HFIP 2011 Annual Review: Preliminary Forecast Verification

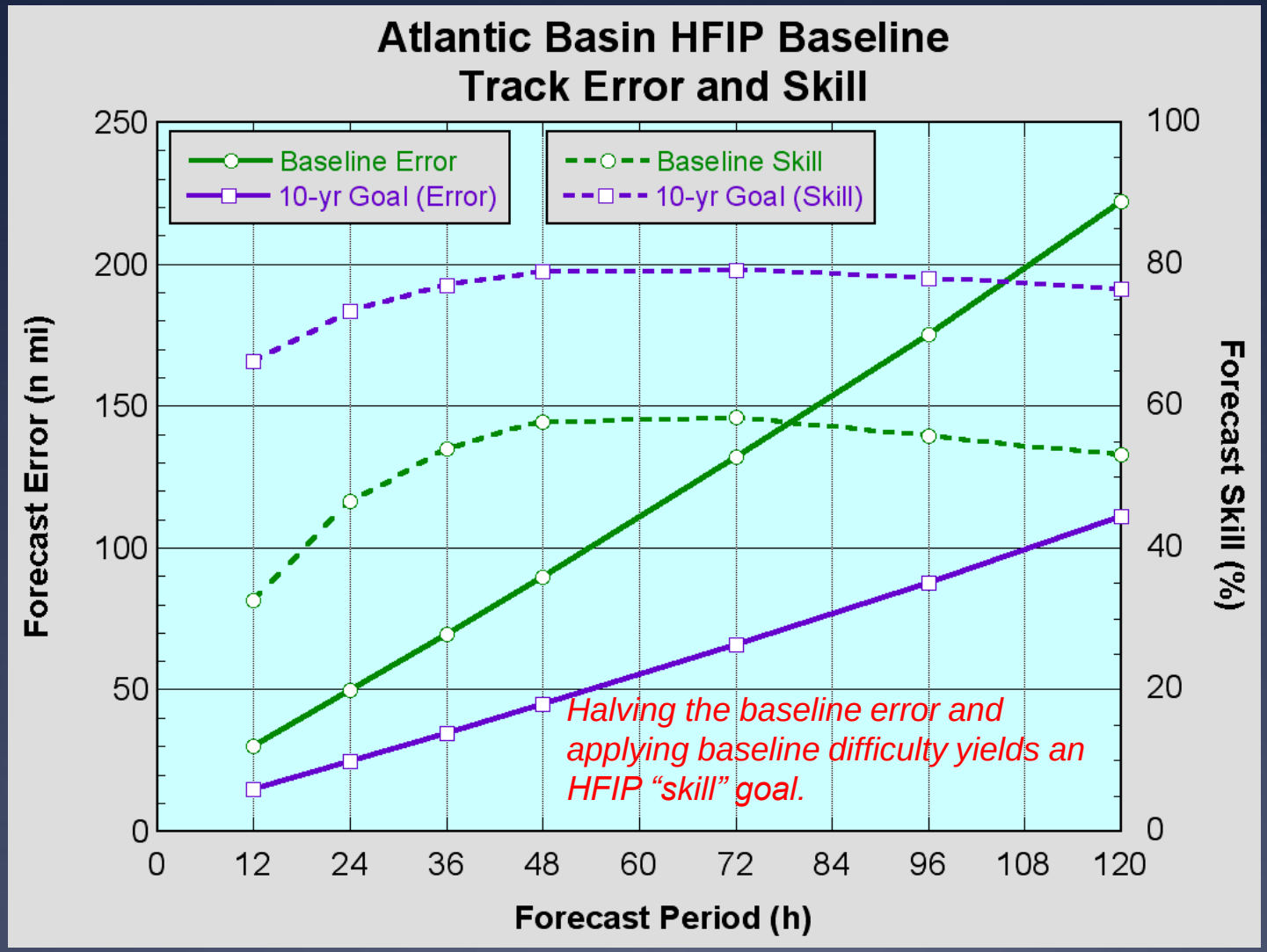
James L. Franklin
Branch Chief, Hurricane Specialist Unit
National Hurricane Center

HFIP Annual Review Meeting
8 Nov 2011

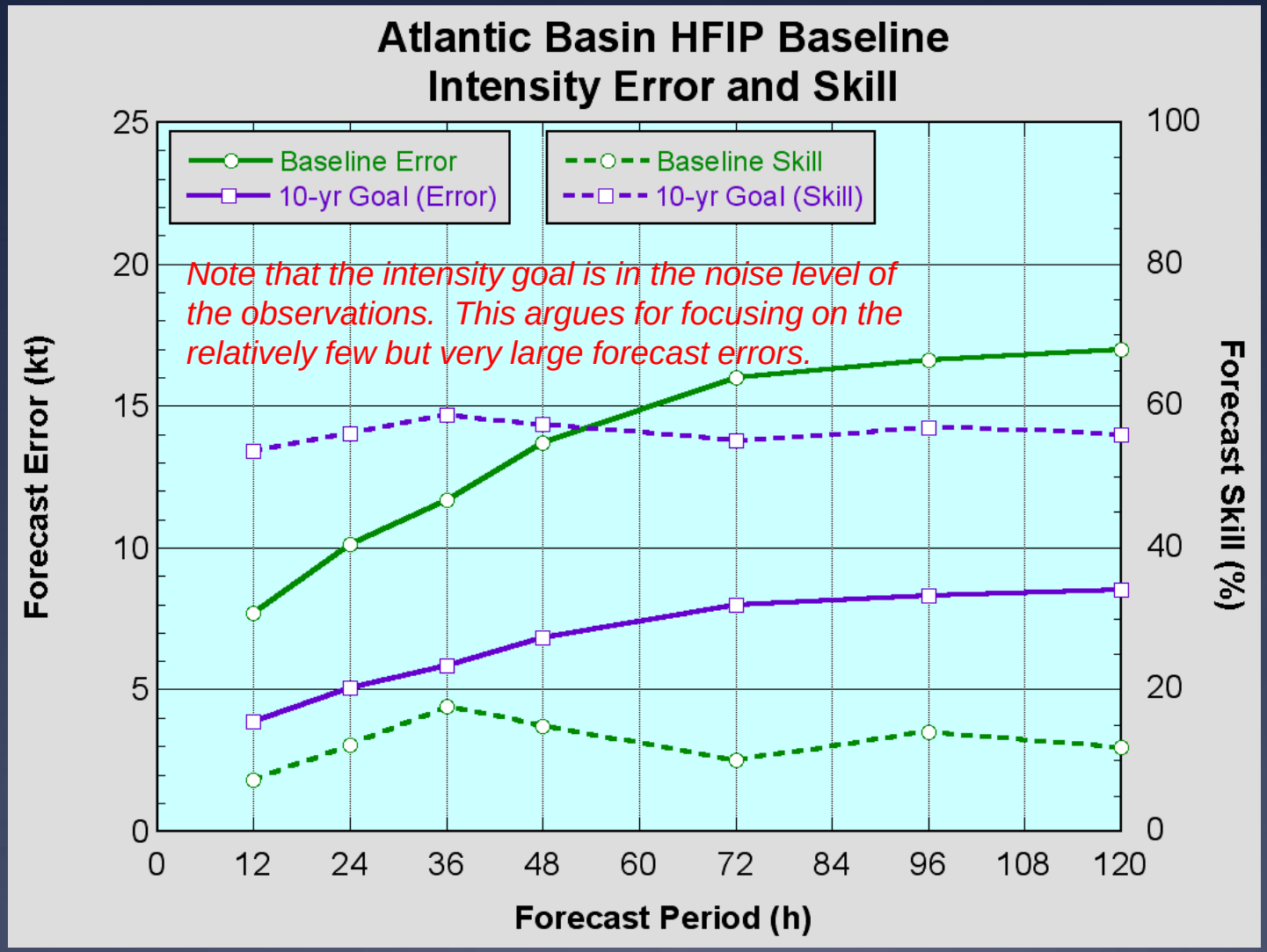
Introductory Remarks

- Evaluations are preliminary, based on NHC best tracks as of 10/31. All storms are included. Only Atlantic results presented here.
- Standard verification rules (system must be a tropical cyclone at the forecast time and at the verification time).
- Except as noted, 12-hr interpolations (e.g., COT2) are verified if the 6-hr interpolation (e.g., COTI) is not available.
- Verifications shown here use data provided to NHC in real time. However, the early (interpolated) Stream 1.5 intensity guidance was regenerated post-storm using the interpolator we had hoped to apply operationally (the GFDL-type interpolator that decays the intensity offset to zero).

HFIP Baselines and Goals



HFIP Baselines and Goals

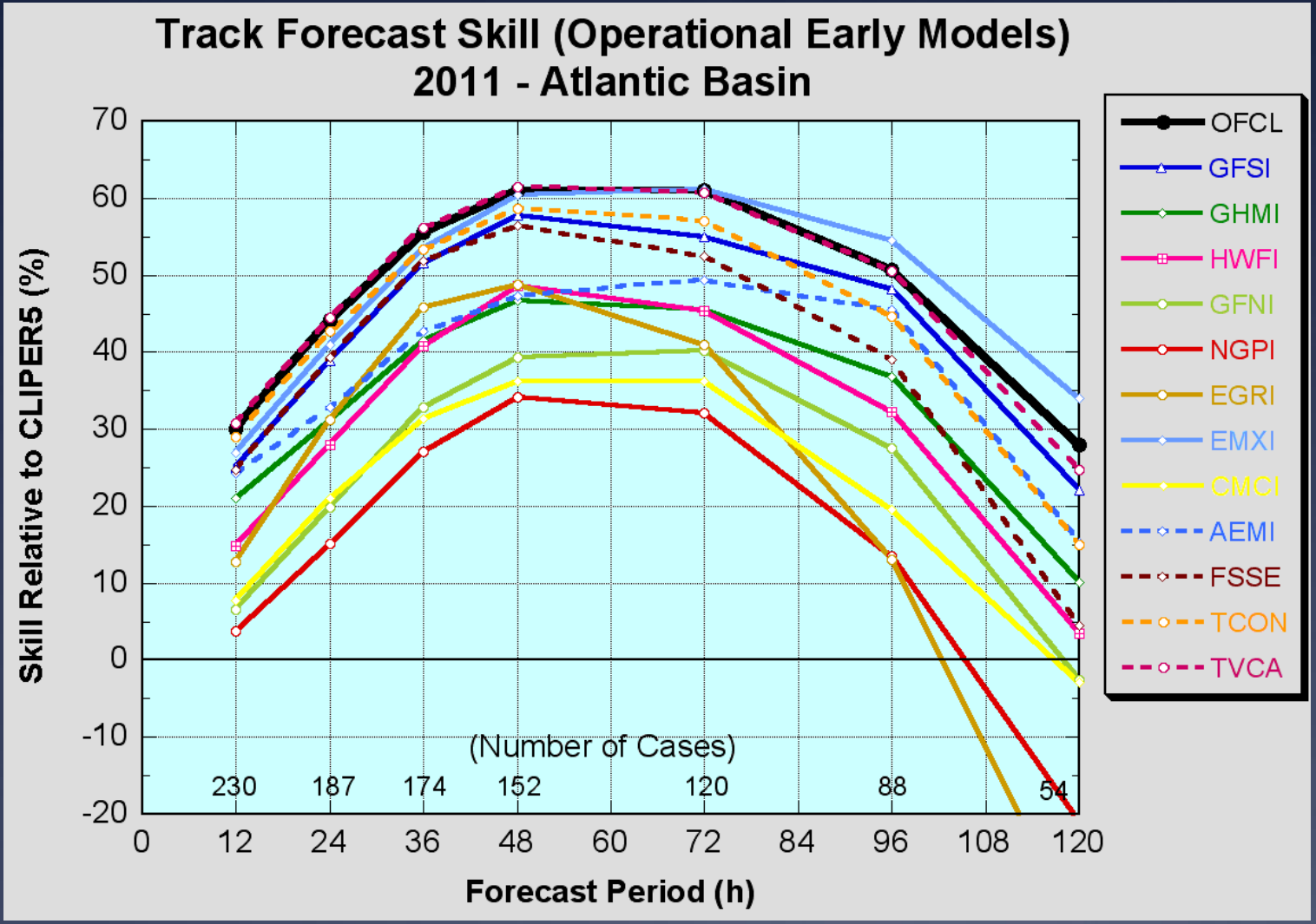


HFIP Goals
In Terms of Skill

Atlantic Basin				
VT (h)	Trk 20%	Trk 50%	Int 20%	Int 50%
0				
12	46.1	66.3	25.8	53.6
24	57.3	73.3	29.7	56.1
36	63.2	77.0	34.1	58.8
48	66.2	78.9	31.9	57.5
72	66.7	79.2	28.1	55.1
96	64.7	77.9	31.2	57.0
120	62.5	76.5	29.5	56.0

East Pacific Basin				
VT (h)	Trk 20%	Trk 50%	Int 20%	Int 50%
0				
12	37.6	61.0	22.3	51.4
24	47.6	67.3	25.7	53.5
36	53.1	70.7	27.8	54.9
48	55.2	72.0	29.3	55.8
72	57.3	73.3	28.9	55.6
96	52.0	70.0	28.7	55.4
120	46.8	66.7	29.7	56.0

2011 Preliminary Verifications



ECMWF and GFS did well (again). ECMWF beat consensus at longer ranges.

TVCA beat FSSE.

AEMI not as good as GFSI, even at 5 days.

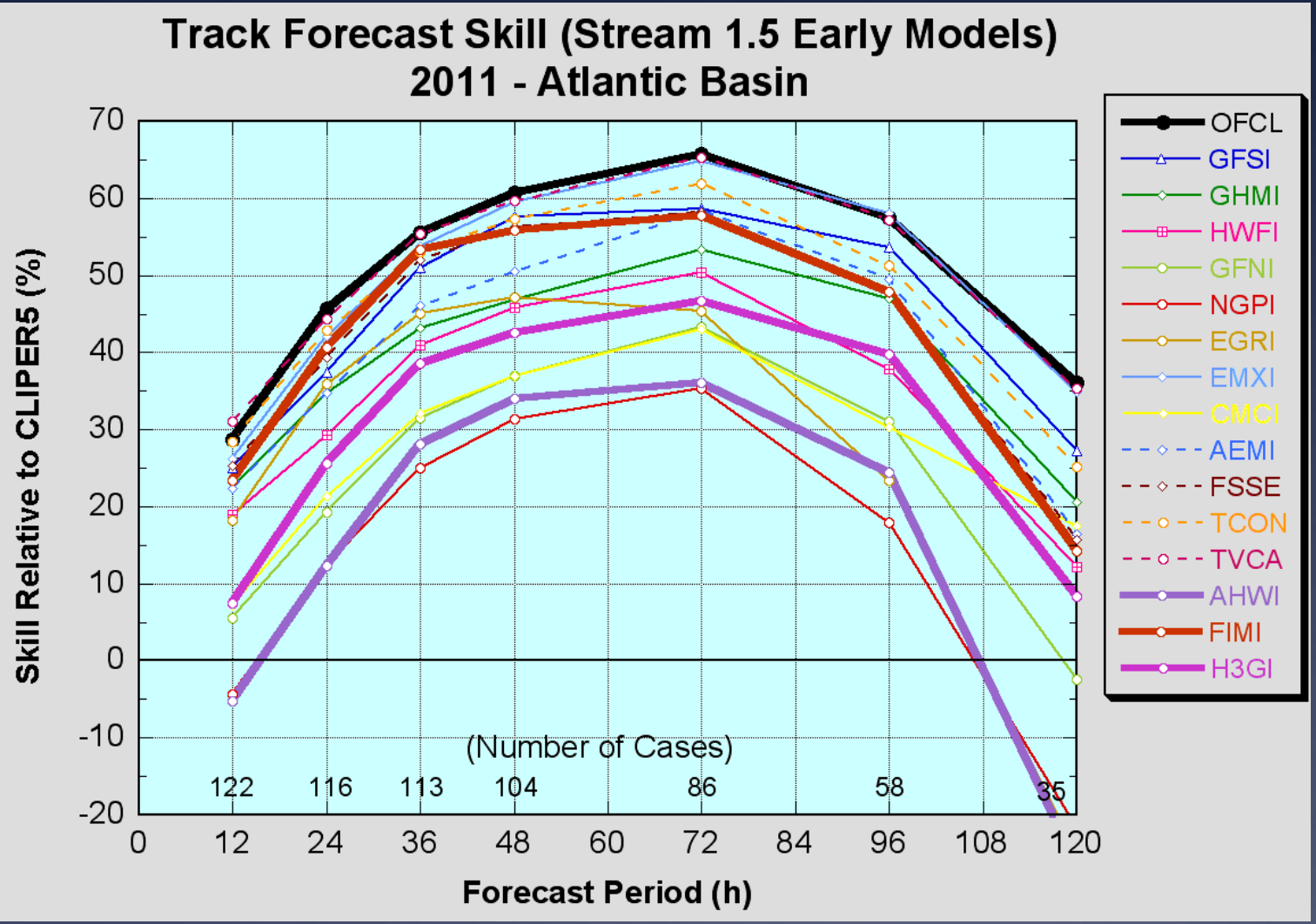
GFDL and HWRF middle of the pack.

NGPS, CMC, UKMET trailed.

2011 HFIP Stream 1.5 Models

- Track:
  - AHWI
  - FIMI
  - H3GI
  - GPMI
  - G01I
- Intensity:
  - AHWI
  - COTI
  - A4PI
  - UWNI
  - SPC3
  - GPMI
  - G01I

2011 Preliminary Verifications

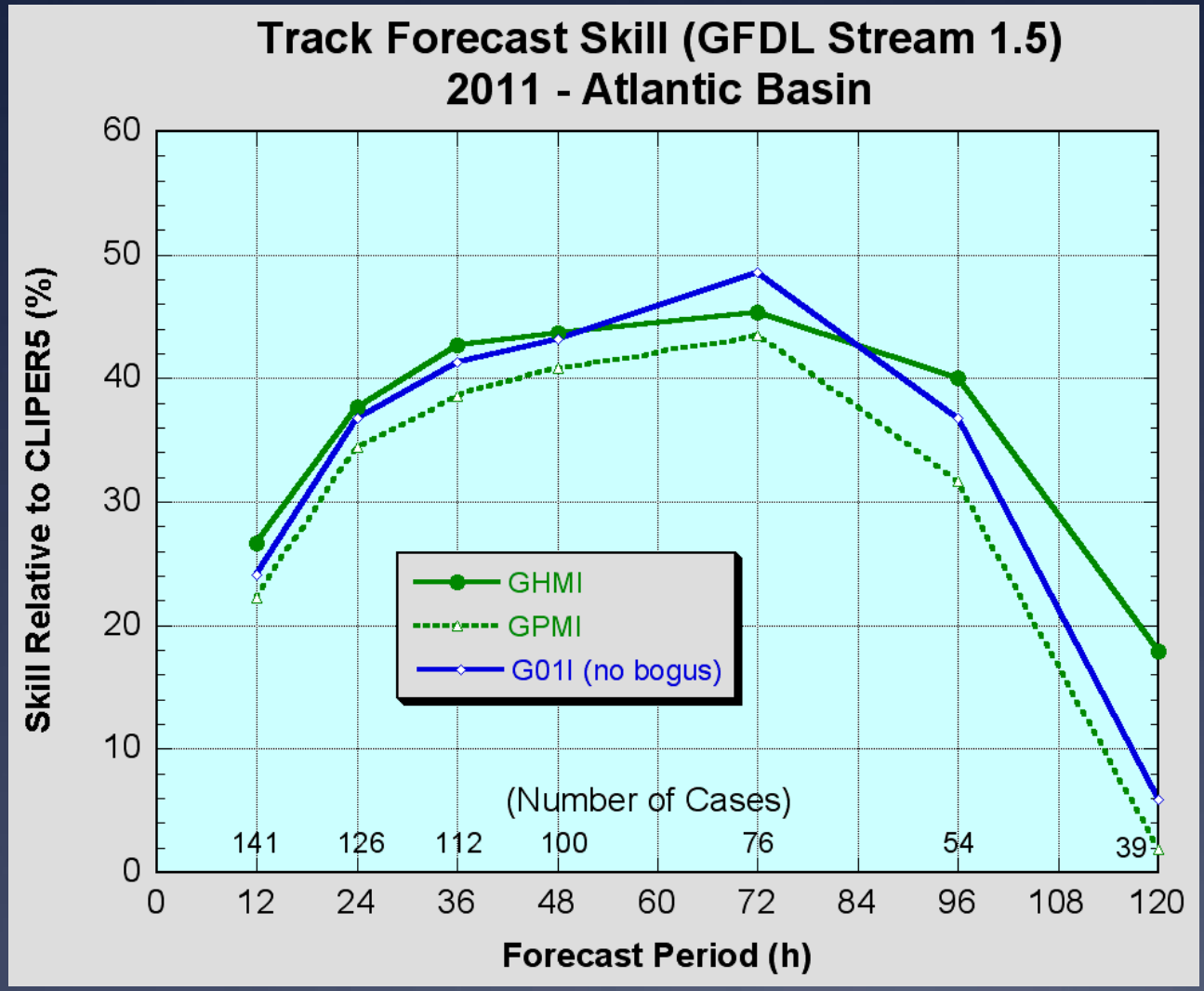


GFDL ensemble excluded due to small sample.

Of the three remaining Stream 1.5 models (FIMI, AHWI, H3GI), only the FIMI was competitive with the best operational models.

H3GI slightly less skillful than HWFI.

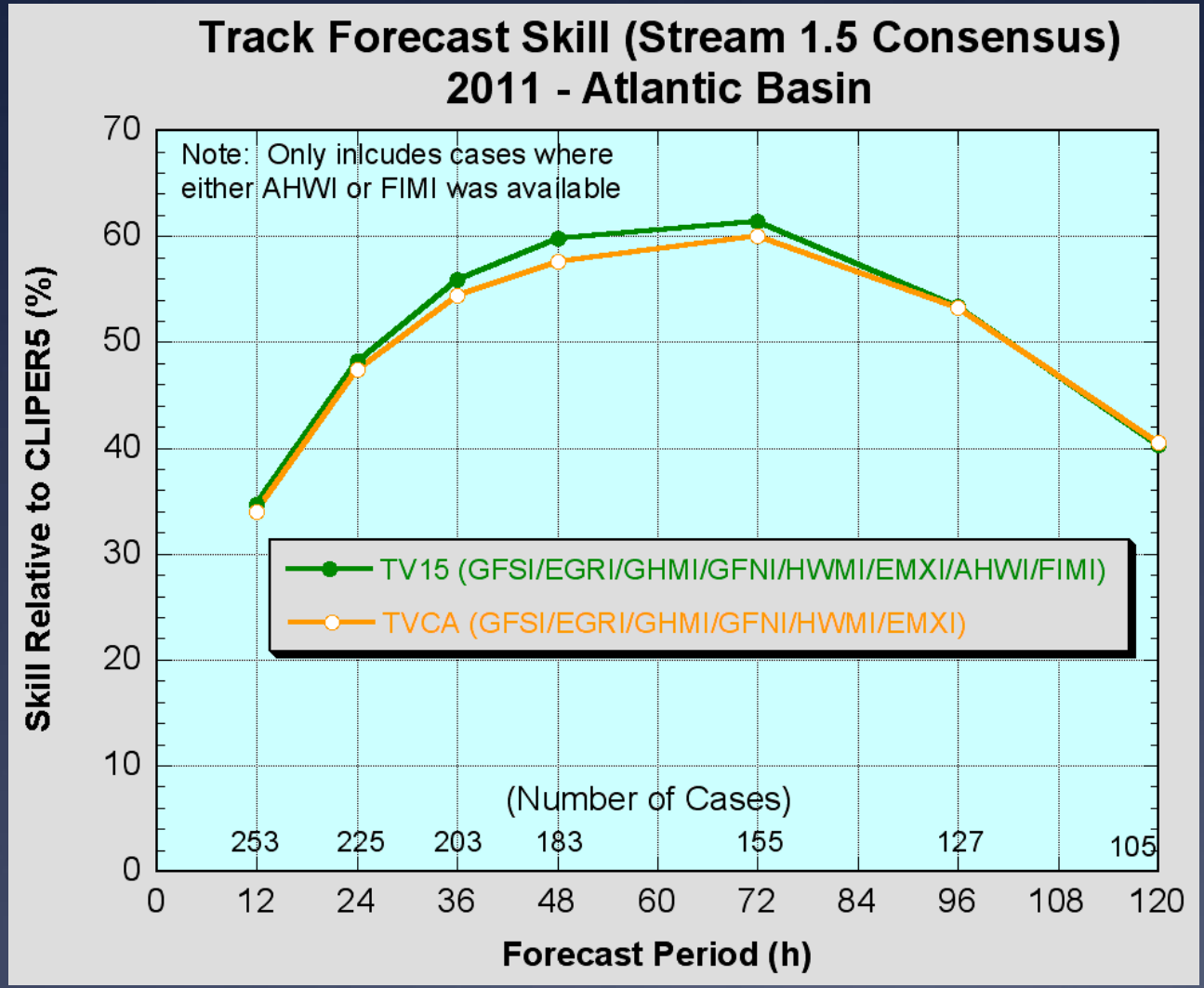
2011 Preliminary Verifications



GFDL ensemble mean not as good as the control.

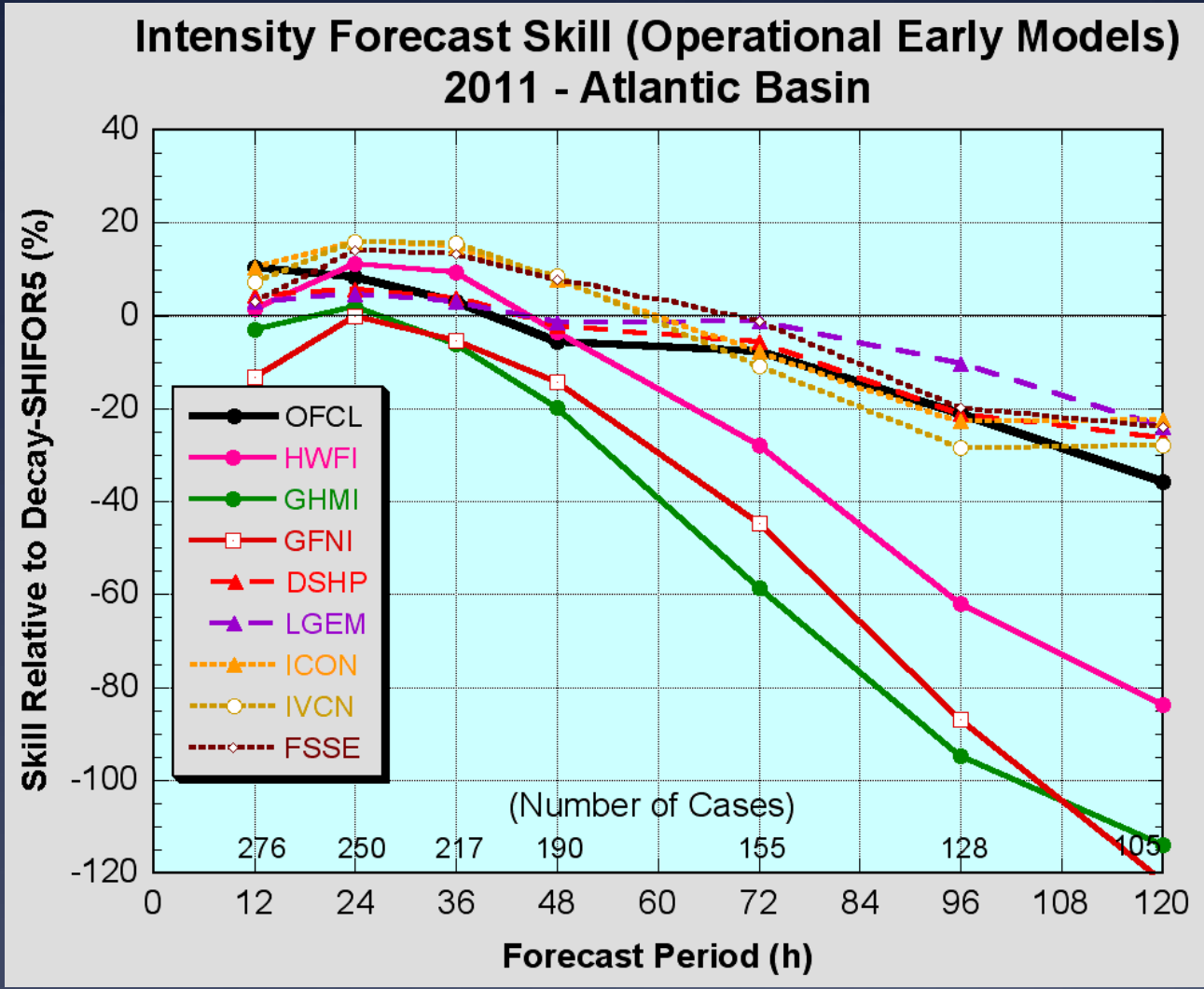
Unbogused ensemble member also not as good as the control.

2011 Preliminary Verifications



Stream 1.5 models admitted for track did contribute positively to the consensus, although the impact was small.

2011 Preliminary Verifications



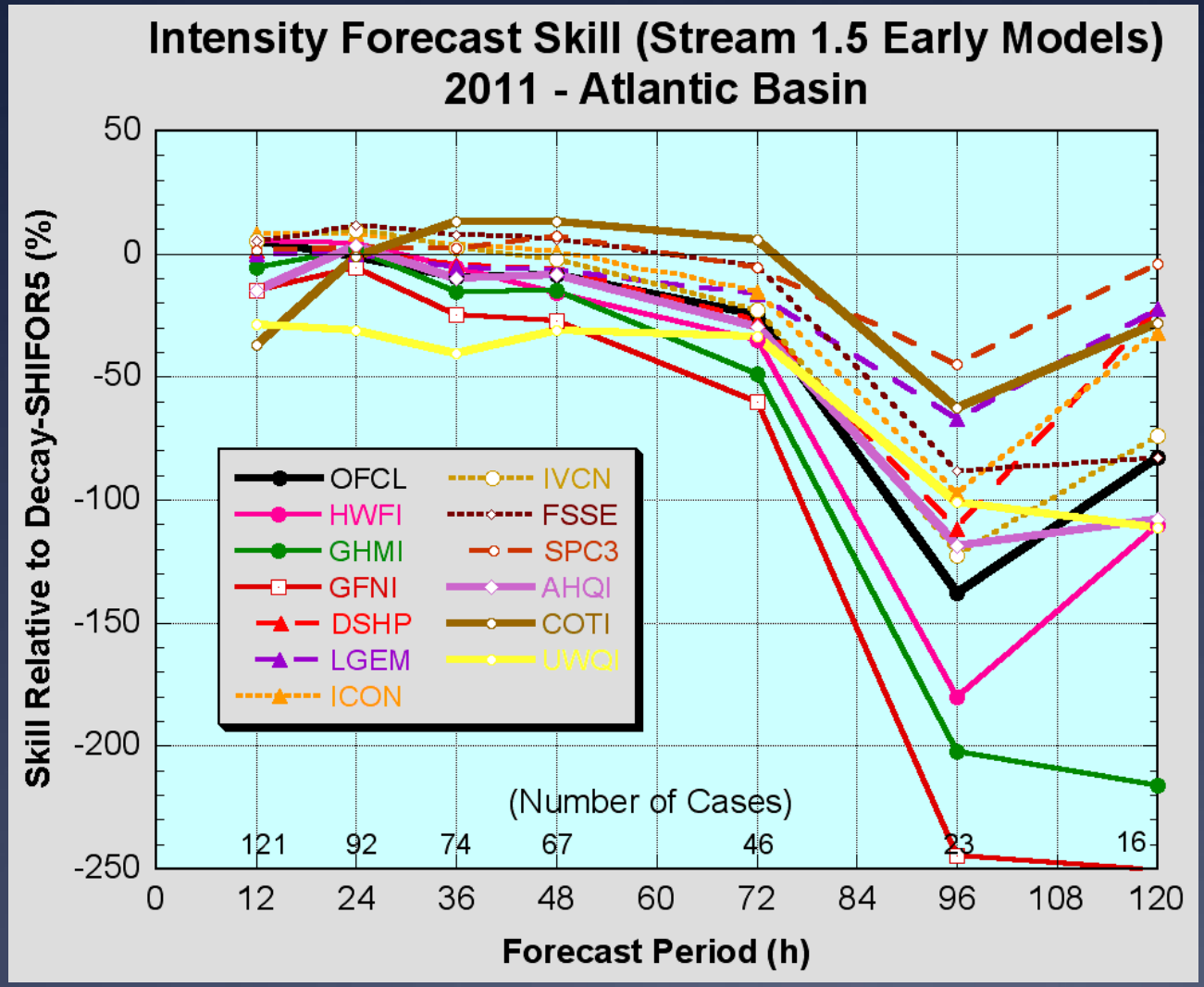
Storms were relatively well-behaved in 2011 -SHIFOR5 errors were well below normal (e.g., 13 kt at 120 h, compared with 5-yr mean of 22 kt).

No operational aid was skillful at or beyond 72 h.

Dynamical models performed very poorly.

LGEM was best individual model overall.

2011 Preliminary Verifications

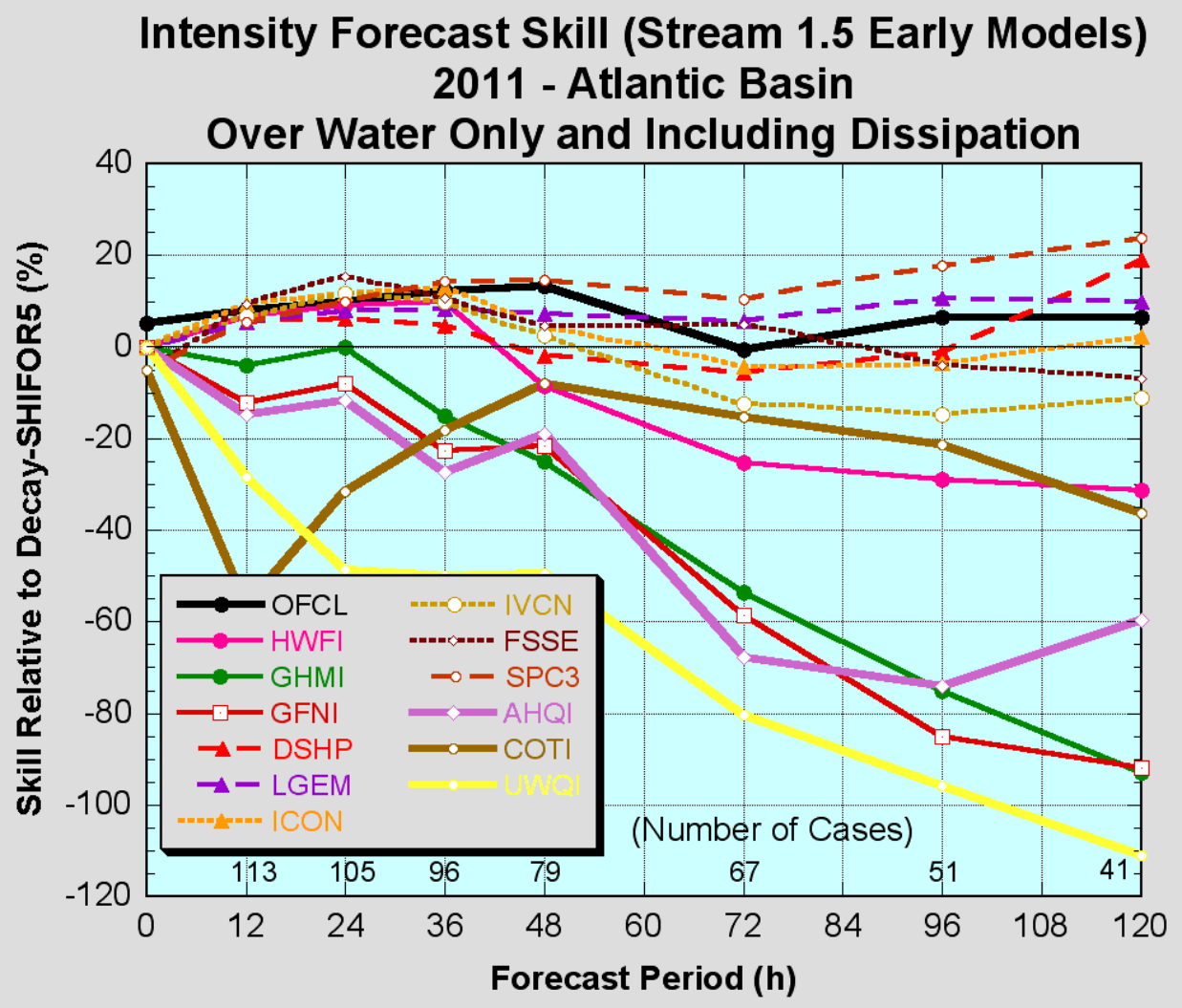


GFDL ensemble and A4PI (radar) omitted due to sample size.

Results shown here with the "Q" designation were regenerated post-storm using the GHMI interpolator (what we had hoped to apply operationally).

Of the Stream 1.5 models (SPC3, AHQI, COTI, UWQI), COTI and SPC3 performed better than the current operational models. SPC3 was better than DSHP or LGEM; COTI was the only dynamical model competitive with statistical guidance.

2011 Preliminary Verifications

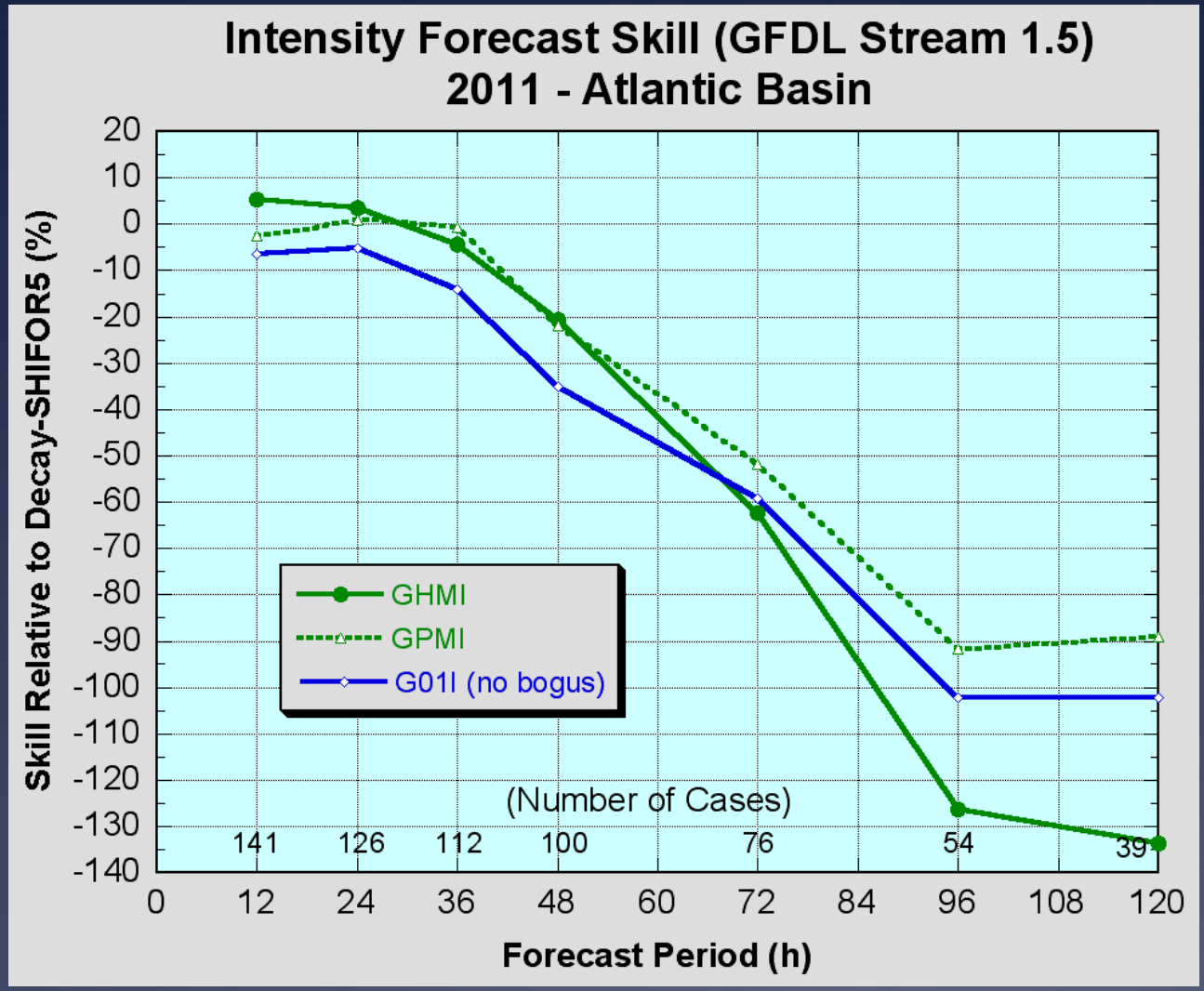


Restrict the sample to over water forecasts only, and include the dissipation stage (assign 15 kt, if necessary to either BT or model).

Now none of the dynamical models are as good as the statistical models.

HWRF best dynamical model overall.

2011 Preliminary Verifications

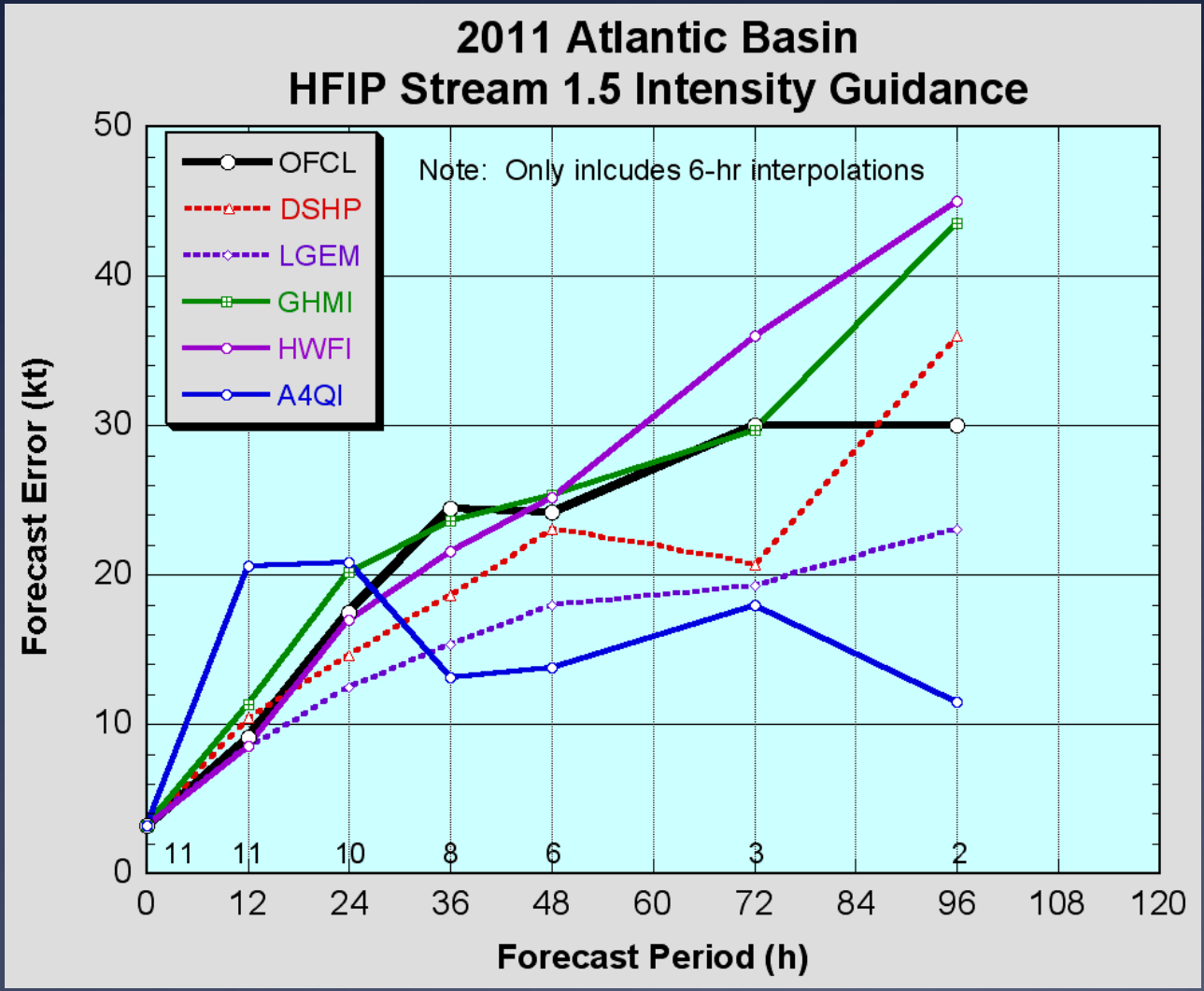


GFDL mostly not skillful.

GFDL ensemble mean mostly better than the control.

Unbogused ensemble member worse early, better late.

2011 Preliminary Verifications

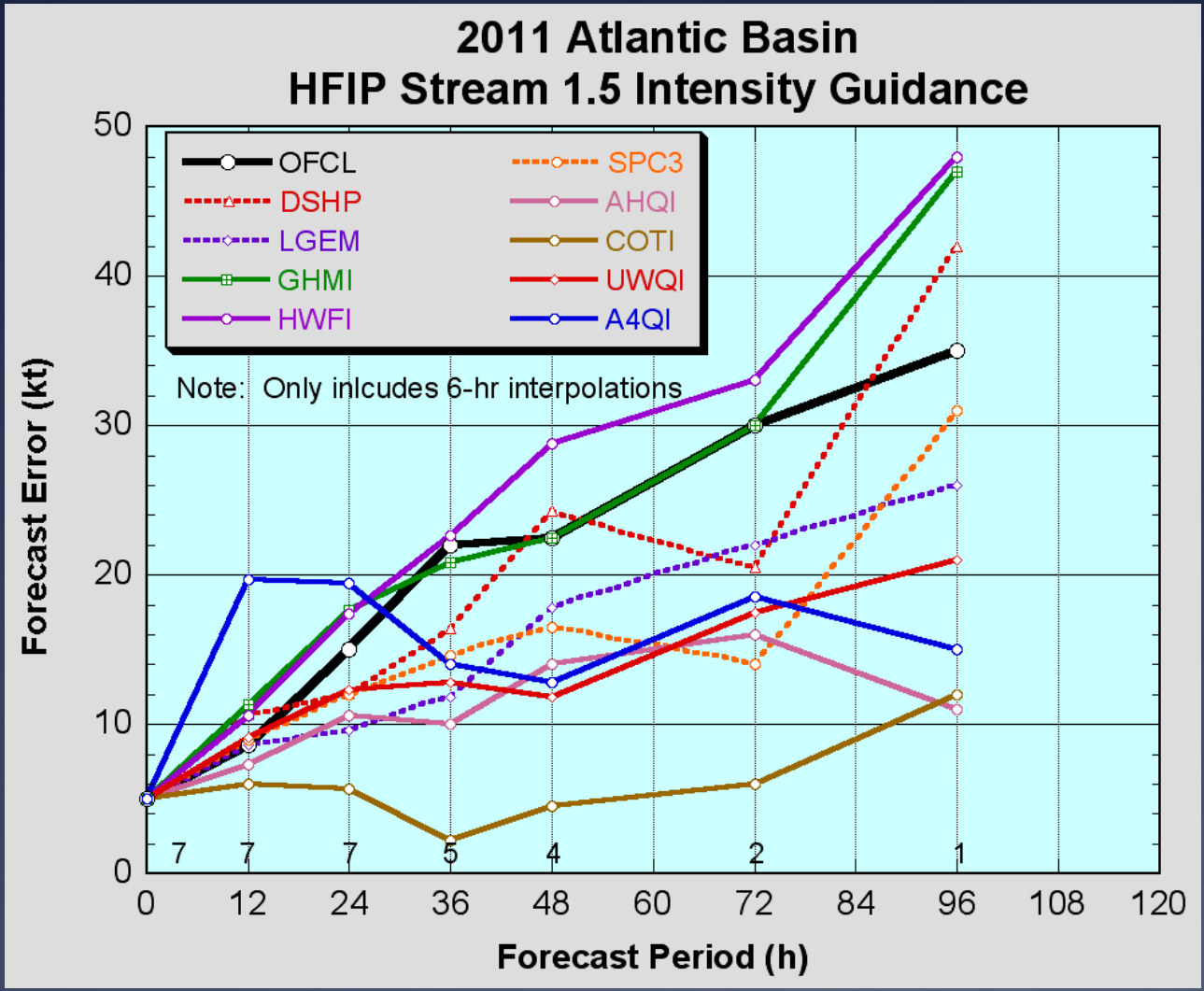


PSU radar runs (A4QI):

Presenting error rather than skill because the sample of radar cases is so small.

A4QI was better than the statistical guidance, except at 12-24 h, which is likely an interpolator issue.

2011 Preliminary Verifications

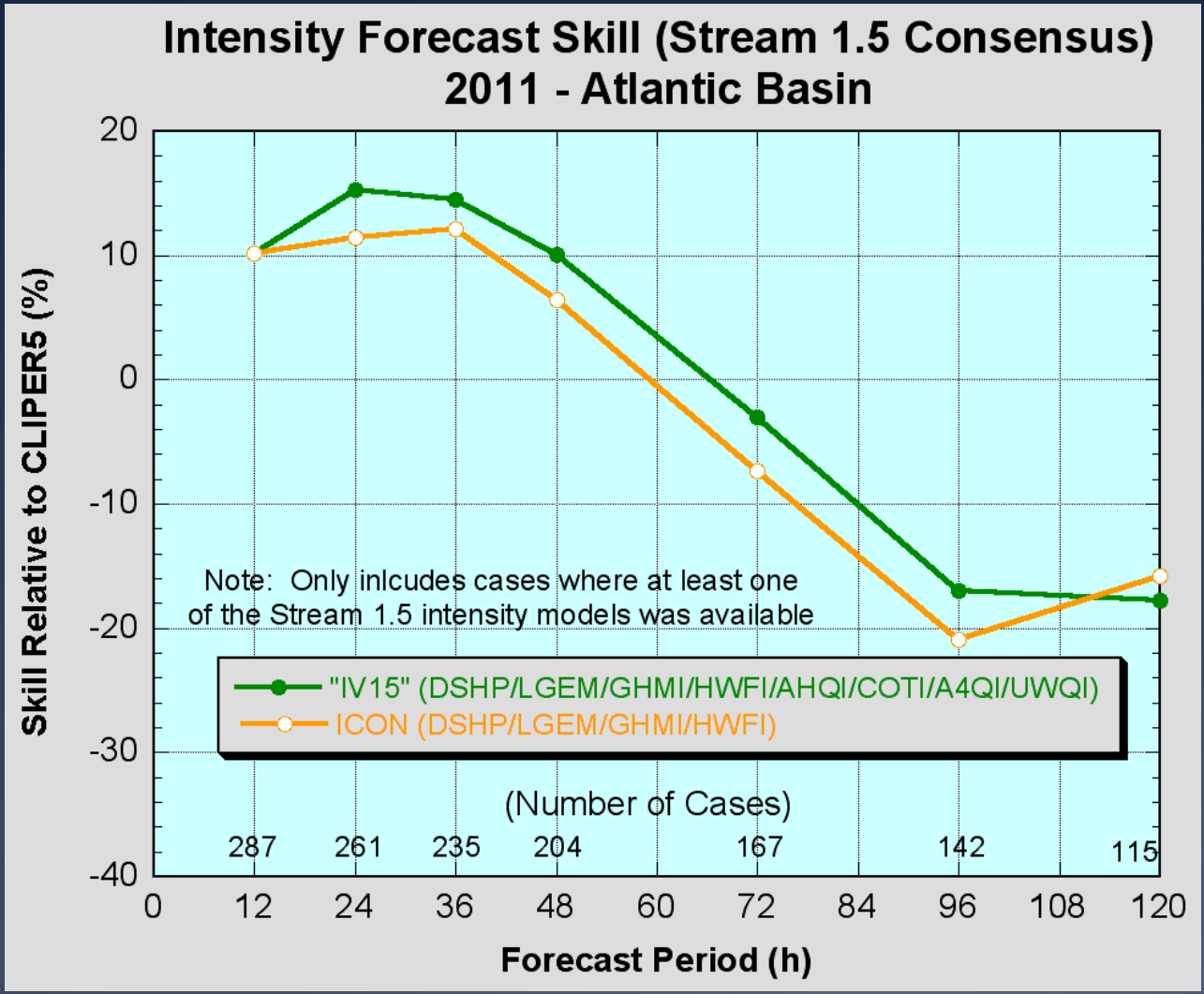


PSU radar runs (A4QI):

Homogeneous (but ridiculously small) sample with the other primary Stream 1.5 intensity models.

Even radar data didn't give errors as low as COTI.

2011 Preliminary Verifications



Stream 1.5 intensity models did contribute positively to the consensus (although the improvements were all less than 1 kt).

Conclusions

- For track, no breakthroughs, but the FIM global model was competitive with the best current operational track guidance.
- For intensity, the consensus aid SPC3 was an improvement over its individual members. COTI performed very well in this season of slowly changing storms.
- Stream 1.5 models did contribute positively to the intensity consensus.
- Allowing storms to move over land in the verification sample makes interpretation of comparative results difficult.