

# Predictive Skill of AGCM Seasonal Climate Forecasts Subject to Different SST Prediction Methodologies

SHUHUA LI, LISA GODDARD, AND DAVID G. DEWITT

*International Research Institute for Climate and Society, The Earth Institute at Columbia University, Palisades, New York*

(Manuscript received 31 August 2006, in final form 29 August 2007)

## ABSTRACT

This study examines skill of retrospective forecasts using the ECHAM4.5 atmospheric general circulation model (AGCM) forced with predicted sea surface temperatures (SSTs) from methods of varying complexity. The SST fields are predicted in three ways: persisted observed SST anomalies, empirically predicted SSTs, and predicted SSTs from a dynamically coupled ocean–atmosphere model. Investigation of relative skill of the three sets of retrospective forecasts focuses on the ensemble mean, which constitutes the portion of the model response attributable to the prescribed boundary conditions. The anomaly correlation skill analyses for precipitation and 2-m air temperature indicate that dynamically predicted SSTs generally improve upon persisted and empirically predicted SSTs when they are used as boundary forcing in the AGCM predictions. This is particularly the case for precipitation forecasts. The skill differences in these experiments are ascribed to the skill of SST predictions in the tropical ocean basins. The multiscenario forecast by averaging the three retrospective experiments performs, overall, as well as or better than the best of the three individual experiments in specific seasons and regions. The advantage of multiscenario forecast manifests both in the deterministic and probabilistic skill. In particular, the multiscenario precipitation forecast for the December–February season demonstrates better skill than the best of the three scenarios over several regions, such as the western United States and southeastern South America. These results suggest the potential value in producing superensembles spanning different SST prediction scenarios.

## 1. Introduction

In recent years, two-tier forecasting systems based on atmospheric general circulation models (AGCMs) have been used routinely at several institutions for making seasonal climate forecasts up to several seasons in advance (e.g., Goddard et al. 2003). Within the two-tier approach, tier 1 is designed to predict sea surface temperatures (SSTs), which are then used as lower boundary forcing in tier 2, the AGCM predictions. In real-time operational climate forecasts, global SST fields can be configured from a number of SST prediction methods, such as dynamical models, statistical models, or persistence of recent observations. Much of the seasonal climate predictability stems from the modulating role of SST on atmospheric circulation. Therefore, more skillful SST predictions portend more skillful

climate predictions (e.g., Bengtsson et al. 1993; Palmer and Anderson 1994; Goddard et al. 2001; Robertson et al. 2004). Errors in SST predictions are one source of errors in AGCM forecasts. Since perfect SST predictions are not possible, assessing the impacts of limited SST predictability on seasonal climate predictability is of critical importance to providing meaningful seasonal forecasts (Barnston et al. 2003; Goddard et al. 2003).

The International Research Institute for Climate Prediction (IRI) began making seasonal forecasts of precipitation and near-surface temperature in late 1997 (Mason et al. 1999), based on an increasingly objective multimodel ensembling forecast technique (Barnston et al. 2003). At present, these forecasts are calibrated using skill information from AGCM simulation runs, in which the AGCM is forced with actual SST fields, which are regarded as “perfect” representations of oceanic boundary conditions. In real-time forecasting, however, it is necessary to predict the SSTs used as lower boundary forcing. Therefore, the skill of forecasts based on model predictions is generally lower than that

---

*Corresponding author address:* Dr. Shuhua Li, International Research Institute for Climate and Society, The Earth Institute at Columbia University, 61 Route 9W, Palisades, NY 10964-8000.  
E-mail: shuhua@iri.columbia.edu

TABLE 1. Summary of one set of simulation runs and three sets of retrospective forecast experiments with the ECHAM4.5 AGCM.

Experiments	Period	No. of years	Ensemble size	Remarks on SST
Simulation	1950–2002	53	24	Observed SST
PST_Retro	1968–2002	35	12	Undamped persistence, globally
CA_Retro	1957–2002	46	24	CA <30°N/S, damped persist >30°N/S
DMT_Retro	1958–2001	44	24	ECMWF <30°N/S, damped persist >30°N/S

of the simulations. In other words, partial loss of predictive skill in operational forecasts compared to the AGCM simulations results because predicted SSTs are always imperfect regardless of the SST prediction method.

Within the global climate system, regional SSTs may be important to regional climate, but they are not the whole story because global atmospheric circulation is influenced by larger-scale patterns of SST. Hence, emphasis must be placed on the broad patterns of SST variability. Given the numerous approaches and models for predicting global SSTs, how does one judge a “more skillful” SST prediction? How superior must it be to become the “preferred” prediction?

Goddard and Mason (2002) suggest that persisted SST anomalies (SSTAs) may provide an appropriate forecast of oceanic conditions for the first 3-month season in many regions. However, predictions of evolving SSTA can improve upon persistence for seasonal forecasting of rainfall anomalies in some regions, such as over East Africa during the September–November rainy season. Over the past several years, coupled atmosphere–ocean general circulation model (CGCM) systems have been increasingly used for seasonal forecasting (e.g., Palmer et al. 2004; DeWitt 2005; Guérémy et al. 2005). Despite their computational expense, CGCM forecast systems appear to have seasonal prediction skill that is comparable to other forecast techniques (DeWitt 2005; Guérémy et al. 2005; L. Goddard et al. 2007, unpublished manuscript). CGCMs can also benefit two-tier prediction by applying their predictions of SST to force one or more AGCMs. It is therefore worthwhile to assess the benefit to climate forecasts of more sophisticated SST predictions than persistence, such as from empirical or dynamical models. To get an estimate of real-time seasonal predictability, we have produced three sets of retrospective forecasts, in which a single AGCM is forced by empirically and dynamically predicted SSTs, as well as persisted observed SST anomalies, over more than 30 years. In addition to identifying which SST prediction yields better climate predictability in specific seasons, we also try to examine the potential benefit attainable by combining the three SST predictions (persisted, empirical, and dynamical SSTs), as opposed to choosing the “best” one.

This article aims to examine the predictive skill that could be realistically expected of precipitation and 2-m temperature seasonal forecasts from a single AGCM using both deterministic and probabilistic measures. The retrospective experiments, data, and methodology are described in the next section. Section 3 presents some results of the skill in the ensemble mean, as measured by correlation, based on these experiments. Probabilistic verification is illustrated by the reliability and resolution components of the examined retrospective experiments and the multiscenario forecasts for precipitation. The skill of the SST predictions is also investigated in section 3 in terms of their pattern statistics over the tropical ocean basins, and linked with skill of seasonal precipitation forecasts over the tropical land area. At the end of section 3 we demonstrate the benefit of multiscenario ensembling forecasts on both correlation skill and probabilistic measures through simple combination of the three retrospective forecast experiments. A summary and discussion are presented in section 4.

## 2. Data, experiments, and methodology

### a. Model and experimental design

The experiments in this study use the AGCM developed at the Max Planck Institute for Meteorology, in Hamburg, Germany. The model version is ECHAM4.5 with a horizontal resolution of T42 spectral truncation (approximately 2.8° in latitude and longitude), and 18 levels in the vertical. Further details of this AGCM can be found in Roeckner et al. (1996). As a primary AGCM within the IRI’s multimodel ensembling forecast system, the ECHAM4.5 AGCM is executed routinely at the IRI for our real-time seasonal forecasting, and it has also been extensively used for climate research (e.g., Barnston et al. 2003; Goddard et al. 2003; Robertson et al. 2004; DeWitt 2005).

Four sets of ensemble experiments are produced with the ECHAM4.5 AGCM (see Table 1 for details). These include a 24-member ensemble simulation (i.e., forced with observed SSTs), a 12-member retrospective experiment using persisted SSTAs, a 24-member retrospective experiment using empirically predicted constructed analog (CA) SSTs (Van den Dool 1994),

and a 24-member retrospective experiment using SSTAs predicted by the European Centre for Medium-Range Weather Forecasts (ECMWF) coupled ocean–atmosphere dynamical model from the Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER) project (Palmer et al. 2004). The common period for the four sets of ensembles is 1968–2001. Over this 34-yr period, these experiments differ only in their prescribed SST forcing, as detailed below. The experiment using empirically predicted SSTs consists of 0–6-month<sup>1</sup> lead forecasts starting from each month, while the experiment using persisted SSTs consists of 0–4-month-lead forecasts. The experiment using dynamically predicted SSTs has only four start times per year (February, May, August, and November) and is available for 0–4-month lead. The initial conditions for the retrospective forecast experiments are taken from the corresponding times in the long simulation runs of the same AGCM. The initial conditions for each of the four sets of experiments at any given start time are identical.

The simulation runs were generated by forcing the ECHAM4.5 with observed monthly mean SSTs from Extended Reconstructed SST, version 1 (ERSSTv1) during 1950–81 (Smith and Reynolds 2003) and Optimum Interpolation, version 1 (OIv1) from 1982 to the present (Reynolds and Smith 1994). Initial atmospheric states for the different long simulation ensemble members are generated by adding random machine precision level perturbations to the wind field during the model spinup period of the 1940s. In other words, no observed initial conditions are used for any of the experiments. The integrations proceeded continuously from 1940s to present-day conditions, and the years before 1950 were discarded.

Retrospective ensemble forecasts were made using one of three types of predicted SSTs as boundary forcing. In the retrospective experiments with persisted SSTAs (hereafter referred to as PST\_Retro), the observed SSTAs from the month preceding the forecast period are persisted on top of the evolving climatological annual cycle of SST to yield the full SST field. The other two sets of retrospective experiments, hereafter referred to as CA\_Retro and DMT\_Retro, respectively, use the empirical CA SSTs and the ECMWF

dynamically predicted SSTs as boundary conditions over the tropical oceans (within 30° of the equator). Over the extratropical oceans, the damped persistence of the preceding month's observed SSTAs are added to the climatological annual cycle of SST from each starting month through the following forecast months; the damping coefficient has an  $e$ -folding time of three months (a scheme employed in the IRI real-time climate forecasts). That is,

$$\text{SSTA}(T) = \text{SSTA}_0 e^{-(T/3)}, \quad (1)$$

where  $\text{SSTA}_0$  is the observed SST anomalies from the month preceding the forecast start time and  $T$  is time in months between each starting month and the forecast months (0–6 for CA\_Retro and 0–4 for DMT\_Retro). A brief summary of the three sets of retrospective forecast experiments, along with the long-run simulations, is provided in Table 1.

### b. Observational data

The observational verification data for both precipitation and near-surface (2 m) air temperature come from the recently updated Climate Research Unit (CRU TS 2.1) dataset of the University of East Anglia (Mitchell and Jones 2005), which is available every month over the 1900–2002 period. The high-resolution gridded data, originally on a 0.5° grid, were linearly upscaled to the T42 Gaussian grid (approximately 2.8° horizontal resolution) for comparison with the model fields.

Since this study focuses on predictive skill of seasonal climate forecasts for the purposes of real-time prediction, our full analysis covers all 12 three-month seasons [January–March (JFM) through December–February (DJF)]; hereafter, all 3-month seasons will be referred to by the first letter of each month]. However, the results presented here focus on just a couple illustrative seasons for brevity.

The SST dataset used for verification of SST predictions is obtained from the combination of monthly mean SSTs from ERSSTv2 (Smith and Reynolds 2004) for 1968–81 and OIv2 (Reynolds et al. 2002) for 1982–2001. Both ERSSTv2 and OIv2 SSTs have modest regional difference from their earlier versions, ERSSTv1 and OIv1, which were used to force the AGCM simulations in this study.

## 3. Results

A basic set of diagnostics indicates performance of the ensemble means from the retrospective experiments, and provides an initial assessment of how the different retrospective experiments compare to each

<sup>1</sup> The lead time is defined here as the difference between the month at the beginning of which the forecast is made and the first month of the period being predicted. For example, a forecast made in early August for the month of September, or the September–November season, has a lead time of one month; the lead time for October, or the October–December season, is thus two months, and so on.

other and to the simulations. The initial hypothesis is that the simulations constitute an upper limit for seasonal climate prediction skill, and that the PST\_Retro constitutes a lower limit. For the purpose of judging the impact of SST accuracy on AGCM climate predictions, we consider the temporal anomaly correlation between observed and model-produced climate variability. The anomaly correlation quantifies the agreement between the observed variability and the model ensemble mean, which is the primary component of the model variability related to the SST forcing, often referred to as externally forced “signal.” For calculation of the ensemble means, 12 members are used for each experiment, because the PST\_Retro experiment has only 12 members. For the simulation runs, CA\_Retro, and DMT\_Retro, all of which contain 24 members, the first 12-ensemble members of each set are used, so that all sets derive from common initial conditions and differ only in their prescribed boundary conditions.

Temporal correlation coefficients are calculated for seasonal-mean 2-m temperature and total seasonal precipitation over the verification period 1968–2001 for the three retrospective forecast datasets and for the simulation runs. Note that the seasonal forecasts from PST\_Retro and DMT\_Retro are only available for up to 2-month lead, because their monthly retrospective experiments were executed for 0–4-months lead (e.g., forecasts from February extended through June, or seasonal forecasts from FMA through AMJ). We thus focus on 1–2-month-lead seasonal predictions, with an emphasis on 1-month lead. In addition to the deterministic measures for the ensemble mean, a probabilistic verification based on the ensemble distribution, as demonstrated by the half-Brier score (Wilks 1995) and its decomposition, is also investigated for the three retrospective experiments and the multiscenario forecasts.

#### a. Summarized correlation skill of retrospective forecasts

Summary graphs of correlation coefficients represent the overall skill of the seasonal forecasts. Here, the skill is defined as the percentage of land grid points within a specific region where the correlation coefficient  $r$  exceeds a given value  $R$ . This can be expressed by

$$S = \frac{N_{\text{land}}|r > R}{N_{\text{land}}} \times 100, \quad (2)$$

where  $R$  is the correlation threshold, which can be any value between  $-1$  and  $1$ , and  $N_{\text{land}}$  represents the number of grid points over land. We have chosen the lower bound to be  $0.3$ , which represents approximately the 95% confidence level for one-tailed (positive values only here) Pearson’s anomaly correlations.

Figures 1 and 2 present the summarized skill of 2-m temperature and precipitation simulations and retrospective experiments over low-latitude land ( $30^{\circ}\text{S}$ – $30^{\circ}\text{N}$ ) for all the 12 seasons. An estimate of the uncertainty due to limited ensemble size is constructed by Monte Carlo resampling the 24-member ensembles of CA\_Retro at 1-month lead for 100 12-member subsets. Similar uncertainty bounds apply to any of the lines in the graphs. Any single 12-member subset may give locally higher or lower skill values because of sampling fluctuations. The uncertainty range provides an indication of where the PST\_Retro or DMT\_Retro curves lie within or notably without these limits, suggesting meaningful differences in skill between the three retrospective experiments.

As expected, the summarized skill from the simulations provides an upper limit of the model’s potential predictability, exceeding the skill from all the three retrospective forecasts for all seasons. This is because the observed monthly SSTs represent the “real” oceanic boundary conditions, and a more accurate response is expected when the model is forced with “perfect” SSTs. It is then assumed that the majority of the skill degradation in the retrospective climate runs results from erroneous boundary conditions introduced by the predicted SSTs. If the skill from the retrospective experiments were significantly larger than those of the simulations, it would suggest that the models were getting the “right” answer for the wrong reasons.

To illustrate the significance of the summarized skill from these experiments, we also estimated the field significance (Livezey and Chen 1983) as the percentage of grid points (over the tropical land area), at 95% confidence level, exceeding given correlation values by chance. This threshold of field significance is shown by dark gray shading at the bottom of each panel in Figs. 1 and 2. This was determined using the Monte Carlo technique of shuffling the time sequence of ensemble-mean model fields from the simulation runs, calculating the anomaly correlation maps relative to the properly time-sequenced observations, and tabulating the summary curves 500 times. The curves from all the four experiments clearly exceed field significance for both precipitation and 2-m temperature, suggesting that the summarized skill from these experiments is statistically significant.

For 1-month-lead forecasts, the skills of CA\_Retro and PST\_Retro compare similarly for most of the 12 seasons, except for Northern Hemisphere winter when CA\_Retro performs better. In comparison to CA\_Retro and PST\_Retro, the DMT\_Retro experiments provide comparable skill for 2-m temperature in SON and DJF, but perform worse in MAM and JJA.

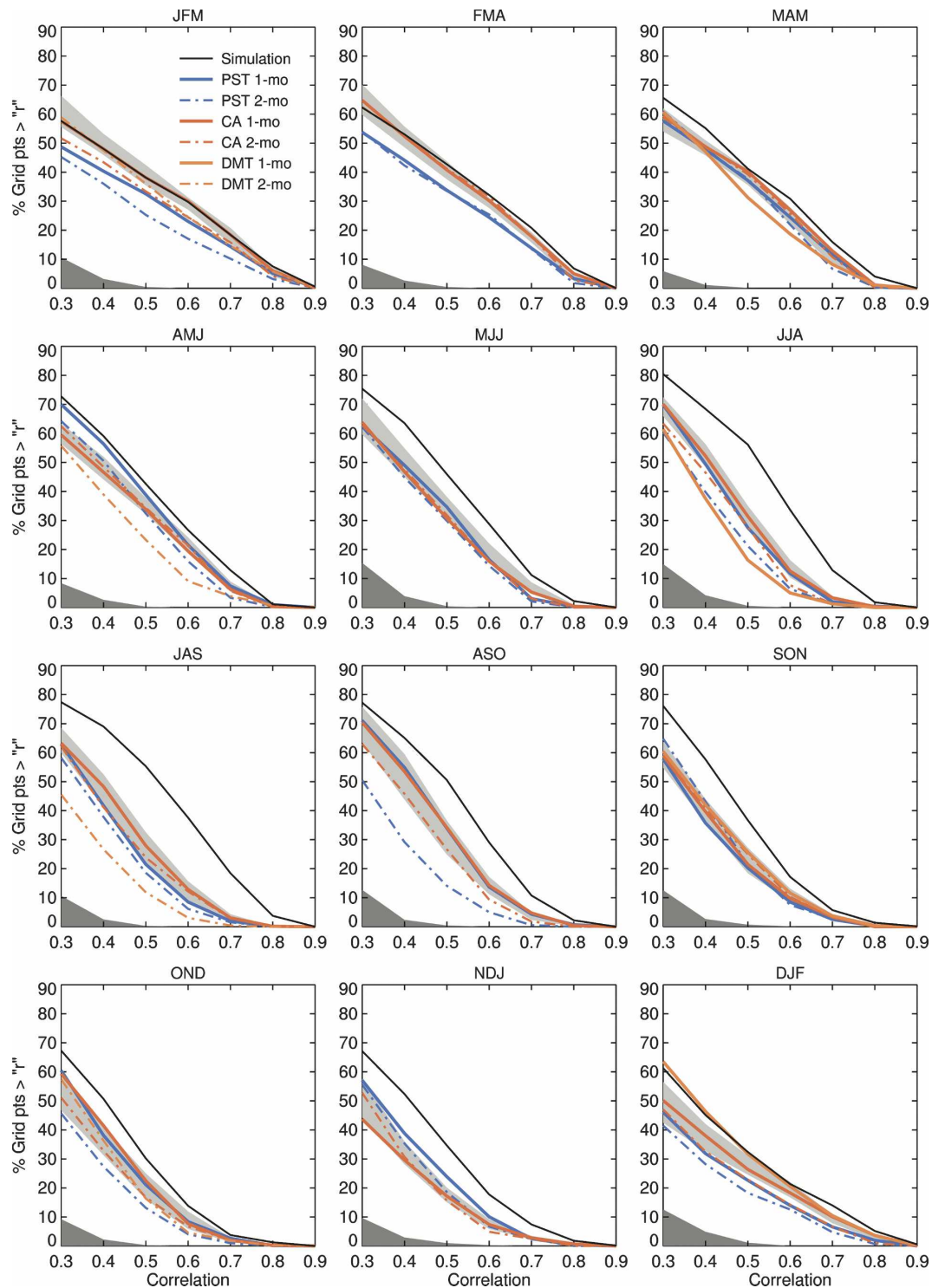


FIG. 1. Summary graphs showing percentage of grid points over tropical land ( $30^{\circ}\text{S}$ – $30^{\circ}\text{N}$ ) exceeding a given anomaly correlation coefficient between model forecast/simulations and observations of seasonal-mean 2-m temperature for all seasons during the 1968–2001 period. Black, blue, red, and orange curves represent the simulations, and the forecast experiments using persistence, CA SSTs, and ECMWF SSTs, respectively. Light gray shaded area denotes the range of uncertainty produced by resampling 12-member ensemble out of 24 CA\_Retro runs. Dark gray shading at the bottom of each panel represents the percentage of grid points, at 95% field significance, exceeding the given correlation values by chance over the tropical land area based on the simulation runs.

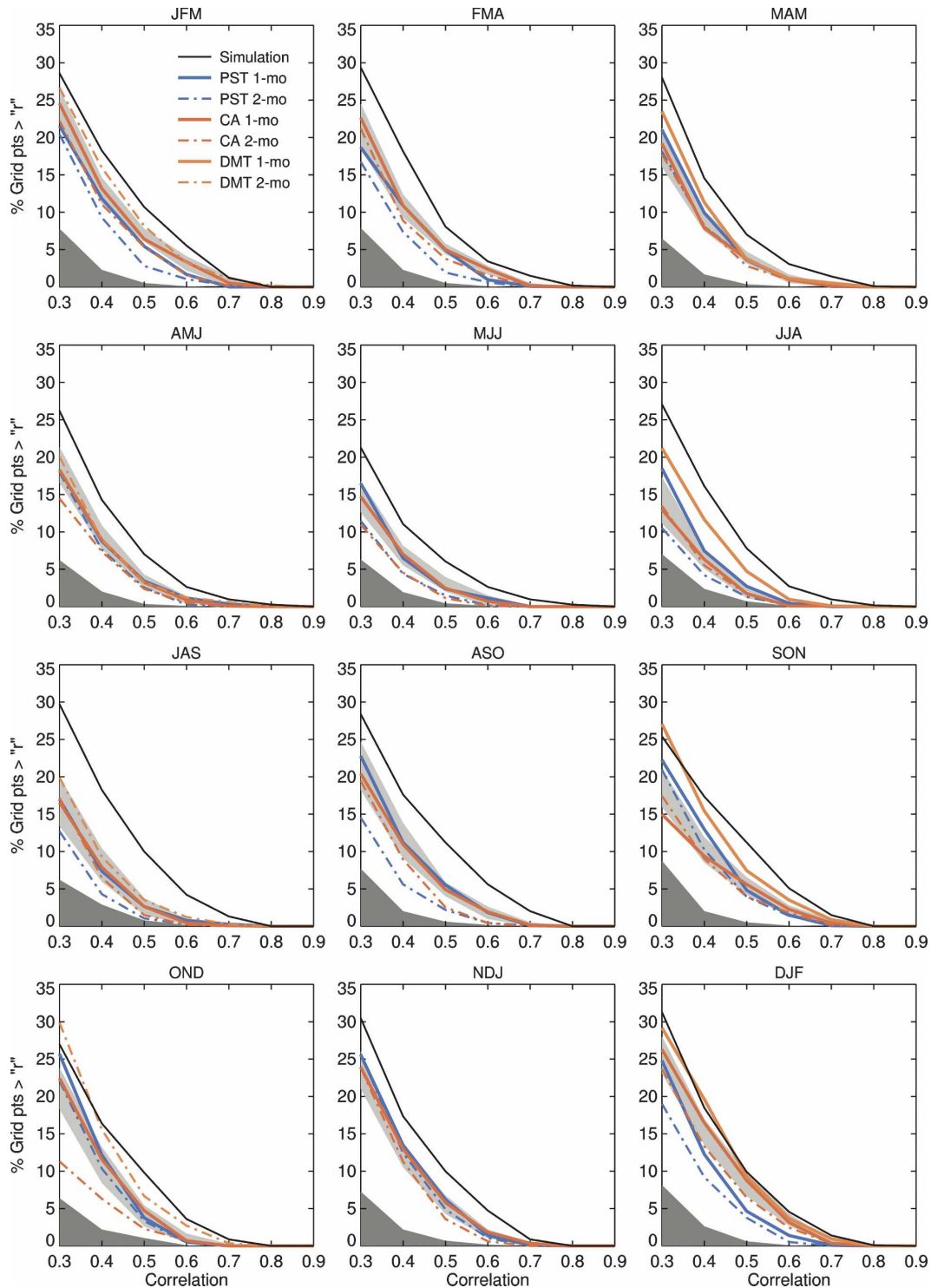


FIG. 2. Similar to Fig. 1, but for seasonal precipitation totals.

For precipitation, the skill of DMT\_Retro appears to be the best of the three retrospective forecasts, especially for JJA and SON. In other words, there are no clear rules for which SST prediction methodology

yields the best model response. One would presume, for example, that forecasts produced with empirically predicted (evolving) SSTs should outperform those using persisted SST anomalies, unless the evolving SST

predictions were consistently worse than persistence, which would not be expected from a statistical prediction tool. However, the forecast skill of temperature from PST\_Retro is slightly better than that from CA\_Retro (Fig. 1) in some of the 12 seasons, such as in AMJ and NDJ when the PST\_Retro curve is mostly outside the CA\_Retro uncertainty envelope.

To investigate the relative performance of longer lead forecasts, we also take into account the 2-month lead for the three retrospective forecast experiments. As expected, skill declines at longer lead times, illustrated by the dot-dashed lines generally lying below the solid lines in Figs. 1 and 2. For PST\_Retro the drop in skill from 1- to 2-month lead is largest in the middle to the end of the year (Figs. 1 and 2). Similarly, CA\_Retro skill drops the most at increasing lead for start times in the last half of the year. In both cases, this may be related to the evolution of ENSO events, which are responsible for much of the realizable skill in climate predictions (Goddard and Dilley 2005). For DMT\_Retro the drop in skill is difficult to judge for a given forecast season because of the limited start times. However, it appears that at some times of year (MAM/AMJ and JJA/JAS) the skill for DMT\_Retro drops more rapidly than in the other experiments. It is assumed that skillful SST predictions, such as those based on empirical models or state-of-the-art CGCMs, should provide more skillful climate predictions than persisted SSTs, but, overall, similar declines in skill with increasing lead time are seen in the different methodologies.

As an *index of performance*, the integrated areas under the curves in Figs. 1 and 2 are calculated for 1-month-lead forecasts (Fig. 3, Table 2) and 2-month-lead forecasts (Table 2). As noted previously, the simulation runs provide the upper limit of potential skill for all the 12 seasons. The ordering of performance among the experiments exhibited in DJF precipitation (Fig. 3) was our incoming expectation: that more sophisticated means of SST prediction would provide more accurate climate predictions. As illustrated in the other cases of Fig. 3, that expectation is not the common reality. In some seasons, PST\_Retro performs better than CA\_Retro, and in some seasons DMT\_Retro is the poorest performer for a given variable. That the CA SSTs do not necessarily lead to more accurate predictions than persisted SST anomalies is evident (Fig. 3, Table 2); however, the ECMWF coupled model SSTs do generally, although not always, lead to more accurate predictions. For precipitation, DMT\_Retro often shows a higher index of performance at 2-month lead than the other experiments have at 1-month

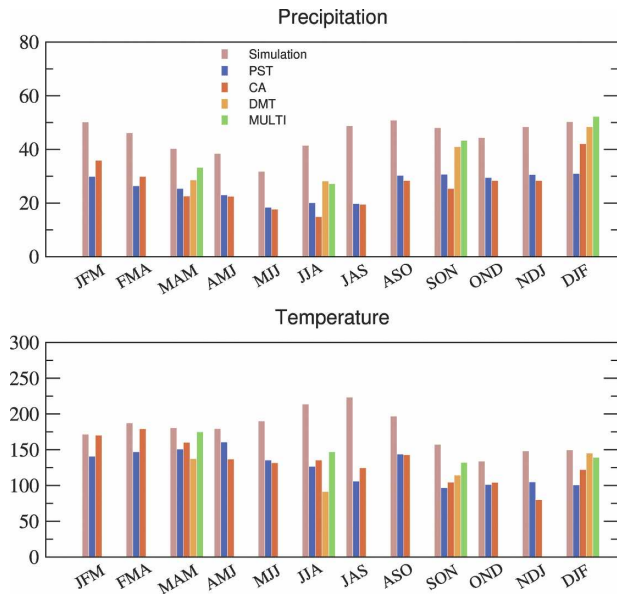


FIG. 3. Integrated percentage of grid points over tropical land area ( $30^{\circ}\text{S}$ – $30^{\circ}\text{N}$ ) with statistically significant anomaly correlations over the period 1968–2001 (i.e., integrated area under curves in Figs. 1 and 2 for 1-month lead). Simulations and forecast experiments consist of 12-member ensembles. MULTI represents the multiscenario forecast with all 36 members of the three forecast experiments. The DMT and MULTI bars are plotted for four seasons only, because the ECMWF SSTs used for DMT\_Retro are available for four forecast months (February, May, August, and November) per year.

lead (Table 2). The same cannot generally be said, though, for 2-m temperature, and in some seasons, the DMT\_Retro is not even better at 1-month lead than the others are at 2-month lead for temperature.

What regions contribute most to the noticeable differences between the three retrospective forecasts? To answer this question, we turn to the geographical distribution of the anomaly correlation skill for the three sets of retrospective experiments. Based on the above discussion, we focus on the seasons JJA and DJF because for these seasons the three retrospective predictions indicate noticeable differences in 1-month-lead forecast skill, particularly for the forecast of precipitation. Most, though not all, attention will be focused on predictive skill of precipitation because of its large social implications for many parts of the world.

#### b. Geographical features of prediction skill

The maps of anomaly correlation coefficients of 2-m temperature for JJA and DJF between  $60^{\circ}\text{S}$  and  $80^{\circ}\text{N}$  are presented in Fig. 4. Generally, the correlation skills of the three sets of retrospective forecasts show similar patterns to that of the simulation runs. High skill is

TABLE 2. (a) Integrated area under the curves in Fig. 1 for the simulations and 1–2-month-lead forecasts of 2-m temperature, as well as the multiscenario of the three forecast experiments. The values in parentheses are for 2-month lead. (b) Similar to (a), but for precipitation totals (i.e., integrated area under the curves in Fig. 2).

	Simulation	PST	CA	DMT	MULTI
(a)					
JFM	171.0	140.1 (114.5)	169.6 (147.1)	(158.0)	(161.9)
FMA	186.7	146.4 (143.6)	178.6 (177.4)		
MAM	180.0	150.2 (145.2)	159.6 (154.4)	136.8	174.4
AMJ	178.7	160.1 (134.5)	136.2 (141.4)	(104.0)	(168.4)
MJJ	189.4	134.8 (122.5)	131.1 (129.9)		
JJA	213.1	126.0 (98.9)	134.9 (114.9)	90.9	146.4
JAS	222.8	150.3 (93.2)	124.0 (111.2)	(64.8)	(122.8)
ASO	196.3	143.1 (74.2)	142.3 (115.2)		
SON	156.8	96.2 (107.3)	103.8 (108.7)	113.8	131.6
OND	133.4	100.6 (68.4)	103.6 (83.8)	(87.9)	(112.4)
NDJ	147.6	104.3 (90.8)	79.5 (80.2)		
DJF	149.0	100.0 (85.3)	121.5 (100.9)	144.6	138.7
(b)					
JFM	50.0	29.7 (23.6)	35.7 (29.3)	(42.0)	(41.4)
FMA	46.0	26.2 (18.3)	29.7 (24.9)		
MAM	40.1	25.2 (22.0)	22.4 (21.0)	28.4	33.1
AMJ	38.3	22.8 (19.4)	22.3 (17.5)	(24.1)	(27.3)
MJJ	31.6	18.2 (11.8)	17.5 (11.3)		
JJA	41.3	19.9 (11.0)	14.7 (14.0)	28.0	27.0
JAS	48.6	19.6 (12.0)	19.3 (16.8)	(24.3)	(24.3)
ASO	50.7	30.1 (15.6)	28.2 (21.6)		
SON	47.9	30.5 (26.2)	25.2 (23.7)	40.8	43.2
OND	44.2	29.3 (25.2)	28.2 (15.0)	(40.3)	(39.1)
NDJ	48.2	34.0 (31.2)	32.9 (27.4)		
DJF	50.1	30.8 (23.0)	41.9 (34.4)	48.2	52.1

consistently retained over the tropics for all experiments in both seasons. The skill comparison for the JJA season shows that the predictions using CA and persisted SSTs yield similar skill, while the forecast using ECMWF SSTs gives somewhat lower skill than the other two retrospective experiments, particularly over Africa (Figs. 4b–d). For DJF, the most noticeable skill differences between the three retrospective experiments (Figs. 4f–h) occur over Africa, and to a lesser extent over eastern Australia, and some regions of North America and Asia. DMT\_Retro may give the best overall performance for DJF 2-m temperature, but regionally the other experiments occasionally perform better. For example, for JJA CA\_Retro performs best over Africa and DMT\_Retro performs worst; for DJF DMT\_Retro performs best over eastern Europe, and CA\_Retro performs worst.

The spatial distributions of skill of precipitation forecasts are noisier than those of 2-m temperature (Fig. 5), and the prediction skill of precipitation is generally lower than that of temperature. As seen in Figs. 5a–d, very little skill is indicated for JJA; what does exist is confined to Indonesia and northern South America. CA\_Retro performs worst of the three experiments at

this time of year, and that is reflected over those two regions. For all the three retrospective experiments some of the potential predictability captured by forcing the AGCM with observed SSTs, specifically that over West Africa and Indonesia/west Pacific, is not realized in the predictions as JJA is a time of the year when SST persistence is weak in both the tropical Atlantic and Pacific Oceans. Predictions for this season also encounter the obstacle of the “spring barrier” (Webster and Yang 1992) noted in ENSO predictions, so predictions of ENSO are often difficult for JJA based on information prior to May. For DJF, the patterns of skill are consistent across experiments, but as suggested in Fig. 3, the magnitude of skill increases with increasing complexity of SST prediction methodology. The correlation skill of CA\_Retro is slightly better than that from PST\_Retro, but worse than that from DMT\_Retro, with most notable difference over southern Africa, the southern United States, the Philippines, and northern South America. It is these regions that dominate the differences in summarized skill between the experiments (Figs. 2 and 3). The CA\_Retro outperforms PST\_Retro only during boreal winter seasons (e.g., DJF and JFM) when ENSO phenomena dominate over



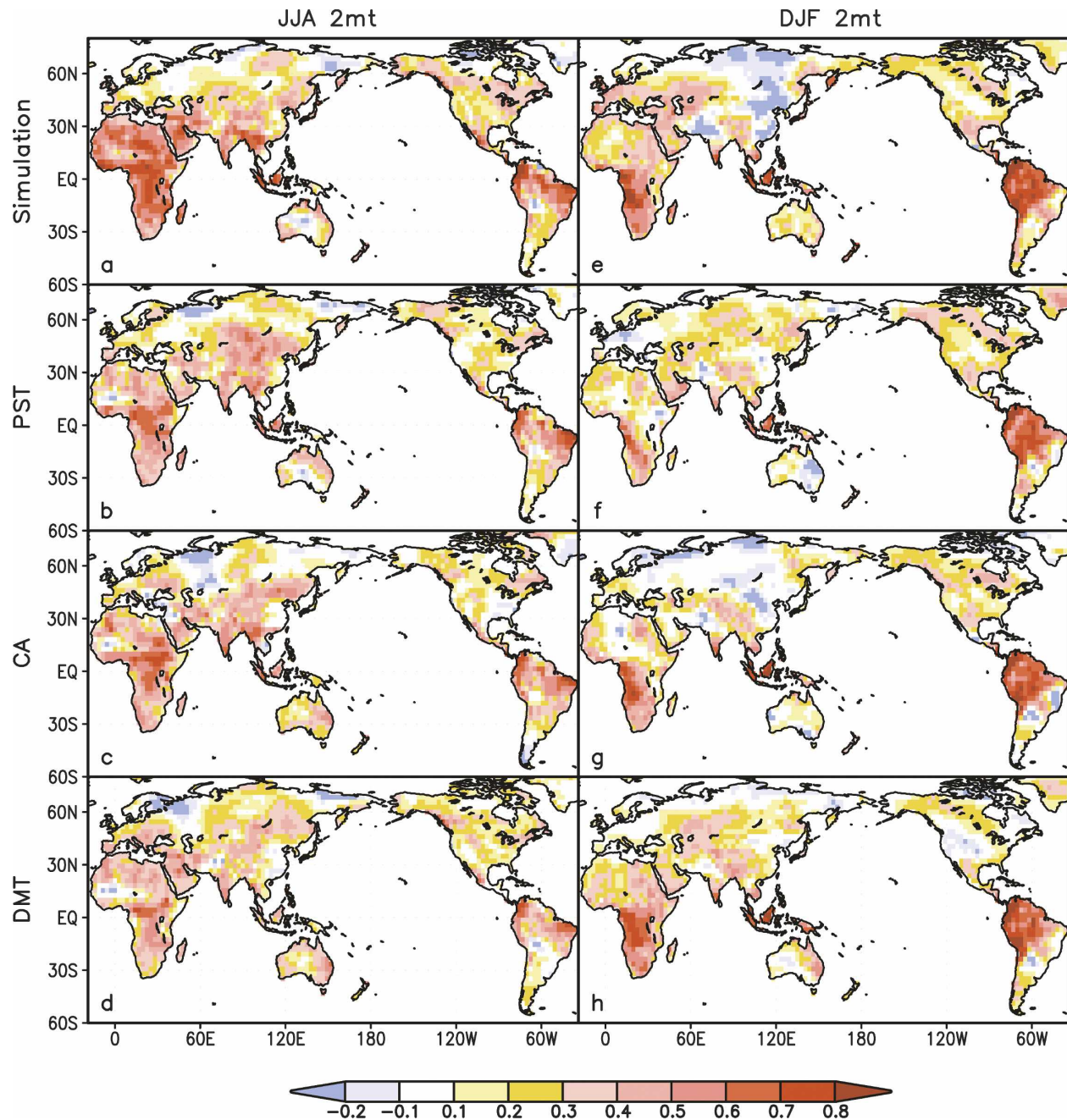


FIG. 4. Anomaly correlation maps of 2-m temperature anomaly correlations for (a)–(d) JJA and (e)–(h) DJF for 1968–2001. (a), (e) Simulations using observed SSTs; (b), (f) 1-month-lead predictions using persisted SST anomalies; (c), (g) 1-month-lead predictions using CA SSTs; (d), (h) 1-month-lead predictions using ECMWF SSTs.

the tropical Pacific. This is somewhat counterintuitive, as it might be expected that persistence would prevail during the peak phase of ENSO and evolving predictions to fare better than persistence during seasons of ENSO growth. Further, the DMT\_Retro precipitation skill is highest for all four seasons (MAM, JJA, SON, and DJF) when 1-month-lead predictions are available

for ECMWF SSTs; for the 2-month-lead forecasts, skill is higher than the 1-month-lead skill of the other experiments in that season (Table 2). The continental regions over the tropical land area most impacted by these differences in SST forcing include mainly Indonesia and northern South America for JJA, and Africa for DJF.

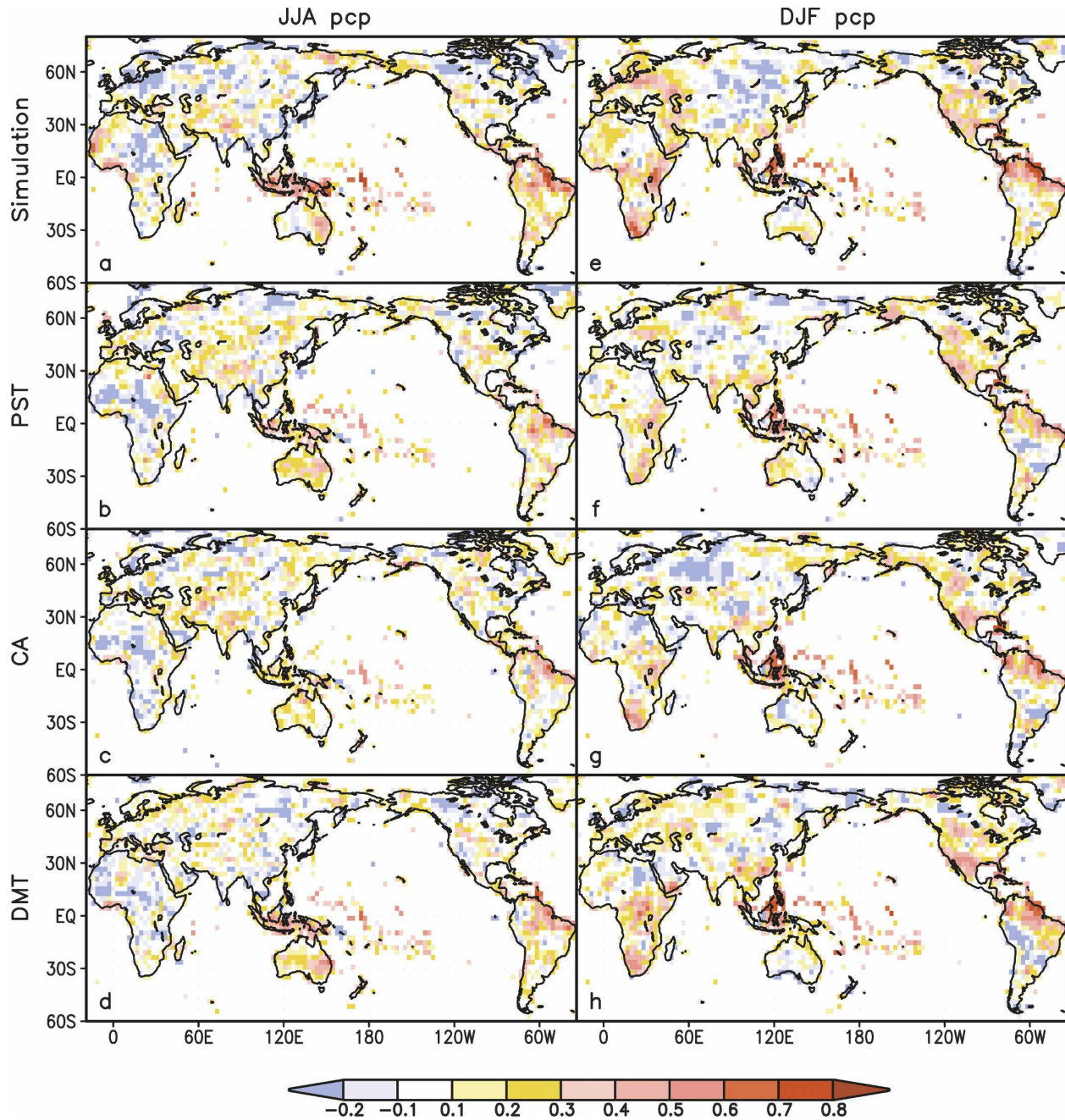


FIG. 5. Similar to Fig. 4, but for seasonal precipitation totals.

Loss of skill between the simulations, which use observed “perfect” SSTs, and the retrospective forecasts that use predicted “imperfect” SSTs is clear. The question is as follows: How much better do predicted SST fields have to be in order to be preferred? During the JJA season, persisted SSTs may be better than CA SSTs in some specific regions (e.g., Africa), but for other seasons (e.g., DJF) empirically predicted SSTs

appear to be a better selection in addition to dynamically predicted SSTs. To identify what differences in regional SSTs are most responsible for differences in regional skill of seasonal precipitation forecast, we now examine the skill of the three SST predictions employed here with emphasis on the seasons for which precipitation forecast skill was just presented: JJA and DJF.

c. Skill of precipitation forecast relative to skill of SST anomalies

1) REGIONAL SST SKILL DEPICTED BY TAYLOR DIAGRAMS

Taylor diagrams (Taylor 2001) are used to investigate the skill characteristics of the SST predictions with the aim of identifying aspects of those predictions that may be responsible for the skill difference in seasonal climate predictions. The Taylor diagrams, often used for model-to-model intercomparison or tracking changes in model performance, combine a few statistical quantities: anomaly pattern correlation coefficients of model and observed fields, root-mean-square error (RMSE), and the normalized ratio of the model and observed variance. These performance metrics are presented in one diagram, and thus provide a comprehensive figure of multimodel or multivariate performance in temporal-spatial scales. Further details of the Taylor diagrams can be found in Taylor (2001). As mentioned above, an individual point in a Taylor diagram corresponds to three statistical skill metrics: anomaly pattern correlation, RMSE, and normalized standard deviation. If a model performs nearly perfect, its position in the diagram is expected to be very close to the reference point, as shown in Fig. 6. In other words, the model fields that have little RMSE, high pattern correlation with respect to observations, and nearly identical-to-observed spatial-temporal variability or standard deviation lie closer to the reference point.

Certainly empirical methods can improve upon persistence. Figure 6 shows better overall skill characteristics for CA SSTAs than persisted SSTAs. The various SST prediction methodologies exhibit similar skill for the Pacific basin, especially in DJF, but not in the other basins. This suggests that some notable fraction of climate prediction skill comes from the accuracy of SST predictions outside the tropical Pacific, or at least that getting the SSTs wrong outside the Pacific can hurt climate prediction skill.

The higher relative predictability of the tropical Pacific basin is evident in all three prediction methodologies for both seasons. The skill characteristic quantifying the relative variance of the predicted SSTs seems to have the least impact on accuracy of the climate predictions. On the other hand, the skill represented by pattern correlation and RMSE, which is highest for the ECMWF SSTAs over all basins and thus the entire tropics, does seem to lead to higher climate prediction skill. The only evident relationship between the skill metrics of the individual basins across prediction methodologies appears to be related to predictive skill in the

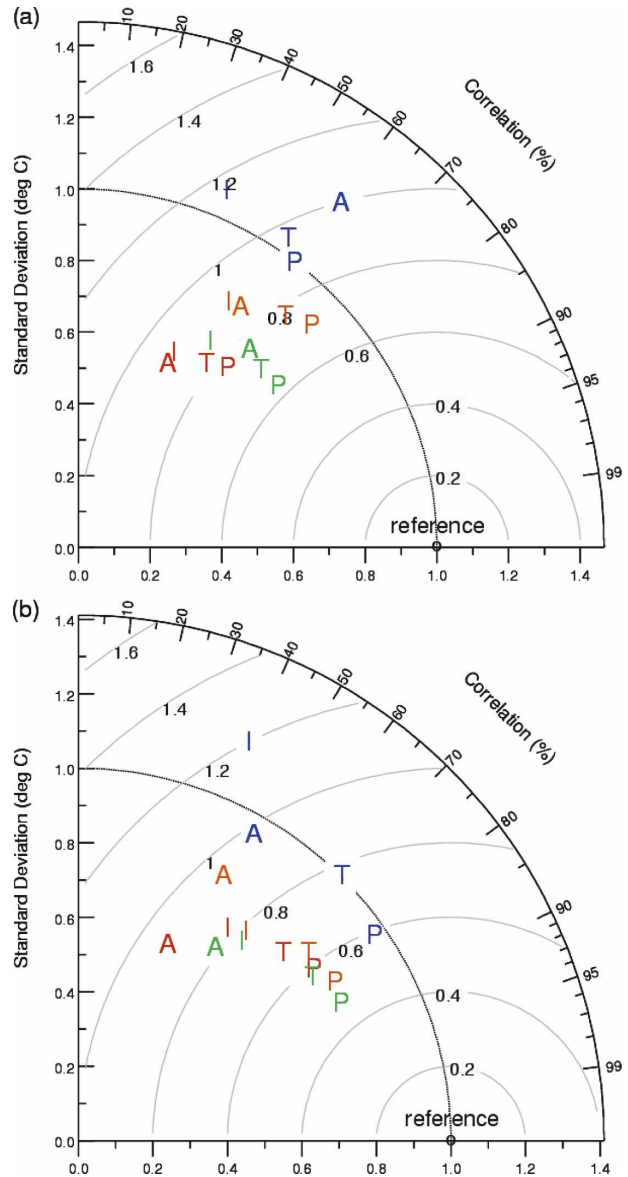


FIG. 6. Taylor diagrams displaying skill characteristics of 1-month-lead SST predictions produced by persistence (blue), CA (red), and ECMWF (orange) for (a) JJA and (b) DJF. Relative to observations (“reference” point) the plots indicate relative variance in units of standard deviation (radial coordinate), pattern correlation in percent (azimuthal coordinate), and RMSE (gray circles). The letters A, I, P, and T represent the Atlantic, Indian, and Pacific Ocean basins and the tropical oceans (20°S–20°N), respectively. Green letters represent multiscenario of the three SST predictions.

tropical Pacific. For example, the relative distance from the reference point for the tropical Pacific predictions in JJA corresponds to the relative distance from the reference point of each of the other basin predictions, which reconfirms previous suggestions that the tropical Pacific SSTs drive much of the SST variability in the

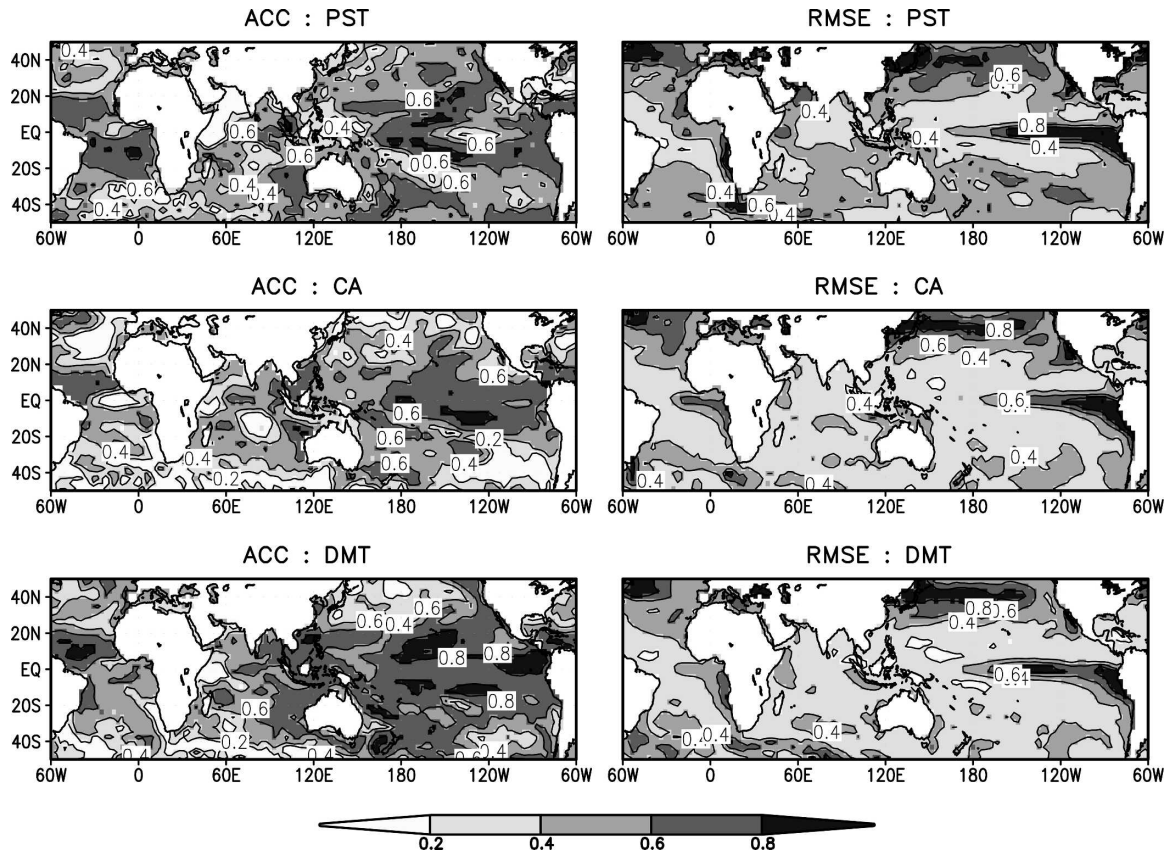


FIG. 7. Correlations and RMSE of SST anomalies in JJA between observations and 1-month-lead predictions from persistence, CA and ECMWF during 1968–2001. The latitude ranges from 50°S to 50°N. Contours for anomaly correlation coefficient (ACC) and RMSE are 0.2, 0.4, 0.6, and 0.8.

other tropical oceans on seasonal-to-interannual time scales. Of particular interest to potential advances in prediction methodologies is the fact that the multisenario SST obtained by simply averaging the three predictions outperforms all the individual SST predictions (according to correlation and RMSE) for all basins and in both seasons. We will come back to the multisenario SST in section 3d.

## 2) GEOGRAPHICAL FEATURES OF SST PREDICTION SKILL

The Taylor diagrams of SST predictions in the various tropical ocean basins provide summarizing information about the performance characteristics of the three SST prediction methodologies. The diagrams indicate to what extent the patterns of variability are well captured. The skill maps of SST predictions are required to gather information on performance of local variability. Figures 7 and 8 illustrate the spatial distribution of the SST anomaly correlation coefficients and RMSE.

In JJA the persisted SSTs give rise to higher correlation in the equatorial Atlantic and Indian Oceans compared to CA SST prediction (Fig. 7). The only exception is northwestern Indian Ocean, where the skill of CA SST forecast is better than that of the persistence. Over the eastern equatorial Pacific persisted SSTs indicate lower correlation than CA SSTs, but the overall skill over the tropical Pacific Ocean appears to be similar for the two types of predictions, as suggested in Fig. 6a. The RMSE patterns show similar error distributions between the two SST predictions, but the tropical Indian Ocean SSTs from persistence have greater errors than CA SSTs. The ECMWF SST predictions exhibit generally higher anomaly correlations and lower RMSE, particularly over the tropical Pacific and Indian Oceans. During DJF, although the anomaly correlation over the tropical Pacific is similar between the three SST predictions, CA SSTs correspond to higher skill than persistence over the tropical Indian Ocean (Fig. 8), particularly near the equator. Moreover, the eastern equatorial Indian Ocean exhibits low and even negative correlations for persisted SSTs, while

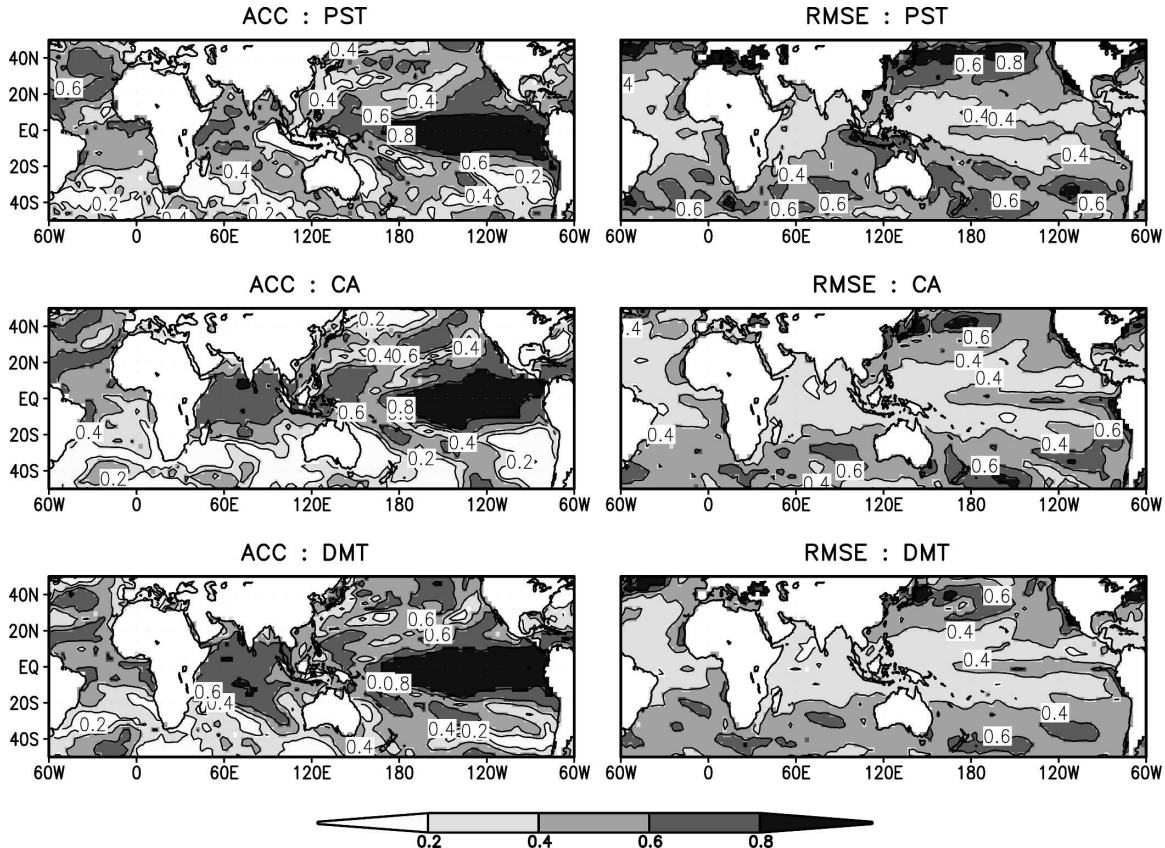


FIG. 8. Same as Fig. 7, but for DJF SST anomalies.

the CA SST prediction presents a high positive correlation in the same region. The RMSE appears similar between these two types of SST predictions (Fig. 8), but relatively greater errors occur for the persistence in the eastern and southern parts of the tropical Atlantic and Indian Oceans. The ECMWF SST predictions again indicate slightly higher correlation and lower RMSE than CA and persisted SSTs over the tropical Pacific and Indian Oceans. Overall, the three SST prediction methods present lower skill and larger errors for JJA than for DJF, consistent with overall higher climate prediction skill in DJF than in JJA.

### 3) SKILL OF PRECIPITATION FORECAST AND SST ANOMALIES

The 1-month-lead climate forecasts of total precipitation obtained by forcing the AGCM with three types of SST predictions differ only in the lower boundary conditions from sea surface. Therefore, skill differences in seasonal forecast of precipitation should be, almost entirely, ascribed to differences in the “fidelity” of the prescribed regional SST anomalies. Connecting skill

differences in precipitation forecasts and the skill patterns of SST predictions allows us to identify what errors in SST anomalies are contributing to skill reduction in seasonal climate forecasts relative to the AGCM’s potential skill.

During JJA the persisted SSTs (based on observed April SSTAs), especially in the tropical Atlantic Ocean, perform better than CA SSTs, with higher anomaly correlation and spatial-temporal variance closer to the observed (Fig. 6a). This corresponds to slightly better skill from PST\_Retro in precipitation forecast over South America as well as parts of North America (Fig. 5). The ECMWF SST predictions show advantages compared to the other two, particularly over the tropical Pacific and Indian Oceans, and thus the entire tropical ocean (Figs. 6a and 7). The response to these differences in SST predictions is that the best predictive skill of precipitation forecasts for JJA comes from using ECMWF SSTs as boundary forcing (Figs. 3 and 5), even though this skill is still much lower than the potential skill as shown by the simulations. Again, the tropical ocean SSTs exhibit low persistence and predictability for this season, in part because of the “spring barrier.”

During the DJF season, the skill of climate forecasts from CA\_Retro is better than that from PST\_Retro (Figs. 1–3), and the skill from DMT\_Retro is better still. This is evident over southern Africa, Australia, and northern South America (Figs. 4 and 5). Corresponding to the lower skill of the climate predictions using persisted SSTs is the lower skill of the SST patterns from persistence compared to CA and ECMWF SST prediction, in particular over the tropical Indian Ocean (Fig. 8). For example, the SST errors over the Indian Ocean from persistence (as shown by the blue letter “I” in Fig. 6b) are contributing to the relatively lower skill in precipitation forecast over some areas of Africa (Fig. 5). Moreover, the ECMWF SST prediction indicates higher skill and smaller RMSE than CA and persisted SSTs, particularly over the tropical Pacific and Indian Oceans (Figs. 6b and 8). This corresponds to the best overall skill of precipitation forecast for DJF by the AGCM forced with ECMWF SSTs.

The results presented above indicate the importance of accurate SST predictions in order to produce as accurate as possible climate forecasts. Since errors in prescribed SSTs can translate into significant losses in predictive climate skill (Goddard and Mason 2002), an appropriate selection of SST prediction methodology is crucial to ensure that the seasonal climate prediction from an AGCM provides the best forecast products for a specific region. How superior must a prediction methodology be to yield notably better climate predictions is difficult to quantify. The skill differences in precipitation forecasts over the tropical land area suggest that, over specific regions, the preferred prediction approach may differ with season and even with the predicted field. While one methodology might perform better for some regions, it likely performs worse for others. Therefore, it seems reasonable to utilize different scenarios of SST predictions varying with month of the year and by specific ocean basins, or better yet, an ensemble of differing SST prediction scenarios, which may be weighted seasonally and regionally based on past performance of SST predictions, similarly to current multimodel ensembling methodologies (e.g., Barnston et al. 2003).

#### d. Skill of forecasts forced with multiple SST scenarios

The benefit of multimodel methods in seasonal forecasting has been increasingly employed in research and operations (e.g., Barnston et al. 2003; Palmer et al. 2004; Robertson et al. 2004; Hagedorn et al. 2005). As a demonstration of concept, the three retrospective experiments are combined into a “multiscenario” retro-

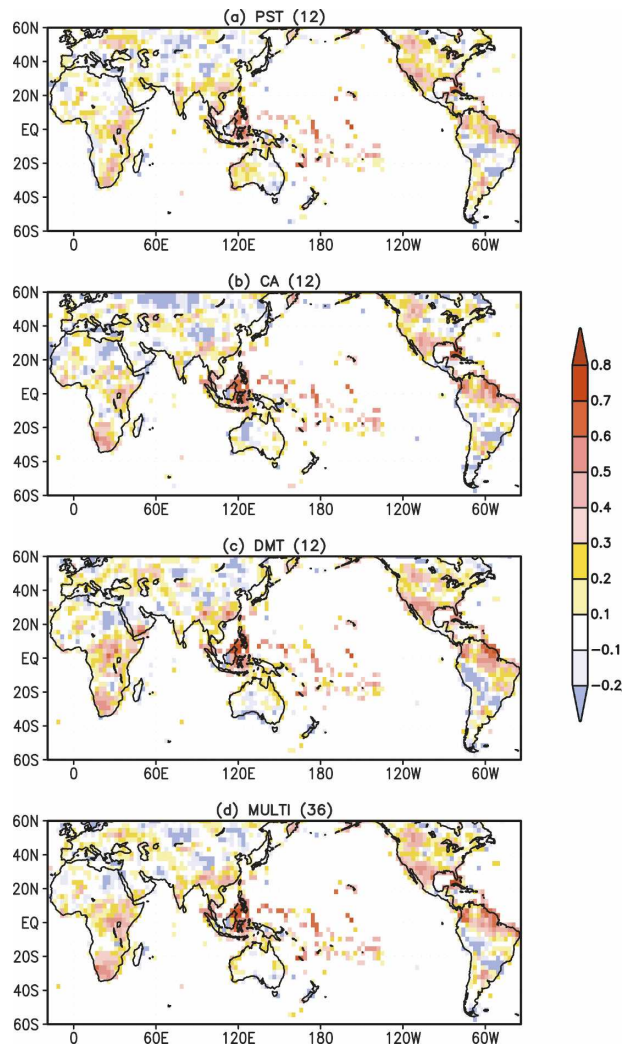


FIG. 9. Anomaly correlations of DJF precipitation 1-month-lead forecasts during the 1968–2001 period for three 12-member ensemble forecasts of (a) PST\_Retro, (b) CA\_Retro, and (c) DMT\_Retro, as well as (d) combined multiscenario of the three experiments.

spective forecast, using simple averaging. We refer to this as MULTI\_Retro, and to the averaged SST prediction as multi-SSTA. The connection between the two assumes that the AGCMs respond predominantly linearly to SST forcing (Barsugli and Sardeshmukh 2002; Barsugli et al. 2006). Figure 9 illustrates the anomaly correlation skill of 1-month-lead precipitation forecast by MULTI\_Retro (36-member ensemble mean) for DJF, based on the same 12-member of PST\_Retro, CA\_Retro, and DMT\_Retro, comparing against these three individual retrospective experiments.

For the DJF season, MULTI\_Retro captures the “best” of the three retrospective forecasts, particularly over southern United States, eastern and southern

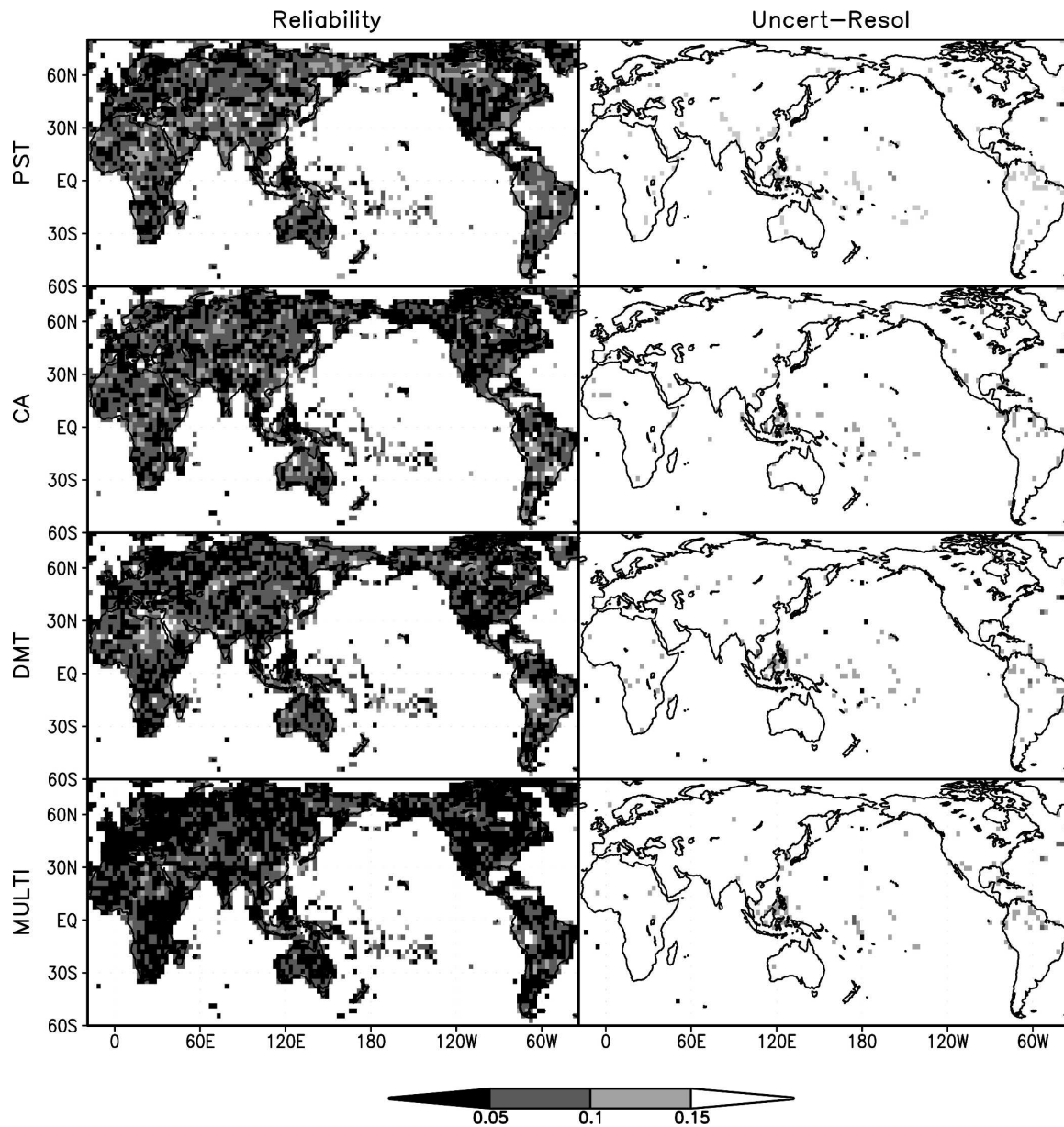


FIG. 10. Half-Brier score decomposition: (left) reliability and (right) uncertainty minus resolution for two-category forecasts of DJF precipitation for the period 1968–2001. The three retrospective experiments and multiscenario forecasts are labeled in the same order as in Fig. 9.

Africa, Indonesia, and northern South America. The multiscenario forecast has better skill overall (Fig. 3) than exhibited individually by the forecasts from the three experiments that went into MULTI\_Retro and also yields slightly better skill than any of the individual scenario predictions over regions, such as over the western United States and southeastern South America (Fig. 9). This suggests the potential advantage of climate predictions using multiscenario ensemble SSTs as boundary conditions.

To further investigate the predictive skill of these individual experiments and multiscenario forecasts, we examine probabilistic measures of skill of the DJF precipitation forecast, as shown in Fig. 10 for the half-Brier score and its decomposition. See Wilks (1995, section 7.4.3) for further information about this skill metric. The terms in the half-Brier decomposition are the following: reliability, which indicates how representative are the probabilities assigned to a category; the resolution, which measures how much the forecast probabili-

ties deviate from climatological expectations; and uncertainty, which is independent of the forecast quality. The half-Brier score, which is the probabilistic equivalent to mean-squared error, is computed for two-category (above and below median) forecasts of DJF precipitation. The half-Brier score is a negatively oriented measure of skill, meaning that smaller values indicate better forecast quality. The reliability skill of precipitation forecasts indicates better forecast skill, overall, for MULTI\_Retro than for any of the three individual experiments (Fig. 10), even though their uncertainty minus resolution term does not show much difference or improvements.

Similar to the comparison of anomaly correlations between the three individual retrospective experiments and the multiscenario forecasts, the reliability skill of the precipitation forecasts using the ECMWF SSTs is slightly better than the skill of two-category probabilistic forecasts using persistence or empirically predicted CA SSTs. Furthermore, the multiscenario forecasts demonstrate better forecast quality than the “best” of the three individuals over some regions, such as the western United States, southeastern South America, and southern Africa (left column in Fig. 10).

The performance of the corresponding multiscenario SST predictions for DJF, along with the three individual scenarios, persisted, CA and ECMWF SSTs, is shown in the Taylor diagram of SST prediction skill (Fig. 6b). Indeed, the multi-SSTA indicates better performance than the three scenarios individually, characterized by smaller RMSE and higher anomaly pattern correlation while keeping a similar magnitude of spatial-temporal SST variability. This represents improvement over all the three tropical ocean basins, particularly over the tropical Pacific and Atlantic Oceans.

The advantages of multiscenario retrospective forecasts of tropical precipitation and temperature for all the four available seasons (MAM, JJA, SON, and DJF) are also shown in Fig. 3. In most of the cases, MULTI\_Retro performs better than the three individual experiments. In the other cases, MULTI\_Retro performs on par with the best of the three, which is not always derived from the same SST prediction methodology. Regionally, MULTI\_Retro may not perform as well as the best experiment individually, but this is likely to be outweighed by the benefit of more robust overall performance. More sophisticated weighting of the forecast methods, which vary regionally and seasonally, based on historical performance, may yield additional improvements in skill.

#### 4. Summary and discussion

Three sets of retrospective forecasts made by forcing the ECHAM4.5 AGCM with SST predictions obtained from methodologies of varying complexity (persisted and empirically and dynamically predicted SSTs) are used to assess the relative skill of seasonal climate forecasts. By comparing the retrospective AGCM forecasts using these prescribed SSTs, along with the simulation runs using observed SSTs, the impacts of imperfect SST predictions on the accuracy of climate forecasts are investigated. Since “perfect” predictions of global SSTs are not possible, skill levels of real-time seasonal forecasts made by AGCM predictions are expected to be lower than that of the simulations. The skill loss in real-time forecasts relative to the simulations depends upon the errors introduced by predictions of regional SSTs. Although the results in this study may be specific to one AGCM, the analysis methodology can be applied to other AGCMs and even other SST prediction scenarios.

Our results demonstrate some seasonally and regionally dependent gain in predictive skill of climate forecasts using dynamically predicted SSTs compared to persistence or empirical SST predictions, based on the examples of those SST predictions chosen for this study. In some regions, the climate predictions based on persisted SSTs beat those based on empirically predicted CA SSTs for a few specific seasons (e.g., JJA). DJF is a typical season when skill of the forecast using CA SSTs exceeds that using persistence. This appears to stem from higher skill of CA SST prediction compared to SSTA persistence for the DJF season, especially over the tropical Indian Ocean. Conversely, during the JJA season, persisted SSTAs outperform CA SSTAs over the tropical oceans, with the largest skill difference occurring over the Atlantic Ocean. This difference in SST pattern statistics corresponds to relatively lower skill for the CA\_Retro precipitation forecast. The ECMWF SSTs indicate better skill and lower errors over most of the tropical oceans, and thus give rise to overall higher skill of seasonal precipitation forecast over the tropical land area than using CA or persisted SSTAs for all the four available seasons.

The multiscenario retrospective forecast obtained by averaging the three individual experiments performs as well as, or better than, the skill exhibited individually by the best of the three experiments. In particular, the DJF precipitation forecast by MULTI\_Retro exhibits slightly better skill than all the three individual scenarios over many parts of the world. This is related to better skill characteristics for the averaged SST predic-



tion than for the three individual SST predictions over the tropical oceans.

Based on our analysis and skill comparison, a few critical points are outlined as follows: 1) The AGCM simulations overestimate the realizable skill of seasonal climate forecasts made with predicted SSTs as boundary forcing; 2) skill of seasonal predictions using dynamically predicted SSTAs from a coupled ocean–atmosphere model generally improves upon the forecast skill using persisted or empirically based SST predictions; 3) the skills of retrospective forecasts using persisted and CA SSTs are generally comparable; 4) skill differences in the tropical ocean SSTs contribute to differences in regional skill of seasonal climate predictions; 5) the multiscenario retrospective forecast created by averaging the three individual scenarios has better skill overall than exhibited individually by the three experiments. The deterministic and probabilistic measures of predictive skill both indicate the advantage of multiscenario or mixed forecasts. The details of some of our results and conclusions may be sensitive to the specific examples of empirical and dynamical SST predictions used here, and it would be worthwhile bringing in results using other examples to test robustness.

Overall, the atmosphere experiences the broad pattern of SST variability, not just the local manifestations. Our diagnostic results imply potential value in producing “superensembles,” or multiscenario combinations, spanning different SST prediction scenarios, which would likely outperform forecasts based on any single scenario of predicted SSTs.

It is expected that these retrospective experiments and their skill evaluation will improve understanding of seasonal climate predictability obtainable from AGCMs using predicted SSTs as boundary conditions. They can also provide a comparative basis for validation of coupled ocean–atmosphere global climate models. The retrospective experiments provide an objective basis for evaluating the utility of the AGCM-based seasonal forecasts for particular climate-sensitive management problems. By coupling retrospective forecasts with impact (e.g., crop or forage yields, disease risk, water reservoir inflow) and economic decision models, potential users can evaluate the likely benefits and risks associated with particular applications of a forecast system.

*Acknowledgments.* We are grateful to two anonymous reviewers for their constructive comments that improved the manuscript significantly. Dr. Xiaofeng Gong assisted in setting up the retrospective experiments. The retrospective runs forced with persisted SSTs were executed by Dr. Hua Sheng. The computing

for this project was partially provided by a grant from the NCAR CSL program to the IRI. We also thank Anthony Barnston and James Hansen for their comments on an earlier draft of this manuscript. This work is funded by a grant/cooperative agreement from the National Oceanic and Atmospheric Administration (NA07GP0123 and NA050AR4311004). The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its sub-agencies.

## REFERENCES

- Barnston, A. G., S. J. Mason, L. Goddard, D. G. DeWitt, and S. E. Zebiak, 2003: Multimodel ensembling in seasonal climate forecasting at IRI. *Bull. Amer. Meteor. Soc.*, **84**, 1783–1796.
- Barsugli, J. J., and P. D. Sardeshmukh, 2002: Global atmospheric sensitivity to tropical SST anomalies throughout the Indo-Pacific Basin. *J. Climate*, **15**, 3427–3442.
- , S.-I. Shin, and P. D. Sardeshmukh, 2006: Sensitivity of global warming to the pattern of tropical ocean warming. *Climate Dyn.*, **27**, 483–492.
- Bengtsson, L., U. Schelese, E. Roeckner, M. Latif, T. P. Barnett, and N. E. Graham, 1993: A two-tiered approach to long-range climate forecasting. *Science*, **261**, 1026–1029.
- DeWitt, D. G., 2005: Retrospective forecasts of interannual sea surface temperature anomalies from 1982 to present using a directly coupled atmosphere–ocean general circulation model. *Mon. Wea. Rev.*, **133**, 2972–2995.
- Goddard, L., and S. J. Mason, 2002: Sensitivity of seasonal climate forecasts to persisted SST anomalies. *Climate Dyn.*, **19**, 619–631.
- , and M. Dille, 2005: El Niño: Catastrophe or opportunity. *J. Climate*, **18**, 651–665.
- , S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basher, and M. A. Cane, 2001: Current approaches to seasonal to interannual climate predictions. *Int. J. Climatol.*, **21**, 1111–1152.
- , A. G. Barnston, and S. J. Mason, 2003: Evaluation of the IRI’s “Net Assessment” seasonal climate forecasts: 1997–2001. *Bull. Amer. Meteor. Soc.*, **84**, 1761–1781.
- Guérémy, J.-F., M. Déqué, A. Braun, and J.-P. Piedelièvre, 2005: Actual and potential skill of seasonal predictions using the CNRM contribution to DEMETER: Coupled versus uncoupled model. *Tellus*, **57A**, 308–319.
- Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting. Part I: Basic concept. *Tellus*, **57A**, 219–233.
- Livezey, R. E., and W. Y. Chen, 1983: Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.*, **111**, 46–59.
- Mason, S. J., L. Goddard, N. E. Graham, E. Yulaeva, L. Sun, and P. A. Arkin, 1999: The IRI seasonal climate prediction system and the 1997/98 El Niño event. *Bull. Amer. Meteor. Soc.*, **80**, 1853–1873.
- Mitchell, T. D., and P. D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.*, **25**, 693–712.

- Palmer, T. N., and D. L. T. Anderson, 1994: The prospects for seasonal forecasting—A review paper. *Quart. J. Roy. Meteor. Soc.*, **120**, 755–793.
- , and Coauthors, 2004: Development of a European multi-model ensemble system for seasonal-to-interannual prediction (DEMETER). *Bull. Amer. Meteor. Soc.*, **85**, 853–872.
- Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate*, **7**, 929–948.
- , N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609–1625.
- Robertson, A. W., U. Lall, S. E. Zebiak, and L. Goddard, 2004: Improved combination of multiple atmospheric GCM ensembles for seasonal prediction. *Mon. Wea. Rev.*, **132**, 2732–2744.
- Roeckner, E., and Coauthors, 1996: The atmospheric general circulation model ECHAM4: Model description and simulation of present-day climate. Max Planck Institute for Meteorology Tech. Rep. 218, 90 pp.
- Smith, T. M., and R. W. Reynolds, 2003: Extended reconstruction of global sea surface temperatures based on COADS data (1854–1997). *J. Climate*, **16**, 1495–1510.
- , and —, 2004: Improved extended reconstruction of SST (1854–1997). *J. Climate*, **17**, 2466–2477.
- Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192.
- Van den Dool, H. M., 1994: Searching for analogues, how long must we wait? *Tellus*, **46A**, 314–324.
- Webster, P. J., and S. Yang, 1992: Monsoon and ENSO: Selectively interactive systems. *Quart. J. Roy. Meteor. Soc.*, **118**, 877–926.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press, 467 pp.

## CORRIGENDUM

Due to a press error, an incorrect version of Fig. 9 was published in “Predictive Skill of AGCM Seasonal Climate Forecasts Subject to Different SST Prediction Methodologies,” by Shuhua Li, Lisa Goddard, and David G. DeWitt, which was published in the *Journal of Climate*, Vol. 21, No. 10, 2169–2186. On p. 2182, Fig. 9 was published with a black and white scale bar, instead of a color bar. The correct version of Fig. 9 is shown below.

The staff of the *Journal of Climate* regret any inconvenience this error may have caused.

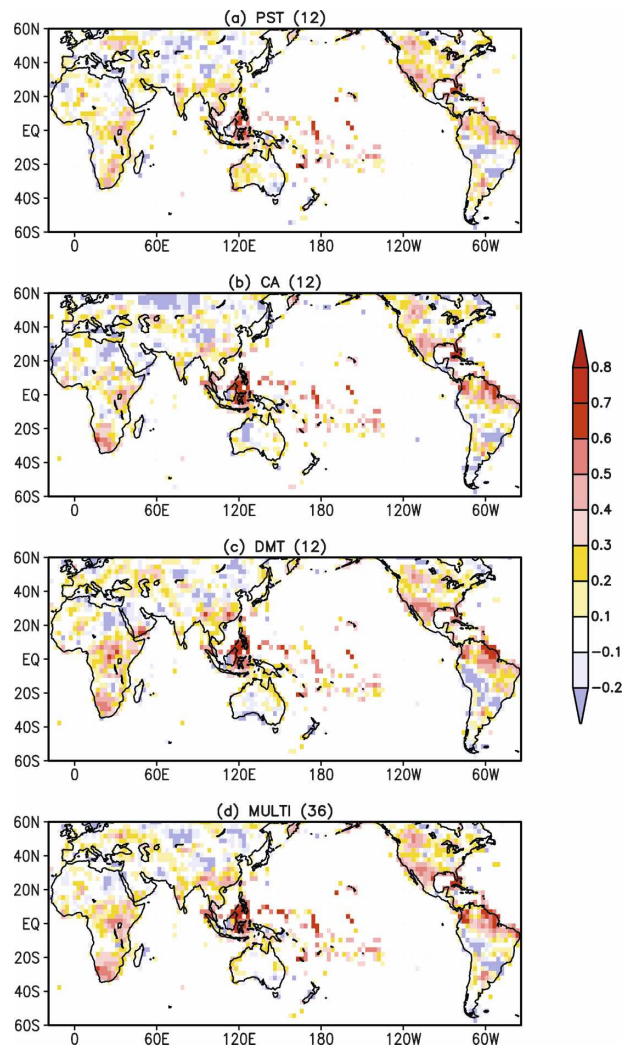


FIG. 9. Anomaly correlations of DJF precipitation 1-month-lead forecasts during the 1968–2001 period for three 12-member ensemble forecasts of (a) PST\_Retro, (b) CA\_Retro, and (c) DMT\_Retro, as well as (d) combined multiscenario of the three experiments.