UGCS Benchmark Testing (with NEMS GSM + MOM5.1 + CICE5)

Suru Saha, Christopher Melhauser, Malaquias Pena
and all of the UGCS Team in EMC.

Special thanks to Huug van den Dool, Innovim-CPC for his advice in computing the scores as well as generating the calibration climatologies.

We gratefully acknowledge the NESII (ESMF team at ESRL) for their significant contributions in setting up the coupled NEMS system.

Models used:

All components of the UGCS are in NEMS:

A. GSM: Spectral T574L64 semi-Lagrangian grid
   gsm_q3fy2017_lamda tag (SVN EXTERNAL:
   https://svnemc.ncep.noaa.gov/projects/gsm/tags/gsm_q3fy2017_lambda)

A. MOM5.1: GFDL Ocean Model. Z-coordinates, Tripolar CFSv2 grid 0.25^o in the tropics and 0.5^o global.

B. CICE5: Los Alamos SeaIce Model. Same grid as MOM5.1 ocean model.

Data used

- CFSR Initial Conditions for all experiments are from:
  Operational CFSv2 CDAS using
  Spectral T574L64 Eulerian grid
  MOM4 GFDL Ocean Model, Z-coordinates, Tripolar grid, 0.25^o in the tropics and 0.5^o global.
  SIS1 GFDL SeaIce Model, same grid as MOM4 ocean model.

- April 2011 to March 2017 (6 years).

- UGCSbench: 35-day coupled forecasts were made from the 1st and 15th of each month, a total of 144 forecasts.

- UGCSuncpl_cfsbc: 35-day uncoupled forecasts, using bias corrected SSTs from the operational CFSv2 from the same set of 144 initial conditions.

- CFSv2ops: 35-day coupled forecasts from the operational CFSv2 from the same set of 144 initial conditions were used for comparison.

Calibration Climatologies

We need climatologies to form anomalies and, more importantly, for systematic error correction (SEC) which may be very large in some variables.

A climatology as an average over just 6 cases (years) would be much too noisy.

Here we produce a smoothly interpolated climatology by fitting the 6 year time series (144 elements, 2 weeks apart) to a sine wave of period 365.24 days plus three overtones. This way, leap days are handled correctly both on the input and output side. The climatology consists of an annual mean plus four harmonics.

This is done for each gridpoint and variable separately. Both for forecasts (as a function of lead, at 6 hour intervals) and verifying data (mainly CFSR).

All forecasts (coupled, uncoupled, control, experiment, etc) were bias corrected in exactly the same manner.

CONUS 2-meter Temperature

CONUS 2-meter temperature AC (CPC daily*)

	UGCSbench	UGCSbench	CFSv2ops	CFSv2ops
	Raw	**Sec**	**Raw**	**Sec**
week1	78.0	**87.5**	79.3	85.9
week2	40.1	**46.7**	41.7	46.4
week3	19.4	**23.3**	17.6	19.9
week4	11.0	**12.6**	0.3	1.8
week3&4	20.8	**26.1**	11.6	14.7

UGCSbench equal or better than the CFSv2ops for all lead times.

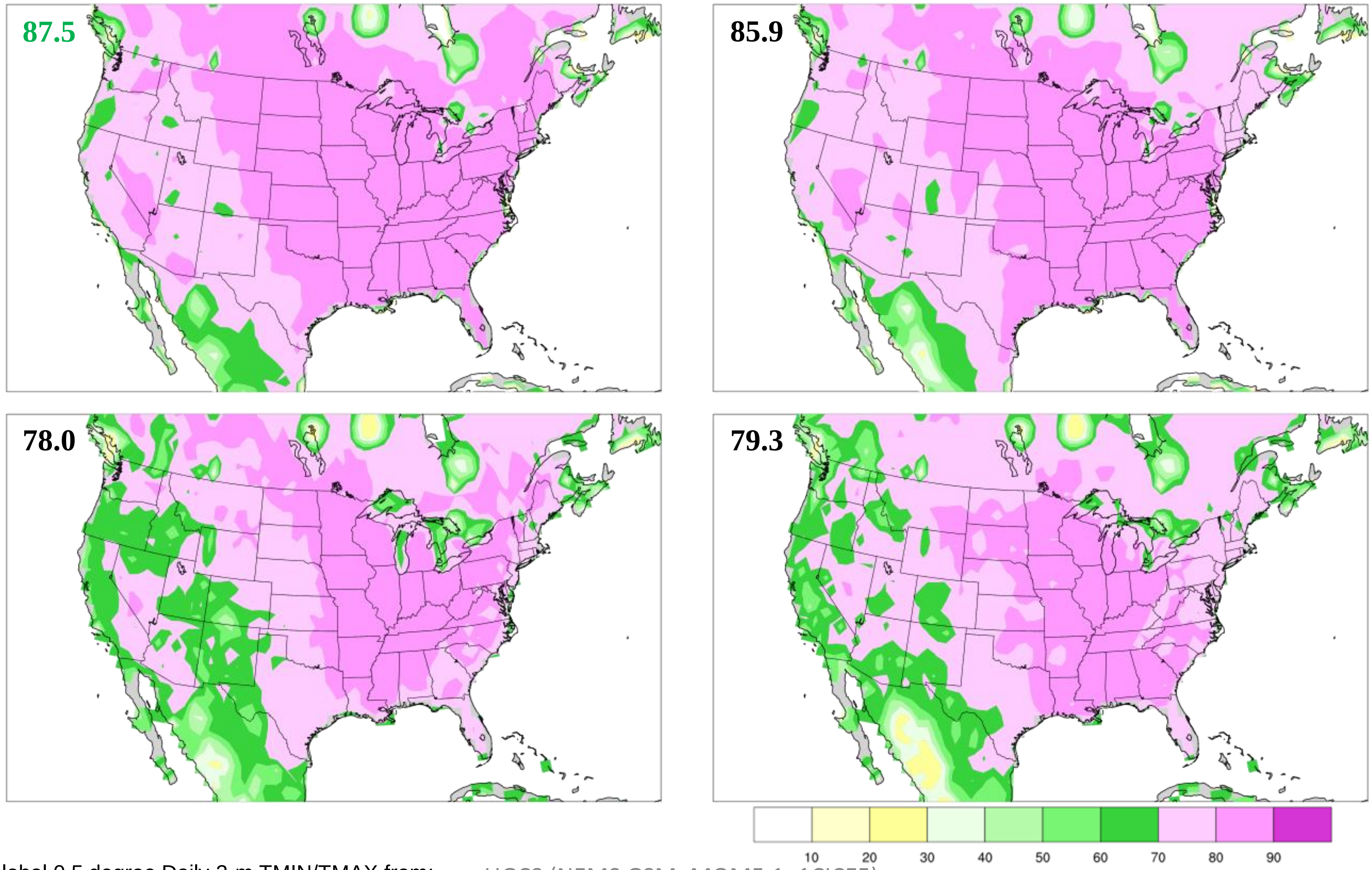
*CPC Global 0.5 degree Daily 2-m TMIN/TMAX from:
ftp://ftp.cpc.ncep.noaa.gov/precip/wd52ws/global_temp/
e.g., CPC_GLOBAL_T_V0.x_0.5deg.lnx.YYYY

Week 1
T2m AC

UGCSbench CFSv2ops

sec

raw



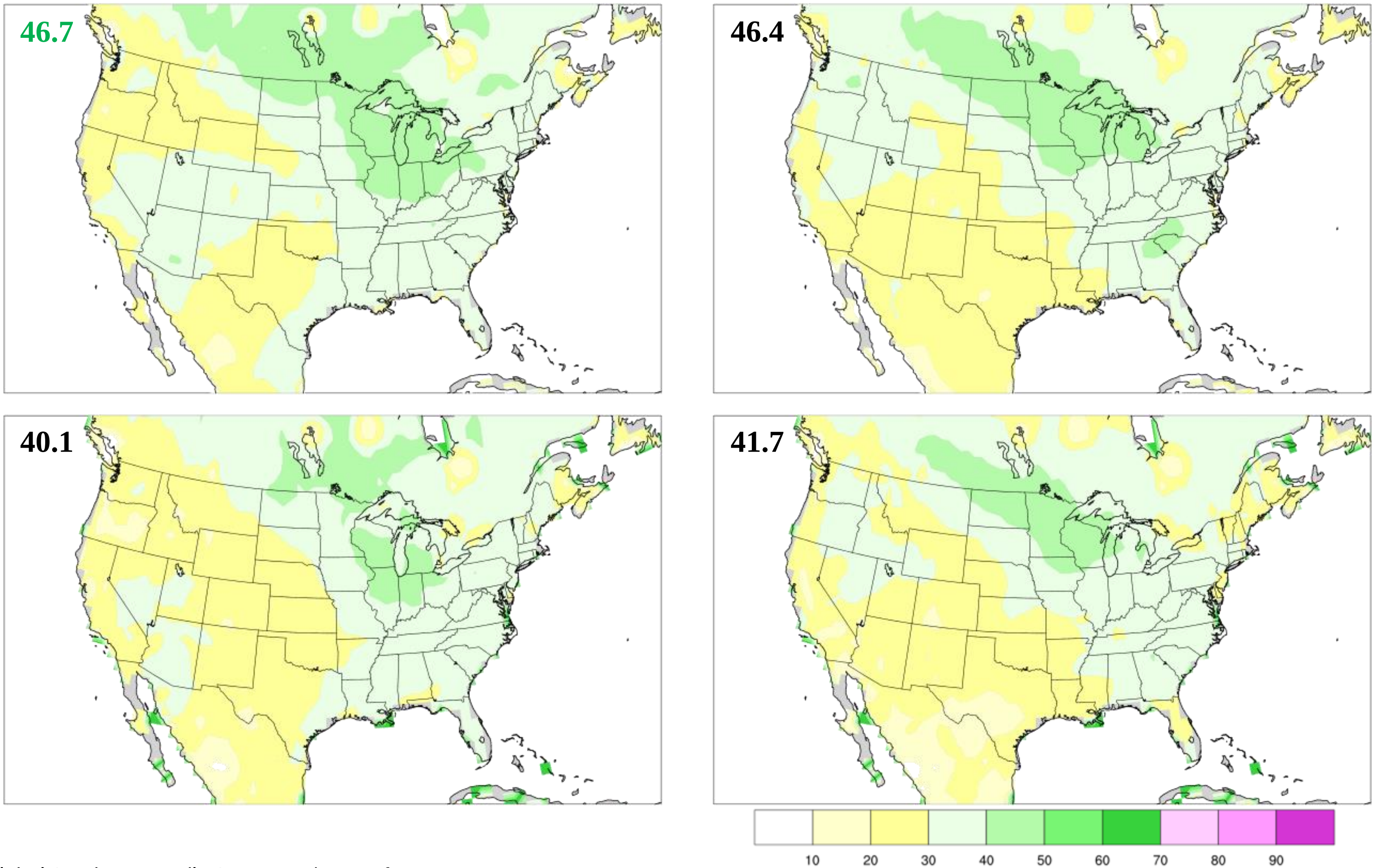
*CPC Global 0.5 degree Daily 2-m TMIN/TMAX from:
ftp://ftp.cpc.ncep.noaa.gov/precip/wd52ws/global_temp/
e.g., CPC_GLOBAL_T_V0.x_0.5deg.lnx.YYYY

Week 2
T2m AC

UGCSbench | CFSv2ops

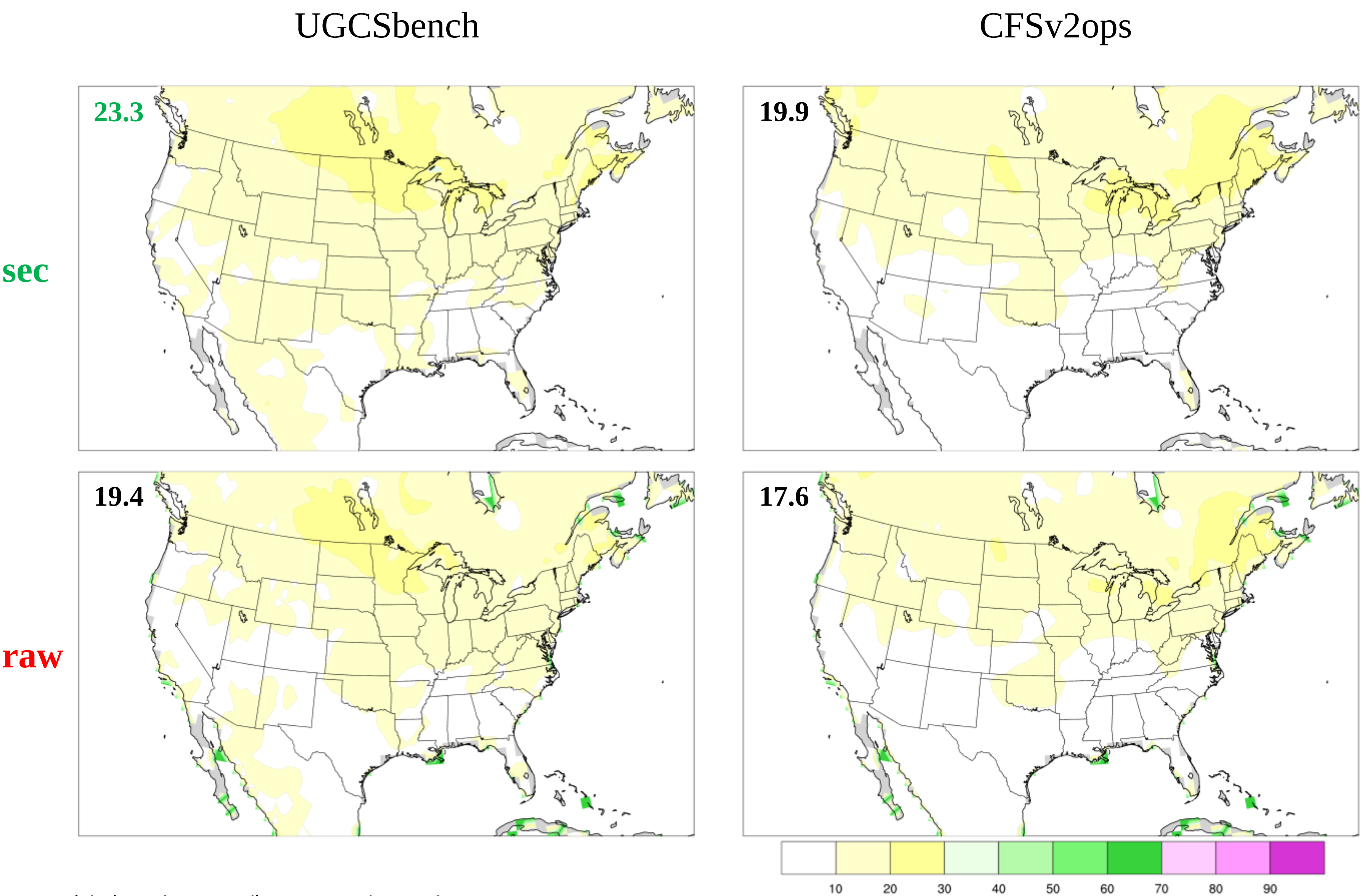
sec

raw



*CPC Global 0.5 degree Daily 2-m TMIN/TMAX from:
ftp://ftp.cpc.ncep.noaa.gov/precip/wd52ws/global_temp/
e.g., CPC_GLOBAL_T_V0.x_0.5deg.lnx.YYYY

Week 3
T2m AC



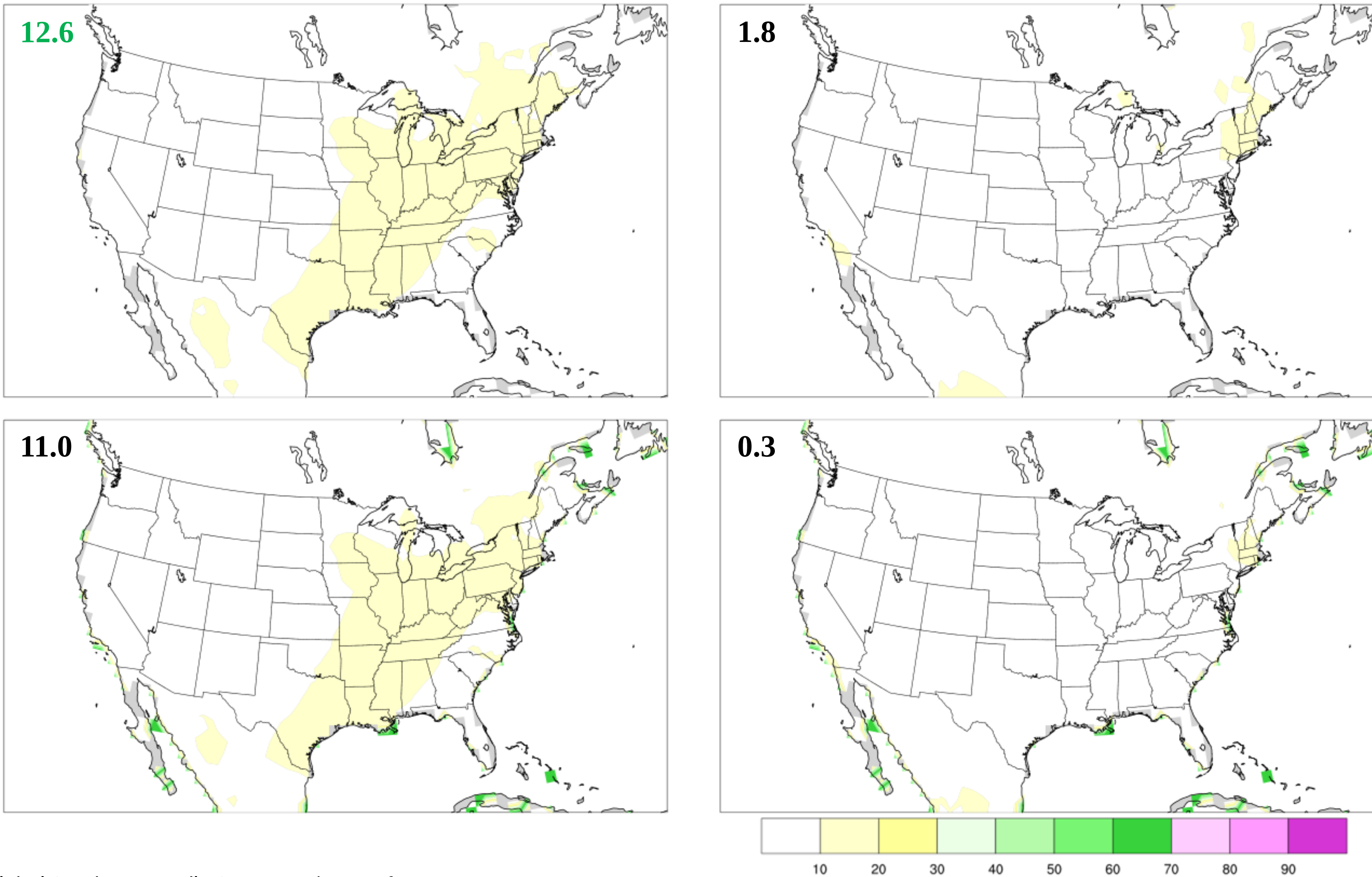
*CPC Global 0.5 degree Daily 2-m TMIN/TMAX from:
ftp://ftp.cpc.ncep.noaa.gov/precip/wd52ws/global_temp/
e.g., CPC_GLOBAL_T_V0.x_0.5deg.lnx.YYYY

Week 4
T2m AC

UGCSbench CFSv2ops

sec

raw



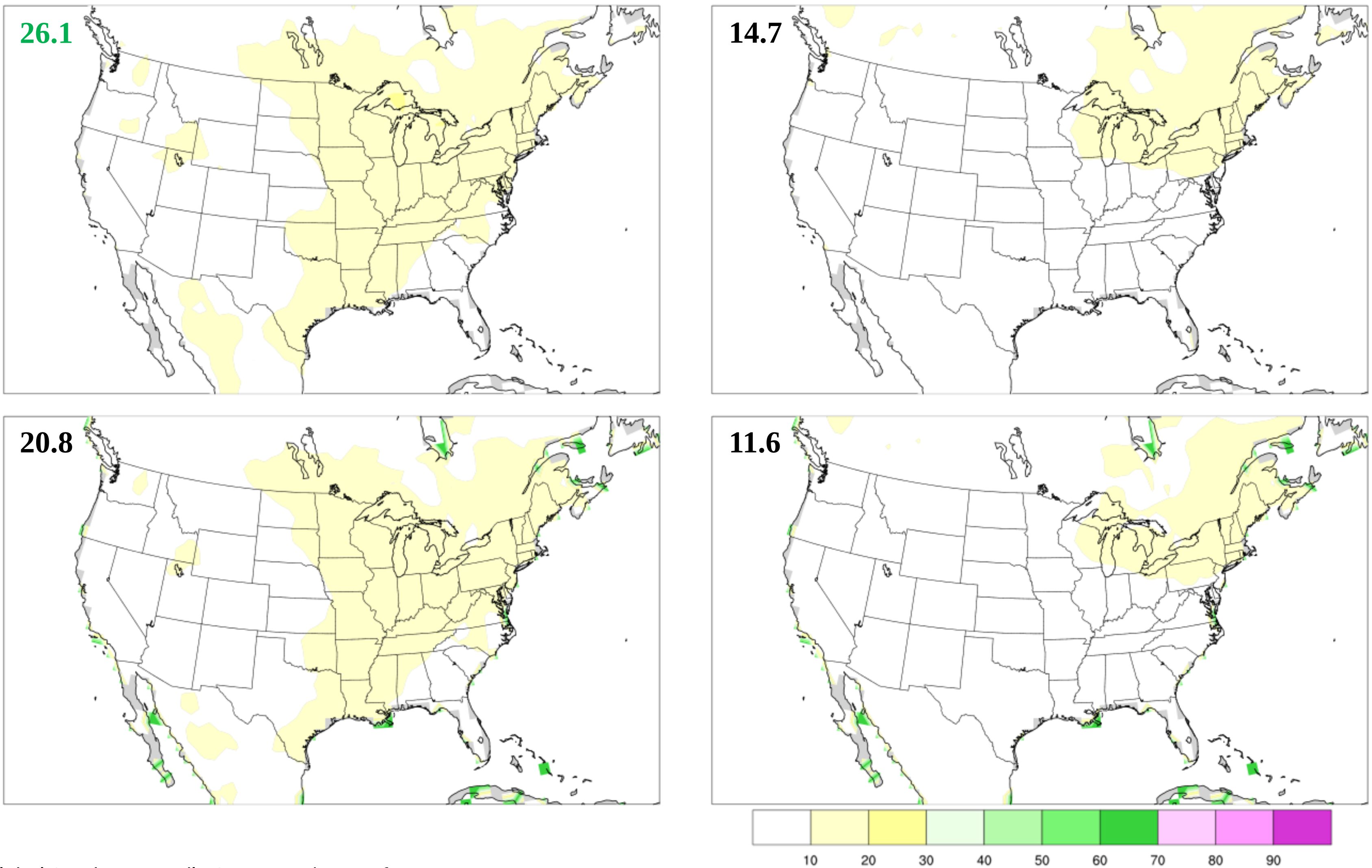
*CPC Global 0.5 degree Daily 2-m TMIN/TMAX from:
ftp://ftp.cpc.ncep.noaa.gov/precip/wd52ws/global_temp/
e.g., CPC_GLOBAL_T_V0.x_0.5deg.lnx.YYYY

Week 3 & 4
T2m AC

UGCSbench CFSv2ops

sec

raw



*CPC Global 0.5 degree Daily 2-m TMIN/TMAX from:
ftp://ftp.cpc.ncep.noaa.gov/precip/wd52ws/global_temp/
e.g., CPC_GLOBAL_T_V0.x_0.5deg.lnx.YYYY

UGCS (NEMS GSM+MOM5.1+1CICE5) - Saha, Melhauser, Pena et al.

CONUS Precipitation

CONUS Precipitation AC (CPC Unified Rain Gauge*)

	UGCSbench	UGCSbench	CFSv2ops	CFSv2ops
	Raw	**Sec**	**Raw**	**Sec**
week1	51.7	**56.0**	48.6	53.0
week2	16.3	18.2	18.0	**19.9**
week3	4.9	**5.4**	3.2	3.5
week4	0.9	**1.0**	0.7	0.7
week3&4	3.1	3.6	3.3	**3.7**

UGCSbench generally performs as well or better than CFSv2ops for most lead times.

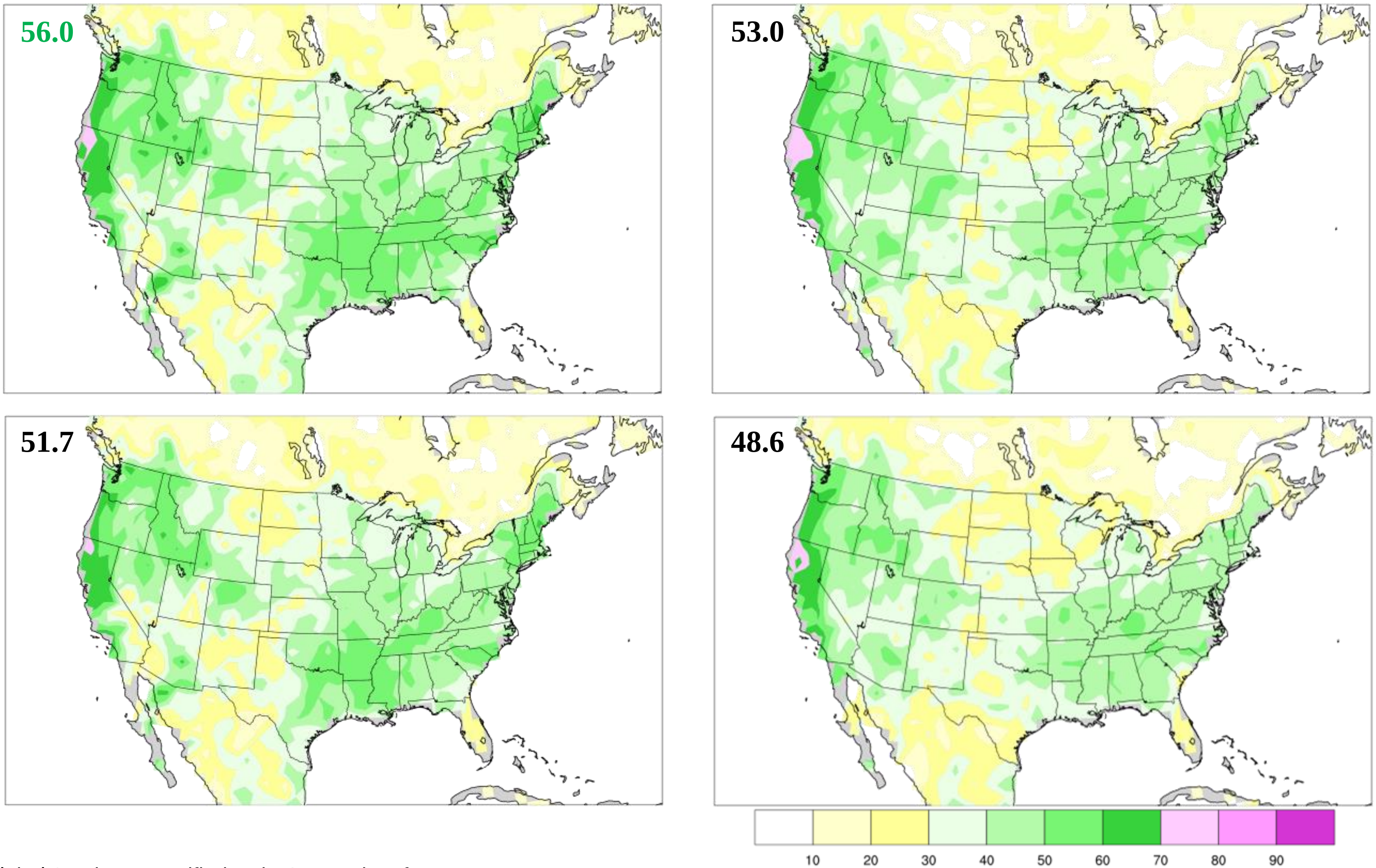
*CPC Global 0.5 degree Unified Rain Gauge data from:
ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB/
e.g., PRCP_CU_GAUGE_V1.0GLB_0.50deg.lnx.YYYYMMDDRT

Week 1
PRATE AC

UGCSbench | CFSv2ops

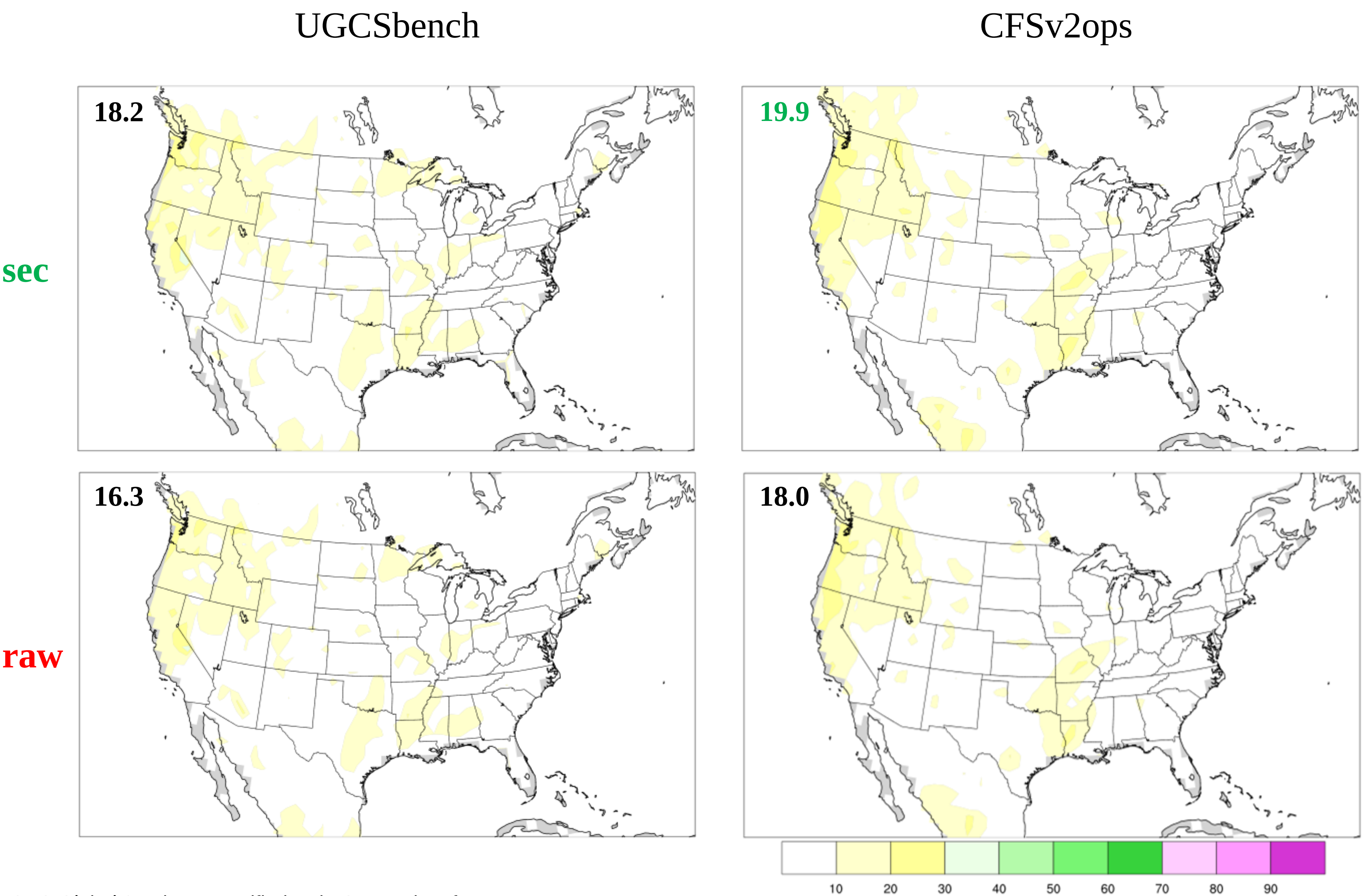
sec

raw



*CPC Global 0.5 degree Unified Rain Gauge data from:
ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB/
e.g., PRCP_CU_GAUGE_V1.0GLB_0.50deg.lnx.YYYYMMDDRT

Week 2
PRATE AC



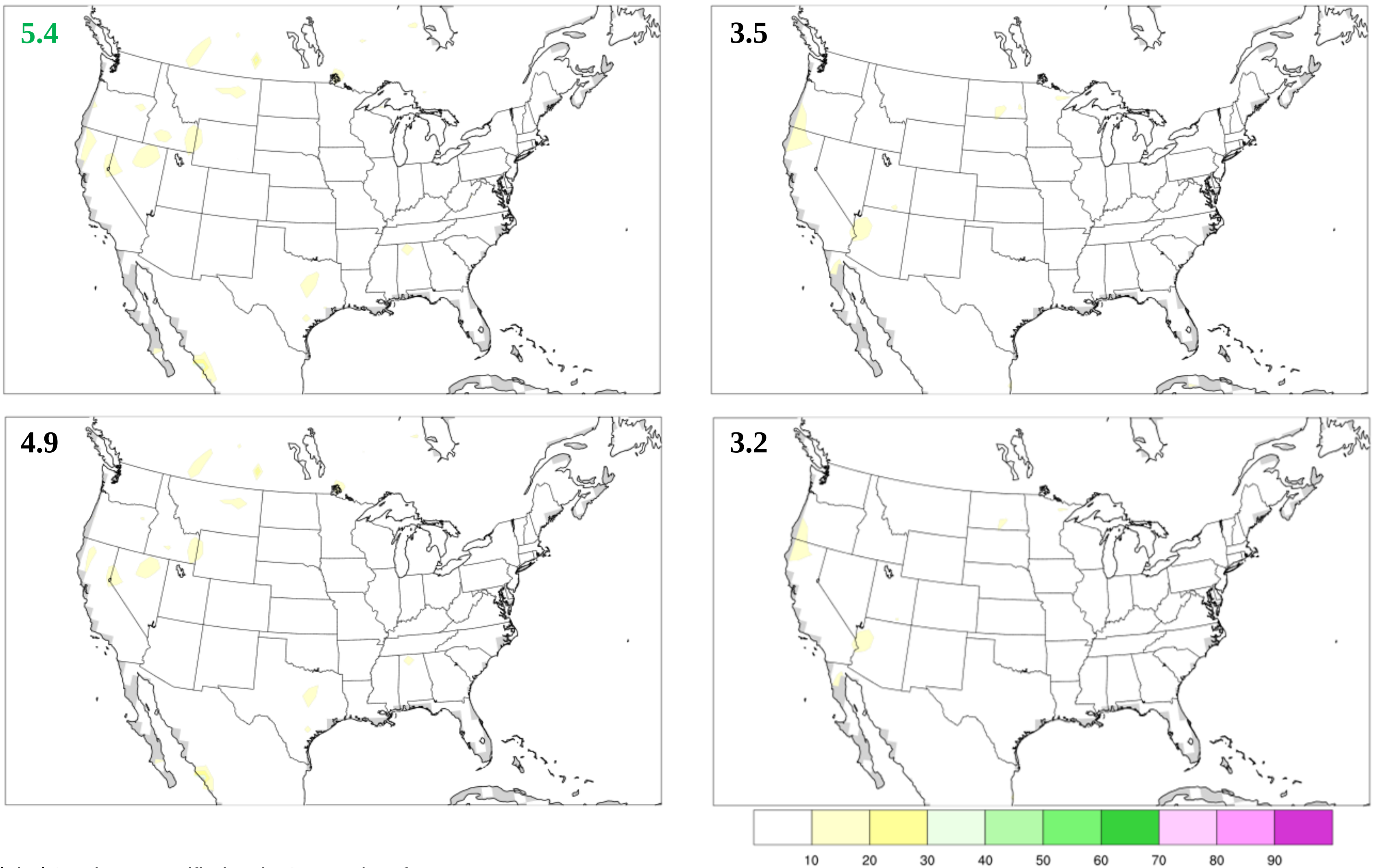
*CPC Global 0.5 degree Unified Rain Gauge data from:
ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB/
e.g., PRCP_CU_GAUGE_V1.0GLB_0.50deg.lnx.YYYYMMDDRT

Week 3
PRATE AC

UGCSbench CFSv2ops

sec

raw



*CPC Global 0.5 degree Unified Rain Gauge data from:
ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB/
e.g., PRCP_CU_GAUGE_V1.0GLB_0.50deg.lnx.YYYYMMDDRT

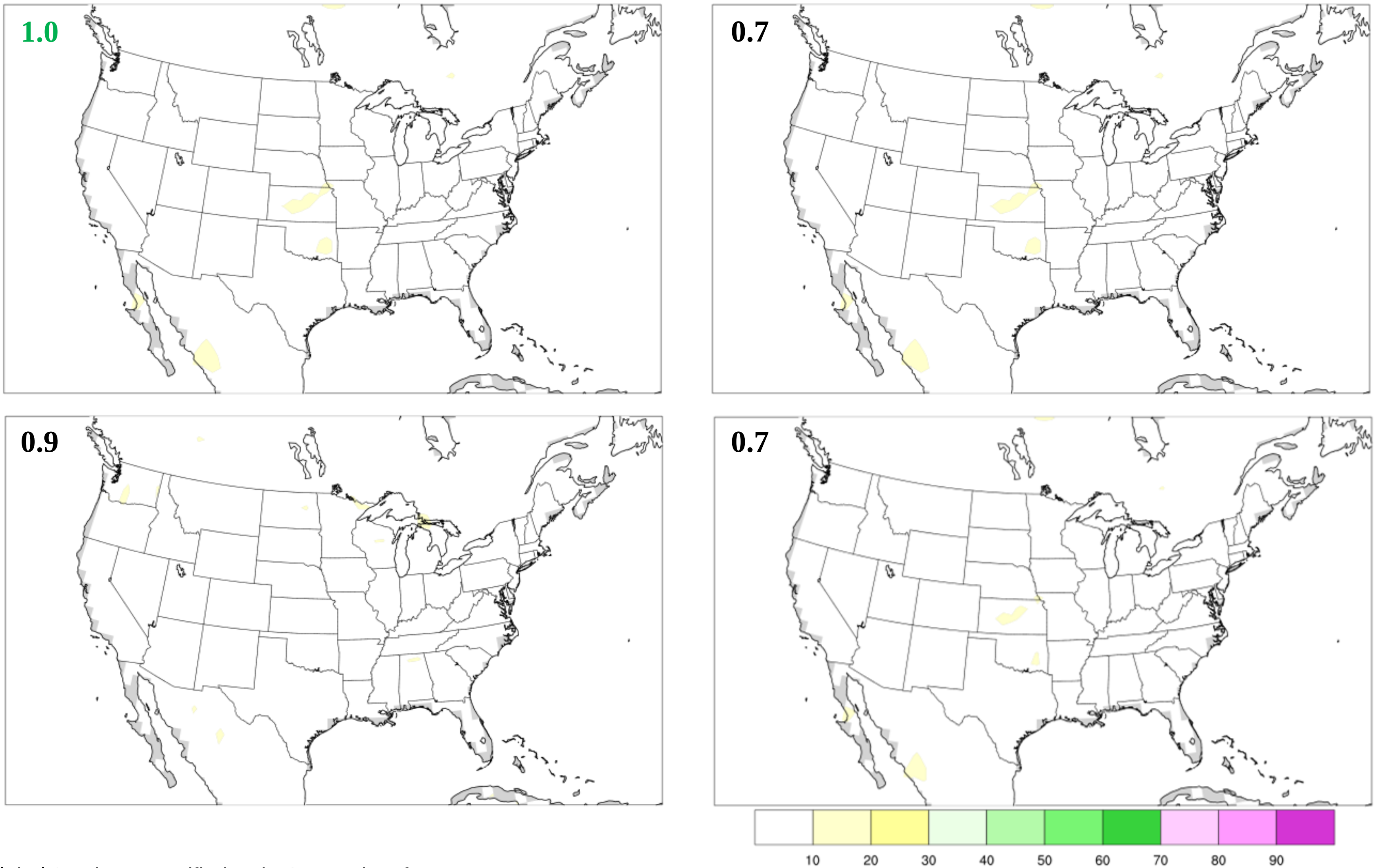
UGCS (NEMS GSM+MOM5.1+1CICE5) - Saha, Melhauser, Pena et al.

Week 4
PRATE AC

UGCSbench CFSv2ops

sec

raw



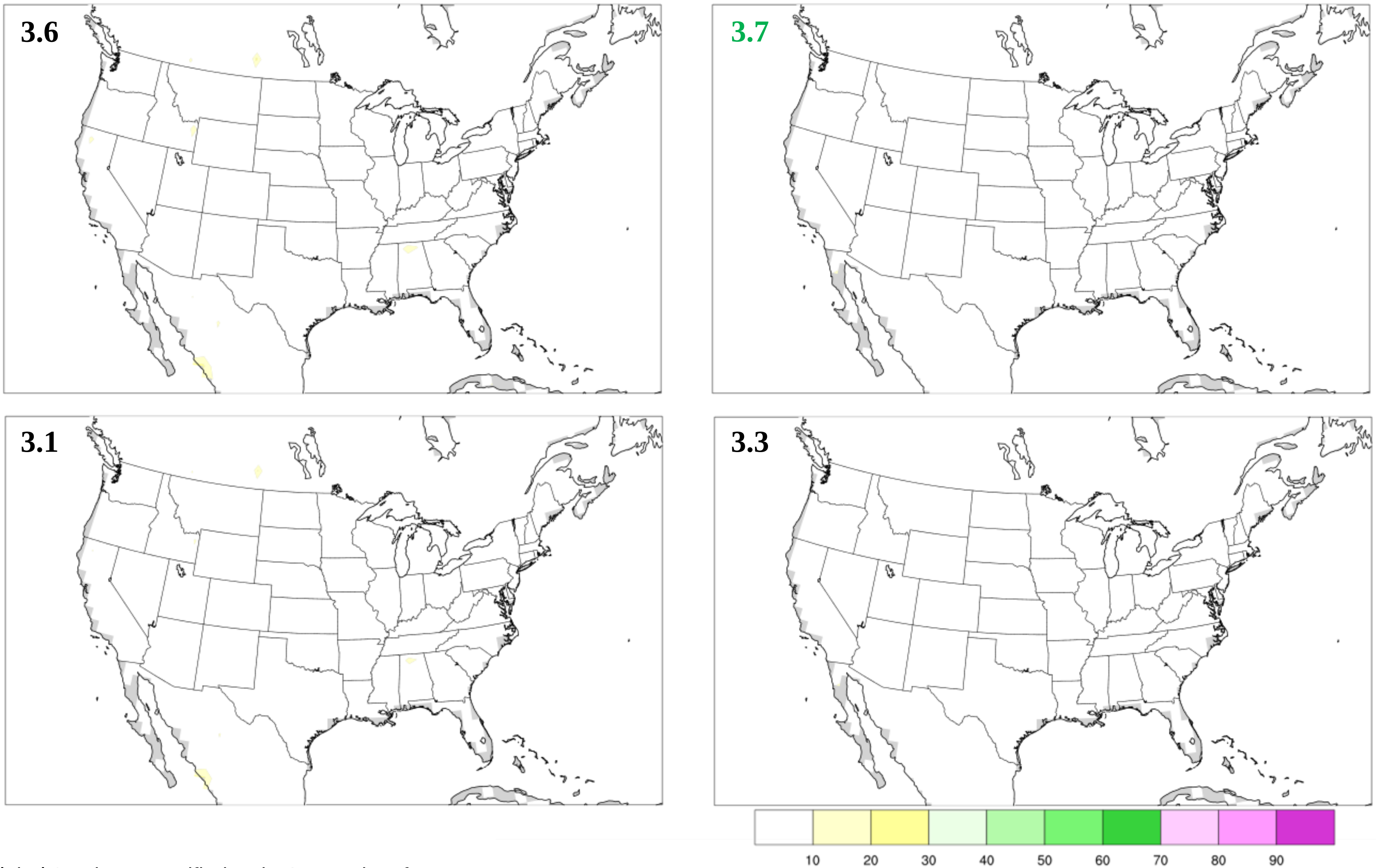
*CPC Global 0.5 degree Unified Rain Gauge data from:
ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB/
e.g., PRCP_CU_GAUGE_V1.0GLB_0.50deg.lnx.YYYYMMDDRT

Week 3 & 4
PRATE AC

UGCSbench CFSv2ops

sec

raw

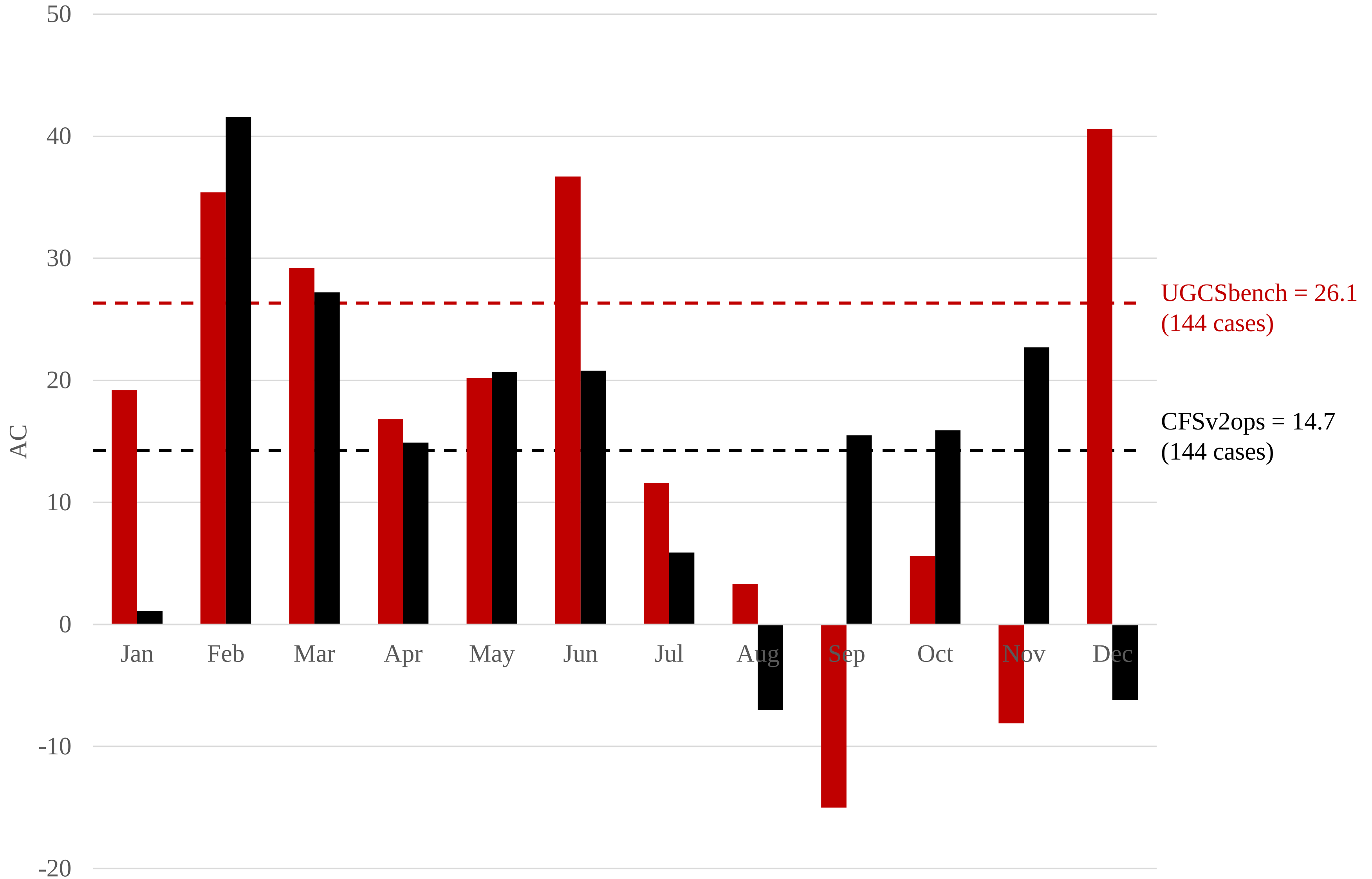


*CPC Global 0.5 degree Unified Rain Gauge data from:
ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB/
e.g., PRCP_CU_GAUGE_V1.0GLB_0.50deg.lnx.YYYYMMDDRT

Box averages for US-land points and Nino3.4 points for week 3 & 4 forecasts

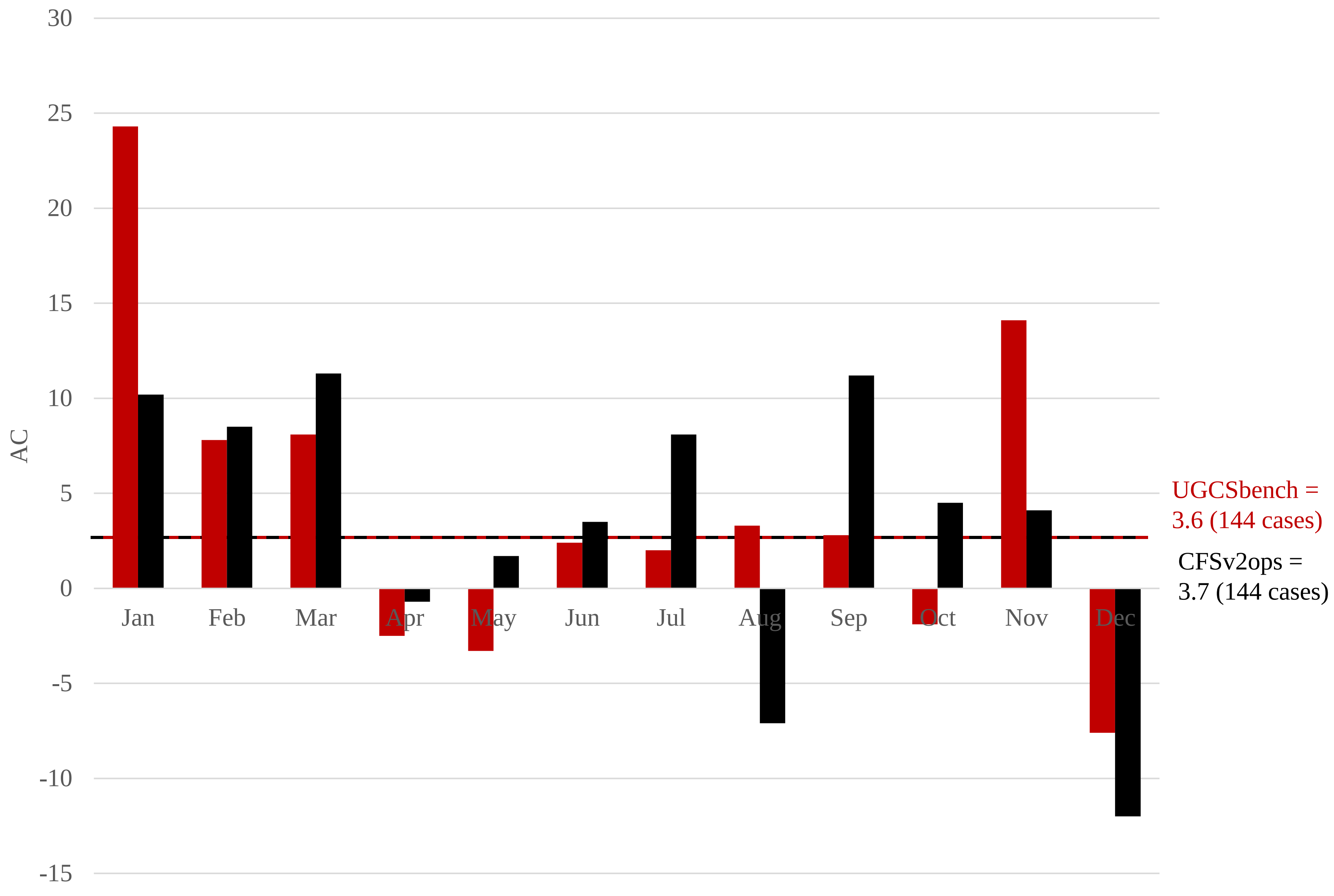
- T2m and Prate for CONUS (verification is CPC Global 0.5 degree Daily 2-m TMIN/TMAX and CPC Global 0.5 degree Unified Rain Gauge data)
- SST and Prate for Nino3.4 (verification is CFSR)
- Bar diagrams with seasonality (warning: not a large enough sample size to study seasonality)
- Averaged scores across all seasons

CONUS average of week 3 & 4 SEC AC for T2m forecast
(Each bar based on 12 cases with IC in the month indicated)



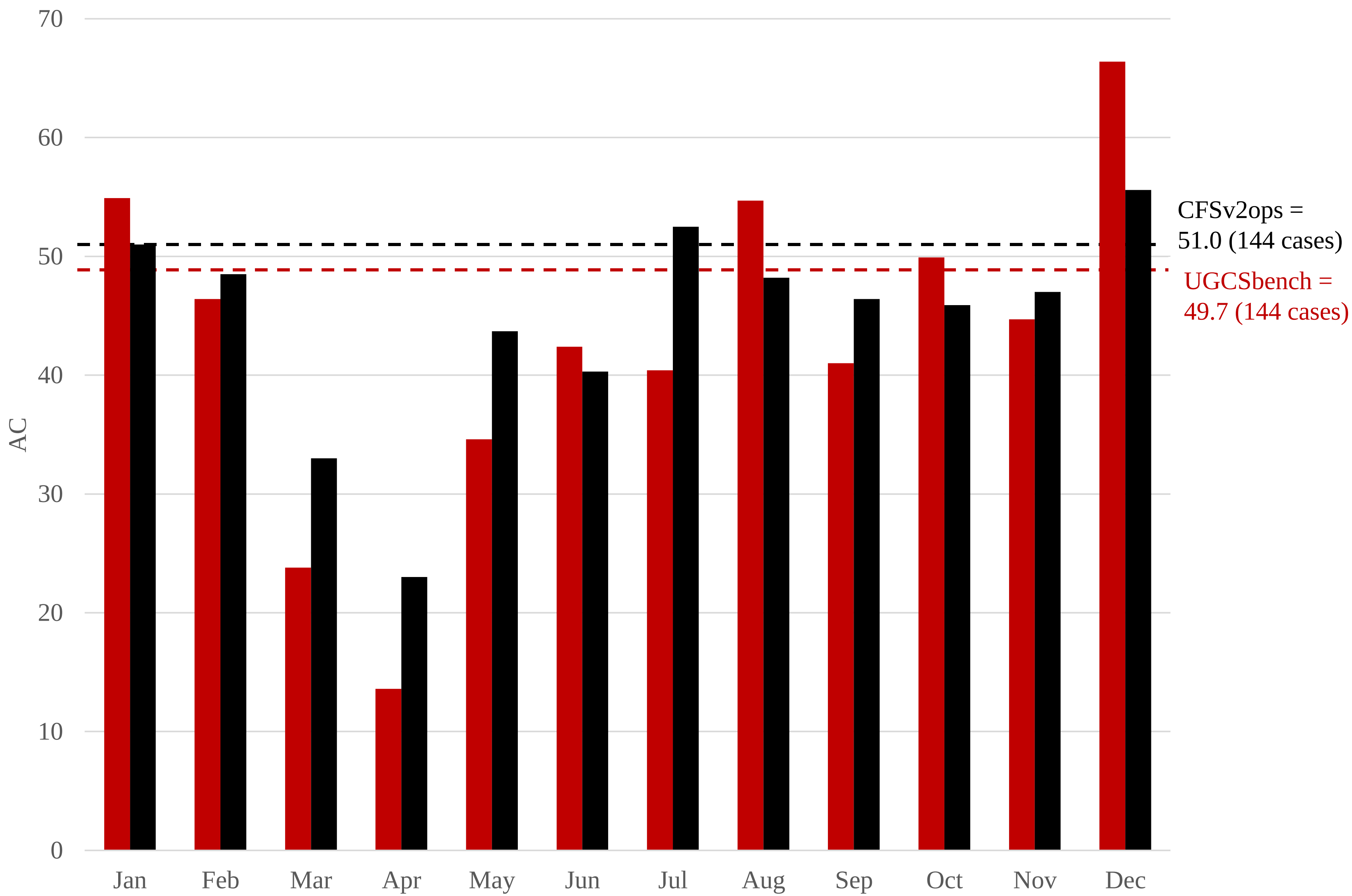
*CPC Global 0.5 degree Daily 2-m TMIN/TMAX from:
ftp://ftp.cpc.ncep.noaa.gov/precip/wd52ws/global_temp/
e.g., CPC_GLOBAL_T_V0.x_0.5deg.lnx.YYYY

CONUS average of week 3 & 4 SEC AC for PRATE forecast
(Each bar based on 12 cases with IC in the month indicated)



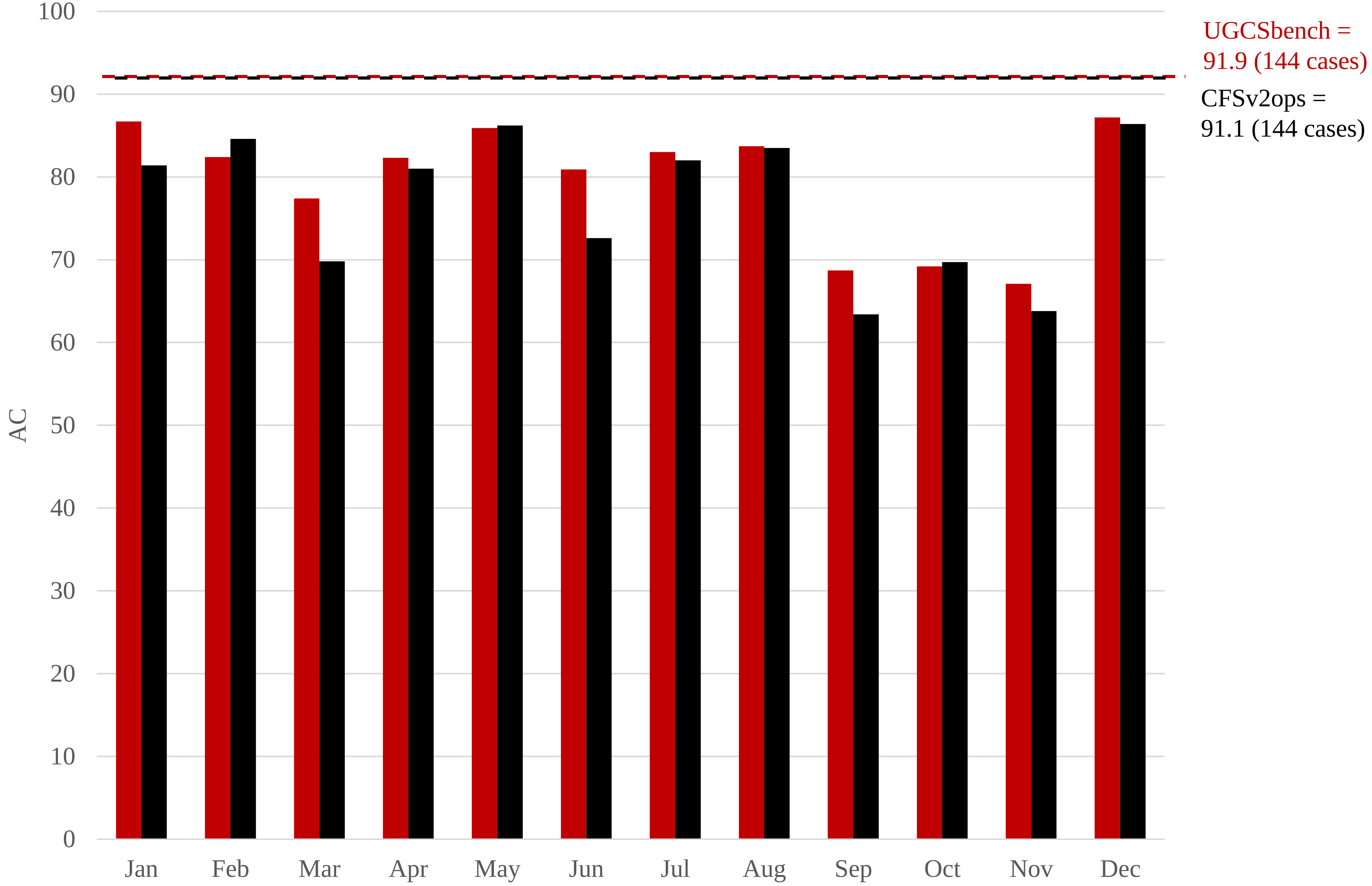
*CPC Global 0.5 degree Unified Rain Gauge data from:
ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB/
e.g., PRCP_CU_GAUGE_V1.0GLB_0.50deg.lnx.YYYYMMDDRT

Nino3.4 average of week 3 & 4 SEC AC for PRATE forecast
(Each bar based on 12 cases with IC in the month indicated)



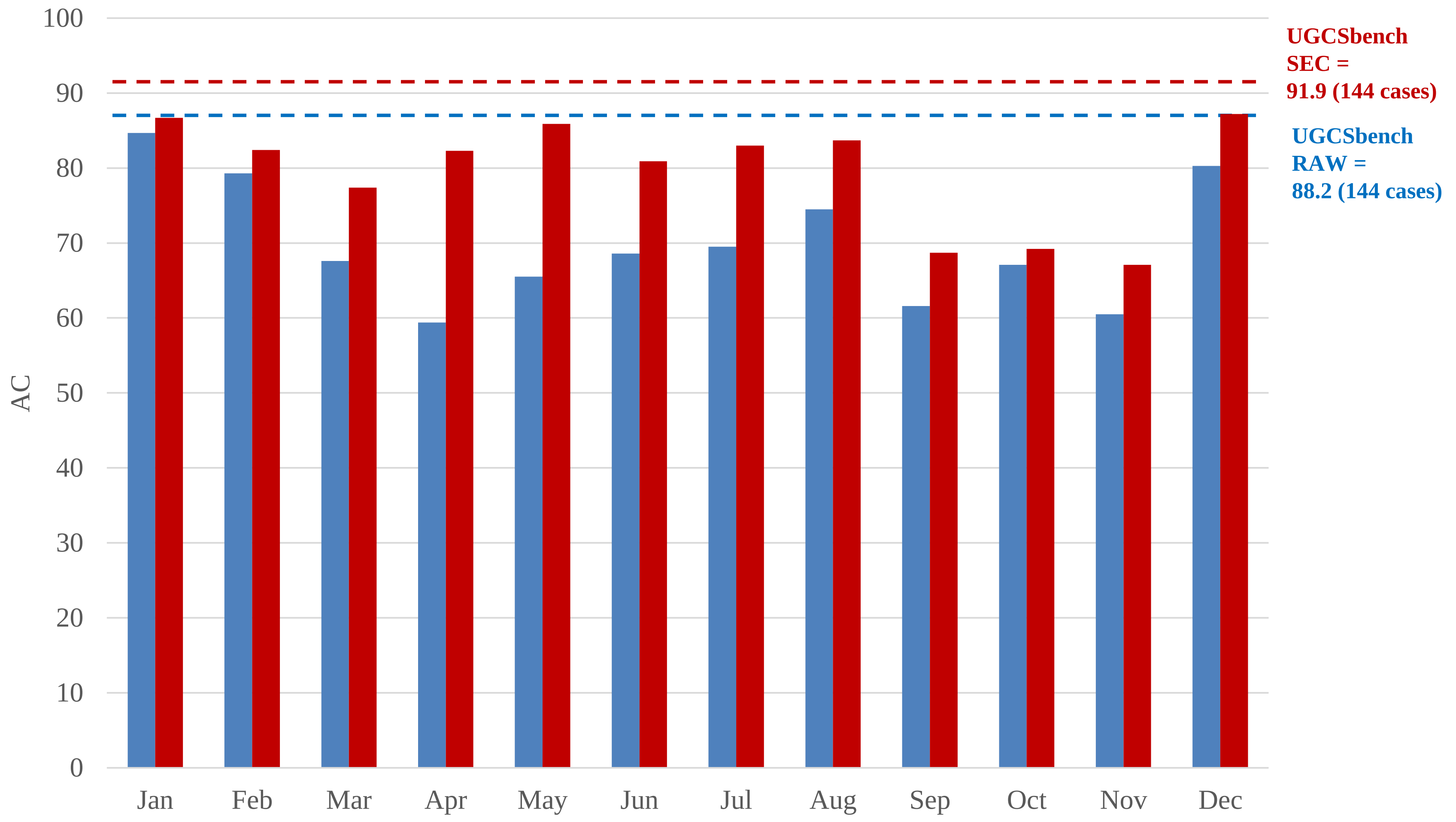
CFSR used for verification

Nino3.4 average of week 3 & 4 SEC AC for SST forecast
(Each bar based on 12 cases with IC in the month indicated)



CFSR used for verification

Nino3.4 average of week 3 & 4 SEC AC for SST forecast
(Each bar based on 12 cases with IC in the month indicated)



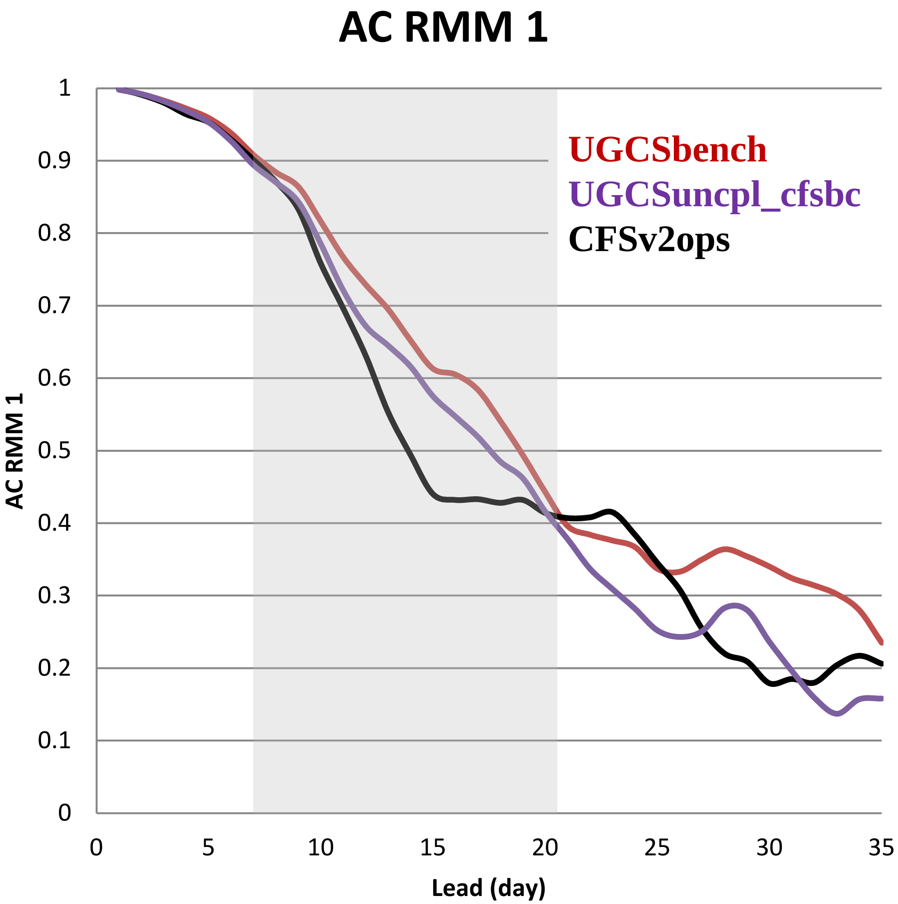
After SEC, the scores for SST in Nino3.4 area are improved.

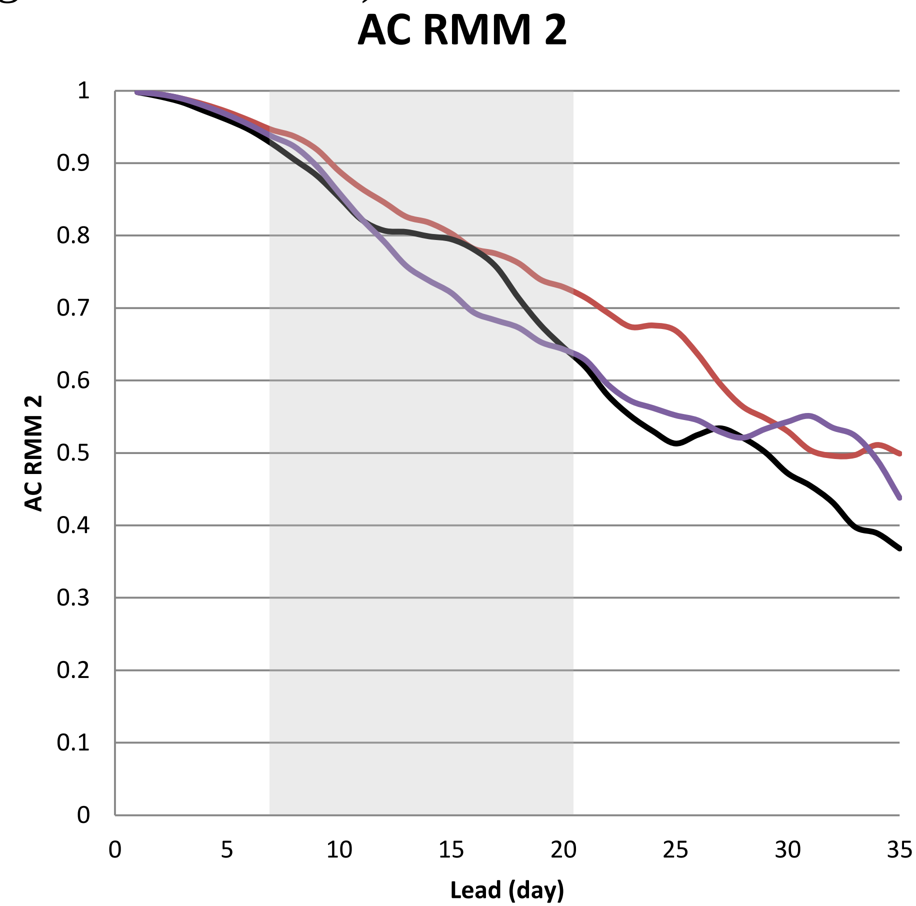
CFSR used for verification

Madden Julian Oscillation (MJO) Leading modes of forecast skill

MJO Anomaly Correlation Skill

(144 cases covering 04/2011-03/2017)





Week 2 & 3: UGCSbench generally has higher skill than uncoupled and CFSv2

All-seasons MJO's two leading modes (RMM1 and RMM2) of the combined timeseries of OLR, U850 and U200 equatorial anomalies. RMM1 series has the largest amplitude in the Maritime Continent and (negative) in the West. Hem. and Africa; RMM2 has largest amplitude in the Western Pacific and (negative) in the Indian Ocean.

CFSR verification u200,u850 and OLR

Conclusions

- Overall interpretation is that the UGCS Coupled run improves the skill of the equatorial zonal wind (for both 850 and 200 hPa), particularly weeks 2 and 3 over the uncoupled.

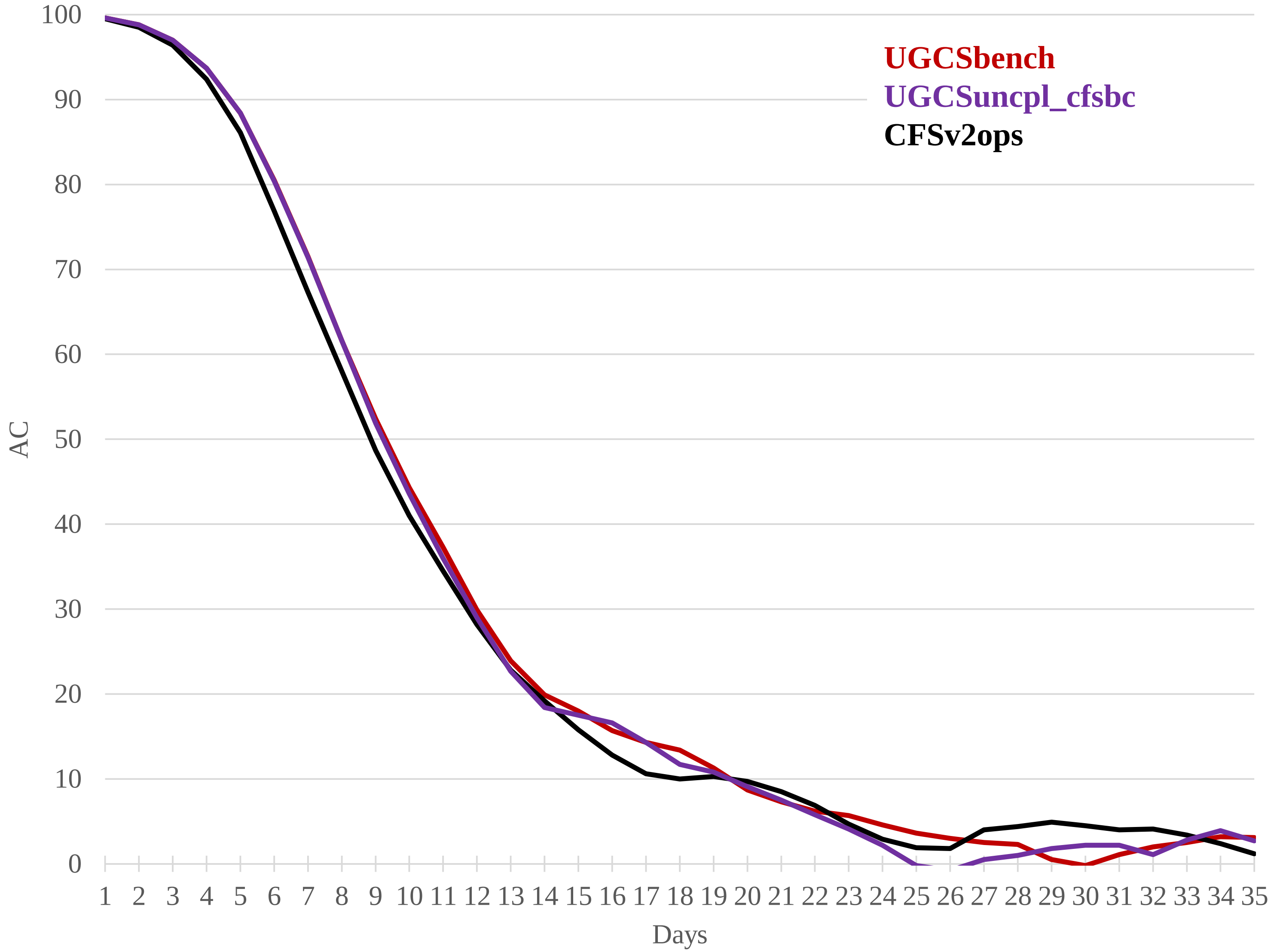
- This in turn improves the RMM1 and RMM2 for that period.

- No conclusions can be made for week 4, due to the large sampling error. The results do suggest that improvement from a coupled system for winds goes beyond week 3.

- For the OLR variable, the results are mixed.

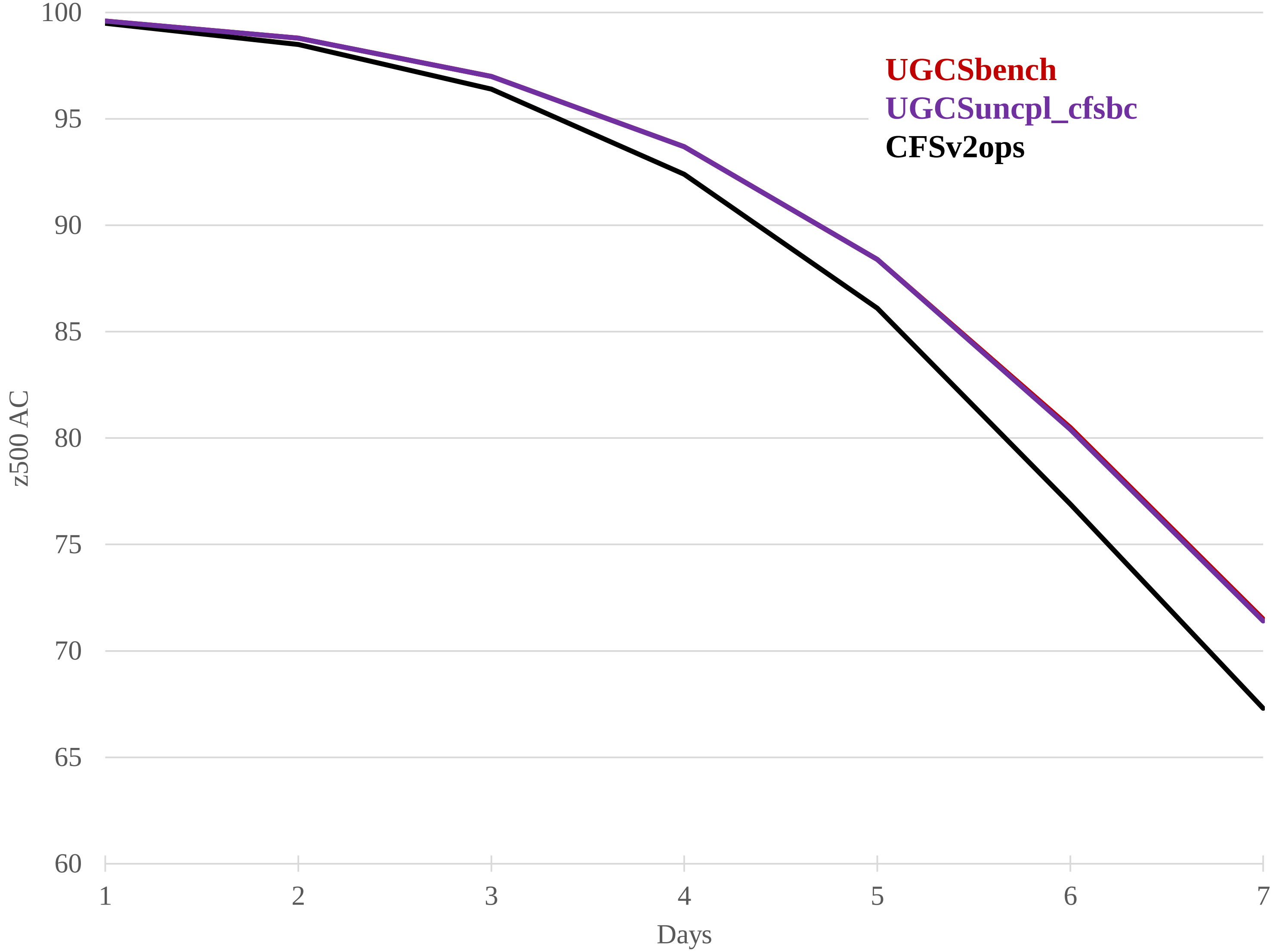
500 hPa Geopotential

500hPa Geopotential NH (20N-80N) AC
Weeks 1-5 (days 1 – 35)



CFSR used for verification

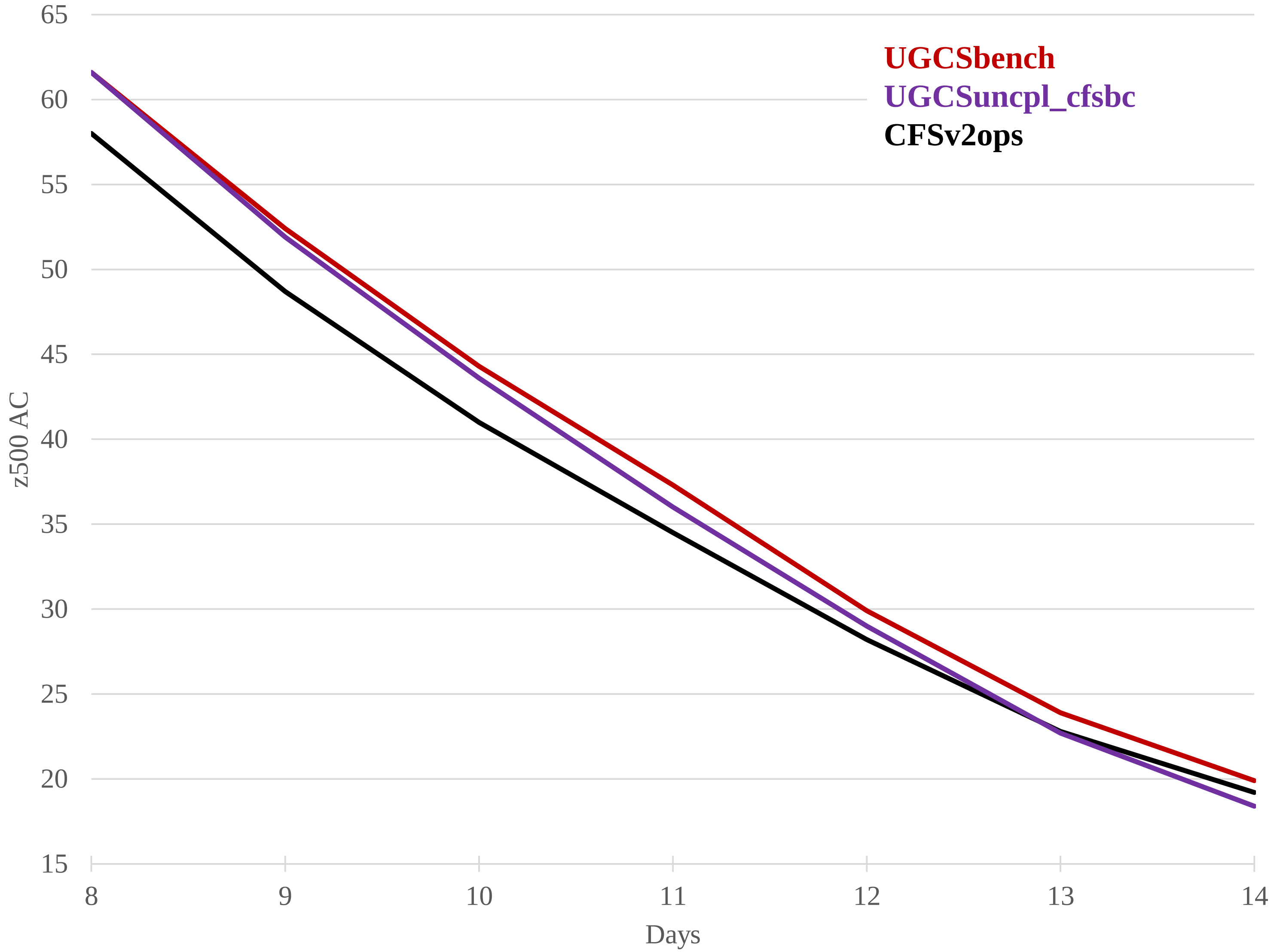
500hPa Geopotential NH (20N-80N) AC
Weeks 1 (days 1 – 7)



CFSR used for verification

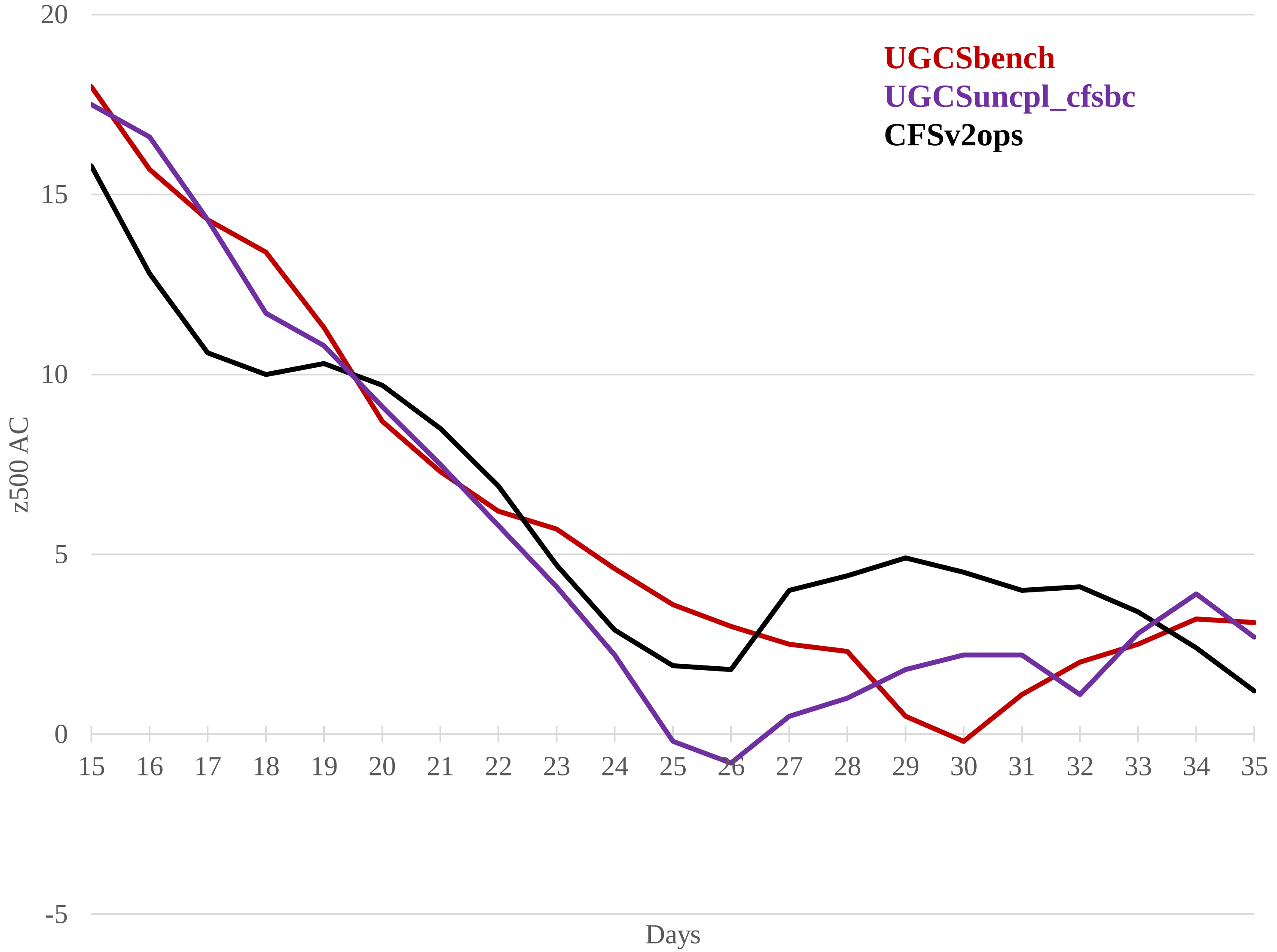
UGCS (NEMS GSM+MOM5.1+1CICE5) -
Saha, Melhauser, Pena et al.

500hPa Geopotential NH (20N-80N) AC
Weeks 2 (days 8-14)



CFSR used for verification

500hPa Geopotential NH (20N-80N) AC
Weeks 3-5 (days 15-35)



CFSR used for verification

500hPa Geopotential NH (20N-80N) AC

	UGCS uncpl_cfsbc	UGCS uncpl_cfsbc	UGCS bench	UGCS bench	CFSv2ops	CFSv2ops
	Raw	**Sec**	**Raw**	**Sec**	**Raw**	**Sec**
week1	96.1	**96.6**	96.1	**96.6**	95.0	95.9
week2	52.2	54.6	52.9	**55.3**	49.7	52.6
week3	19.3	**20.9**	18.8	20.2	17.0	18.7
week4	3.1	3.3	7.1	**7.7**	5.8	6.5
week3&4	12.9	14.2	15.9	**17.5**	13.9	15.6

Conclusions:

- UGCSbench does not hurt the uncoupled UGCSuncpl_cfsbc scores at week1.
- UGCSbench is generally better than the uncoupled UGCSuncpl_cfsbc scores after week1.
- UGCSbench is better than the CFSv2ops for all lead times.

CFSR used for verification

Annually aggregated week3&4 AC scores
UGCSbench vs CFSv2ops, Raw vs SEC

	UGCSbench	CFSv2ops	Region
Raw	20.8	11.6	T2m-land-CONUS (CPC)
SEC	**26.1**	14.7	**T2m-land-CONUS (CPC)**
Raw	3.1	3.3	Prate-land – CONUS (CPC)
SEC	3.6	**3.7**	**Prate-land – CONUS (CPC)**
Raw	46.7	47.2	Prate-Nino34 (CFSR)
SEC	49.7	**51.0**	**Prate-Nino34 (CFSR)**
Raw	88.2	87.9	SST-ocean-Nino34 (CFSR)
SEC	**91.9**	91.1	**SST-ocean-Nino34 (CFSR)**
Raw	15.9	13.9	Z500 –ext- NH (CFSR)
SEC	**17.5**	15.6	**Z500 –ext- NH (CFSR)**

Conclusions

- All variables studied (Z500, SST, T2m land, Prate-land and Prate-ocean) show that UGCSbench is generally equal to, or better than CFSV2ops, over the extratropical NH.

- In particular, UGCSbench is generally equal or better than CFSv2ops for week3&4 over CONUS land points, for both T2m and Prate.

- In the Tropics (Equatorial Pacific Ocean), UGCSbench is worse than CFSv2ops for Prate, and slightly better for SST, but only after correcting for a substantial systematic error in SST.

Conclusions

- The Systematic Error Correction (SEC) appears to function satisfactorily, even with runs only every two weeks for 6 years.

- Prate and Z500 have little SE, or little “correctable” SE.

- In contrast, T2m and SST have larger SE, and the need for generating hindcasts for SEC is pressing.

- Die-off curves day 1 to 35 for Z500, T2m, Prate-land, Prate-ocean and SST, all for extra-tropical NH, provides evidence that UGCSbench is generally equal or better than CFCsv2ops.

- Obviously, we need a much larger sample size to compute more accurate calibration climatologies and skill estimates. We hope to fill in the days of each month in this 6-year sample.

Bottom Line

The current UGCS configuration (NEMS GSM+MOM5.1+CICE5) is a working coupled model that is already a candidate to replace CFSv2 in operations.

The following future enhancements will only serve to make it even more competitive:

1. Replacing the spectral model with the GFDL FV3 dynamic core for the atmospheric model component (work underway)
2. Replacing the MOM5.1 with the more advanced GFDL MOM6 for the ocean model component (work underway)
3. Working towards improving the coupling physics with the new FV3 dynamic core (work underway)
4. Working towards an FV3 based weakly coupled data assimilation system, based on the hybrid EnKF approach to all component systems (work underway).
5. Working towards a full ensemble of coupled model members with consistent initial perturbations to all components.
6. Reanalysis and Retrospective forecasts for consistent and appropriate systematic error correction, as well as skill estimation.
7. Working towards a full end-to-end workflow infrastructure that includes full validation metrics (work underway).

This is the normal path for any major coupled modeling system to be developed