraction of seasons that beat the baseline was 55% before HFIP, and 60% after HFIP.

72-h OFCL Intensity Skill



Very noisy in skill space. Is 2013 representative? Probably not.

48-h Consensus Intensity Skill



SPC3 and HWRF improvements offer hope that 2013 is a turning point, but time will tell.