Validation/Evaluation

1) Baseline evaluation of Forecasts
- skill (anomaly correlation, rmse, BSS, etc)
- potential predictability
- forecast spread and consistency with obs (e.g., rmse, rank histograms)
- realism of variance, autocorrelation, extremes
- climate drift

2) Forecasts of opportunity (assess conditional skill)
- ENSO/other teleconnections (e.g., soil moisture teleconnections?)
- In-depth probabilistic evaluation of forecasts
    - (reliability, resolution, ROC, etc)

3) Ability of model to simulate key modes of climate variability
- ENSO and other SST modes
- MJO, PNA,NAO, etc (important for subseasonal as signal, important for seasonal and longer as noise)

S. Schubert: 18 January 2017

Results from Becker et al. with a focus on the GEOS-5 S2S-1.0 Model Performance

Becker, E., H. van den Dool, Q. Zhang, 2014: Predictability and Forecast Skill in NMME. J. Climate, 27, 5891-5906.

Various Measures

(see Appendix of Becker et al)

- Skill
  - anomaly correlation
  - rmse
- Potential predictability
  - homogeneous (heterogeneous) anomaly correlation
  - homogeneous (heterogeneous) rms
- Forecast Spread
  - rms about ensemble mean
  - compare to rmse to assess adequacy of dispersion
- Variance of single ensemble member
  - $(\text{singmember\_SD})^2 = (\text{EM\_SD})^2 + \text{Spread}^2$
  - can compare to observed variance

NINO3.4 region SST: Lead 1 month, aggregated over all seasons (3month): 1982-2010; 1982-2009 for cfsv1

Estimates of potential predictability

Estimates of skill

A. SST Nino3.4 Region Season 1										
	cfsv1	cfsv2	cmc1	cmc2	gfdl	nasa	ncar	obs (EM AC)	EM RMSE (C)	*EM SD*
cfsv1 EM	94	74	85	86	78	87	82	82	0.61	*1.12*
cfsv2 EM	69	93	78	76	80	86	80	82	0.58	*1.03*
cmc1 EM	84	76	96	93	82	90	83	87	0.43	*0.83*
cmc2 EM	83	73	92	97	80	88	82	85	0.57	*1.13*
gfdl EM	76	78	83	82	94	87	80	80	0.65	*1.13*
nasa EM	83	84	88	87	84	98	85	88	0.56	*1.17*
ncar EM	79	78	82	82	79	85	95	80	0.67	*1.14*
singmem AC	76	76	84	83	75	85	78	NMME=89		
singmem RMSE	0.74	0.70	0.46	0.63	0.73	0.62	0.70			
singmem & obs SD	*1.19*	*1.12*	*0.85*	*1.16*	*1.17*	*1.22*	*1.15*	*0.89*		

A	B
0.65	1.3
0.75	1.3
0.41	1.0
0.45	1.3
0.46	1.3
0.62	1.4
0.22	1.3

FIG. 9. As in Fig. 1, but for sea surface temperature in (a) the Niño-3.4 region (aggregated results for all grid points in the area 5°S–5°N, 170°–120°W) and (b) the Niño-3.4 index (area-averaged SST).

A. dispersion/rmse
B. Single member SD/obs SD

All models are under dispersive

Most models have too much variability

NH SST: Lead 1 month, aggregated over all seasons (3month): 1982-2010 (NMME models), 1982-2009 for cfsv1

Estimates of potential predictability

Estimates of skill

SST Northern Hemisphere Season 1

	cfsv1	cfsv2	cmc1	cmc2	gfdl	nasa	ncar	obs (EM AC)	EM RMSE (C)	*EM SD*
cfsv1 EM	49	28	38	37	33	27	23	29	0.63	*0.39*
cfsv2 EM	23	68	38	40	38	47	14	41	0.62	*0.52*
cmc1 EM	26	34	71	64	44	35	23	44	0.55	*0.40*
cmc2 EM	26	35	62	74	45	36	21	46	0.55	*0.45*
gfdl EM	22	32	42	43	78	34	17	42	0.59	*0.46*
nasa EM	19	43	33	36	36	74	12	35	0.66	*0.55*
ncar EM	16	11	21	19	16	12	81	15	0.76	*0.54*
singmem AC	15	27	35	36	35	27	12	NMME=50		
singmem RMSE	0.83	0.82	0.65	0.66	0.69	0.81	0.82			
singmem & obs SD	*0.64*	*0.75*	*0.54*	*0.58*	*0.57*	*0.71*	*0.63*	*0.62*		

FIG. 7. As in Fig. 1, but for sea surface temperature aggregate scores over Northern Hemisphere ocean, between 23° and 75°N.

A	B
0.81	1.0
0.87	1.2
0.65	.87
0.67	.94
0.58	.92
0.68	1.1
0.42	1.0

A. dispersion/rmse
B. Single member SD/obs SD

All models are under dispersive

Most models have realistic variability

T2m NH land: Lead 1 month, aggregated over all seasons (3month): 1982-2010; 1982-2009 for cfsv1

Estimates of potential predictability

Estimates of skill

TMP2m Northern Hemisphere Season 1

	cfsv1	cfsv2	cmc1	cmc2	gfdl	nasa	ncar	obs (EM AC)	EM RMSE (C)	*EM SD*
cfsv1 EM	29	12	12	11	10	16	8	12	1.41	*0.59*
cfsv2 EM	14	38	16	24	26	28	1	29	1.32	*0.62*
cmc1 EM	11	14	30	19	18	21	5	17	1.38	*0.60*
cmc2 EM	10	23	21	38	27	27	3	27	1.36	*0.72*
gfdl EM	9	23	16	25	36	26	0	25	1.38	*0.77*
nasa EM	12	25	19	24	25	39	2	23	1.37	*0.69*
ncar EM	6	1	4	6	3	3	19	0	1.62	*0.84*
singmem AC	3	13	7	16	13	11	0	NMME=29		
singmem RMSE	1.91	1.79	1.77	1.76	1.90	1.79	1.99			
singmem & obs SD	*1.41*	*1.35*	*1.26*	*1.38*	*1.54*	*1.34*	*1.46*	*1.38*		

A	**B**
0.91	1.0
0.91	1.0
0.80	.91
0.81	1.0
0.96	1.1
0.84	1.0
0.74	1.1

FIG. 1. Area-aggregated results for 2-m temperature, land-only Northern Hemisphere, 23°–75°N, averaged over the 12 lead-1 seasons. The seven rows and columns in black are results for predictability. The orange highlighted

A. dispersion/rmse
B. Single member SD/obs SD

All models are somewhat under dispersive

Variance is about right for most

Precip NH land: Lead 1 month, aggregated over all seasons (3month): 1982-2010; 1982-2009 for cfsv1

Estimates of potential predictability

Estimates of skill

PRATE Northern Hemisphere Season 1										
	cfsv1	cfsv2	cmc1	cmc2	gfdl	nasa	ncar	obs (EM AC)	EM RMSE	*EM SD*
cfsv1 EM	24	11	8	12	11	10	4	10	0.41	*0.21*
cfsv2 EM	13	20	7	11	9	10	4	12	0.38	*0.15*
cmc1 EM	6	6	16	14	9	7	6	9	0.40	*0.16*
cmc2 EM	10	6	13	25	13	10	5	11	0.40	*0.18*
gfdl EM	10	8	9	14	22	11	5	12	0.40	*0.20*
nasa EM	9	8	6	10	10	18	4	9	0.40	*0.19*
ncar EM	4	3	5	6	5	3	12	4	0.45	*0.25*
singmem AC	3	4	3	6	5	4	2	NMME=16		
singmem RMSE	0.67	0.61	0.56	0.56	0.60	0.61	0.65			
singmem & obs SD	*0.57*	*0.50*	*0.42*	*0.43*	*0.48*	*0.48*	*0.54*	*0.37*		

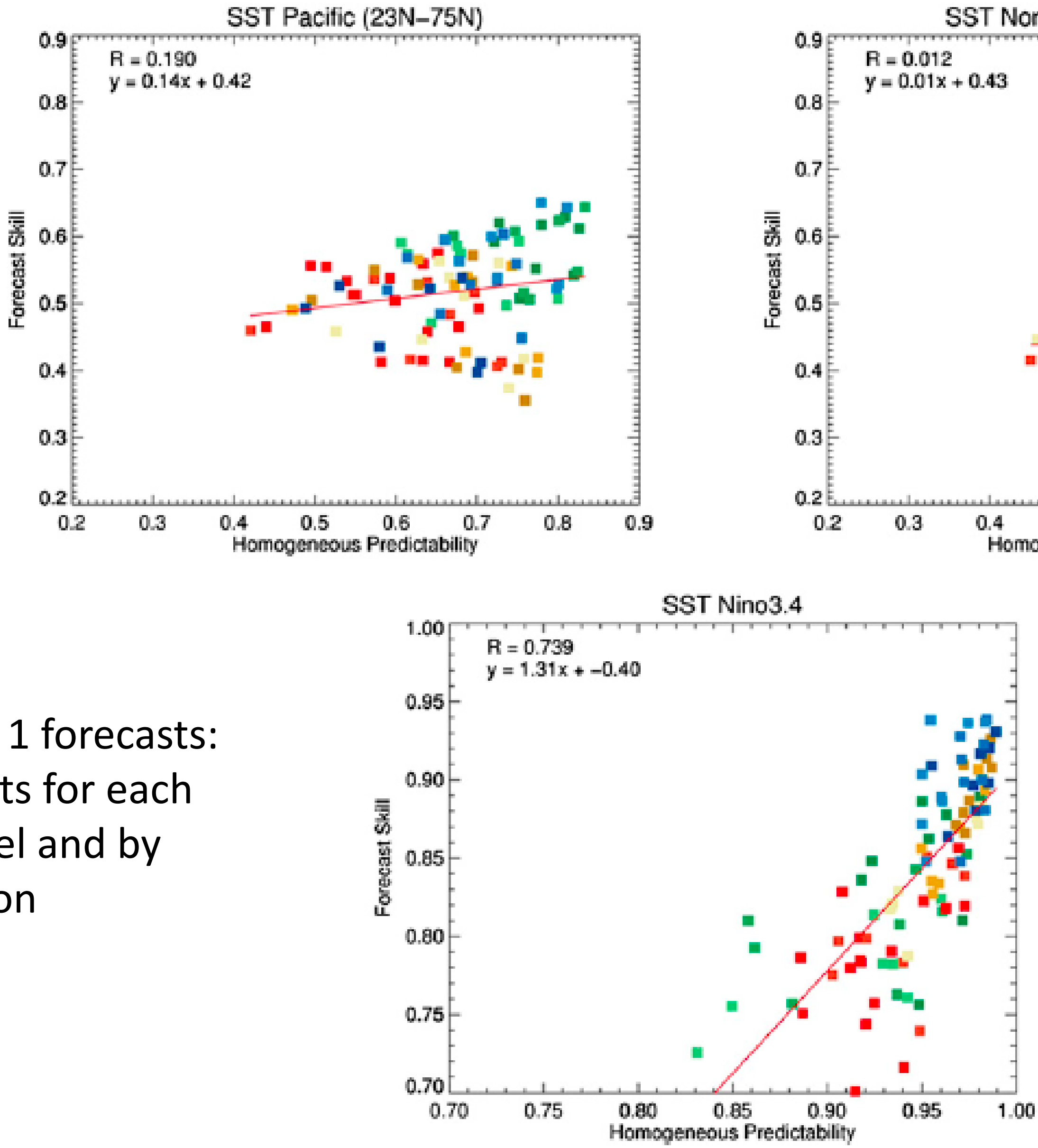
FIG. 4. As in Fig. 1, but for precipitation rate over Northern Hemisphere land. RMSE is in millimeters per day.

A	B
1.54	1.3
1.35	1.3
1.14	1.0
1.16	1.0
1.3	1.1
1.3	1.1
1.46	1.1

All models are over dispersive

Variance is about right for most

A. dispersion/rmse
B. Single member SD/obs SD



Lead 1 forecasts: results for each model and by season

Blue – NDJ,DJF,JFM
GREEN - FMA,MAM,AMJ
RED - MJJ,JJA,JAS
ORANGE – ASO,SON,OND

FIG. 10. As in Fig. 3, but for sea surface temperature: (top left) Pacific Ocean (23°–75°N); (top right) Atlantic Ocean (45°–75°N); and (bottom) Niño-3.4 region.

Blue – NDJ,DJF,JFM
GREEN - FMA,MAM,AMJ
RED - MJJ,JJA,JAS
ORANGE – ASO,SON,OND

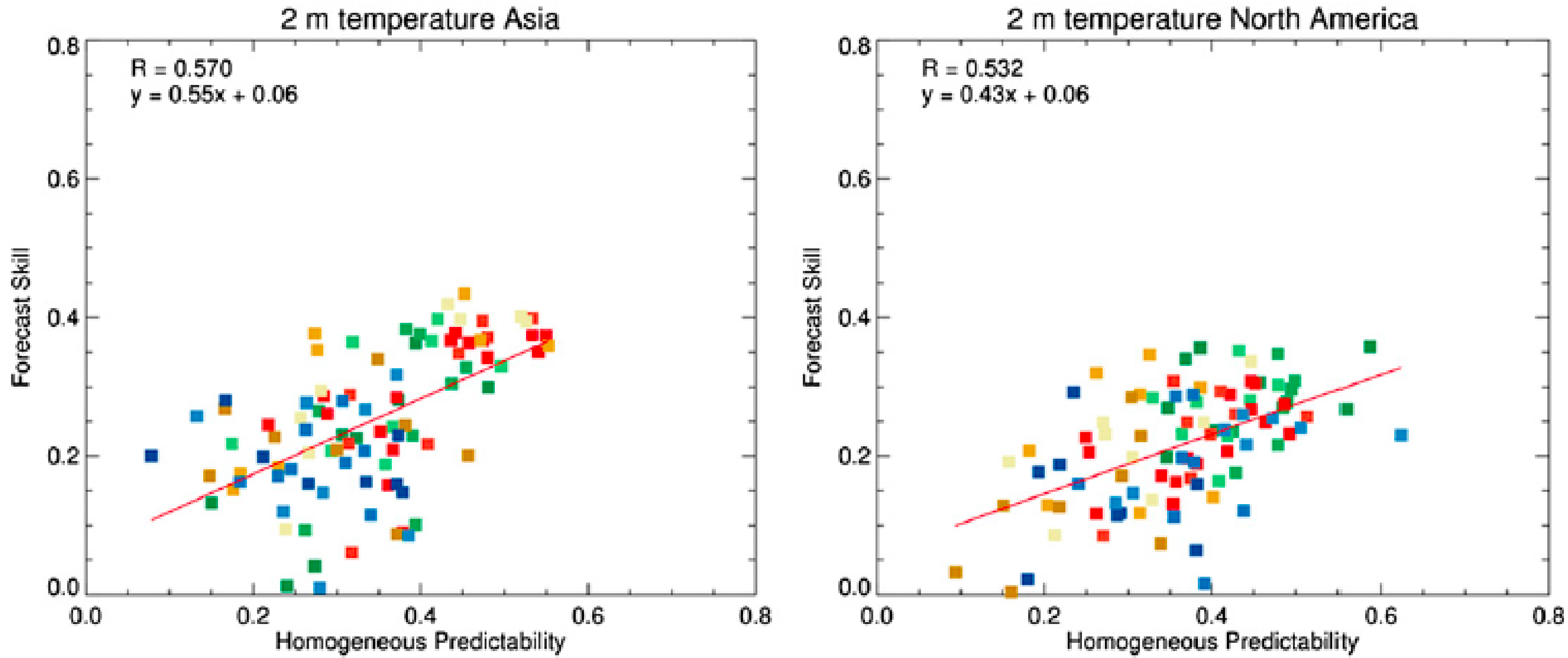


FIG. 3. The 2-m temperature homogeneous predictability AC (*x* axis) vs forecast skill AC (*y* axis). Colors indicate season: boreal winter seasons (November–January, DJF, and January–March) are blue, spring seasons (FMA, MAM, and April–June) are green, summer seasons [May–July (MJJ), JJA, and July–September (JAS)] are red, and autumn seasons (August–October, SON, and October–December) are orange. The shades of each color vary from lighter (first 3-month period; e.g., MJJ) to darker (third 3-month period; e.g., JAS). Each individual season has seven points, one for each model. The linear fit is depicted by the red line. (left) Southeastern Asia (5°–50°N, 70°–145°E) and (right) extratropical North America (north of 23°N; Greenland not included).

Lead 1 forecasts: results for each model and by season

Blue – NDJ,DJF,JFM
GREEN - FMA,MAM,AMJ
RED - MJJ,JJA,JAS
ORANGE – ASO,SON,OND

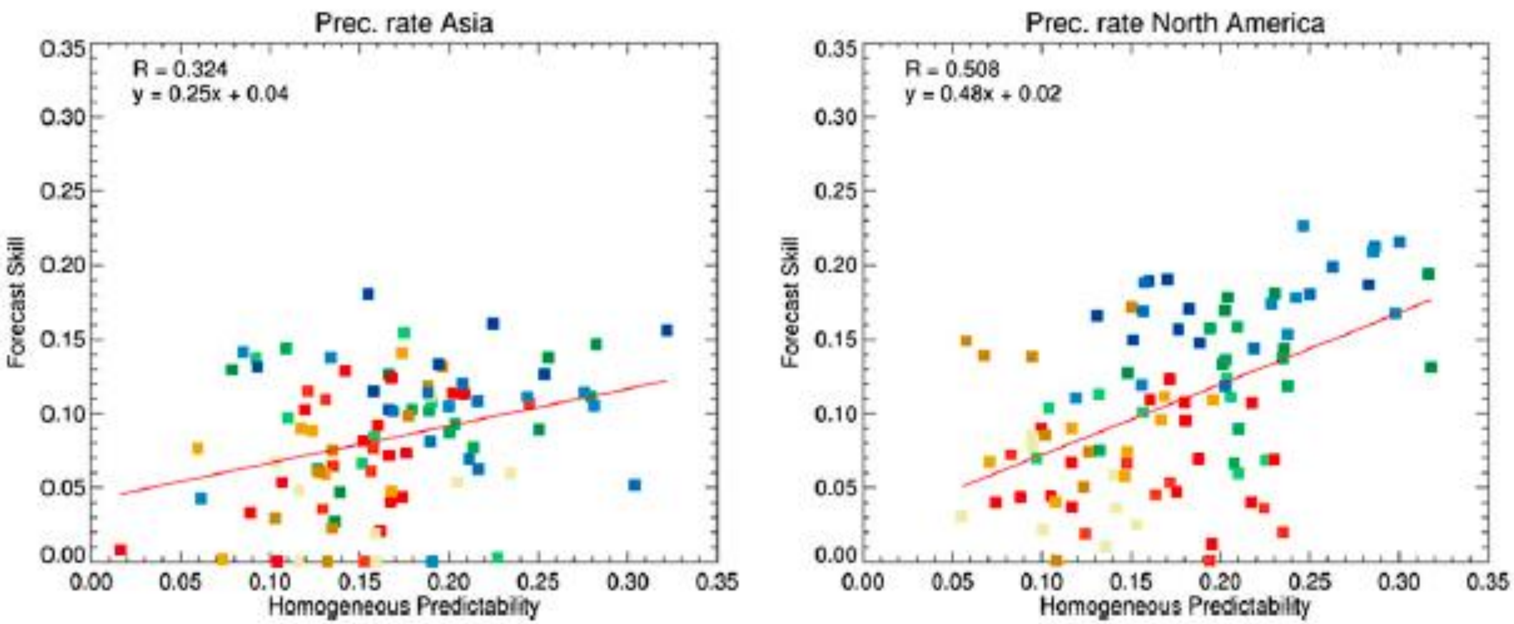


FIG. 6. As in Fig. 3, but for precipitation rate.

Lead 1 forecasts: results for each model and by season

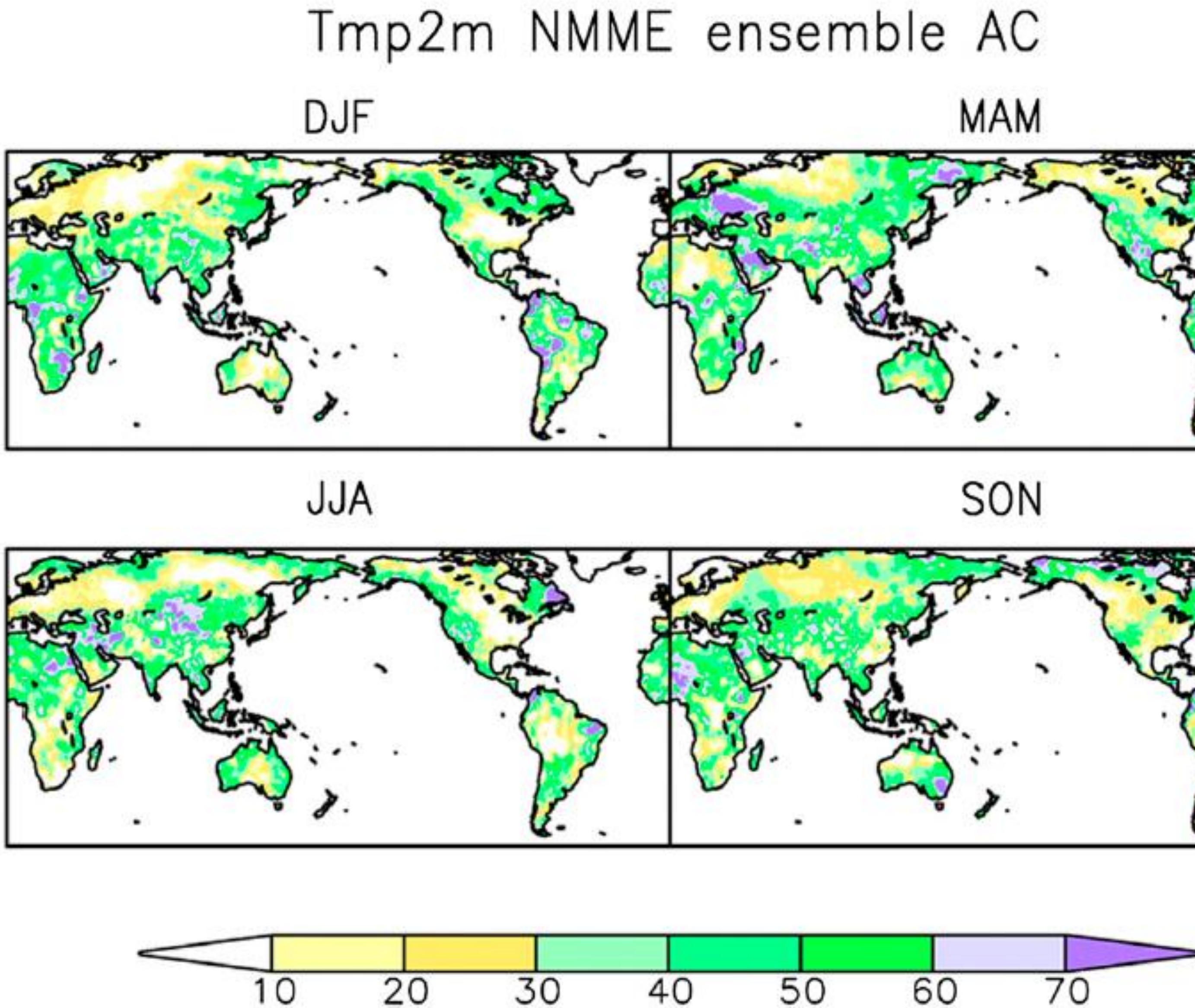


FIG. 2. Forecast skill measured by the anomaly correlation for NMME 7-model EM prediction of 2-m temperature. Four seasons are shown: lead-1 DJF, MAM, JJA, and SON. ACs are multiplied by 100. Note that the domain is near global: the NMME score of AC = 0.29 quoted in Fig. 1 corresponds to the extratropical Northern Hemisphere portion of the above maps.

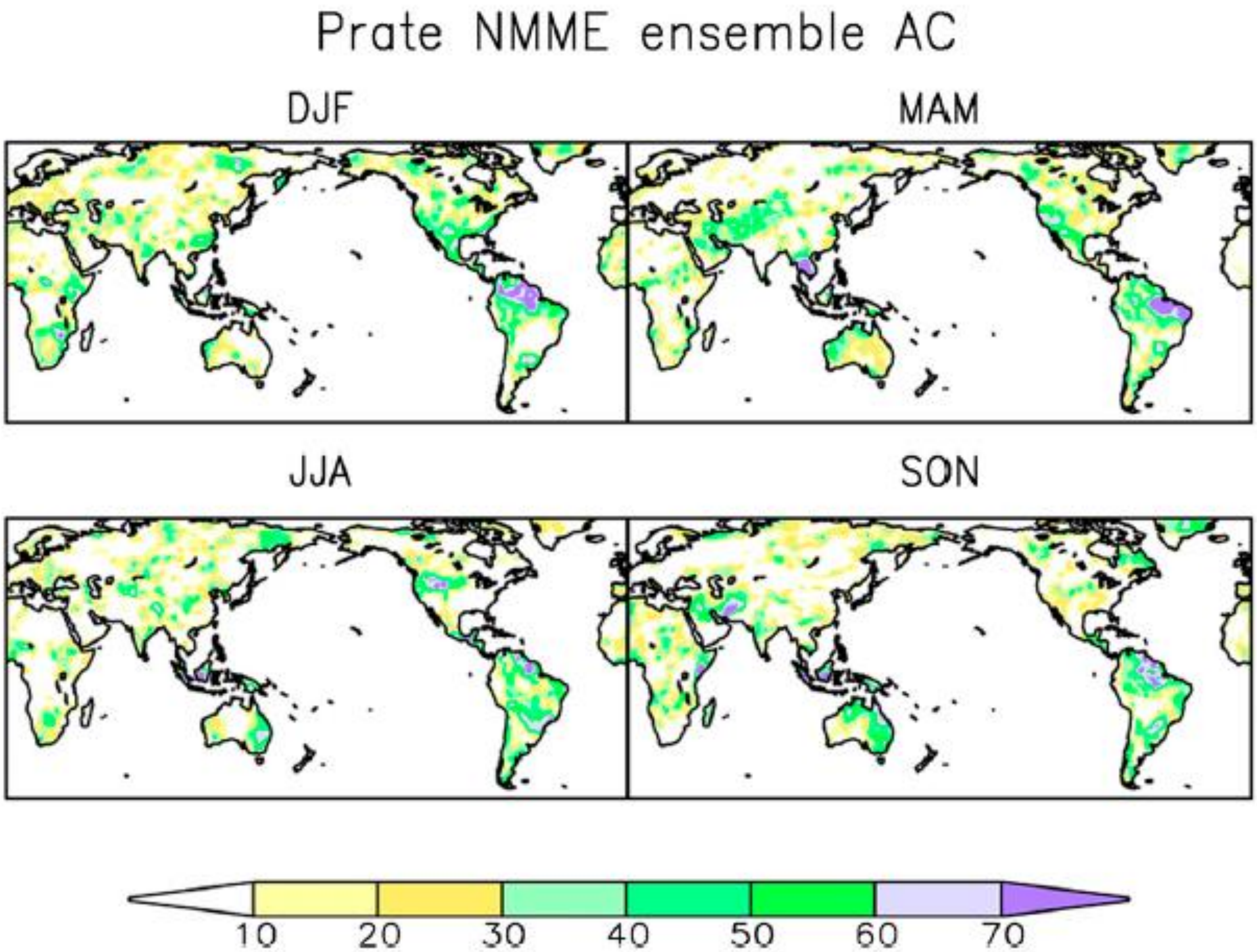


FIG. 5. As in Fig. 2, but for precipitation rate. The NMME skill of AC = 0.16 quoted in Fig. 4 refers to the extratropical NH portion of the above maps.

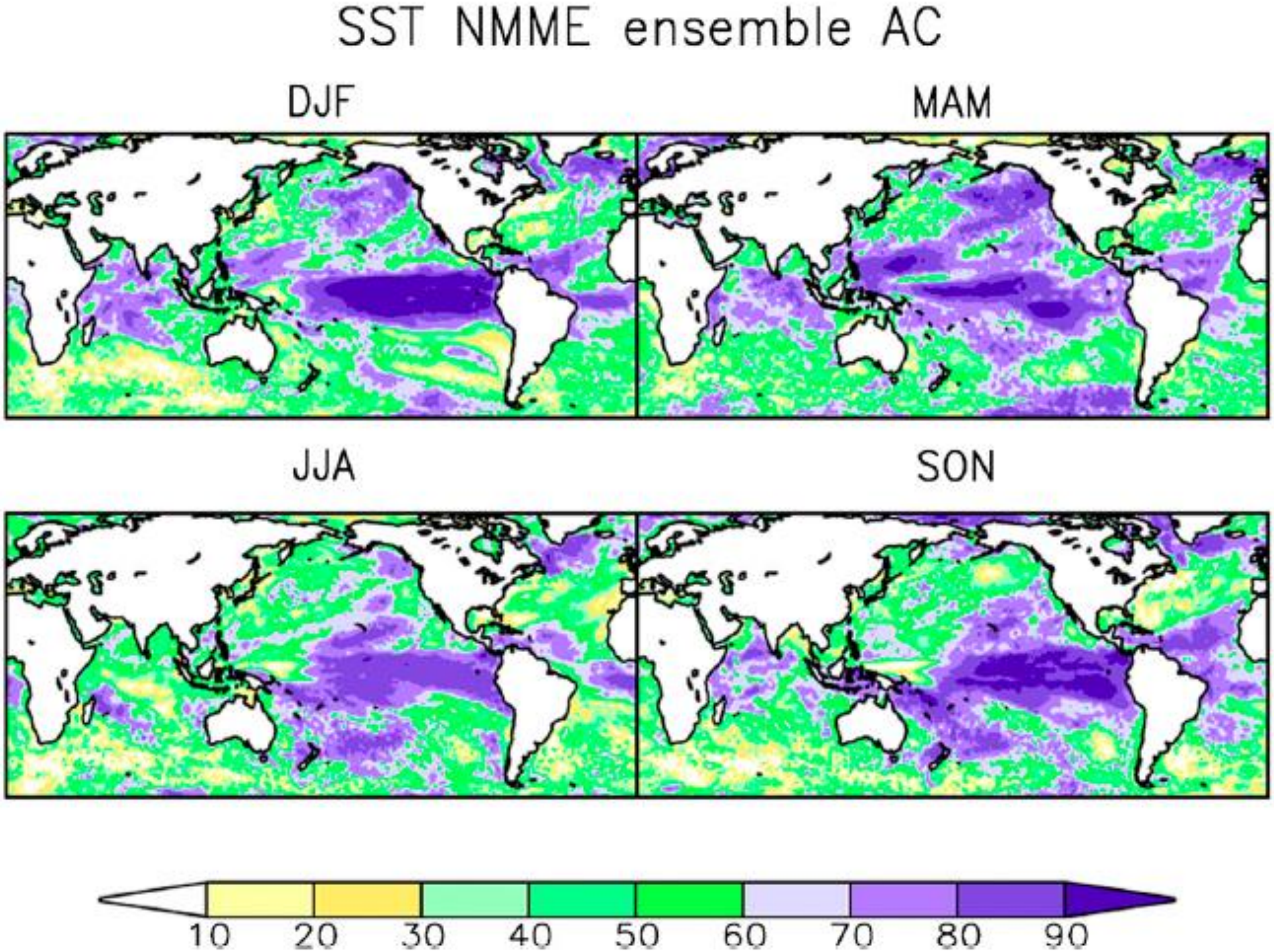
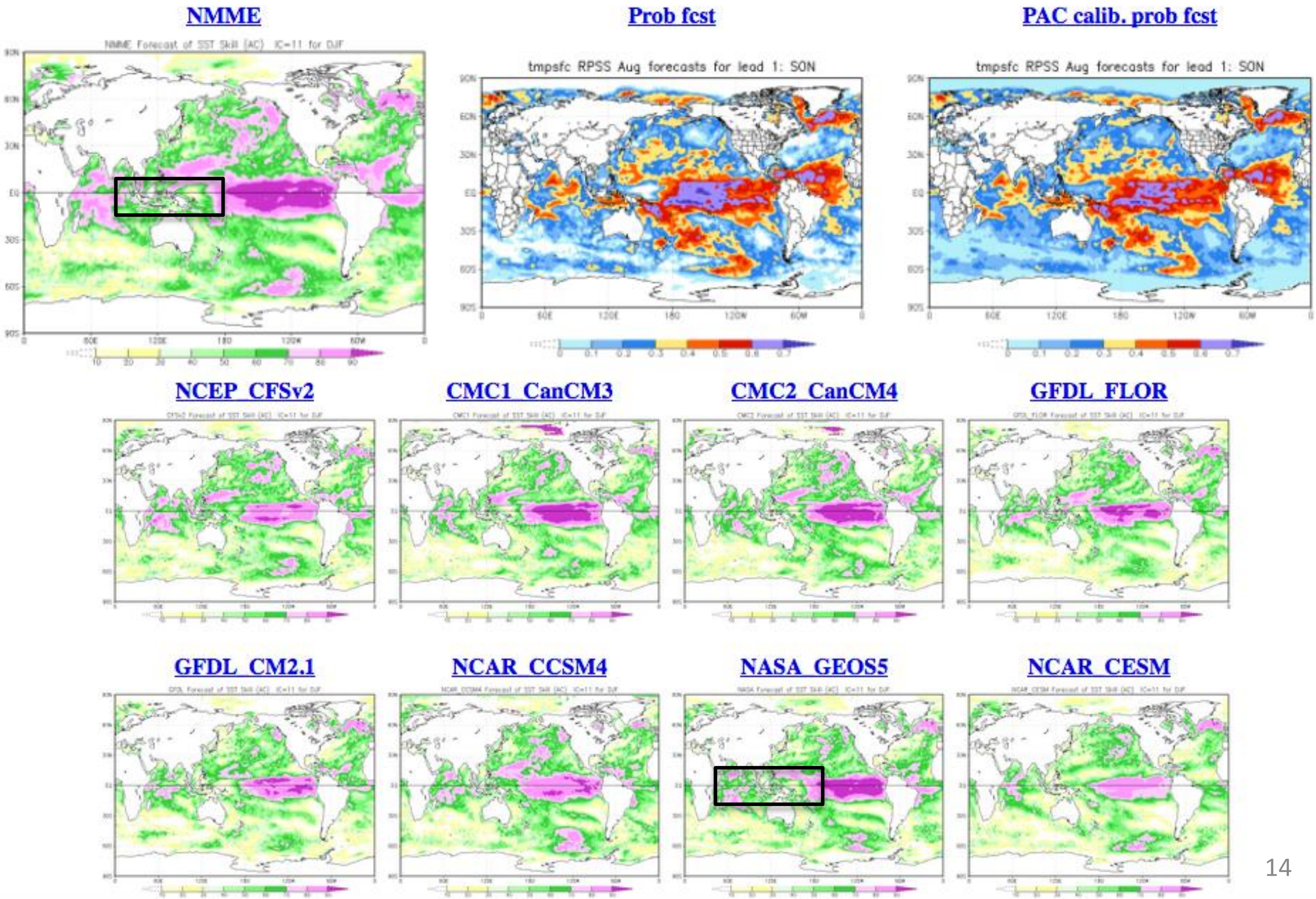


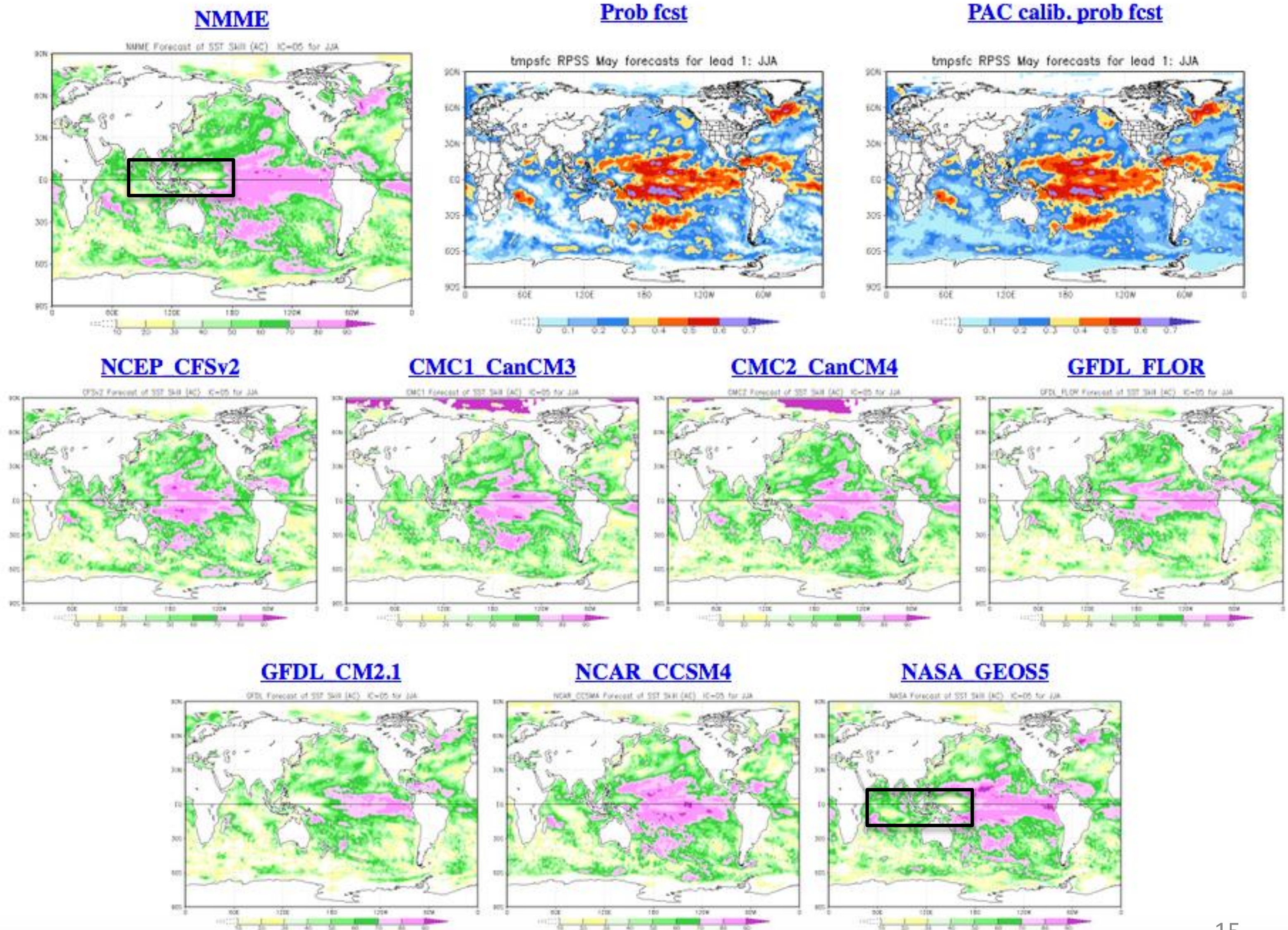
FIG. 8. As in Fig. 2, but for sea surface temperature. The NMME skill of AC = 0.50 quoted in Fig. 7 refers to the extratropical NH portion of the above maps.

Other issues (from NMME website)

Season 1 tmpsfc forecast **DJF**



Season 1 tmpsfc forecast JJA



Comments

- GEOS-5 model is in the mix of the other NMME models in terms of various basic skill measures
  - It is one of the best for NINO3.4 (potential predictability suggests some room for improvement though not much)
  - model is under-dispersive for NINO3.4 SST (like all the models) despite having too much variance – impacts reliability and consistency of forecasts (how do we increase the spread?)
  - other ocean basin SST show considerably less skill but also much more room for improvement based on potential predictability
  - The warm pool and eastern Indian Ocean SST show almost no skill for all models – an area where the atmosphere is known to be sensitive to SST
  - skill of T2m over North America for NMME is surprisingly poor in winter and summer – likely linked to errors in ENSO teleconnections in winter, and poor or no land initialization for some of the NMME models in summer
  - Skill of precipitation forecasts is generally low in NMME models (JJA over North America skill is especially low and surprising – again suggests need for improved land initialization

- The generally low skill at seasonal time scales over North America and elsewhere (T2m and Precipitation) suggests the need to focus on forecasts of opportunity- conditional skill can be much higher
  - Requires a focus on assessing impacts and predictability of various teleconnections including ENSO
  - Requires an in-depth analysis of the probabilistic quality of the forecasts (reliability, resolution, spread, BSS, etc)

- Suggest developing a comprehensive web-based baseline validation/evaluation capability (with flexibility to easily assess different regions, lags, seasons, quantities, models)