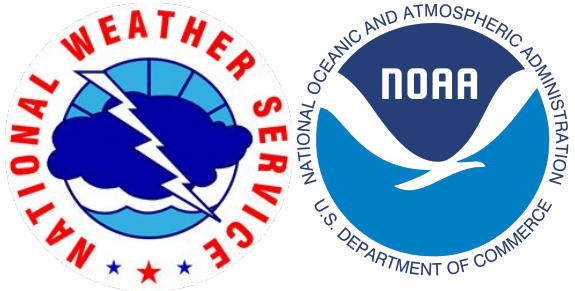
CPC Evaluation of GEFSv12

Mingyue Chen
Mike Charles
Lindsey Long
Craig Long
Kyle MacRitchie
Hui Wang
Matt Rosencrans

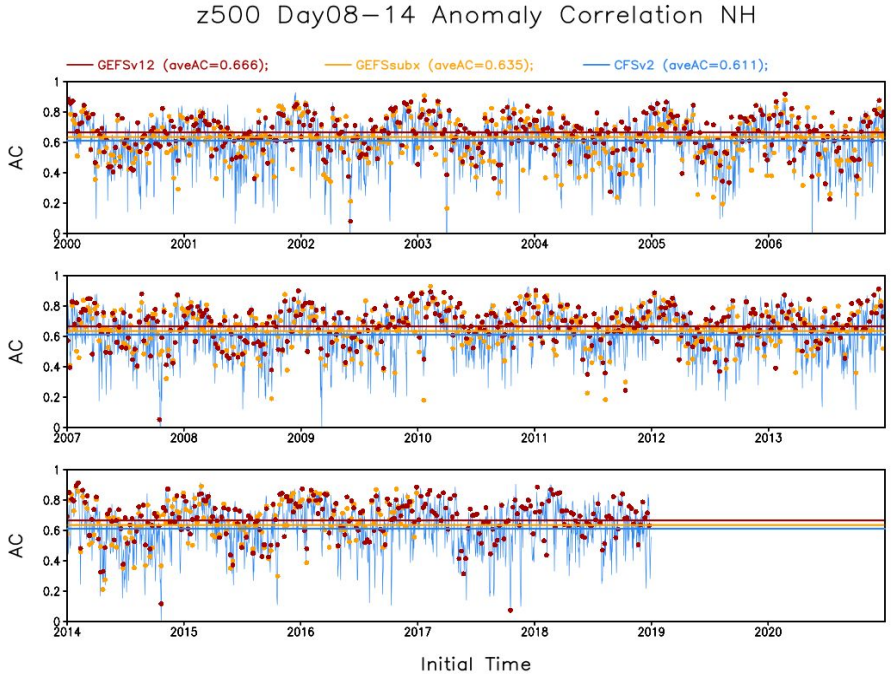


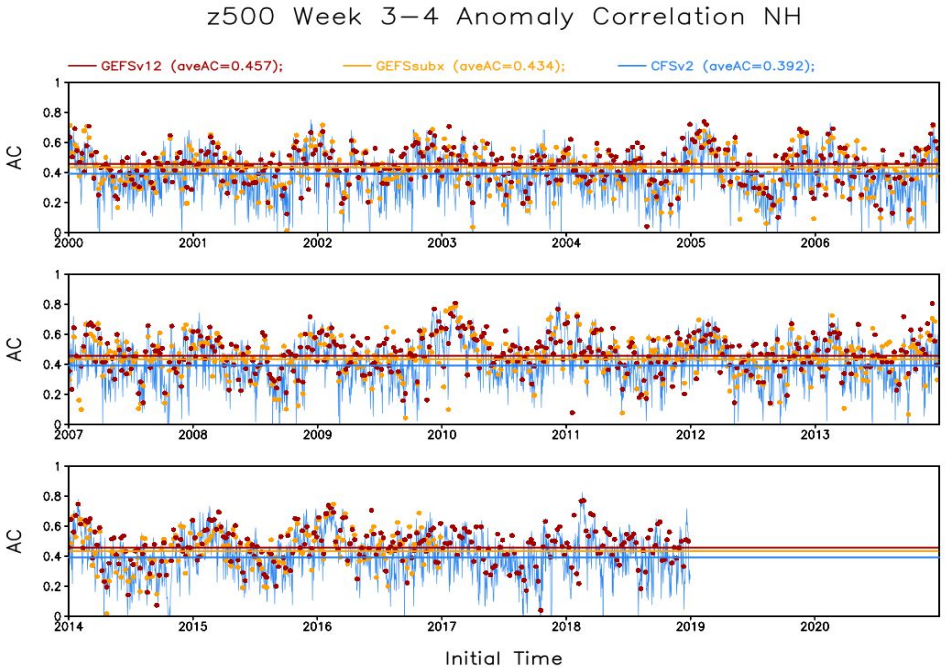
Constraints

- Week - 2
  - Most of GEFSv12 retrospective unusable for CPC (stop at fhr 240)
    - Used last years of GEFSv12 reforecast data instead. Adv: V12
    - Subset of 5 out of 21 GEFSv10 real-time members was used to make the two datasets more homogeneous.
- Week - 3/4
  - No major issues noted

500 hPa Heights

Mingyue Chen

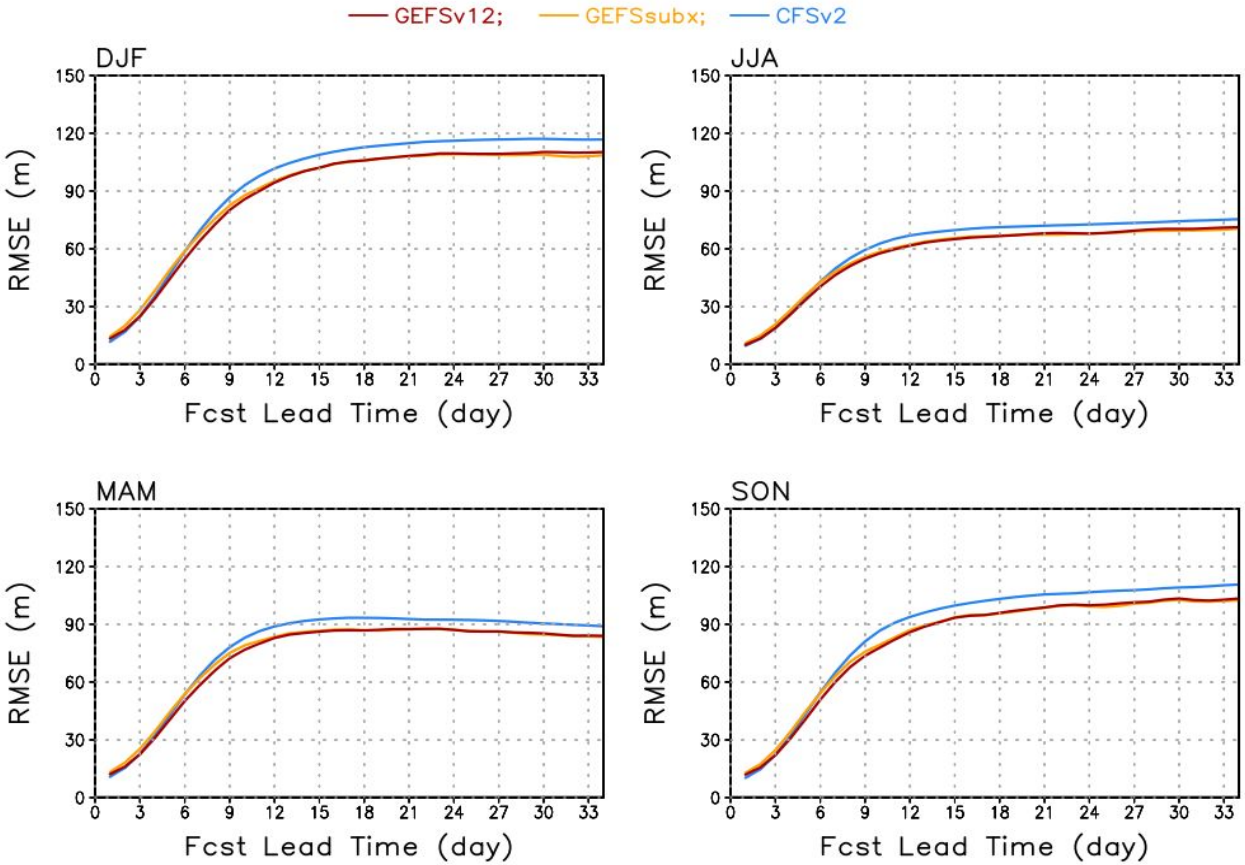


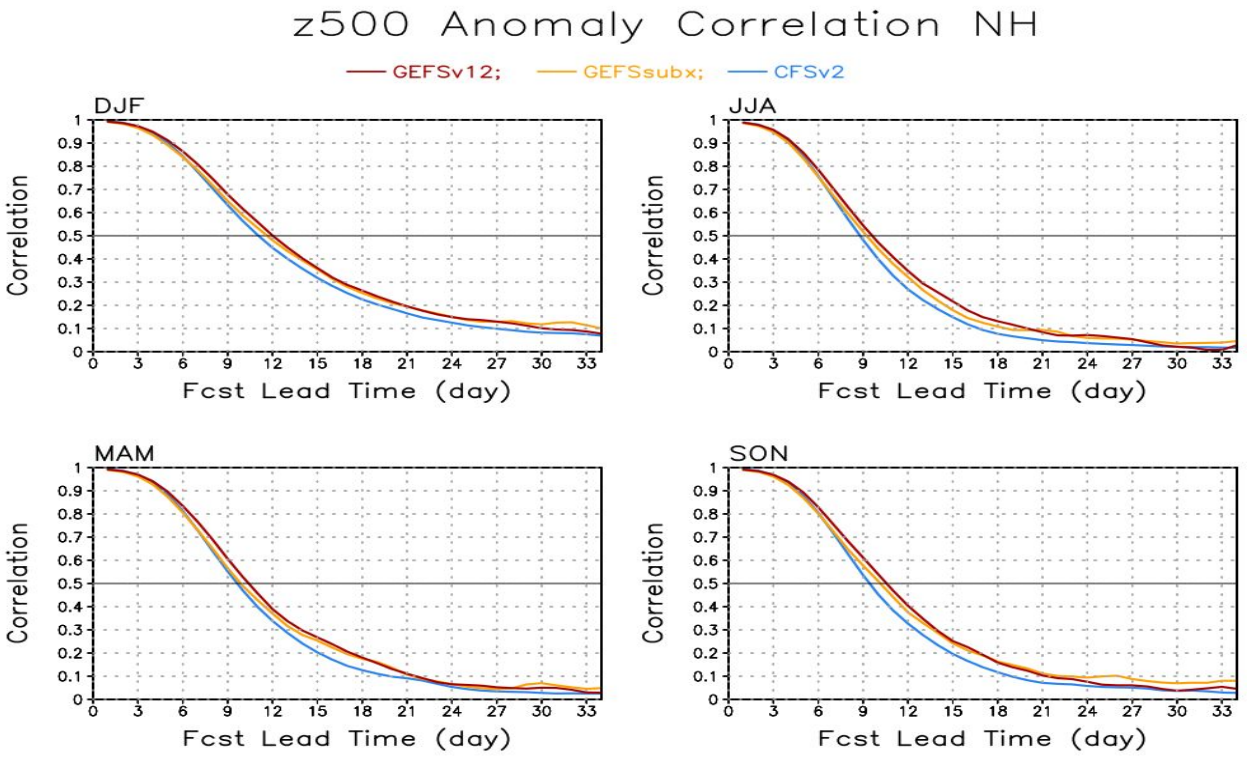


GEFSv12	0.666
GEFSSubX	0.635
CFSv2	0.611

GEFSv12	0.457
GEFSSubX	0.434
CFSv2	0.392

z500 RMSE(m) NH



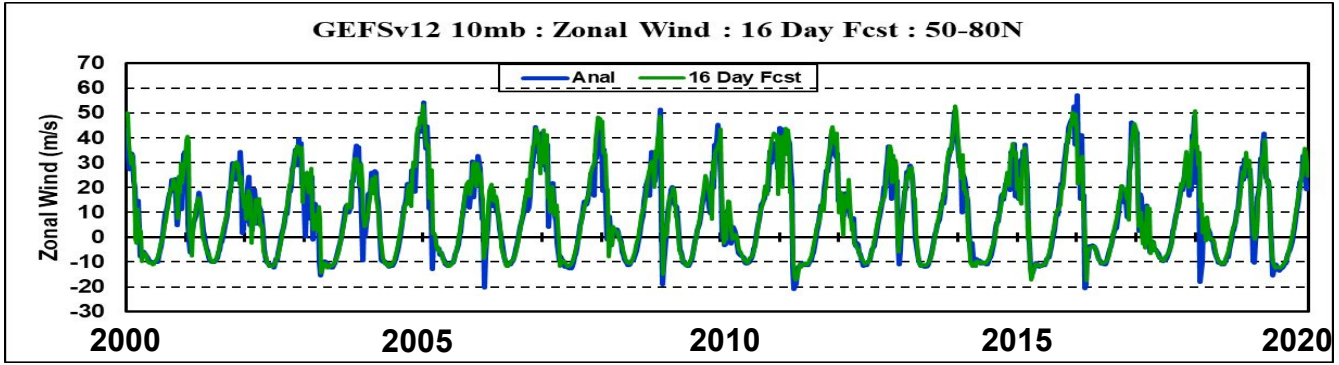


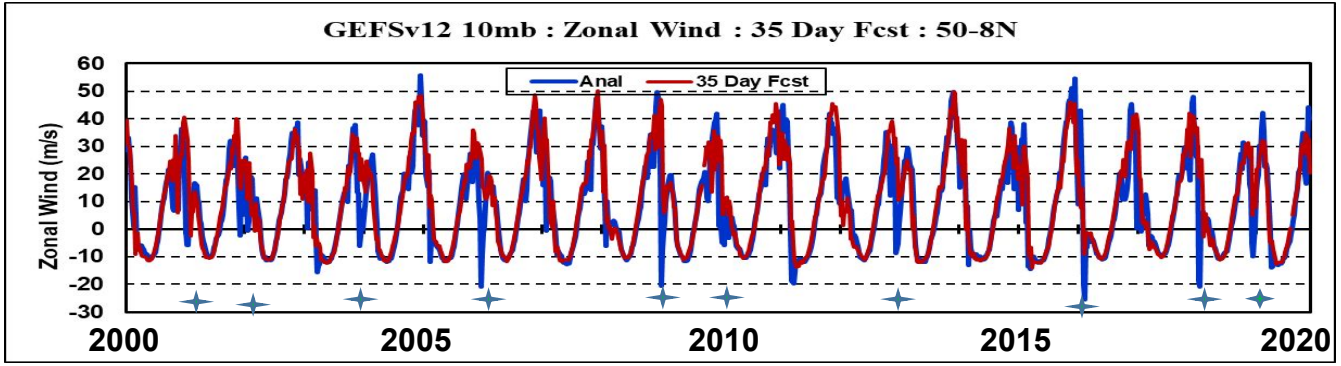
Day which AC=0.5	NH				PNA			
	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
GEFSv12	12	10	9	10	12	10	9	10
GEFSsubx	11	9	9	10	12	9	8	10
CFSv2	11	9	8	9	11	9	8	9

Stratosphere

Craig Long

35 Day ReFcsts : NH Polar Jet : 10mb





Both 10 & 50 mb time series show:

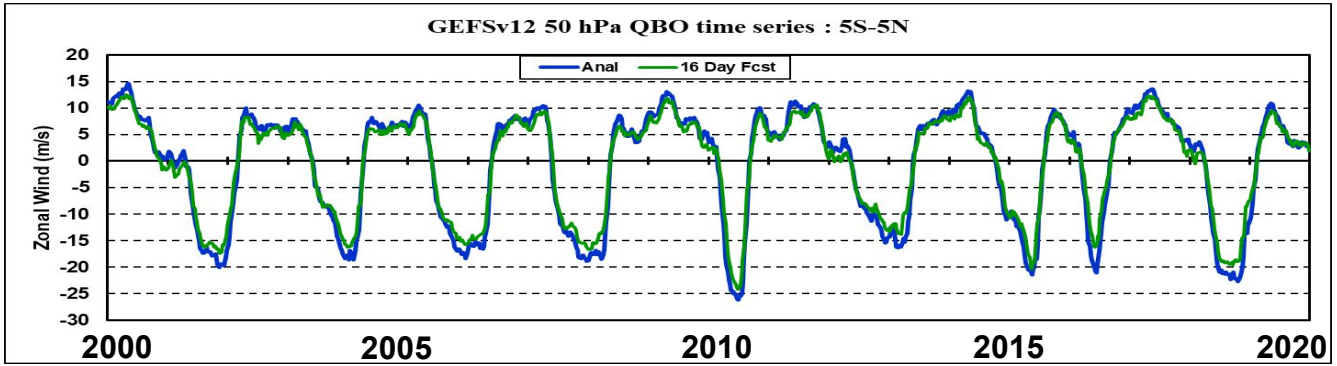
- Winter max winds are under fcst
- Sudden warmings (wind decelerate/reverse) under forecast/not forecasted

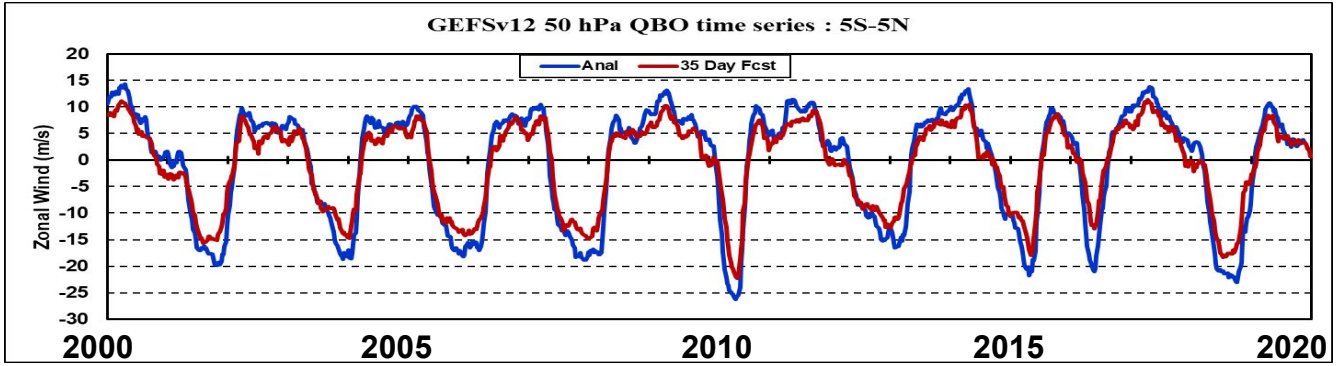
Similar time series set up as NH Polar Temps:

- Analysis and 16 day fcsts are shown in top plot.
- analysis and 35 day fcsts are shown in bottom plot.

- During Sudden Warmings polar winds will decelerate and may become Easterly for a short period of time.
- Commonly, 60N at 10 mb is used to denote a Major SSW if the winds become Easterly (denoted by ✦).

35 Day ReFcsts : QBO zonal winds : 50mb





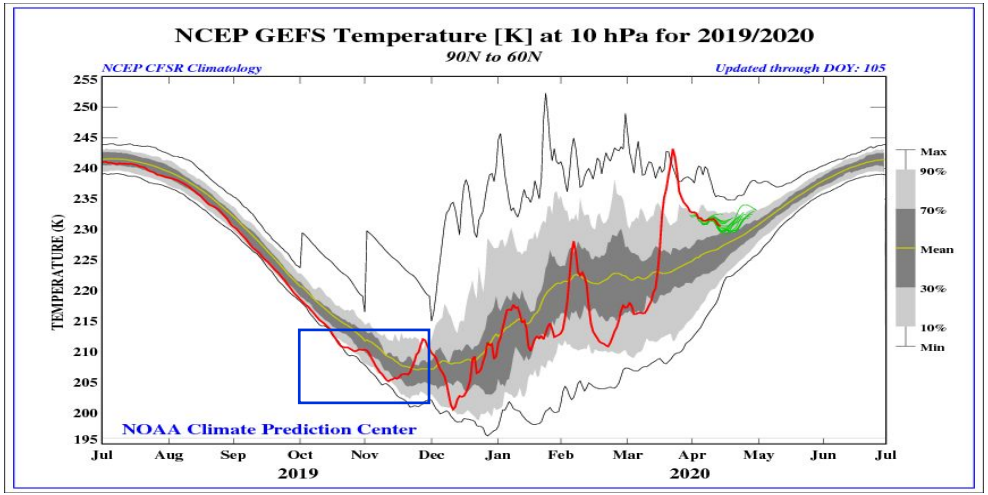
Retention of QBO structure is good. Some S2S models relax their QBO winds to easterlies by 35 days.

Notes:

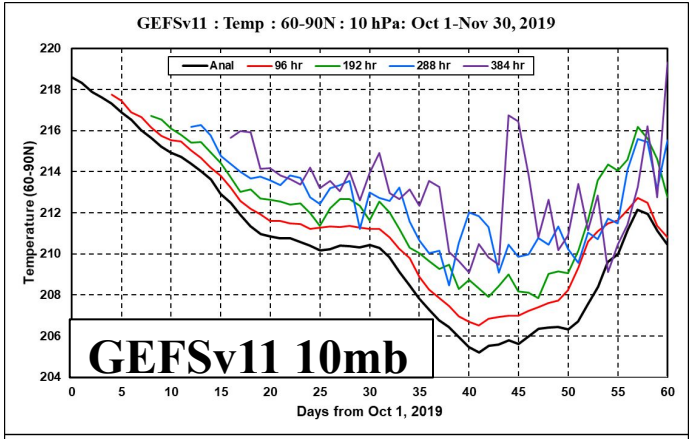
- Capturing QBO state well preserved out to 35 days.
- Westerlies become more under forecast with time which did not happen at 10mb.
- Easterlies also become more under-forecast with time.

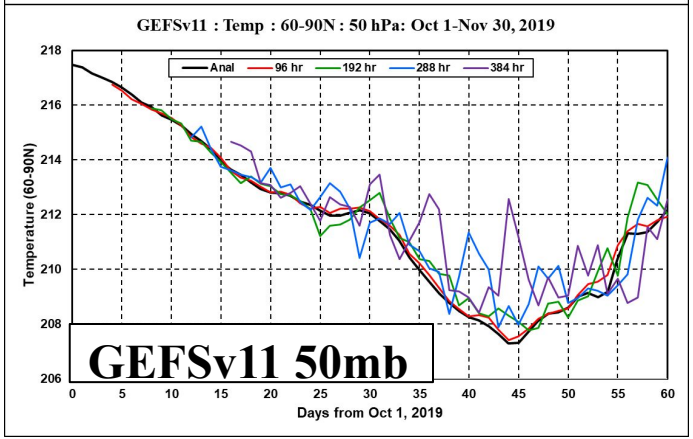
Evaluation of Retrospectives

- In Oct and Nov of 2019 NH polar temps progressively became colder but at the end of Nov there was a period of warming.
- Daily analysis and forecasts (4, 8, 12, & 16 day) are examined to see how well this event was captured.

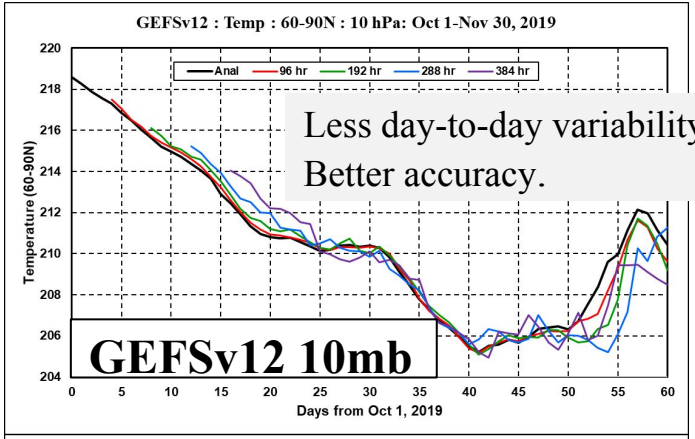


Retrospectives : Temperature : 60-90N

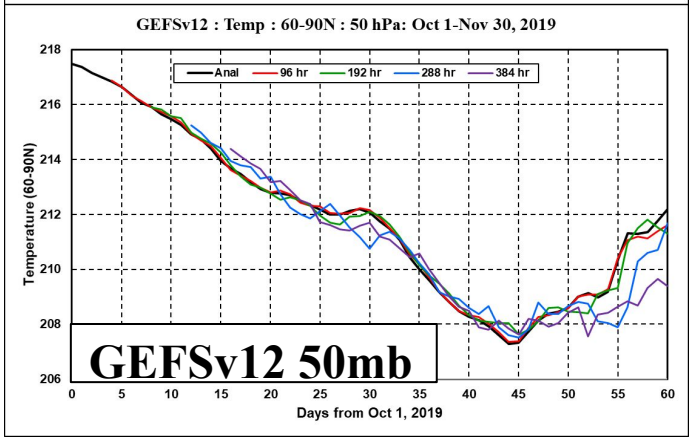




Anl
96 hr
192 hr
288 hr
384 hr



Less day-to-day variability.
Better accuracy.



"Vast" improvement of GEFSv12 12 & 16 day fcsts wrt GEFSv11

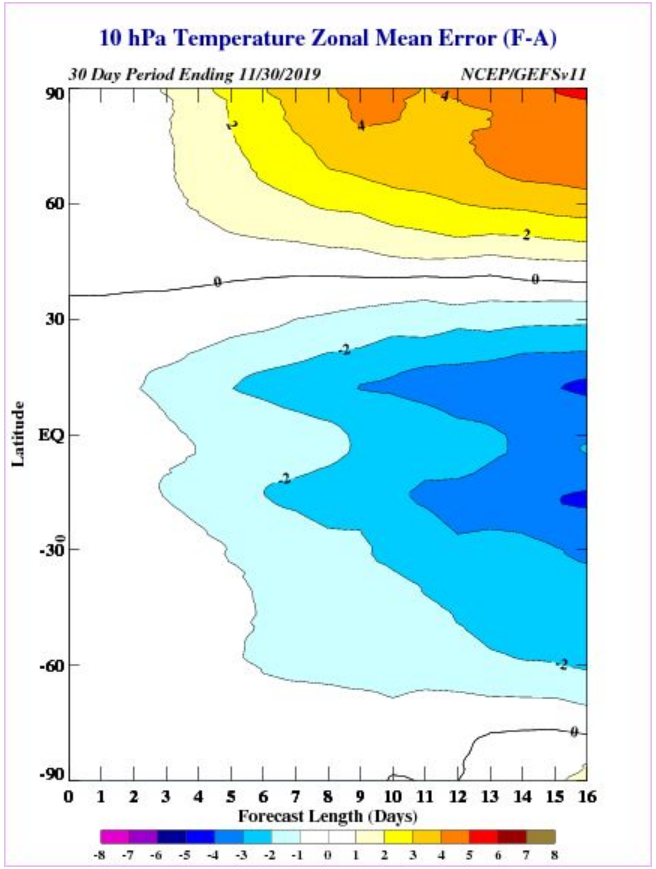
Next slides show zonal mean forecast errors from day 1 to day 16 for the month of November 2019.

Shown are T and u at pressure levels 10, 50, and 100 mb.

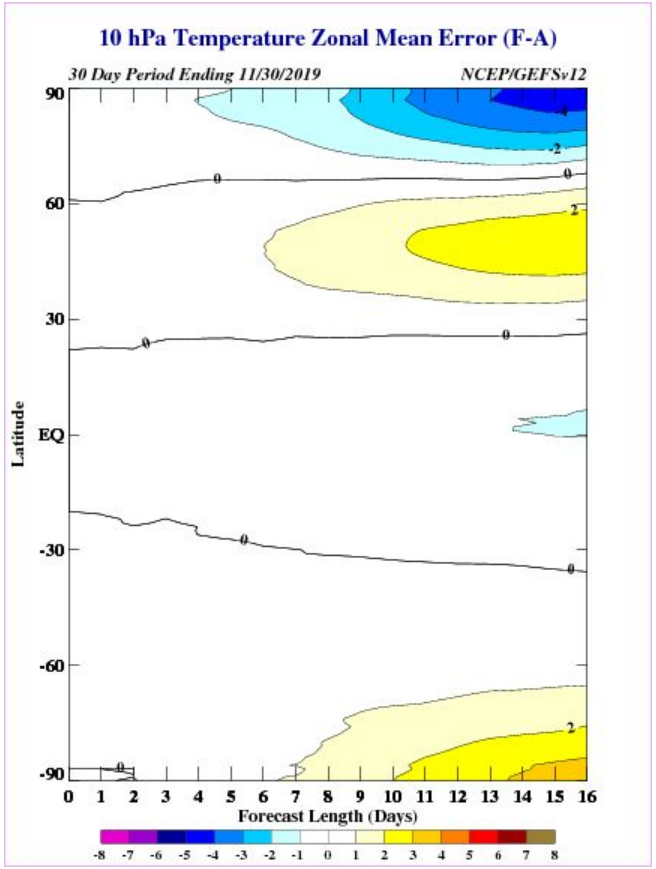
All slides show that the monthly mean forecast errors for GEFSv12 are smaller at most latitudes than GEFSv11.

10 hPa Temperature Fcst Errors
30 day errors ending Nov 30, 2019

GEFSv11

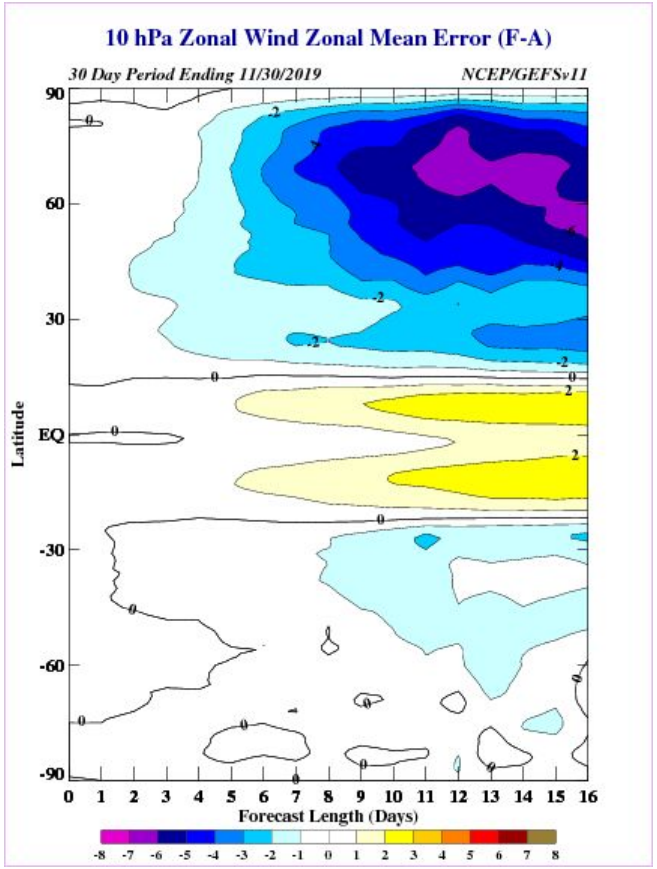


GEFSv12 Retro

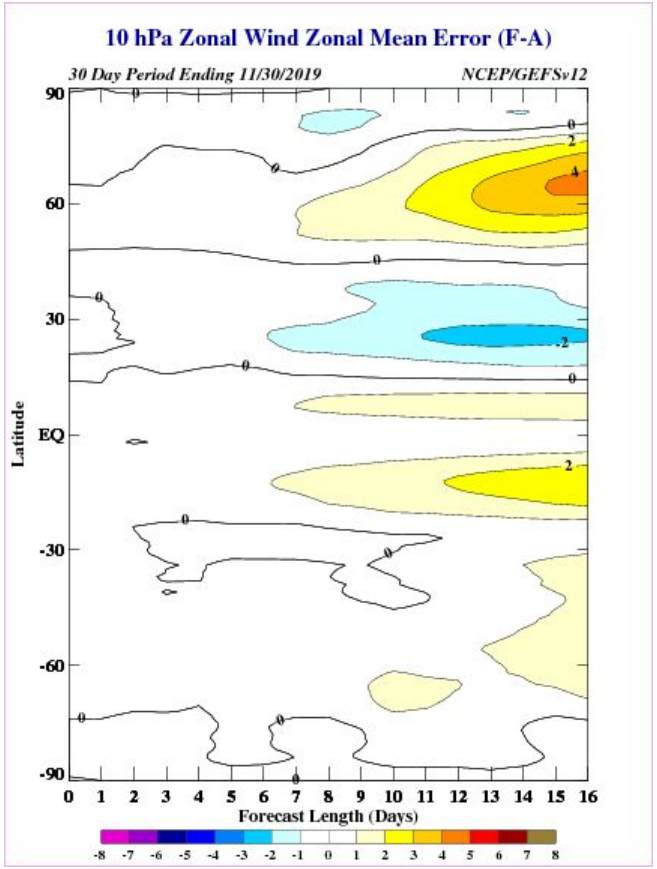


10 hPa Zonal Wind Fcst Errors
30 day errors ending Nov 30, 2019

GEFSv11

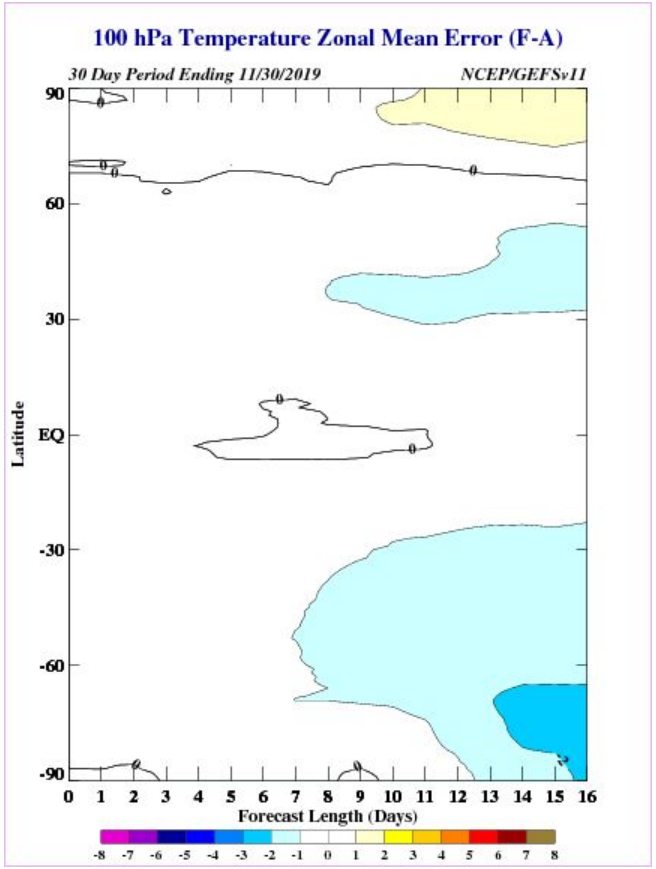


GEFSv12 Retro

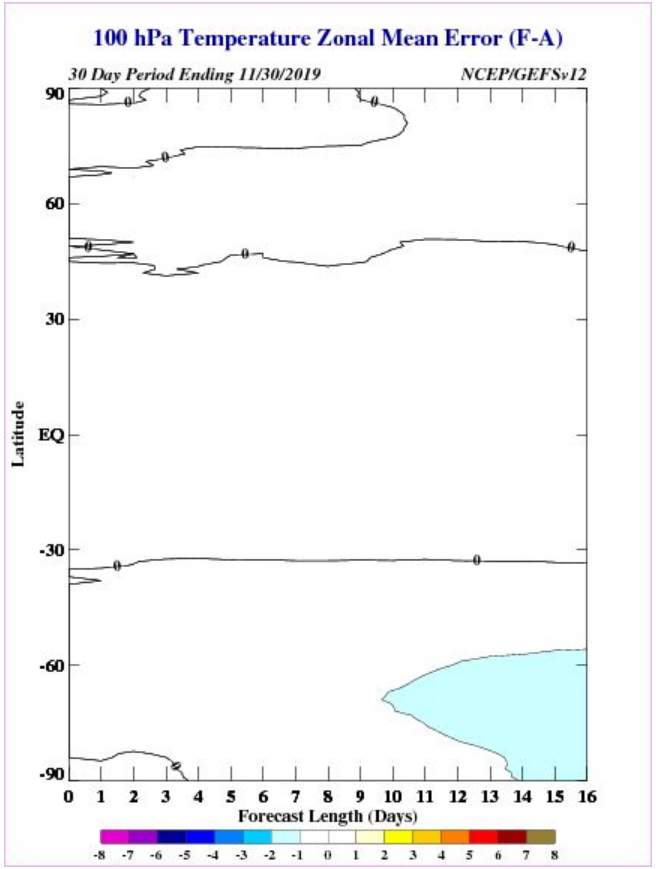


100 hPa Temperature Fcst Errors
30 day errors ending Nov 30, 2019

GEFSv11

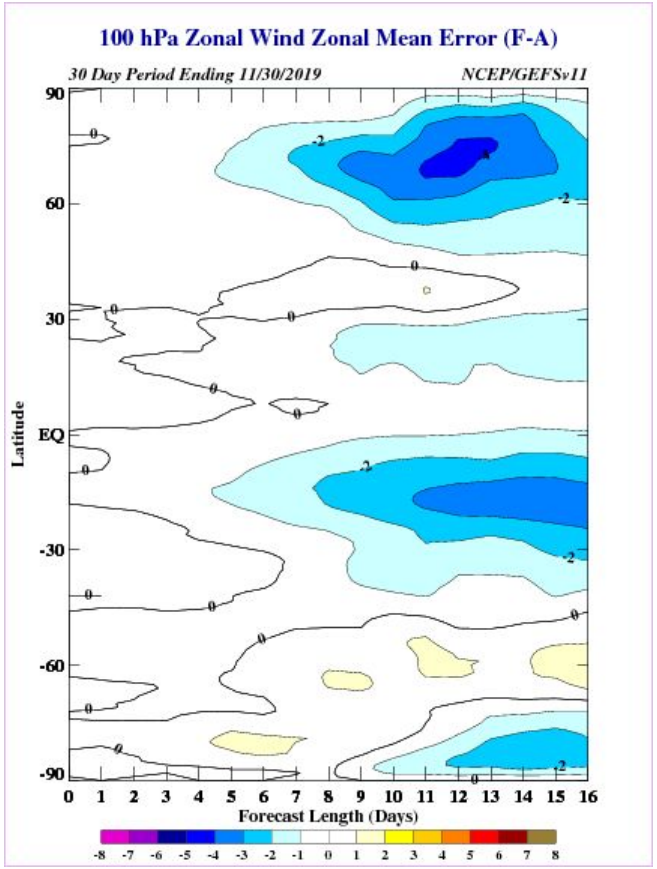


GEFSv12 Retro

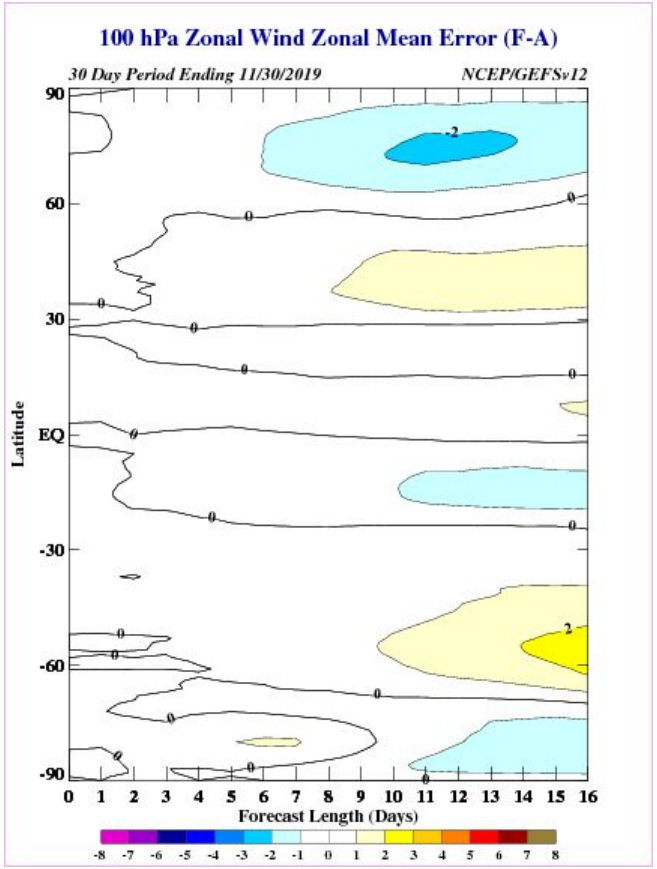


100 hPa Zonal Wind Fcst Errors
30 day errors ending Nov 30, 2019

GEFSv11

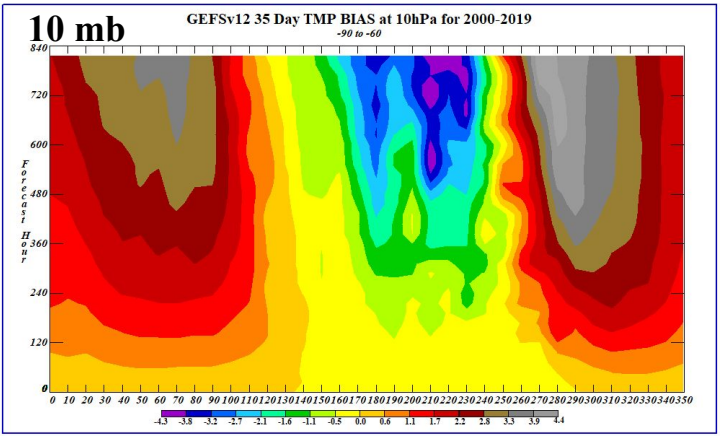


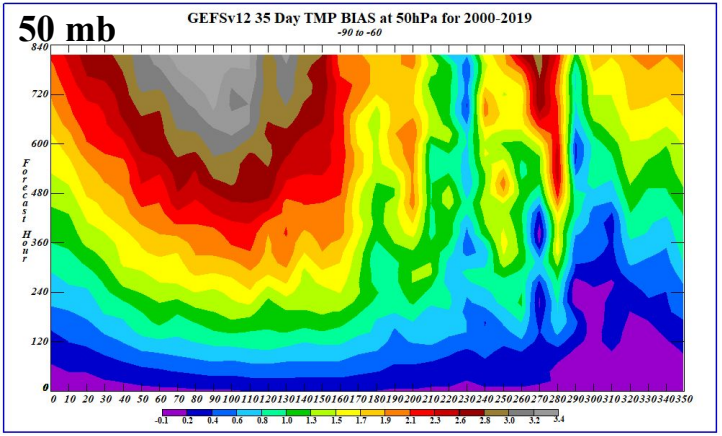
GEFSv12 Retro



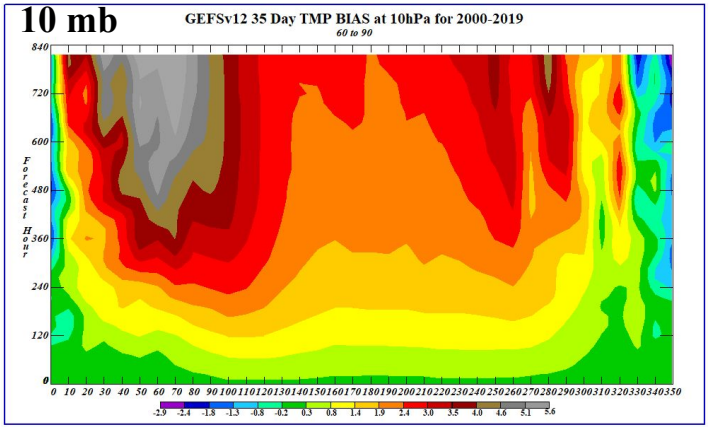
0-35 Day Temperature Bias (F-A) Growth (2000-2019)

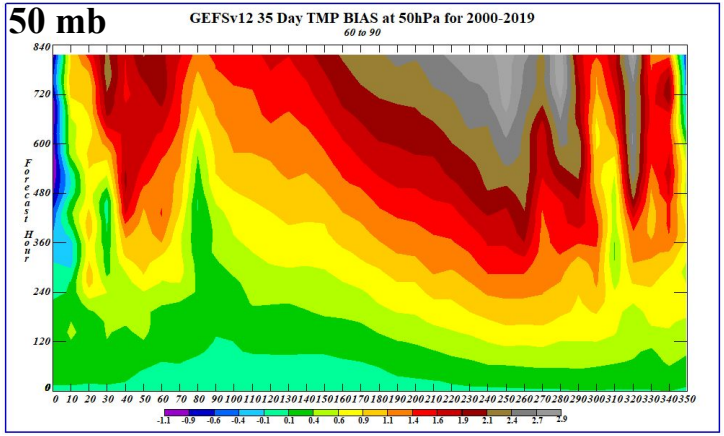
SH 60-90S



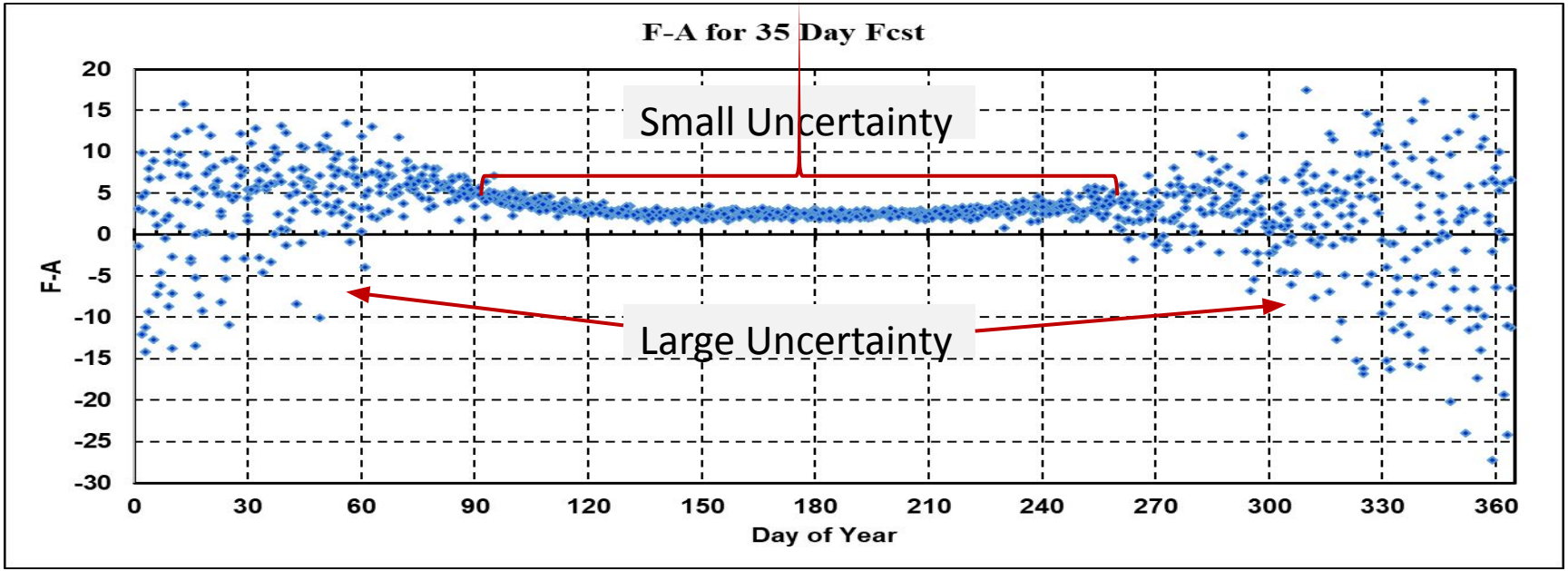


NH 60-90N





Uncertainty in Determining Bias for Any Single Day



Stratosphere Summary

- Temperatures:
  - NH & SH summer fcst temps increasingly are warmer than analysis
  - NH winter fcst temps also increasingly warmer than analysis
  - Longer forecasts miss sudden warmings in NH.
- Winds:
  - NH winter max winds are under-forecast
  - NH sudden warming wind deceleration not caught in longer forecasts
- QBO:
  - Structure is maintained throughout the 35 forecast day period
  - Intensity of both westerlies and easterlies become smaller with forecast time.
- Retrospectives:
  - T and u - Zonal mean forecast errors are smaller for GEFSv12 compared to GEFSv11.
- Long Term Statistics:
  - Performing one 35 day run per week may hinder ability to determine Bias and rms errors
  - 10 Member spread increases during NH winter months and SH winter-to-spring transition.

Week-2 Temperature

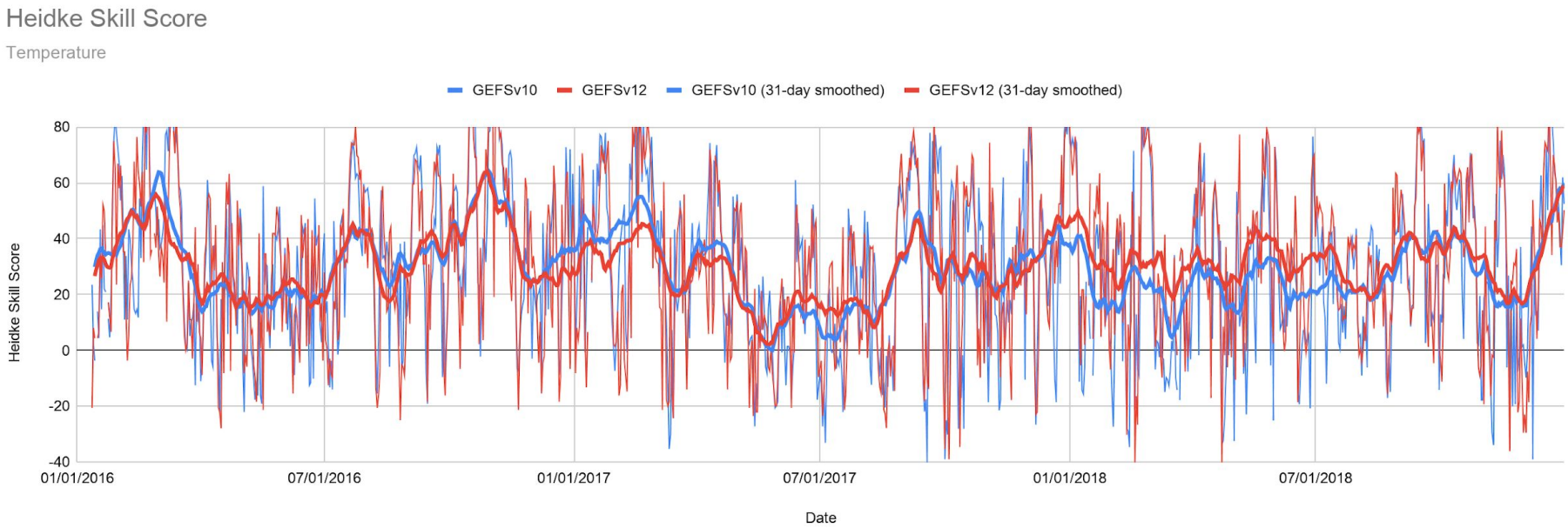
Mike Charles

Data

- **Forecast lead**: Week-2 (fhrs 198 to 360)
- **Spatial resolution**: 1° (GEFSv12 interpolated to 1°)
- **Calibration training data**:
  - *GEFSv10*: GEFSv10 reforecasts (11 members) from 1986 to 2015
  - *GEFSv12*: GEFSv12 reforecasts (5 members) from 1989 to 2018
- **Forecasts verified**:
  - Verification for 2016 to 2018
  - *GEFSv10*: Independent real-time GEFSv10 forecasts (daily, subset of 5 members) *
  - *GEFSv12*: Non-independent GEFSv12 reforecasts (daily, 5 members) *
- **Observations**: Subset of CADB stations available in the VWT

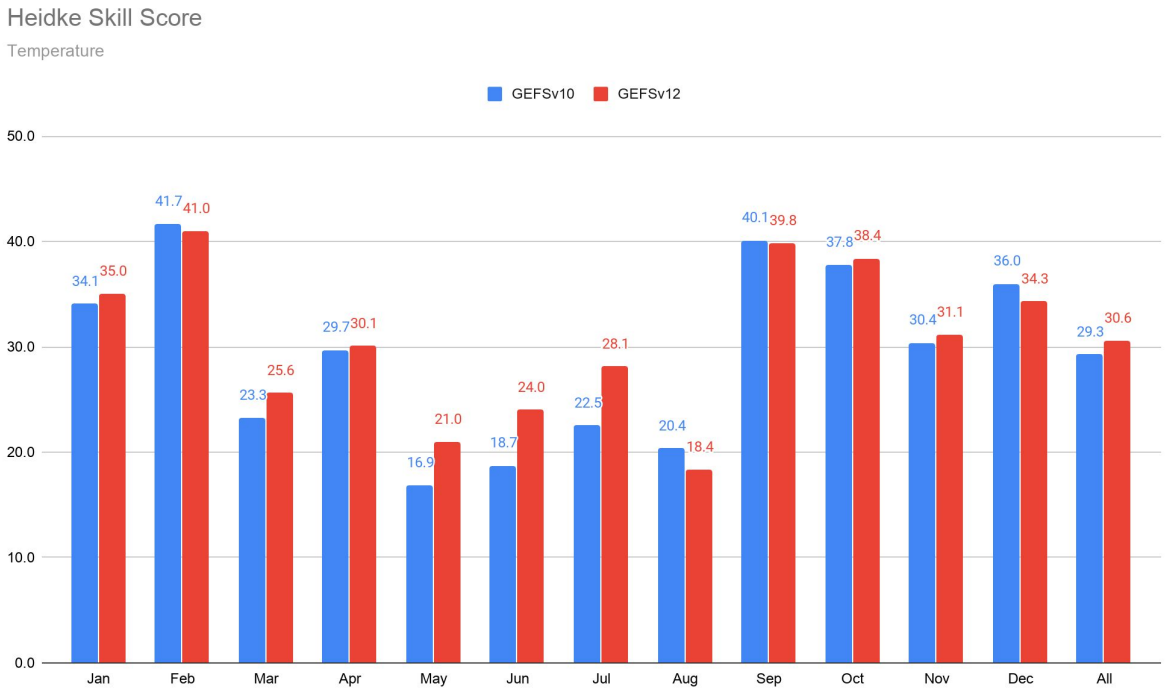
* ***GEFSv12 retrospective runs unusable for CPC (stop at fhr 240)*** *– needed to use GEFSv12 reforecast data instead. This gives GEFSv12 an advantage as the forecasts verified are not independent. A subset of 5 out of 21 GEFSv10 real-time members was used to make the two datasets more homogeneous.*

Heidke Skill Score Timeseries



Temperature

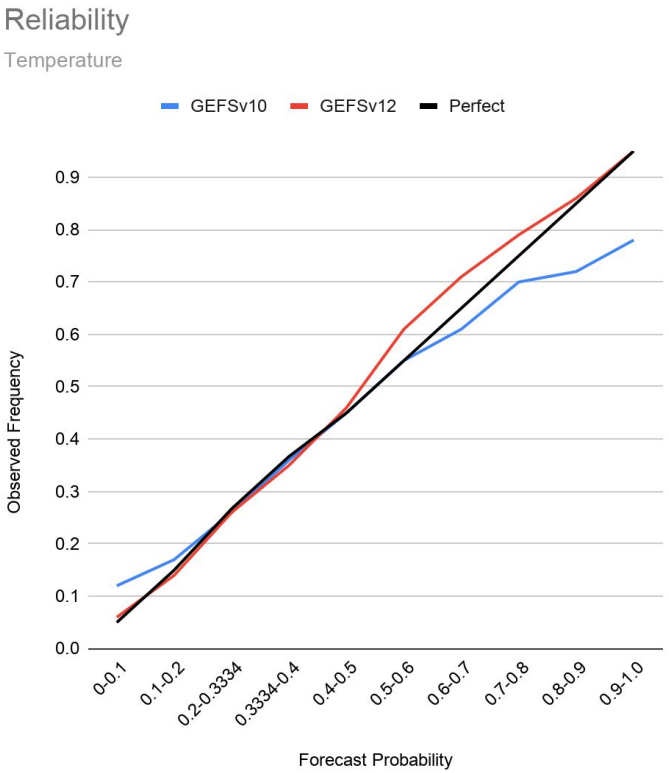
Average Heidke Skill Score



- GEFSv12 HSS is higher in 8 out of 12 months – especially in May, June, and July
- Overall GEFSv12 skill higher than GEFSv10 (95% sig.)

Temperature

Reliability



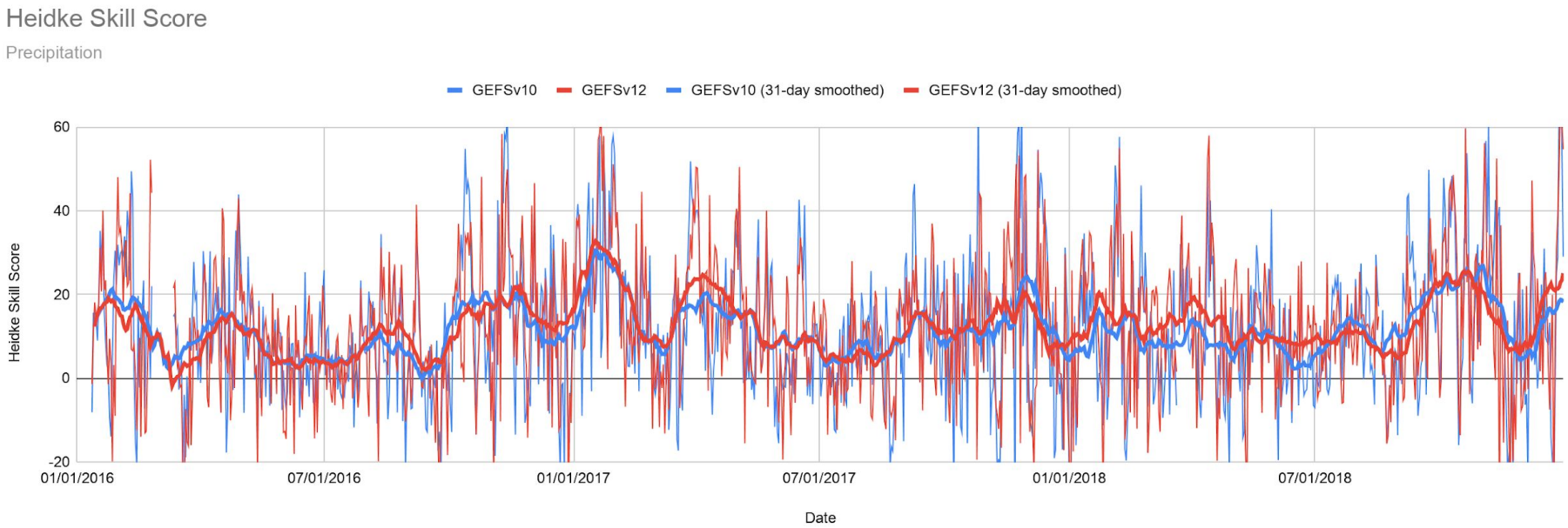
- GEFSv12 slightly under-confident above 50% probs.
- GEFSv10 overconfident above 60% probs.
- **CAVEAT** – Both reliabilities are expected to be better if all members from each ensemble were used

Temperature

Week-2 Precipitation

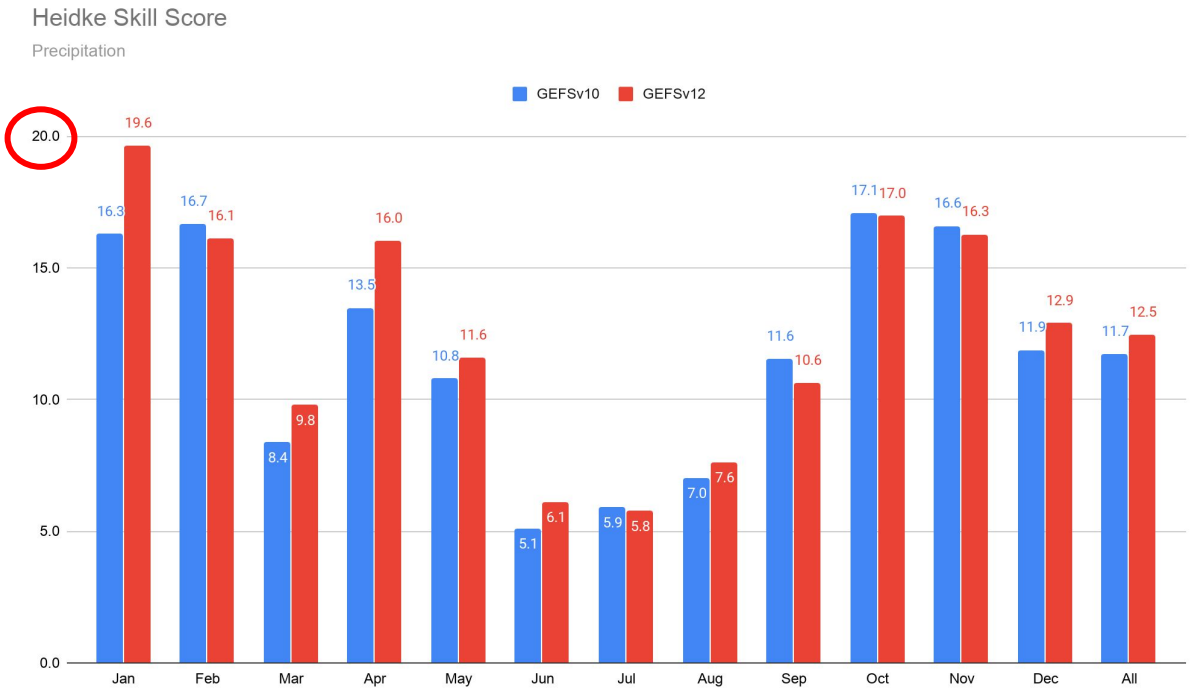
Mike Charles

Heidke Skill Score Timeseries



Precipitation

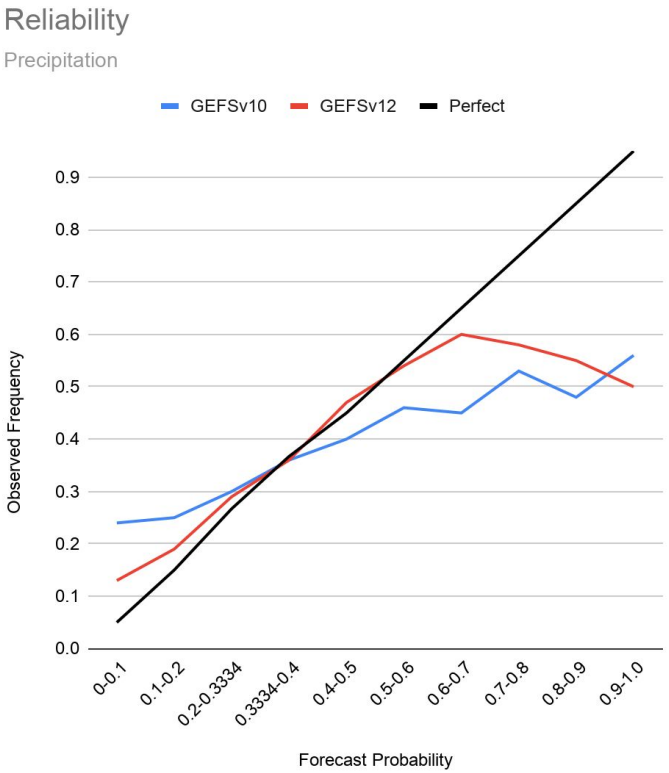
Average Heidke Skill Score



- GEFSv12 HSS is during 8 out of 12 months
- Overall GEFSv12 skill higher than GEFSv10, but only 87% statistically significant

Precipitation

Reliability



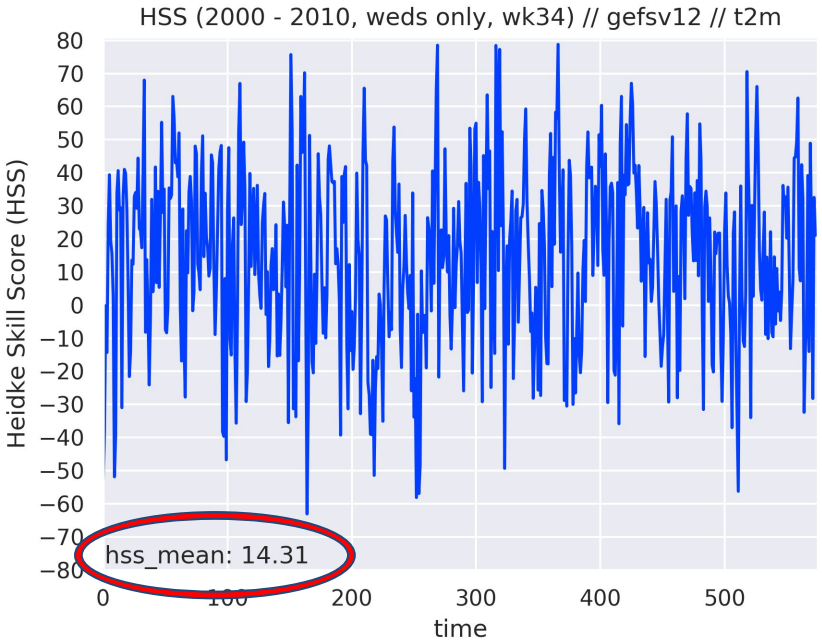
- Both models are overconfident
- GEFSv10 less overconfident for probabilities less than 60%
- **CAVEAT** – Both reliabilities are expected to be better if all members from each ensemble were used

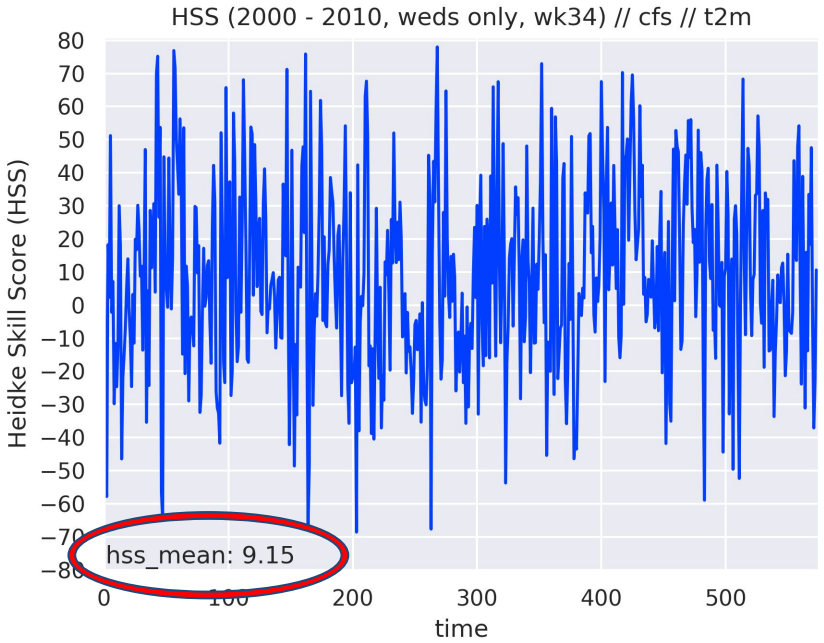
Precipitation

Week 3/4 Temperature

Kyle MacRitchie

t2m: GEFSv12 vs CFS for 2000-2010

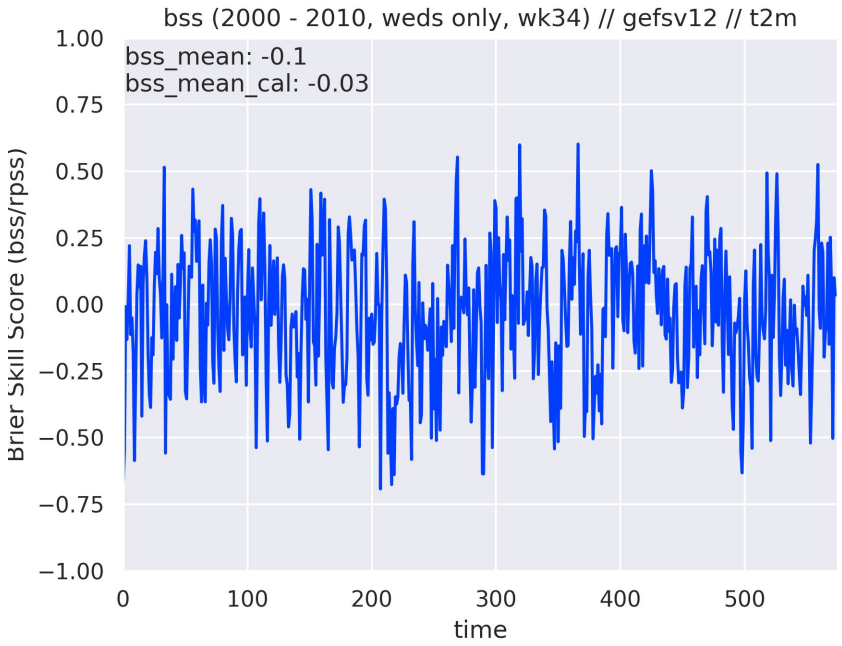


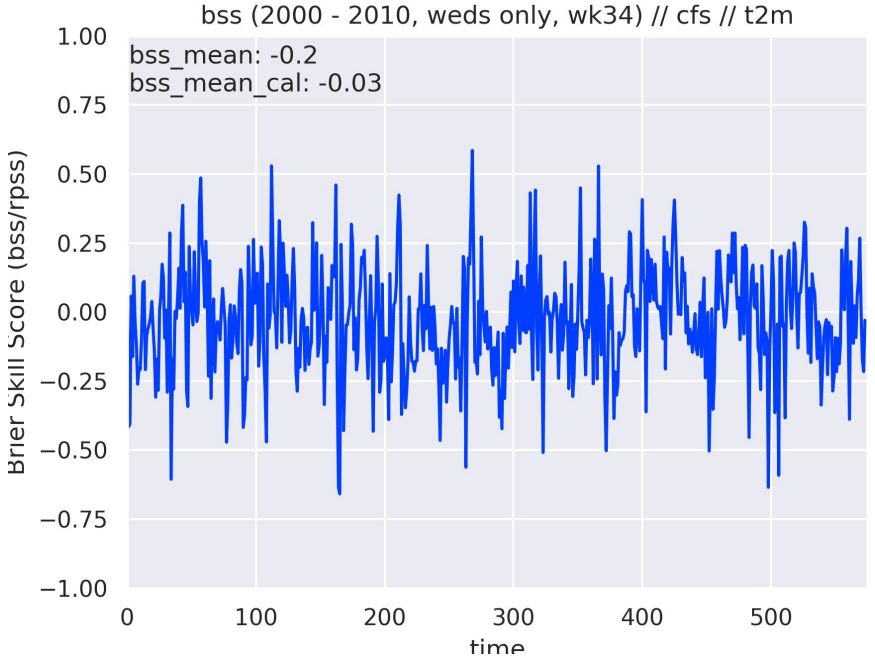


Summary:

- The GEFSv12 is noticeably better than the CFS over this time period.
- The difference between these means passes a t-test at 95%.

t2m: GEFSv12 vs CFS for 2000-2010

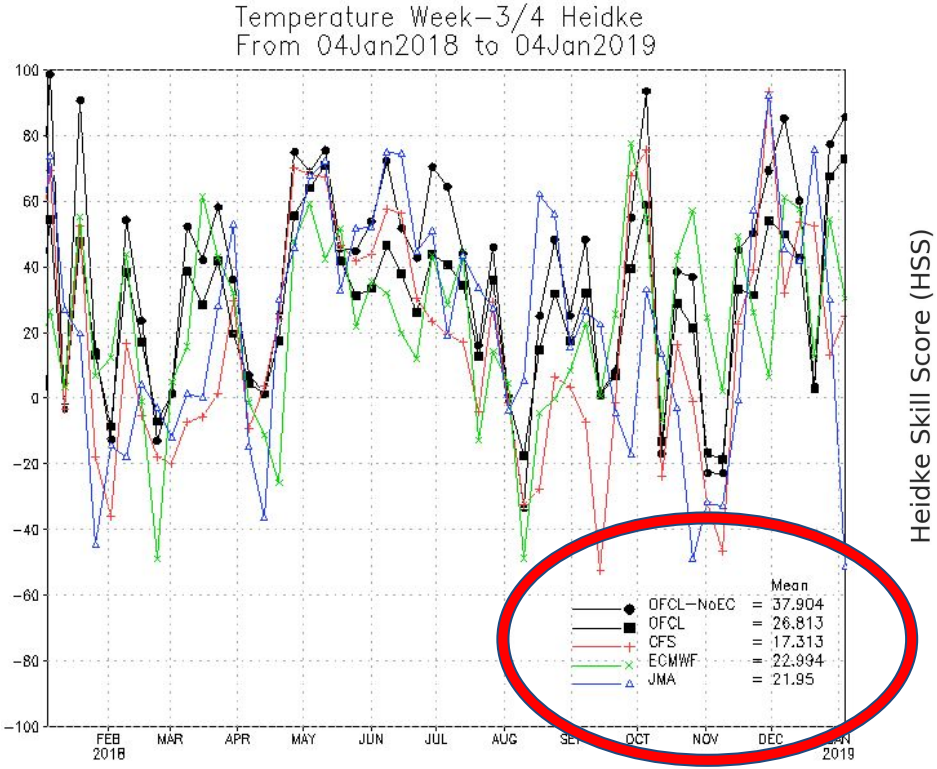


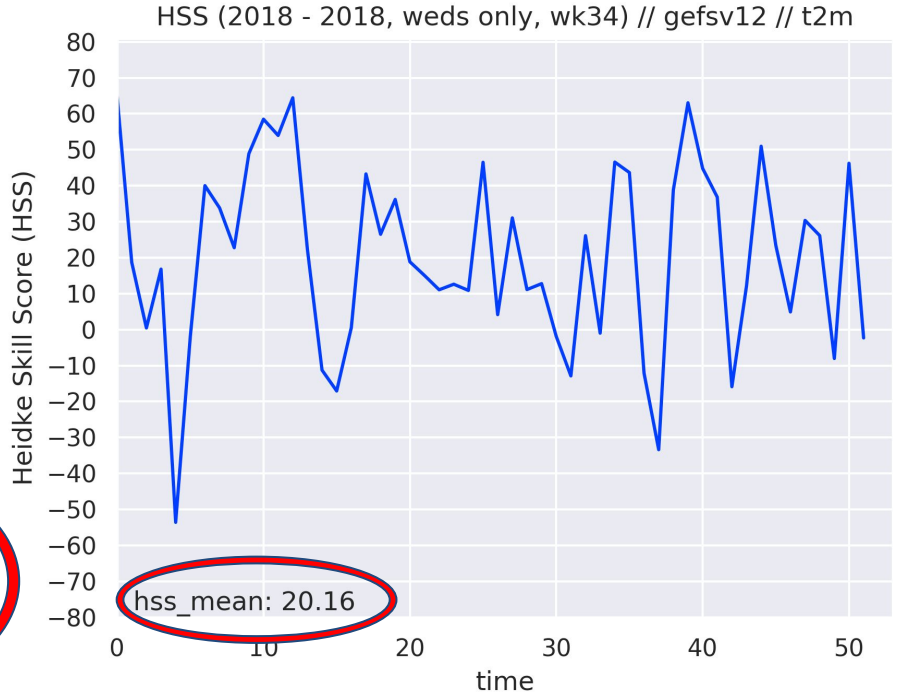


Summary:

- The GEFSv12 is slightly better uncalibrated, and identical when calibrated.

Temperature: 2018 Comparison





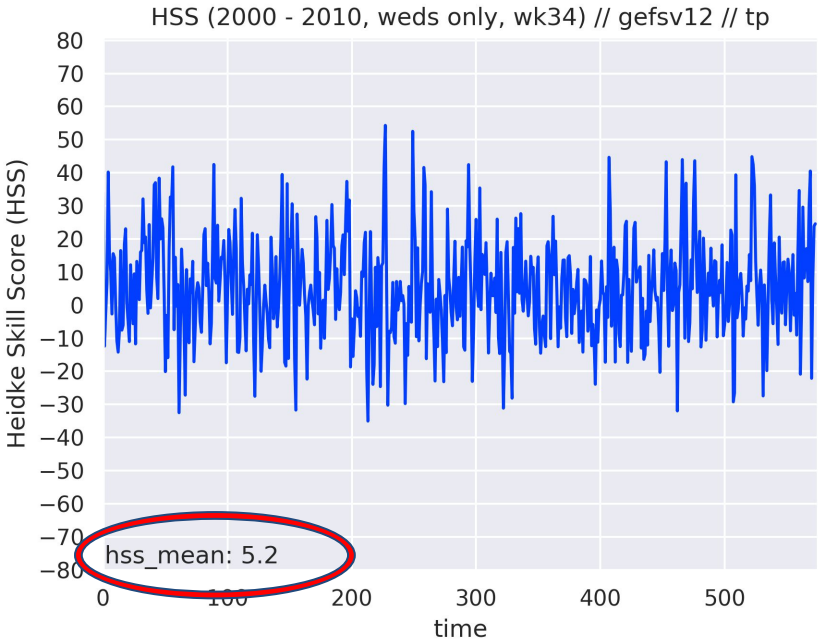
Summary:

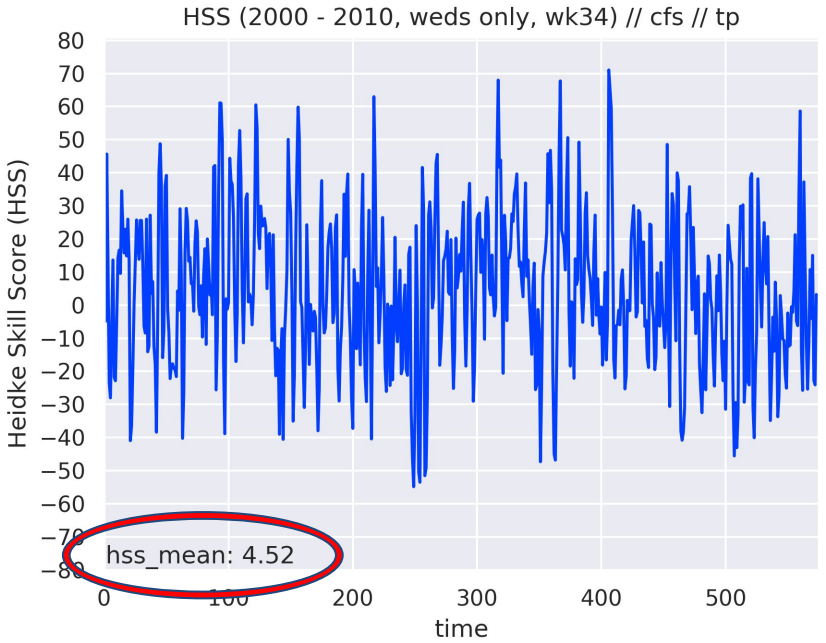
- The GEFSv12 holds its own among the other models and bests the CFS by about 3 points.

Week 3/4 Precipitation

Kyle MacRitchie

precip: GEFSv12 vs CFS for 2000-2010

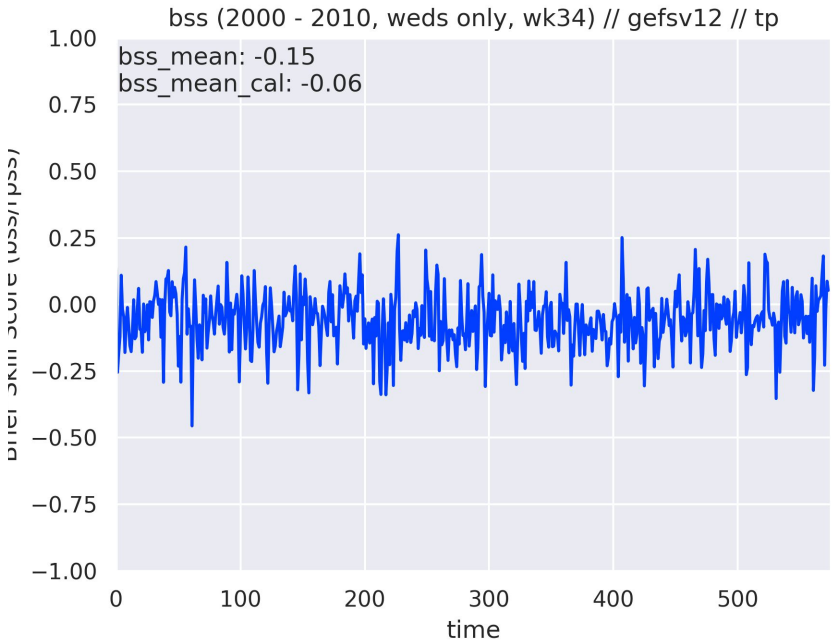


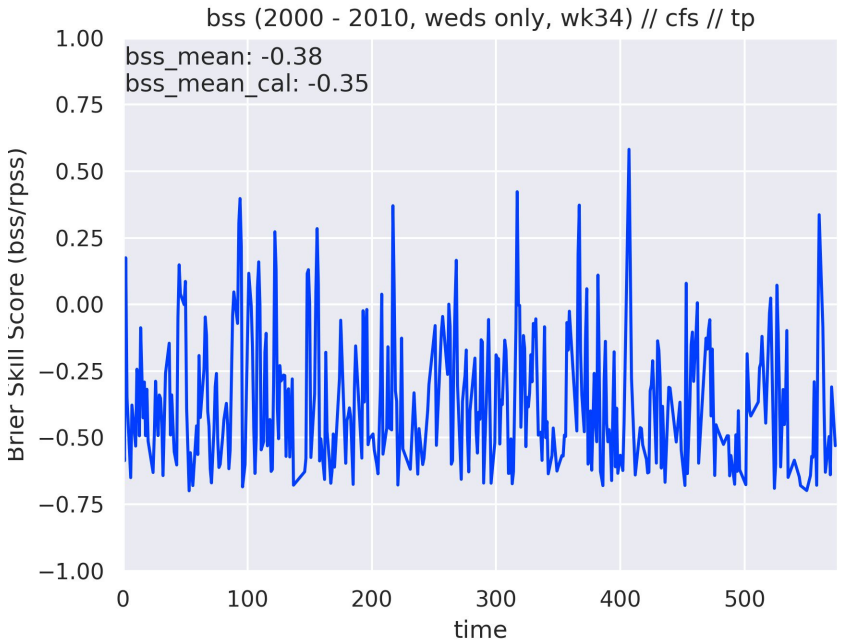


Summary:

- The GEFSv12 is slightly better than the CFS over this time period.
- The difference between these means **<u>does not</u>** pass a t-test at 95%.

precip: GEFSv12 vs CFS for 2000-2010

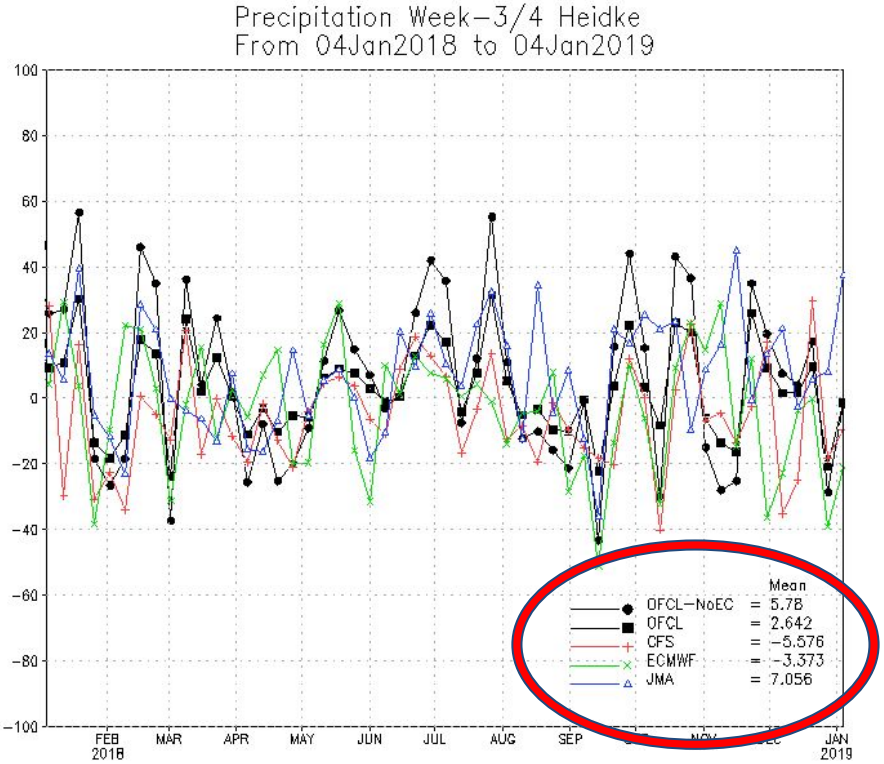


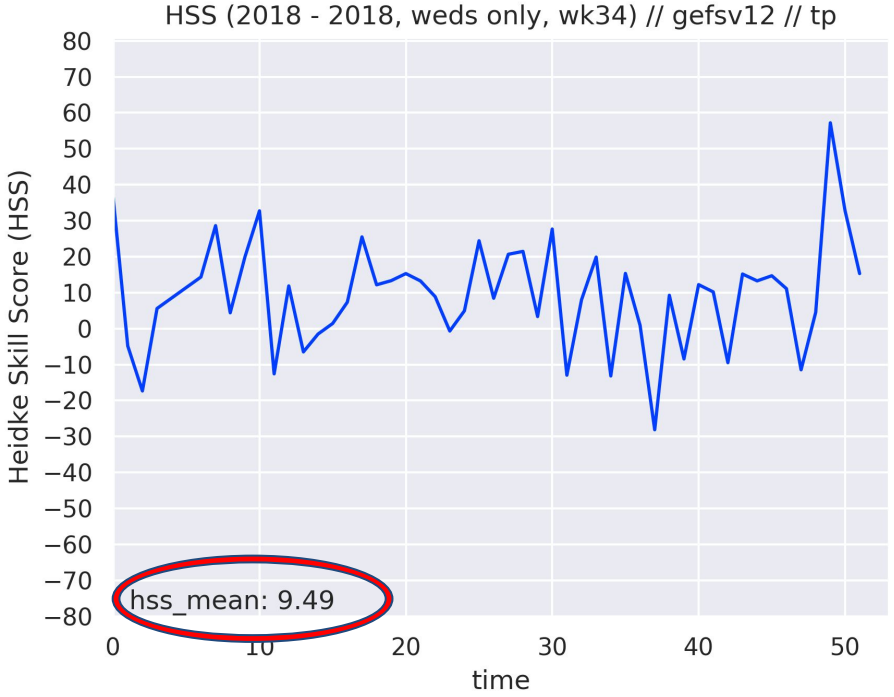


Summary:

- The GEFSv12 has higher BSS than the CFS and less variance.

Precipitation: 2018 Comparison





Summary:

- The GEFSv12 seems to do well with precipitation. One reason for this that the plot on the left measures precipitation against a climatology of median values whereas the one on the right uses mean values.

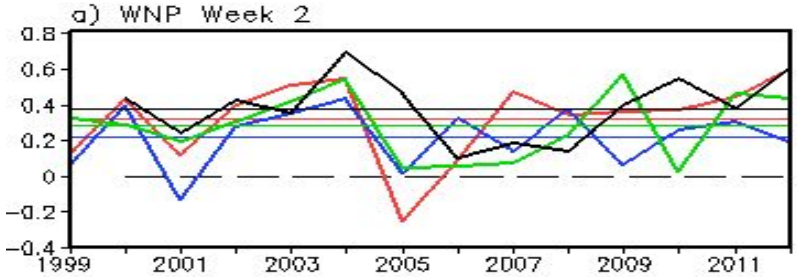
Tropical Cyclone Counts / Tracks

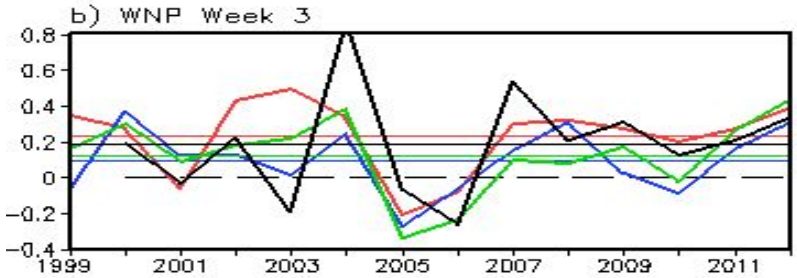
Lindsey Long

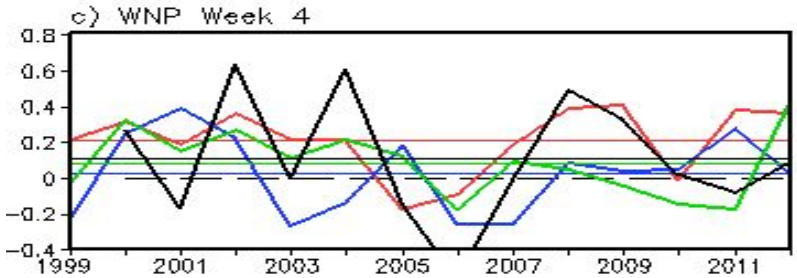
Weekly Storm Count and Tracks

- Data are grouped into weekly forecast periods starting on Day +2, out to week 4.
  - Each week forecast represents how many storms are present during that weekly period, not just genesis.
  - One storm can be counted in multiple weeks.
- Climatology based on week number (Weeks 1-22)
  - Wednesday different each year
  - ie. Wk1: May 31 – Jun 6… Wk22: Oct 24 – Oct 31
- Tracks filtered using false alarm clim from reforecasts

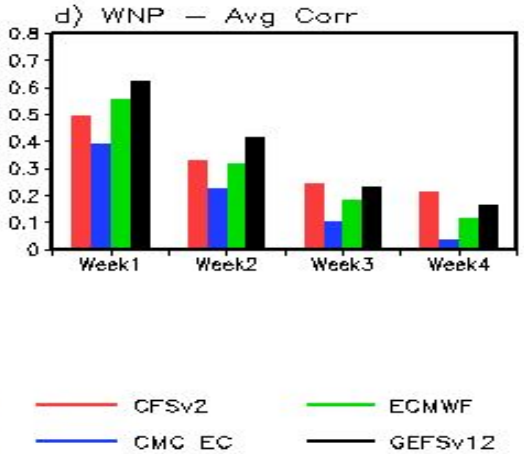
Anomaly Correlations - WNP







- GEFS outperforms the other models in Weeks 1-2 and is on par with CFS in weeks 3-4.



- Years of Note
- Good: 2004, 2005?
- Bad: 2006

CFSv2 ECMWF
CMC GEFSv12

Anomaly Corr.

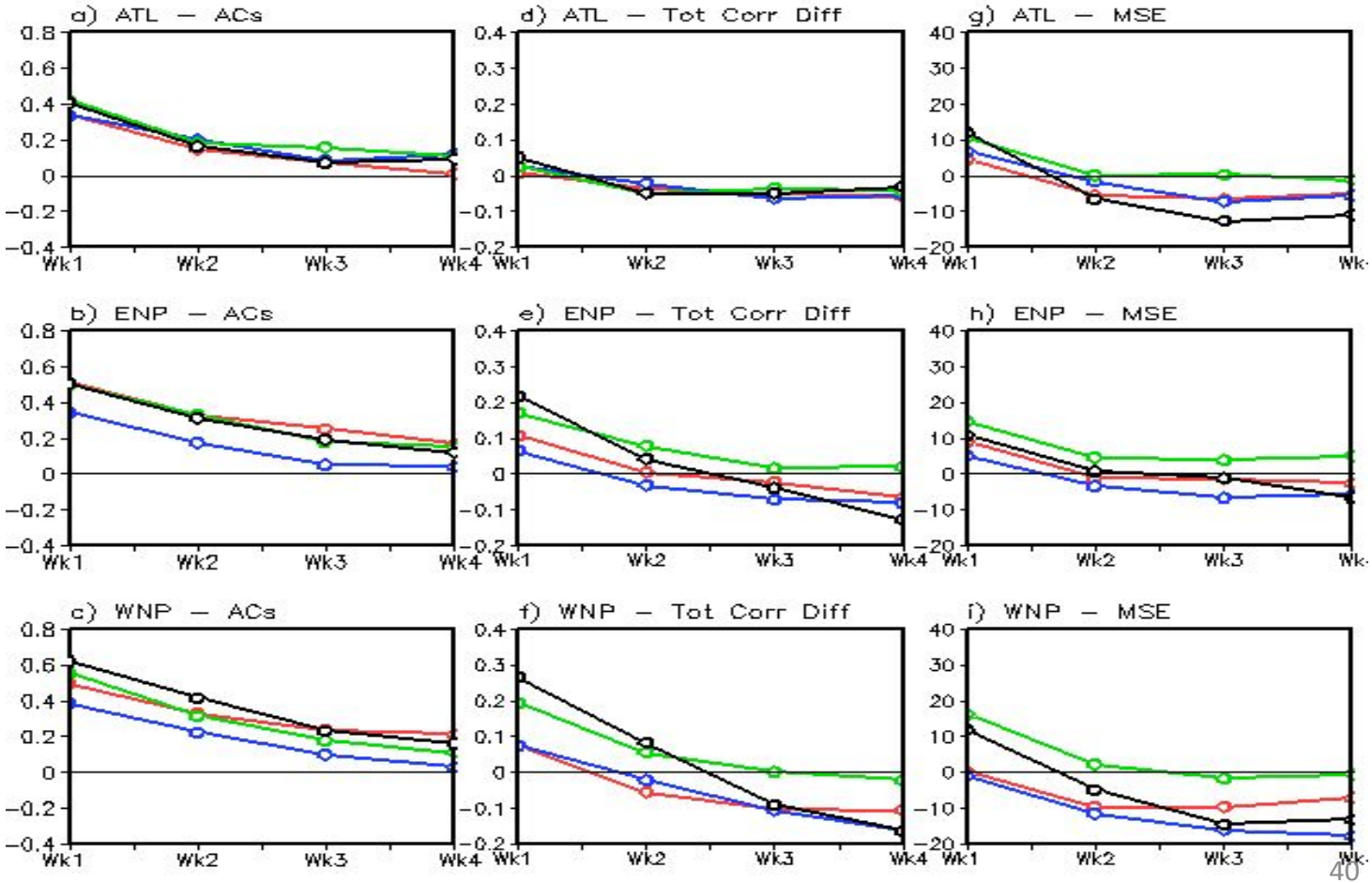
AC Difference from obs. seasonal cycle

MSE

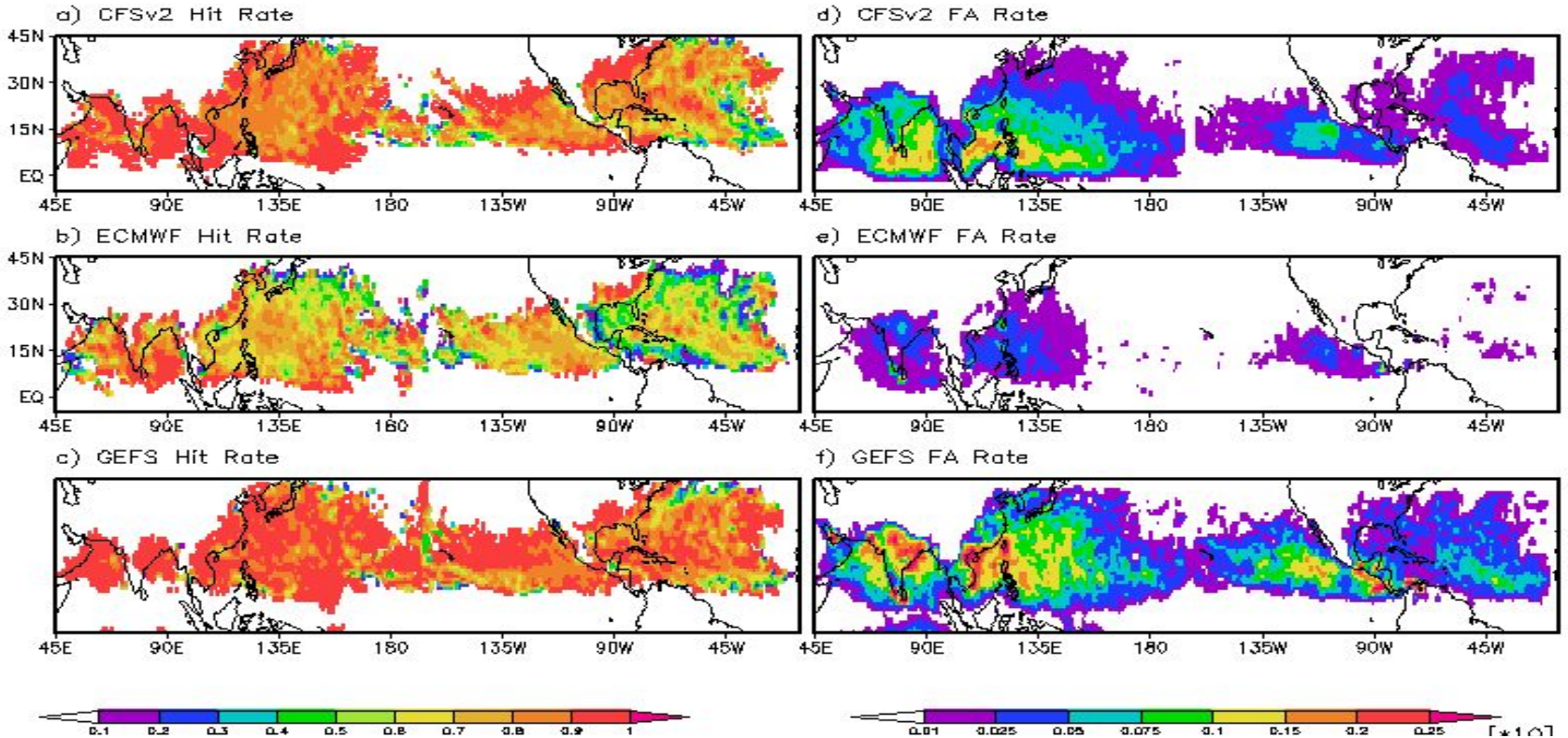
Atlantic

E. Pacific

W. Pacific

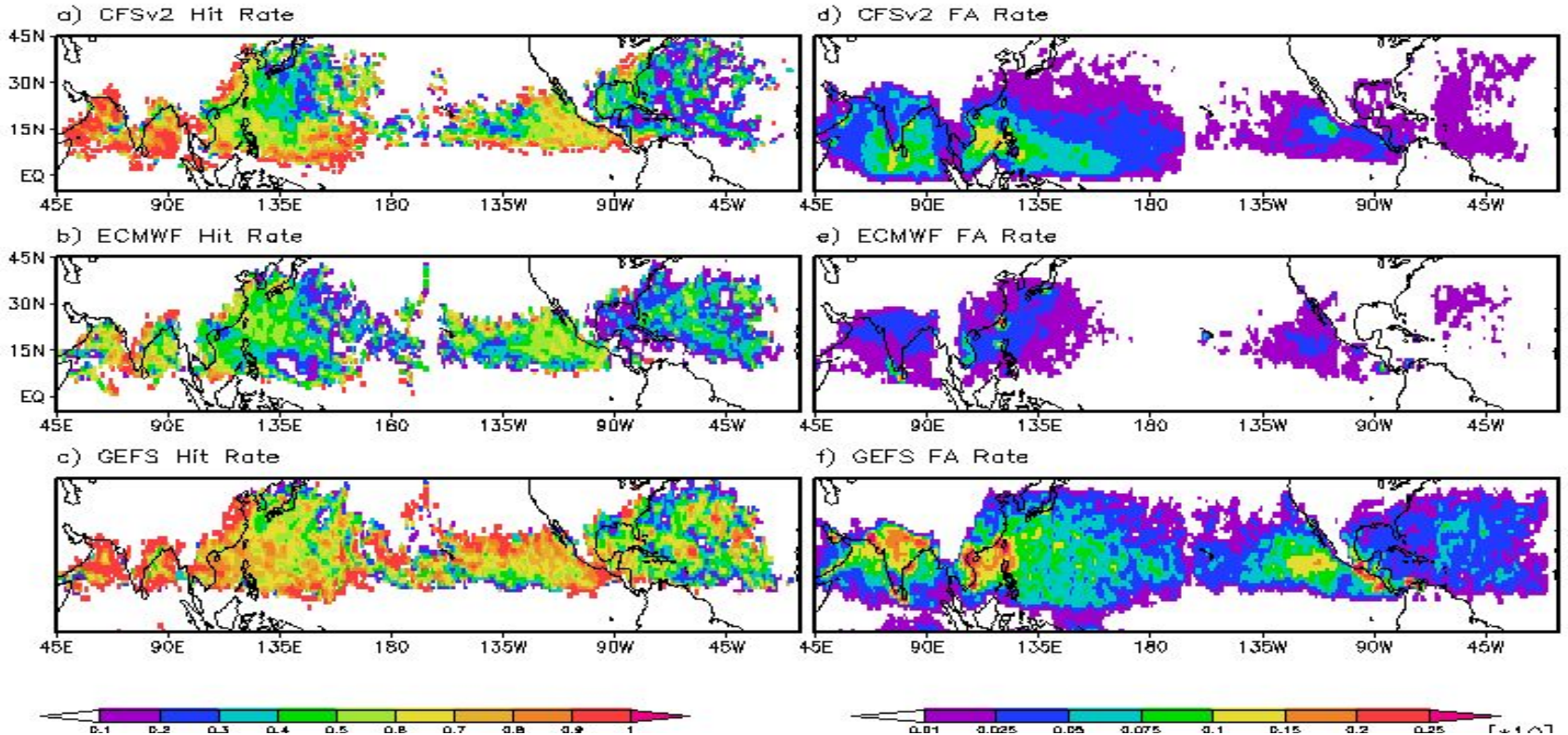


Hit and False Alarm Rates – Week1



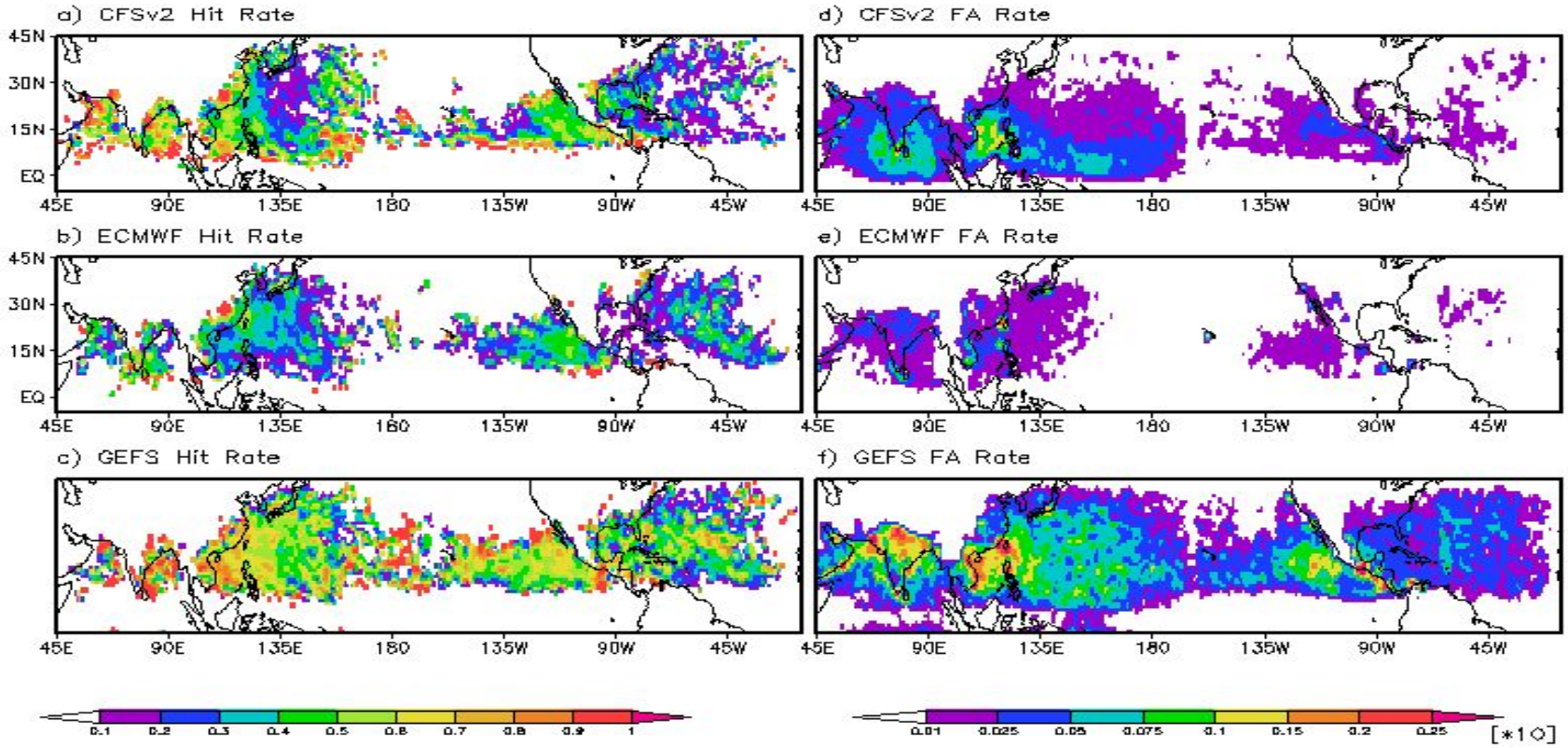
- GEFSv12 has very high hit rates, but also highest FA rates - Asian coast/N.Philippines. High skill in WNP consistent with storm count scores. Atlantic "patchy" compared to other basins.

Hit and False Alarm Rates – Week2

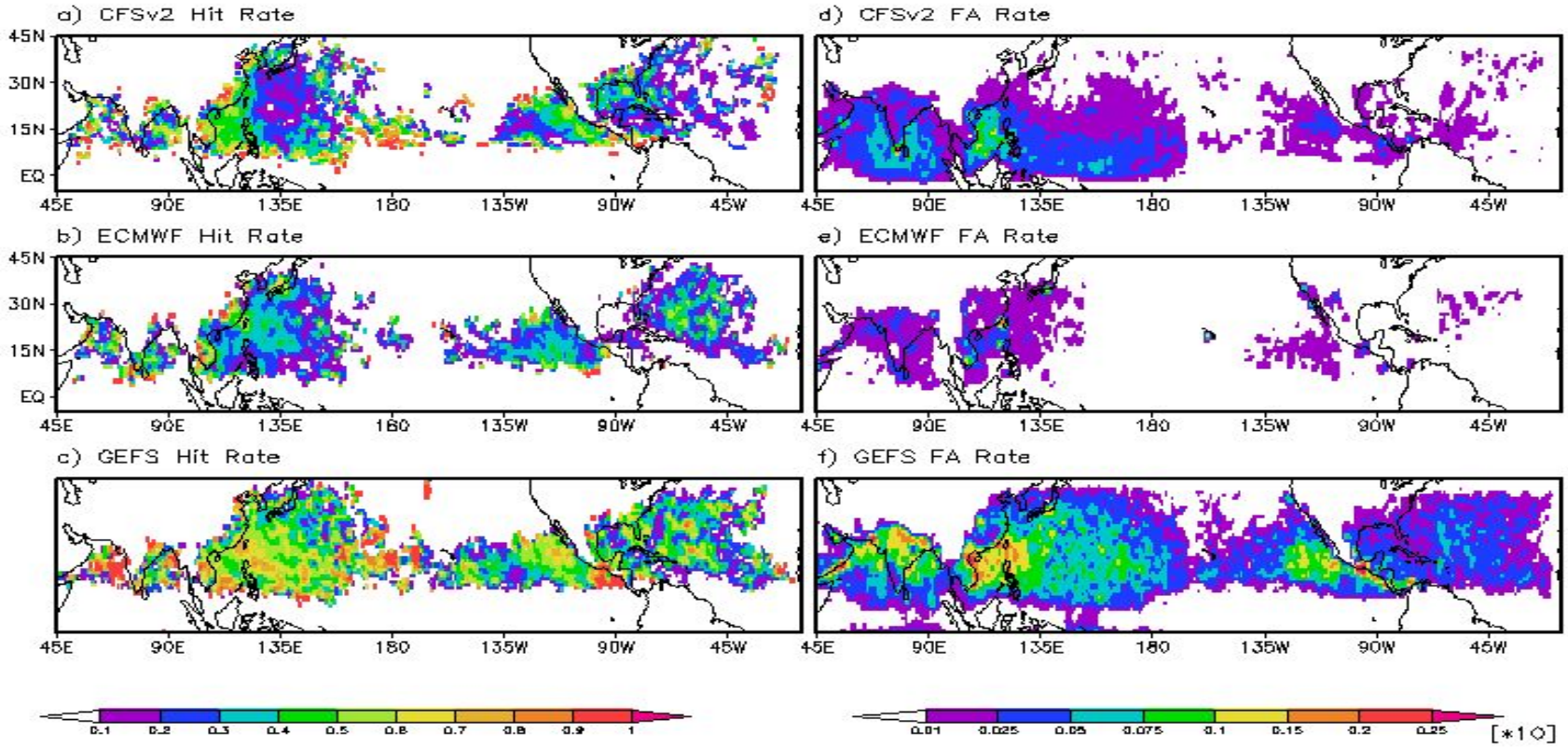


- GEFSv12 maintains better hit rate, but forecasts unreliable because of high FA rate, versus ECMWF which has a very low FA rate. Trend continues in weeks 3 and 4 (next slides).

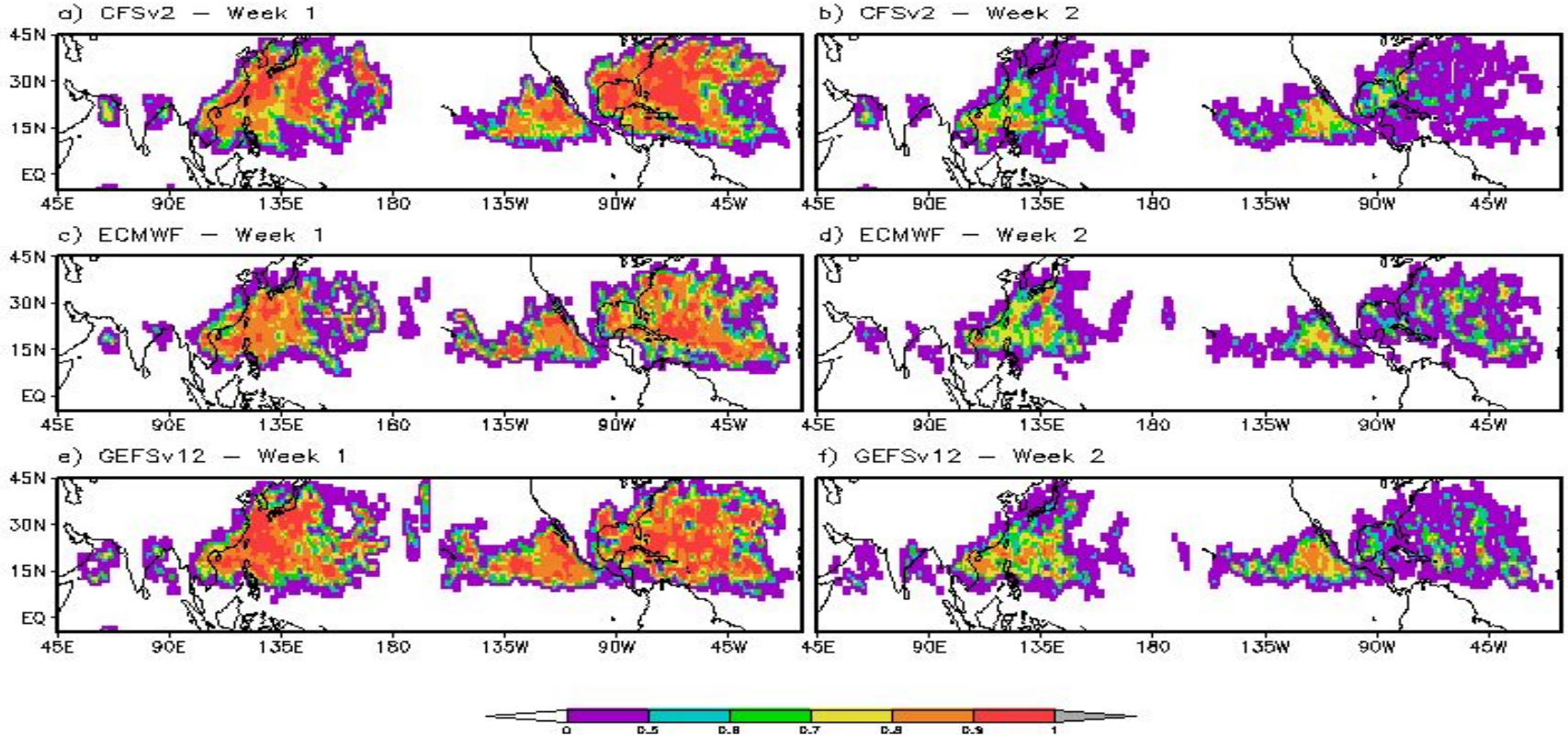
Hit and False Alarm Rates – Week3



Hit and False Alarm Rates – Week4



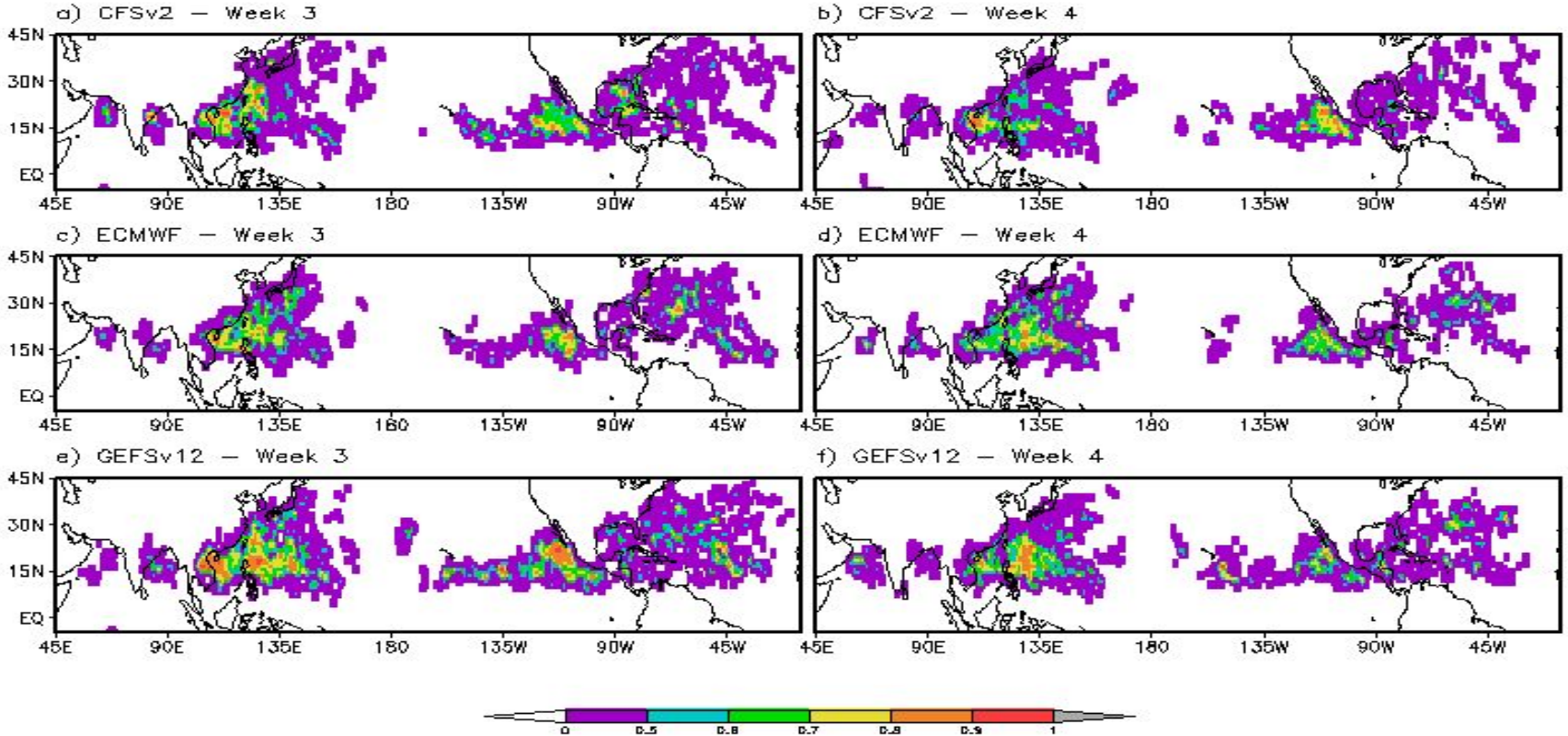
SEDS – Weeks 1&2 for Sept



- SEDS combines hit and FA rates to better portray usefulness of forecasts.
- GEFS is consistent with other models in forecasting presence of a storm.

Note, other months have similar results.

SEDS – Weeks 3&4 for Sept



- Models show continued drop in skill in later leads.
- Still pockets of skill in Pacific.

Supercell Convective Parameter

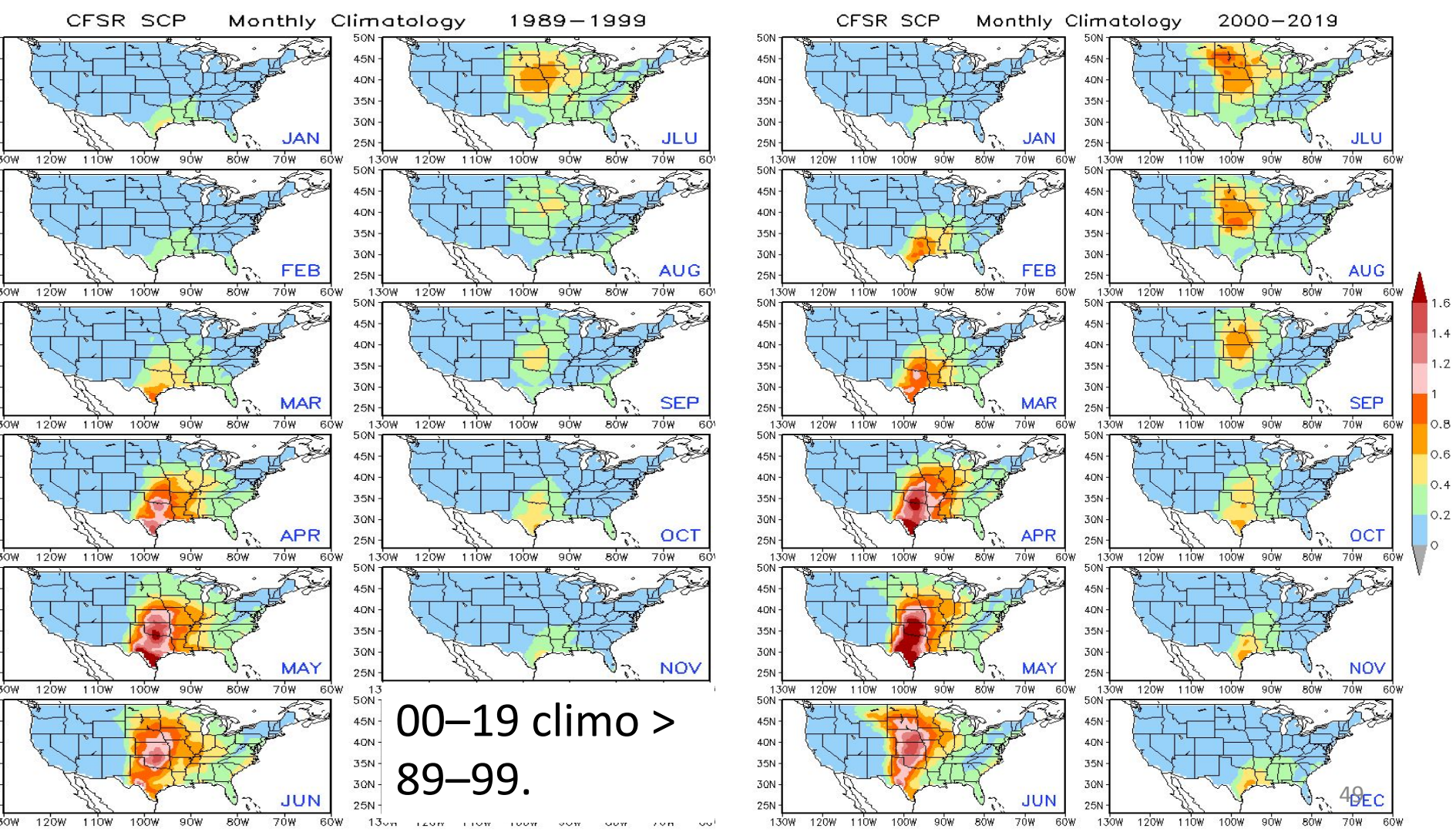
Hui Wang

Derived SCP (Supercell Composite Parameter) from

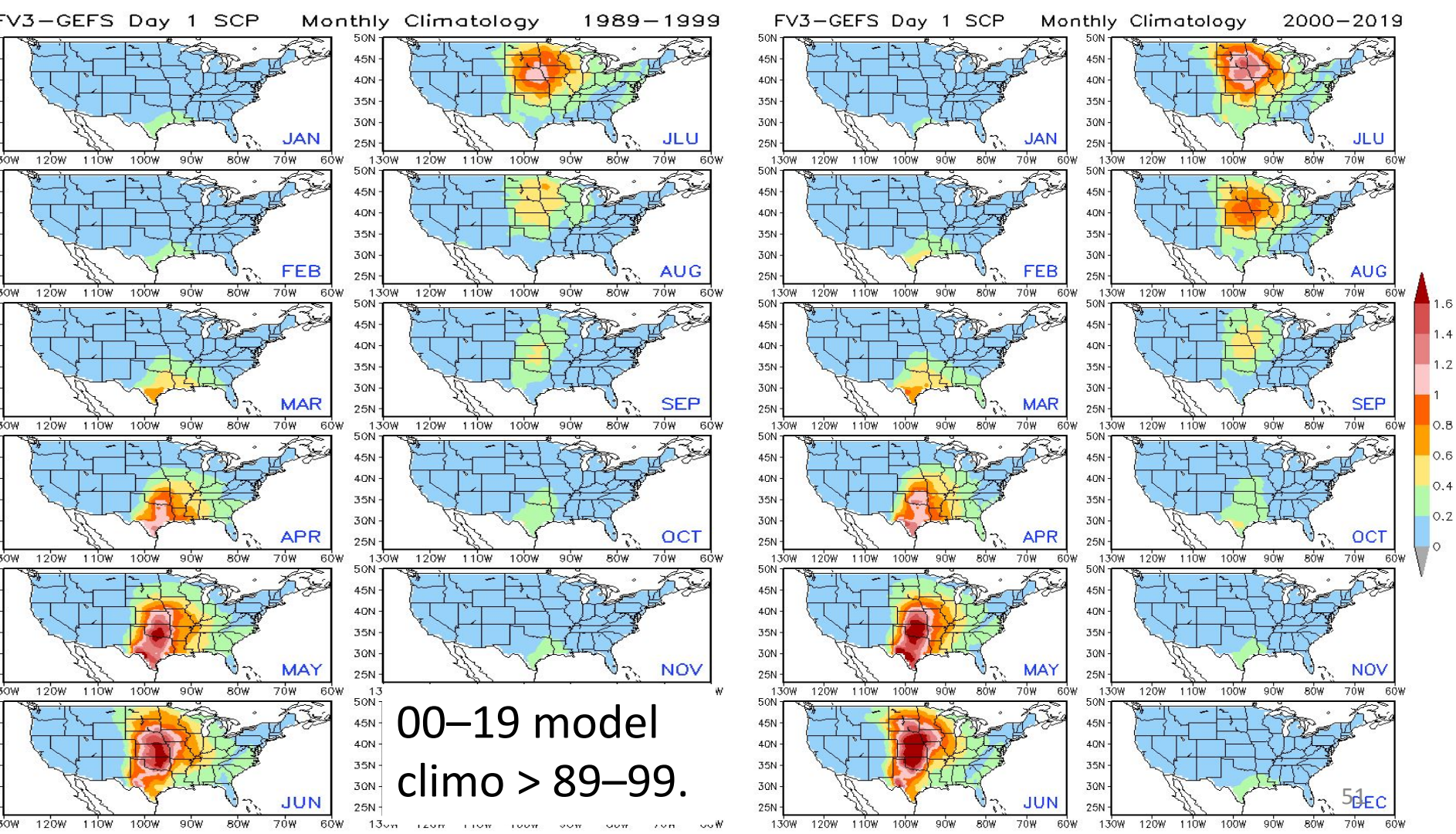
- OBS (CFSR)
- FV3 GEFS hindcast
  - 1989 – 2019 (31 years)
  - 6 hourly data
  - Daily 14-day hindcast
  - 5 ensemble members

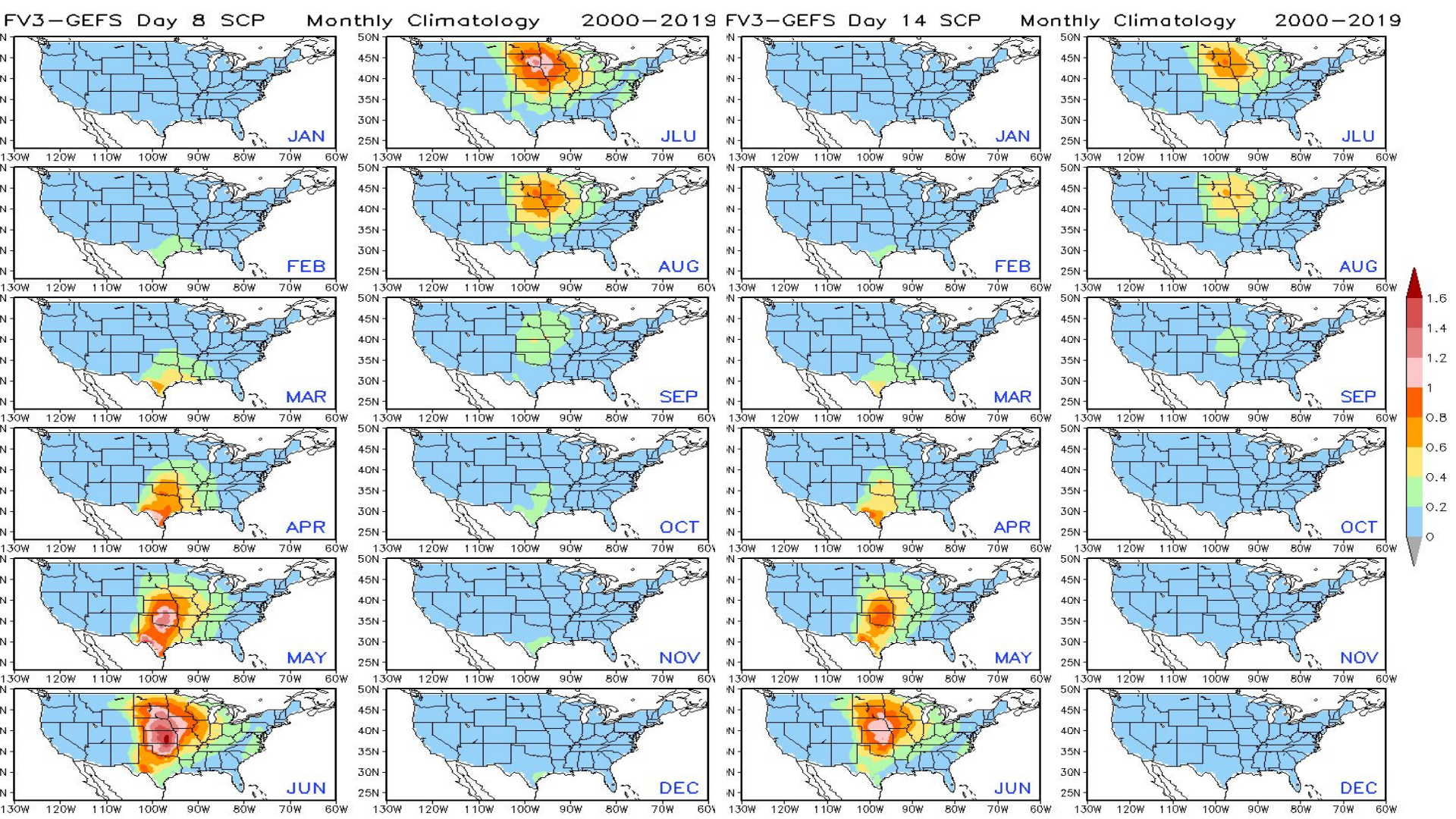
Compared monthly climatology between GEFS and OBS over

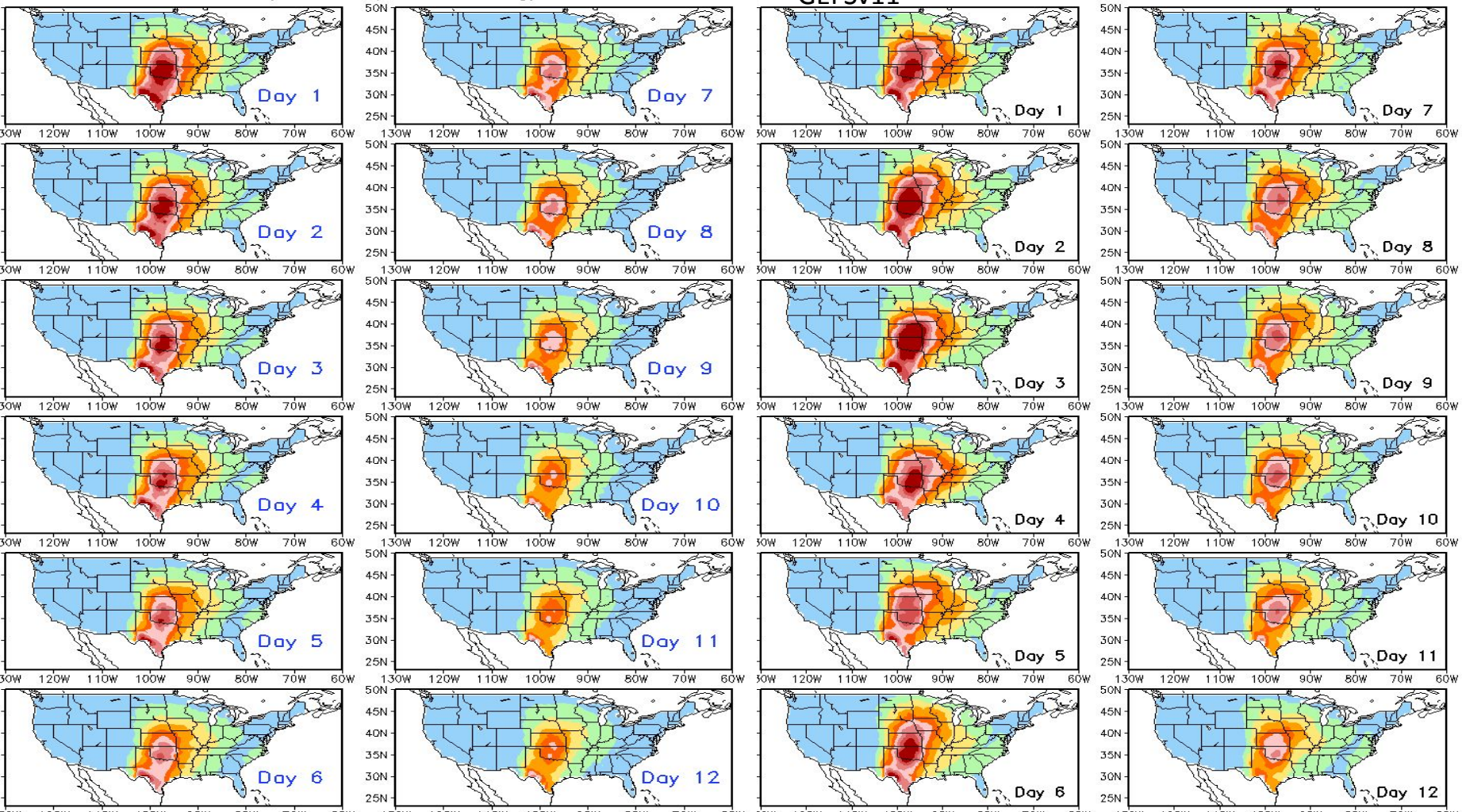
- 1989 – 1999
- 2000 – 2019


CFSR SCP Monthly Climatology 1989–1999
CFSR SCP Monthly Climatology 2000–2019
JAN
FEB
MAR
APR
MAY
JUN
JLU
AUG
SEP
OCT
NOV
DEC
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
00–19 climo > 89–99.

FV3 GEFS Climatology
1989 – 1999
2000 – 2019
Day-1 forecast


FV3–GEFS Day 1 SCP Monthly Climatology 1989–1999
FV3–GEFS Day 1 SCP Monthly Climatology 2000–2019
JAN
JLU
FEB
AUG
MAR
SEP
APR
OCT
MAY
NOV
JUN
DEC
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
00–19 model climo > 89–99.


FV3-GEFS Day 8 SCP Monthly Climatology 2000-2019
FV3-GEFS Day 14 SCP Monthly Climatology 2000-2019
JAN
FEB
MAR
APR
MAY
JUN
JLU
AUG
SEP
OCT
NOV
DEC
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0


FV3-GEFS SCP May Monthly Climatology 2000-2019
GEFSv11 SCP Monthly Climatology MAY 1996-2012
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Day 7
Day 8
Day 9
Day 10
Day 11
Day 12
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0

Lessons Learned

- Retrospectives are the independent data - need 3 years of retrospective, out to Week4

- Need to better agree upon metrics ahead of time.
  - Even if agreed upon, familiarity is part of it.

- CPC needs to systematize our evaluations to where we can drop new GEFS gribs into folder (rfcst, retro) and obs into another, and just let it run. - Facilitate faster feedback.

Summary

- T & P
  - Week 2 - improved*
  - Week 3/4 T & P - improved
- Z500
  - Week 2,3/4 - improved
- Tropical Cyclones
  - Week 2 - improved vs CFSv2, similar to ECWMF
  - Week 3/4 - improved - all models struggle
- Stratosphere
  - GEFSv12 - largely an improvement, still work to be done
- **Overall - Recommend implementation**