HFIP 2012 Annual Review: Preliminary Forecast Verification

John Cangialosi and James L. Franklin
National Hurricane Center
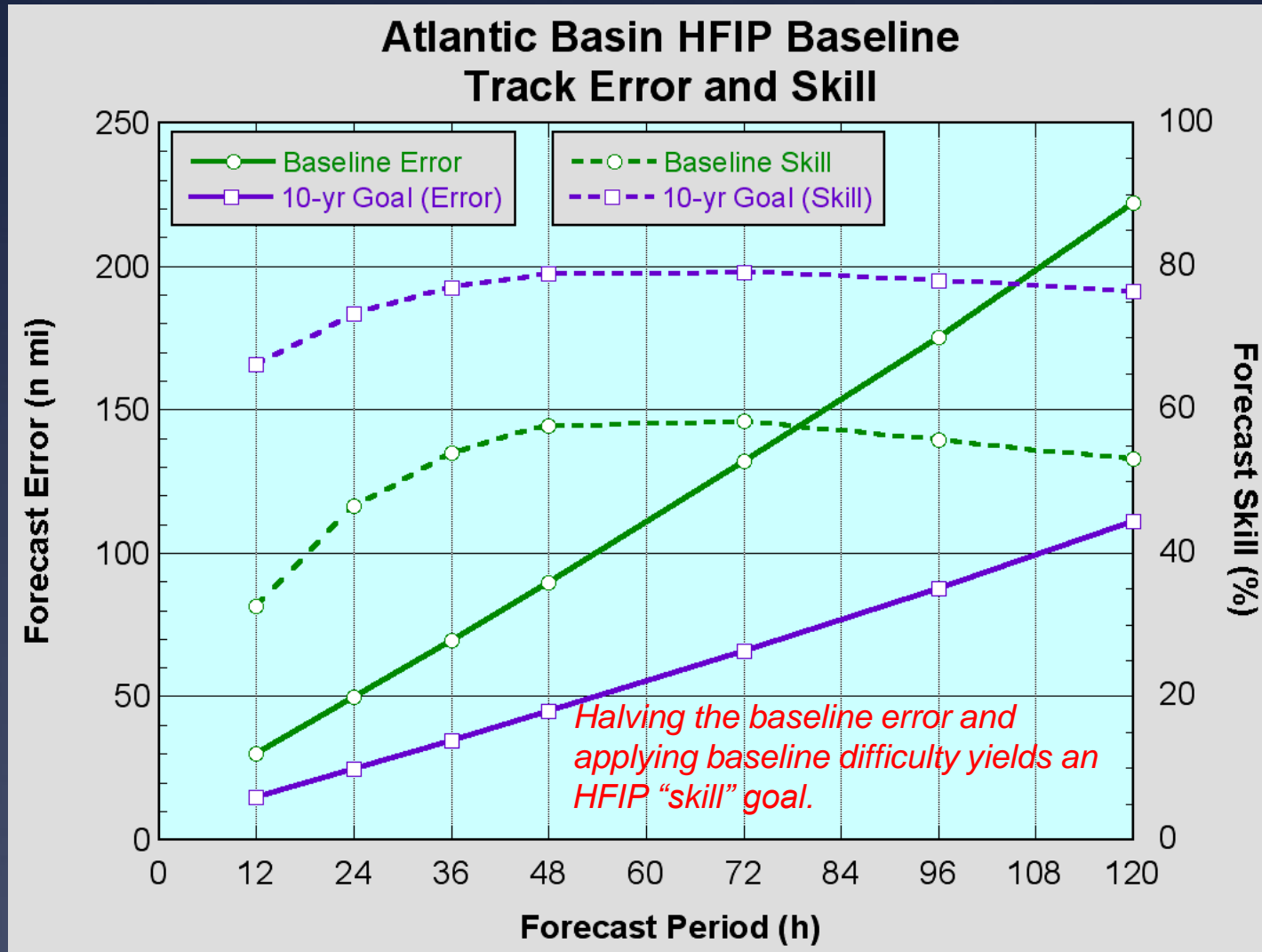
December 2012
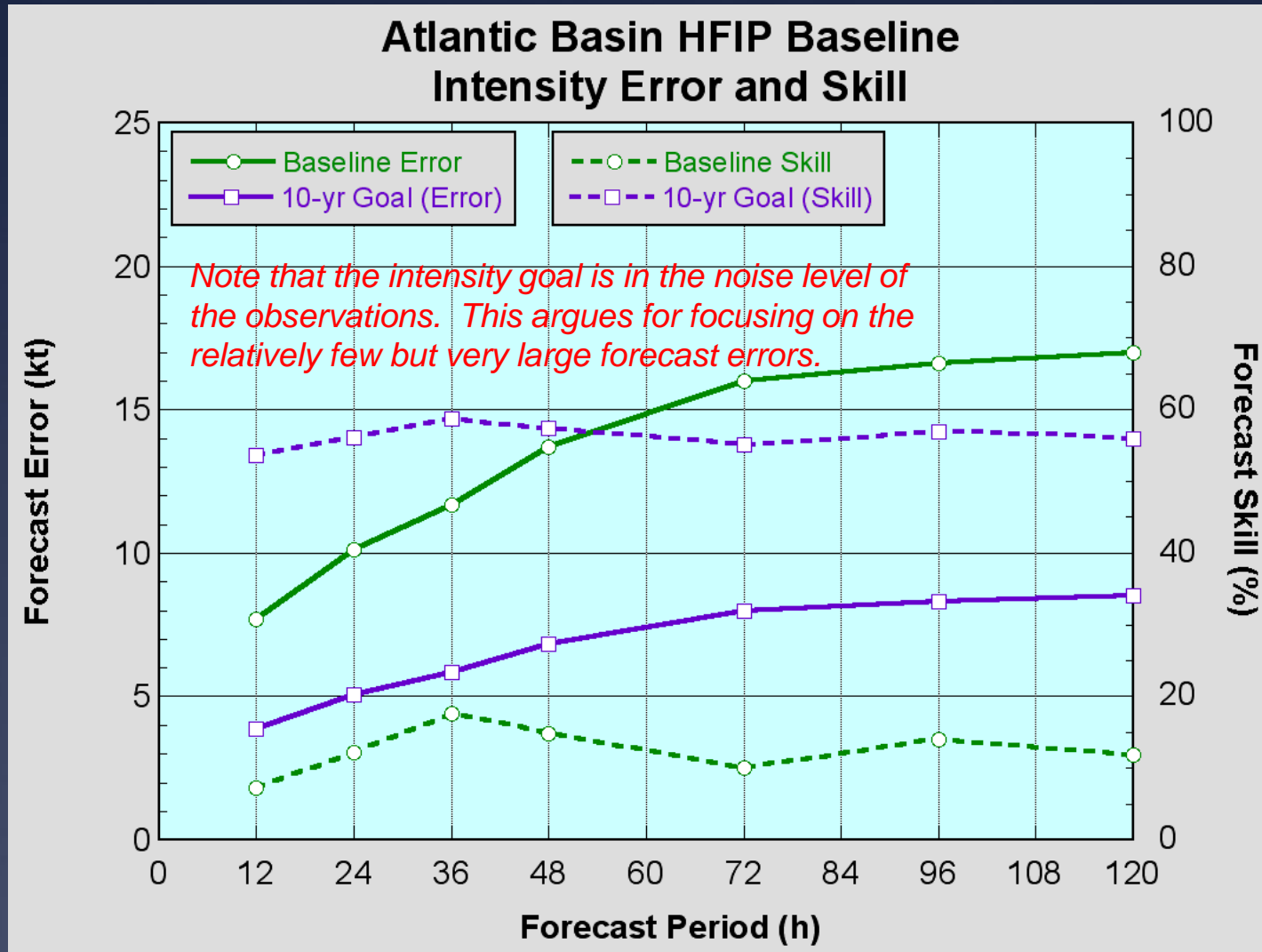
Introductory Remarks

* Evaluations are based on preliminary best track data. 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

Atlantic Basin HFIP Baseline Intensity Error and Skill

Legend: Baseline Error, Baseline Skill, 10-yr Goal (Error), 10-yr Goal (Skill)

Note that the intensity goal is in the noise level of the observations. This argues for focusing on the relatively few but very large forecast errors.

X-axis: Forecast Period (h); Left Y-axis: Forecast Error (kt); Right Y-axis: Forecast Skill (%)

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

2012 HFIP Stream 1.5 Models

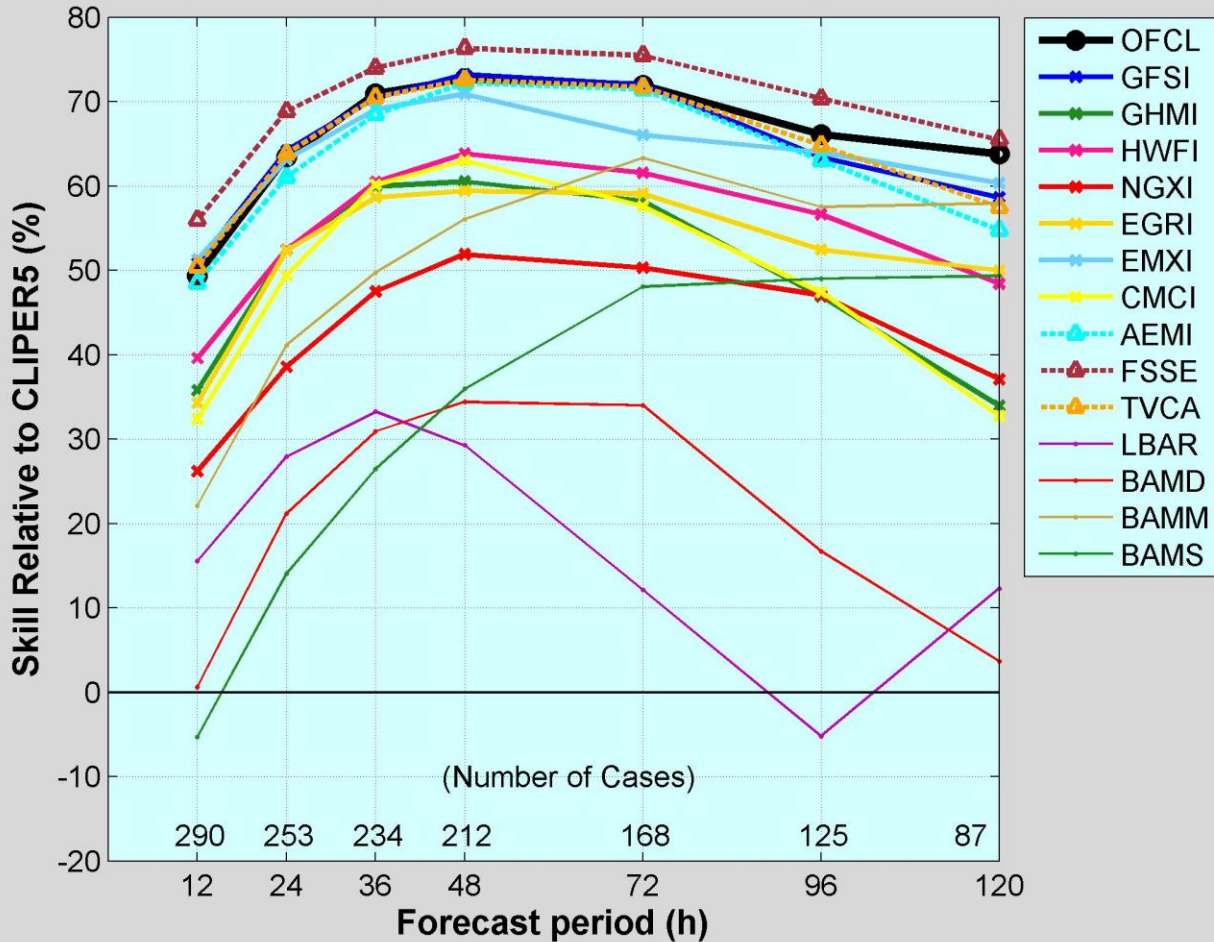* Track:
  * AHWI
  * FM9I
  * GPMI
  * G01I
  * APSI

* Intensity:
  * AHWI
  * COTI
  * APSI
  * UWNI
  * SPC3
  * GPMI
  * G01I

Shading indicates models meant to be used in consensus only.

2012 Preliminary Verifications

Track Forecast Skill (Early Models)
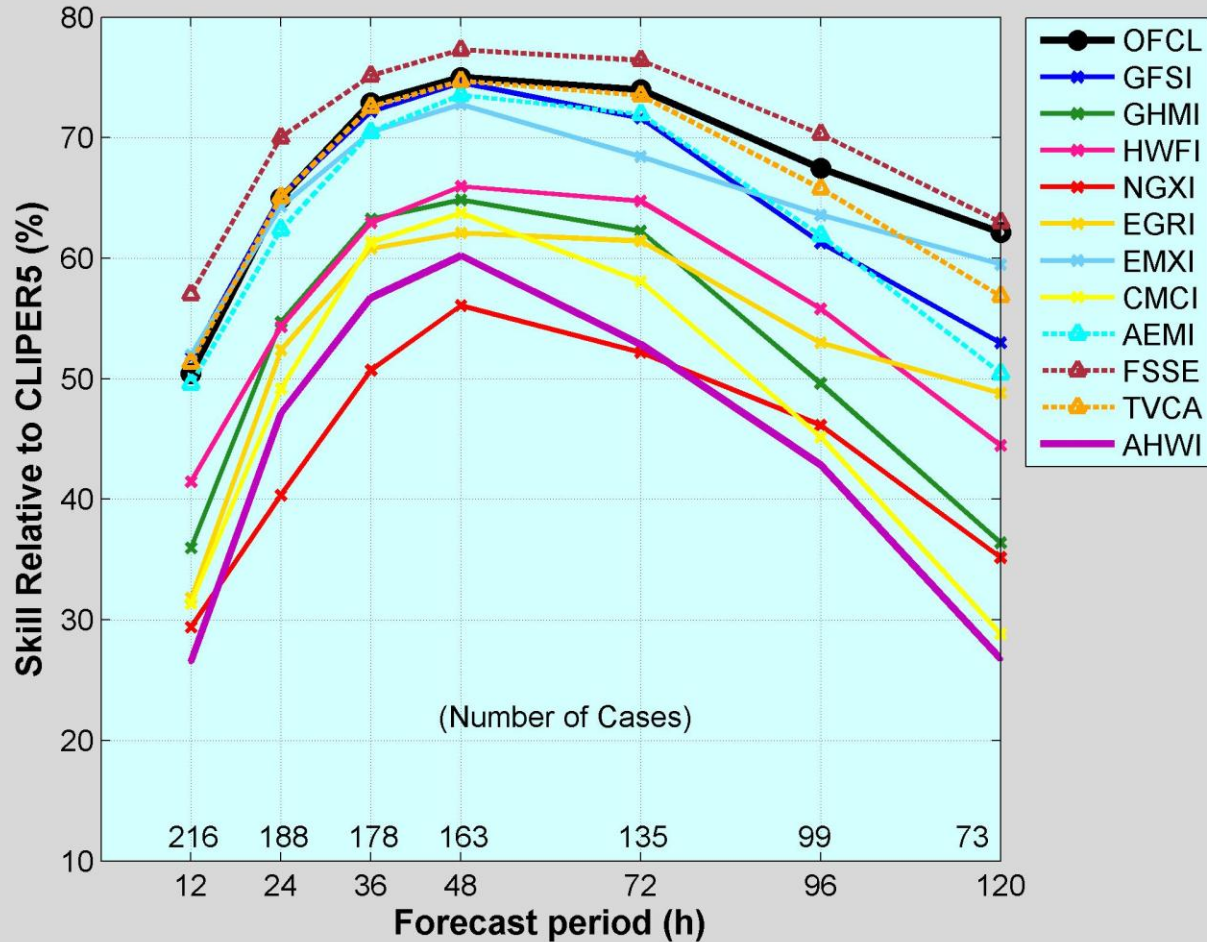**2012 - Atlantic Basin**

Legend: OFCL, GFSI, GHMI, HWFI, NGXI, EGRI, EMXI, CMCI, AEMI, FSSE, TVCA, LBAR, BAMD, BAMM, BAMS

Among the operational dynamical models, GFS was the best performer, with ECMWF close behind.

Second tier comprises the regional models, CMC and UKMET. NOGAPS trails. BAMM beat the more sophisticated members of this group at longer ranges.

FSU Superensemble best consensus model.
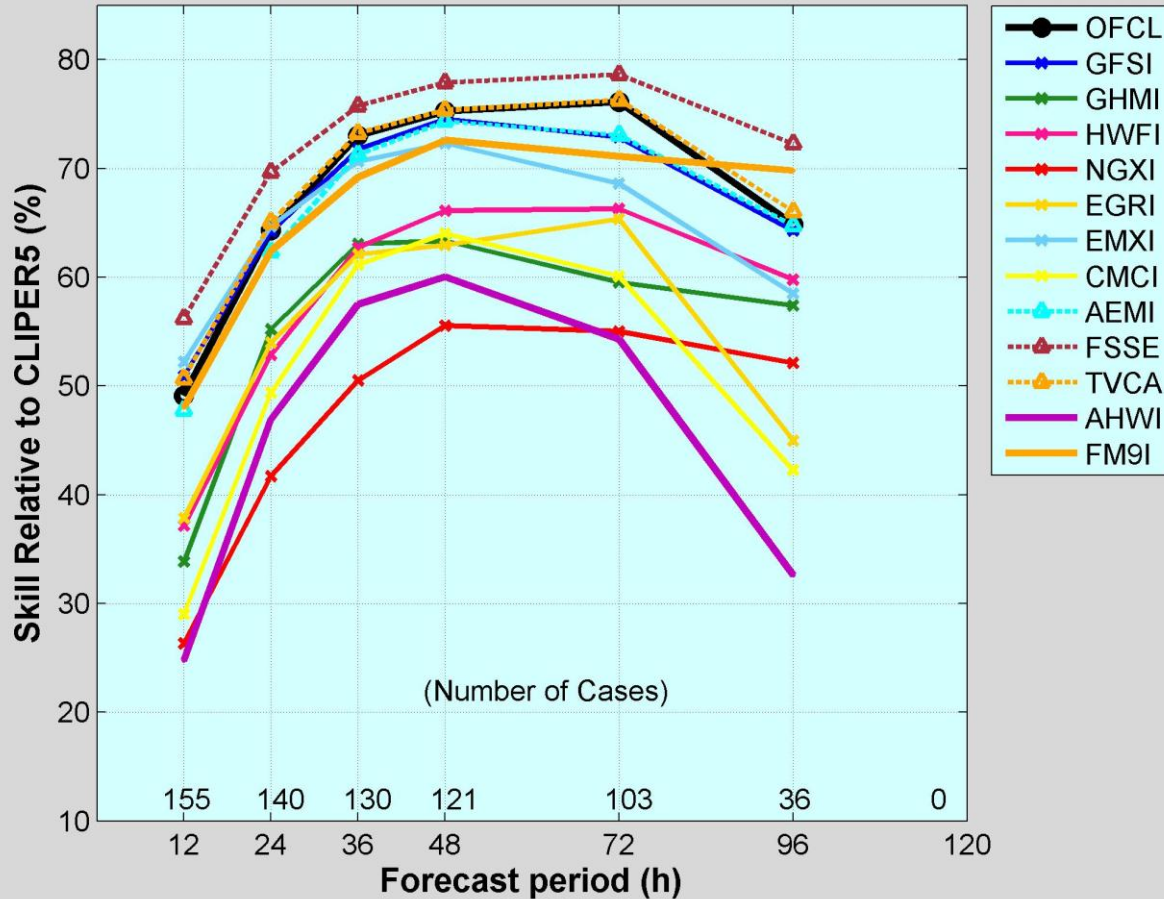
2012 Preliminary Verifications

FM9I was excluded because of the small sample (n=2) at 120 h.

AHWI was not competitive with the best operational models.
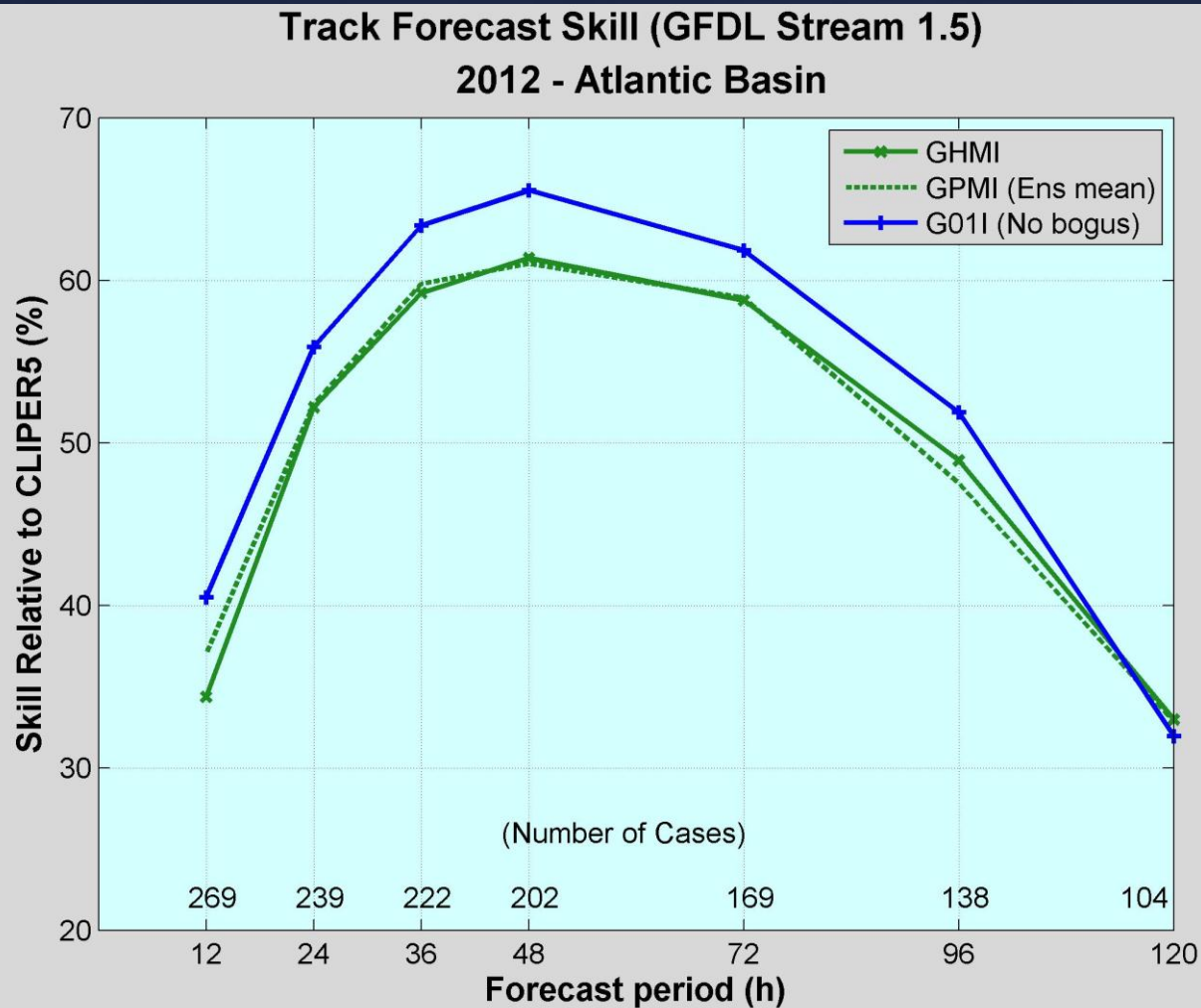
2012 Preliminary Verifications

Track Forecast Skill (Stream 1.5 Early Models) 2012 - Atlantic Basin

Adding the FM9I (note no sample at 120 h).

For this sample, FM9I was a good performer, better than the other Stream 1.5 models and close to the skill of the best operational models.
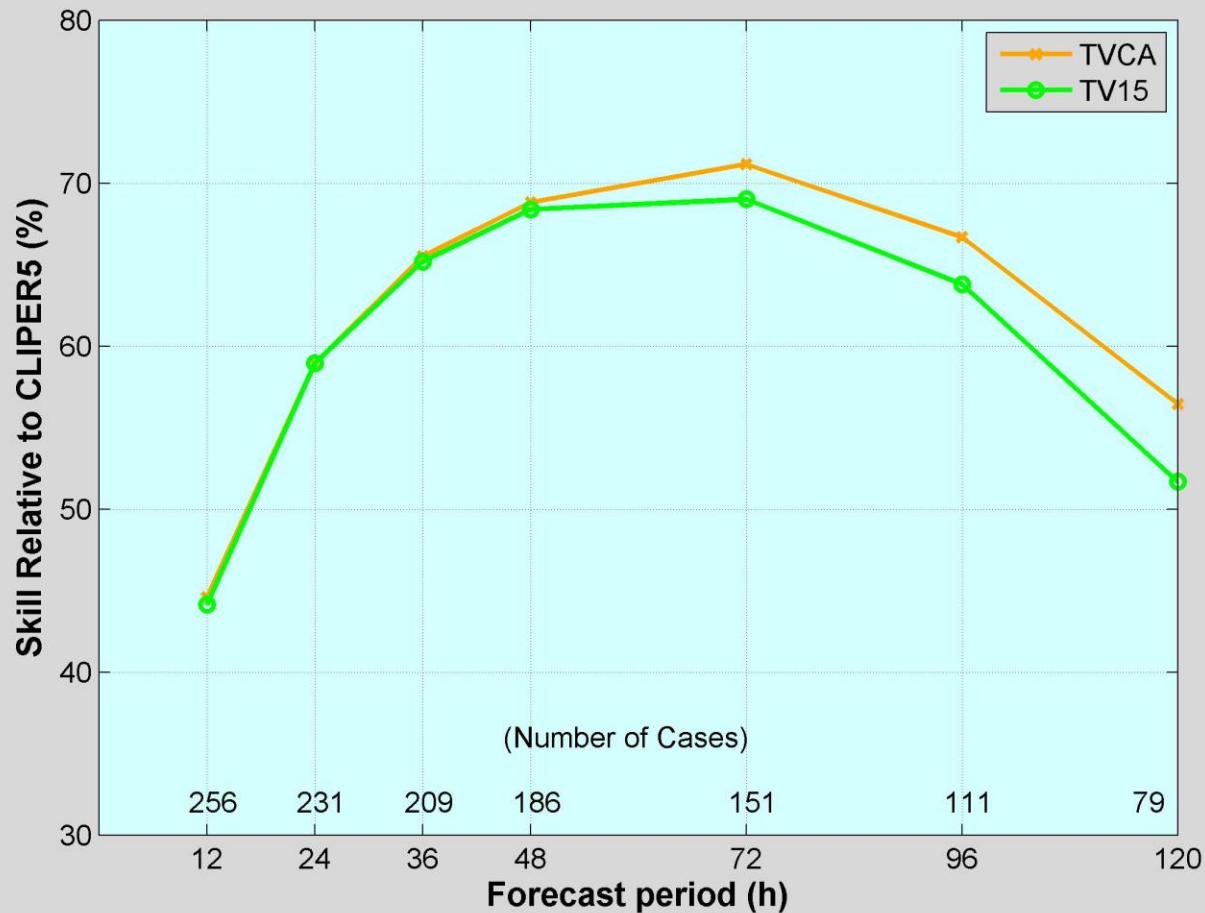
2012 Preliminary Verifications

Track Forecast Skill (GFDL Stream 1.5)
2012 - Atlantic Basin

Legend:
- GHMI
- GPMI (Ens mean)
- G01I (No bogus)

GFDL ensemble mean was very similar to the control.

Unbogused ensemble member had a little more skill than the control.
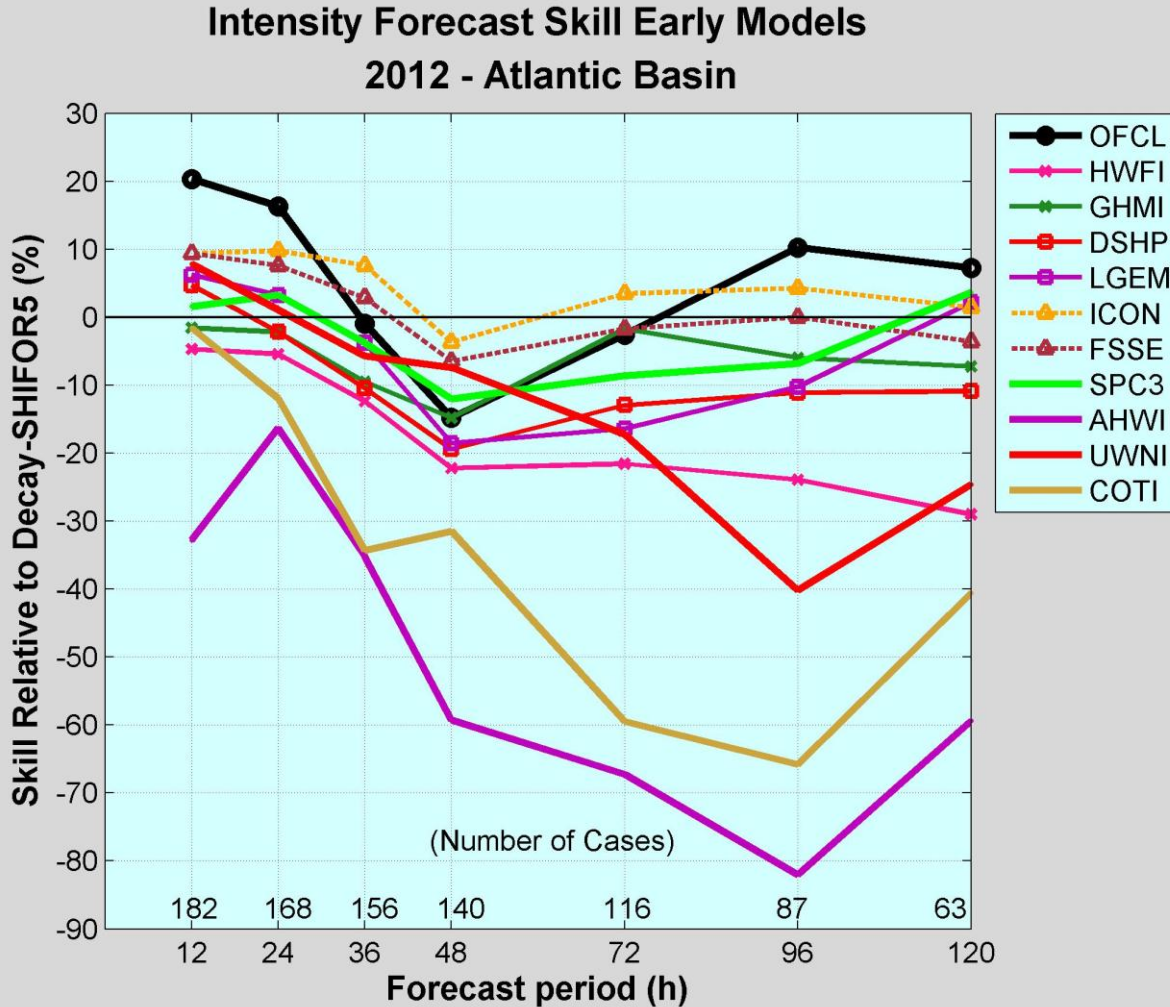
2012 Preliminary Verifications

Track Forecast Skill (Stream 1.5 Consensus)
2012 - Atlantic Basin

Evaluated only for cases where there was a Stream 1.5 track model.

The skill of TV15 was similar to TVCA through 48 h, then slightly worse.

2012 Preliminary Verifications

Intensity Forecast Skill Early Models
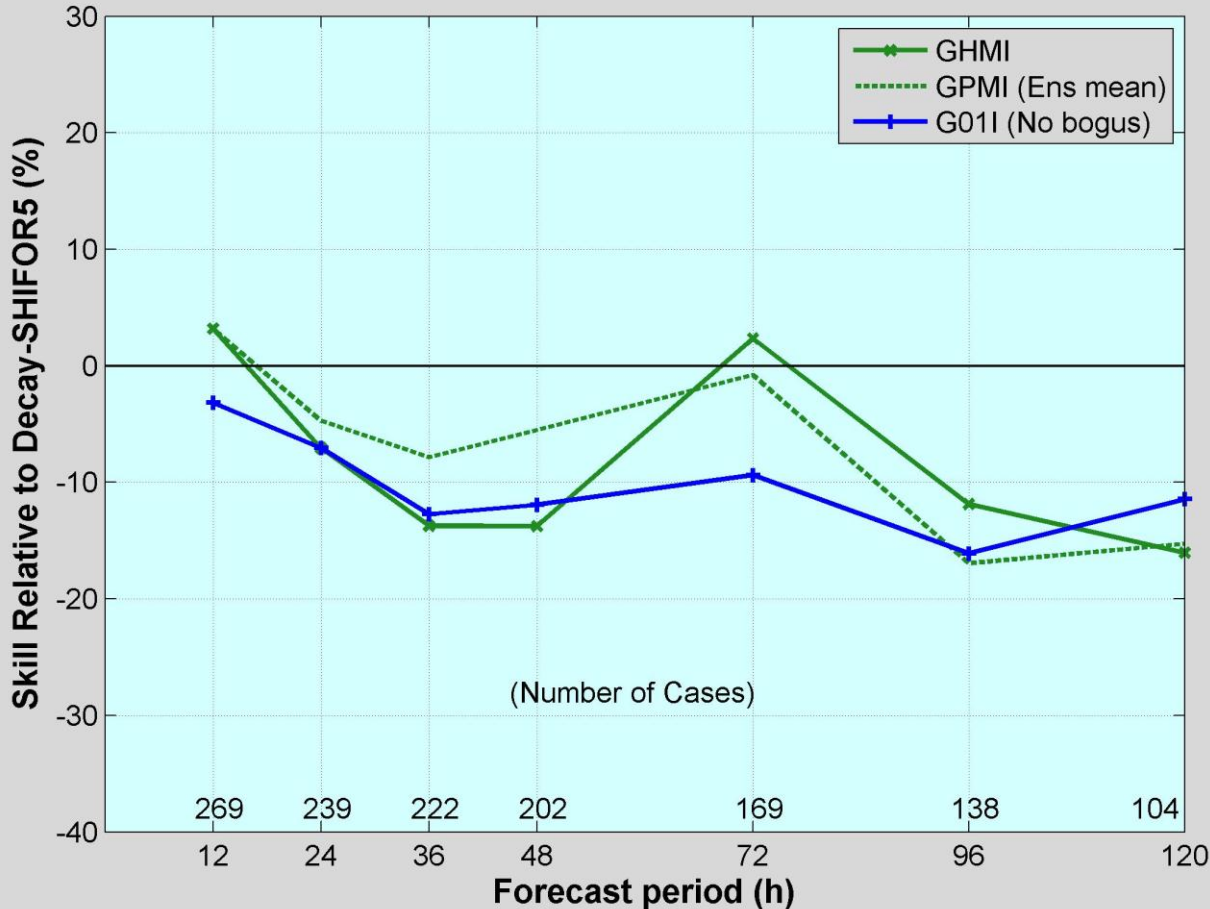2012 - Atlantic Basin

A4PI (radar) omitted due to sample size.

UWNI, COTI were poor performers and much worse than all of the operational models.

SPC3 beat DSHP and LGEM but still had little skill.

2012 Preliminary Verifications

Intensity Forecast Skill (GFDL Stream 1.5)
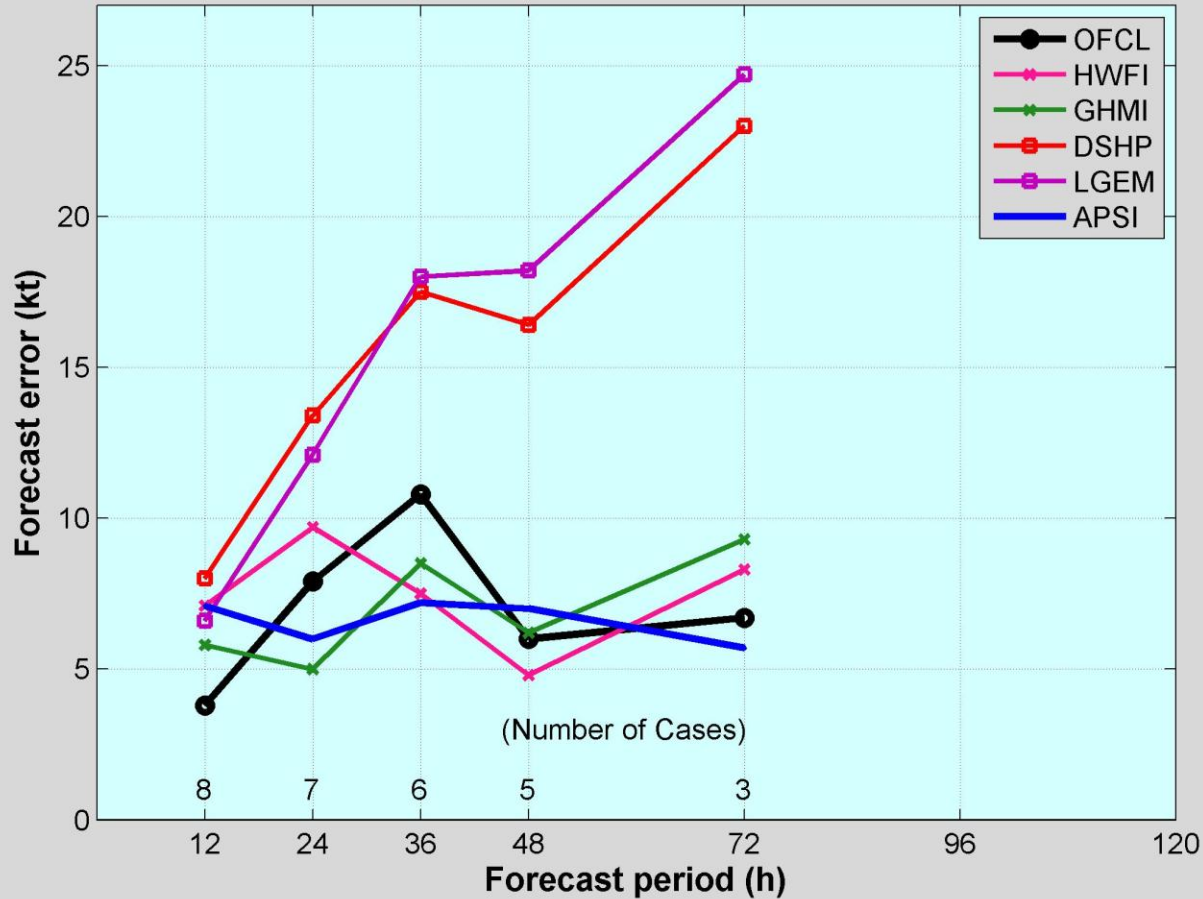**2012 - Atlantic Basin**

GFDL mostly was not skillful.

GFDL ensemble mean not consistently better than the control (better early, worse late).

Unbogused ensemble member a little worse than the control.

2012 Preliminary Verifications

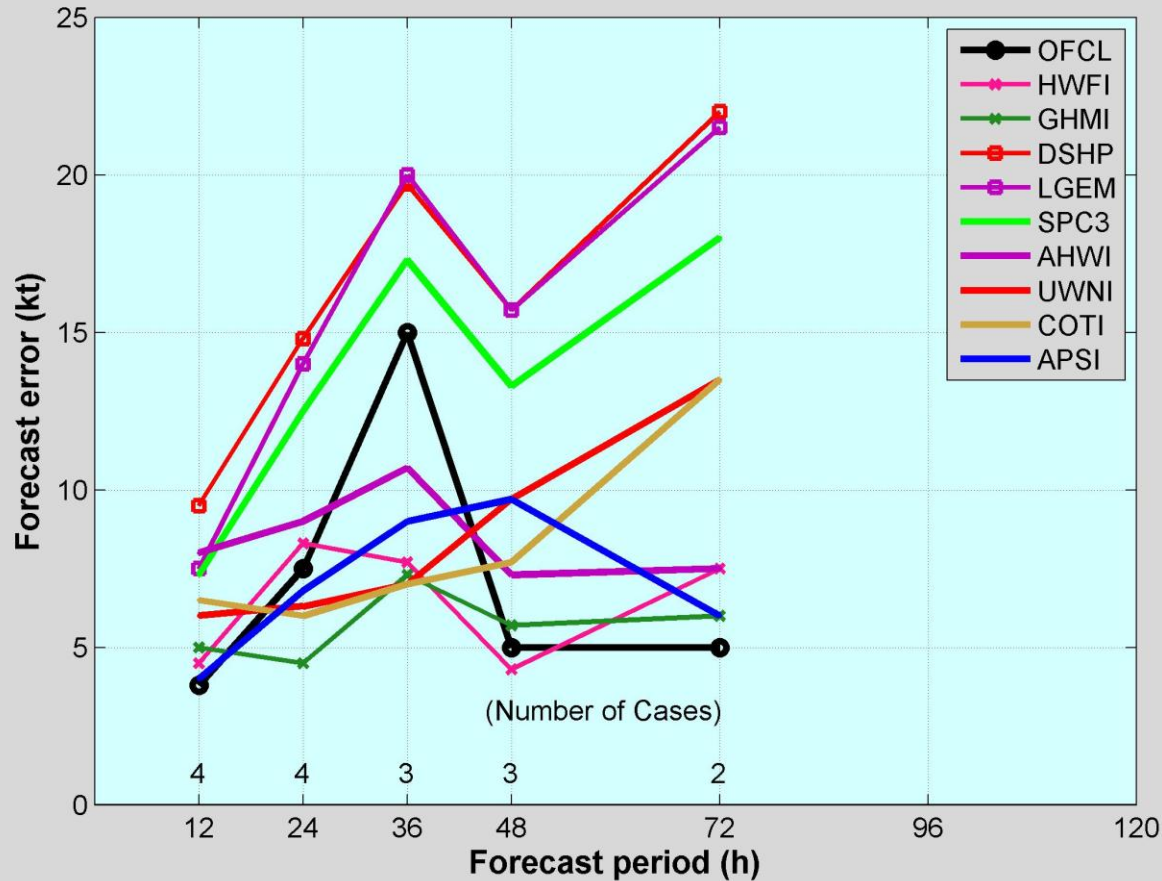Intensity Forecast Errors (Stream 1.5 Early Models)
2012 - Atlantic Basin

PSU radar runs (APSI):

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

Although APSI was much better than the statistical guidance, it was similar to the dynamical guidance.

2012 Preliminary Verifications

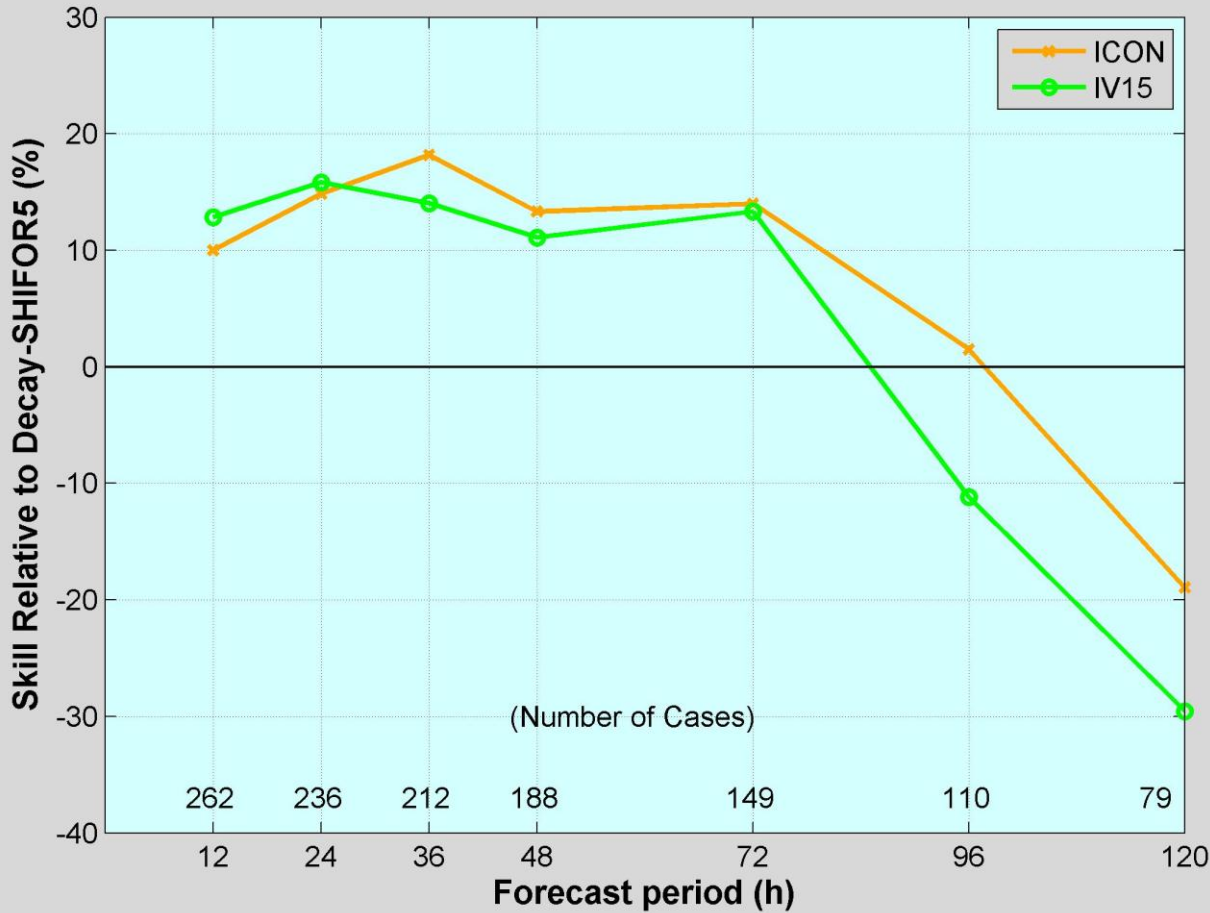Intensity Forecast Errors (Stream 1.5 Early Models)
2012 - Atlantic Basin

Adding the rest of the Stream 1.5 model suite.

Homogeneous (but ridiculously small) sample.

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

2012 Preliminary Verifications

Intensity Forecast Skill (Stream 1.5 Consensus)
2012 - Atlantic Basin

Stream 1.5 intensity models did contribute positively (although improvements were tiny) to the consensus early, but degraded it late.

Conclusions

* For track, no breakthroughs, but the FM9I global model was competitive with the best current operational track guidance.

* For intensity, the consensus aid SPC3 was an improvement over its individual members. The other Stream 1.5 models generally performed poorly.

* Stream 1.5 models did contribute positively to the intensity consensus early.