

Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004)

Bin Wang · June-Yi Lee · In-Sik Kang · J. Shukla · C.-K. Park · A. Kumar · J. Schemm · S. Cocke · J.-S. Kug · J.-J. Luo · T. Zhou · B. Wang · X. Fu · W.-T. Yun · O. Alves · E. K. Jin · J. Kinter · B. Kirtman · T. Krishnamurti · N. C. Lau · W. Lau · P. Liu · P. Pegion · T. Rosati · S. Schubert · W. Stern · M. Suarez · T. Yamagata

Received: 31 January 2008 / Accepted: 18 August 2008 / Published online: 16 September 2008
© The Author(s) 2008. This article is published with open access at Springerlink.com

Abstract We assessed current status of multi-model ensemble (MME) deterministic and probabilistic seasonal prediction based on 25-year (1980–2004) retrospective forecasts performed by 14 climate model systems (7 one-tier and 7 two-tier systems) that participate in the Climate Prediction and its Application to Society (CliPAS) project sponsored by the Asian-Pacific Economic Cooperation Climate Center (APCC). We also evaluated seven DEMETER models' MME for the period of 1981–2001 for comparison. Based on the assessment, future direction for improvement of seasonal prediction is discussed. We found that two measures of probabilistic forecast skill, the Brier Skill Score (BSS) and Area under the Relative Operating

Characteristic curve (AROC), display similar spatial patterns as those represented by temporal correlation coefficient (TCC) score of deterministic MME forecast. A TCC score of 0.6 corresponds approximately to a BSS of 0.1 and an AROC of 0.7 and beyond these critical threshold values, they are almost linearly correlated. The MME method is demonstrated to be a valuable approach for reducing errors and quantifying forecast uncertainty due to model formulation. The MME prediction skill is substantially better than the averaged skill of all individual models. For instance, the TCC score of CliPAS one-tier MME forecast of Niño 3.4 index at a 6-month lead initiated from 1 May is 0.77, which is significantly higher than the

B. Wang (✉) · J.-Y. Lee · X. Fu · P. Liu
Department of Meteorology and International Pacific Research Center, IPRC, School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, HI 96822, USA
e-mail: wangbin@hawaii.edu

I.-S. Kang · J.-S. Kug
School of Earth and Environmental Sciences, Seoul National University, Seoul, South Korea

J. Shukla · E. K. Jin · J. Kinter · B. Kirtman
Climate Dynamics Program, George Mason University and COLA, Calverton, MD, USA

C.-K. Park
APEC Climate Center, Busan, South Korea

A. Kumar · J. Schemm
Climate Prediction Center/NCEP, Camp Springs, MD, USA

S. Cocke · T. Krishnamurti
Florida State University, Tallahassee, FL, USA

J.-J. Luo · T. Yamagata
Frontier Research Center for Global Change, Yokohama, Japan

T. Zhou · B. Wang
LASG/Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

W.-T. Yun
Korean Meteorological Administration, Seoul, South Korea

O. Alves
Bureau of Meteorology Research Center, Melbourne, Australia

N. C. Lau · T. Rosati · W. Stern
Geophysical Fluid Dynamics Laboratory/NOAA, Princeton University, Princeton, NJ, USA

W. Lau · P. Pegion · S. Schubert · M. Suarez
Godard Space Flight Center/NASA, Greenbelt, MD, USA

T. Yamagata
University of Tokyo, Tokyo, Japan

corresponding averaged skill of seven individual coupled models (0.63). The MME made by using 14 coupled models from both DEMETER and CliPAS shows an even higher TCC score of 0.87. Effectiveness of MME depends on the averaged skill of individual models and their mutual independency. For probabilistic forecast the CliPAS MME gains considerable skill from increased forecast reliability as the number of model being used increases; the forecast resolution also increases for 2 m temperature but slightly decreases for precipitation. Equatorial Sea Surface Temperature (SST) anomalies are primary sources of atmospheric climate variability worldwide. The MME 1-month lead hindcast can predict, with high fidelity, the spatial-temporal structures of the first two leading empirical orthogonal modes of the equatorial SST anomalies for both boreal summer (JJA) and winter (DJF), which account for about 80–90% of the total variance. The major bias is a westward shift of SST anomaly between the dateline and 120°E, which may potentially degrade global teleconnection associated with it. The TCC score for SST predictions over the equatorial eastern Indian Ocean reaches about 0.68 with a 6-month lead forecast. However, the TCC score for Indian Ocean Dipole (IOD) index drops below 0.40 at a 3-month lead for both the May and November initial conditions due to the prediction barriers across July, and January, respectively. The MME prediction skills are well correlated with the amplitude of Niño 3.4 SST variation. The forecasts for 2 m air temperature are better in El Niño years than in La Niña years. The precipitation and circulation are predicted better in ENSO-decaying JJA than in ENSO-developing JJA. There is virtually no skill in ENSO-neutral years. Continuing improvement of the one-tier climate model's slow coupled dynamics in reproducing realistic amplitude, spatial patterns, and temporal evolution of ENSO cycle is a key for long-lead seasonal forecast. Forecast of monsoon precipitation remains a major challenge. The seasonal rainfall predictions over land and during local summer have little skill, especially over tropical Africa. The differences in forecast skills over land areas between the CliPAS and DEMETER MMEs indicate potentials for further improvement of prediction over land. There is an urgent need to assess impacts of land surface initialization on the skill of seasonal and monthly forecast using a multi-model framework.

1 Background

In the past two decades, climate scientists have made ground-breaking progress in dynamic seasonal prediction. The advent of dynamic climate prediction can be traced back to El Niño forecast that used an intermediate-

complexity coupled ocean-atmosphere model (Cane et al. 1986). In the early part of the 1990s, Bengtsson et al. (1993) proposed a “two-tier” approach for dynamical seasonal forecast, in which the global SST anomalies are first predicted, and an atmospheric general circulation model (GCM) is subsequently forced by the pre-forecasted SST to make a future seasonal prediction. At that time, ENSO was recognized as the major source of the predictability of the tropical and mid-latitude climate variations through ENSO teleconnection, which depends critically on the correct simulations of mean climatology. However, the coupled atmosphere-ocean GCMs (CGCMs) then had considerable errors in simulating the observed mean climatology as well as anomalous conditions of the tropical ocean and atmosphere (Mechoso et al. 1995). Thus, the two-tier system had an obvious advantage over the direct use of the CGCMs.

While the two-tier approach was a useful strategy to capture better teleconnection, recent research advances using CGCMs suggest that prediction of certain phenomena (e.g., summer monsoon precipitation) may require taking into account local monsoon-warm pool ocean interactions (Wang et al. 2000, 2003; Wu and Kirtman 2005; Kumar et al. 2005). It has been shown that the low performance of atmospheric GCMs forced by observed SST in simulation of the Asian summer monsoon variability is partially attributed to the neglect of atmospheric feedback on SST (Wang et al. 2004a). In the absence of the monsoon-ocean interaction, all models yield positive SST-rainfall correlations that are at odds with observations in the heavily precipitating summer monsoon region (Wang et al. 2005).

Toward the end of the twentieth century, a new era of seasonal forecast with coupled GCMs (also known as the one-tier approach) began, due to rapid progress made in coupled climate models (Latif et al. 2001; Davey et al. 2002; Schneider et al. 2003) and due to a concerted international effort (through the Tropical Ocean-Global Atmosphere program) to monitor tropical ocean variations. Although the CGCMs still have significant systematic errors, many have demonstrated their capacity to reproduce realistic characteristics of ENSO (e.g., Latif et al. 1994; Ji et al. 1994; Rosati et al. 1997; Kirtman and Zebiak 1997; Vintzileos et al. 1999a, b; Guilyardi et al. 2004) and the major modes of interannual variability for the Asian-Australian monsoon system (Wang et al. 2008). It has been increasingly recognized that the CGCMs are the most promising ultimate tools for seasonal prediction. The unprecedented 1997 El Niño was fairly well predicted 3–6 months in advance using a CGCM (Anderson et al. 2003). Since the beginning of the twenty-first century, a number of meteorological centers worldwide have implemented routine dynamical seasonal predictions using

coupled atmosphere–ocean–land climate models (Alves et al. 2003; Palmer et al. 2004; Saha et al. 2006).

For the two-tier systems, the physical basis for seasonal prediction lies in slowly varying lower boundary forcing, especially the anomalous SST (as well as the land surface) forcing (Charney and Shukla 1981; Shukla 1998). For the one-tier systems, prediction of ENSO and associated climate variability is essentially an initial value problem (Palmer et al. 2004). The slowly varying lower boundary of the atmosphere is evolving as a result of feedback among various components of the climate system. The climate predictability in nature and in CGCMs comes from “slow” coupled (atmosphere–ocean–land–ice) dynamics and initial memories in ocean and land surfaces.

Atmospheric chaotic dynamics may cause seasonal forecast errors, inherently limiting seasonal climate predictability. Since the seasonal predictability does not depend on atmospheric initial conditions, an ensemble forecast with different atmospheric initial conditions was developed to reduce the errors arising from atmospheric chaotic dynamics. Another considerable source of seasonal forecast errors arises from uncertainties in model parameterizations of unresolved sub-grid scale processes. In an individual model, stochastic physics schemes have been developed to alleviate the uncertainty arising from the sub-grid scales (Buizza et al. 1999; Shutts 2005; Bowler et al. 2008), which are now operationally used in European Centre for Medium-range Weather Forecast (ECMWF) and United Kingdom Meteorological Office for medium-range forecast. Meanwhile, a more effective way, the multi-model ensemble (MME) approach, was designed for quantifying forecast uncertainties due to model formulation near the turn of this century (Krishnamurti et al. 1999, 2000; Doblas-Reyes et al. 2000; Shukla et al. 2000; Palmer et al. 2000). The idea behind the MME is that if the model parameterization schemes are independent of each other, the model errors associated with the model parameterization schemes may be random in nature; thus, an average approach may cancel out the model errors contained in individual models.

In general, the MME prediction is superior to the predictions made by any single-model component for both two-tier systems (Krishnamurti et al. 1999, 2000; Palmer et al. 2000; Shukla et al. 2000; Barnston et al. 2003) and one-tier systems (Hagedorn et al. 2005; Doblas-Reyes et al. 2005; Yun et al. 2005). A number of international projects have organized multi-model intercomparison and synthesis, among which the most comprehensive projects are the European Union-sponsored “Development of a European Multi-model Ensemble System for Seasonal to Inter-Annual Prediction (DEMIETER; Palmer et al. 2004) and the Climate Prediction and its Application to Society (CliPAS) project, sponsored by the Asian-Pacific Economic Cooperation (APEC) Climate Center (APCC).

The APCC/CliPAS project was formally established in April 2005 as a research and development component of APCC. One of the objectives of CliPAS is to develop a well-validated MME prediction system and to study the predictability of the seasonal and sub-seasonal climate variations. Some of CliPAS models aim to become operational. The current CliPAS team is a coordinated research body consisting of 12 institutions and involving a large group of climate scientists from United States, South Korea, Japan, China, and Australia. The CliPAS team has analyzed historical retrospective predictions made by seven DEMETER one-tier systems for the 1980–2001 period and 14 CliPAS model systems for the 1980–2004 period. A number of published papers have documented the error growth and predictability of ENSO (Jin et al. 2008), the predictability of the major modes of Asian-Australia monsoon variability (Wang et al. 2008), the predictability and prediction skill of the intraseasonal variations (Kim et al. 2008), the performance of coupled models on mean states and its relation to seasonal prediction skills (Lee et al. 2008), and the optimal MME method for seasonal climate prediction (Kug et al. 2008). Here we present an overall assessment of the seasonal forecast skills of the state-of-the-art MME by bringing together high-quality retrospective forecast data issued from both the DEMETER and CliPAS project in order to gain a better understanding of the factors that limit our capability to improve seasonal prediction. Issues important for MME approach and methodology are discussed.

2 Models and evaluation methods

2.1 The models

The CliPAS project has 14 climate prediction models. Table 1 lists the acronyms of the institutions and models mentioned in the text. Tables 2 and 3 present a brief summary of model specifications for the seven two-tier and seven one-tier models, respectively, as well as the current status of retrospective forecasts. The APCC/CliPAS models have generated ensemble retrospective forecasts for the approximately common period of 1980–2004. Each model has a different forecast length and ensemble size (Tables 2, 3), but all models were integrated from around 1 May to at least 30 September for the boreal summer season (JJA hereafter), and integrated from 1 November to at least 31 March for the boreal winter season (DJF hereafter).

All two-tier models except NCEP GFS were forced by the global SST field pre-forecasted by the Seoul National University (SNU) statistical–dynamical forecast model (Kug et al. 2007a). The SNU statistical–dynamical SST forecast system for global SST prediction is based on four different

Table 1 Acronym names of institutions and models used in the text

Acronym	Full names
APCC	Asia-Pacific Economic Cooperation Climate Center
BMRC	Bureau of Meteorology Research Center (BMRC)
CES	Climate Environment System Research Center
CFS	Climate Forecast System
CliPAS	Climate Prediction and its Application to Society
COLA	Center for Ocean–Land–Atmosphere Studies
DEMIETER	Development of European Multimodel Ensemble system for seasonal-to-interannual prediction
FRCGC	Frontier Research Center for Global Change
FSU	Florida State University
GAMIL	Grid-point Atmospheric Model of IAP/LASG
GCPS	Global Climate Prediction System
GFDL	Geophysical Fluid Dynamic Lab
GFS	Global Forecast System
IAP	Institute of Atmospheric Physics
KMA	Korean Meteorological Administration
LASG	Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics
NASA	National Aeronautics and Space Administration
NCEP	National Center for Environmental Prediction
POAMA	Predictive Ocean–Atmosphere Model for Australia
SINTEX-F	Scale INTERaction Experiment-FRCGC
SNU	Seoul National University
UH	University of Hawaii

Table 2 Description of APCC/CliPAS two-tier prediction models

Institute	AGCM	Ensemble member	SST BC	Period and forecast lead time	Reference
FSU	FSUGSM T63 L27	10	SNU SST forecast	1979–2004 5-month	Cocke and LaRow (2000)
GFDL	AM2 2°lat × 2.5°lon L24	10	SNU SST forecast	1979–2003 5-month	GFDL GAMD Team (2004)
LASG/IAP	GAMIL 2.8°lat × 2.8°lon L26	6	SNU SST forecast	1979–2004 6-month	Wang et al. (2004b)
NCEP	GFS T62 L64	15	CFS SST forecast	1981–2004 6-month	Saha et al. (2006)
SNU/KMA	GCPS T63 L21	6	SNU SST forecast	1979–2002 4-month	Kang et al. (2004)
UH	CAM2 T42 L26	10	SNU SST forecast	1979–2003 6-month	Liu et al. (2005)
UH	ECHAM4 T31 L19	10	SNU SST forecast	1979–2003 6-month	Roeckner et al. (1996)

models: an intermediate dynamic model for tropical Pacific SST, a lagged linear regression model for Indian Ocean SST (Kug et al. 2004), a pattern projection model for global SST, and a persistent prediction model. Detailed description of this method is referred to Kug et al. (2007a). The NCEP GFS

was forced by the forecast SST from its coupled version CFS. All models used the same initial conditions from NCEP/DOE Reanalysis-2 and have 10 ensemble members except SNU GCPS, LASG/IAP GAMIL and NCEP GFS, which have 6, 6, and 15 members, respectively.

Table 3 Description of APCC/CliPAS one-tier prediction models

Institute	AGCM	OGCM	Ensemble member	Period and forecast lead time	Reference
BMRC	BAM 3.0d	ACOM2	10	1980–2002	Zhong et al. (2005)
POAMA1.5	T47 L17	0.5–1.5°lat × 2°lon L31		9 months	
FRCGC	ECHAM4	OPA 8.2	9	1982–2004	Luo et al. (2005)
SINTEX-F	T106 L19	2°cos(lat) × 2°lon L31		6 months	
GFDL	AM2.1 2°lat × 2.5°lon L24	OM3.1 (MOM4) 1/3°lat × 1°lon L50	10	1979–2005 12 months	Delworth et al. (2006)
NASA	NSIPP1 2°lat × 2.5°lon L34	Poseidon V4 1/3°lat × 5/8°lon L27	3	1980–2004 5 months	Vintzileos et al. (2003)
NCEP	GFS	MOM3	15	1981–2004	Saha et al. (2006)
CFS	T62 L64	1/3°lat × 1°lon L40		9 months	
SNU	SNU T42 L21	MOM2.2 1/3°lat × 1°lon L32	6	1980–2001 6 months	Kug et al. (2007b)
UH	ECHAM4 T31 L19	UH Ocean 1°lat × 2°lon L2	10	1982–2003 12 months	Fu and Wang (2004)

All seven coupled models (one-tier systems) do not apply any flux correction. Three coupled models (BMRC, POAMA1.5, NCEP CFS, and GFDL CM2.1) use ocean data assimilation for initialization; other coupled models use either a SST nudging scheme (FRONTIER SINTEX-F, NASA, SNU) or a SST and thermocline-depth nudging scheme (UH). No coupled model has land initialization schemes. The NCEP CFS and SNU models use the NCEP reanalysis data as land surface initial conditions, and other coupled models use climatological land surface condition as the initial condition.

Persistent forecast was performed as a benchmark on predicting SST and 2 m air temperature since a skillful forecast must be significantly superior to the persistent forecast. In the case of May (November) initial condition, the observed April (October) anomaly was used as an anomaly persistent forecast for all forecast lead time.

The observed data used for verification were obtained from the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) data set (Xie and Arkin 1997) and from the NCEP/DOE (department of Energy) reanalysis data (Kanamitsu et al. 2002) for atmospheric variables, and from the improved Extended Reconstructed Sea Surface Temperature Version 2 (ERSST V2) data (Smith and Reynolds 2004) for SST.

2.2 Forecast quality measures

To measure the forecast quality of a deterministic forecast, MME prediction was made using simple average of 14 models' ensemble means. For probabilistic forecast, we selected ten models that have nine or more ensemble members. Forecast probabilities were derived from simple

democratic counting using 109 individual realizations from the ten models (5 one-tier and 5 two-tier models) after normalizing each simulation with respect to its own mean and standard deviation. All skill measures for deterministic and probabilistic prediction were cross-validated (refer to Appendix A in Saha et al. 2006).

The metrics used to measure prediction skill of MME mean forecast includes the anomaly pattern correlation coefficient (PCC) and the root mean square error (RMSE) normalized by the corresponding observed standard deviation. A temporal correlation coefficient (TCC) was used for a specific time series of a predictand. For convenience of comparison, we also calculated the time-averaged anomaly PCC and RMSE over the global tropics (30°S – 30°N , 0 – 360°E) and its sub-domains for 23 years (1981–2003). To make unbiased estimates of the mean PCC and RMSE, we firstly averaged quadratic measures, such as variance, covariance, and mean square error and then calculated PCC and RMSE. A similar way was taken to calculate the zonal mean and area mean of TCC.

The probabilistic forecast skills were evaluated using the Brier Skill Score (BSS) and the Relative Operating Characteristic (ROC; Mason 1982; Wilks 1995; Richardson 2000; Zhu et al. 2002). The definition of BSS is based on the Brier Score (BS), which is a scalar measure of the accuracy of a probabilistic forecast of a dichotomous event and is defined by:

$$\text{BS} = \frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2,$$

where n is the number of forecasts, f_i the forecast probability of occurrence for the i th forecast, and o_i is the i th observed probability, which is defined to be 1 if the

event occurs and 0 otherwise. The BS can be decomposed as three terms related to uncertainty, reliability, and resolution as follows (Wilks 1995):

$$\begin{aligned} \text{BS} &= \bar{o}(1 - \bar{o}) + \frac{1}{n} \sum_{k=1}^m n_k(f_k - \bar{o}_k)^2 - \frac{1}{n} \sum_{k=1}^m n_k(\bar{o}_k - \bar{o})^2, \\ &= \text{BS}_{\text{unc}} + \text{BS}_{\text{rel}} - \text{BS}_{\text{res}}, \end{aligned}$$

where \bar{o} indicates the climatological probability of the event, m indicates the number of probability bins, f_k represents the forecast probability for bin k , and \bar{o}_k denotes the relative frequency of occurrence of the event when the forecast probability is f_k . Brier score of a climatological forecast $\text{BS}_{\text{clim}} = \text{BS}_{\text{unc}}$. Then, the Brier Skill Score is

$$\text{BSS} = 1 - \frac{\text{BS}}{\text{BS}_{\text{clim}}} = \frac{\text{BS}_{\text{res}} - \text{BS}_{\text{rel}}}{\text{BS}_{\text{unc}}}.$$

Thus, for a climatological prediction the BSS is 0. In the present study, a probabilistic forecast of an event for each tercile category was performed. The three categories are “below-normal”, “normal”, and “above-normal” based on climatological terciles.

Another measure of probabilistic forecast is the area under the ROC curve (AROC), which is calculated by integrating the area beneath the ROC curve in the graph of hit rates against false-alarm rates within a range of probability threshold (Green and Swets 1966). It is equal to the unit for a perfect deterministic forecast, while it is equal to 0.5 for a no skill forecast in which the hit rate and false-alarm rate are equal.

3 SST forecast quality

In this section, the current status of SST forecast is assessed using seven coupled models in APCC/CliPAS project and their MME prediction. In some occasions, 14 coupled model MME is examined, which includes seven DEMETER and seven CliPAS coupled models.

3.1 Equatorial SST

The SST anomalies along the equator are of central importance for determining tropical and global teleconnection, and thus, they have been recognized as a major source of global atmospheric climate variability. An effort is made here to evaluate the coupled model performance in forecast of equatorial SST anomalies averaged between 10°S and 5°N where SST variability is largest. To facilitate evaluation, an Empirical Orthogonal Function (EOF) analysis of the seasonal mean SST anomalies along the global equatorial region was performed during the common hindcast period of 1983–2002. The observed first mode

accounts for 77% (62%), while the second mode accounts for 12% (17%) of the total variance during the DJF (JJA) season. Thus, the first two modes account for 89% (79%) of the total variance during the DJF (JJA).

To what extent can the seven one-tier CGCM MME 1-month lead hindcasts capture these two leading modes? We found that the hindcasts replicate realistic spatial-temporal structures of the first two leading modes for both the JJA and DJF seasons, and especially for the DJF season (Figs. 1, 2). For the DJF season, the observed first mode represents equatorial SST anomalies (SSTA) associated with the mature phase of ENSO: The maximum positive anomaly is located in the Niño 3.4 region with a complementary negative anomaly in the western Pacific and a weak positive anomaly in the Indian Ocean (Fig. 1a). The longitudinal distribution of SSTA in the Indo-Pacific Oceans, along with its temporal evolution, is captured very well by the MME forecast except for a slight westward shift of the SSTA in the western Pacific. The observed second mode shows a maximum SSTA located near the dateline with negative SSTAs in both the eastern Pacific (east of 130°W) and far western Pacific-eastern maritime continent (100°E–150°E), which resembles a so-called “Modoki” mode (Weng et al. 2006) (Fig. 1b). The observed second principal component shows variability on decadal time scale: a negative phase in the prolonged cold events during the period of 1998–2001 and a positive phase around the lasting warm event in the early 1990s (1990–1994). The CGCM MME captured the major features of the second mode in spatial pattern reasonably well, but the spatial phase has a significant westward shift in the western Pacific by about 15° of longitude.

During the northern summer (JJA), the observed leading mode exhibits a peak SSTA in the Niño 3 region, rather than the Niño 3.4 region (Fig. 2a). An accompanied SSTA dipole pattern occurs with a negative SSTA in the western Pacific and a positive SSTA in the western Indian Ocean. The eastern Indian Ocean SST is nearly normal. The first principal component suggests that this mode most often occurs in the developing phase of El Niño/La Niña but sometimes also occurs in the decaying phase of El Niño (in 1983, for instance) or during prolonged warm or cold events. The observed second mode shows a wide peak in the Niño 3.4 region with complementary cooling in both the eastern and western Pacific (Fig. 2b). The MME prediction captures the gross longitudinal distribution, but an evident westward shift of the SSTA exists in the western Pacific between 120°E and 180°E for both modes. Large errors are seen over the Indian Ocean in JJA for the second mode, indicating that the MME model has difficulty in predicting JJA Indian Ocean SSTA in the years of a decaying strong El Niño (e.g., 1983, 1998).

Fig. 1 Spatial patterns (*upper panels*) and principal components (*lower panels*) of the first (**a**) and second (**b**) empirical orthogonal mode of the equatorial SST variations in DJF, derived from observation, the CliPAS coupled model MME prediction, the SNU dynamic-statistical SST prediction, the persistence forecast and the individual coupled models (*dashed lines*)

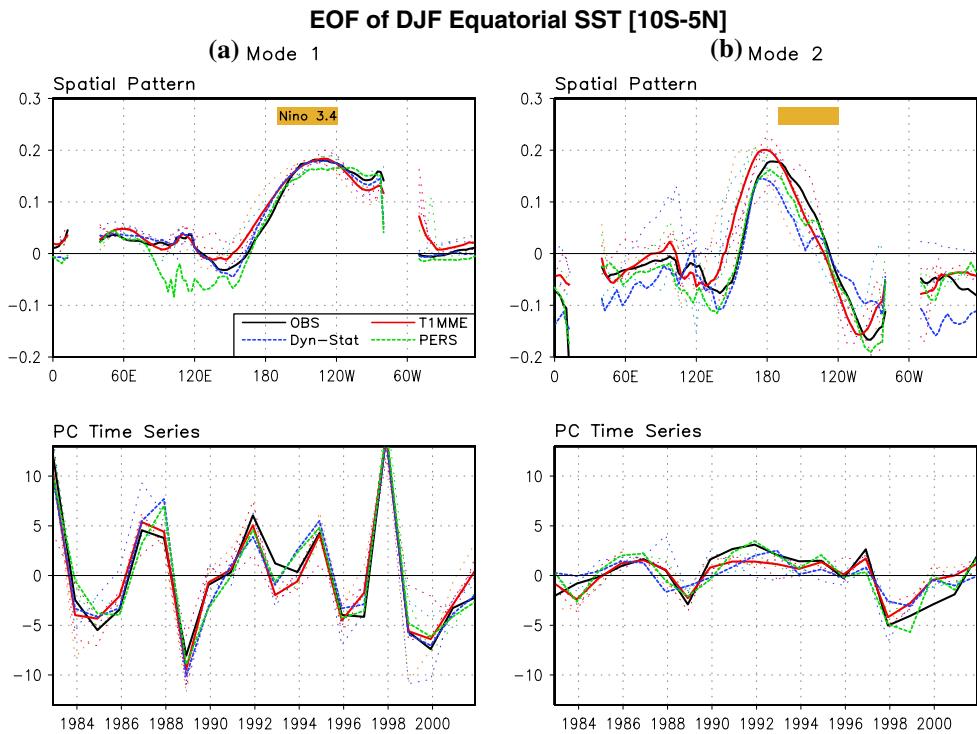
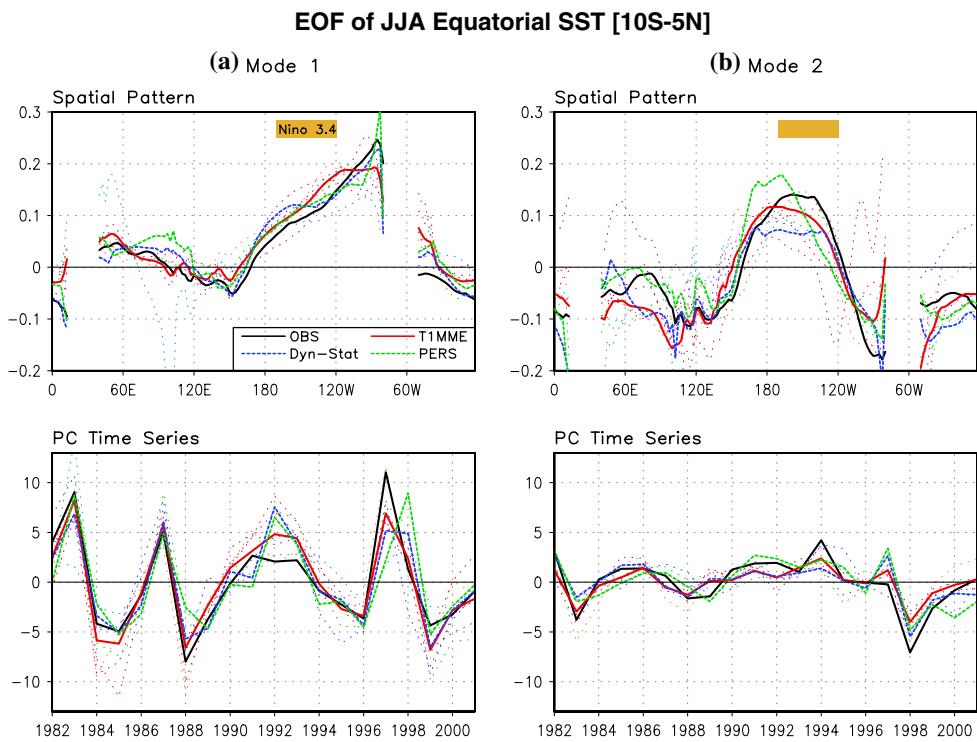


Fig. 2 Same as in Fig. 1 except for JJA



Although the individual models show a large spread in skill, especially for longitudinal distribution, the current coupled MME prediction generally outperforms the dynamic-statistical prediction for the two leading modes of SST variability except in the mature phase of ENSO. In particular, the current MME captures the temporal

variation of the two leading modes realistically. The temporal correlation skill of PC time series of the coupled MME prediction (dynamical-statistical model) is 0.94 (0.87) for the first mode and 0.90 (0.78) for the second mode in JJA. However, the spatial shift of the MME prediction could potentially cause errors in the global

teleconnection that is associated with equatorial SSTA, degrading seasonal climate prediction skills over both the tropics and extratropics. Over the Atlantic, the observed SSTA for all modes is uniformly negative except for the first mode in DJF, for which SST is normal. The predicted SSTA in the Atlantic is generally good, but large errors are seen off the coast of Brazil (Figs. 1, 2).

In terms of the fractional variance, the hindcast leading mode of MME accounts for a larger percentage of the total variance, i.e., 88% (83%), for DJF (JJA), while the second mode accounts for a smaller fractional variance, 8% (9%) for DJF (JJA) compared to the corresponding observational counterparts. Most of individual models have more realistic values of the fractional variance although the model spread is large. The range of the percentage variances for the individual coupled models is 77–87% (55–85%) in DJF (JJA) for the first mode and 6–13% (5–18%) in DJF (JJA) for the second mode.

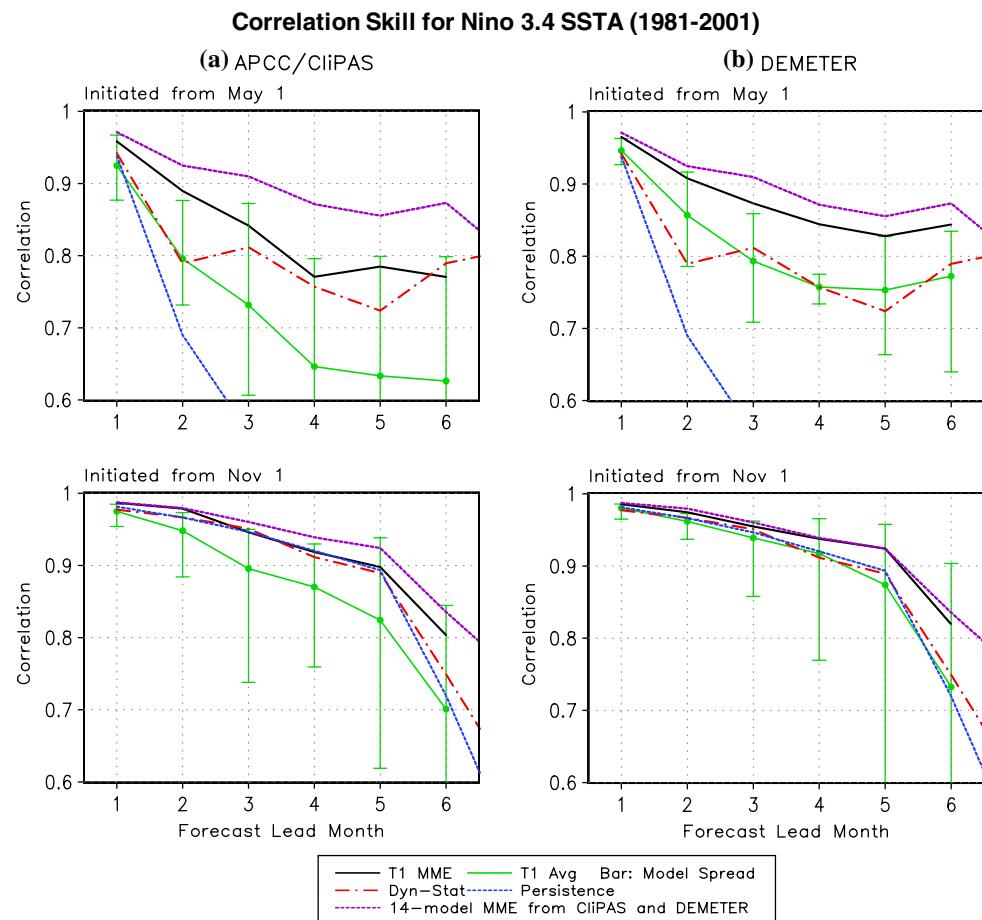
The fractional variance accounted for by the first two modes of MME is 96% (92%) for the DJF (JJA) season. This suggests that the MME hindcast tends to overestimate the variance contribution of the first mode and is unable to capture the third and higher modes of variability. The

reason is partially due to the effect of the multi-model ensemble mean, which tends to suppress the higher modes that may be more associated with stochastic processes.

3.2 ENSO

ENSO forecast is at the heart of the seasonal prediction. Here the Niño 3.4 SST anomaly (averaged over the region 5°S–5°N, 120°W–170°W) is used as an ENSO index for gauging prediction skill. Figure 3 shows TCC skill of MME prediction of the Niño 3.4 SST anomaly for 21 years of 1981–2001 as a function of forecast lead time, initiated from May 1 and November 1, respectively. The average skill of the seven individual coupled models and the range of the models' spread are also presented. Here the range of spread is denoted by the distance from the worst to the best model skill. For comparison, Fig. 3b shows the counterparts made by seven different DEMETER coupled models for the same period. The correlation skill of the CliPAS MME forecast at a 6-month lead reaches 0.77 and 0.81 for predictions starting from 1 May and 1 November, respectively. These skills are slightly lower than the DEMETER MME prediction (Palmer et al. 2004). The reason is that the

Fig. 3 **a** Temporal correlation skill of prediction of Niño3.4 SST index as a function of forecast lead time, initiated from 1 May and 1 November for the period of 1981–2001 derived from 7 CliPAS coupled models. **b** The same as in **a** but derived from 7 DEMETER coupled models. The green lines indicate the averaged skill of the individual models and the bars show the range of the best and worst coupled model skills. For comparison, the skills of persistent forecast (blue), the SNU dynamic-statistical model forecast (red), and the 14-model MME from both DEMETER and CliPAS (purple) are also shown



averaged skill of the individual models in DEMETER is better than CliPAS. If both DEMETER and CliPAS models were used, the 6-month lead forecast skill reaches 0.86 for the average of all May and November initial conditions. Using CliPAS and DEMETER coupled models, Jin et al. (2008) found that the forecast skill depends strongly on season, ENSO phases, and ENSO intensity. A stronger El Niño or La Niña is more predictable and ENSO-neutral years are far less predictable than warm and cold events. Given that only about a half a dozen warm and cold events occurred during the hindcast period, the conclusions that have been drawn here need verification using a longer hindcast record.

3.3 Indian Ocean SST

Recent studies have reported that the Indian Ocean Dipole (IOD) mode has a considerable influence on climate variability in the surrounding continental regions including South Asia, eastern Africa, Australia, and even East Asia (Guan and Yamagata 2003; Saji and Yamagata 2003). Although some IOD events concur with ENSO events, the IOD events can be independent of ENSO, because IOD development depends on Indian Ocean air-sea interaction, including the Bjerknes (1969) feedback in the equatorial Indian Ocean (Saji et al. 1999; Webster et al. 1999; Saji and Yamagata 2003) and the off-equatorial moist Rossby wave-SST dipole interaction (Wang et al. 2000, 2003). An effort is made to assess the performance of coupled models on the prediction of Indian Ocean SST.

Figure 4a and b show the TCC skill as a function of the forecast lead month for the West Indian Ocean (WIO; 50°E–70°E, 10°S–10°N) and the East Indian Ocean (EIO; 90°–110°E, 10°S–0) SSTAs, respectively. For the EIO SST prediction, the correlation skill reaches 0.68 at a 6-month lead forecast for both initial forecast times (1 May and 1 November). But there is a considerable dip in skill in July for the forecast initiated from early May and a dip in December for the forecast initiated from early November. For the western Indian Ocean, the MME skill is 0.80 for a 6-month lead forecast with November initial condition, but only 0.43 for the 5-month lead forecast with May initial condition. The MME initiated from 1 November has generally better skill than the forecasts initiated from 1 May for both the EIO and WIO. These skills surpass the corresponding persistent skill and statistical-dynamical forecast, especially for predictions with November initial condition.

While the SST predictions in the WIO and EIO show some useful skills, the skill for prediction of the IOD SST index (SST at EIO minus SST at WIO) is reduced (Fig. 4c). The TCC skill of IOD forecast drops below 0.4 at a 3-month lead forecast for both May and November initial

conditions. The results indicate the existence of a July prediction barrier and a severe, unrecoverable January prediction barrier for the IOD index prediction. The winter barrier for prediction of IOD and the EIO SSTAs is related to its strong phase lock to the annual reversal of the monsoon (Luo et al. 2007). For the forecasts started from early May, while a July barrier exists, there is a robust bounce-back after July, suggesting that the mature phase of IOD in October–November is more predictable (if it starts from early May) probably due to the predictability of the EIO pole where the SST dominates the mature phase of IOD. We suggest that the EIO SSTAs should be predicted rather than IOD index.

In summary, the results in Figs. 1 through 4 indicate that (1) significant improvement of forecast skill can be obtained through MME approach, and (2) the seven coupled models' MME prediction skill outperforms the SNU statistical-dynamical model and is far better than persistence forecast.

4 Spatial and seasonal dependence of prediction skills

In this section, we examine the MME skills in the 1-month lead seasonal prediction of 2 m air temperature, precipitation, and three-dimensional circulation fields. The seven one-tier and seven two-tier predictions are all used to make MME prediction. The first two variables have been chosen because they are important surface variables. In order to understand the transition of predictable signal between the tropical phenomena into global fields, we further evaluate forecast skills of stream function at 850 and 200 hPa, and geopotential height at 500 hPa.

4.1 Two-meter air temperature

Figure 5a and b (upper panels) show the spatial distribution of the 1-month lead seasonal forecast skills for 2 m air temperature in terms of TCC at each grid point for the period of 1981–2003. The significance of temporal correlation coefficient was tested using *t* test. The region of statistical significant TCC at the confidence level of 0.05 and 0.01 are outlined. Since the near surface air temperature over the ocean has well known persistence due to the influence of the underlying SST, a skillful forecast must be significantly superior to the persistence forecast. For this reason, the differences between the MME prediction and persistence forecast are shown in lower panels of Fig. 5. During the DJF season, useful skill is seen in a horseshoe region extending from the Maritime continent and the eastern Indian Ocean toward the extratropics in the north-eastern and southeastern Pacific, and also in North America, Europe, the Middle East, southern Africa, and

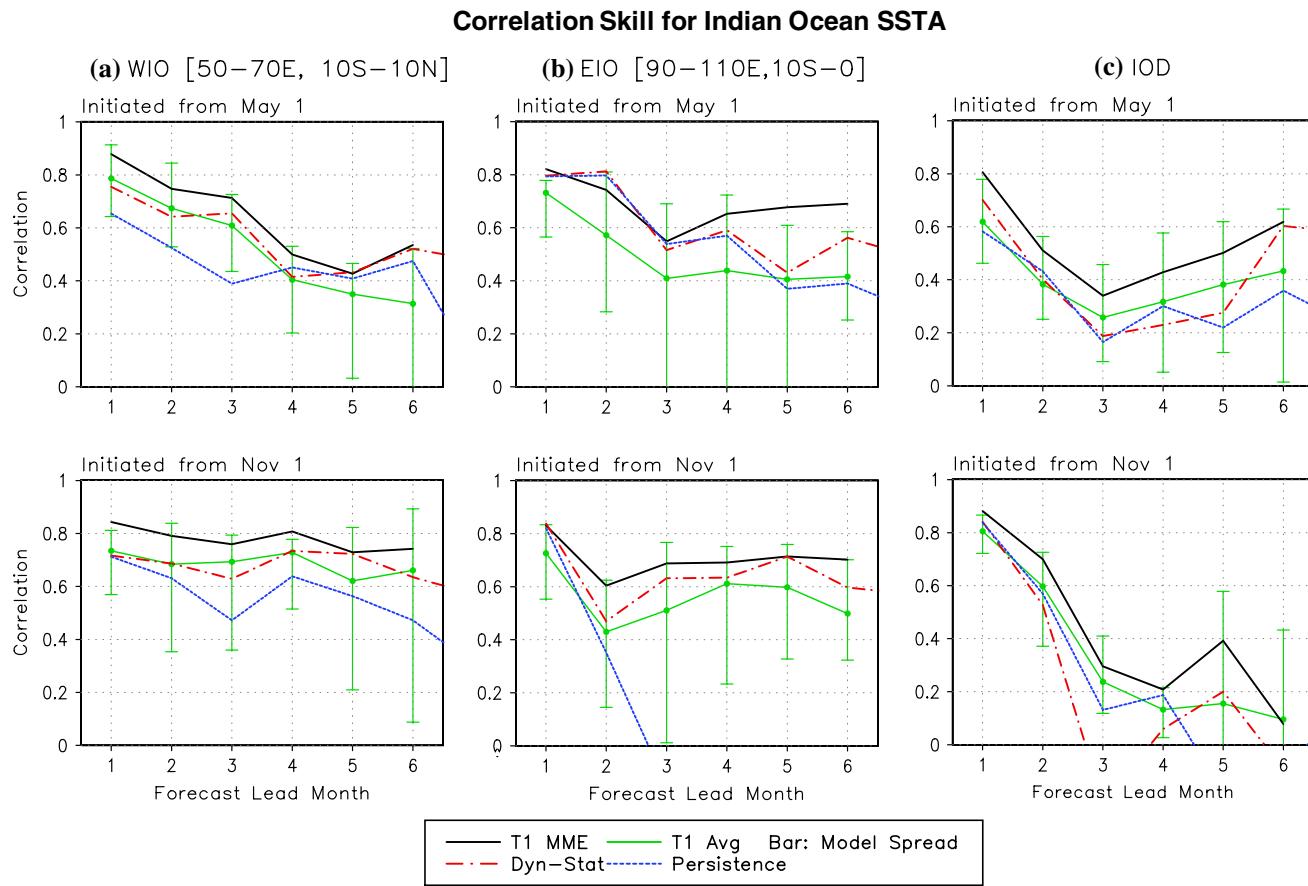


Fig. 4 **a** Same as in Fig. 3a except for the West Indian Ocean (WIO, 10S–10N, 50–70E) SST anomaly. **b** Same as in **a** except for the East Indian Ocean (EIO, 10S–0, 90–110E) SSTA. **c** Same as in **a** except for

the Indian Ocean Dipole (IOD) SST index, which is difference between WIO and EIO SSTA

Southeast Asia. Most of these additional skills originate from the influence of ENSO via teleconnection during its peak phases. During JJA, the MME temperature prediction outperforms persistence over the northern Indian Ocean, northeastern Asia, and the tropical eastern North Pacific–Caribbean Sea. Note, however, in many regions of the world, MME prediction may not necessarily be better than persistence, suggesting that persistence forecast can be a complementary tool for JJA temperature prediction. In particular, persistence is relatively strong after 1998 (not shown).

In the tropical Pacific between 10°N and 20°N (north of the ITCZ), the JJA skill is higher than the DJF skill (upper panels of Fig. 5). But, this higher skill is mainly due to persistence (lower panels of Fig. 5), suggesting that as the thermal equator moves northward in JJA, the northern hemisphere's tropics are more predictable than in the boreal winter due to SST persistence. In the Indian Ocean, the SST is better predicted in the local summer than in the local winter season. But, in the South Pacific Convergence Zone, the local winter prediction is better than prediction in summer.

4.2 Precipitation

Figure 6a and b show the geographic distribution of the TCC skill for 1-month lead seasonal precipitation prediction. High skill (0.5–0.7) in both JJA and DJF is observed over the tropical Pacific and Atlantic between 10°S and 20°N, the Maritime continent, northeastern Brazil, and the subtropical South Pacific Convergence zone.

During DJF, the high skill regions expand in the global tropics and subtropics and over both the ocean and land, particularly in the following regions: (a) the subtropical North Pacific and North Atlantic between 20°N and 40°N; (b) the equatorial Indian Ocean and the east coast of equatorial Africa; (c) tropical South America, (d) southern subtropical Africa; (e) Mexico and the southern United States, and (f) Southeast Asia. The overall expansion of the good forecast skill regions is a result of the model's capacity to capture the ENSO teleconnection. During DJF, ENSO events mature and exert robust influence on remote regions through atmospheric teleconnection. While DJF prediction is generally better than JJA prediction, there are exceptions. The most evident exceptions are seen over the

Fig. 5 Temporal correlation coefficients for 2 m air temperature between observation and 1-month lead seasonal prediction for 1981–2003 obtained from 14 CliPAS models' MME system in **a** JJA and **b** DJF, respectively. The lower panels indicate the skill difference between MME prediction and persistence for each season. The thin (thick) solid contours represent statistically significance of the correlation coefficients at 0.05 (0.01) confidence level. One-month lead persistence was obtained from the observed anomalies in April for JJA forecast and those in October for DJF forecast, respectively

APCC/CliPAS MME Skill for 2m Air Temperature (1981-2003)

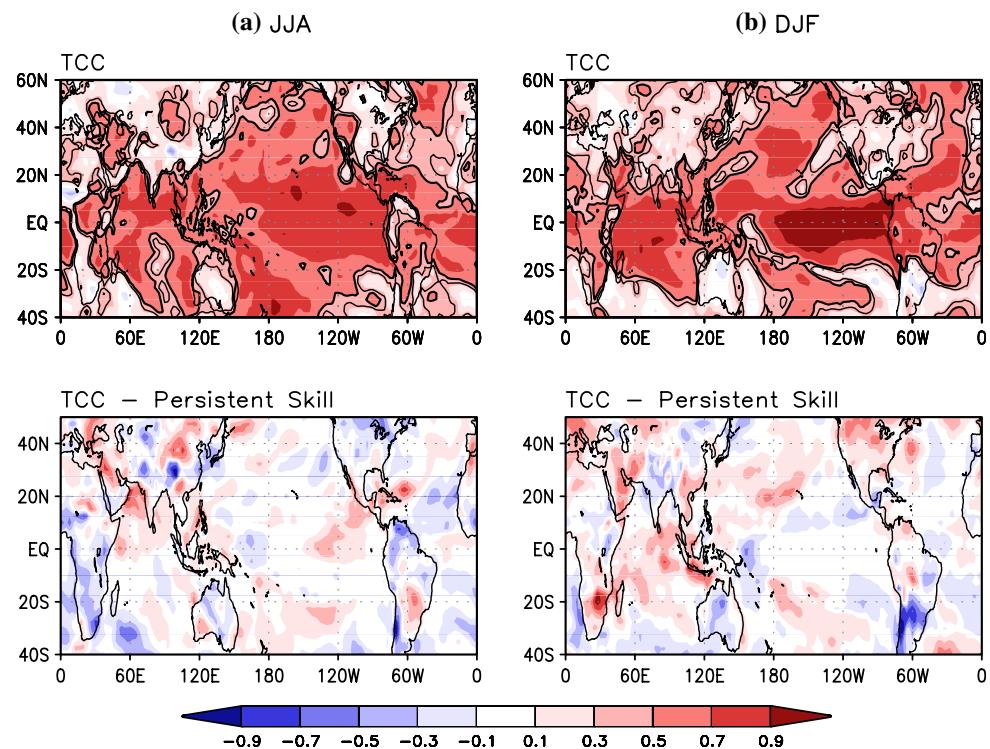
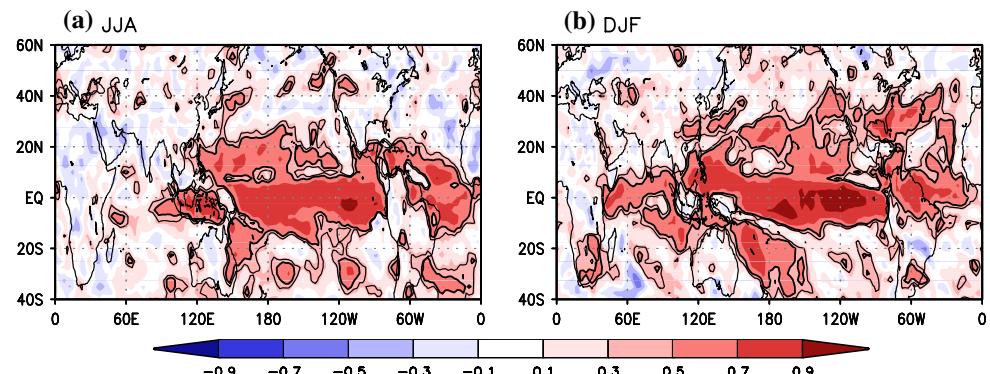


Fig. 6 Temporal correlation coefficients for precipitation between observation and 1-month lead seasonal prediction for 1981–2003 obtained from 14 APCC/CliPAS models' MME system in **a** JJA and **b** DJF, respectively. The thin (thick) solid contours represent statistical significance of the correlation coefficients at 0.05 (0.01) confidence level

APCC/CliPAS MME Skill for Precipitation (1981-2003)



subtropical southeastern South America and the southwestern Atlantic Ocean (20°S – 35°S , 10°W – 50°W), and over the eastern Australia. These areas are under local winter monsoon regimes.

Forecast skill (as well as predictability) is a function of latitude and longitude. The zonally averaged TCC skill shows that the highest mean skill over the equator exceeds 0.6 at 5°S in JJA and reaches 0.7 at 5°N in DJF (Fig. 7a). The prediction of DJF precipitation is superior to that of JJA precipitation primarily in the northern hemisphere between the equator and 40°N , and especially between 20°N and 40°N . However, in the southern hemisphere, precipitation prediction for JJA is better than that for DJF

south of 30°S . We suggest that during the ENSO-developing phases the teleconnection associated with convective anomalies over the Maritime continent enhances the Austral winter teleconnection in the southern subtropics, thus increasing the prediction skills. The message from Fig. 7a is that precipitation prediction in the local winter is better than in the local summer in both the southern and northern hemispheres. Figure 7b shows how the correlation skills averaged over the tropics (30°S – 30°N) for precipitation in both JJA and DJF decrease moving away from the El Niño/La Niña region. The highest mean skill exceeding 0.6 is found near the dateline from 150°E to 170°W in JJA and from 150°E to 140°W in

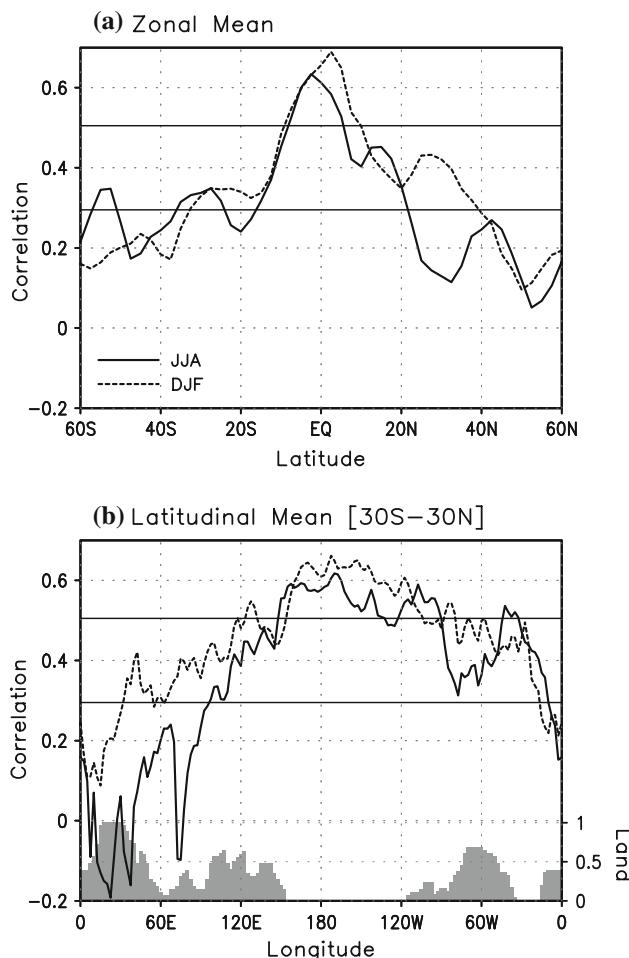


Fig. 7 **a** Zonal mean temporal correlation skill of precipitation predicted by APCC/CliPAS MME system in JJA (solid line) and DJF (dashed line), respectively. **b** Same as in **a** except for latitudinal mean temporal correlation skill between 30S and 30N. The light shaded bar indicates the fraction of land between 30S–30N at each longitude

DJF. The lowest skill is found over tropical Africa. The DJF prediction skill is considerably higher than prediction skill in JJA mainly in the Asian–Australian monsoon sector from 40°E to 140°E and over the tropical American sector between 60°W and 90°W. Skill over land regions is generally lacking except in some specific regions during DJF.

4.3 Atmospheric circulation fields

In general, prediction for atmospheric circulation fields shows higher skill than that for temperature and precipitation. One-month lead seasonal prediction of the 850 hPa streamfunction field shows high skill over the western Pacific and Asian continents in JJA, and in the eastern Pacific (east of 180°E) and North America, as well as over the maritime continent in DJF (Fig. 8). Prediction of the 200 hPa streamfunction shows good TCC skill almost

everywhere between 40°S and 60°N except in the equatorial region. High prediction skill for the 500 hPa geopotential height is confined to the global tropics with a meridional seasonal migration of the high skill region. Note also that DJF skill is considerably higher than JJA skill for all of the variables at the three levels.

The season-dependence and spatial patterns of the circulation forecast skills can be reasonably explained in terms of ENSO impact (Kumar and Hoerling 2003). Skill tends to increase from JJA to DJF because ENSO forcing increases from JJA to DJF. The remarkable eastward shift of the high skill region in 850 hPa rotational flow anomalies from JJA to DJF is attributed to the eastward shift in the teleconnection pattern associated with ENSO-induced maximum equatorial convective anomalies from the developing (JJA) to mature (DJF) phases of ENSO. In the decaying ENSO phase (also in JJA), the 850 hPa rotational flow remains strong in the eastern hemisphere mainly due to local warm pool-atmosphere interaction (Wang et al. 2000; Lau and Wang 2006). The off-equatorial 200 hPa streamfunction is the atmospheric Rossby wave response and teleconnection to the equatorial dipole heat source/sink associated with El Niño and La Niña, which have a wide meridional scale and global zonal scale. During El Niño, the entire tropics warm up due to rapid propagation of the equatorial Kelvin and Rossby waves excited in the eastern Pacific warming. The opposite is true during La Niña. As a result, the 500 hPa geopotential height rises up in El Niño and decreases during La Niña in accord with temperature changes. Thus, nearly all of the high skill regions are due to the influence of ENSO teleconnection through atmospheric internal dynamics.

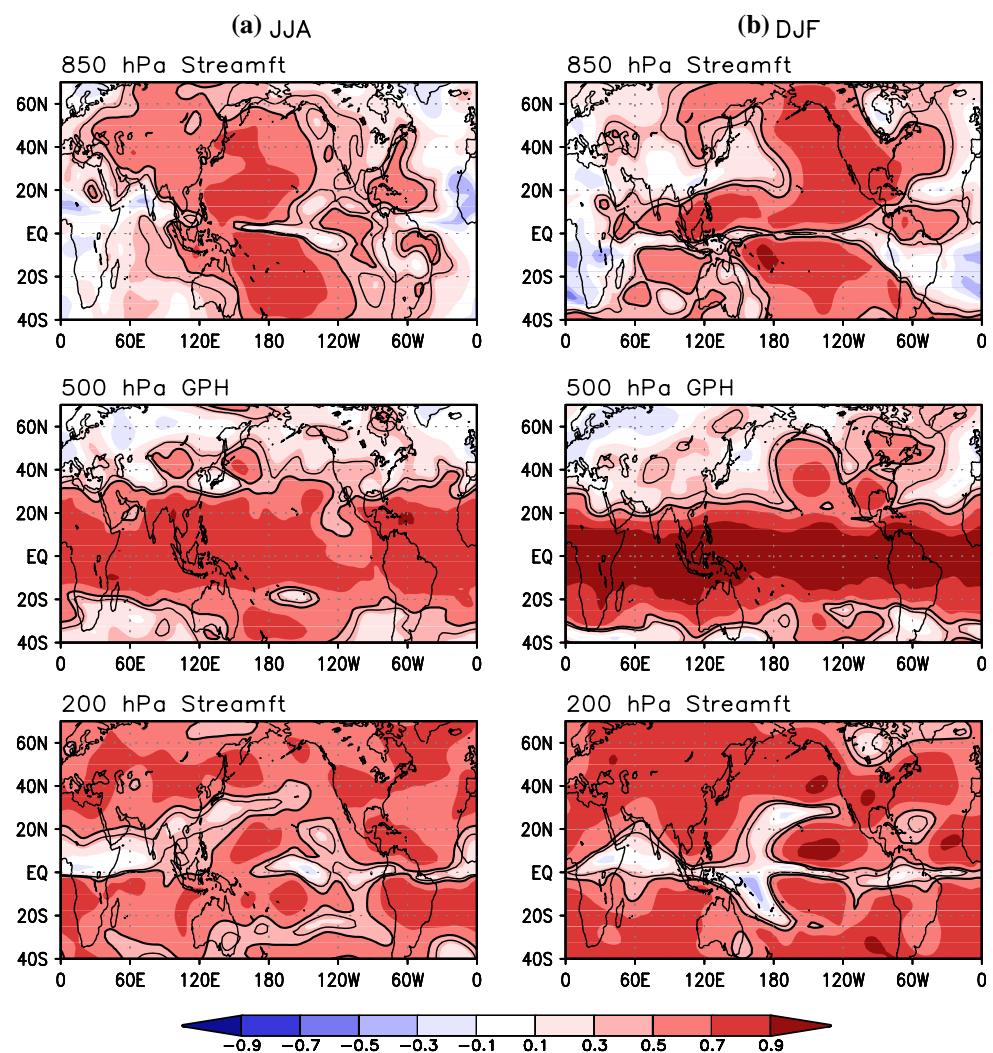
In summary, the prediction skills vary with location and season. The variations in the spatial patterns and the seasonality of the correlation skills suggest that ENSO variability is the primarily source of global seasonal prediction skill. Winter monsoon precipitation in both hemispheres is more predictable due to teleconnection associated with ENSO. Precipitation predictions over land and the local summer monsoon region show little skill.

5 Interannual variation of MME skill

5.1 Dependence on ENSO amplitude and season

Figure 9 shows that the anomaly PCC between the observed 2 m air temperature and the MME's 1-month lead temperature prediction in the global tropics (30°S–30°N) varies from year-to-year and ranges from 0.25 to 0.70 in JJA and from 0.20 to 0.80 in DJF. Surprisingly, the time-mean PCC score during 24 years is better for JJA (0.53) than for DJF (0.50). The DJF correlation skill decays more

Fig. 8 Temporal correlation coefficients for 850 hPa (*upper panels*) and 200 hPa (*lower panels*) streamfunction and 500 hPa geopotential height (*middle panels*) between observation and 1-month lead seasonal prediction obtained from APCC/CliPAS MME system in **a** JJA and **b** DJF seasons, respectively. The thin (thick) solid contours represent statistical significance of the correlation coefficients at 0.05 (0.01) confidence level



sharply poleward away from the equator than during JJA, although DJF skill is higher than JJA skill over the equatorial band between 15°S and 10°N. The reason is that as the thermal equator moves northward in JJA, the northern hemisphere tropics are more predictable than the boreal winter due to SST persistence. Note that the range of interannual variation in the PCC score in DJF (0.18–0.80) is larger than that in JJA (0.25–0.70). In DJF of 1982/1983 and 1997/1998, the correlation skill reaches about 0.80, while in DJF of 1981/1982 and 1989/1990, the skill is only around 0.20.

Obviously, the year-to-year variation in overall skill depends on ENSO variability. The MME PCC skill has a clear relationship with the amplitude of Niño 3.4 SST variation especially in the boreal winter with a correlation coefficient of 0.76. The worst years tend to occur during transition or normal ENSO phases. The major failures in the prediction of DJF temperature occurred in the 1980s (in the northern winters of 1981/1982 and 1989/1990) and were either 1-year before the mature

phase of El Niño or 1-year after the mature phase of a La Niña.

In contrast to the surface air temperature skill, the tropical mean anomaly PCC score for precipitation prediction is higher in DJF (0.57) than in JJA (0.46) over the global tropics (Fig. 10). Similar to the air temperature, the range of interannual skill variation in DJF (−0.2 to 0.8) is much larger than in JJA (0.2–0.7). The DJF of 1989/1990 shows extremely low skill of −0.2, which deserves a special case study. The MME anomaly PCC of precipitation also shows strong relationship with the amplitude of Niño 3.4 SST variation especially in the boreal winter with a correlation coefficient 0.75.

Figures 9 and 10 indicate that although the MME prediction does not necessarily outperform the best model during each individual year, the MME's overall skill is superior to any individual model in terms of the time-averaged PCC score for all years. There is no best model that is always better than the other models in every year studied.

Fig. 9 Time series of anomaly pattern correlation coefficients in 2 m air temperature between observation and 1-month lead MME prediction over the global tropics in **a** JJA and **b** DJF, respectively. The values for individual model ensemble predictions are also plotted with grey square marks. The dashed line indicates the amplitude of Niño 3.4 SST anomaly

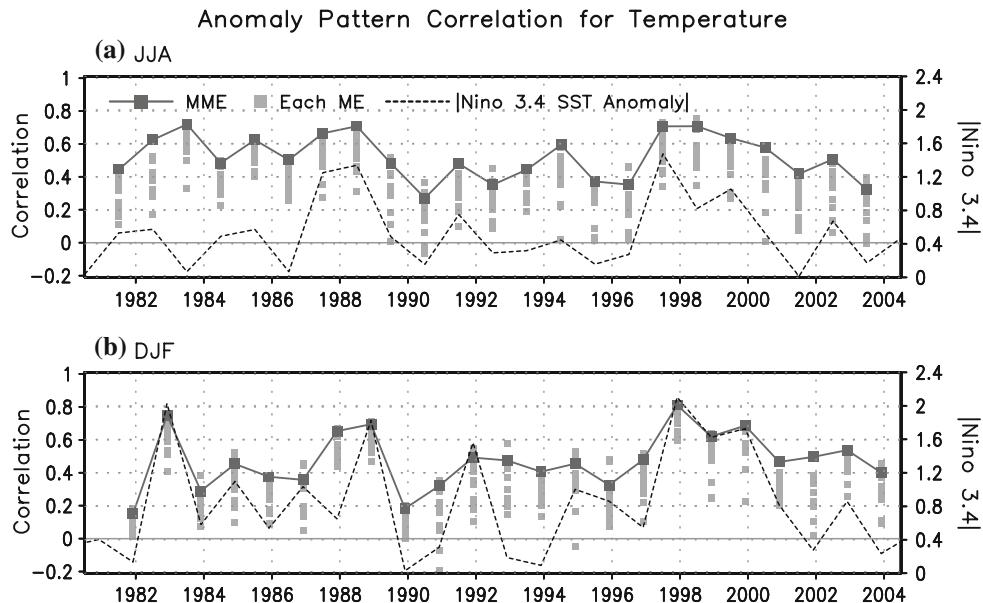
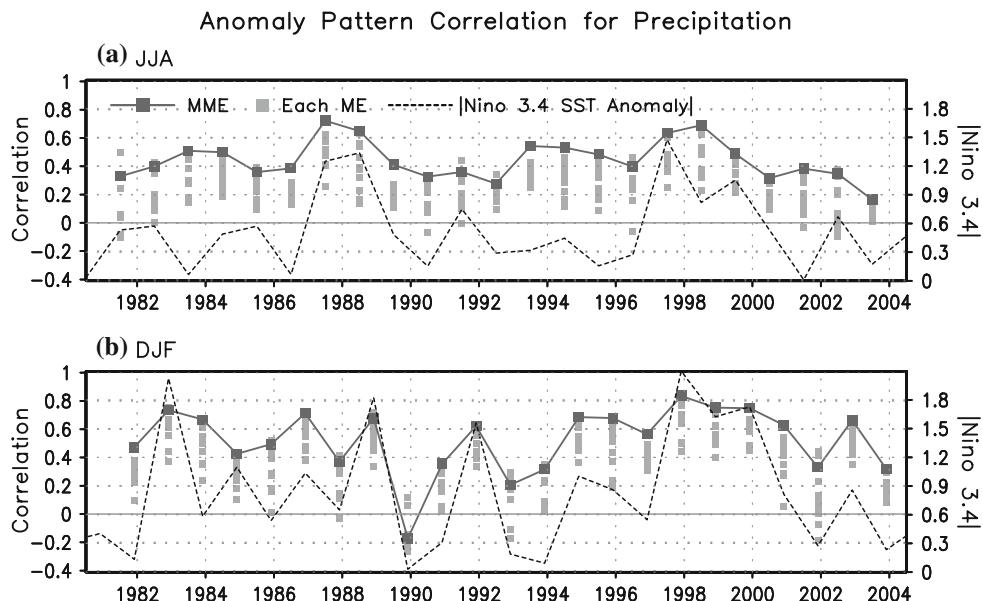


Fig. 10 Same as in Fig. 9 except for precipitation



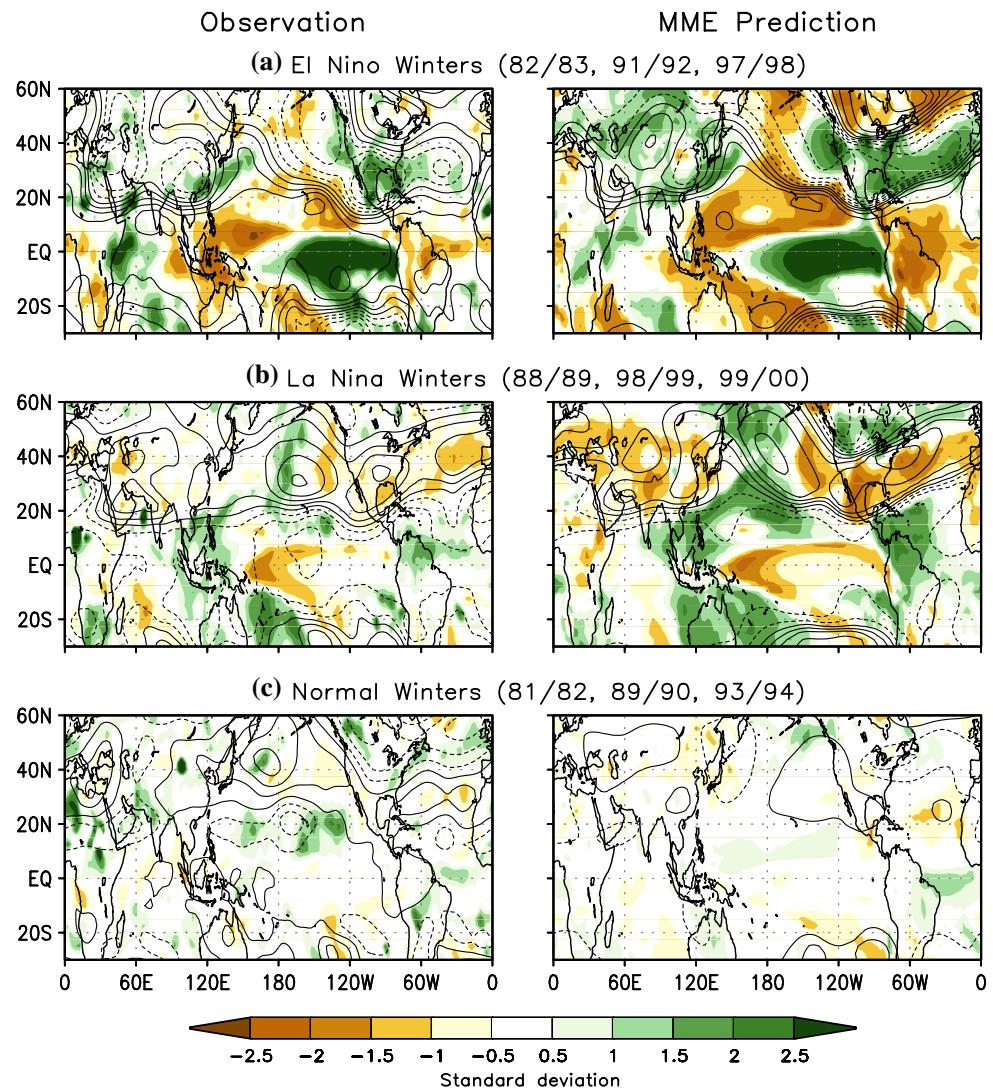
5.2 Asymmetries with respect to El Niño and La Niña and development and decay phases

The observed anomalies associated with ENSO's mature phases are not a mirror image when comparing El Niño and La Niña (e.g., Hoerling et al. 1997). In the Asian–Australian monsoon region, the atmospheric responses to a developing and a decaying ENSO event are also nearly out of phase (Wang et al. 2001). So, do MMEs capture those features faithfully?

Figure 11 shows composite maps of precipitation and 500 hPa geopotential height anomalies normalized by their standard deviation. The composites were made by using three El Niño, three La Niña, and three normal DJF in

observation and in 1-month lead seasonal MME prediction. The mean SST amplitude for El Niño (La Niña) composite is 1.9 (1.7) degree and the anomaly PCC of precipitation prediction are 0.73 (0.74) over the global Tropics, respectively. For normal years, the predicted anomalies are weaker than observed and the anomaly PCC is only 0.18, suggesting that without ENSO forcing the MME does not have useful skill. For both El Niño and La Niña events, the predicted normalized anomalies of precipitation and geopotential height agree well with observed anomalies, especially over the tropics and the western Hemisphere. Significant errors are found over Eurasian continent in both precipitation and circulation. However, the prediction tends to overestimate anomalies over most of the regions. Observation shows

Fig. 11 Precipitation (shaded) and 500 hPa geopotential height (contoured) anomalies composit for three El Niño (upper panels), three La Niña (middle panels) and three normal (lower panels) boreal winters. The left and right panels are made from observation and 1-month lead CliPAS MME prediction, respectively. All anomalies were normalized by their own standard deviations. The contour interval is 0.5



rather asymmetric anomalies with strong anomaly in El Niño and weak anomaly in La Niña during DJF. But the prediction shows more symmetric pattern of anomaly between El Niño and La Niña except in the tropical eastern Pacific.

It was found that anomalous precipitation and circulation are predicted better in the El Niño-decaying JJA than in the El Niño-developing JJA (Fig. 12). The averaged PCC is 0.52 for the three selected decaying JJAs and 0.47 for the three developing JJAs. In the El Niño decaying JJA, the precipitation seems to be quite predictable over subtropical and extratropical Asia and North America. The MME prediction realistically captures the dryness over the Philippine Sea and the South Asian monsoon trough and the wetness over the Maritime continent, the equatorial Indian Ocean, East Asia, and western North America but fails to capture the subtle location of the wet-dry boundary over the Indian subcontinent. Further, the MME prediction overestimates the precipitation anomaly over Europe and Africa. During the developing JJA, weak anomalies in both

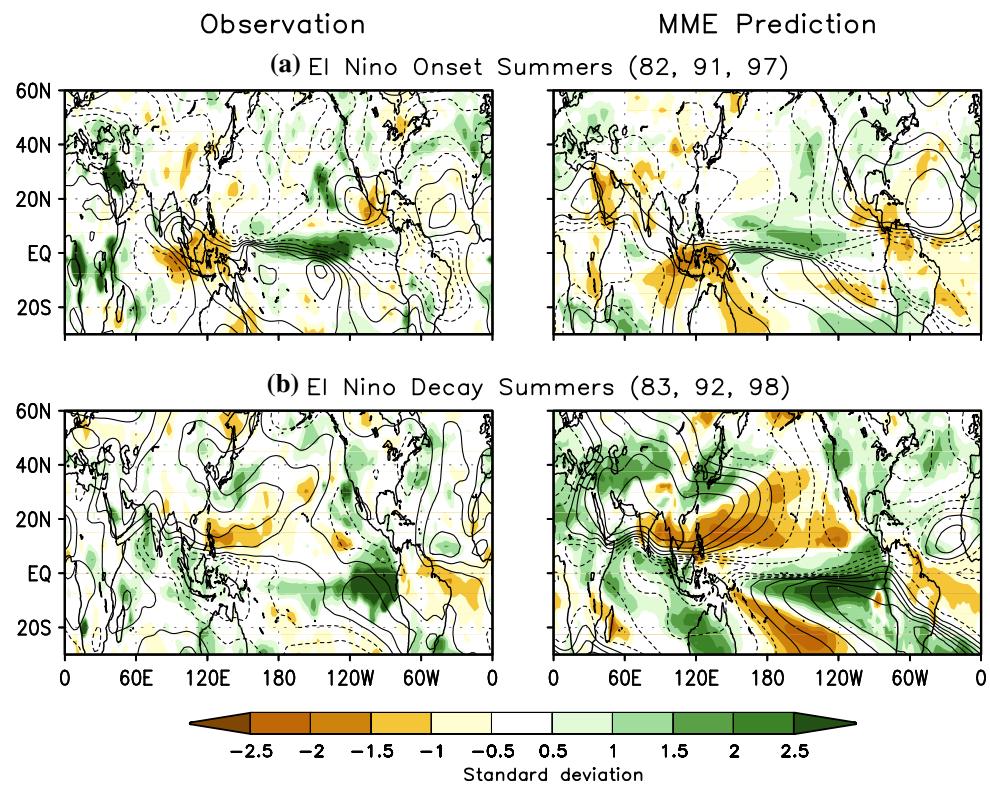
the observation and prediction are seen over subtropical and extratropical Asia. The MME prediction realistically captures the dryness over the Maritime continent, Mexico, and northern East Asia and the wetness over the tropical eastern Pacific and North America but misses the strong wetness over Europe and the equatorial African continent. These findings are in dynamical agreement with the results of Kumar and Hoerling (2003). They showed that a strong asymmetry in the strength of the zonal mean tropical 200 mb height response is stronger in an ENSO-decaying JJA than in the preceding JJA.

6 Probabilistic forecast

6.1 Reliability diagram and BSS

The probabilistic forecast skill was examined for three categorical forecasts using climatological terciles in terms

Fig. 12 Precipitation (*shading*) and 850 hPa stream function (*contoured*) anomalies compositing for **a** three El Niño onset and **b** three El Niño decay JJA seasons. The *left* and *right* panels are made from observation and 1-month lead CliPAS MME prediction, respectively. All anomalies were normalized by their corresponding standard deviation. The contour interval is 0.5



of the BSS and the area under the ROC curve (AROC) using 109 individual realizations from the five-one-tier and five-two tier systems.

Figure 13 shows the reliability diagram, with the forecast probability as abscissa and the observed frequency for the corresponding forecast as ordinate, at each probability bin for the above-normal categorical forecast of 2 m air temperature and precipitation in both JJA and DJF over the global tropics. In general, the reliability curve of MME prediction (dashed line with round dot) is much closer to the diagonal line (dashed-dotted line) than the curves of the individual models. This result suggests that the forecast reliability of the MME prediction is considerably higher than any individual model especially for the cases of very low and very high forecast probability, although the MME probabilistic forecast still tends to overestimate the observed frequency in the case of high forecast probability. Note that in terms of reliability the precipitation forecast is slightly better than the 2 m temperature forecast, and the precipitation prediction is most reliable in the DJF season, as indicated by the reliability term of BSS (0.76).

Although the probabilistic prediction of precipitation using the multi-model system is more reliable than that of temperature, it has poor resolution. The bars in Fig. 13 represent the relative frequency with which the upper tercile was predicted with different levels of probability, the so called sharpness (Palmer et al. 2000). For the precipitation, the probability distribution function is strongly

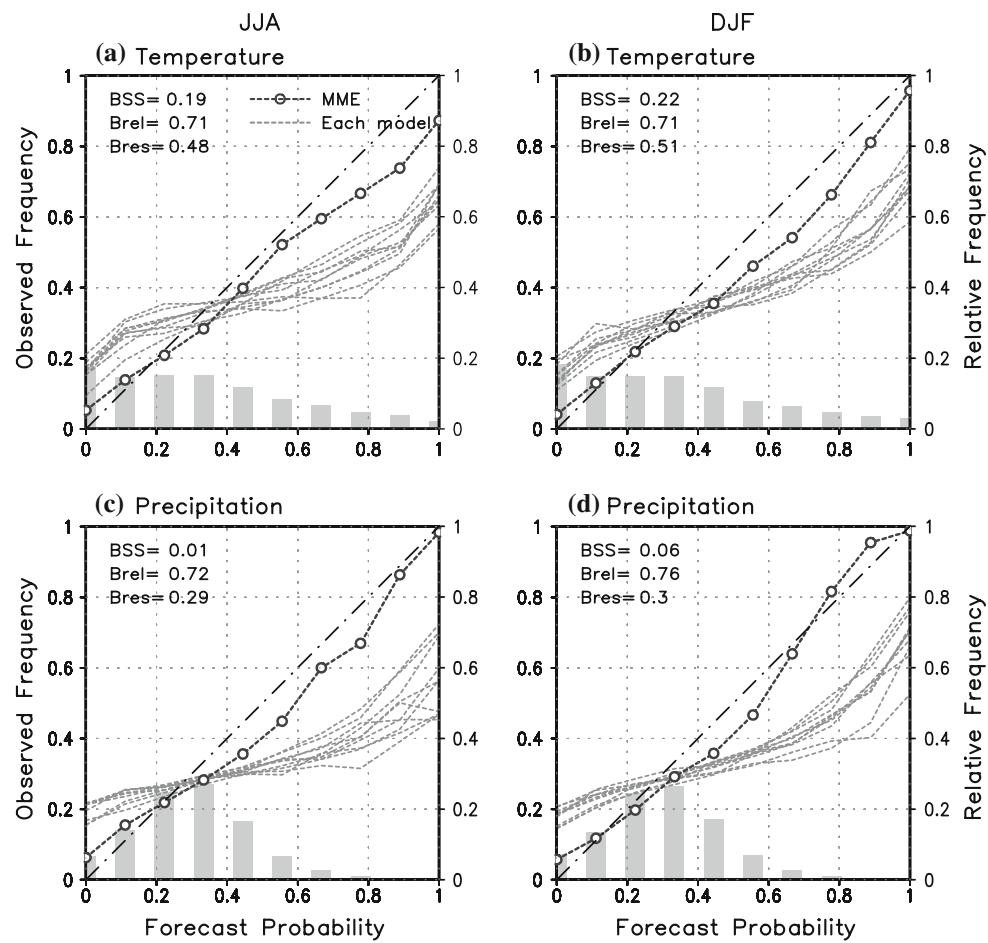
weighted towards the climatological frequency of the upper tercile event. The results indicate that the prediction is not better than a forecast based on the climatology. The resolution terms of BSS for precipitation have low skills that degrade its total BSS. As a result, the BSS of precipitation is 0.01 for JJA and 0.06 for DJF. In terms of BSS, the temperature prediction in DJF has the highest value (0.22), while precipitation prediction in JJA has the lowest score (0.01) (Fig. 13).

The results for the below-normal categorical forecast are very similar to those for the above-normal cases. However, the BSSs for normal cases are below zero for both precipitation and temperature and for both the JJA and DJF seasons.

6.2 Relationship between probabilistic and deterministic forecasts

Is the skill of the multi-model probabilistic forecast related to the MME deterministic forecast? By definition, a value of 0.5 for the area under the ROC curve (AROC) indicates that the hit rate equals the false-alarm rate, and the zero value of BSS indicates that the probabilistic forecast skill is equal to the skill of the forecast based on climatology. Figure 14 shows that (1) the spatial distributions of the BSS and AROC scores agree with each other very well, (2) the spatial patterns of the two probabilistic skill measurements are very similar to the MME TCC scores for

Fig. 13 Reliability diagrams for above-normal categorical forecast of 2 m air temperature (*upper panels*) and precipitation (*lower panels*) over the global tropics in JJA (left panels) and DJF (right panels). The probabilistic forecast was made by CliPAS MME prediction system. The *thick dashed lines with circles* indicate the reliability of multi-model prediction and *thin dashed lines* indicate that of each model prediction. The *bars* represent the forecast sharpness, which is the relative frequency with which the upper tercile was predicted with different levels of probability. The Brier Skill Score (BSS), reliability term of BSS (Brel), and resolution term of BSS (Bres) are also shown in each panel



temperature (Fig. 5b) and precipitation (Fig. 6b), and (3) an AROC score of 0.7 roughly corresponds to a value of 0.1 in BSS and 0.6 in TCC for the deterministic forecast. The results also show that normal events are difficult to predict in both temperature and precipitation, but above-normal and below-normal events can be predicted using the current multi-model prediction system (not shown).

Figure 15 shows the general relationship between the deterministic TCC and the probabilistic BSS and AROC scores. The data examined are DJF forecasts of precipitation at each grid point over the global tropics. Obviously, the relationships are nonlinear, but the relationships tend to be linear when the skill is reasonably high—for instance, when TCC exceeds 0.6, AROC exceeds 0.7, and BSS exceeds 0.1.

7 Effectiveness of MME prediction

7.1 Efficiency of MME prediction

The MME prediction skill was compared to each of the individual model's skill and to the averaged skill of all

individual models by using the PCC-RMSE diagram for 2 m air temperature (Fig. 16) and precipitation (Fig. 17) in JJA over the global tropics and five tropical sub-domains including Africa (0–50°E), the Indian Ocean (50°E–110°E), the western Pacific and the Maritime continent (110°E–180°E), the eastern Pacific (180°W–80°W), and the Atlantic Ocean (80°W–0). All skills were obtained by first computing the PCC score for each year and then making a 24-year time-mean in an unbiased way (Refer to Sect. 4). To quantify the MME's effectiveness, a MME efficiency index (δ) was defined by the non-dimensional distance between the point representing MME skill and the point representing the averaged skill of all individual models in the PCC-RMSE diagram. The larger the efficiency index, the more effective the MME forecast is, compared to the individual models.

It is noted that the mean PCC and the normalized RMSE have a good linear relationship in terms of temperature (Fig. 16) and a significantly weaker linear relationship in terms of precipitation (Fig. 17). This implies that the pattern-related errors are dominant in the temperature prediction, but more random errors are contained in the precipitation prediction in addition to the pattern-related

Fig. 14 Spatial distribution of the Area under ROC (AROC) curve (*upper panels*) and Brier Skill Score (*lower panels*) for three categorical probabilistic forecast of 2 m air temperature (*left panels*) and precipitation (*right panels*) in DJF season

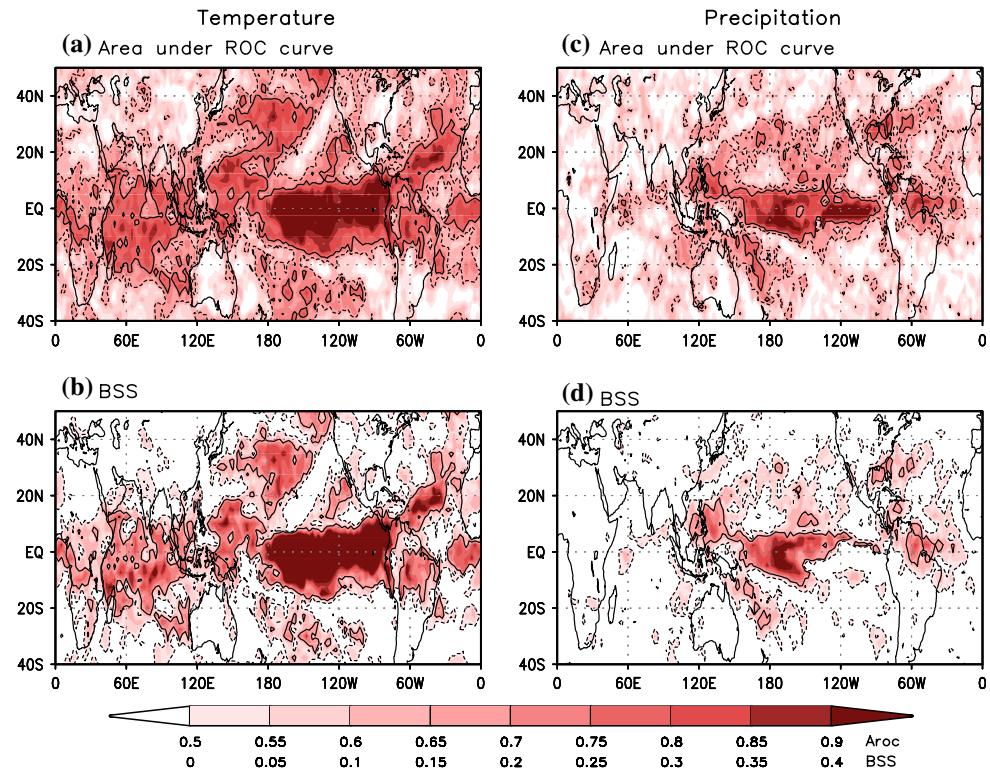
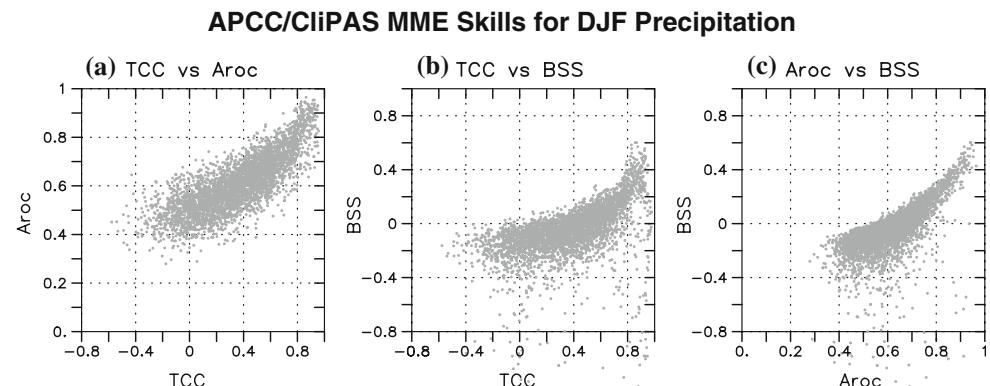


Fig. 15 Scatter diagram of forecast skills of DJF precipitation between **a** TCC and AROC, **b** TCC and BSS, and **c** AROC and BSS at each grid points over the global tropics



error. In general, the linear relationship between the mean PCC and the normalized RMSE strengthens as the mean PCC score increases.

Figures 16 and 17 show that the MME skill is better than any individual model's skill, and it is also better than the all models' average skill over all regions in terms of normalized RMSE. The same is true in terms of mean PCC except over the African and Indian Ocean sectors, where the averaged skills of the individual models are negligibly small (less than 0.1). Over the global tropics, the mean PCC for the MME prediction is 0.53, which is considerably higher than the averaged value of the individual model skills (0.41) for 2 m air temperature. Similarly, for precipitation prediction, the MME skill (0.46) is also

significantly better than the averaged skill of the entire member models (0.31).

The value of the MME efficiency index in the global tropics is 0.88 for 2 m air temperature and 1.32 for precipitation. Although the averaged skill for temperature is higher than for precipitation, the MME prediction for precipitation is more effective than that for temperature. The increased effectiveness is mainly due to the fact that the precipitation predictions among member models have a higher degree of mutual independence. We note that for temperature prediction, there are six-two-tier systems that were driven by the same SNU SST prediction, making the 2 m temperature predictions dependent on each other. For precipitation prediction, on the other hand, different

Fig. 16 The anomalous PCC-RMSE diagram for 2 m temperature prediction in JJA over **a** the global tropics, **b** Africa, **c** Indian Ocean, **d** western Pacific, **e** eastern Pacific, and **f** Atlantic Ocean sectors, respectively. Filled and open red squares represent, respectively, the MME skill and the averaged skill of 14 APCC/CliPAS models. δ indicates the effectiveness index of MME prediction with reference to the averaged skill of all models

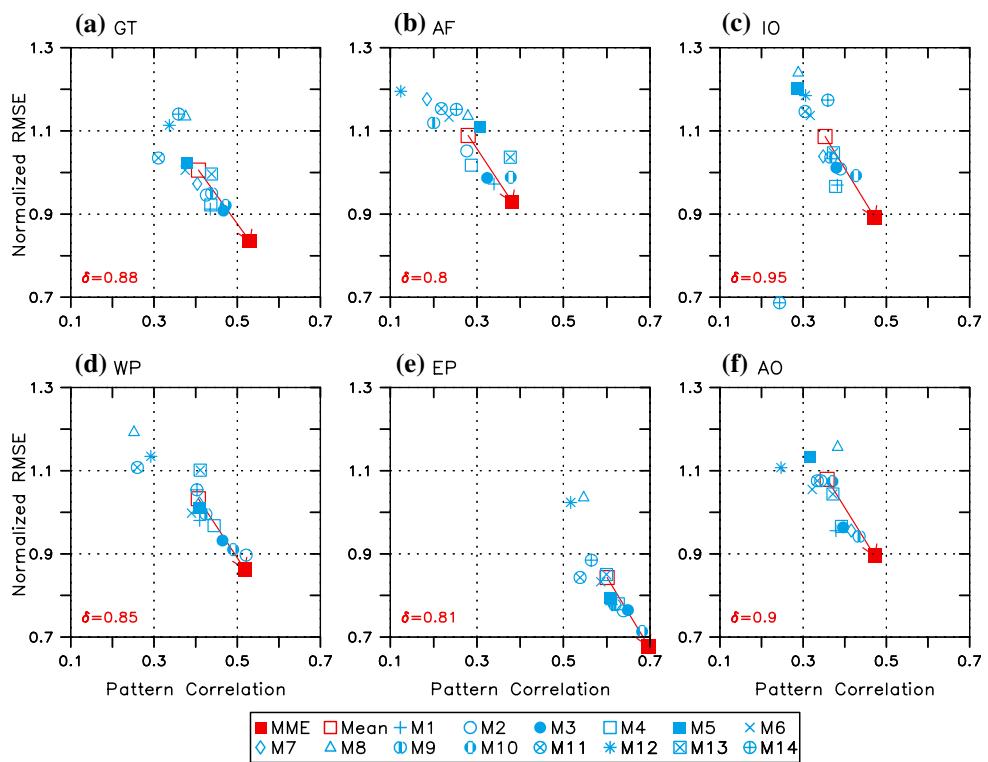
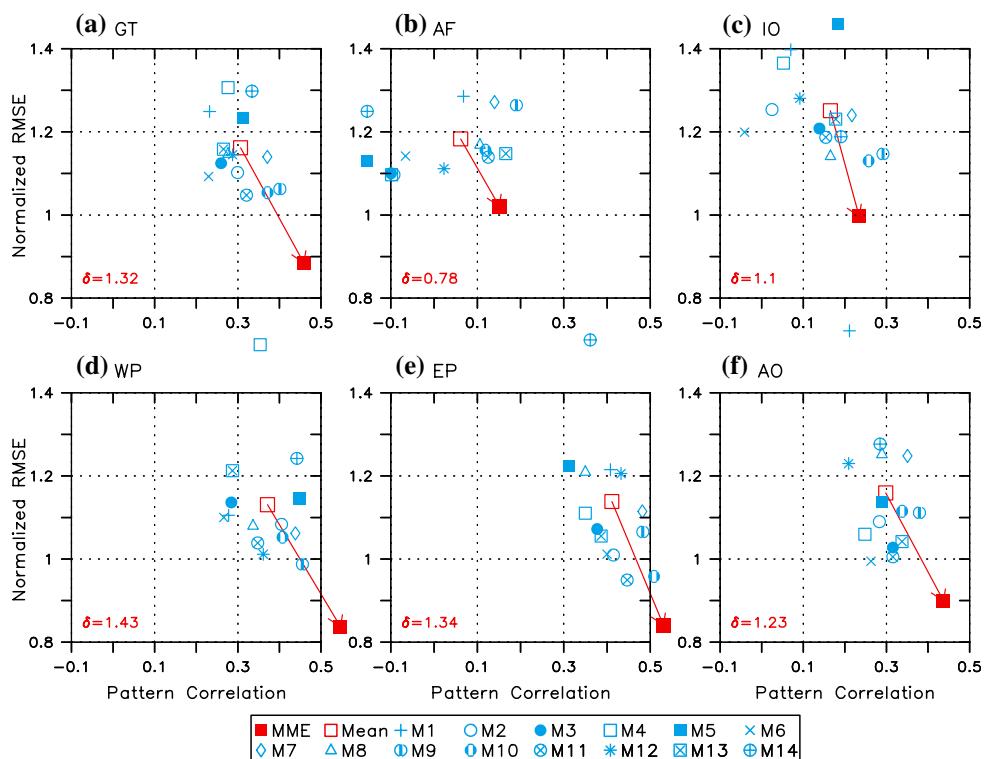


Fig. 17 Same as in Fig. 16 except for JJA precipitation



cumulus parameterization schemes were used in different models, making the precipitation predictions more independent of each other than the corresponding temperature predictions. This result agrees well with that of Yoo and Kang (2005), who pointed out that MME skill depends on

the averaged skill of individual model predictions and their mutual independency.

The most effective region for MME prediction among the five sub-domains is the Indian Ocean for 2 m air temperature and the western Pacific for precipitation, as

shown by the δ values in Figs. 16 and 17. For precipitation prediction, although the individual models have a higher averaged skill over the eastern Pacific than over the western Pacific, the MME prediction shows comparable skills over the two regions. Temperature and precipitation over Africa turn out to be the most difficult to predict. It is also shown that a good performance in temperature prediction doesn't guarantee a good skill in precipitation prediction over the same region. Although temperature forecast skill over the Indian Ocean is comparable to that over the Atlantic Ocean, the precipitation forecast skill is considerably worse. On the contrary, the precipitation skill over the western Pacific is comparable to that over the eastern Pacific, but the temperature skill in the western Pacific is much worse than that in the eastern Pacific. These results suggest that the remote forcing via teleconnection may control the accuracy of precipitation prediction over the Indian Ocean and western Pacific regions rather than local SST forcing in the current MME system.

7.2 Effect of the number of models on MME prediction

One of the open questions concerning the MME prediction is the number of models that should be used to achieve the MME's optimal performance. To address this question, the dependence of the MME correlation skill on the number of member models was examined in terms of the 24-year average of the mean PCC in JJA and DJF over the global tropics. As in Sect. 13, only ten models that have nine or more ensemble runs were selected.

Figure 18 shows how MME skill depends on the number of member models used. At first, the PCC skill increases when the number of models increases, but then saturates after five to six models are used, depending on variable and season. In Fig. 18, the upper (lower) cross mark indicates the skill that would be obtained by using the models having

the best (worst) performance. Interestingly, the combination of the best models doesn't always guarantee the highest MME skill. Similarly, the combination of the worst models may not always yield the lowest skill. Many studies have been carried out to find the optimal combination of MME to improve forecast skill (Krishnamurti et al. 1999, 2000; Kang et al. 2002; Yun et al. 2005; Yoo and Kang 2005; Doblas-Reyes et al. 2005; Kug et al. 2008). The highest MME skill may be achievable by an optimal choice of a subgroup of models, drawing upon an individual model's skill and the mutual independence among the chosen models (Yoo and Kang 2005). However, it is important to mention that the skill of MME prediction possibly depends on the length of the training period. Further discussion for optimal combination of MME can be found in Kug et al. (2008).

Palmer et al. (2004) showed that the largest contribution to the multi-model skill improvement for probabilistic forecast is due to increased reliability. Here, the impact of the number of models used for multi-model probabilistic forecast was also investigated by using the reliability and resolution terms of BSS. In general, the reliability skill increases as the number of models being used increases for the above-normal categorical prediction in JJA (Fig. 19). The probability forecast of precipitation shows a large degree of improvement for the forecast reliability at a modest expense of degraded resolution skill. In contrast, the resolution skill of temperature forecast remains almost the same when the number of models increases. The temperature prediction has better skill than the precipitation prediction, and the DJF season is more predictable than JJA for the multi-model probabilistic forecast. The result of a below-normal case is very similar to that of an above-normal case presented here, and there is practically no skill for a normal event using the current APCC/CliPAS multi-model probabilistic forecast system (not shown).

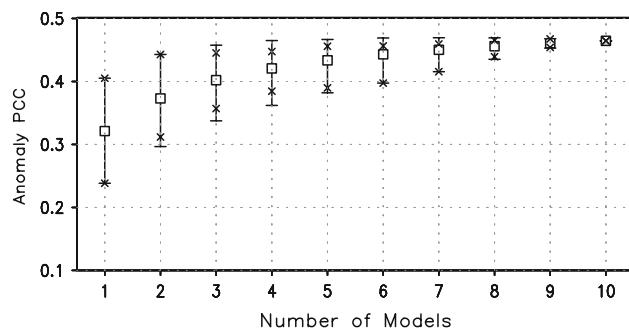


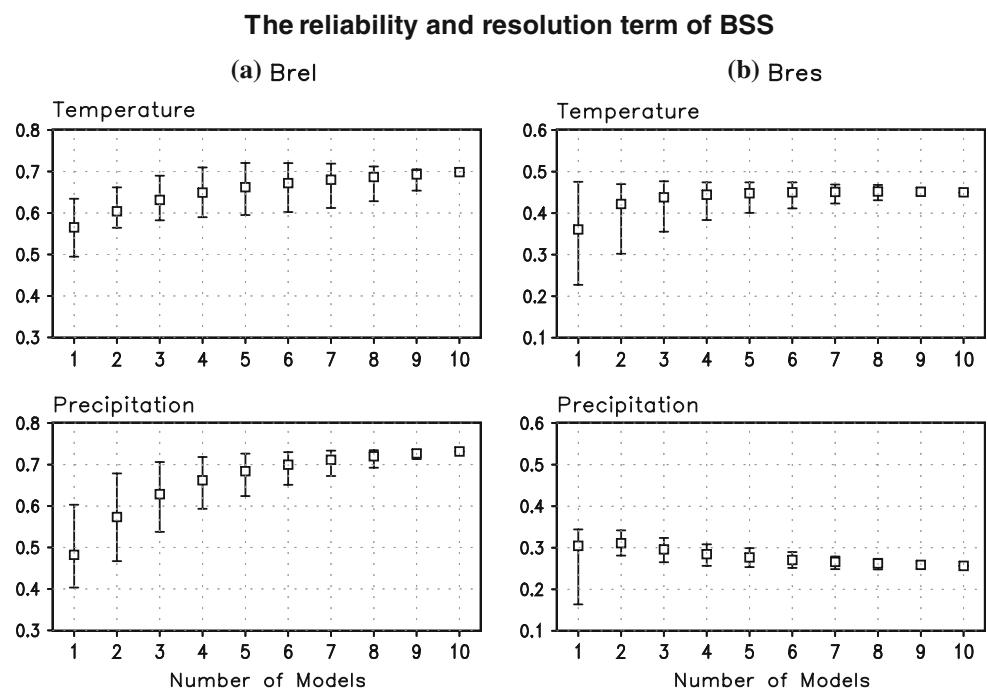
Fig. 18 Dependence of MME correlation skills on the number of models used for 1-month lead JJA precipitation forecasts over the global tropics. The vertical line segments indicate the range of the anomaly pattern correlation coefficients for various combinations of the models. The upper (lower) cross marks denote the skills obtained by selecting the best (worst) models

8 Conclusion and prospectus

In the past two decades, climate scientists have made tremendous advance in understanding the variability and predictability of the earth's climate system. Prediction of seasonal variations and associated uncertainties using multiple dynamical models has become operational. While this is a major breakthrough in the history of numerical weather prediction, state-of-the-art climate prediction is still in its infancy.

One of the purposes of the present study was to assess the state-of-the-art seasonal prediction skills of the multi-model ensemble (MME) mean and probabilistic forecast based on 25-year (1980–2004) retrospective predictions made by 14 climate models that participated in the APCC/

Fig. 19 Range of **a** reliability and **b** resolution terms of Brier skill score for the above-normal categorical forecast of temperature (*upper panels*) and precipitation (*lower panels*), respectively, over global Tropics in JJA using different number of the model being used in APCC/CliPAS predictions. Marks indicate the average value of the each terms of Brier skill score for various combinations of the models composed



CliPAS project. We also evaluated seven DEMETER (Palmer et al. 2004) models' MME for the period of 1981–2001 for comparison. To save space, we have primarily focused on the results derived from the CliPAS models. We should mention that the DEMETER and CliPAS MMEs have comparable skills for both precipitation and 2 m air temperature, although the average skill of the individual CliPAS models is lower than that of the DEMETER models. We also found that the two MME skills show great similarity in spatial structure over the oceans (not shown), suggesting that the two MMEs capture the same predictable part of temperature and precipitation in association with ENSO. The differences in forecast skill over land areas between the two MMEs indicate potentials for further improvement of predictability over these regions.

9 Conclusions

We found that two measures of probabilistic forecast, the BSS and the AROC yield similar spatial distribution of skills, and they are also similar to the spatial pattern of the temporal correlation coefficient (TCC) skill, which is a measure of deterministic MME skill (Fig. 14). While these skills have a nonlinear relationship, an AROC score of 0.7 approximately corresponds to BSS of 0.1 and TCC score of 0.6, and beyond these critical values, they are linearly correlated (Fig. 15). Thus, the spatial distribution of the TCC score also provides valuable information about the spatial distributions of the skill scores for the probabilistic forecast (BSS and AROC).

MME method is demonstrated to be a useful and practical approach for reducing errors and quantifying forecast uncertainty due to model formulation. The MME prediction skill is substantially better than the averaged skill of all individual models. For instance, the TCC skill for Niño 3.4 index forecast at a 6-month lead initiated from May 1 is 0.77 for CliPAS 7-coupled model ensemble, which is significantly higher than the corresponding averaged skill of all individual coupled models (0.63). The MME made by using 14 coupled models from both DEMETER and CliPAS shows an even higher TCC skill of 0.87 (Fig. 3). Over the global tropics (30°S–30°N) and during JJA, the time-mean Pattern Correlation Coefficient (PCC) for MME prediction of 2 m air temperature (0.53) is considerably higher than the averaged skill of the individual models (0.41) (Fig. 16a). Similarly, for precipitation prediction, the MME skill (0.46) is also significantly better than the averaged skill of all member models (0.31) (Fig. 17a). Although the MME does not necessarily outperform the best model during each individual year, the MME's overall skill is superior to any individual model in terms of the time-mean PCC score for all years (Figs. 9, 10). The MME approach shows greater advantages in probabilistic forecast than deterministic forecast. Results of Fig. 19 show that the resolution of BSS score increases from 0.48 (single model's averaged score) to 0.73 when ten models are used. For probabilistic prediction, the largest contribution to MME improvement is due to increased reliability; the resolution score also increases for 2 m temperature but slightly decreases for precipitation forecast (Fig. 19).

The seven CliPAS CGCMs' MME SST forecast skills outperform the SNU statistical-dynamical model's performance and are far better than persistence forecast. The 1-month lead MME prediction for the SST anomalies along the equator can capture, with high fidelity, the spatial-temporal structures of the first two leading empirical orthogonal modes for both the JJA and DJF seasons, which account for 80–90% of the total variance (Figs. 1, 2). The major common deficiencies include a westward phase-shift in SSTA in the central-western Pacific, which leads to significant errors in the western Pacific and may potentially degrade global teleconnection associated with ENSO. The TCC for SST predictions over the equatorial eastern Indian Ocean (EIO) reaches about 0.68 at a 6-month lead forecast, although there is a major dip in skill across August for the forecast initiated in early May, and there is a skill dip in January for the forecast initiated in early November (Fig. 4b). The TCC of SST prediction for the western equatorial Indian Ocean (WIO) is about 0.8 for November initiation due to large persistence but drops below 0.5 at the 4-month lead (Fig. 4a).

While the SST predictions in the WIO and EIO have useful skills, prediction of Indian Ocean Dipole (IOD) index (SST at EIO minus SST at WIO) shows a reduced skill (Fig. 4c). The TCC skills of IOD forecast drop below 0.4 at the 3-month lead forecast for both May and November initial conditions. The results indicate existence of a July prediction barrier and a severe, unrecoverable January prediction barrier for IOD index prediction. However, there is a robust bounce-back for the forecasts initiated from early May, suggesting that mature IOD in October–November is more predictable (if it starts in early May), probably due to the predictability of the EIO pole where SST dominates the mature phase of IOD.

What is the current level of precipitation and temperature prediction skills with the 14 CliPAS models' MME? Here we measured the global-scale MME forecast skill during each season by the Pattern Correlation Coefficient (PCC) between the observed and 1-month lead MME predicted anomaly fields and then make a time-mean PCC over the entire hindcast period in order to quantify the overall MME hindcast skill.

Prediction skills vary by season. For 1-month lead MME seasonal prediction of 2 m air temperature, the mean PCC score over the global tropics (30°S – 30°N) is 0.53 for JJA, which is slightly better than that for DJF (0.50). The higher skill in boreal summer is due to increased persistence. In contrast, the tropical mean PCC score for precipitation in DJF (0.57) is significantly higher than that in JJA (0.46) over the global tropics (Fig. 10). The higher DJF prediction skill is mainly found between the equator and 40°N , and especially in the northern subtropics between 20°N and 40°N (Fig. 7a) from 40°E to 140°E in the Asian–Australian

monsoon sector and from 60°W and 90°W in the tropical American sector (Fig. 7b). The large SST anomalies during the mature phase of ENSO make the DJF precipitation forecast better than during JJA.

Prediction skills highly depend on the strength and phases of ENSO. The MME PCC skills for both temperature and precipitation are well correlated with the amplitude of NIÑO 3.4 SST variation especially in boreal winter with a correlation coefficient 0.75–0.76 (Figs. 9, 10). The performance in El Niño years is better than in La Niña years (Fig. 11). There is virtually no useful skill in the ENSO-neutral years. It is of interest that the anomalous precipitation and circulation were predicted better in ENSO-decaying JJA than in the ENSO-developing JJA (Fig. 12).

In general, prediction of circulation fields shows higher skill than temperature and precipitation predictions. The 200 hPa streamfunction shows very good correlation skill almost everywhere between 40°S and 60°N except in the equatorial region (Fig. 8). The high-skill region in prediction of 850 hPa streamfunction shifts eastward from JJA to DJF. The 500 hPa geopotential height shows high prediction skill confined to the global tropics with a north–south seasonal migration. The DJF skill is considerably higher than JJA for the circulation prediction at all three levels. The season-dependence and the spatial patterns of circulation forecast skills can be well explained in terms of ENSO impacts. The variations in spatial patterns and the seasonality of correlation skills strongly suggest that ENSO variability is the primarily source of the global seasonal prediction skill.

9.1 Prospectus

How do we move forward with seasonal prediction? Two aspects need to be considered. First, given the current levels of the climate models, how do we get the best forecast through MME? Second, from a long-run, what are the priorities we should take in improving our climate models' physics?

The MME deterministic forecast shown in the present study is simple arithmetic average and MME probabilistic forecast is simple democratic counting. Results in Fig. 18 indicate that a combination of the best models doesn't always guarantee highest MME skill; similarly, a combination of the worst models may not always yield the lowest skill. But, this conclusion really depends on models' performance. When individual models have poor skills, such as in the African sector (Fig. 16b), use of the top four models makes a much better MME than all models are taken into account. It is speculated that the highest MME skill may be achievable by an optimal choice of a subgroup of models, drawing upon an individual model's skill and the mutual independence among the chosen models.

Our evaluation confirms that ENSO is the primary source of MME skill for global seasonal prediction. Further, the results of Figs. 9 through 12 suggest that forecast skill for tropical precipitation depends on accurate forecast of the amplitude, spatial patterns, and detailed temporal evolution of ENSO cycle. This is particular true for a long-lead seasonal forecast, because as forecast lead time increases, the model forecast tend to be determined by the model ENSO behavior (Jin et al. 2008).

Therefore, the foremost factor leading to successful seasonal prediction is the model's capability to accurately forecast the amplitude, spatial pattern and detailed temporal evolution of ENSO. Continuing improvement of the one-tier climate model's slow coupled dynamics in reproducing a realistic ENSO mode is a key for long-lead seasonal forecast.

Forecast of monsoon precipitation remains a major challenge. We have showed that seasonal precipitation predictions over land and during local summer have little skill. The TCC for precipitation forecast averaged over the tropics (30°S – 30°N) decreases away from the central-eastern Pacific, with the highest mean skill exceeding 0.5 found near the dateline (150°E – 170°W) in JJA and from 150°E to 140°W in DJF; in contrast, the lowest skill is found over the tropical Africa (Fig. 7b). We speculate that outside of the tropical Pacific, seasonal prediction of monsoon rainfall depends on tropical atmospheric teleconnection associated with ENSO forcing, monsoon-ocean interaction in the Indian and Atlantic Oceans, as well as land-atmosphere interaction. In addition, it has been suggested that correction of inherent bias in the predicted mean states is critical for improving the long-lead seasonal prediction of monsoon precipitation (Lee et al. 2008). The differences in forecast skills over land areas between the CliPAS and DEMETER MMEs indicate potentials for further improvement of predictability over land. The fact that MME has little skill in predicting precipitation over the continental region and during local summer season suggests potential importance of atmosphere-land interaction. Unfortunately, in the current CliPAS models, lack of land surface initialization and use of fixed sea ice makes it impossible to evaluate the impact of atmosphere-land interaction and the atmosphere-ice interaction. There is an urgent need to assess the impact of land surface initialization on the skill of seasonal and monthly forecast using a multi-model framework.

Over mid-latitudes, seasonal rainfall prediction skill shows wavelike patterns in both the southern and northern hemispheres (Fig. 6), suggesting important influences from tropical-extratropical teleconnection and Rossby wave energy propagation. Since the atmospheric teleconnection both within the tropics and between the tropics and extratropics is a major source of predictability for the region

outside of the eastern tropical Pacific, and since teleconnection is sensitive to mean climatology, continuing improvement of the mean state and seasonal cycle as well as statistical behavior of the transient atmospheric circulation in coupled models is also of importance. However, to what extent seasonal predictions depend on nonlinear rectification of high-frequency atmospheric and oceanic processes (so-called "noises") is not well known.

Since the primary memory affecting slow coupled dynamics is stored in ocean subsurface layers and land surfaces, continuing improvement of coupled model initialization is an urgent task. Another direction for future improvement of the seasonal forecast is to increase climate models' resolution. This is absolutely necessary and critical for improvement of prediction of precipitation and statistical behavior of extreme events. However, it remains to be demonstrated whether increased resolution and improved simulation of high-frequency perturbations would improve slow coupled dynamics in coupled climate models. Inclusion of anthropogenic (especially aerosols) and natural forcing (solar, volcanic and aerosol) and a better representation of sea-ice may also benefit accurate seasonal forecast.

Acknowledgments This study is supported by APEC Climate Center (APCC). Wang and Lee acknowledge support from IPRC, which is in part supported by JAMSTEC, NOAA and NASA. This is the SOEST publication number 7543 and IPRC publication number 548.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Alves O, Wang G, Zhong A, Smith N, Tseitkin F, Warren G et al (2003) POAMA: Bureau of meteorology operational coupled model seasonal forecast system. In: Proceedings of national drought forum, Brisbane, April 2003, pp 49–56
- Anderson D, Stockdale T, Balmaseda M et al (2003) Comparison of the ECMWF seasonal forecast system 1 and 2, including the relative performance for the 1997/8 El Niño. Tech. Memo. 404, ECMWF, Reading, UK, 93 pp
- Barnston AG, Mason SJ, Goddard L, Dewitt DG, Zebiak SE (2003) Multimodel ensembling in seasonal climate forecasting at IRI. Bull Am Meteorol Soc 84:1783–1796. doi:[10.1175/BAMS-84-12-1783](https://doi.org/10.1175/BAMS-84-12-1783)
- Bengtsson L, Schlese U, Roeckner E, Latif M, Barnett TP, Graham NE (1993) A two-tiered approach to long-range climate forecasting. Science 261:1027–1029. doi:[10.1126/science.261.5124.1026](https://doi.org/10.1126/science.261.5124.1026)
- Bjerknes J (1969) Atmospheric teleconnections from the equatorial Pacific. Mon Weather Rev 97:163–172. doi:[10.1175/1520-0493\(1969\)097<0163:ATFTEP>2.3.CO;2](https://doi.org/10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2)
- Bowler NE, Arribas A, Mylne KR, Robertson KB, Beare SE (2008) The MOGREPS short-range ensemble prediction system. Q J R Meteorol Soc 134:703–722. doi:[10.1002/qj.234](https://doi.org/10.1002/qj.234)

- Buizza R, Miller M, Palmer TN (1999) Stochastic representation of model uncertainties in the ECMWF EPS. *Q J R Meteorol Soc* 125:2887–2908. doi:[10.1256/smsqj.56005](https://doi.org/10.1256/smsqj.56005)
- Cane MA, Zebiak SE, Dolan SC (1986) Experimental forecasts of El Nino. *Nature* 321:827–832. doi:[10.1038/321827a0](https://doi.org/10.1038/321827a0)
- Charney JG, Shukla J (1981) Predictability of monsoons. Paper presented at the monsoon symposium in New Delhi, India, 1977, and published in Sir Lighthill J, Pearce RP (eds) *Monsoon dynamics*, Cambridge University Press, London
- Cocke S, LaRow TE (2000) Seasonal predictions using a coupled ocean atmospheric regional spectral model. *Mon Weather Rev* 128:689–708. doi:[10.1175/1520-0493\(2000\)128<0689:SPUARS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<0689:SPUARS>2.0.CO;2)
- Davey M, Huddleston M, Sperber KR, Braconnot P, Bryan F et al (2002) A study of coupled model climatology and variability in tropical ocean regions. *Clim Dyn* 18:403–420. doi:[10.1007/s00382-001-0188-6](https://doi.org/10.1007/s00382-001-0188-6)
- Delworth TL et al (2006) GFDL's CM2 global coupled climate models—Part I: Formulation and simulation characteristics. *J Clim* 19:643–674. doi:[10.1175/JCLI3629.1](https://doi.org/10.1175/JCLI3629.1)
- Doblas-Reyes FJ, Deque M, Piedfleuve JP (2000) Multi-model spread and probabilistic seasonal forecasts in PROVOST. *Q J R Meteorol Soc* 126:2069–2088. doi:[10.1256/smsqj.56704](https://doi.org/10.1256/smsqj.56704)
- Doblas-Reyes FJ, Hagedorn R, Palmer TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting—II. Calibration and combination. *Tellus* 57A:223–252
- Fu X, Wang B (2004) The boreal-summer intraseasonal oscillations simulated in a hybrid coupled atmosphere–ocean model. *Mon Weather Rev* 132:2628–2649. doi:[10.1175/MWR2811.1](https://doi.org/10.1175/MWR2811.1)
- GFDL Global Atmospheric Model Development Team (2004) The new GFDL global atmosphere and land model AM2/LM2: evaluation with prescribed SST simulation. *J Clim* 17:4641–4673. doi:[10.1175/JCLI-3223.1](https://doi.org/10.1175/JCLI-3223.1)
- Green DM, Swets JA (1966) *Signal detection theory and psychophysics*. Wiley, New York
- Guan Z, Yamagata T (2003) The unusual summer of 1994 in East Asia: IOD teleconnections. *Geophys Res Lett* 30:1544. doi:[10.1029/2002GL016831](https://doi.org/10.1029/2002GL016831)
- Guilyardi E, Gualdi S, Slingo J, Navarra A, Delecluse P, Cole J et al (2004) Representing El Niño in coupled ocean–atmosphere GCMs: the dominant role of the atmospheric component. *J Clim* 17:4623–4629. doi:[10.1175/JCLI-3260.1](https://doi.org/10.1175/JCLI-3260.1)
- Hagedorn R, Doblas-Reyes FJ, Palmer TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting I. Basic concept. *Tellus* 57A:219–233
- Hoerling MP, Kumar A, Zhong M (1997) El Nino, La Nina and the nonlinearity of their teleconnections. *J Clim* 10:1769–1786. doi:[10.1175/1520-0442\(1997\)010<1769:ENOLNA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<1769:ENOLNA>2.0.CO;2)
- Ji M, Kumar A, Leetmaa A (1994) An experimental coupled forecast system at the National Meteorological Centre. Some early results. *Tellus* 46A:398–418
- Jin EK, Kinter JL III, Wang B et al (2008) Current status of ENSO prediction skill in coupled ocean–atmosphere model. *Clim Dyn.* doi:[10.1007/s00382-008-0397-3](https://doi.org/10.1007/s00382-008-0397-3)
- Kanamitsu M et al (2002) NCEP dynamical seasonal forecast system 2000. *Bull Am Meteorol Soc* 83:1019–1037. doi:[10.1175/1520-0477\(2002\)083<1019:NDSFS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<1019:NDSFS>2.3.CO;2)
- Kang I-S, Jin K et al (2002) Intercomparison of GCM simulated anomalies associated with the 1997–1998 El Nino. *J Clim* 15:2791–2805. doi:[10.1175/1520-0442\(2002\)015<2791:IOAGSA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<2791:IOAGSA>2.0.CO;2)
- Kang I-S, Lee J-Y, Park C-K (2004) Potential predictability of a dynamical seasonal prediction system with systematic error correction. *J Clim* 17:834–844. doi:[10.1175/1520-0442\(2004\)017<0834:PPOSMP>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<0834:PPOSMP>2.0.CO;2)
- Kim H-M, Kang I-S, Wang B, Lee J-Y (2008) Interannual variations of the boreal summer intraseasonal variability predicted by ten atmosphere and ocean coupled model. *Clim Dyn* 30:485–496. doi:[10.1007/S00382-007-0292-3](https://doi.org/10.1007/S00382-007-0292-3)
- Kirtman BP, Zebiak SE (1997) ENSO simulation and prediction with a hybrid coupled model. *Mon Weather Rev* 125:2620–2641. doi:[10.1175/1520-0493\(1997\)125<2620:ESAPWA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<2620:ESAPWA>2.0.CO;2)
- Krishnamurti TN, Kishtawal CM, Zhang Z, LaRow TE, Bachiochi DR et al (1999) Improved weather and seasonal climate forecasts from multi-model superensemble. *Science* 285:1548–1550. doi:[10.1126/science.285.5433.1548](https://doi.org/10.1126/science.285.5433.1548)
- Krishnamurti TN, Kishtawal CM, Shin DW, Williford CE (2000) Multi-model superensemble forecasts for weather and seasonal climate. *J Clim* 13:4196–4216. doi:[10.1175/1520-0442\(2000\)013<4196:MEFFWA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<4196:MEFFWA>2.0.CO;2)
- Kug J-S, Lee J-Y, Kang I-S, Wang B, Park CK (2008) Optimal multi-model ensemble method in seasonal climate prediction. *Asia-Pacific J Atms Sci* 44:233–247
- Kug J-S, Lee J-Y, Kang I-S (2007a) Global sea surface temperature prediction using a multimodel ensemble. *Mon Weather Rev* 135:3239–3248. doi:[10.1175/MWR3458.1](https://doi.org/10.1175/MWR3458.1)
- Kug J-S, Kang I-S, Choi D-H (2007b) Seasonal climate predictability with tier-one and tier-two prediction system. *Clim Dyn.* doi:[10.1007/s00382-007-0264-7](https://doi.org/10.1007/s00382-007-0264-7)
- Kug J-S, Kang I-S, Lee J-Y, Jhun J-G (2004) A statistical approach to Indian Ocean sea surface temperature prediction using a dynamical ENSO prediction. *Geophys Res Lett* 31:L09212. doi:[10.1029/2003GL019209](https://doi.org/10.1029/2003GL019209)
- Kumar KK, Hoerling M, Rajagopalan B (2005) Advancing Indian monsoon rainfall predictions. *Geophys Res Lett* 32:L08704. doi:[10.1029/2004GL021979](https://doi.org/10.1029/2004GL021979)
- Kumar A, Hoerling MP (2003) The nature and causes for the delayed atmospheric response to El Nino. *J Clim* 16:1391–1403. doi:[10.1175/2771.1](https://doi.org/10.1175/2771.1)
- Lau N-C, Wang B (2006) Interaction between the Asian monsoon and the El Nino/Southern Oscillation. In: Wang B (ed) *The Asian monsoon*, Chap 12. Praxis/Springer, Berlin/Heidelberg, pp 479–512
- Latif M, Barnett TP, Cane MA, Flügel M, Graham NE, von Storch H et al (1994) A review of ENSO prediction studies. *Clim Dyn* 9:167–179. doi:[10.1007/BF00208250](https://doi.org/10.1007/BF00208250)
- Latif M et al (2001) ENSIP: The El Nino simulation intercomparison project. *Clim Dyn* 18:255–276. doi:[10.1007/s003820100174](https://doi.org/10.1007/s003820100174)
- Lee JY, Wang B, Kang IS, Shukla J et al (2008) How are seasonal prediction skills related to models' performance on mean state and annual cycle? Accepted to *Clim Dyn*
- Liu P, Wang B, Sperber KR, Li T, Meehl GA (2005) MJO in the NCAR CAM2 with the Tiedke convective scheme. *J Clim* 18:3007–3020. doi:[10.1175/JCLI3458.1](https://doi.org/10.1175/JCLI3458.1)
- Luo J-J, Masson S, Behera S, Shingu S, Yamagata T (2005) Seasonal climate predictability in a coupled OAGCM using a different approach for ensemble forecast. *J Clim* 18:4474–4497. doi:[10.1175/JCLI3526.1](https://doi.org/10.1175/JCLI3526.1)
- Luo J-J, Masson S, Behera S, Yamagata T (2007) Experimental forecasts of the Indian Ocean Dipole using a coupled OAGCM. *J Clim* 20:2178–2190. doi:[10.1175/JCLI4132.1](https://doi.org/10.1175/JCLI4132.1)
- Mason I (1982) A model for assessment of weather forecasts. *Aust Meteorol Mag* 30:291–303
- Mechoso CR, Robertson AW, Barth N, Davey MK, Delecluse P et al (1995) The seasonal cycle over the tropical Pacific in coupled atmosphere ocean general-circulation models. *Mon Weather Rev* 123:2825–2838
- Palmer TN, Brankovic C, Richardson DS (2000) A probability and decision-model analysis of PROBOST seasonal multi-model ensemble integrations. *Q J R Meteorol Soc* 126:2013–2034. doi:[10.1256/smsqj.56702](https://doi.org/10.1256/smsqj.56702)

- Palmer TN, Alessandri A, Andersen U, Cantelaube P, Davey M et al (2004) Development of a European multi-model ensemble system for seasonal to interannual prediction (DEMIETER). *Bull Am Meteorol Soc* 85:853–872. doi:[10.1175/BAMS-85-6-853](https://doi.org/10.1175/BAMS-85-6-853)
- Richardson DS (2000) Skill and economic value of the ECMWF ensemble prediction system. *Q J R Meteorol Soc* 126:649–668. doi:[10.1002/qj.49712656313](https://doi.org/10.1002/qj.49712656313)
- Roeckner E, Arpe K, Bengtsson L, Christoph M, Claussen M, Dumenil L et al (1996) The atmospheric general circulation model ECHAM4: model description and simulation of present-day climate. Max Planck Institut fur Meteorologie, Report No. 218, Hamburg, Germany
- Rosati A, Miyakoda K, Gudgel R (1997) The impact of ocean initial conditions on ENSO forecasting with a coupled model. *Mon Weather Rev* 125:754–772
- Saha S, Nadiga S, Thiaw C, Wang J et al (2006) The NCEP climate forecast system. *J Clim* 19:3483–3517. doi:[10.1175/JCLI3812.1](https://doi.org/10.1175/JCLI3812.1)
- Saji HN, Goswami BN, Vinayachandran PN, Yamagata T (1999) A dipole mode in the tropical Indian Ocean. *Nature* 401:360–363
- Saji HN, Yamagata T (2003) Interference of teleconnection patterns generated from the tropical Indian and Pacific Oceans. *Clim Res* 25:151–169. doi:[10.3354/cr025151](https://doi.org/10.3354/cr025151)
- Schneider EK, DeWitt DG, Rosati A, Kirtman BP, Ji L, Tribbia JJ (2003) Retrospective ENSO forecasts: sensitivity to atmospheric model and ocean resolution. *Mon Weather Rev* 131:3038–3060. doi:[10.1175/1520-0493\(2003\)131<3038:REFSTA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<3038:REFSTA>2.0.CO;2)
- Shukla J (1998) Predictability in the midst of chaos: a scientific basis for climate forecasting. *Science* 282:728–731. doi:[10.1126/science.282.5389.728](https://doi.org/10.1126/science.282.5389.728)
- Shukla J et al (2000) Dynamical seasonal prediction. *Bull Am Meteorol Soc* 81:2493–2606. doi:[10.1175/1520-0477\(2000\)081<2493:DSP>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<2493:DSP>2.3.CO;2)
- Smith TM, Reynolds RW (2004) Improved extended reconstruction of SST (1854–1997). *J Clim* 17:2466–2477. doi:[10.1175/1520-0442\(2004\)017<2466:IEROS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2466:IEROS>2.0.CO;2)
- Shutts G (2005) A kinetic energy backscatter algorithm for use in ensemble prediction system. *Q J R Meteorol Soc* 131:3079–3102. doi:[10.1256/qj.04.106](https://doi.org/10.1256/qj.04.106)
- Vintzileos A, Rienecker MM, Suarez MJ, Miller SK, Pegion PJ, Bacmeister JT (2003) Simulation of the El Niño-Interannual prediction project coupled general circulation model. *CLIVAR Exchanges* 8:25–27
- Vintzileos A, Delecluse P, Sadourny R (1999a) On the mechanisms in a tropical ocean–global atmosphere coupled general circulation model. Part I: mean state and the seasonal cycle. *Clim Dyn* 15:43–62. doi:[10.1007/s003820050267](https://doi.org/10.1007/s003820050267)
- Vintzileos A, Delecluse P, Sadourny R (1999b) On the mechanisms in a tropical ocean–global atmosphere coupled general circulation model. Part II: interannual variability and its relation to the seasonal cycle. *Clim Dyn* 15:63–80. doi:[10.1007/s003820050268](https://doi.org/10.1007/s003820050268)
- Wang B, Kang I-S, Lee J-Y (2004a) Ensemble simulations of Asian–Australian monsoon variability by 11 AGCMs. *J Clim* 17:803–818. doi:[10.1175/1520-0442\(2004\)017<0803:ESOAMV>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<0803:ESOAMV>2.0.CO;2)
- Wang B, Ding QH, Fu XH, Kang I-S, Jin K, Shukla J et al (2005) Fundamental challenge in simulation and prediction of summer monsoon rainfall. *Geophys Res Lett* 32:L15711. doi:[10.1029/2005GL022734](https://doi.org/10.1029/2005GL022734)
- Wang B, Lee JY, Kang IS, Shukla J et al (2008) How accurately do coupled climate models predict the Asian–Australian monsoon interannual variability? *Clim Dyn* 30:605–619. doi:[10.1007/s00382-007-0310-5](https://doi.org/10.1007/s00382-007-0310-5)
- Wang B, Wu R, Fu X (2000) Pacific-East Asia teleconnection: How does ENSO affect East Asian climate? *J Clim* 13:1517–1536. doi:[10.1175/1520-0442\(2000\)013<1517:PEATHD>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1517:PEATHD>2.0.CO;2)
- Wang B, Wu R, Lau K-M (2001) Interannual variability of Asian summer monsoon: contrast between the Indian and western North Pacific-East Asian monsoons. *J Clim* 14:4073–4090. doi:[10.1175/1520-0442\(2001\)014<4073:IVOTAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<4073:IVOTAS>2.0.CO;2)
- Wang B, Wu R, Li T (2003) Atmosphere–warm ocean interaction and its impact on Asian–Australian monsoon variation. *J Clim* 16:1195–1211. doi:[10.1175/1520-0442\(2003\)16<1195:AOIAII>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)16<1195:AOIAII>2.0.CO;2)
- Wang B, Hui W, Zongzhen J, Xin Z, Rucong Y, Yongqiang Y et al (2004b) Design of a new dynamical core for global atmospheric models based on some efficient numerical methods. *Sci China Ser A* 47:4–21. doi:[10.1360/04za0001](https://doi.org/10.1360/04za0001)
- Webster PJ, Moore A, Loschnigg J, Leban M (1999) Coupled ocean–atmosphere dynamics in the Indian Ocean during 1997–1998. *Nature* 401:356–360. doi:[10.1038/43848](https://doi.org/10.1038/43848)
- Weng H, Ashok K, Behera SK, Rao SA, Yamagata T (2006) Impacts of recent El Niño Modoki on droughts/floods in the Pacific Rim during Boreal summer. *Clim Dyn* 29:113–129. doi:[10.1007/s00382-007-0234-0](https://doi.org/10.1007/s00382-007-0234-0)
- Wilks DS (1995) Statistical methods in the atmospheric sciences. Academic Press, Dublin, p 467
- Wu R, Kirtman B (2005) Roles of Indian and Pacific Ocean air–sea coupling in tropical atmospheric variability. *Clim Dyn* 25:155–170. doi:[10.1007/s00382-005-0003-x](https://doi.org/10.1007/s00382-005-0003-x)
- Xie P, Arkin PA (1997) Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull Am Meteorol Soc* 78:2539–2558. doi:[10.1175/1520-0477\(1997\)078<2539:GPAYMA>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2)
- Yoo J-H, Kang I-S (2005) Theoretical examination of a multi-model composite for seasonal prediction. *Geophys Res Lett* 32:L15711. doi:[10.1029/2005GL023513](https://doi.org/10.1029/2005GL023513)
- Yun WT, Stefanova L, Mitra AK, Kumar VV, Dewar W, Krishnamurti TN (2005) A multi-model superensemble algorithm for seasonal climate prediction using DEMETER forecasts. *Tellus* 57A:280–289
- Zhong A, Hendon HH, Alves O (2005) Indian Ocean variability and its association with ENSO in a global coupled model. *J Clim* 18:3634–3649. doi:[10.1175/JCLI3493.1](https://doi.org/10.1175/JCLI3493.1)
- Zhu Y, Toth Z, Wobus R, Richardson D, Mylne K (2002) The economic value of ensemble-based weather forecasts. *Bull Am Meteorol Soc* 83:73–83. doi:[10.1175/1520-0477\(2002\)083<0073:TEVOEB>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<0073:TEVOEB>2.3.CO;2)